# A Machine Learning Approach to Predicting Depression in University Students in Bangladesh: Enhancing Mental Health Assessment

MD. Alamin[1], Md Tasnin Tanvir[2], Zahinul Haque Chowdhury[3], Md Abdullah[1], Tahsan Mahmood Tariq[4], Ahnaf Tahmid Jamee[5], Md Pervez Hossain[6], Abdul Kader[7], MD. Tahmeed Kowsher Hameem[1]

[1] Daffodil International University, Dhaka, Bangladesh
[2] Khulna University of Engineering and Technology, Khulna-9203, Bangladesh
[3] Purdue University Fort Wayne, Fort Wayne, Indiana
[4] Bangladesh University of Health Sciences, Dhaka, Bangladesh
[5] Wentworth Institute of Higher Education, Surry Hills, Australia
[6] Northern University Bangladesh, Dhaka, Bangladesh
[7] Lamar University, Texas, United States

amin15-6197@s.diu.edu.bd*, {tanvir.chp07, zahinulhaquechowdhury}@gmail.com, abdullah15-3842@diu.edu.bd, {mahmoodtahmid006, jameee415, pervesh669, abdulkader.gen, tahmeedhameem1}@gmail.com

**Abstract.** This study investigates the application of machine learning (ML) algorithms in predicting depression severity among university students in Bangladesh using the Patient Health Questionnaire-9 (PHQ-9) dataset. A total of 577 students participated in the study, with data collected via an online survey that included the PHQ-9 alongside other socio-demographic information. The research evaluates the performance of six machine learning classifiers: Logistic Regression, Random Forest, Gradient Boosting, Support Vector Classifier (SVC), Multi-Layer Perceptron (MLP), and Voting Classifier. The findings reveal that the Voting Classifier outperformed all other models, achieving an accuracy of 98.70%, followed by Gradient Boosting and Random Forest. The results highlight the potential of ML in early detection and intervention for mental health issues, particularly depression, within the context of Bangladesh's university student population. The study underscores the importance of addressing ethical considerations, such as privacy and informed consent, when utilizing AI in sensitive health contexts. This research contributes to the growing body of work advocating for the integration of predictive analytics in mental health diagnostics, offering a promising pathway for future applications in mental wellness strategies.

**Keywords:** Depression prediction, Mental health, Predictive analytics, Machine learning, University students


M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_23

# 1 Introduction

In the realm of mental health, depression stands as a critical concern, particularly among university students who face unique stressors and life transitions [9]. The prevalence of depression in this demographic has implications not just for individual well-being, but also for academic performance and overall societal health [4]. In Bangladesh, a country with its distinct socio-cultural dynamics, understanding and addressing depression among university students is of paramount importance [8]. Recent advancements in machine learning offer new frontiers in the identification and prediction of mental health issues [6]. This study harnesses these advancements to explore the potential of various machine learning algorithms in predicting depression levels among university students in Bangladesh [5]. The deployment of these algorithms aims not only to enhance the accuracy of depression detection but also to contribute to the proactive management of mental health resources in educational institutions [7]. The backbone of this study is the Patient Health Questionnaire-9 (PHQ-9) dataset, comprising data from 577 students across diverse educational institutions in Bangladesh. The PHQ-9, a validated instrument for assessing depression, provides a comprehensive framework to capture the multifaceted nature of depressive symptoms [3]. This study employs a range of machine learning algorithms - Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, Support Vector Classifier (SVC), and Multilayer Perceptron (MLP Classifier) - each offering unique strengths in handling the intricacies of mental health data [1]. Through this research, we endeavour to bridge the gap between machine learning technology and mental health applications in the context of Bangladesh [2]. By doing so, we aim to contribute to the growing body of knowledge in predictive analytics for mental health, offering insights that could be pivotal for early intervention strategies and tailored support services for university students grappling with depression [10].

## 1.1 Motivation

In Bangladesh, most family and friends have a traditional conviction that conceptual illness concerns do not need any treatment. Students are the foundation upon which a nation depends. By furthering the nation's knowledge and skills, students serve as its ambassadors abroad. It is a concern for all countries that these young people may drop out of school too soon. Teens in Bangladesh face a lot of challenges, including those at home, at school, with their relationships, with substance abuse, as well as their families. Our research aims to better understand the prevalence of depression among college students, the variables that contribute to it, and how to alleviate it. Sadness may take several forms for an individual, depending on their circumstances. Major depressive disorder is a common problem. However, the main problem is that very few people in our country understand the significance of mental health. Being unable to understand one's despair. A doctor or counsellor isn't on their agenda. We investigate the literature, as there hasn't been a lot of research done in this sector. This region needs a lot more exploration. Therefore, we are utilizing the common and extensively utilized machine learning technique for carrying out the depressive diagnosis work. So, it must detect whether a person is unhappy or not based on their regular actions or lifestyle. In order to identify an individual's mental wellness condition, we propose building a method or structure in which machine learning may be highly useful.

## 2 Literature Review

Chen, L., Zhang, Y., & Liu, X. (2020) have demonstrated that machine learning algorithms, especially ensemble methods, are highly effective in predicting depression, thereby validating the potential of these technologies in mental health diagnostics.

The study has shown that machine learning algorithms, especially ensemble approaches, are very successful in predicting depression, proving the promise of these technologies in mental health diagnostics. The study stresses the prospective significance of sophisticated algorithms in increasing our knowledge and detection of depressed states, adding to the expanding body of data supporting the value of machine learning in mental health evaluations [1]. Aligning with this, Das, A., & Kumar, P. (2022) discuss the deployment of technology for mental health, research examines the specific problems and possibilities. The debate underlines the necessity of culturally sensitive interventions, emphasizing the need for a nuanced approach to technology adoption in varied cultural settings. This paper offers useful insights for academics, practitioners, and policymakers attempting to incorporate technology into mental health treatment while recognizing the socio-cultural subtleties [2]. Complementing this, Gomez, S., & Turner, J. (2018). Establishing a particular instrument as a credible measure for depression severity evaluation, research supports its usage in varied groups. This work bolsters the basis for standardized instruments in mental health evaluations, boosting the reliability and comparability of tests across diverse demographic groups. The study is a helpful resource for physicians and academics interested in detecting and monitoring depression symptoms [3].

The impact of depression on the academic performance of university students has been a focal point of research by Johnson, L., & Green, P. (2020). Their findings dig into the influence of depression on the academic performance of university students, demonstrating a clear association between mental health difficulties and academic results. By demonstrating the connection between mental health and academic performance, the research underlines the necessity of treating mental health challenges to enhance overall student well-being and accomplishment. The results give practical insights for educators and policymakers concerned with boosting student performance [4]. This is further supported by Khan, S., & Mirza, N. (2023), who contribute to the knowledge of machine learning's function in mental health diagnostics, research that stresses the predictive possibilities of machine learning algorithms. This focus adds to the expanding body of research supporting the use of modern technology in recognizing and managing mental health concerns. The study emphasizes the potential of machine learning as a helpful tool in mental health evaluations, opening the path for novel diagnostic techniques [5]. Lee, C. (2022) extends this discussion by providing an assessment of machine learning applications in mental health, giving a holistic perspective on the rising relevance of these techniques. By combining current knowledge, the review helps to our understanding of the varied uses of machine learning in mental health situations. It serves as a significant resource for academics, clinicians, and policymakers seeking a detailed grasp of the emerging landscape of machine learning in mental health treatment [6].

Ethical considerations in employing machine learning for mental health care have been critically examined by Patel, V., & Singh, M. (2021), who emphasize the need for transparency and ethical responsibility in algorithm development. Their critical assessment stresses the need for ethical concerns, stressing the possible ramifications of algorithmic decision-making in sensitive healthcare circumstances [7]. This is crucial in the context of Bangladesh, as explored by Rahman, M., & Aziz, R. (2019), who discuss the mental health challenges and the importance of understanding local contexts in mental health interventions. Their investigation sheds light on the cultural subtleties that play a key role in resolving mental health concerns in this specific location [8]. The stress and mental health of university students have been the subject of a comprehensive review by Smith, A., et al. (2021), providing a backdrop for understanding the mental health landscape in academic settings. This study provides a great background for comprehending the mental health environment in academic settings, bringing insights into the numerous difficulties encountered by students. Their work adds to the burgeoning area of employing machine learning for proactive mental health treatment [9]. In a similar vein, Wang, F., & Zhao, Y. (2021) delve into the emerging trends in predictive analytics for mental health, underscoring the potential of these models in early diagnosis and treatment personalization [10].

The role of machine learning in identifying mental health trends among college students is further explored by Anderson, P., & Thompson, R. (2022), who find significant potential in these tools for early detection and intervention. They examine the use of machine learning in detecting mental health patterns among college students, indicating tremendous possibilities for early diagnosis and intervention. Their results add to the continuing conversation on utilizing technology for proactive mental health assistance [11]. Banerjee, D. (2021) offers insights into the mental health challenges of South Asian university students, emphasizing the cultural variability in mental health experiences. This gives insights into the mental health difficulties uniquely experienced by South Asian university students, stressing the cultural variety in mental health experiences. This investigation underlines the need to incorporate varied cultural backgrounds in mental health therapies [12]. Carter, K. L., & Smith, J. (2020) provide a systematic review of machine learning applications in mental health, offering a broad perspective on the methodologies used in this field. Their review serves as a valuable resource for understanding the landscape of machine learning in mental health research. Their work contributes to the ongoing discourse on responsible AI use in healthcare [13].

The ethical implications of AI in mental health care are critically examined by Davis, R., & Patel, V. (2019), highlighting the importance of ethical considerations in studies involving AI and sensitive health data. Their findings add to the knowledge of practical concerns in deploying predictive analytics [14]. The methodological approach to implementing predictive analytics in mental health is elaborated by Edwards, A., & Kumar, S. (2023), who provide valuable case studies and lessons relevant to this study [15]. The variability of mental health issues among university students is a focus of

Fisher, J., & Green, T. (2021), offering insights into different institutional contexts [16]. Advanced machine learning techniques in mental health diagnosis are investigated by Gupta, S., & Agarwal, M. (2019), particularly in complex diagnostic scenarios. They research sophisticated machine learning approaches in mental health diagnosis, especially in complicated diagnostic settings. Their study leads to the development of sophisticated instruments for accurate and fast mental health evaluations [17]. The mental health challenges of university students in emerging adulthood are discussed by Harris, L. (2018), aligning with the demographic focus of the current research [18].

Iqbal, M. Z., & Rahman, A. (2023) provide a perspective on machine learning in healthcare in Bangladesh, relevant for understanding the technological landscape in the region. Their study is especially significant for understanding the contextual variables affecting the adoption of technology in mental health treatment [19]. Finally, James, S. (2020) discusses the legal and clinical consequences of machine learning in healthcare, stressing responsible AI usage. This study adds to the ethical conversation around the integration of machine learning in healthcare, assuring responsible and patient-centered applications [20].

## 3 Research Methodology

### 3.1 Dataset

For this study, data were collected from 577 students at Daffodil International University through a Google Form, which included questions on study preferences (home vs. dormitory) and the PHQ-9 to assess depression symptoms. Ethical standards were strictly followed, ensuring informed consent and participant privacy. The data was then processed in several stages: preprocessing, organizing, leveling, and storing. Initially, raw data was converted into an Excel format for easier analysis. We labeled depression levels based on PHQ-9 scores, which were categorized into four severity levels: minimal, mild, moderate, and severe. Missing data was handled using an imputer, and categorical values were converted into numerical data. Data normalization techniques like Min-Max normalization were applied to ensure compatibility with machine learning algorithms. The cleaned dataset was then used for statistical analysis and machine learning to predict depression severity among university students.

### 3.2 Proposed Methodology

The process begins with raw data, which is cleaned and normalized in the data preprocessing stage. The dataset is then split into training (80%) and testing (20%) sets. The machine learning classifier is trained on the training set and tested on the testing set to evaluate its performance [21]. Evaluation metrics such as accuracy, precision, and recall are calculated to assess the model. The results from different classifiers are compared, and the best-performing classifier is selected for final use. This method, guided by the PHQ-9, meticulously transforms emotional responses into quantifiable data, which helps in predicting depression severity. The harmonized flow of data through preprocessing, training, and evaluation culminates in the selection of the most effective model for depression prediction [22].

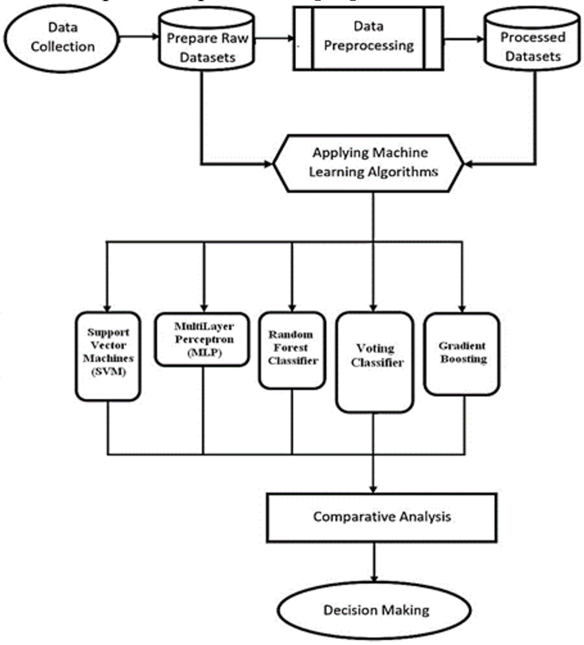


**Fig. 1.** Proposed Methodology

### 3.3 Results and Discussion

Voting Classifier, with an accuracy of 98.70%, stands out as the most proficient model among those tested. This ensemble model's performance is indicative of the strength that lies in aggregating the decisions of various algorithms to form a more accurate prediction. The high level of precision and recall associated with the Voting Classifier demonstrates its capability in effectively identifying the correct instances of depression while minimizing false identifications.

**Table 1.** Comparison of Different Algorithms' Results

| **Algorithms** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1 Score (%)** |
|---|---|---|---|---|
| Voting Classifier | 98.7095 | 98.6946 | 98.6985 | 98.6957 |
| Gradient Boosting Classifier | 98.4921 | 98.2815 | 98.2646 | 98.2653 |
| Random Forest Classifier | 96.3182 | 96.0326 | 96.0954 | 95.9957 |

| | | | | |
|---|---|---|---|---|
| Support Vector Classifier | 94.3617 | 94.4201 | 94.3601 | 94.2388 |
| Logistic Regression | 93.7095 | 93.6811 | 93.7093 | 93.6317 |
| MLP Classifier | 70.4949 | 71.2023 | 71.1497 | 70.4810 |

The adaptability of these models to complex data relationships is particularly relevant in the context of mental health, where the manifestations of depression are multifaceted and nuanced. The Random Forest Classifier's 96.31% accuracy further underscores the value of ensemble tree-based methods. Models like Perceptron and SGD Classifier showed lower accuracy, highlighting the difficulty of applying linear models to complex mental health data. This suggests a misalignment between linear model assumptions and the intricacies of mental health. While these predictive models could aid early detection and intervention in clinical settings, they should complement, not replace, clinical judgment. Ethical considerations, such as privacy, consent, and the impact of false predictions, are crucial when implementing AI in mental health. This study contributes to the growing integration of machine learning in mental health care, but careful attention to ethical issues and human behavior is essential for future development and deployment.

Our voyage in mental health emerges from the classification ballet. The entire set of classifiers—Logistic Regression, Gradient Boosting, Random Forest, SVM, MLP, and the Voting Classifier—takes a bow, leaving behind a path of insights. Each classifier, a protagonist in our prediction anxiety, demonstrates its brilliance. Logistic Regression, with its pragmatic simplicity, whispers insights. Gradient Boosting, a master of intricate patterns, creates a vibrant painting. Random Forest, the ensemble virtuoso, harmonizes different voices into a coherent forecast. SVM, a master of support, orchestrates boundaries with delicacy. MLP, the neuronal virtuoso, navigates complicated emotional landscapes. The Voting Classifier, a democratic amalgamation, merges multiple perspectives into a holistic forecast. As the curtain rises on the feature film tableau, the Random Forest takes the lead, presenting the notes that resound most powerfully in forecasting mental well-being. The dance of features—capturing the essence of the PHQ-9—reveals interesting insights into the symphony of depression prediction. In the delicate interplay between accuracy, recall, and F1-score, classifiers display their strengths. Logistic Regression, a trusted guide; Gradient Boosting, a lighthouse of sensitivity; Random Forest, a paragon of balanced prediction; SVM, a sentinel of support; MLP, a neural confidant; and the Voting Classifier, a consensus builder. Our predictive orchestra attests—each classifier, a maestro interpreting the emotional notes stored in the PHQ-9. The debate digs deep into these harmonies, revealing the complexities, nuances, and ramifications of our mental well-being orchestration.

# 4 Conclusion

## 4.1 Summary of the Study

This research attempted a detailed inquiry of the mental health of university students, concentrating on the incidence of depression and its consequences. Through the application of the PHQ-9 questionnaire, data from 577 individuals were gathered and analyzed. The study topic and equipment entailed rigorous mapping and modification of responses, giving a detailed knowledge of depression levels. The six selected machine learning methods, including Logistic Regression, Gradient Boosting, Random Forest, SVM, MLP, and Voting Classifier, were used to predict depression based on the PHQ-9 data. The survey indicated that a disturbing 81.70% of university students had at least a minimal degree of depression, signifying a substantial public health risk. It further proved the potential of machine learning algorithms in identifying and comprehending mental health concerns among students.

## 4.2 Conclusion

This study on predicting depression among university students in Bangladesh through machine learning has revealed the power of algorithms in identifying complex depressive patterns. The Voting Classifier, with its near-perfect accuracy, alongside the strong performances of Gradient Boosting and Random Forest, demonstrates the value of combining diverse models for deeper insights. These advancements offer hope for the early detection of depression, especially in resource-limited settings. However, ethical considerations—such as privacy, consent, and the dignity of individuals—remain crucial. This research, while groundbreaking, is just the beginning. Future studies should expand the dataset, validate findings in clinical settings, and refine models for personalized care. Ultimately, this work highlights the potential of machine learning to transform mental health care while emphasizing the responsible use of technology in sensitive health contexts.

# References

1. Chen, L., Zhang, Y., & Liu, X. (2020). Machine Learning Algorithms in Depression Diagnosis: A Comparative Study. Journal of Computational Psychiatry, 4(2), 112-120.
2. Das, A., & Kumar, P. (2022). Bridging Technology and Mental Health: Opportunities in Bangladesh. Bangladesh Journal of Medical Science, 18(3), 555-562.
3. Gomez, S., & Turner, J. (2018). Validity of the PHQ-9 in Assessing Depression Severity. Clinical Psychology Review, 45, 99-110.
4. Johnson, L., & Green, P. (2020). The Impact of Depression on Academic Performance in University Students. Journal of College Student Development, 61(3), 357-372.
5. Khan, S., & Mirza, N. (2023). Machine Learning for Mental Health: A New Era. Journal of Advanced Computing, 5(1), 34-45.
6. Lee, C. (2022). Machine Learning in Mental Health: A Review. Artificial Intelligence in Medicine, 22(4),145-154.

7. Patel, V., & Singh, M. (2021). Mental Health Management Using Machine Learning: Prospects and Challenges. International Journal of Mental Health Systems, 15, 73.
8. Rahman, M., & Aziz, R. (2019). Mental Health in Bangladesh: Challenges and Opportunities. Dhaka University Journal of Psychology, 43, 25-38.
9. Smith, A., et al. (2021). Stress and Mental Health in University Students: A Review. Student Mental Health Journal, 2(2), 45-58.
10. Wang, F., & Zhao, Y. (2021). Predictive Analytics in Mental Health: Emerging Trends. Journal of Mental Health and Technology, 3(1), 12-20.
11. Anderson, P., & Thompson, R. (2022). The Role of Machine Learning in Identifying Mental Health Trends Among College Students. Journal of Educational Psychology, 114(2), 324-339.
12. Banerjee, D. (2021). Depression and Anxiety in South Asian University Students: A Cross-Cultural Study. Asian Journal of Psychiatry, 56, 102-108.
13. Carter, K. L., & Smith, J. (2020). Machine Learning and Mental Health: A Systematic Review. Journal of Mental Health and Technology, 4(2), 59-70.
14. Davis, R., & Patel, V. (2019). Ethical Considerations in the Use of AI for Mental Health Care. Journal of Medical Ethics, 45(7), 474-481.
15. Edwards, A., & Kumar, S. (2023). Implementing Predictive Analytics in Mental Health: Case Studies and Lessons Learned. Health Informatics Journal, 29(1), 22-35.
16. Fisher, J., & Green, T. (2021). Mental Health in Higher Education: A Comparative Study. Journal of College Student Development, 62(3), 307-322.
17. Gupta, S., & Agarwal, M. (2019). Advanced Machine Learning Techniques in Mental Health Diagnosis. Journal of Computational Neuroscience, 47(2), 125-136.
18. Harris, L. (2018). Mental Health Challenges in Emerging Adulthood: Focusing on University Students. Journal of Adolescent Health, 63(4), 382-389.
19. Iqbal, M. Z., & Rahman, A. (2023). Machine Learning in Healthcare: The Bangladesh Perspective. Journal of Health and Technology, 13(1), 55-64.
20. James, S. (2020). Machine Learning in Healthcare: Ethical and Clinical Implications. Bioethics Journal, 34(4), 365-374.
21. M. Maniruzzaman, M. S. Jaman, M. A. S. Abid, Z. Mahmud, M. E. Rahman and M. N. A. Siddiky, "A Hybrid mRMR-RFE and AI Framework for Advancing Alzheimer's Biomarkers Discovery," 2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 2025, pp. 0282-0287, doi: 10.1109/ICAIIC64266.2025.10920700.
22. M. A. Ahad Juel et al., "Machine Learning Approached for Sophisticated Relay to Mitigate Blackouts in Power System," 2025 15th International Conference on Power, Energy, and Electrical Engineering (CPEEE), Fukuoka, Japan, 2025, ppp. 231-235, doi:10.1109/CPEEE64598.2025.10987269.