Received: 27 December 2025 | Revised: 18 March 2026 | Accepted: 8 May 2026 | Published online: 21 May 2026

RESEARCH ARTICLE

Medinformatics
2026, Vol. 00(00) 1–11
DOI: 10.47852/bonviewMEDIN62028924

BON VIEW PUBLISHING

Explainable AI-Driven Identification of Depression Levels for Early Mental Health Intervention in IT Students

M.d Rakeen Islam Nahin[1], Safiul Haque Chowdhury[2,*], Mohammed Ibrahim Hussain[2], Mohammad Minoar Hossain[3] and Mohammad Mamun[2,4]

[1] *Department of Computer Science and Engineering, Military Institute of Science and Technology, Bangladesh*

[2] *Department of Computer Science and Engineering, Jahangirnagar University, Bangladesh*

[3] *Department of Electrical Engineering and Computer Science, Florida Atlantic University, USA*

[4] *Department of Computer Science and Engineering, Bangladesh University, Bangladesh*

Abstract: In today's rapidly evolving technological era, the IT industry has emerged as one of the most in-demand sectors among students. Students in IT faculties face significant academic pressure, leading to an increase in mental health problems. This research is related to the prediction of IT students' mental health status by means of machine learning (ML) approaches and explainableAI for early prevention and diagnosis. The proposed approach involves preprocessing the dataset, circumscribing various parameters of IT students and IT education. The primary objective of this research is to determine the factors that have a comparatively greater influence on the depression of students. Upon employing several ML models for outcome explanation, such as Random Forest Regressor, Gradient Boosting Regressor, AdaBoost Regressor, Linear Regression (LR), and Histogram-based Gradient Boosting Regressor, the best result was obtained from the LR with R^2 = 0.581867, mean squared error = 0.025011, and mean absolute error = 0.1197, which was later cross-verified using 10-fold results. After the cross-validation, LR obtained the mean R^2 value of 0.6008. Additionally, applying K-means clustering, a positive relationship was visualized between anxiety and depression, suggesting a proportional relationship between them. Lastly, Local Interpretable Model-agnostic Explanations and Shapley Additive Explanations were used for the outcome explanation to provide insights into the decisions of the models. These findings highlight the important aspects that influence the depression level of IT students, so that proper initiatives can be taken to improve the overall mental condition.

Keywords: machine learning, regression, explainable artificial intelligence, mental health, depression

1. Introduction

Globally, approximately 5% of graduates from tertiary education are from an IT background [1]. Among South Asian countries, India alone produces 550,000 graduates each year, making it the highest in the world [2]. It is worth mentioning that the number is increasing over the years. In Pakistan, more than 1 million students are enrolled in STEM programs [3]. In Bangladesh, the IT industry currently supports more than 1 million jobs [4]. Despite the sector's global advancement, concerns have emerged regarding their mental well-being. Statistics show that more than 75% of Bangladeshi university students experienced mental health problems due to academic pressure in the post-COVID-19 period; 34% reported suicidal thoughts, and 2.44% had attempted suicide [5]. The bar chart in Figure 1 visualizes the statistics on mental health problems among Bangladeshi students.

Systematic reviews show high levels of anxiety and depression among computer science and computing students across the education pipeline, including undergraduate and doctoral levels [6]. In the current situation, students' mental health in the education sector is an exacerbated issue due to academic stress, financial worries, and prospects of their future. Furthermore, the area that is most in demand for the (tech) industry has become one of the most preferred fields for students and their parents. Depression among IT students is a spectrum, and it can have a negative effect on their studies and their general condition. Early and precise prediction of depression severity is critical for timely support and intervention of depression.

Over the years, research has been carried out to identify the main cause of mental health issues among individuals. Madububambachu's study [7] focused on reviewing the machine learning (ML) techniques that were used for predicting mental

***Corresponding author:** Safiul Haque Chowdhury, Department of Computer Science and Engineering, Jahangirnagar University, Bangladesh. Email: 4169safiulhaque@juniv.edu

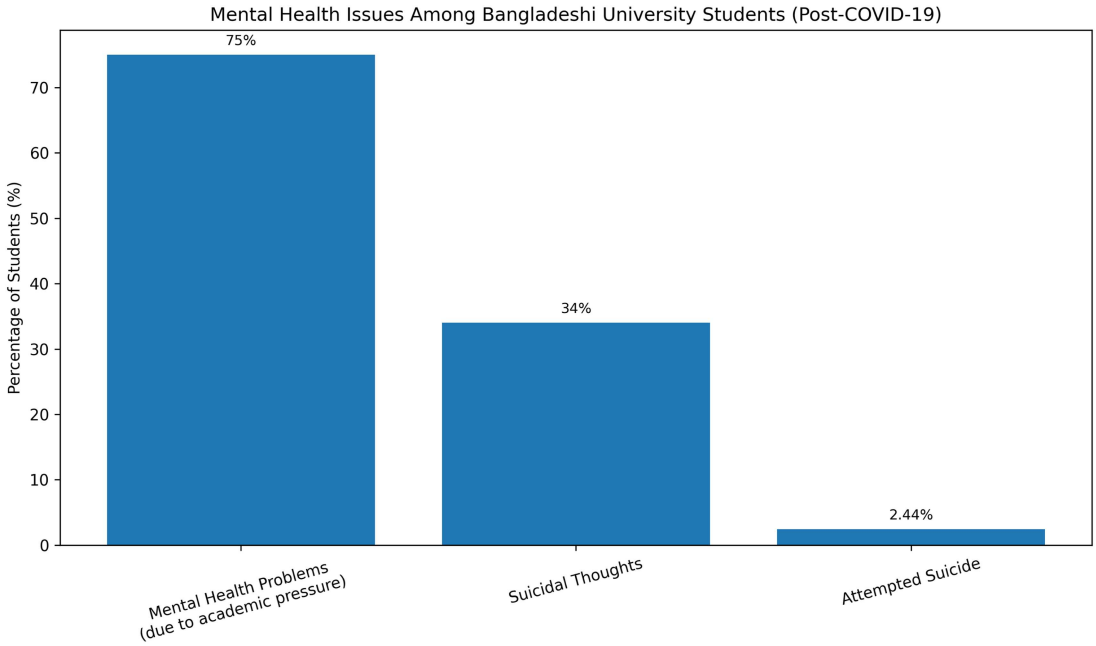

Figure 1. Mental health issues among Bangladeshi University students (post-COVID-19 period)

health diagnoses; moreover, his paper highlighted the commonly used algorithms, datasets, and performance challenges. Mutalib et al. [8] aimed to develop and evaluate the ML-based mental health prediction models specifically within higher education institutions to support early student intervention. Besides, Abdur Rahman's paper emphasized assessing mental well-being from a psychological and clinical perspective by examining the key indicators of mental health [9]. Furthermore, Pathak's review [10] investigated the pervasive effects of psychological distress on the general public and caregivers, emphasizing social, emotional, and caregiving burdens that were associated with mental illness. This narrative review considers a focused set of recent studies (2020–2025) on mental health analysis and prediction, considering both ML-based approaches and psychological assessment perspectives. Moreover, the review emphasizes the mental-healthcare sector and predictive modeling techniques to enhance interpretability. Although a significant number of studies addressed the issue of mental health, explicit research regarding mental health among IT students is still limited. Additionally, the implementation of explainable AI (XAI) provides insights for parents, teachers, students, and mental health practitioners, which is still sparse.

Motivated by the need for early identification and intervention of mental health conditions among IT students, the proposed system develops a method with a dataset of 1977 students encompassing 39 features in the IT sector. The dataset went through initial preprocessing and feature selection, which reduced the student entities from 1977 to 1906 and features from 39 to 7, making the dataset more compatible and lighter for further analysis and explanations. Several models were utilized; among them, Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), AdaBoost Regressor (ABR), Linear Regression (LR), and the Histogram-based Gradient Boosting Regressor (HGBR) were significant, with LR performing the best. Interpretable AI methods, such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP), demonstrated that anxiety is the most important predictor for patients' level of depression, followed by stress value, current CGPA, age, and gender.

Major objectives of this research can be summarized as follows:

- Depression severity level prediction using ML models.
- Identifying the most accurate regression model (RFR, GBR, ABR, LR, HGBR) upon comparison.
- Identification of key mental health indicators by applying XAI (LIME, SHAP).

Scope: The study focuses on predicting mental health conditions of IT students using supervised ML regression models.

2. Materials and Methods

In recent years, various ML techniques have been developed to predict and analyze mental health conditions of students. Madububambachu et al. [7] applied regression models for the prediction of mental health effects, without using XAI. S Mutalib et al. [8] undertook an extensive survey of mental health models looking at their conceptual frameworks but used only logistic regression. Abdul Rahman et al. [9] analyzed mental well-being via surveys and psychological assessments, with measurement rather than predictive modeling being the focus. Pathak et al. [10] discussed the use of ML for mental health more generally, which focused on psychological distress in the general public. The papers mainly focused on college and university students. However, studies specifically targeting IT students were absent. The proposed method overcomes the limitation of predicting mental health conditions of IT students and analyzes the performance using a regression method. Table 1 presents a comparative illustration of the reviewed studies.

2.1. Methodology

The main aim of this study was to examine and evaluate several socio-economic and academic variables on the mental health of IT students. Prediction using ML algorithms and the application of XAI are the primary focus in the overall research framework. The research focuses on predicting the depression level of IT students based on various socio-economic factors.

Table 1. Comparative illustration of literature review

Author(s)	Focus area	Method used	Key findings	Target population
Madububambachu et al. [7]	Prediction of mental health outcomes	Systematic literature review	Summarized ML-based prediction techniques; highlighted regression (logistic) and other ML methods for mental health classification	General students
Mutalib et al. [8]	Mental health prediction models in higher education	ML modeling (Decision Tree, SVM, Neural Network, Naïve Bayes, logistic regression)	Applied multiple ML classifiers to predict stress, depression, and anxiety levels; logistic regression used for classification	Higher education students
Abdul Rahman et al. [9]	Assessment of well-being	Surveys and psychological assessments	Measured mental well-being; focused on assessment rather than predictive modeling	General students
Pathak [10]	Psychological distress in caregivers and general public	Literature review/general discussion of ML applications	Provided broad overview of psychological distress and potential ML approaches for prediction	General public and caregivers of persons with mental illness

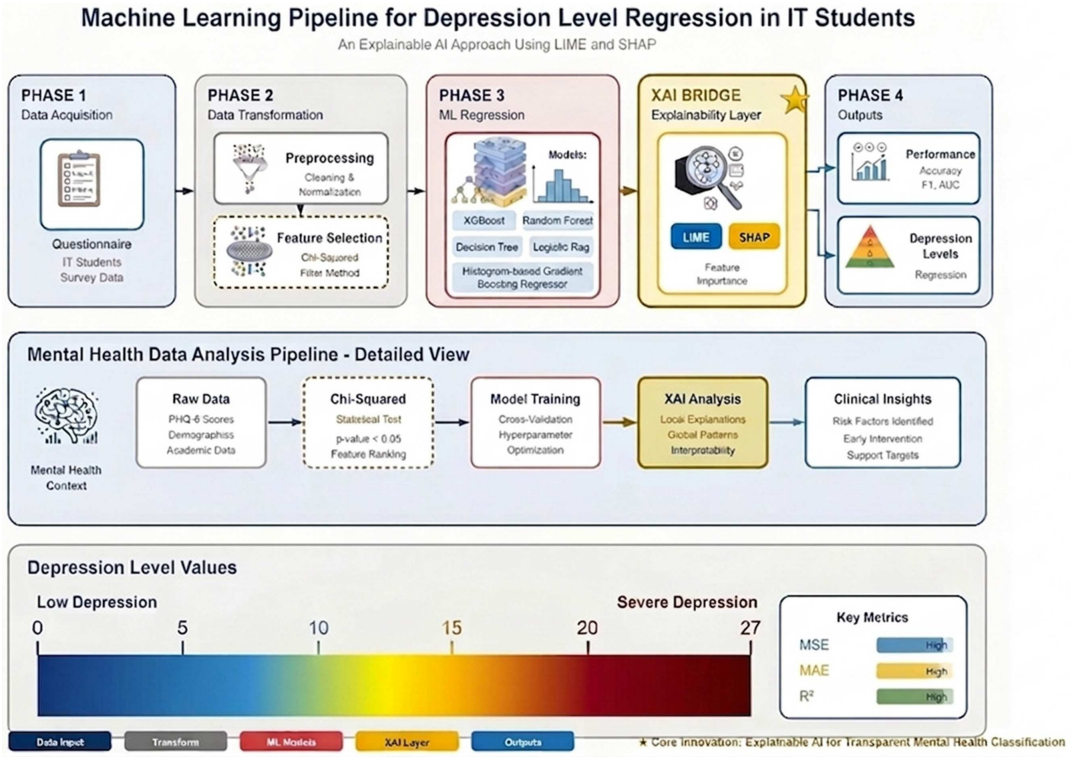

Figure 2. Detailed overview of the proposed research framework illustrating the machine learning pipeline and explainable AI integration for depression level regression in IT students

The steps include the collection of data from Kaggle, performing data analysis for visualization, and preprocessing the data to ensure compatibility for model training. A brief description of the research architecture is shown in Figure 2.

The figure shows the steps that are included in the data analysis. The steps include data collection, visual analysis, dataset preprocessing, model construction, training and testing, evaluation and selection, and lastly, outcome explanation.

2.1.1. Dataset

The dataset, which was used in this study, is publicly available on Kaggle under the title "University Students Mental Health" [11]. An external dataset used for validation is also publicly accessible on Kaggle under the title "Student Mental Health Survey" [12]. Personal identifiers were not accessed. Furthermore, the data were used in accordance with the platform's usage policies. The primary dataset consists of 1977 data samples and 39 features.

Among all features, depression is used as the target variable. Table 2 provides a concise summary of the dataset used.

2.1.2. *Data examination and visual analysis*

Among all the features, age, anxiety value, stress value, depression value, current CGPA, and gender (male and female) were finally selected for the data analysis as all non-numeric survey questions were excluded. Subsequently, the features are scaled using min-max scaling.

$$\text{Xscaled} = (\text{X} - \text{X min})/(\text{X max} - \text{X min}) \quad (1)$$

Visual analysis was performed to analyze the distribution of data. Visual analysis covered illustration through violin plot, box plot, and heat map. Figure 3 is a neat visualization for the distribution of data across the features using a violin plot and a box plot. A correlation heat map is a graphical representation of the correlation matrix. Correlation value varies from −1 (strong negative) to +1 (strong positive), with 0 indicating no correlation [13]. Figure 4 shows the correlation heat map.

2.1.3. *Data preprocessing*

In data preprocessing, the primary concern was to prepare the data to apply ML models. The data preprocessing phase included dropping incompatible columns like recreational activities that did not have binary classification, and due to overlapping, one-hot encoding was also not found to be effective. In the features like universities, major subjects, and sports engagement activities, one-hot encoding was applied. Boolean values were converted to binary, and a small number of data points were eliminated due to missing values. Finally, to decrease dimensionality and considering the feature importance with respect to depression value, 7 features were selected along with the depression value for the subsequent steps. The significant features are shown in Figure 5. The figure represents all the six features along with their importance in the dataset, keeping the depression value as the target column. The important features are explained in Table 3.

2.1.4. *Model construction*

In the present study, regression equations were developed to predict the continuous dependent variable. To make predictions, both linear and ensemble approaches were used. These models

Table 2. Concise summary of the dataset used for analysis

Category	Feature type	Description	No. of features
Demographic information	Age, gender, university, department, academic year, CGPA, scholarship/waiver	Basic background information about the student	7
Anxiety indicators	Anxiety Q1–Q7	Questions measuring academic anxiety symptoms during a semester	7
Anxiety outcome	Anxiety value, anxiety label	Numerical anxiety score and severity category	2
Stress indicators	Stress Q1–Q10	Questions measuring academic stress and coping ability	10
Stress outcome	Stress value, stress label	Calculated stress score and classification	2
Depression indicators	Depression Q1–Q9	Questions assessing depressive symptoms (interest, mood, sleep, energy, etc.)	9
Depression outcome	Depression value, depression label	Final depression score and severity level	2

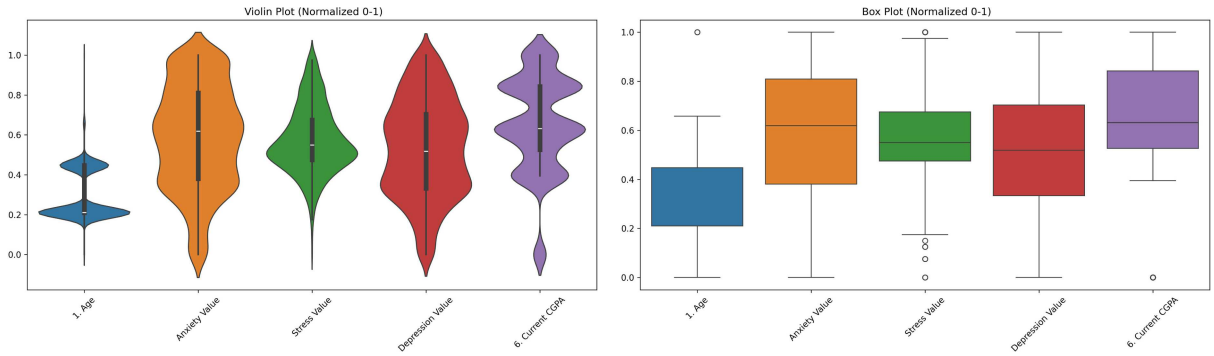

Figure 3. Violin plot and box plot illustrating the concentration and outlier range of dataset, respectively

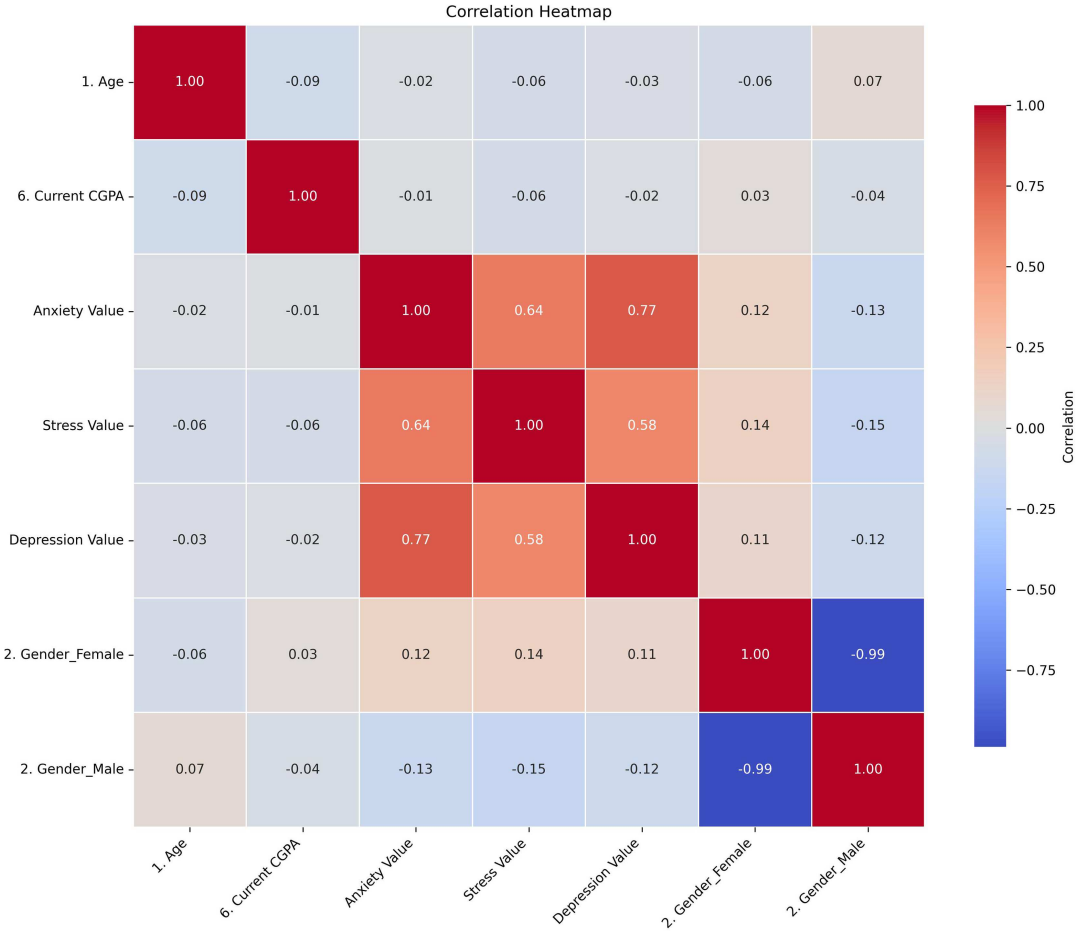

Figure 4. The heatmap illustrates the strength and direction of relationships among the features, with colors indicating the intensity of correlation.

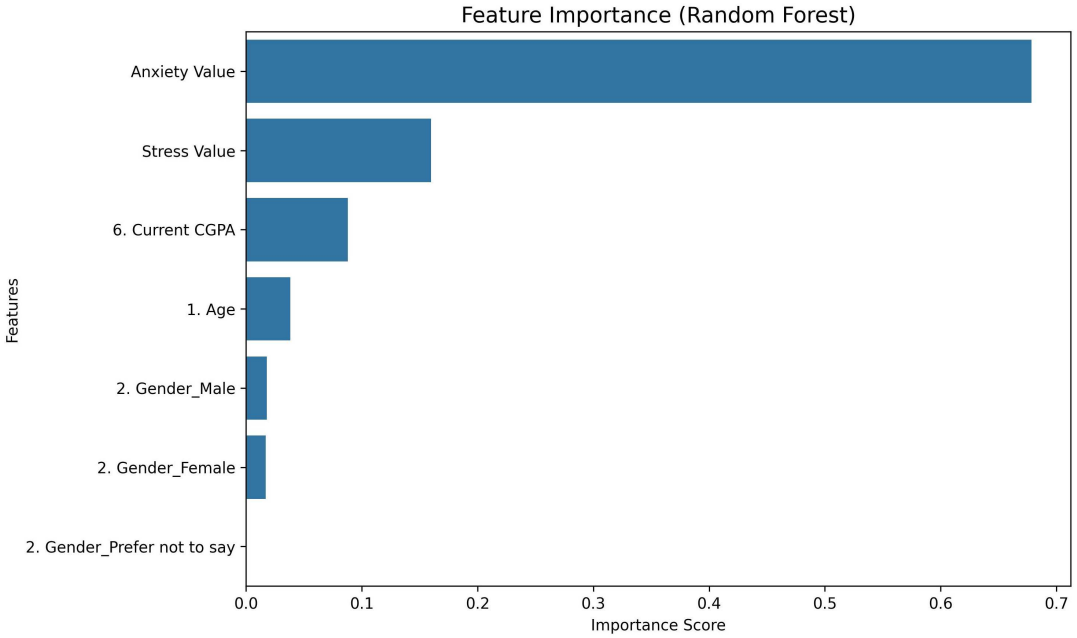

Figure 5. The bar diagram shows the relative importance of each feature in the dataset, with taller bars indicating higher influence on the target correlation.

Table 3. Summary of the selected features used for analysis

Feature	Type	Description/notes	Value distribution/stats
Age	Numeric/categorical	Age of participants, grouped in ranges	18–22: 64%, 23–26: 34%, Other: 2%
Gender	Categorical	Gender of participants	Male: 69%, female: 30%, other: 1%
Current CGPA	Numeric/categorical	Current academic performance (GPA)	3.00–3.39: 29%, 3.40–3.79: 28%, other: 43%
Anxiety value	Numeric	Anxiety score derived from survey	0–4: 6, 4–8: 10, 8–12: 42, . . . (bins given in the dataset)
Stress value	Numeric	Stress score derived from survey	0–2.7: 76, 2.7–5.4: 97, . . . (bins given in the dataset)
Depression value	Numeric	Depression score derived from survey	0–2.1: 100, 2.1–4.2: 58, . . . (bins given in the dataset)

were chosen because they are widely established for capturing various patterns and relationships in data.

Linear Regression (LR) [14]: This models the relationship between input features and target variable using a straight-line equation. The basic equation of LR, where y is the dependent variable and X is the independent variable, is

$$Y = \mathrm{m}X + \mathrm{C} \tag{2}$$

Random Forest Regressor (RFR) [15]: RFR uses an ensemble of decision trees (uses a tree-like model for making decisions and predicting outcomes). RFR uses an ensemble of decision trees (uses a tree like model for making decisions and AQ5 predicting outcomes) to make predictions by averaging outputs. In the equation, g(x) represents the final predicted value, M is the total number of decision trees, and rm(x) is the prediction of the m-th individual decision tree.

$$g(x) \;=\; (1/\mathrm{M}) \;\times\; \Sigma[m = 1 \text{ to } \mathrm{M}]\ \mathrm{rm}(x) \tag{3}$$

Gradient Boosting Regressor (GBR) [16]: Builds sequential trees where each new tree corrects errors of previous ones. In this algorithm, decision trees are built successively. Each successive tree is used to correct the errors made by the previous trees in the sequence. Predictions are made by summing the contributions of all the trees after training all the trees. The final prediction is given by the formula:

$$ypred = y1 + \eta \cdot r1 + \eta \cdot r2 + \cdots + \eta \cdot rN \tag{4}$$

AdaBoost Regressor (ABR) [17]: This model uses a sequential combination of multiple weak learners, giving higher weight to instances mis-predicted by earlier models. It is based on a loss function and re-weights the training data at each iteration. The equation represents the final predicted value. Here, $y(x)$ is the final prediction after M iterations, which is determined by the weighted median of the individual weak learner predictions fm (x).

$$\begin{aligned} y(x) \;=\; &\text{Weighted Median}(\{\mathrm{f1}(x), \ldots, \mathrm{fM}(x)\}; \\ &\{\log(1/\beta 1), \ldots, \log(1/\beta M)\}) \end{aligned} \tag{5}$$

Histogram-based Gradient Boosting Regressor (HGBR) [18]: This model uses a faster gradient boosting variant that bins the continuous features into histograms for efficient training. This algorithm does regression over negative gradients. The equation of HGBR is stated below, where M is the total number of trees and the learning rate is represented by η.

$$\text{Prediction} = \text{mean} + \eta(\text{Tree1} + \text{Tree2} + \cdots + \text{Tree}M) \tag{6}$$

2.1.5. Testing and training

To evaluate the models, the following performance metrics were used: mean squared error (MSE), mean absolute error (MAE), coefficient of determination (R^2), confidence interval (CI), and standard deviation (SD). R^2 depicts how well a model explains variance in the target, whereas prediction errors are measured by MSE and MAE. Additionally, CI provides a range of values within which the true parameter or estimate is expected to lie with a certain level of confidence, typically 95%. CI reflects the reliability of model predictions and statistical estimates, indicating the degree of uncertainty and helping to quantify the precision of the results. Lastly, SD determines the dissemination of data points around the mean. Lower SD indicates stability of the result, whereas higher SD indicates more variance. These quantities yield information regarding the reliability, stability, and resilience of the model predictions. Table 4 presents the formula and preferred value of each of these metrics [19].

After testing all models with the aforementioned performance metrics (MSE, MAE, R^2), it has been observed that LR provided the best outcome compared to the other regression methods. It worked well as the data are simple, linear, and low-noise. Finally, K-means clustering will be used for exploratory analysis to examine the relationship between anxiety and depression across different datasets, as anxiety appears to be the most significant feature that affects the change in depression level from the feature importance bar graph.

K-means clustering: This technique is an unsupervised ML algorithm that groups similar data into clusters, where each cluster is represented by a centroid. Figure 6 illustrates that the relationship between anxiety and depression is positive in the current dataset, indicating a proportional relationship between the two variables.

Anxiety: Anxiety is represented by tremendous fear and worry. It can also include heightened physiological arousal, accompanied by restlessness. Additionally, this phenomenon causes increased heart rate and difficulty concentrating. When these symptoms continue and become incongruent to real threats, they may illustrate an anxiety disorder [20].

Depression: Depression can be defined as a mood disorder, which is indicated by continuous sadness, loss of interest in usual activities, and reduced energy. According to the WHO, this can further lead to impaired concentration and significant changes in sleep or appetite, which ultimately affects daily functioning as well as psychological well-being.

A similar pattern was also observed between anxiety and depression when the evaluation was carried out using an exter-

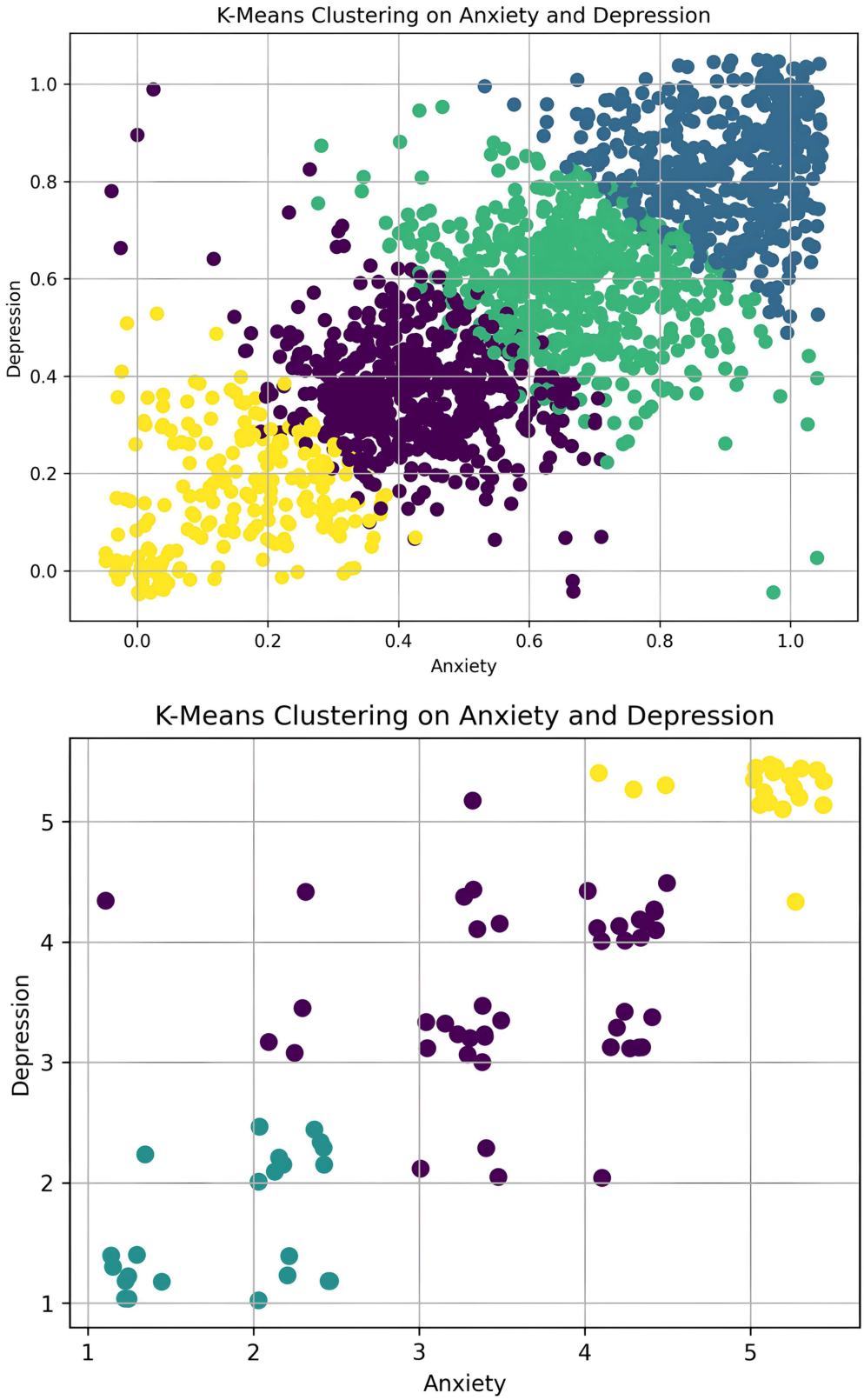

Figure 6. Visualization of the relationship between anxiety and depression using K-means clustering analysis in the current and external dataset

Table 4. Some important metrics for evaluation with their formulas

Metrics	Measurement	Preferred value
MSE	$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	Lower
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert Y_i - \hat{Y}_i\rvert$	Lower
R^2 value	$1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2}$	Closer to 1
CI	Sample-mean ± margin of error	Higher
SD	$\sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x})^2}{n-1}}$	Lower

Table 6. 10-fold results of Linear Regressor

Fold/metric	MSE	MAE	RMSE	R^2
1	0.0208	0.1149	0.1442	0.6163
2	0.0292	0.1208	0.1709	0.5232
3	0.0254	0.1249	0.1595	0.5978
4	0.0209	0.1156	0.1444	0.6261
5	0.0226	0.1175	0.1503	0.6050
6	0.0300	0.1276	0.1732	0.5388
7	0.0239	0.1224	0.1547	0.5960
8	0.0249	0.1238	0.1578	0.5702
9	0.0223	0.1231	0.1493	0.6266
10	0.0196	0.1121	0.1401	0.7078
Mean	0.0240	0.1203	0.1544	0.6008
SD	0.0035	0.0050	0.0111	0.0515
95% CI (L)	0.0215	0.1167	0.1465	0.5640
95% CI (U)	0.0265	0.1238	0.1624	0.6376

nal dataset that constitutes data of 88 students across various universities in Pakistan, resembling a coherent result across the subcontinent.

2.1.6. Outcome explanation

To further elucidate the best-performing regression model, XAI approaches were applied to analyze the impact of single features on the model output response. This interpretability is very attractive as it adds transparency to the prediction, and it makes the output more credible.

SHAP values (reflect from cooperative game theory) quantify the degree to which each feature impacts the prediction for both overall and individual instances [21]. Furthermore, local explainability techniques such as LIME are used in XAI to interpret individual predictions of complex ML models [22, 23]. For better assimilation of the model, SHAP and LIME were implemented for the explainability of the outcome.

3. Results and Discussion

Table 5 represents the overall performance of the models. This table shows that LR has the lowest MSE, MAE, and highest R^2 of 0.025011, 0.119797, and 0.581867, respectively. All the evaluation metrics have been explained in Section II (F). Furthermore, fold-wise performance of LR was evaluated for better accuracy, which is illustrated in Table 6. It was observed that Fold 10 came out with the best results.

The 95% CI for the R^2 value helps avoid interpreting the result using just a single value and provides an understanding of the model's reliability across a range of values. This outcome shows the range within which the true performance of the model is likely to fall. This means the model is not performing by chance but is performing consistently, with its accuracy within a definite range. If this range is narrow, it indicates that the model is consistent and stable, whereas a wider range would suggest more uncertainty. Therefore, the CI makes it easier to interpret how much we can trust the model's predictions in practical situations. The model predicted a mean value of R^2 as 0.6008 with a 95% CI of 0.5640–0.6376; thus, 95% CI(L) = 0.5640 and 95% CI(U) = 0.6376, meaning that it is 95% confident that the value of R^2 lies between 0.5640 and 0.6376. On the other hand, SD indicated that R^2 values vary on average by ±0.0515 from the mean value of 0.6008, which reflects the stability and reliability of the prediction. The R^2 value of approximately 0.6008 reflects a satisfactory level of explanatory power. Unlike categorical classification approaches, the regression model captures the continuous variation in depression levels, providing a more precise and realistic representation of patient condition, which may also contribute to the moderate R^2 value. As a result, the model preserves the granularity of depression severity rather than categorical segregation. From a clinical perspective, this approach can provide valuable support to medical practitioners by enabling them to monitor subtle changes in depression severity with respect to the change in psychological factors such as anxiety as well as academic and social factors like CGPA, gender and age, these findings will help to identify early risk trends, and make more informed decisions regarding intervention and treatment planning, thereby complementing traditional assessments and enhancing personalized patient care.

Table 5. Performance evaluation of all the models

Model	MSE	MAE	R^2
Random Forest Regressor	0.031727	0.136562	0.469594
Gradient Boosting Regressor	0.025700	0.121302	0.570365
AdaBoost Regressor	0.026890	0.128368	0.550463
Linear Regression	0.025011	0.119797	0.581867
Histogram-based Gradient Boosting Regressor	0.029318	0.131097	0.509878

XAI plays a vital role in the overall framework of the dataset that helps interpret model predictions through training and testing. Using the LR model, XAI shows the impact of anxiety, stress, gender, age, and CGPA in fluctuating the depression level compared to other features. SHAP and LIME were used as XAI techniques, where SHAP explains the overall effect of the features across the dataset while LIME explains feature significance for individual cases. Figure 7 represents the LIME and SHAP plot for the model.

The SHAP and LIME analyses indicate that anxiety is the most influential factor, followed by stress, in driving variations in depression levels. Overall, it was observed that psychological factors have more influence in causing depression.

This research not only studies the use of ML models and XAI but also dives into the psychiatry domain of medical science,

education, and current psychological and social conditions of IT students. By enhancing the predictive capabilities, significant steps can be taken toward improving the mental health conditions of IT students and the education curriculum in IT sectors. The methodology demonstrates promising results in predictive capability, with LR emerging as the most accurate model. Moreover, this study presents a proof-of-concept framework demonstrating the potential of integrating ML and XAI for depression prediction. Since the dataset consists mainly of universities in Bangladesh, this prediction might not be similar in all regions of the world, leading to potential biases in the dataset. Besides, the findings are illustrative and require further validation using larger and more diverse datasets. Despite the aforesaid limitations, the result will serve well in predicting the mental health condition of the people of South Asia. Feature optimization techniques like principal component analysis can be implemented to further enhance the accuracy of the results in the future. Table 7 provides a comparison between the proposed methods and existing methods.

In the above table, it can be illustrated that the proposed method describes the IT students' mental health condition by using the regression method and XAI, which is not present in related studies. Most of the papers used classification and measured performance using accuracy; thus, R^2 remained absent.

The identification of anxiety, stress, gender, age, and CGPA as key contributors to depression among IT students aligns with earlier research on student mental health, while extending insights through the use of XAI. However, unlike previous studies, which were based on classification models, this research is a combination of regression with interpretable ML techniques. The results obtained from the research underscore the importance of targeted interventions in higher education, enabling universi-

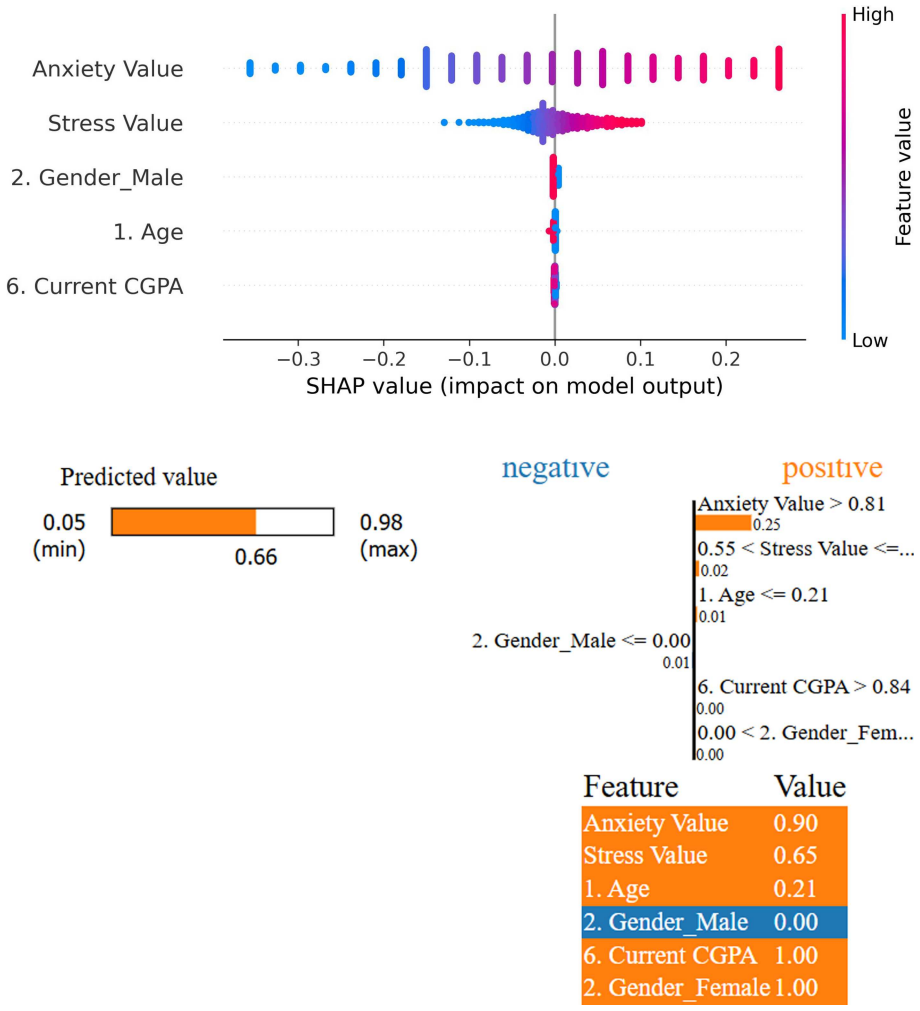

Figure 7. Relative contribution and impact of each feature on the final model prediction demonstrated by SHAP and LIME plot

Table 7. Comparison between the proposed method and existing methods

Study (short)	IT students	Regression	XAI	R^2
Madububambachu et al. [7]	No	No	No	N/A
Mutalib et al. [8]	Yes	Yes (logistic regression)	No	N/A
Abdul Rahman et al. [9]	Yes	No	No	N/A
Pathak [10]	No	No	No	N/A
Proposed method	Yes	Yes	Yes	0.6008 (after cross-validation)

ties, guardians, and mental health practitioners to address critical stressors and promote well-being in IT-focused learning environments. Overall, this research provides a combined, explainable outlook of the factors affecting mental health.

4. Conclusion

The paper successfully combines ML and XAI in predicting mental health conditions of IT students. Furthermore, the paper demonstrates promising results in predicting mental health conditions among IT students. The paper highlighted the main reasons for depression among IT students, as identified by SHAP interpretation of the LR model, where anxiety and stress are identified as the main reasons. This finding underscores the need of universities and guardians to remain attentive to students regarding their mental condition and well-being in order to maintain a healthy environment in the IT-education domain. The method used a dataset of 1977 students, which was further reduced to 1906 in the data preprocessing stage. Upon applying feature extraction, a total of 7 features including the target feature depression were selected from 39 features based on the impact and suitability of the prediction. Among all the models, LR came out as the best performer, achieving an R^2 of 0.581867. This model was further evaluated by a 10-fold cross-verification, which resulted in achieving an R^2 of 0.6008. Lastly, SHAP and LIME plots were used as a part of XAI to explain the outcome of the result.

Ethical Statement

The study was conducted using secondary data, which are publicly available and sourced from Kaggle. Additionally, the primary dataset predominantly represents students from South Asian regions, particularly Bangladesh, which may cause regional or cultural bias. Moreover, the data are self-reported survey data, which are subject to response bias. However, the effects of these drawbacks are limited, provided the information is meticulously applied. Future work should focus on more diverse primary datasets and a larger number of samples. The study used human data available from public datasets, and the data were collected under the ethical approvals reported in the original study/dataset.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are publicly available in at:

1. https://www.kaggle.com/datasets/mohsenzergani/bangladeshi-university-students-mental-health
2. https://www.kaggle.com/datasets/abdullahashfaqvirk/student-mental-health-survey

Author Contribution Statement

Md. Rakeen Islam Nahin: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Safiul Haque Chowdhury:** Methodology, Validation, Writing – review & editing, Supervision. **Mohammed Ibrahim Hussain:** Supervision. **Mohammad Minoar Hossain:** Supervision. **Mohammad Mamun:** Supervision.

References

[1] Antoninis, M., Alcott, B., Al Hadheri, S., April, D., Fouad Barakat, B., Barrios Rivera, M., . . . , & Weill, E. (2023). Global Education Monitoring Report 2023: Technology in education: A tool on whose terms? *GEM Report UNESCO*. https://doi.org/10.54676/UZQV8501

[2] Loor Vinces, L. W., & Galarza López, J. (2024). Mental health and academic performance in university students: Systematic review. *Salud, Ciencia y Tecnología*, *4*, 599. https://doi.org/10.56294/saludcyt2024.599

[3] Deblonde, J., De Koker, P., Hamers, F. F., Fontaine, J., Luchters, S., & Temmerman, M. (2010). Barriers to HIV testing in Europe: A systematic review. *The European Journal of Public Health*, *20*(4), 422–432. https://doi.org/10.1093/eurpub/ckp231

[4] Paiva, U., Cortese, S., Flor, M., Moncada-Parra, A., Lecumberri, A., Eudave, L., . . . , & Arrondo, G. (2025). Prevalence of mental disorder symptoms among university students: An umbrella review. *Neuroscience & Biobehavioral Reviews*, *175*, 106244. https://doi.org/10.1016/j.neubiorev.2025.106244

[5] Hossain, M. K., & Halder, A. (2025). Prevalence and determinants of stress among university students in Bangladesh: Insights from a study at Gopalganj Science and Technology University. *International Journal of Statistical Sciences*, *25*(1), 23–37. https://doi.org/10.3329/ijss.v25i1.81042

[6] Paiva, U., Cortese, S., Flor, M., Moncada-Parra, A., Lecumberri, A., Eudave, L., . . . , & Arrondo, G. (2025). Prevalence of mental disorder symptoms among university students: An umbrella review. *Neuroscience & Biobehavioral Reviews*, *175*, 106244. https://doi.org/10.1016/j.neubiorev.2025.106244

[7] Madububambachu, U., Ukpebor, A., & Ihezue, U. (2024). Machine learning techniques to predict mental health diagnoses: A systematic literature review. *Clinical Practice & Epidemiology in Mental Health*, *20*(1), e17450179315688. https://doi.org/10.2174/0117450179315688240607052117

[8] Mutalib, S., Shafiee, N. S. M., & Abdul-Rahman, S. (2021). Mental health prediction models using machine learning in higher education institution. *Turkish Journal of Computer and Mathematics Education*, *12*(5), 1782–1792. https://doi.org/10.17762/turcomat.v12i5.2181

[9] Abdul Rahman, H., Kwicklis, M., Ottom, M., Amornsriwatanakul, A., Abdul-Mumin, K. H., Rosenberg, M., & Dinov, I. D. (2023). Machine learning-based prediction of mental well-being using health behavior data from university students. *Bioengineering*, *10*(5), 575. https://doi.org/10.3390/bioengineering10050575

[10] Pathak, A. (2023). Pervasive effect of psychological distress on general public and caregivers of a person with mental illness: A review. *Antrocom: Online Journal of Anthropology*, *19*(1), 125–134.

[11] Zergani, M. (2023). *University Students Mental Health [Date set]. Kaggle*. https://www.kaggle.com/datasets/mohsenzergani/bangladeshi-university-students-mental-health

[12] Ashfaq, A. (2024). *Students Mental Health Survey [Date set]. Kaggle*.

[13] Singh, N. K., & Nagahara, M. (2024). LightGBM-, SHAP-, and Correlation-Matrix-Heatmap-based approaches for analyzing household energy data: Towards electricity

self-sufficient houses. *Energies*, *17*(17), 4518. https://doi.org/10.3390/en17174518

[14] Wang, Y. A., Huang, Q., Yao, Z., & Zhang, Y. (2024). In a class of linear regression methods. *Journal of Complexity*, *82*, 101826. https://doi.org/10.1016/j.jco.2024.101826

[15] Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/BJML/2024/007

[16] Gunasekara, N., Pfahringer, B., Gomes, H. M., & Bifet, A. (2025). Gradient boosted bagging for evolving data stream regression. *Data Mining and Knowledge Discovery*, *39*(5), 65. https://doi.org/10.1007/s10618-025-01147-x

[17] Xie, Y., & Tummala, S. P. (2025). Machine learning for sensor analytics: A comprehensive review and benchmark of boosting algorithms in healthcare, environmental, and energy applications. *Sensors*, *25*(23), 7294. https://doi.org/10.3390/s25237294

[18] Salehi, M. M., & Zarei, H. (2025). A review of the structure and application of scikit-learn datasets in machine learning model development. *International Journal of Operations Research and Artificial Intelligence*, *1*(2), 88–98. https://doi.org/10.48314/ijorai.v1i2.64

[19] Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, *14*(1), 6086. https://doi.org/10.1038/s41598-024-56706-x

[20] American Psychiatric Association. (Ed.). (2022). *Diagnostic and statistical manual of mental disorders (DSM-5-TR)*. American Psychiatric Association Publishing. https://doi.org/10.1176/appi.books.9780890425787

[21] Ortigossa, E. S., Dias, F. F., Barr, B., Silva, C. T., & Nonato, L. G. (2025). T-explainer: A model-agnostic explainability framework based on gradients. *IEEE Intelligent Systems*, *40*(5), 34–44. https://doi.org/10.1109/MIS.2025.3564330

[22] Seo, B., & Li, J. (2024). Explainable machine learning by SEE-Net: Closing the gap between interpretable models and DNNs. *Scientific Reports*, *14*(1), 26302. https://doi.org/10.1038/s41598-024-77507-2

[23] Basha, S. E., Gull, M., Alquqa, E. K., Mahmoud, K., & Harhash, A. (2025). The role of AI in university students' mental health: A bibliometric review. *Discover Social Science and Health*, *5*(1), 131. https://doi.org/10.1007/s44155-025-00300-7

How to Cite: Nahin, M. R. I., Chowdhury, S. H., Hussain, M. I., Hossain, M. M., & Mamun, M. (2026). Explainable AI-Driven Identification of Depression Levels for Early Mental Health Intervention in IT Students. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN62028924