An Explainable AI-Driven Machine Learning Approach for Student Depression Detection

Conference Paper · February 2025

DOI: 10.1109/ECCE64574.2025.11013941

5 authors, including:

Nazim Uddin
Chandpur Science and Technology University
5 PUBLICATIONS **52** CITATIONS
SEE PROFILE

Md. Jahidul Islam
Chandpur Science and Technology University
15 PUBLICATIONS **50** CITATIONS
SEE PROFILE

2025 International Conference on Electrical, Computer and Communication Engineering
(ECCE), 13-15 February 2025, CUET, Chattogram-4349, Bangladesh

An Explainable AI-Driven Machine Learning Approach for Student Depression Detection

Nazim Uddin
Department of ICT
Chandpur Science and Technology University
Chandpur, Bangladesh
nazim.uddin.cse.jnu@gmail.com

Md Sajib Mia
Department of CSE
Sylhet Engineering College
Sylhet, Bangladesh
sajibsec18@gmail.com

Sohel Rana
Department of ICT
Chandpur Science and Technology University
Chandpur, Bangladesh
sohelranacse052@gmail.com

Prince Mahmud
Department of CSE
Chandpur Science and Technology University
Chandpur, Bangladesh
m.princecse@gmail.com

Md. Jahidul Islam
Department of CSE
Chandpur Science and Technology University
Chandpur, Bangladesh
jahid.jabed@gmail.com

***Abstract*—Student mental health has emerged as a major global concern, with depression being a common problem that has a considerable influence on academic performance and personal well-being. This paper presents an Explainable AI (XAI) enabled machine learning framework for effectively diagnosing depression in students. The model uses advanced machine learning algorithms to examine the dataset which includes behavioral, academic, and demographic aspects, with an impressive accuracy of 98.41%. To provide interpretability and transparency, the framework incorporates XAI techniques like LIME (Local Interpretable Model-agnostic Explanations) and ELI5 (Explain Like I'm 5) to uncover crucial elements that influence predictions. The findings show that specific behavioral patterns and academic indicators are critical in identifying depressed tendencies and providing actionable information for educators and mental health providers. This method promotes trust, accountability, and successful intervention techniques by striking a balance between explainability and predictive performance, with the ultimate goal of improving student's mental health outcomes.**

***Index Terms*—Mental Health, Depression, Explanaible AI, LIME, ELI5, Machine Learning**

I. Introduction

A. Introduction

Depression is a significant mental health concern that affects millions of students globally, leading to adverse effects on academic performance, social engagement, and overall well-being. The World Health Organization (WHO) estimates that depression is one of the leading causes of disability worldwide, with young adults being particularly vulnerable to its impact. The increasing prevalence of depression among students necessitates innovative methods for early detection and intervention, enabling timely support and care. Traditionally, mental health assessments rely on surveys and clinical evaluations, which, while effective, are often resource-intensive and subject to bias or underreporting [1].

Machine learning (ML) techniques have emerged as a promising alternative for depression detection, offering data-driven approaches to identify patterns and predict mental health conditions with high accuracy. However, the complexity of ML models often creates a trade-off between predictive power and interpretability, posing challenges for their application in sensitive domains like mental health. To bridge this gap, Explainable Artificial Intelligence (XAI) methods provide transparency by revealing how models make predictions, fostering trust among educators, psychologists, and other stakeholders. Despite this advancement, traditional ML and XAI methods often require extensive computational resources and feature engineering, limiting their efficiency. The use of Explainable AI (XAI) techniques in machine learning processes has emerged as a crucial development to overcome this constraint. This paper suggests an explainable AI-driven method for detecting sadness in students that makes use of ELI5 (Explain Like I'm 5) and LIME (Local Interpretable Model-agnostic Explanations). Both local interpretability, which explains specific predictions, and global interpretability, which provides insights into the behavior of the entire model, are provided by these XAI tools. While ELI5 offers a thorough grasp of feature importance and their contributions to the predictions, LIME uses a local model approximation to produce human-readable explanations for particular cases. Our approach assures that the machine learning model's predictions are not only precise but also understandable and useful. The method helps mental health professionals and educational institutions understand the root causes of depression by identifying important elements such as age, study hours, financial hardship, and academic pressure. Furthermore, explainability increases confidence in the AI system, which qualifies it for practical uses where morality and openness in decision-making are crucial.

B. Contribution

The objectives of this research are as follows:

- To determine important factors and trends linked to student depression, providing practical advice for prevention

and treatment.
- To assess how well each ML model performs in identifying student depression.
- Enhancing stakeholder trust by guaranteeing the interpretability of predictions using XAI frameworks.

This work attempts to solve the urgent problem of student mental health by combining XAI with machine learning pipelines in a way that is reliable, interpretable, and scalable. This innovative paradigm is in line with the increasing focus on addressing real-world issues by fusing human-centric concepts with powerful analytics.

II. Related Work

Student mental health, especially depression, has become a focal point of research due to its widespread impact on academic performance, social relationships, and overall well-being. Traditional mental health assessments, though valuable, are resource-intensive and often subjective. Machine learning (ML) has emerged as a promising solution, offering data-driven methods to detect and predict mental health conditions. However, the interpretability of these models remains a critical challenge, particularly in sensitive applications like mental health. To address this, Explainable AI (XAI) methods, along with tools like LIME and ELI5 are being explored to make ML models more transparent and actionable.

To improve model interpretability, several XAI methods have been developed. Lundberg and Lee (2017) [2] introduced SHAP (SHapley Additive exPlanations), which provides consistent and accurate explanations for ML predictions. Grad-CAM++, proposed by Chattopadhyay et al. (2019) [3], improved visual explanations for convolutional networks, making it easier to understand the decision-making processes of complex models. Surveys by Burkart and Huber (2021) [4] and Zhou and Smola (2018) [5] further emphasized the necessity of XAI in domains like mental health, where trust and accountability are paramount. Liu et al. (2020) [6] provided a review of ML techniques for adolescent depression prediction, underlining the importance of integrating XAI to enhance model trustworthiness. A technique called LIME (Local Interpretable Model-agnostic Explanations) [7] uses simpler, interpretable models to approximate complex machine learning models locally around a particular prediction. It works with any machine learning model because it is model-agnostic. LIME enhances model transparency and trust by assisting in the identification of the key factors affecting individual forecasts, especially in delicate domains like healthcare and finance. By offering clear, understandable explanations of intricate model behaviors, Google Research's ELI5 (Explain Like I'm 5) [8] model enhances machine learning interpretability. It seeks to improve models' readability and transparency, particularly for non-experts. ELI5 promotes accountability and transparency by decomposing decision-making into clear explanations, especially in delicate fields like healthcare, finance, and education.

Shatte et al. [9] reviewed the application of ML in mental health, focusing on its ability to uncover patterns and predict conditions like depression. Their findings underscored the potential of ML in enhancing early detection and intervention. Uddin (2019) [10]applied ML techniques to analyze social media data for predicting student depression, highlighting the need for interpretable models to ensure actionable insights. Similarly, Ghosh and Majumder (2020) [11] demonstrated the utility of XAI in the early prediction of student stress and depression using physiological signals, emphasizing the importance of transparency in such sensitive domains. Machine learning (ML) techniques were used by Dayeon et al. (Shin et al., 2020) to create prediction models for postpartum depression. Ten-fold cross-validation and nine different machine-learning techniques were used to evaluate the models. These techniques included logistic regression, support vector machines (SVM), K-nearest neighbor (KNN), random forest (RF), naïve Bayes, recursive partitioning, logistic regression, and stochastic gradient boosting. The RF approach yielded the best result, with an area under the curve (AUC) value of 0.884. Shin et al. (2020) [12] used machine learning (ML) techniques to create prediction models for postpartum depression. Stochastic gradient boosting, random forest (RF), naïve Bayes, SVM, KNN, neural networks, and ten-fold cross-validation were among the nine techniques they used. With an area under the curve (AUC) of 0.884, the RF approach produced the best outcome. A study by Andersson et al. (2021) [13] used information from 4,313 residents of Uppsala, Sweden. The randomized trees approach outperformed the other seven machine learning techniques in terms of accuracy (73%), specificity (75%), sensitivity (72%), negative predictive value (94%), positive predictive value (33%), and area under the curve (81%).

These works collectively provide a strong foundation for integrating XAI-enhanced ML techniques to address student depression detection. By leveraging the strengths of interpretable models and advanced ML techniques, researchers can develop robust, transparent, and scalable solutions to tackle the growing mental health crisis among students.

III. Proposed System

Fig 1 illustrates the procedure flow diagram outlined in the proposed methodology. We have collected the data examined in this study from a public source kaggle, and preprocessed the data to make the best fit for the machine learning model. For the preprocessing procedure, we cleaned the data, and filled the missing values with a median filter. Then, the best features are found by feature engineering for a better outcome. Then divided the data into training and testing portions of 75% and 25% respectively. After dividing the data, we trained the data with machine learning models like KNN, SVM, ExtraTree Classifier, Logistic Regression, DecisionTreeClassifier, RandomForestClassifier, GaussianNB, Perceptron etc. The model's performance was evaluated in terms of accuracy, precision, recall, and F1_score. Then, we found the best-performing model and then analyzed this model with the interpretation model like LIME and ELI5.

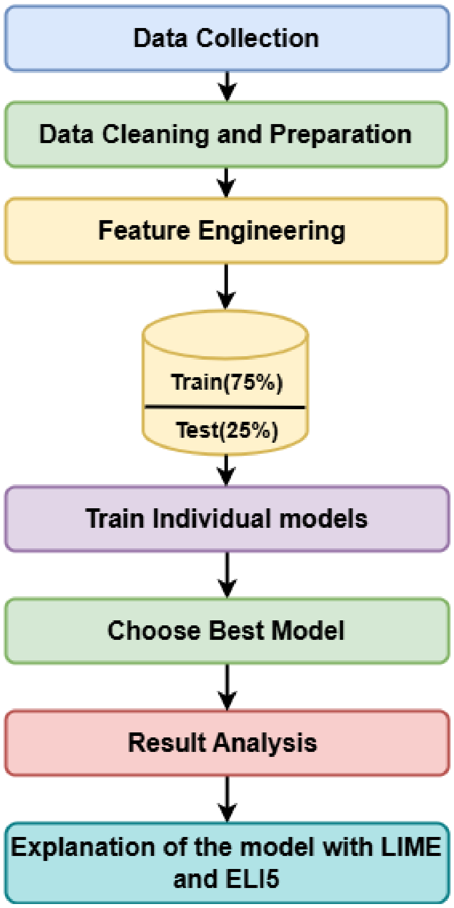


Fig. 1: A complete explanation of the proposed methodology

A. Data Collection

The dataset for this study has been collected from a Kaggle repository named Depression Student Dataset [14]. The data set has 11 features named Gender, Age, Academic Pressure, Study Satisfaction, Sleep Duration, Dietary Habits, Have you ever had suicidal thoughts?, Study Hours, Financial Stress, Family History of Mental Illness, and Depression where Depression is the target feature and there are 502 records of the students. Table I shows the detailed description of the dataset used in this study.

B. Data Cleaning and Preparation

Data preparation transforms unprocessed data into a comprehensible and useful format for analysis. Imperfections, absent trends, inconsistencies, and inconsistent patterns are just a few of the problems that raw datasets frequently offer. Furthermore, preparation is necessary to properly handle missing values and remove data discrepancies. During the preparation process, we first clean the data, and fill the missing values with a median scaler. For better training with the models the numerical features of the dataset have been normalized with a log normalization technique and the qualitative features have been encoded with a labelencoder. Figure 2 shows the log normalization of the features of the dataset used in the study.

TABLE I: Summary of the features in our dataset.

Number	Name of Attribute	Attribute Description	Data Type
0	Gender	Gender of students	object
1	Age	Age of the students in years	int64
2	Academic Pressure	Academic Pressure Low(1) to High(5)	float64
3	Study Satisfaction	Study Satisfaction Low(1) to High(5)	float64
4	Sleep Duration	Sleep Duration in Hours	object
5	Dietary Habits	Healthy, Moderate, Unhealthy	object
6	Have you ever had suicidal thoughts?	Yes/No	object
7	Study Hours	Study hours per day	int64
8	Financial Stress	Financial Stress Low(1) to High(5)	int64
9	Family History of Mental Illness	Yes/No	object
10	Depression	Yes/No	object

The original dataset was imbalanced. Machine Learning models perform well with balanced dataset [15], that's why we have balanced the data with an augmentation technique named SMOTE. Table II shows the depression class label values before and after SMOTE techniques have been used.

TABLE II: Class values before and after SMOTE Analysis

Class Label	Before SMOTE	After SMOTE
0	191	191
1	185	191

C. Feature Engineering

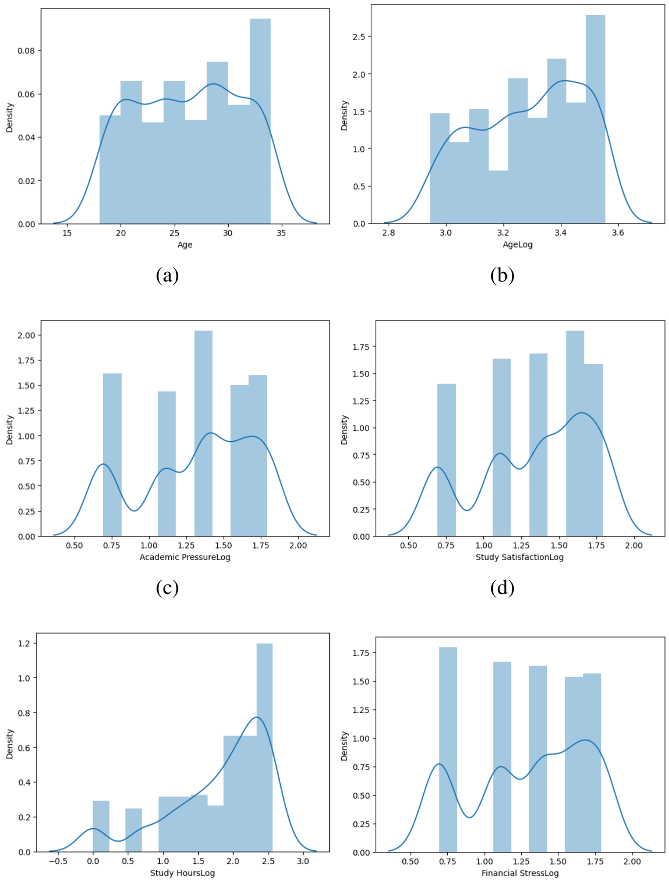
For better performances from the models trained in the process, key features have been identified by using the correlation method and Chi-Square test. Figure 3 and Table III show the features and their importance in accordance with the target value. The correlation matrices and the chi-square test both show that the feature named 'having suicidal thoughts' has the highest significance to depression, also Academic Pressure, Financial Stress, Study Satisfaction, Dietary Habits, Age, and Study Hours are the key features in this study and the rest are negligible according to the significance of the target feature.

D. Data Splitting

After the preprocessing stage, the dataset is divided into 75% for training and 25% for testing data, and the data is used to train the model with precision, recall, f1_score, and accuracy. Figure 4 shows the dataset splitting in the training and testing set.

E. Model Selection

For our study, we have employed 9 machine learning models named LogisticRegression, DecisionTreeClassifier, RandomForestClassifier, ExtraTreesClassifier, Support Vector Machine, KNeighborsClassifier, GaussianNB, AdaBoostClassifier



(a) (b)

(c) (d)

(e) (f)

Fig. 2: Log Transformation of the features

TABLE III: Chi-Sqaure Analysis of the features

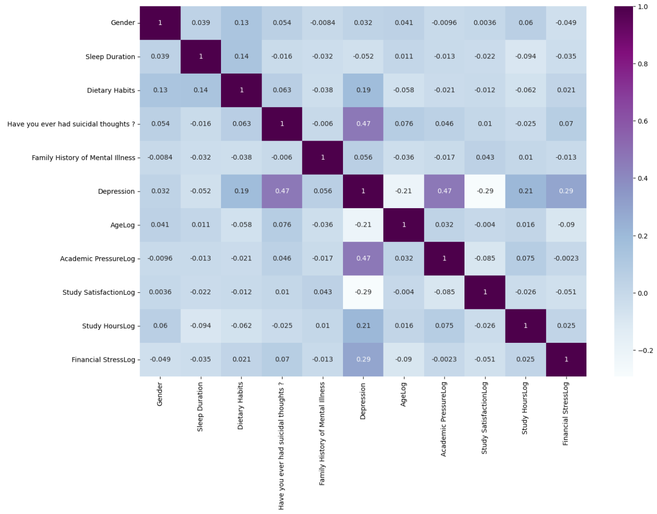
Number	Feature Name	Chi Value
0	Gender	0.235899
1	Sleep Duration	0.370347
2	Dietary Habits	5.658086
3	Have you ever had suicidal thoughts?	52.618569
4	Family History of Mental Illness	0.833620
5	AgeLog	3.717618
6	Academic PressureLog	23.294198
7	Study SatisfactionLog	8.445542
8	Study HoursLog	2.387694
9	Financial StressLog	10.051047

and GradientBoostingClassifier. Each model has been trained and analysed the performances of these models and, finally find the best model. Finally, the selected model has been explained with LIME and ELI5.

IV. EXPERIMENTAL RESULTS

A. Evaluation of Performance Matrices

Four performance measures are used in this study to assess how well the suggested architecture performs. Accuracy, precision, sensitivity (recall), and f1_score are the evaluated metrics. The mathematical formulae for these metrics have been provided below in the equation: 1, 2, 3, and 4, respectively.



	Gender	Sleep Duration	Dietary Habits	Have you ever had suicidal thoughts ?	Family History of Mental Illness	Depression	AgeLog	Academic PressureLog	Study SatisfactionLog	Study HoursLog	Financial StressLog
Gender	1	0.039	0.13	0.054	-0.0084	0.032	0.041	-0.0096	0.0036	0.06	-0.049
Sleep Duration	0.039	1	0.14	-0.016	-0.032	-0.052	0.011	-0.013	-0.022	-0.094	-0.035
Dietary Habits	0.13	0.14	1	0.063	-0.038	0.19	-0.058	-0.021	-0.012	-0.062	0.021
Have you ever had suicidal thoughts ?	0.054	-0.016	0.063	1	-0.006	0.47	0.076	0.046	0.01	-0.025	0.07
Family History of Mental Illness	-0.0084	-0.032	-0.038	-0.006	1	0.056	-0.036	-0.017	0.043	0.01	-0.013
Depression	0.032	-0.052	0.19	0.47	0.056	1	-0.21	0.47	-0.29	0.21	0.29
AgeLog	0.041	0.011	-0.058	0.076	-0.036	-0.21	1	0.032	-0.004	0.016	-0.09
Academic PressureLog	-0.0096	-0.013	-0.021	0.046	-0.017	0.47	0.032	1	-0.085	0.075	-0.0023
Study SatisfactionLog	0.0036	-0.022	-0.012	0.01	0.043	-0.29	-0.004	-0.085	1	-0.026	-0.051
Study HoursLog	0.06	-0.094	-0.062	-0.025	0.01	0.21	0.016	0.075	-0.026	1	0.025
Financial StressLog	-0.049	-0.035	0.021	0.07	-0.013	0.29	-0.09	-0.0023	-0.051	0.025	1

Fig. 3: Correlation matrices of the features.

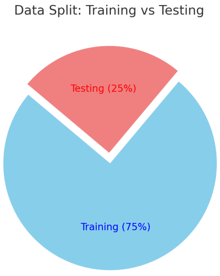


Fig. 4: Dataset Splitting

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$F1\_Score = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \tag{4}$$

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

B. Result of Proposed Model

We have trained each of the selected models and calculated the performances of these models with precision, recall, f1_score, and accuracy. The precision, recall, and F1-score of several machine learning models are compared in table IV and applied to two classes: ”No” and ”Yes.” With an F1-score of 0.95, Gradient Boosting and AdaBoost perform best overall, while Logistic Regression comes in second with 0.96 for ”Yes.” AdaBoost is the most accurate (0.98) and recall-efficient (0.98) algorithm for ”No.” On the other hand, K-Neighbors (KNN) have the lowest F1-score (0.71). Because of their exceptional balance across all criteria, Gradient Boosting and AdaBoost stand out as the best-performing models overall.

TABLE IV: Precision, recall and f1-score performance for all model.

Model	Precision		Recall		F1_score	
	No	Yes	No	Yes	No	Yes
Logistic Regression	0.97	0.96	0.95	0.97	0.96	0.96
Decision Tree	0.82	0.83	0.8	0.85	0.81	0.84
Random Forest	0.93	0.91	0.9	0.94	0.91	0.93
Extra Trees	0.9	0.92	0.92	0.91	0.91	0.92
SVM	0.93	0.89	0.86	0.94	0.89	0.91
KNeighbors	0.67	0.82	0.85	0.63	0.75	0.71
GaussianNB	0.92	0.93	0.92	0.93	0.92	0.93
AdaBoost	0.98	0.99	0.98	0.99	0.98	0.99
Gradient Boosting	0.95	0.94	0.93	0.96	0.94	0.95

The accuracy of several machine learning models with imbalanced and balanced data is contrasted in the table V. With balanced data AdaBoost model performs well, achieving the best accuracy of 98.41% which is 1.58% greater when the model trained with the imbalanced data, followed by Gradient Boosting (94.44%) and Logistic Regression (96.03%). With 92.06% accuracy, Random Forest and GaussianNB also exhibit strong performance. On the other hand, Decision Trees have a moderate accuracy of 82.54%, while K-Neighbors (KNN) has the lowest accuracy at 73.02%. All things considered, AdaBoost is the most accurate model, followed closely by Gradient Boosting and Logistic Regression.

Based on total accuracy, the model with the best performance was identified, as shown in the table V. The AdaBoost outperformed the other models, with a maximum accuracy of 98.41%. Figure 5, which displays each model’s accuracy performance, makes this AdaBoost classifier’s advantage visually apparent.

TABLE V: Model Wise Accuracy Performance (Trained with imbalanced and balanced data)

Model	**Accuracy(%)**	
	Imbalanced data	**Balanced data**
Logistic Regression	92.86	96.03
Decision Tree	80.95	82.54
Random Forest	88.10	92.06
Extra Trees	89.68	91.27
SVM	88.89	90.48
KNeighbors	71.42	73.02
GaussianNB	91.27	92.06
Gradient Boosting	93.65	94.44
AdaBoost	96.83	**98.41**

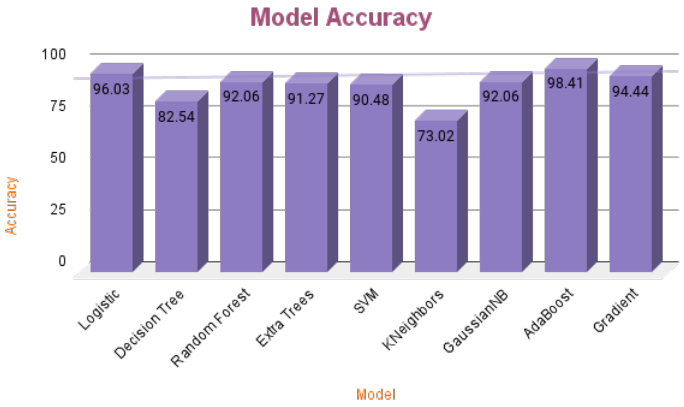


Fig. 5: Model Wise Accuracy Compariosn

Figure 6 shows the confusion matrix of the proposed model. By displaying the distribution of predictions, the confusion matrix assesses a binary classification model’s performance. The model demonstrated a significant capacity to reliably categorize both classes by correctly identifying 59 occurrences as class 0 (true negatives) and 65 examples as class 1 (true positives). Nevertheless, it incorrectly identified one case as class 0 rather than class 1 (false negatives) and one instance as class 1 rather than class 0 (false positives). This indicates that, in comparison to the examples that were correctly predicted, the model performs well generally, with a low number of classification errors.

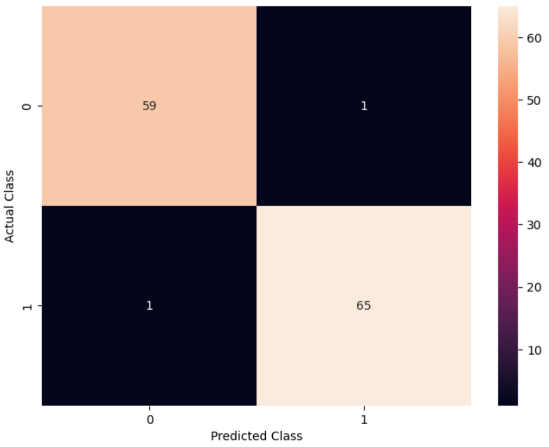
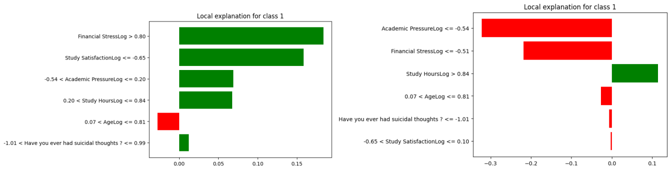


Fig. 6: Confusion matrix

C. *Explanation with LIME*

Figure 7 explained the interpretation for both class ”Yes” and ”No” for Depression prediction. The main characteristics affecting forecasts for the ”Yes” and ”No” classes are revealed by the LIME plots. For class ’Yes’ in Fig (a), the most significant contributors are Study SatisfactionLog ¡= -0.65 and Financial StressLog ¿ 0.80, with Academic PressureLog ¡= 0.20 and 0.20 ¡ Study HoursLog ¡= 0.84 providing modest support. While 0.07 ¡ AgeLog ¡= 0.81 marginally opposes ’Yes,’ features such as Have you ever had suicidal thoughts? ¡= 0.99 somewhat support it. For class ’No’ in Fig (b) the primary drivers are Financial StressLog ¡= -0.51 and Academic PressureLog ¡= -0.54, whereas Study HoursLog ¿ 0.84 provides moderate support for ’Yes’. Features such as Have you ever had suicidal thoughts? ¡= -1.01 and 0.07 ¡ AgeLog ¡= 0.81 provide minor contributions to the answer ”No.” Plotting the variables affecting the model’s predictions for both groups together provides a thorough analysis.

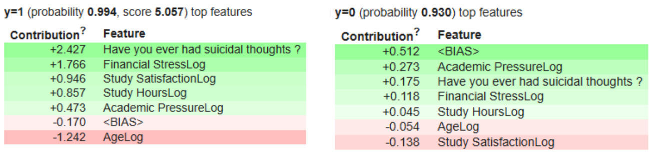


(a) (b)

Fig. 7: LIME interpretation for (a) class ”Yes” and (b) class ”No” of proposed model

D. Explanation with ELI5

Figure 8 provides the details interpretations of the model with ELI5.In Figure (a), Strongly positive contributions from Suicidal Thoughts (+2.427), Financial StressLog (+1.766), and Study SatisfactionLog (+0.946) drive the model’s 99.4% likelihood prediction of y=1. AgeLog’s (-1.242) and ¡BIAS¿’s (-0.170) negative contributions have little effect because the prediction is dominated by factors associated with stress and mental health. In Figure (b), Academic PressureLog (+0.275), Suicidal ThoughtsLog (+0.169), and Financial StressLog (+0.116) are the key drivers of the model’s 92% likelihood prediction that y=0. Stress-related factors dominate the prediction, whereas negative contributions such as AgeLog (-0.062) and Study SatisfactionLog (-0.138) somewhat lower the chance.



y=1 (probability **0.994**, score **5.057**) top features

Contribution?	Feature
+2.427	Have you ever had suicidal thoughts ?
+1.766	Financial StressLog
+0.946	Study SatisfactionLog
+0.857	Study HoursLog
+0.473	Academic PressureLog
-0.170	<BIAS>
-1.242	AgeLog

(a)

y=0 (probability **0.930**) top features

Contribution?	Feature
+0.512	<BIAS>
+0.273	Academic PressureLog
+0.175	Have you ever had suicidal thoughts ?
+0.118	Financial StressLog
+0.045	Study HoursLog
-0.054	AgeLog
-0.138	Study SatisfactionLog

(b)

Fig. 8: ELI5 interpretation for (a) class ”Yes” and (b) class ”No” of proposed model

V. Conclusion and Future works

This study offers a new Explainable AI (XAI)-driven machine learning method for identifying student depression with a remarkable 98.41% accuracy rate. By including XAI approaches like LIME and ELI5, the model’s predictions are guaranteed to be both accurate and interpretable, resolving the issues of transparency and trust that are frequently connected to AI-driven solutions. The model offers practical insights that can help parents, educators, and clinical counselors create focused mental health interventions by highlighting important elements like social interactions, academic achievement, and behavioral patterns. The model’s explainability guarantees the ethical and responsible use of AI, while its high accuracy shows how well it captures tiny yet important patterns suggestive of sadness. The strategy is appropriate for real-world applications like automated early warning systems and student-specific support mechanisms because it strikes a compromise between interpretability and performance. In summary, the suggested framework provides a solid and understandable way to deal with the increasing problem of student depression.

The original dataset has only 502 samples which is relatively too low for practical application. To further improve the model’s accuracy and applicability, future research could examine integrating additional data sources, such as physiological signals and real-time behavioral monitoring. Additionally, different data balancing techniques, like Random Oversampling, SMOTETomek, SMOTEENN, or ADASYN, could be examined, and hyperparameter tuning could be changed to enhance the model’s performance.

References

[1] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters, “Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication,” *Archives of general psychiatry*, vol. 62, no. 6, pp. 593–602, 2005.

[2] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30. Curran Associates, Inc., 2017, pp. 4765–4774.

[3] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, “Grad-cam++: Improved visual explanations for deep convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, pp. 9178–9186.

[4] N. Burkart and M. F. Huber, “A survey on the explainability of supervised machine learning,” *Journal of Artificial Intelligence Research*, vol. 70, pp. 245–317, 2021.

[5] X. Zhou and A. J. Smola, “Explainable artificial intelligence for human decision-making,” *Nature Communications*, vol. 9, no. 1, pp. 1–15, 2018.

[6] W. Liu, Z. Zhang, and L. Yang, “Machine learning for predicting adolescent depression: A review of recent advances,” *Frontiers in Psychiatry*, vol. 11, p. 595, 2020.

[7] D. Garreau and U. Luxburg, “Explaining the explainer: A first theoretical analysis of lime,” in *International conference on artificial intelligence and statistics*. PMLR, 2020, pp. 1287–1296.

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, “” why should i trust you?” explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[9] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, “Machine learning in mental health: a scoping review of methods and applications,” *Psychological medicine*, vol. 49, no. 9, pp. 1426–1448, 2019.

[10] M. Z. Uddin, “Machine learning-based student mental health and depression prediction using social media data,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 5, pp. 82–89, 2019.

[11] S. Ghosh and S. Majumder, “An explainable ai approach for early prediction of student stress and depression using physiological signals,” *Journal of Biomedical Informatics*, vol. 108, p. 103495, 2020.

[12] D. Shin, K. J. Lee, T. Adeluwa, and J. Hur, “Machine learning-based predictive modeling of postpartum depression,” *Journal of clinical medicine*, vol. 9, no. 9, p. 2899, 2020.

[13] S. Andersson, D. R. Bathula, S. I. Iliadis, M. Walter, and A. Skalkidou, “Predicting women with depressive symptoms postpartum with machine learning methods,” *Scientific reports*, vol. 11, no. 1, p. 7877, 2021.

[14] Ikynahidwin, “Depression student dataset,” https://www.kaggle.com/datasets/ikynahidwin/depression-student-dataset?resource=download, 2023, accessed: 2024-11-26.

[15] N. Uddin, M. K. U. Ahamed, M. A. Uddin, M. M. Islam, M. A. Talukder, and S. Aryal, “An ensemble machine learning based bank loan approval predictions system with a smart application,” *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 327–339, 2023.