IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

Predicting Student Well-Being Through Academic Performance and Daily Habits

Priyanka Mohan[1], Gopika R[2], Deekshitha S[3]

Assistant Professor, Department of Master of Computer Applications, Surana College (Autonomous), Bengaluru, India[1]

Student, Master of Computer Applications, Surana College (Autonomous), Bengaluru, India[2]

Student, Master of Computer Applications, Surana College (Autonomous), Bengaluru, India[3]

Abstract: Over the past several years,there,has been an increase in the number of cases where students report higher levels of mental health strain due to various reasons such as pressure in studies,changes in lifestyle,and economical issues among others.It is necessary to find out who the students are that might be depressed and give them appropriate treatment.In this study,an attempt was made to develop a predictive model for the probability of depression in students as per the academic and lifestyle indicators of the Kaggle Student Depression Data Set.The Kaggle data set was pre-processed with missing value treatment,encoding of labels and normalization of features.Broadly,machine learning is categorized into four types namely Logistic Regression,SVM,Random Forest and XGBoost were developed using the Kaggle Student Depression Data Set.The metrics analyzed were accuracy,precision,recall,and F1-score.The accuracy of logistic regression model was the highest (83.61%) among four machine learning models.

Keywords: Machine Learning,Student Depression Prediction,Student Mental Health,Academic Factors, Lifestyle Factors,Logistic Regression,Predictive Modeling,Early Risk Detection.

I. INTRODUCTION

Nowadays,the problem of student mental health is a key major problems affecting the students studying at educational institutions.Stress and mental disorders can be exacerbated among students due to academic stress,financial stress,lifestyle changes and other stress factors.Students mental disorders do not just impact their academic achievement,but also their general welfare and their interactions with other students[1,2].But,the conventional evaluation of mental health is done by questionnaires,interviews and clinical assessment.However,such approach requires more time to conduct such evaluation on a regular basis.As a result,many cases of mental disorders are not diagnosed[4].Mental condition of the students can be identified by using machine learning technology,it is possible to analyze tons of data,and see some patterns.Machine learning prediction technology can be used to make inferences about the vulnerability of students to depression.Some of the factors affecting mental health of the students can be related to stress,CGPA,nutrition,economic status of students and genetic factors [7],[8].Machine learning approaches such as logistic regression,SVM,random forest,and XGBoost are widely used for mental health risk prediction [9, 10].This study aims to build a predictive model for student depression using specific Academic and Lifestyle variables from the Student Depression Dataset. The overarching aim of this research is to design a machine learning-based model for deployment in educational settings to facilitate early prevention of depression among students [11, 12].Due to the increasing need for support services,the importance of mental health evaluations increases significantly.This is because sometimes the pupils experiencing stress,anxiety or depression receive no support and it is more difficult for the educational setting to detect pupils who need support.Machine learning techniques enable the analysis of student behavior using available data.Predictive modeling has demonstrated strong capability in detecting risk factors linked to student mental health [13]-[15]. In recent years,data-driven approaches have been adopted by educational institutions,resulting in improvements in both academic performance and student well-being.Because data can be gathered relative to academic outcomes and lifestyle regarding the student,it is possible to determine strategies to predict mental issues.This algorithms can take an information and make suggestions about what to do.Educational institutions can also utilize the prediction of mental issues apart from the psychologists.The students at risk will be identified in advance,thus avoiding any additional psychological issues.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

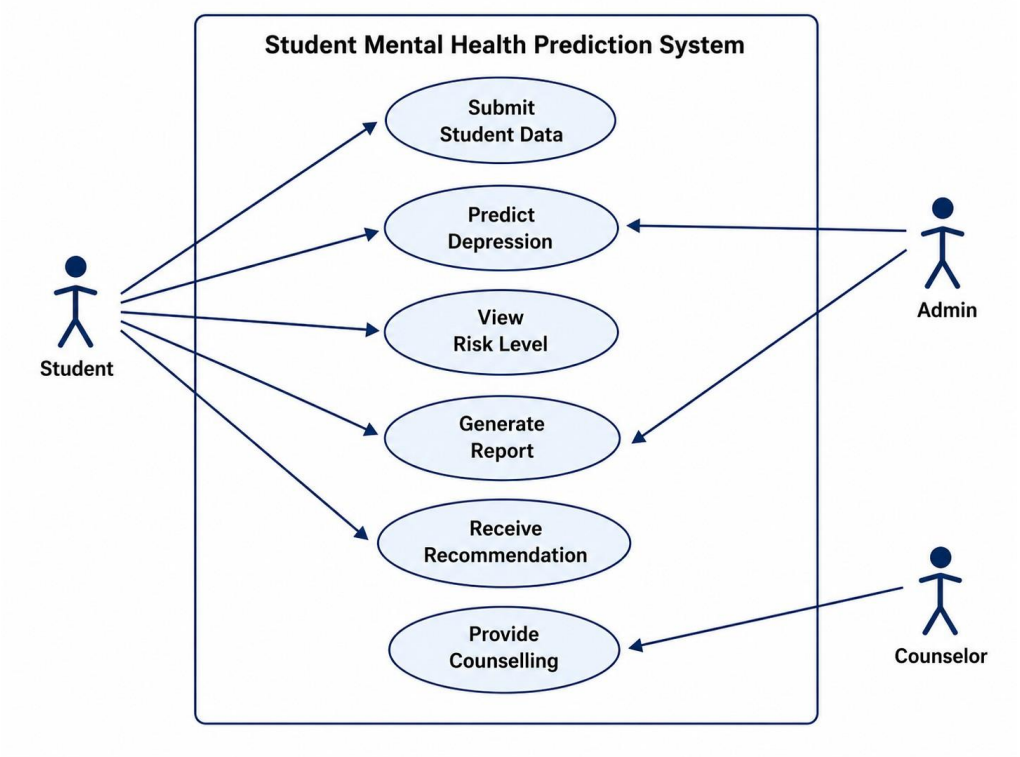

Fig. 1 Use Case Diagram of the Proposed Student Mental Health Prediction System

The student's mental health prediction system has been illustrated in the use case of Figure 1. Information pertaining to the student is fed into the system,and the machine learning algorithm trained is used to make a prediction about the risk of depression. Prediction is then employed to decide risk level and reports are produced.The administrator can see the predictions,but the counsellor can use the risk information to guide him or her in giving the appropriate advice to students.

II. LITERATURE REVIEW

Machine learning approaches have been highly popular recently due to their ability to recognize students with various mental issues such as depression, anxiety, stress and others through academic,behavioral,and lifestyle parameters.Several studies have been carried out by machine learning models analyzing risk factors in the case of students with mental disorders.According to Rois et al.[1],academic stress,quality of sleep,and life styles were recognized as significant factors of high perceived stress levels among university students,where machine learning was used as detection tools.Machine learning study conducted in Lebanon [2]This paper addresses the issue of anxiety amongst university students.Data-driven analysis has been conducted in order to prove that machine learning algorithms could be used to identify students who need psychological support.By applying artificial intelligence analysis,Rezapour and Elmshaeuser [3] have investigated the influence of the COVID-19 outbreak and online studying on the mental health of students.It was found that such variables as academic pressure,the virtual learning space,and social isolation affected the emotional state of students. Additionally,Sharda et al.[4] have developed the Student Health Risk Evaluator (SHRE), which is a framework based on the principles of a predictive model aimed at determining mental health risks of students by means of using academic and behavioral characteristics.The role of XAI in predicting mental health has been discussed in [5].The experiment provided accurate predictive results as well as important interpretations of variables affecting the state of the student's well-being.An algorithm for screening depression based on machine learning using data of the student population not containing any clinical parameters was developed by Nison et al.[6],and the result received shows that the machine learning algorithm is capable of predicting the risk of depression regardless of whether only clinical parameters are used.Nayan et al.[7] used machine learning methods to compare the predictive results of various algorithms of depression and anxiety in university students.It was found that the models of ensemble learning exhibited comparable accuracy of prediction compared to other classification algorithms.Applications of machine learning for mental health predictions as well as challenges related to it were considered by Chung and Teo [8].Models for assessment of mental well-being have been created on the basis of student health behavior data because of the application of machine learning algorithms by Rahman et al. [9].It has been recognized that the lifestyles of people are influencing their mental well-being to a great extent.Mental states of international students are analyzed using machine learning-based techniques by Rahman and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

Kohli [10].It has been realized that the factors such as academics,society,and culture play important roles in the assessment of mental well-being of the students.Assessment of depression,anxiety,and stress of university students in the time of pandemic has been done through machine learning techniques by Morr et al.[11].Intelligent approach for the assessment of mental distress using demographic,behavioral,and lifestyle factors has been proposed by Zhang et al. [12], which provided accurate predictions.Iparraguirre-Villanueva et al.[13] conducted research on the efficiency of various machine learning classifiers in prediction of depression in university students,achieving good results.The study by Chung and Teo [14] contrasts single classifier learning with ensemble learning methods when predicting mental illnesses.The findings indicate that an ensemble approach is better in terms of making predictions compared to the single classifier method.Systematic review of Schaab et al.[15]explores the application of machine learning algorithms for the detection of depression,anxiety,and stress among the undergraduate students.As is clear from the findings from the study,machine learning algorithms are getting more effective but the data size requirements are increasing.From the literature review of academic sources above,it can be seen that there is a clear need for machine learning algorithms for the prediction of mental disorders.Nevertheless,there are some limitations with regard to the size of data set,geography,types of features,and result interpretations.Hence,there is a need to create a framework that will consider both academic and lifestyle features. In this regard,a comprehensive review of academic literature on student mental well-being prediction was conducted in order to identify the most popular machine learning algorithms and their limitations in this field.Machine learning algorithms models have been reviewed to predict mental illnesses like depression,anxiety,stress,etc.,among the students.The trends that can be identified in the literature have been shown in Fig.2. Despite having received positive results through these studies,it is necessary to develop an approach which will integrate both educational as well as lifestyle-based parameters within the same model.This is due to the reason that most of the studies have taken into account a very limited set of parameters and/or students only.

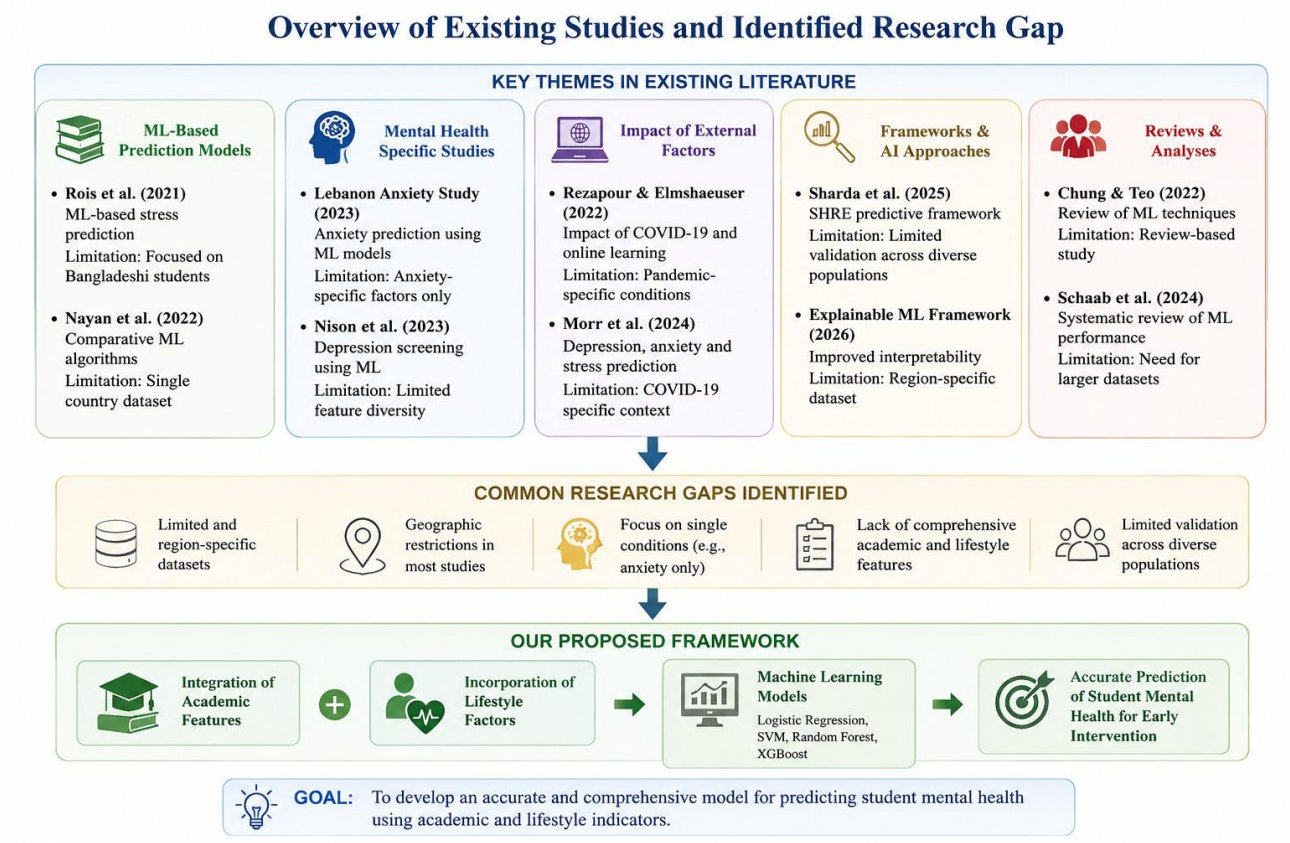

Fig. 2 Overview of Existing Studies and Identified Research Gap

The key findings and limitations of previous studies on prediction of student mental health are summarized in figure 2.There is a lot of research conducted that can be seen, that has adopted machine learning methods to identify mental illnesses among students such as depression,anxiety,stress,etc.However,most of the existing studies have limitations in their geographical focus,limited sets of predictor variables,focus on specific mental health problems,and limited sample validation.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

III. RESEARCH GAP

Prior studies have demonstrated the potential of machine learning algorithms to infer students mental states.Various ML models have been applied to examine the underlying factors of depression,anxiety,stress,and overall psychological well-being among students [1]–[15].Though there has been progress in this area recently,there are some questions to be answered.Firstly,the data sets employed in researches are specific and are related to particular country,institution,students [1],[7],[10],[13].Moreover,a number of scientific publications do not pay attention to the importance of educational and lifestyle characteristics of students that influence their psychological condition [6],[9],[12].There is another issue of some of the predictive models being hard to interpret.While recent advances in machine learning have yielded accurate predictions,it is difficult for experts in education and psychology to understand the causes of mental disorders of students [5,8].In addition,there are no publications that have an analysis of several models built in a similar vein dataset [7],[14].Despite the promising outcomes reported in earlier work,the need for building a universal model which would be able to work both with academic and non-academic features and be extremely accurate and clear remains.Hence,the prediction of students mental state based on their academic and non-academic characteristics using machine learning is proposed.

IV. PROPOSED FRAMEWORK AND METHODOLOGY

This research introduces an intelligent machine learning framework for predicting mental health disorders in students using academic and lifestyle attributes.The proposed workflow consists of three stages:data acquisition,data preprocessing, and predictive modeling for psychological conditions.The core objective is the early identification and diagnosis of depression within the student population.

A. Dataset Collection and Feature Selection

Data set is obtained and features are selected.In order to carry out our research,we have selected the Student Depression Dataset that can be accessed in Kaggle Data Repositories.The Student Depression Dataset includes information regarding the academic performance of students,their life style and mental wellbeing.The size of this dataset equals 27,901 observations and 18 features.The number of features enables us to consider a range of factors that could affect pupils mental health.The features concerning academic performance and life style of the students have been chosen to make an adequate model, which will aid to make prediction about some conditions of students.Relevant variables such as CGPA and Academic Stress,Study Satisfaction,and Work-Study Hours were selected for the model due to their bearing on the student academic environment.Additionally,other factors like Sleep Duration,Diet,Financial Stress,Family History of Mental Disorders,Age and Gender have been chosen as they are linked with the mental state of the individual.Depression has been chosen as the feature to be targeted as there is a relationship with the symptoms of depression among students.

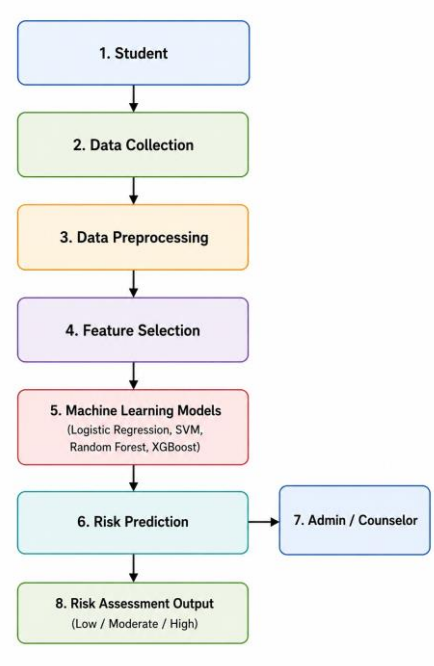

Fig. 3 Process Flow of the Proposed Student Depression Prediction System

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

B. Data Preprocessing

High data quality and reliability of machine learning models are the key factors behind the importance of data preprocessing.Before developing the machine learning models,the dataset was examined for missing values, duplicates,and other errors that could adversely affect prediction performance.The features (Gender,Sleep Duration,Dietary Habits,Family History of Mental Illness) were encoded using the Label Encoding method,since they were categorical data.When needed,features were normalized and scaled in order to guarantee the consistency between the features and enable the effective learning of the machine learning models.After preprocessing,the dataset was partitioned into training and testing subsets.The training portion was utilized to build the machine learning models,whereas the testing portion was employed to assess their performance.

C. Model Development

In the process of predicting the occurrence of psychological problems in college students.Below is the list of the algorithms used in this research as they are effective when classifying and widely applied in predictions. Logistic Regression:For binary classifications,Logistic Regression algorithm was the first algorithm used in this research due to its simplicity and effectiveness.It was possible to predict the likelihood of occurrence of depression risk factors in students using the logistic regression. In this research,Support Vector Machine (SVM) algorithm was utilized to determine the boundary between the two groups.SVM is especially good in complex classification in which the relationship between the data sets is not linear. Random Forest:This techniques used to create multiple decision trees and achieve high classification accuracy and low chances of overfitting.In the random forest technique,decision trees are used to get a better result. Extreme Gradient Boosting:XGBoost was selected for its capacity to capture complex feature interactions and nonlinear relationships.It is widely adopted due to its strong predictive performance.All models were trained on the preprocessed training data and evaluated on the held-out test data.Model performance was compared across algorithms,and the top-performing model was adopted for student depression detection.

D. Model Evaluation

Some metrics of classification have been used in order to assess the efficiency of the machine learning algorithms used to predict depression prevalence among the students.These metrics offer insights into the effectiveness of models,and can also guide the choice of prediction algorithms.In addition to these metrics a confusion matrix was also developed to analyze the output of the models with their true value.Confusion matrix is used to gain insight to the correct classification as well as incorrect classification of the models.

The following evaluation metrics were considered in this case study:

1.Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

2.Precision:

$$\frac{TP}{TP + FP}$$

3.Recall:

$$\frac{TP}{TP + FN}$$

4.F1Score:

$$\frac{2 \times (Precision \times Recall)}{Precision + Recall}$$

Other than the use of these metrics,a confusion matrix was also used to test the performance of each model through actual and predicted class labels.From the confusion matrix,one is able to identify which elements have been correctly classified and which ones have not. From the results obtained from the above evaluation metrics,the performance of Logistic

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

Regression,Support Vector Machine (SVM),Random Forest,and XGBoost algorithms was compared and the best one chosen for the predictions.

E. Proposed Framework

The designed framework follows a systematic approach to forecasting the chances of mental disease amongst students by means of behavioral and educational data.The first phase of this framework involves extracting the relevant features related to students behaviour and health from the Student Depression Dataset by collecting data from students.Data preprocessing consists of data cleaning,label encoding,and normalization to address data imperfections.Once it is done with the pre-processing,the data will be divided into two sets–training set and testing set.Some machine learning algorithms will be applied to create the framework of predicting the probability of mental diseases.These machine learning algorithms used are:Logistics regression,Support Vector Machine (SVM),Random Forest,XGBoost.They will be trained using the dataset and will be evaluated based on their performance on prediction of mental diseases.Performance metrics will be accuracy, precision,recall and F1 scores.

F. Workflow of Proposed Framework

Workflow Process The technique described here includes a workflow process that includes the collection and selection of features in the data.The collected data will be pre-processed,which includes data cleaning,data encoding and data normalization.The data set was then divided into training and test sets.A number of machine learning algorithms were designed and evaluated using the chosen performance metrics.Finally,the algorithm that was most predictive was chosen for the detection of depression in students

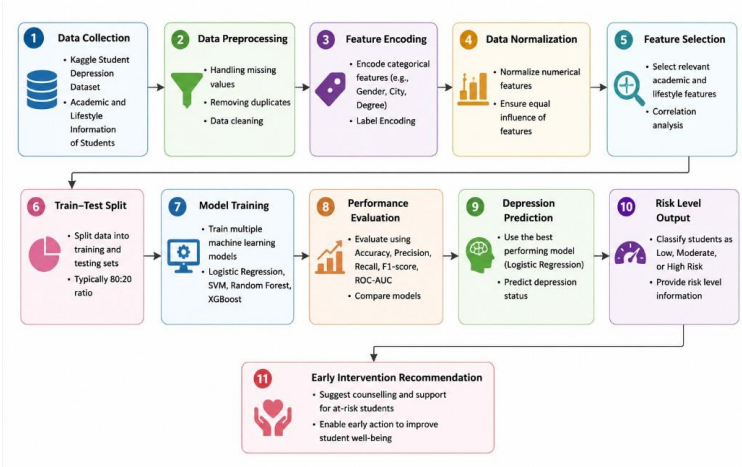

Fig. 4 Workflow of the Proposed Framework

V. IMPLEMENTATION

A.Dataset Loading

Dataset was obtained for student depression,and the data was imported into Jupyter Notebook using Pandas.Data analysis was conducted to gain information regarding the structure of the data,features and target variables.It helps to predict the status of mental health based of features.

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress
0	2	Male	33.0	Visakhapatnam	Student	5.0	0.0	8.97	2.0	0.0	5-6 hours	Healthy	B.Pharm	Yes	3.0	1.0
1	8	Female	24.0	Bangalore	Student	2.0	0.0	5.90	5.0	0.0	5-6 hours	Moderate	BSc	No	3.0	2.0
2	26	Male	31.0	Srinagar	Student	3.0	0.0	7.03	5.0	0.0	Less than 5 hours	Healthy	BA	No	9.0	1.0
3	30	Female	28.0	Varanasi	Student	3.0	0.0	5.59	2.0	0.0	7-8 hours	Moderate	BCA	Yes	4.0	5.0

Fig. 5 Student Depression Dataset Sample

B.Data Preprocessing Implementation

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

Data preprocessing is a necessary step before building a machine learning model.It was important to explore the data and find missing data points and duplicate data that might reduce the effectiveness of the model.High-quality data has been achieved by applying different data cleaning techniques to the data.To encode textual features into numerical values,Label Encoding was performed on categorical data (gender,sleep duration,eating habits and family history of mental disorders).The data set is now divided into two sets-training set and test set.Furthermore,features were normalized to ensure the equal influence of all numeric features to the process of model creation.It is very crucial to perform data preprocessing because it helps algorithm remain stable,and it ensures that there is no imbalance between features operating at different scales.The preprocessed data set was first verified for correctness prior to developing the model.In addition to this,correlation analysis was performed to identify relationships between the selected features.The type of relationship between academic,lifestyle and mental health variables is presented in the correlation matrix listed below (see Figure 7).The figure shows an association between academic performance and life satisfaction.These include Academic Stress which has a high relationship with depression to the extent that its correlation value is 0.47.Likewise,there is financial stress,also dealing with depression and here the correlation value is positive at 0.36.There are also factors such as study satisfaction and age which have a negative correlation with depression.

```
X_train, X_test, y_train, y_test = train_test_split(
    X,
    y,
    test_size=0.2,
    random_state=42
)

print(X_train.shape)
print(X_test.shape)
```

```
(22320, 17)
(5581, 17)
```

Fig. 6 Train-Test Split of the Preprocessed Dataset

The correlation analysis was conducted to determine the nature of relationships among the chosen variables. From Figure 7 below,it is clear that the correlation matrix determines the nature of the relationship between academic,lifestyle,and psychological wellbeing variables.

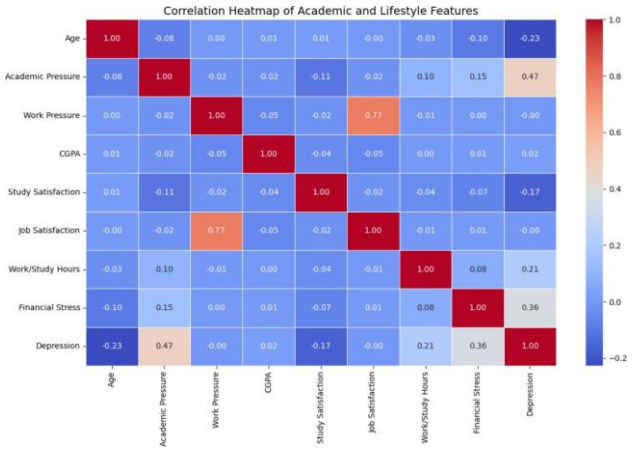

Fig. 7 Correlation Heatmap of Academic and Lifestyle Features

Figure above illustrates the correlation between different academic and lifestyle parameters.Out of these,Academic stress shows high correlation with Depression having a positive correlation value of 0.47,whereas,Financial stress also shows high correlation with depression having a positive correlation value of 0.36.It clearly indicates that both these parameters have a considerable impact on mental well-being of students.At the same time,Study Satisfaction and Age show negative correlation with Depression.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

C. Model Training

The four machine learning models that have been used include:

- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- XGBoost

Training was carried out using the training data set,and the evaluation was carried out using the test data set.The purpose of this comparison was to find out which technique is best suited for student classification.

D.Performance Evaluation

The algorithms were measured based on four different measures;these measures included accuracy,precision,recall,and F1-score.This provided an understanding of algorithm performance while allowing efficiency to be evaluated through additional metrics.The confusion matrix and visualization results were performed to give an insight into the prediction process.

VI.RESULTS AND DISCUSSION

The chapter below contains results of the study,namely,modeling of machine learning algorithms in order to predict the depression in the students.The following machine learning algorithms have been used for this purpose:Logistic Regression,Support Vector Machine (SVM), Random Forest and XGBoost.These machine learning algorithms have beentrained using Student Depression Dataset.The efficiency of machine learning algorithms has been estimated in accordance with their ability to detect students that are prone to getting depressed using such criteria as Accuracy,Precision,Recall and F1-Score.

To find out the most influential features on the prediction,the feature importance of the machine learning algorithms was analyzed in the Random Forest algorithm.

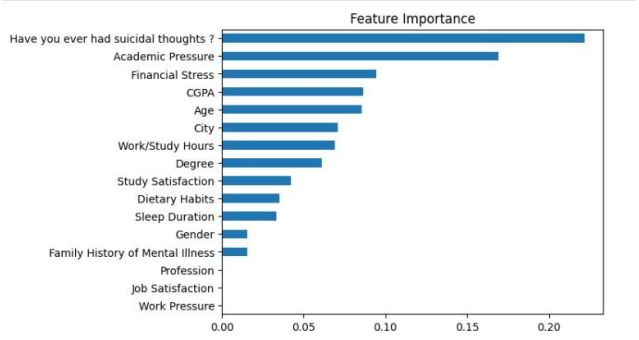

Fig. 8 Feature Importance Analysis Using Random Forest

Figure 8 presents The output shows that some of the parameters which affect the prediction of depression are parameter "Have you ever had suicidal thoughts?,"academic pressure,financial Stress and CGPA.Furthermore,the effect of more important variables is even larger on the prediction of the models.The models are compared in terms of their efficiency in predicting the accuracy.Accuracy can be considered as an evaluation criterion that assesses how well a model is able to predict the result for students with the help of their characteristics.The greater accuracy value means the higher ability of the model to differentiate between students with and without depression.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

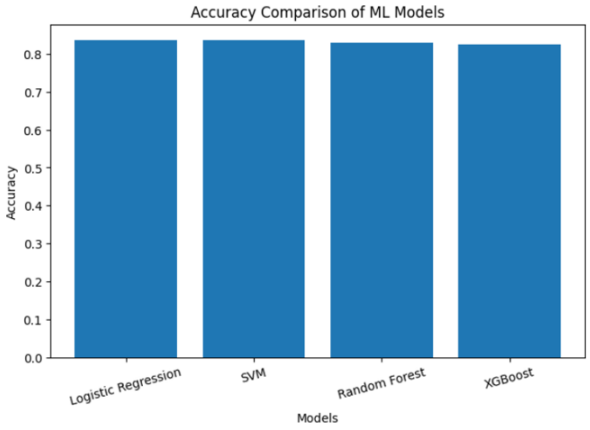

Fig. 9 Accuracy Comparison Obtained from the Executed Machine-Learning Models

The accuracy performance of the four machine learning models were demonstrated by running the test code in Figure 9.The accuracy of the models can be deduced from the results and it is found that Logistic Regression model has the highest accuracy of 83.61% followed by Support Vector Machine model which achieved an accuracy of 83.53%. The other 2 models did not have quite as high accuracy

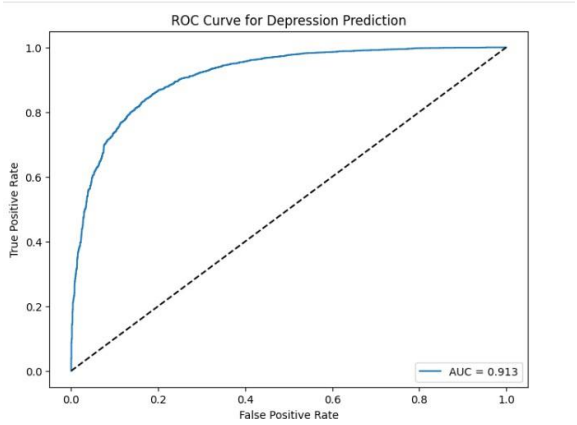

Fig. 10 Receiver Operating Characteristic Curve for the Logistic Regression Model

Figure 10 shows the Receiver Operating Characteristic (ROC) plot that was created for the Logistic Regression Model.The ROC plot shows the True Positive Rate (TPR) against the False Positive Rate (FPR),at various threshold levels.The area under the curve of the model is about 0.913 which indicates that the model has a good discriminatory power between depression and non-depression students.The higher the value of the area under the curve,the better is the classification performance of the model.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

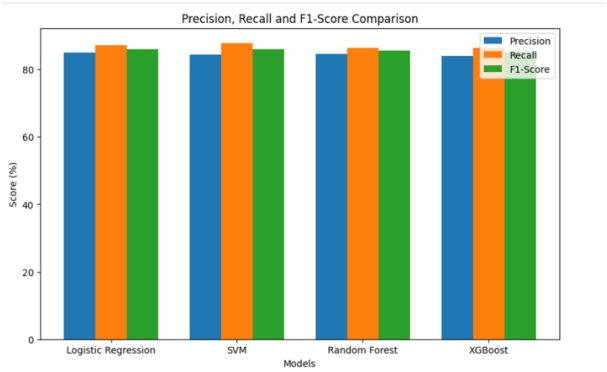

Fig. 11 Logistic Regression and SVM demonstrated competitive performance across precision, recall and F1-score.

The graph above (Figure 11) shows the Precision,Recall and F1 Score of the four machine learning models used.From the above figure 11,it is observed that all the models were compared directly across the evaluation measures,and it is found that all of them have performed competitively.Logistic regression model and SVM model showed better performances than other models in the evaluation metrics because of high precision and recall scores.

A. Discussion

All the algorithms applied for machine learning showed a good performance in predicting the depression cases from the students.But when it comes to performance,Logistic Regression is the best algorithm,because the algorithm has a balanced performance despite all the factors taken into account.The high scores of Precision and Recall indicate that the algorithm is fairly effective in identifying at risk students for depression and in reducing errors in the classification process.This is significant in regard to the kind of psychological difficulty involved;if there is no prediction as to the need for treatment,the consequences of this may be very serious.High values of precision also mean that most of the students that were classified as being at risk of depression are actually experiencing depression.It's a plus point of Logistic Regression,although this algorithm assumes the relations between some parameters of the input data set,which can be accounted for when using a linear decision boundary.Even though they are complicated,random forest and XGBoost were not able to outperform other algorithms.

B. Confusion Matrix Analysis

What's more,another technique applied is estimating the efficiency of the algorithm and that's a confusion matrix,which has been applied for the Logistic Regression model.As it is clear from the confusion matrix,the model is able not only to predict positive,but also negative cases. The low error rate shows that the model can effectively identify people who are experiencing depression problems.To conclude,the results show that it is possible to predict human psychological states through certain characteristics using machine learning techniques.

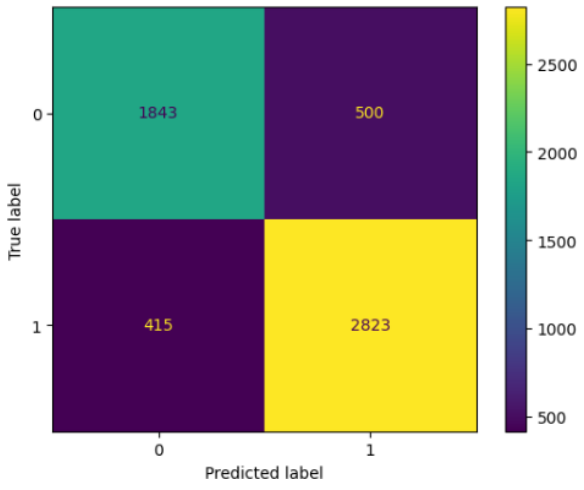

Fig. 12 Confusion Matrix for Logistic Regression

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026
DOI: 10.17148/IJARCCE.2026.15741

Logistic Regression predicted the confusion matrix as displayed below in Figure 12.The Logistic Regression model has correctly classified 1,843 patients who were not depressed,and 2,823 patients who were depressed.It, however,labeled 500 patients that were not depressed as depressed,and 415 patients that were depressed as not depressed. The results above show that Logistic Regression has been able to classify positive and negative patients but there are some relatively few errors.This indicates that Logistic Regression is a good technique to identify depression from students academic and lifestyle variables.

VII.CONCLUSION AND FUTURE SCOPE

In this research paper,we will see the overview of machine learning techniques that can be used to predict depression in a student by analyzing the academic and lifestyle data of the students.For the study in this research paper,the Student Depression Data set obtained from Kaggle is used and four approaches for classification are used like logistic regression,SVM,random forest and XGBoost.For accurate results,it is important to preprocess the selected data set including data cleaning,encoding,normalizing the data.The efficiency of the machine learning methods was measured by using the Accuracy,Precision,Recall and F1Score.The result obtained shows that logistic regression is the most efficient machine learning method with an accuracy rate of 83.61% and SVM is also very efficient with an accuracy rate of 83.53%. The findings of this study suggest that academic stress,financial stress,sleep quality,food consumption,CGPA,background of mental illness in the family of the students are the main factors affecting the mental wellbeing of students studied.It is also clear that the application of machine learning in order to predict depression in students is quite successful.This model forms the foundation of an efficient system that can be used to examine the health of the students and determine those that need interventions. If the symptoms of depression can be detected at an early stage,there will be no doubt about the health of the students.The aspects that can be improved are big data,live behavior data,deep learning and XAI technology.

REFERENCES

[1]. Rois, R., Ray, M., Rahman, A., et al. (2021). Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms. Journal of Health Population and Nutrition. https://doi.org/10.1186/S41043-021-00276-5

[2]. A Machine Learning Study to Predict Anxiety on Campuses in Lebanon. (2023). Studies in Health Technology and Informatics. https://doi.org/10.3233/shti230430

[3]. Rezapour, M., & Elmshaeuser, S. K. (2022). Artificial intelligence-based analytics for impacts of COVID-19 and online learning on college students' mental health. PLOS ONE. https://doi.org/10.1371/journal.pone.0276767

[4]. Sharda, D., Bansal, R., Kaur, N., et al. (2025). Student Health Risk Evaluator (SHRE): A Predictive Machine Learning Framework for Assessing Mental Health Risks in Students. Cureus Journal of Computer Science. https://doi.org/10.7759/s44389-025-08088-y

[5]. An explainable machine learning framework for analyzing and predicting mental health problems among university students in Bangladesh. (2026). Scientific Reports/Nature. https://www.nature.com/articles/s41598-026-52351-8

[6]. Nison, P., Vuttipittayamongkol, P., Boonyapuk, P., & Kemavuthanon, K. (2023). A Machine Learning Approach for Depression Screening in College Students Based on Non-Clinical Information. IEEE Conference Proceedings. https://doi.org/10.1109/CyMaEn57228.2023.10051001

[7]. Nayan, M., Uddin, M. A., Hossain, M. I., et al. (2022). Comparison of the Performance of Machine Learning-Based Algorithms for Predicting Depression and Anxiety Among University Students in Bangladesh. Asian Journal of Social Health and Behavior. https://doi.org/10. 4103/shb.shb_38_22

[8]. Chung, J., & Teo, J. (2022). Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. Applied Computational Intelligence and Soft Computing. https://doi.org/10.1155/2022/9970363

[9]. Rahman, H. A., Kwicklis, M., Ottom, M. A., et al. (2023). Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students. Bioengineering. https://doi.org/10.3390/bioengineering10050575

[10]. Rahman, M. A., & Kohli, T. (2024). Mental Health Analysis of International Students Using Machine Learning Techniques. PLOS ONE. https://doi.org/10.1371/journal.pone.0304132

[11]. Morr, C. E., Jammal, M., Bou-Hamad, I., et al. (2024). Predictive Machine Learning Models for Assessing Lebanese University Students' Depression, Anxiety, and Stress During COVID-19. Journal of Primary Care & Community Health. https://doi.org/10.1177/21501319241235588

[12]. Zhang, L., Zhao, S., Yang, Z., Zheng, H., & Lei, M. (2024). An Artificial Intelligence Tool to Assess the Risk of Severe Mental Distress Among College Students in Terms of Demographics, Eating Habits, Lifestyles, and Sport Habits. BMC Psychiatry. https://doi.org/10.1186/s12888-024-06017-2

[13]. Iparraguirre-Villanueva, O., Paulino-Moreno, C., Epifanía-Huerta, A., & Torres-Ceclén, C. (2024). Machine Learning Models to Classify and Predict Depression in College Students. International Journal of Interactive Mobile Technologies. https://doi.org/10.3991/ijim.v18i14.48669

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15741

[14]. Chung, J., & Teo, J. (2023). Single Classifier vs. Ensemble Machine Learning Approaches for Mental Health Prediction. Brain Informatics. https://doi.org/10.1186/s40708-022-00180-6

[15]. Schaab, B. L., Calvetti, P. Ü., Hoffmann, S., et al. (2024). How Do Machine Learning Models Perform in the Detection of Depression, Anxiety, and Stress Among Undergraduate Students? A Systematic Review. Cadernos de Saúde Pública. https://doi.org/10.1590/0102-311xen029323