SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I8P116

Volume 12 Issue 8, 176-187, August 2025


*Original Article*

# EDUSTACK-MH: An Intelligent Ensemble-Based Model for University Student Depression Screening

Ruby Faizan[1], Saurabh Dhyani[2], Divya Pandey[3], Gaurav Choudhary[4]

[1,2,3,4]*Uttaranchal School of Computing Sciences, Uttaranchal University, Dehradun, Uttarakhand, India.*

[1]*Corresponding Author : faizan.ruby@gmail.com*

Received: 09 June 2025 Revised: 10 July 2025 Accepted: 11 August 2025 Published: 30 August 2025

***Abstract** - The increasing mental health crisis amongst university students has made early detection of depression a pressing concern. Despite being clinically validated, the current diagnostic tools are frequently subjective, time-consuming, and have limited scalability. This study presents EDUSTACK-MH, a novel ensemble-based machine learning framework that uses structured data from the DASS-21 scale and sociodemographic parameters to identify students' depression severity. Data were collected from 1,120 university students in Dehradun, India, and preprocessed to build predictive models capable of identifying five levels of depression severity-ranging from normal to extremely severe. A comparative analysis of baseline classifiers (KNN, SVC, Random Forest, Gradient Boosting, Logistic Regression) and ensemble approaches (Optimized Voting and Stacking) demonstrated that the Optimized Stacking model outperformed all others, achieving 98.23% accuracy with near-perfect classification performance. The findings demonstrate that early mental health screening and intervention techniques in educational environments can be significantly enhanced by combining intelligent ensemble learning techniques integrated with validated psychological tests. By providing a dependable, scalable, and decent method for determining the full extent of depression, this study addresses a significant gap and has been specifically developed to meet the needs of Indian college students.*

***Keywords** - Depression, Students, DASS-21, Machine Learning, SVM, Stacking, Voting, Mental Health.*

## 1. Introduction

The high rate of depression among college students has an adverse impact on each person's academic progress, interpersonal communication, and physical health [1]. One in eight individuals globally struggles with a mental illness [2]. Persistent sadness for at least two weeks is a diagnostic feature of depression, a psychological disease. It makes it difficult to carry out daily tasks, and people who are depressed lose interest in and enjoyment from the things they normally like [3]. University students are as susceptible as the general public to suffering from mental illness, and the prevalence and severity of these conditions seem to be rising [4]. Depression results in pessimistic thinking, diminished focus at work, and lower productivity. It also has an impact on the human reproductive system. Early mental illness detection enhances the quality of life for the patient and their family by enabling therapy to begin earlier [5]. It has a profound psychological impact on individuals. Depression will significantly impair a person's capacity to focus and learn, as well as their ability to work efficiently, all of which will have a significant impact on their lives. Five mental illnesses make up the top 10 major diseases that incapacitate or impair people worldwide, with depression coming in first [6]. Due to the unique problems they face, such as academic stress, social isolation, financial difficulties, and the transition to adulthood, university students are more susceptible to mental health concerns. Depression can have serious consequences if left untreated, including academic failure, dropout, and even more serious outcomes like suicide or self-harm. Around 264 million people worldwide suffer from depression [7]. ML has been used in various fields of psychological interventions and holds great potential for the prediction and control of mental health disorders and other similar health outcomes [8]. The latest developments in ML provide promising alternatives to conventional statistical methods, the ability to process voluminous amounts of data with high dimensions and reveal complex patterns that can suggest the occurrence of psychiatric disorders [9]. Machine learning applications have the incredible capacity to scan enormous amounts of data, find patterns and predict results [10]. Medical professionals can better anticipate the occurrence of mental diseases and guarantee successful treatment outcomes by using ML algorithms to uncover potential behavioural biomarkers. The algorithms have an advantage in interpreting and displaying complex medical data. The visualization assists in deriving an effective hypothesis about the diagnosis of a mental disorder [11]

### *1.1. Research Gap*

While traditional methods of diagnosing depression rely heavily on manual assessments and questionnaires, Machine Learning (ML), despite its potential, still has a notable gap in current research. Many existing studies have focused only on binary classification-labelling students as either depressed or not depressed. Others have relied on unstructured data like social media content, which may not be reliable or easily applicable in institutional settings. What's missing are robust and accurate models that can identify different levels of depression severity, particularly using well-established, structured tools like the DASS-21 questionnaire, and focusing on Indian university students, a group often overlooked in global datasets. To bridge this gap, our study introduces EDUSTACK-MH-an intelligent ensemble-based machine learning framework designed to detect and classify depression in students across five severity levels. By assembling psychological data (from DASS-21), along with advanced ML techniques like stacking and voting, the model aims to provide more accurate, reliable, and practical solutions for early mental health screening in educational institutions

### *1.2. Problem of Statement*

Mental health issues among university students are rising, but early detection remains difficult due to traditional methods being slow and subjective. While machine learning offers potential, most existing models only perform binary classification and do not measure depression severity.

Many rely on unstructured data like social media, which is not always practical or ethical for institutional use. There is a lack of robust models that use validated tools like DASS-21 and structured student data. Few studies focus on Indian students or use optimized ensemble techniques for better accuracy. A reliable, multiclass prediction system is urgently needed to support timely intervention and better mental health outcomes in academic settings. Our goal is not just to predict whether a student is struggling, but how severely, so that timely and appropriate interventions can be offered. The results show that our optimized ensemble approach significantly outperforms traditional ML models, achieving a remarkable 98.23% accuracy-proving its potential to impact student mental health care.

## 2. Related Work

This study investigates depression among Bangladeshi university undergraduates, identifying key contributing factors through a psychologist-informed survey. It aims to predict depression early for timely psychiatric intervention. Among the three tested models, Random Forest outperformed the others with around 75% accuracy and lower false negatives. The research emphasizes early detection to prevent severe outcomes like suicide [12]. This study identified depression by data mining and machine learning techniques applied to social media posts, comments, and texts. Six classifiers were employed to determine if users were depressed or not depressed. Among these, the Support Vector Machine (SVM) was the most accurate, with an accuracy of 75.15% compared to the random forest having an accuracy of 69.97%. The strategy highlights the importance of social media analysis in the early detection of depression [13]. Ensemble techniques for better accuracy. A reliable, multiclass.

With the use of DASS-21 data gathered from people who are working and those who are not, researchers applied five different machine learning algorithms to predict levels of anxiety, depression, and stress. These predictions were categorized into five levels of severity. Among the models tested, Naïve Bayes emerged as the most accurate in identifying mental health conditions, the 78% accuracy. Due to an imbalance in the dataset-where some severity levels had significantly more samples than others-the F1 score was chosen as the primary metric to evaluate the models' performance better. Additionally, a specificity analysis showed that the models were particularly good at correctly identifying individuals who were not experiencing significant mental health issues (the negative class) [14]. In order to predict depressive symptoms, this study uses information from 138 college students' smartphones and fitness trackers. Post-semester depression was detected with 85.7% accuracy. The strategy emphasizes how longitudinal behavioural data can be used to avoid mental health issues [15]. In order to determine significant depression risk factors among 10,043 Chinese undergraduate students, this study used the Random Forest Algorithm. Academic stress, BMI, and physical fitness metrics were the next most significant indicators, followed by suicidal thoughts, anxiety, and substance use. The model's accuracy was 87.5%, and its AUC was 0.927. There were differences between the sexes, with men placing greater value on physical fitness and women on BMI. The results highlight the value of gender sensitive mental health treatments and more long-term research [16].

This study aimed to predict mental illness using machine learning algorithms in 2121 Bangladeshi university students. An online survey that combined behavioural and sociodemographic evaluations, such as the PHQ-9 and GAD-7, was used to gather data. Random Forest had the highest accuracy (89%) in predicting depression. According to the study, RF and SVM are useful tools for accurately predicting mental health in academic settings [17]. This work targets rising depression rates among university students by integrating the GHQ12 questionnaire and machine learning techniques. Data was collected from 804 students at Bangladeshi universities, and the results revealed that more than 60% were depressed. The questionnaire had psychological, sociodemographic, and career-related items developed with expert psychiatric assistance. Sixteen machine learning models were evaluated in order to discover important predictors of depression. The Extremely

Randomized Tree model had the highest accuracy (90.26%). The focus of the study emphasizes the machine learning potentials in early depression diagnosis, as well as the impact of socioprofessional obligations on student mental health [18]. This study includes techniques of machine learning to forecast risk factors for depression and anxiety among 3984 schoolchildren aged 10–15 in the West Bank. Five models were tested, along with SVM reaching (over 92%) the highest accuracy for both conditions, followed by Random Forest. Key risk factors included school bullying, home violence, academic performance, and family income. The findings suggest machine learning can effectively support school mental health interventions [19]. For the purpose of forecasting the occurrence of depression in the elderly, five ML methods were evaluated: Multilayer Perceptron (MLP), Decision Tree, Logistic Regression, Bayesian Networks and Sequential Minimal Optimization (SMO). They discovered that Bayesian networks and Decision Trees produced the highest results (Accuracy = 95.00%, Precision = 95.00%), with the Network additionally reaching a high ROC Area value of 99.00%. They concluded that the prediction model improves the effectiveness of the treatment procedure for health specialists [20]. This study used a YMM dataset5 to determine the elements impacting depression in children and adolescents. Its goals are to find the optimal machine learning model, identify depressed patients, and comprehend contributing elements. They employed Weighted Voting7, Gradient Boosting, KNN, Bagging, and AdaBoost. When utilizing Select K best, AdaBoost achieved the best Accuracy (92.56%), Precision (95.77%), and F1Score (93.79%), indicating that variable selection strategies enhance the efficacy of models and the accuracy of detection [21]. In order to predict sadness in older persons who live alone, this study created a machine learning system. They used algorithms such as Logit, Boosted Tree and RF with the best evaluation criteria (Precision = 92.90%, Accuracy = 91.10% and AUC = 96.00%). The Logit model was found to be useful in identifying underdiagnosed subgroups for suitably treated care [22].

M. M. T. Ayyalasomayajula (2024) analysed Reddit posts using an NLP and ML model, reaching an F1 score of about 70%**,** showing stronger performance than earlier baseline approaches. This highlights its potential for detecting depression from Reddit posts by combining text features with posting patterns [23]. Hasan et al (2024) took a different approach by analysing the impact of mobile gaming addiction on the mental health of Bangladeshi children using SVM, Naïve Bayes, KNN, and Random Forest. The accuracy of their study was 92.75% with SVM, but it was limited to children and did not address depression severity levels [24]. This study created a user-friendly web tool that uses machine learning to spot signs of depression and possible suicide risk based on people's background information. By using a smart combination of algorithms (AdaBoost with Extra Trees) and balancing the data, the system reached 78.69% accuracy for detecting depression and 90.60% for suicide attempts. It shows how technology can genuinely help identify those who are struggling and guide them toward timely support, all while being mindful of ethical concerns [25]. This study explores how socioeconomic factors contribute to depression and uses machine learning to detect it early. A custom questionnaire and feature selection techniques helped identify key risk indicators. The Random Forest model performed best, achieving 96.85% correctness, demonstrating it is a strong tool for proactive support in mental health [26].

## 3. Methods: Participant and Procedure

### *3.1. Questionnaire*

A structured questionnaire was created to gather information from private university students in order to evaluate their mental health, with a particular emphasis on symptoms of depression. Sociodemographic factors like gender, age, relationship status, academic course, and substance usage were all included in the tool A Likert scale with four points: 0 for "did not apply to me at all" and 3 for "applied to me very much most of the times", was also used to rate seven self-reported items (Q-3, Q-5, Q1-0, Q-13, Q-16, Q-17, and Q-21) that make up the Depression subscale. These questions assessed the participants' feelings of dejection, unhappiness, and low self-esteem. 1120 student answers in all were gathered, creating the dataset that was utilized for analysis and predictive modelling in this study

### *3.2. Dataset*

A total of 1,200 responses were gathered through the university. After preprocessing, 1,120 valid responses were retained for analysis. The survey was disseminated extensively via academic forums, social media, and university networks. Informed consent was obtained, participation was voluntary, and anonymity was maintained. One thousand one hundred twenty computer engineering and management students from a private university in Dehradun, Uttarakhand, are included in the dataset. Every student filled out a questionnaire based on the DASS-21 scale, which is used to measure depression symptoms only. Across the world, people use the Depression, Anxiety, and Stress Scale21 (DASS21) to measure their symptoms of stress, anxiety, and depression [27]. The sociodemographic details of the 1200 university students who took part are shown in Table 1. The primary sociodemographic features of the 1120 college students who took part in the study are listed in this table. Age Groups (18–20, 21–24, 25-30), Relationship Status (Single, Married, In a Relationship), and Gender (Female, Male, Transgender) are among the variables. The participants are pursuing various academic programs, including BCA, BBA, MBA, BSc (IT), and MCA. The frequency of substance usage is also documented in the table under headings like "Do Not Apply on Me," "Sometimes," "Often," and "Most of the Time". The table offers a thorough demographic framework for examining the study's mental health trends.

**Table 1. The participants' sociodemographic attributes (n = 1120)**

| Variable | Group | Values |
|---|---|---|
| Gender | Female<br>Male<br>Transgender | Selection |
| Age | 18–20<br>21–24<br>25-30 | Selection |
| Relationship Status | Single<br>Married<br>Relationship | Selection |
| Course | BCA,<br>BBA<br>MBA<br>BSC(IT)<br>MCA | Selection |
| Substances Used | Do Not Apply to me<br>Sometime<br>Often<br>Most of the time | Categorical |

Table 1 presents characteristics and explanations of the sociodemographic Questions among college students, consisting of 1120 entries. Table 2 describes the DASS-21 self-reported items, divided into three psychological dimensions-stress, anxiety, and depression- which are the basis for the dataset used in this study. Only the seven items connected to depression were taken into consideration for this study in order to evaluate university students' mental health.

During Q-3, I experienced a complete lack of motivation to do anything. During Q-5, I struggled to find anything to get excited about. During Q-10, I thought there was nothing to anticipate. During Q-13, I was depressed and hopeless. During Q-16, I still couldn't find anything to get excited about. Lastly, during Q-17, I felt like my life did not have much purpose.

A four-point Likert scale measuring from 0 (did not apply to me at all) to 3 (applies to me very often or most of the time) was used to rate each question. The depression score, which was the main factor in assessing and forecasting depressive symptoms among people, was calculated using these particular indicators.

**Table 2. Attributes and details of the Student University DASS 21 dataset (n =1120)**

| Cod. | Attribute | Description | Data Type | Values |
|---|---|---|---|---|
| **Q1A** | Q1(s) | I found it hard to wind down | Categorical | 0,1,2,3 |
| **Q2A** | Q2(a) | I was aware of the dryness of my mouth | Categorical | 0,1,2,3, |
| **Q3A** | Q3(d) | I could not seem to experience any positive feelings at all | Categorical | 0,1,2,3 |
| **Q4A** | Q4(a) | I Experienced Breathing difficulty. | Categorical | 0,1,2,3, |
| **Q5A** | Q5(d) | I found it difficult to work up the initiative to do things. | Categorical | 0,1,2,3 |
| **Q6A** | Q6(s) | I tended to overreact to situations | Categorical | 0,1,2,3 |
| **Q7A** | Q7(a) | I experienced trembling in my hands. | Categorical | 0,1,2,3 |
| **Q8A** | Q8(s) | I felt that I was using much nervous energy. | Categorical | 0,1,2,3 |
| **Q9A** | Q9(a) | I was worried about situations in which I might panic. | Categorical | 0,1,2,3, |
| **Q10A** | Q10(d) | I felt that I had nothing to look forward to. | Categorical | 0,1,2,3 |
| **Q11A** | Q11(s) | I found myself getting agitated. | Categorical | 0,1,2,3 |
| **Q12A** | Q12(s) | I found it difficult to relax. | Categorical | 0,1,2,3 |
| **Q13A** | Q13(d) | I felt downhearted and blue | Categorical | 0,1,2,3 |
| **Q14A** | Q14(s) | I was intolerant of anything. | Categorical | 0,1,2,3 |
| **Q15A** | Q15(a) | I felt I was close to panic. | Categorical | 0,1,2,3 |
| **Q16A** | Q16(d) | I was unable to become enthusiastic about anything. | Categorical | 0,1,2,3 |
| **Q17A** | Q17(d) | I felt I was not worth much as a person. | Categorical | 0,1,2,3 |
| **Q18A** | Q18(s) | I felt that I was rather touchy. | Categorical | 0,1,2,3 |
| **Q19A** | Q19(a) | I was aware of the action of my heart in the absence of physical exertion. | Categorical | 0,1,2,3 |
| **Q20A** | Q20(a) | I felt scared without any good Reason. | Categorical | 0,1,2,3, |
| **Q21A** | Q21(d) | I felt that life was meaningless. | Categorical | 0,1,2,3 |

### 3.3. Data Analysis

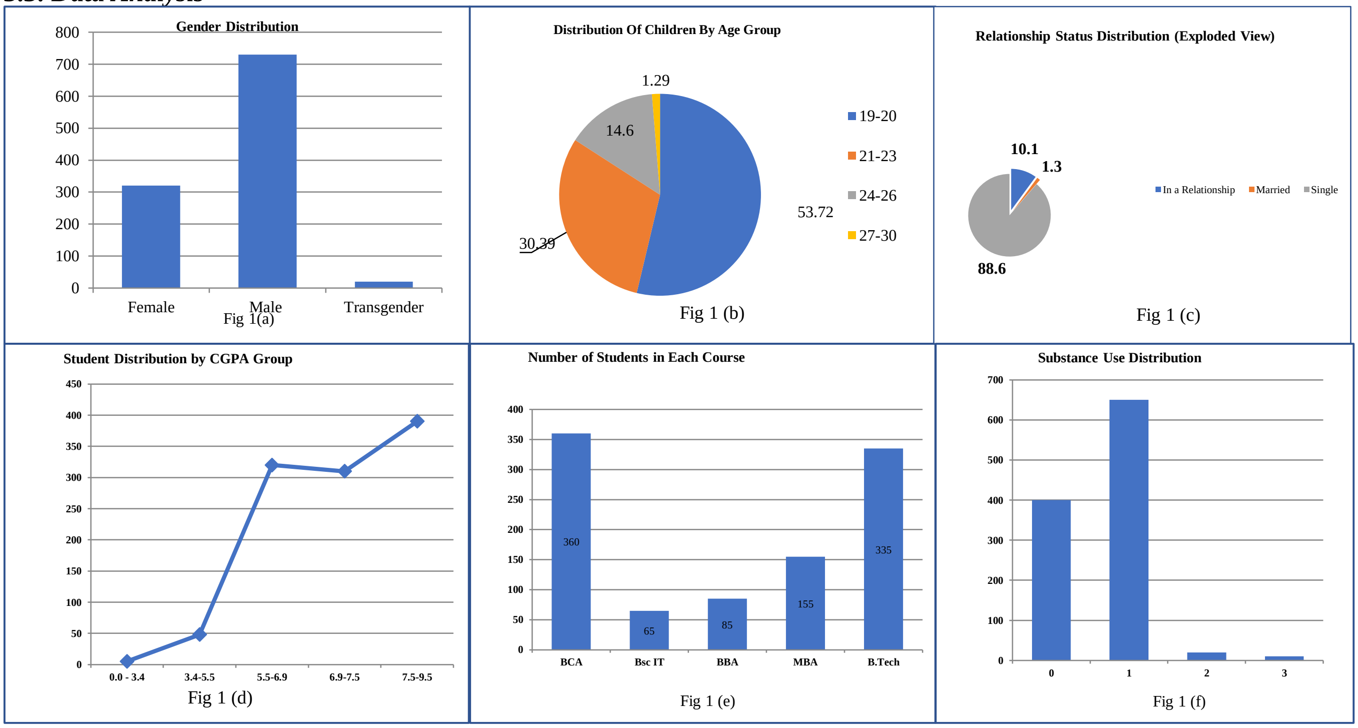


**Fig. 1 Sociodemographic data analysis**

Descriptive demographic, lifestyle, and well-being data analyses were among the preliminary analyses. All variables' frequency distributions were looked at in the preliminary analysis. Since this was an exploratory study, post hoc analysis was performed to look into the high frequency of participant-reported drowsiness [28]. A thorough overview of the academic and sociodemographic characteristics of the student participants is given in Figure 1. With male students making up the majority, female students come in second, and a very small minority identify as other Figure 1(a) demonstrates a stark gender imbalance. Figure1(b) represents the age-wise distribution of students. The age group majority, 53.72%, falls in the 19–20 range, followed by 30.39% in the 21–23 range and 14.6% in the 24– 26 range. A very small proportion (about 1.29%) is aged 27 or above. Figure1(c) displays the relationship status of students. The chart shows that a majority (88.6%) are single, while only a small portion are in a relationship (8.4%), and an even smaller fraction are married (3%). This suggests that most students are not in a committed relationship, which may influence their emotional and mental well-being differently compared to those who are. Figure 1(d) shows the distribution of students across CGPA ranges. The graph clearly highlights that the majority of students fall within the higher CGPA brackets, particularly between 3.5 and 4.0, with over 400 students in this range. The next largest group is in the 3.0–3.5 range, with slightly fewer students, followed by a sharp drop in the number of students in the 2.5–3.0 and 2.0–2.5 ranges. The course-wise distribution is shown in Figure 1(e), which shows that the majority of responders (53.72%) are in the engineering field, followed by management (30.39%) and computer science (14.6%). Figure 1(f) shows the course distribution chart. Out of all the courses, most students are enrolled in BCA (350 students) and B.Tech (344 students). With 158 students, MBA comes in second, with comparatively lower participation rates for BBA (98), MCA (94), and BScIT (75). This suggests that students with technical backgrounds are overrepresented in the dataset. Figure 1(f) depicts the distribution of substance use among students. The majority report either occasional (1) or no use (0) of substances like tobacco or alcohol, while frequent use (2 or 3) is very low, indicating limited substance dependency in the sample.

Figure 1 provides an overview of the academic and sociodemographic characteristics of students. Most participants are female, aged 21–23, and single, indicating a young and predominantly unpartnered population. Academically, students generally perform well, with the majority having a CGPA above 3.0. The sample is primarily composed of engineering and management students. Additionally, substance use is low, with most students reporting no or occasional use of tobacco or alcohol.

### 3.4. Data Preprocessing

Data preparation was an essential step in this investigation to ensure that the analysis was accurate and of high quality. Our next step is to put the dataset into a readable format for machine learning [29]. To make sure raw data is suitable for machine learning applications, the data preparation process involves cleaning, organizing, and structuring it. Data cleaning includes identifying or eliminating outliers, smoothing noisy data, filling in missing values, and resolving inconsistencies [30].

The label encoder function has been used to encode the data set into relevant data and features. To prepare material for statistical modeling, categorical responses were translated into numerical values [31]. Every question was given a variable name, such as Q-1A, Q-2A, Q-3A, …Q-7A. The sum of the scores for each of the seven categories is used to calculate the overall score for each state (such as depression). With the options "never," "occasionally," "often," and "nearly often" (scoring 0, 1, 2, and 3, respectively), the user can indicate the degree to which their current situation resembles that of the previous week in response to the experience statements provided in each question [32]. The severity of each circumstance is categorized as none, mild, moderate, severe, or extremely severe based on the threshold scores associated with each stage. Additionally, for the objectives of this study, the individuals are separated into two classes: rest (high depression) and none or moderate (low depression).

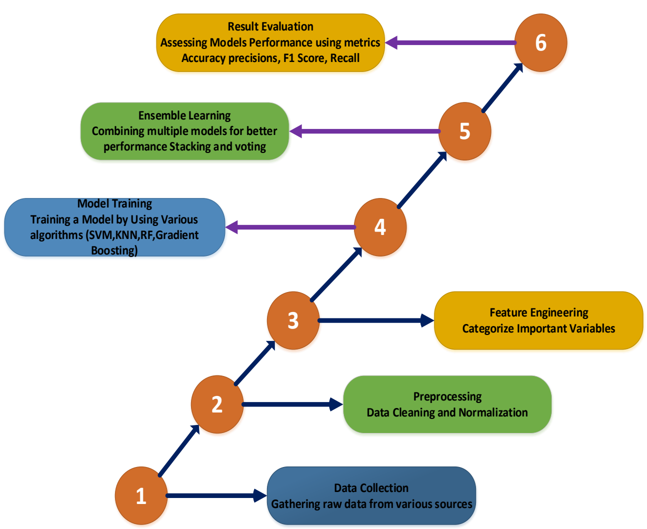


**Fig. 2 EDUSTACK-MH MODEL**

Using sociodemographic information and DASS-21 scores, the model predicts 1,120 college students' stress, anxiety, and depression levels. Out of the five machine learning algorithms used, Random Forest demonstrated the highest accuracy. Cross-validation and the F1 score were employed to guarantee a balanced assessment, particularly in light of the class imbalance.

### *3.5. Factor Analysis*

Researchers can use correlational analysis to help formulate complex real-world relationships [33]. Examining DASS21 models that represent the field's changing conceptualisations of psychopathology is crucial. Significant inter-item correlations were found in the factor analysis of the DASS21 questionnaire responses, as the heatmap produced from the correlation matrix in Figure 3 illustrates. Positive correlations between the Depression Score and each of the seven items (Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, and Q21A) validate their respective contributions to the total assessment of depressive symptoms. With the highest correlation of 0.68 among these, Q17A ("I thought that life was pointless") is a highly relevant item in evaluating depression tendencies. Q10A and Q13A come in second and third, respectively, with a correlation of 0.63, indicating their significance in assessing mental health. A moderately positive correlation exists between the overall score and even the lowest correlation value, Q16A, at 0.50. The depression scale's internal structure is validated by the consistent pattern of moderate to strong correlations across all items, which also highlights the scale's efficacy and dependability for use in student mental health screening. The Bartlett's Test of Sphericity confirmed the dataset's eligibility for factor analysis, obtaining a chi-square value of 5233.6317 with a p-value of 0.02, demonstrating that an identity matrix is not the same as the population correlation.

Furthermore, the Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy confirmed this conclusion by producing a high overall KMO value (anticipated to be greater than 0.6), showing that the data structure was suitable for extracting underlying components. These findings support the DASS-21 scale's validity in assessing mental health aspects among surveyed participants.

Bartlett’s Test Results:
Chi-square value: 5233.6317
p-value: 0.002
Kaiser-Meyer-Olkin (KMO) Test Result: KMO Model value: 0.9061914566968292

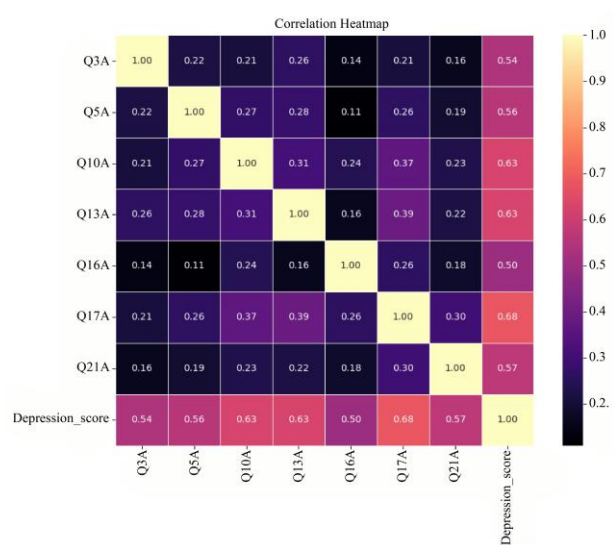


**Fig. 3 Heatmap of depression questions**

### 3.6. Feature Engineering

Finding the most crucial information in the original dataset is the aim of feature engineering, and reducing it to variables that may be used by machine learning algorithms [34]. The DASS-21 Questionnaire's feature engineering consists of 21 questions, seven of which will be on depression. The DASS-21 scale will be used to determine the answers to these questions, including:

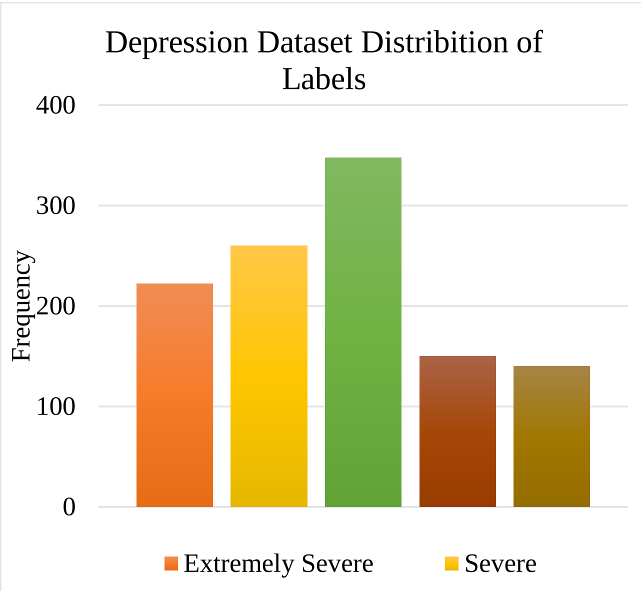


**Fig. 4 Depression Data Set Distribution**

Total Depression_count = SUM (Q3A, Q5A, Q10A, QA13, QA16, Q17A, QA21) *2

Using the DASS21 approach, the total depression score was calculated by multiplying the sum of the scores from questions Q1A, Q3A, Q7A, Q10A, and Q13A by two. Figure 4 shows that the participants' depression severity scores were used to categorize them into five groups. The most prevalent category in the examined dataset was moderate depression, which was found in 347 students (31.0%). Furthermore, 222 pupils (19.8%) were categorized as extremely severe, and 260 students (23.2%) as severe. There were 150 students (13.4%) in the mild category and 140 students (12.5%) in the normal range. With more than 74% of students reporting at least moderate degrees of despair, this distribution highlights a serious mental health issue.

### 3.7. Model Evaluation

As multiple supervised learning techniques are appropriate for creating reliable predictive models, we used them [35]. Based on students' psychological information, Figure 2 presents the EDUSTACK-MH Model. It is a customized machine learning framework designed to forecast and assess mental health issues like stress, anxiety, and depression. To increase prediction accuracy and dependability, this model combines sophisticated ensemble approaches with conventional machine learning workflows. A thorough, data-driven framework called the EDUSTACK-MH Model was created to forecast mental health problems among college students. A cutting-edge idea in the field of health science predictive modelling is machine learning technology/Data collection and preparation are the first steps of a six-phase pipeline that includes feature engineering, model training with algorithms like SVM, Random Forest, and Logistic Regression, and improvement with ensemble methods like stacking and voting [36]. Using performance criteria including accuracy, precision, recall, and F1score, the model ends with an evaluation of the results. This strategy guarantees a reliable and scalable solution for early mental health intervention and identification in educational settings.

*Algorithm*

The algorithm first iterates over all data sources, where for each i, data source[i] is loaded and appended to the raw data. Next, for each record j in raw data, preprocessing is carried out to handle missing values and normalize features, and the result is stored in cleaned data[j]. Then, for each j, the depression score is computed as the sum of the selected DASS-21 items (Q3, Q5, Q10, Q13, Q16, Q17, Q21) and stored in cleaned data[j][i]'Depression Score']. The data is split into training and test subsets. For each base model i, the model is trained on X_train and y_train and stored in trained models[i]. For each test instance j, predictions from all the trained models are obtained, and a majority vote offers vote_preds[j]. In stacking, each training instance j is mapped to a meta_train[j] vector of predictions from all the base models. A meta-model is trained on these vectors. Likewise, each test instance j is utilized to create meta_test[j], and the final stacked prediction is calculated as stack_preds[j] with the help of the meta-model.

Step 1: for i =1 to n
Rawdata_append (load (data_source [ i]
Step 2: for j=1 to length (Rawdata)
data_cleaned [ j] = preprocess (Rawdata[j])
Step 3: for j=1 to length (data_cleaned)
Data_cleaned[j][‘Depression_Score] = sum ([Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21])
Step 4: Split data_cleaned into X_train, Y_train, X_test, Y_test
Step 5: for i=1 to K
Trained_model[i]=base_model[i]. train(X_train, y_train)
Step 6: for j=1 to length(X_test)
Vote_pred[j]=majoritity([model_predict(X_test[j]) for model n trained_models])
Step 7: for j= 1 to length (X_train)
Meta_train[j]=[model_predict[X_train[j]) for model in trained_model
Step 8: Meta_model. train (meta_train, Y_train)
Step 9: for j=1 to length (X_test)
Meta_test[j]= [model_predict(X_test[j]) for model in trained_model]
Step 10: for j=1 to length(meta_test)
Stack_predict [j]= meta_model_predict(meta_test[j])

### *3.8. Results*

This section presents the results, highlighting trends and disclosures that the machine learning models extracted from the data. It thoroughly analyses the findings based on several evaluation criteria, including accuracy, recall, precision, and F1-score, and it discusses the mental health conditions and their severity [37]. Table 3 shows the evaluation metrics, which comprehensively understand each model's classification effectiveness.

Support Vector Classifier (SVC) and Logistic Regression were the two best-performing individual models; SVC had an accuracy of 96.03% and an F1 score of 97.12%, while Logistic Regression fared somewhat better with an accuracy of 97.40% and an F1 score of 97.24%. On the other hand, models with accuracies between 73.20% and 76.01% and lower F1 scores, like KNN, Random Forest, and Gradient Boosting, showed moderate to poor performance. Ensemble approaches were used to improve overall forecast accuracy and resilience. By using soft voting to combine high-performing models, the Optimized Voting Classifier achieved 97.03% accuracy and 97.25% recall, demonstrating its efficacy in merging complementary predictions.

Notably, the Optimized Stacking Classifier achieved 98.24% accuracy, 98.08% precision, 98.00% recall, and an F1 score of 97.80%, outperforming all other methods.

**Table 3. EDUSTACK-MH model results**

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) |
|---|---|---|---|---|
| KNN | 76.01 | 75.03 | 77.02 | 77.03 |
| SVC | 96.03 | 97.01 | 97.12 | 97.12 |
| Random Forest | 76.01 | 76.10 | 75.12 | 76.01 |
| Gradient Boosting | 73.20 | 74.18 | 73.02 | 73.0 |
| Logistic Regression | 97.40 | 96.32 | 97.18 | 97.24 |
| Optimized Stacking | 98.23 | 98.08 | 98.00 | 97.800 |
| Optimized voting | 97.03 | 96.02 | 97.25 | 96.25 |

This notable improvement in performance demonstrates the effectiveness of ensemble-based architectures in mental health prediction challenges, especially when it comes to identifying subtle patterns across different degrees of depression severity.

### *3.9. Ensemble Learning*

Ensemble methods demonstrated significant effectiveness, especially the Optimized Stacking Classifier, which outperformed all other models with a perfect classification performance: 98.23% across all metrics. This suggests a high degree of synergy between the base models (Logistic Regression and SVC) and the meta-learner (Gradient Boosting), enabling the model to generalize exceptionally well over the testing set. While not as high-performing as stacking, the Optimized Voting Classifier still delivered strong results with 97.00% accuracy and high precision and recall values. This indicates that an intelligent combination of top-performing models can significantly enhance predictive performance even without a learning meta-layer.

These results validate the utility of ensemble learning in student mental health prediction and highlight the potential of stacking techniques for building robust and highly accurate classification systems. All five categories of depression severity-Very Severe, Severe, Moderate, Mild, and Normal-show excellent classification performance in the confusion matrix derived from the optimal voting ensemble model, as shown in Figure 5. With just two incorrect classifications into the Normal group, the program correctly identified 67 people as Extremely Severe. 43 people were correctly classified as being in the Severe category, with one case each of Extremely Severe and Moderate being incorrectly labeled. With 86 accurate predictions and little confusion, just one case was incorrectly labeled as severe, and two as normal.

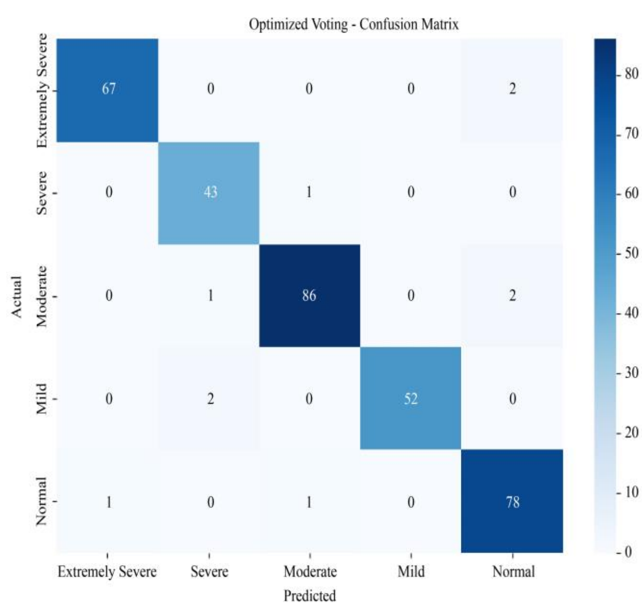


**Fig. 5 Voting classifier confusion matrix**

The Moderate category demonstrated exceptionally high accuracy. In a similar vein, only two people were incorrectly labeled as severe in the Mild category, which had 52 valid classifications. The Normal group had 78 correct predictions, whereas the Extremely Severe and Moderate categories each had one misclassification. The confusion matrix shows how reliable and strong the optimal voting strategy is at differentiating between depression severity levels.

Due to overlapping symptom patterns in borderline cases, it is expected that the majority of misclassifications occurred between adjacent categories, such as Severe and Moderate or Moderate and Normal.

The model had little confusion between nonadjacent classes, indicating that it successfully distinguishes between markedly different severity levels. These results demonstrate the ensemble method's potential for accurate mental health assessment, especially in clinical or educational settings where prompt action depends on early and precise diagnosis.

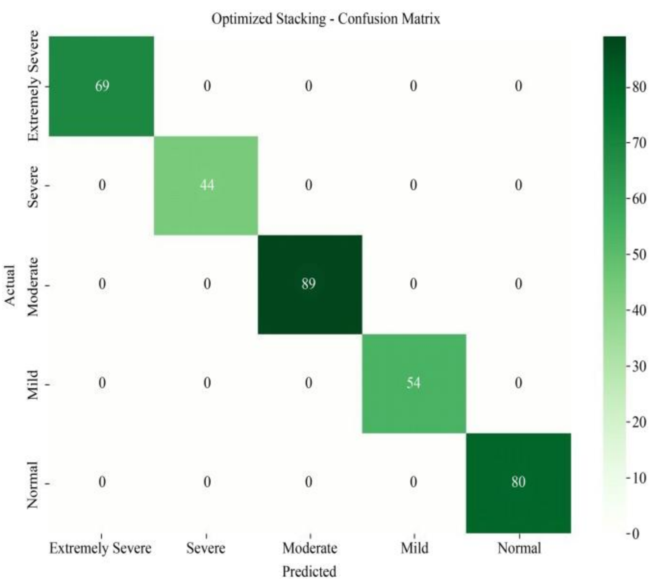


**Fig. 6 Stacking Classifier Confusion Matrix**

All five levels of depression severity, Very Severe, Severe, Moderate, Mild, and Normal, in Figure 6 show perfect classification performance in the confusion matrix for the optimal stacking ensemble model.

The model accurately classified every case in every category, with no misclassifications. In particular, the model produced a precisely diagonal matrix by correctly predicting 69 cases of Extremely Severe, 44 cases of Severe, 89 Moderate, 54 Mild, and 80 Normal instances This flawless alignment shows that the stacking model maintained zero misunderstanding between classes in addition to achieving excellent accuracy, which is especially important in clinical settings where precise mental health assessment is critical.

This degree of accuracy shows that the stacking ensemble successfully combines the advantages of its metaclassifier and base learners to provide extremely dependable predictions. Even in situations when symptoms may overlap, the model's capacity to clearly distinguish between adjacent depression severity levels is validated by the lack of any off-diagonal values.

This performance demonstrates the optimal stacking approach's resilience and possible use in real-world mental health monitoring systems, where accurate psychological state classification is essential for the right kind of care and intervention

### *3.10. Comparison*

When comparing existing studies in the field of depression detection, it is clear that most have either focused on binary classification-simply identifying whether someone is depressed or not-or used unstructured data sources such as social media, which come with privacy concerns and limited real-world application in educational settings.

For instance, Gupta & Kumar (2023) achieved good accuracy (91%) using Reddit posts, but their approach is not practical for institutional use due to ethical and data quality issues.

Similarly, Hasan et al. (2024) explored the link between mobile gaming and depression among children and achieved 92.75% accuracy, yet the study was limited to a younger demographic and did not address the severity of depression.

On the other hand, some structured studies, like the AdaBoost web tool project, attempted to predict depression based on background data but fell short with just 78.69% accuracy and no classification of severity levels.

In contrast, this study-EDUSTACK-MH-takes a comprehensive and clinically relevant approach. It uses structured and validated data (DASS-21), applies advanced ensemble techniques like optimized stacking and voting, and most importantly, classifies students across five levels of depression severity, not just a binary label.

With a standout accuracy of 98.23%, your model not only outperforms others in terms of raw performance but also offers greater depth, practical scalability, and precision in identifying mental health needs in an academic context. This makes your work statistically superior and substantially more impactful and deployable in real-world university settings.

**Table 4. Summary of comparative analysis**

| Study | Tool/Scale Used | Classification Type | Model Used | Accuracy |
|---|---|---|---|---|
| A. A. Choudhury [12] | Custom Psych Survey | Binary | Random Forest | 75% |
| Z. N. Vasha [13] | Text-Based Data | Binary | SVM, RF | 75.15% (SVM) 69.97%(RF) |
| A. Priya[14] | DASS-21 | Multiclass (5 Levels) | Naïve Bayes | 78% |
| P. Chikersal *et al.* [15] | Wearables + Mobile | Binary & Trend Detection | Passive Sensing + ML | 85.7% |
| L. Luo *et* al. [16] | Custom Survey | Binary (Risk Factor Detection) | Random Forest | 87.5% |
| I. Haq [17] | PHQ-9 + GAD-7 | Binary | Random Forest | 89% |
| N. Mumenin *et al* [18] | GHQ-12 | Multiclass | Extremely Randomize d Trees | 90.26% |
| R. Qasrawi [19] | Custom Psych Survey (Schoolchildren) | Binary | SVM, RF | 92% |
| A. Sau and I. Bhakta [20] | Elderly Data Patient | Binary | Bayesian Networks, Decision Tree | 95% |
| EDUSTACK-MH (This Study) | DASS-21 + Demographics | Multiclass (5 Levels) | Optimized Stacking & Voting (SVC + LR + GB) | 98.23% |

## 4. Conclusion

In order to determine the severity of depression among college students, this study created and assessed a novel machine learning framework called EDUSTACK-MH using DASS-21 responses and structured sociodemographic data.

After a thorough comparison of two ensemble strategies and five baseline classifiers, the study concluded that although SVC and Logistic Regression both performed well on their own, the ensemble approaches-in particular, optimized stacking-provided the best results, obtaining perfect classification and 98.24% accuracy in confusion matrix evaluations.

Our method handled multiclass depression intensity and used validated psychological measures, assuring reliability and applicability in contrast to earlier studies that relied on social media data or binary categorization. With little misclassification, the confusion matrix analysis showed good predictive value across all depression severity groups. Our model significantly improves classification accuracy, generalization, and clinical relevance as compared to previous studies. The findings open the door for scalable, data-driven mental health screening systems at educational institutions and highlight the value of ensemble learning in mental health prediction. In order to enable prompt interventions, future research can investigate real-time deployment via mobile applications and integration with counselling support systems.

## References


[1] Xin-Qiao Liu et al., “Influencing Factors, Prediction and Prevention of Depression in College Students: A Literature Review,” *World Journal of Psychiatry*, vol. 12, no. 7, pp. 860-873, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Fauste Ndikumana et al., “Machine Learning-Based Predictive Modelling of Mental Health in Rwandan Youth,” *Scientific Reports*, vol. 15, no. 1, pp. 1-14, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Md. Sabab Zulfiker et al., “An in-Depth Analysis of Machine Learning Approaches to Predict Depression,” *Current Research in Behavioral Sciences*, vol. 2, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Prathamesh Muzumdar, Ganga Prasad Basyal, Piyush Vyas, “An Empirical Comparison of Machine Learning Models for Student’s Mental Health Illness Assessment,” *Asian Journal of Computer and Information Systems*, vol. 10, no. 1, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Purude Vaishali Narayanrao, and P. Lalitha Surya Kumari, “Analysis of Machine Learning Algorithms for Predicting Depression,” *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Gunupur, India, pp. 1-4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yan Ding et al., "A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm," *IEEE Access*, vol. 8, pp. 75616-75629, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Nasirul Mumenin et al., "DDNet : A Robust, and Reliable Hybrid Machine Learning Model for Effective Detection of Depression among University Students," *IEEE Access*, vol. 13, pp. 49334-49353, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] Ngumimi Karen Iyortsuun et al., "A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis," *Healthcare*, vol. 11, no. 3, pp. 1-27, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Christian Janiesch, Patrick Zschech, and Kai Heinrich, "Machine Learning and Deep Learning," *Electronic Markets*, vol. 31, pp. 685-695, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Muhammad Azizur Rahman, and Tripti Kohli, "Mental Health Analysis of International Students Using Machine Learning Techniques," *PLoS One*, vol. 19, no. 6, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Shumaila Aleem et al., "Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions," *Electronics*, vol. 11, no. 7, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Ahnaf Atef Choudhury et al., "Predicting Depression in Bangladeshi Undergraduates using Machine Learning," *2019 IEEE Region 10 Symposium (TENSYMP)*, Kolkata, India, pp. 789-794, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Zannatun Nayem Vasha et al., "Depression Detection in Social Media Comments Data Using Machine Learning Algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 987-996, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Anu Priya, Shruti Garg, and Neha Prerna Tigga, "Predicting Anxiety, Depression and Stress in Modern Life Using Machine Learning Algorithms," *Procedia Computer Science*, vol. 167, pp. 1258-1267, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Prerna Chikersal et al., "Detecting Depression and Predicting its Onset Using Longitudinal Symptoms Captured by Passive Sensing: A Machine Learning Approach with Robust Feature Selection," *ACM Transactions on Computer-Human Interaction*, vol. 28, no. 1, pp. 1-41, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Lin Luo et al., "Predictors of Depression among Chinese College Students: A Machine Learning Approach," *BMC Public Health*, vol. 25, no. 1, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Md. Iqbal Hossain Nayan et al., "Comparison of the Performance of Machine Learning - Based Algorithms for Predicting Depression and Anxiety among University Students in Bangladesh : A Result of the First Wave of the COVID - 19 Pandemic," *Asian Journal of Social Health and Behavior*, vol. 5, no. 2, pp. 75-84, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Nasirul Mumenin et al., "Screening Depression among University Students Utilizing GHQ-12 and Machine Learning," *Heliyon*, vol. 10, no. 17, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Radwan Qasrawi et al., "Assessment and Prediction of Depression and Anxiety Risk Factors in Schoolchildren: Machine Learning Techniques Performance Analysis," *JMIR Formative Research*, vol. 6, no. 8, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Arkaprabha Sau, and Ishita Bhakta, "Predicting Anxiety and Depression in Elderly Patients Using Machine Learning Technology," *Healthcare Technology Letters*, vol. 4, no. 6, pp. 238-243, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[21] Graziella Orrù et al., "Machine Learning in Psychometrics and Psychological Research," *Frontiers in Psychology*, vol. 10, pp. 1-10, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Heejung Kim et al., "Depression Prediction by Using Ecological Momentary Assessment, Actiwatch Data, and Machine Learning: Observational Study on Older Adults Living Alone," *JMIR mHealth and uHealth*, vol. 7, no. 10, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] Madan Mohan Tito Ayyalasomayajula, Akshay Agarwal, and Shahnawaz Khan, "Reddit Social Media Text Analysis for Depression Prediction: Using Logistic Regression with Enhanced Term Frequency-Inverse Document Frequency Features," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 5, pp. 5998-6005, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Tanveer Hasan et al., "Prediction of Bangladeshi Urban Children's Mental Health for the Effect of Mobile Gaming Using Machine Learning," *International Journal on Advanced Science, Engineering & Information Technology*, vol. 14, no. 3, pp. 859-865, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] O.T Akinyele, "Development of a Web-Based System for Predicting Depression and Suicide Attempt Using Ensemble Machine Learning Model," *Journal of Science and Technology*, vol. 30, no. 6, pp. 98-106, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[26] Mohammad Tarek Aziz et al., "Textual and Numerical Data Fusion for Depression Detection: A Machine Learning Framework," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 38, no. 2, pp. 1231-1244, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[27] Cristian Zanon et al., "Examining the Dimensionality, Reliability, and Invariance of the Depression, Anxiety, and Stress Scale–21 (DASS-21) Across Eight Countries," *Assessment*, vol. 28, no. 6, pp. 1531-1544, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Rafat Hussain et al., "Physical and Mental Health Perspectives of First Year Undergraduate Rural University Students," *BMC Public Health*, vol. 13, pp. 1-11, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[29] Jetli Chung, and Jason Teo, "Single Classifier vs. Ensemble Machine Learning Approaches for Mental Health Prediction," *Brain Informatics*, vol. 10, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[30] Theodoros Iliou et al., "ILIOU Machine Learning Preprocessing Method for Depression Type Prediction," *Evolving Systems*, vol. 10, pp. 29-39, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[31] Shaurya Bhatnagar, Jyoti Agarwal, and Ojasvi Rajeev Sharma, "Detection and Classification of Anxiety in University Students through the Application of Machine Learning," *Procedia Computer Science*, vol. 218, pp. 1542-1550, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[32] Srinath Kudlapura Shivaiah et al., "Prediction of Mental Illness Using Ensemble Model and Grid Search Hyperparameter Optimization," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3726-3736, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[33] Malaika Pandit et al., "Examining Factors for Anxiety and Depression Prediction," *International Journal on Perceptive and Cognitive Computing*, vol. 9, no. 1, pp. 70-79, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[34] Mariam Ahmed Sheha, Mai S. Mabrouk, and Amr A. Sharawy, "Feature Engineering : Toward Identification of Symptom Clusters of Mental Disorders," *IEEE Access*, vol. 10, pp. 134136-134156, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[35] Yusen Zhai et al., "Machine Learning Predictive Models to Guide Prevention and Intervention Allocation for Anxiety and Depressive Disorders Among College Students," *Journal of Counseling & Development*, vol. 103, no. 1, pp. 110-125, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[36] Arkaprabha Sau, and Ishita Bhakta, "Informatics in Medicine Unlocked Screening of Anxiety and Depression Among Seafarers using Machine Learning Technology," *Informatics in Medicine Unlocked*, vol. 16, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[37] Md Jawadul Hasan et al., "Enhanced Classification of Anxiety, Depression, and Stress Levels: A Comparative Analysis of DASS21 Questionnaire Data Augmentation and Classification Algorithms," *Proceedings of the 3rd International Conference on Computing Advancements*, Dhaka Bangladesh, pp. 435-442, 2024. [CrossRef] [Google Scholar] [Publisher Link]