Predicting and Analyzing Student Mental Health with Machine Learning: A Data-Driven Approach to Early Intervention

Priyansha Pati[1]

Dr. Fabiola Hazel Pohrmen[2]

[1,2]Department of Computer Science, Christ (Deemed to be University), Bengaluru, India

I. Abstract

The mental health of students has become a fast-developing area of research, given its impact on academic performance, interpersonal relationships, and overall health. This study utilizes machine learning techniques, specifically Random Forest for classification and K-Means clustering with Principal Component Analysis (PCA), to analyze key factors influencing student mental health, including self-esteem, sleep quality, study load, social support and anxiety levels. A mental health analyzer was developed to sort and analyze student data, identifying distinct groups, including those with high stress and severe anxiety and depression due to academic pressure, those with moderate stress but with some coping capacity, those with stable mental health and minor issues, and those with high well-being, good academic performance, and good social support. The findings emphasize the importance of early intervention, personalized support strategies, and mental health support in educational settings. By integrating machine learning for mental health assessment, this research yields valuable insights to educators and policymakers in designing evidence-based, individualized interventions to improve student well-being and academic performance.

Keywords: Mental health, Machine Learning, Random Forest, K-Means clustering, Key factors, Mental health support, Early Intervention, Personalized support, Student well-being.

II. Introduction

Mental well-being is an important part of overall health, affecting the way people think, feel, and behave. It affects how individuals respond to stress, get along with others, and make choices. In students, especially those between the ages of 15 - 29, problems like anxiety and depression are increasingly common, yet they are not given much attention. Mental disorder illness, and falls, report the World Health Organization (WHO), are 3 of the leading causes of death worldwide. Suicidal motives are numerous, with influence derived from a mixture of social, cultural, biological, psychological, and environmental conditions throughout life-course. Such underlines a strong need for early intervention and awareness of mental illness, with strong predictive methods in place for evaluation.

There has been increased alarm at the increasing rate of depression among youth in India. According to the Times of India, the United Nations classifies youth as persons between 15–24 years, who account for 10% of the population of India. WHO states that depression is estimated to afflict more than 300 million individuals worldwide and will be the number one cause of disability by 2030. Research also shows that Indian youth have depression rates of between 31% and 57%. Since there is no mental health support and awareness, the situation is worsened by the need to incorporate technology-based solutions to address this gap. Furthermore, the Centers for Disease Control and Prevention (CDC) indicates that one in every six adults gets depressed at some point during their lifetimes, impacting about 16 million American adults each year. Anxiety disorders usually occur together with depression, resulting in constant and overwhelming fears, worries, and panics. Such feelings disrupt daily functioning, sometimes on a long-term basis if left untreated.

Even with increasing numbers of students suffering from mental health problems, awareness and support systems are weak. Conventional mental health evaluation tends to depend upon self-reporting or clinical judgment, which might not be feasible for all students. The Economic Times (2023) reports the low number of mental health professionals in India—just 0.3 psychiatrists, 0.07 psychologists, and 0.07 social workers for every 100,000 people—adding to the problem. On the other hand, developed countries have around 6.6 psychiatrists for every 100,000 individuals. Additionally, the global average of mental hospitals is 0.04 per 100,000, while India only has 0.004. Considering these differences, there is an urgent need for a predictive model that can evaluate and analyze student mental health from various factors that affect it.

This study seeks to build a machine learning (ML)-based model to evaluate student mental health by recognizing major predictors for anxiety, depression, and stress. As concern over student mental health continues to gain momentum globally, this study hopes to present a data-driven solution towards understanding drivers of well-being. The key goals are to find meaningful predictors of mental health among students, create strong ML models to classify levels of anxiety, depression, and stress, and offer in-depth insights through interactive dashboards. Moreover, the ML analyzer should be able to conclude when intervention should be initiated to enhance mental health care systems. Through the use of data, the study classifies contributing factors into five broad domains—Psychological, Physiological, Social, Environmental, and Academic. These areas encompass variables like self-esteem, sleep, study load, peer pressure, and living conditions, which are examined to check their relationship with mental health outcomes. In a bid to make the model more accessible, an AI-based mental health analyzer will be created that accepts input of symptoms and provides a diagnosis of one's mental health. This system will not only give a clearer picture of mental health patterns among students but also be a useful tool in facilitating early intervention and mental health awareness.

Student mental health needs to be addressed through early intervention, awareness, and policy reforms since mental health influences decision-making, stress tolerance, and social relationships. Governments, healthcare providers, and communities need to prioritize mental health services that are affordable and accessible. India requires additional mental health infrastructure in the form of psychiatrists, psychologists, and counselors in schools. Stigma reduction is essential, as only about 40% of Indian youth between 15 and 24 years of age subscribe to seeking mental health care. This highlights the necessity for awareness campaigns, better policies, and AI-based early diagnosis and intervention tools. Mental illness among students is on the rise, but awareness and support are still lacking.

This research stresses the importance of evidence-based solutions to close the mental health assessment and intervention gap. Using machine learning, it is now possible to catch students at risk early on and deliver targeted recommendations. The application of AI-powered mental health analyzers can transform mental health responses, making help more accessible. Blending predictive analytics with policies on mental health and campaigns of awareness can develop a preventive measure, providing the students with assistance they need prior to when the problems aggravate.

III. Literature Review

Stress, Anxiety and Depression among a Cohort of Health Sciences Undergraduate Students: The Prevalence and Risk Factors

Muhammad Faris Fauzi et al. examined the occurrence and risk determinants of stress, anxiety, and depression (SAD) in health sciences undergraduates. With the use of the DASS-21 scale, 449 students with a response rate of 93.9% were evaluated, and it was found that 65% reported stress, 85.1% anxiety, and 51.4% depression. Though most stress (74.6%) and depression (66.2%) cases were mild, 74.6% had moderate-to-severe anxiety. Poor quality of sleep and tiredness were significant predictors of depression and anxiety, whereas low-grade fever and headache were associated with stress and anxiety. The highest SAD scores were reported by medical imaging students. The research demands early screening, monitoring, and focused interventions to address students' mental well-being.

Social Media Discussions Predict Mental Health Consultations on College Campuses

Koustuv Saha, et al. investigated social media as a passive sensor for measuring college students' mental health. They used a large U.S. public university dataset (2011–2016), which included 66,000 Reddit posts and ground-truth mental health consultation records. Machine learning and natural language processing methods detected mental health terms, and SARIMA models with social media data enhanced consultation forecasting precision by 41% ($r = 0.86$, SMAPE = 13.30). Language analysis indicated that high-consultation months were concerned with academics and career stress, whereas low-consultation months were concerned with socialization and positive effects. Results indicate the potential of social media in forecasting mental health trends and guiding campus support strategies.

Factors that influence mental health of university and college students in the UK: a systematic review

Fiona Campbell, et al. undertook a systematic review of factors that affect the mental health of higher education students. Examining 31 UK-based studies between 2010 and 2020, they recognized key risk factors as childhood trauma, LGBTQ identity, and autism, whereas protective factors were high social networks and flexibility in adjusting to academic change. Disengagement from learning and low mental health literacy were associated with poor mental health. Their report highlights the necessity for specific interventions and support strategies to enhance student well-being.

Influencing factors, prediction and prevention of depression in college students: A literature review

Xin-Qiao Liu et al. summarized nonpathological determinants of depression in college students, classifying them into biological, psychological, college experience, and lifestyle determinants. The COVID-19 pandemic exacerbated depressive levels because it led to distortions in routine life and heightened stress. New technologies such as AI and big data now complement the standard psychological tests used in forecasting mental health risks. The research emphasizes the significance of general and professional nonpharmacological interventions, such as family support, cognitive behavior therapy, and lifestyle adjustment. Higher learning institutions are encouraged to implement sophisticated screening techniques, eliminate stigma, and provide individualized, evidence-based mental health care for students.

Social and Individual Factors Predicting Students' Resilience: A Multigroup Structural Equation Model

Wassilis Kassis et al. examined predictors of students' resilience during the COVID-19 pandemic using a longitudinal study of 713 Greek, German, and Swiss students. The research tested both individual (self-esteem, self-efficacy) and social (parental, teacher, and social resources) factors to establish their influence on resilience prior to and during the pandemic. Through multigroup structural equation modeling, the researchers found two different resilience trajectories: a low-anxiety group, with stable mental health, and a high-anxiety group, with more depression and anxiety. The results emphasize that although parents are providing significant support to high-anxiety or depressed students, teachers tend to neglect them. In addition, the research deconstructs the prevailing narrative of migrant students' resilience problems resulting from cultural deficiencies, demonstrating the stability of processes of resilience among various demographic categories. In total, the research highlights the value of a systems approach to understanding resilience, calling attention to how self-efficacy, self-esteem, and social support structures interact to facilitate students' psychological health.

Anxiety, Depression, and Suicide in Youth

Ned H. Kalin et al. investigated the risk factors and early onset of anxiety disorders and depression in adolescents. Their research is notable for pointing out the significant comorbidity of these psychiatric conditions, where anxiety tends to precede depression. They pointed to genetic and environmental factors, including early childhood trauma and persistent stress, as major risk factors. The study also points to the adolescent transition phase as an important window for intervention. They emphasize the importance of increased accessibility to treatment and novel therapeutic strategies, such as cognitive-behavioral therapy (CBT), neuroimaging-based predictions, and new pharmacological strategies to enhance mental health outcomes in vulnerable youth.

Academic Stress, Anxiety and Depression among College Students- A Brief Review

Narasappa Kumaraswamy reviewed the prevalence and influence of academic stress, anxiety, and depression among university students. The study emphasizes that 10-20% of the students suffer from psychological distress at any point in time, which impairs their academic progress and general well-being. The research summarizes three decades of literature focusing on student stress, emotional distress, and psychiatric morbidity. The article focuses on the significance of counseling in providing solutions for mental health issues and recommends preventive strategies like opening student counseling centers, making mental health awareness programs a priority, and initiating mentor-mentee schemes. It also promotes the establishment of student health committees with professional mental health practitioners, conducting workshops for students and teachers, and establishing a healthy learning environment. The research summarizes that enhancing teaching processes, appreciating educators, and supporting an appropriate system is the key to ensuring college students' mental welfare.

Mental Health Issues and Challenges in India: A Review

Venkata Shiva Reddy carried out a review to evaluate the burden of mental disorders and the problems encountered at the community level. The research examined different electronic databases to establish the prevalence of psychiatric morbidities and factors related to them. Mental and behavioral disorders, as per the World Health Organization, account for about 12% of the world's disease burden. In India, the prevalence varies from 9.5 to 102 per 1000 people. The study puts forth that the apparent burden of mental disorders is merely the "tip of the iceberg." Research in the area suggests that mental disorders are more common among females, older persons, survivors of disasters, factory workers, children, adolescents, and patients with long-term medical illnesses. The study places an emphasis on enhancing the living conditions of people, making efforts to ensure political commitment, ensuring strengthened primary healthcare, as well as women's empowerment. The results indicate that mental disorders change over time, impacting healthcare planning, funding, and service provision. Community surveys are important in getting a better picture of mental health issues among various groups.

IV. Methodology

The research process is structured into a series of important steps to investigate the mental well-being of students in terms of various psychological, physiological, social, environmental, and academic variables. The data processing involves data collection, preprocessing, exploratory data analysis (EDA), feature selection, implementation of the model, and performance assessment.

1. Data Collection

The information utilized for this study was sourced from Kaggle and was from a survey conducted across the country by Chhabi Acharya, a Machine Learning Engineer, between June 2022 and October 2022 in Dharan, Nepal. The information was gathered through visits to schools and colleges, where students were interviewed to measure the impact of mental health awareness. The data set has around 20 numeric attributes that represent the most significant mental health variables: psychological (anxiety level, self-esteem, mental health history, depression), physiological (headache, blood pressure, sleep quality, breathing problem), social (social support, peer pressure, extracurricular activity, bullying), environmental (noise level, living quality, safety, basic needs), and academic (academic performance, study load, teacher-student relationship, future career problems). Anxiety was measured using GAD-7 (0-21), self-esteem was measured using Rosenberg Self-Esteem Scale (0-30), and depression was measured using PHQ-9 (0-27). Most other variables were 0-5, and they ranked low (0-1), mid (2-3), and high (4-5).

2. Data Preprocessing

The data was preprocessed for consistency and quality. Mean, median, or mode imputation was used to handle missing values. Min-Max normalization was used to normalize feature ranges. Encoding was not required since all features were numerical. Pearson correlation and heatmaps were utilized to identify variable relationships. The data consisted of 1,100 instances with 21 numerical features of psychological, physiological, social, environmental, and academic factors that affect student mental health.

3. Exploratory Data Analysis (EDA)

EDA was conducted to gain insights into the dataset and detect patterns related to mental health. Key visualizations and analyses included:

A. Distribution Analysis

Histograms and box plots were used in distribution analysis to compute the range of various features. While factor-wise analyzing the negative experiences, it was understood that the academic factor was predominantly where students have experienced negatives, followed by the psychological factor. Figure 1. Box plot results showed quite a high number of significant trends: Anxiety, self-esteem, and depression were distinct, in the sense that very wide variation was documented on these symptoms among the students with some reporting significantly higher ones. Sleep quality, headache, and blood pressure varied moderately, evidenced by relatively broad interquartile ranges (IQR), suggesting a more stable yet variable response across students. Outliers were also noted for study load, living condition, and noise level, with some experiencing most extreme conditions in these. These when contrasted with others, social and educational variables such as extracurricular activity, teacher- student relations and teachers' abuse were less variable and could be predictive of relatively more comparable experience for subjects on these variables.

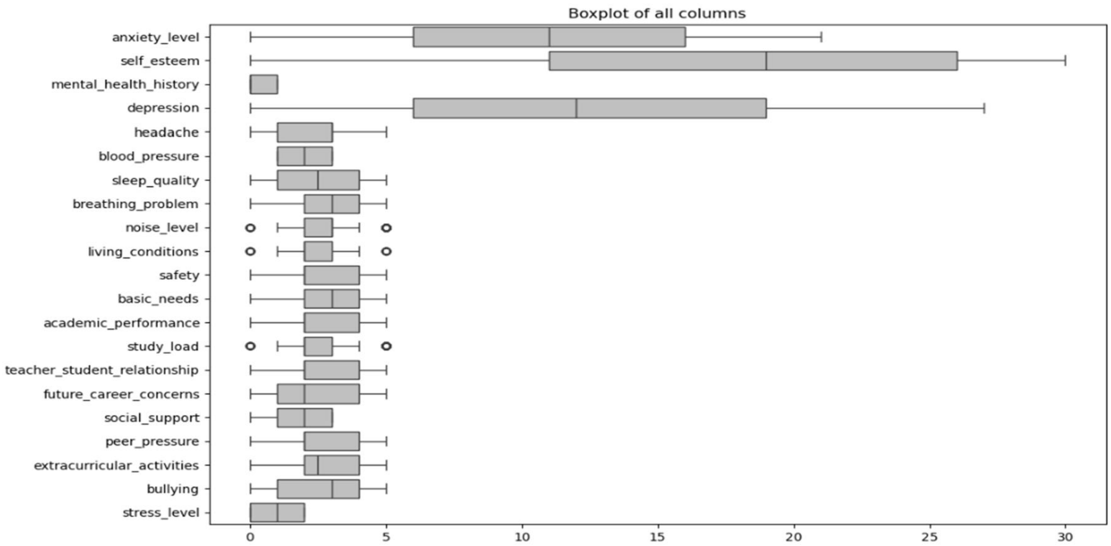

Figure 1. Box plot

B. Correlation Heatmap Analysis

Correlation heatmaps were used to understand relationships between variables and their impact on anxiety and depression and noted in Table1.

Correlation Pair	Correlation with Anxiety	Correlation with Depression
Stress Level	0.74	0.73
Future Career Concerns	0.72	0.68
Bullying	0.71	0.67
Peer Pressure	0.64	0.61
Sleep Quality	-0.71	-0.69
Self-Esteem	-0.70	-0.63
Safety	-0.65	-0.59
Teacher-Student Relationship	-0.67	-0.67
Social Support	-0.57	-0.62

Table 1. Key Correlations with Anxiety and Depression

Some of the main findings from the analysis are that depression and anxiety are related to each other, meaning that by attacking one, the other would also be reduced. Additionally, stress pressure, fear of future career, and bullying play a major role in causing anxiety and depression, hence the need to address these pressures for better mental health. On the other hand, resilience determinants like quality sleep, high self-esteem, and healthy relationships are factors in the relief of mental illness burdens, prioritizing the formation of environments that foster these positive attributes to support overall mental health.

C. Pairwise Comparisons

The research reports strong positive correlations that confirm that the levels of depression as well as anxiety increase as stress levels increase. The comparability of Pearson and Spearman correlations as shown in Table 2 confirms a linear relationship with normal rank ordering, which is also substantiated by Figure 2 having an upward trend in confirmation of this relationship and Figure 3 showing a sting correlation between the 3 factors .

Correlation Pair	Pearson Correlation	Spearman Correlation
Stress Level & Anxiety	0.74	0.74
Stress Level & Depression	0.73	0.75
Anxiety & Depression	0.69	0.71

Table 2. Pairwise Comparisons

Statistical analyses, t-tests and ANOVA, provide additional details on the determinants of depression and anxiety. The results of t-tests reveal that all the determinants have a statistically significant relationship with anxiety and depression. In other words, self-esteem, stress, history of mental health, headache, blood pressure, sleep, noise level, living, and performance all possess significant t-statistics, showing their strong correlation with mental health outcomes. For instance, heightened self-esteem is associated with lower levels of anxiety and depression, while heightened stress and mental health history result in higher levels of both. Poor sleep quality, higher exposure to noise, and physiological symptoms such as headaches and hypertension also relate to heightened levels of depression and anxiety, therefore pointing to the necessity to treat the same.

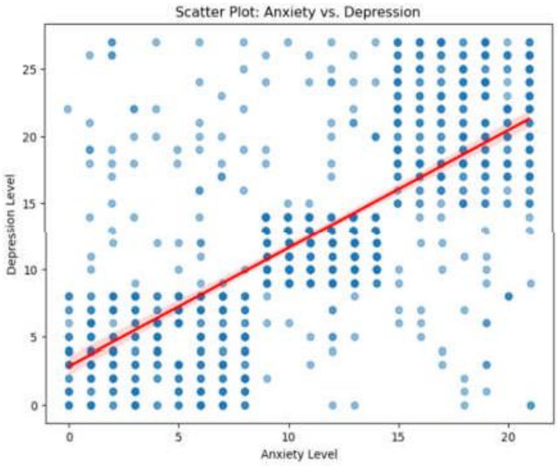

Figure 2. Anxiety vs. Depression

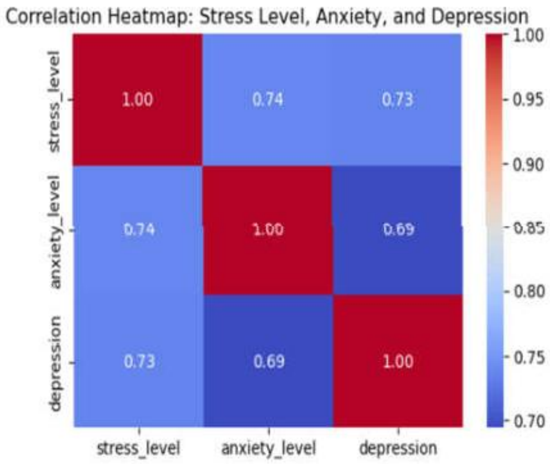

Figure 3. Correlation Heatmap: Stress, Anxiety and Depression

Furthermore, ANOVA tests indicate that past history of mental health, social support, stress, and blood pressure have the most significant effects on depression and anxiety, with F-values of more than 2000. The highest F-values are found for past history of mental health, indicating that previous mental health problems are significant predictors of present depression and anxiety levels. The evaluation also reveals that stress and social support have a significant part to play, with low social support and high stress levels influencing mental health issues considerably. Variables such as headaches, noise level, self-esteem, sleep quality, and concerns about future jobs also influence anxiety and depression significantly, though not as much as stress and social support.

The exploratory analysis reveals key insights into the risk and protective factors of depression and anxiety among students. Career problems, bullying, peer pressure, and high rates of stress are primary culprits of poor mental health, with anxiety and depression seemingly highly correlated, and thus managing one condition might possibly lower the prevalence of the other. Protective factors such as good quality of sleep, high self-esteem, security, and good teacher-student relationships are found to prevent mental distress. Practice implications following these findings are the use of stress management programs, increased mental support services, anti-bullying measures, and fostering self-esteem and resilience-building activities. These results are essential to guide mental health practitioners, educators, and policymakers to develop targeted interventions to enhance student well-being and reverse anxiety and depression.

4. Machine Learning Model Implementation & Model Evaluation and Interpretation

For classification and prediction of mental health illness (stress, depression, and anxiety), multiple models of machine learning were utilized. The dataset was split into 80% train and 20% test set separately for depression and anxiety to facilitate learning and testing in an efficient manner. Feature importance was investigated following training of models for the identification of factors most crucial for influencing mental health. Every model was cross-compared and ranked based on accuracy, predictiveness, and interpretability. Results provided valuable information regarding significant risk factors of mental health, adding to better understanding and response towards student well-being.

Random Forest Regressor

Random Forest Regression is used to forecast continuous values, such as mental health scores, by averaging predictions from a set of decision trees. It provides feature importance, which identifies large risk factors for mental health. Here, the model for anxiety is more precise than that for depression, having a lower Mean Absolute Error (MAE) of 2.43 compared to depression at 3.61. R^2 for anxiety is 0.72, explaining 72% of the variance, while the depression model gives $R^2 = 0.63$ and explains 63%. The higher error for depression predictions may be due to genetic or personal factors not included. Both models, however, perform well in identifying key risk factors, validating strong feature selection.

The feature importance analysis determines the most significant factors that influence both anxiety and depression. Figure 4 indicates for anxiety, the most significant factors are bullying, sleep quality, teacher-student relationships, future career concerns, and breathing problems, with bullying being the most significant predictor. Safety and self-esteem also play a role in increasing anxiety. From Figure 5 for depression, bullying is once more the most significant factor, followed by sleep quality, teacher-student relationships, and future career concerns. Furthermore, study burden and headache are strong factors towards depression. From these results, the strong association of sleep and bullying emerges regarding both disorders, together with significant impact from student-teacher relationship and career planning problems in the future.

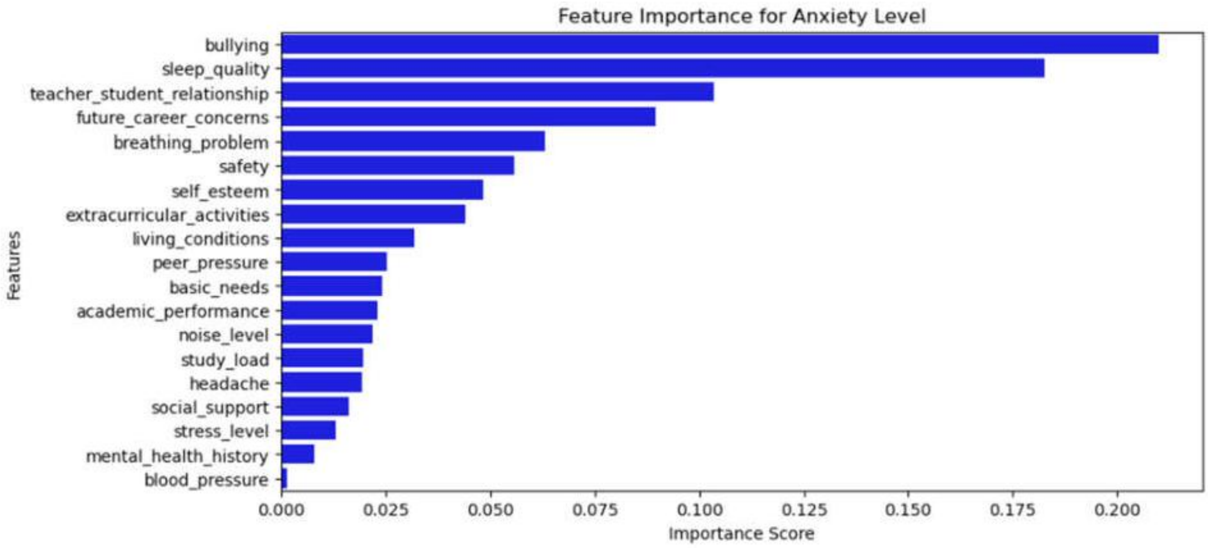

Figure 4 .Feature Imortance For Anxiety Level

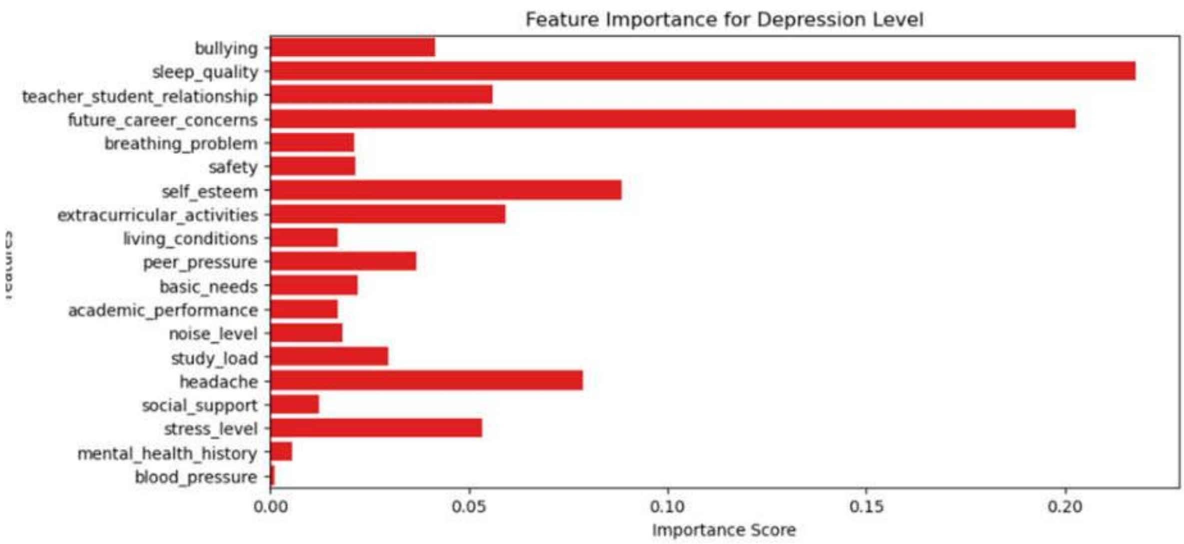

Figure 5.Feature Imortance For Depression Level

SHAP (SHapley Additive exPlanations) feature importance analysis shows the average contribution of every feature to the prediction of anxiety and depression by the model. Bullying and sleep quality are most important in predicting anxiety, followed by breathing difficulties, residential conditions, and student-teacher relationships. Basic needs and self-esteem moderately affect anxiety levels, and blood pressure contributes the least. In forecasting depression, future career problems and sleep are the most important characteristics, followed by headache, stress level, and self-esteem, which play very high influential roles. Teacher-student rapport and study burden also play very influential roles, while blood pressure and safety play very low influence in forecasting depression.

Random Forest Classifier

Random Forest Classifier is used for classification, where the categorical labels are predicted through majority voting over decision trees. It can be used for situations where the target variable is categorical, such as the prediction of risk for anxiety ("Yes" or "No"). Analysis from Figure 6 and Figure 7 illustrates that the anxiety model has a good balance in its prediction with 95 true positives (TP) and 77 true negatives (TN) but similarly with 21 false positives (FP) and 27 false negatives (FN), indicating the tendency towards the slight misclassifying of cases of high anxiety. On the other hand, the depression model performs excellently, correctly predicting 105 depression cases (TP) and 69 non-depression cases (TN) with only 16 false positives (FP) and only 30 false negative (FN), displaying high accuracy and good prediction ability for depression. The anxiety prediction model shows balanced performance with little bias towards low anxiety (Class 0), indicated by higher precision and recall for this class. The 75% overall accuracy and F1-score of around 75% suggest the model is well-generalized but may misclassify some cases of high anxiety.

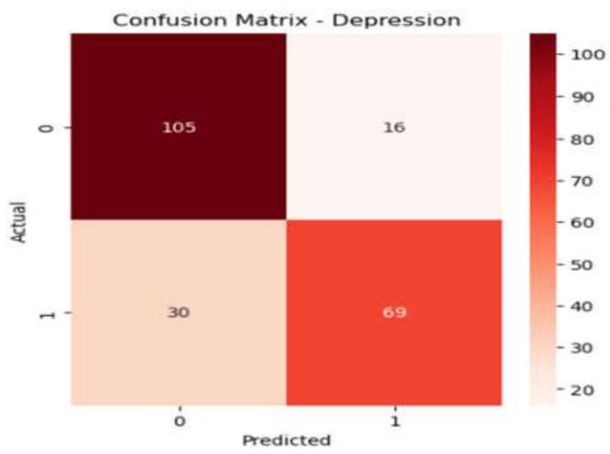

Figure 6. Confusion Matrix – Depression

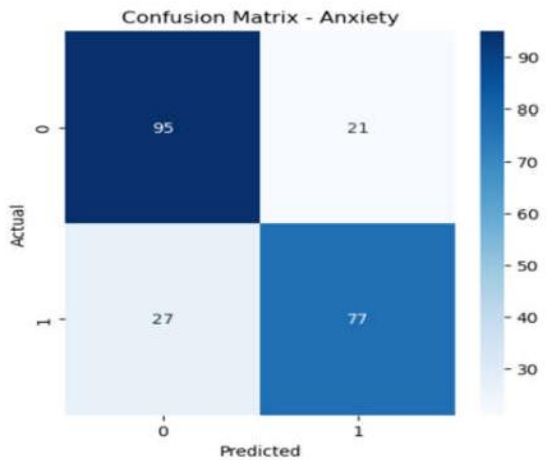

Figure 7. Confusion Matrix – Anxiety

The Random Forest model's feature importance for anxiety and depression classification reveals some key factors. Bullying, Sleep Quality, and Self-Esteem are the most significant factors leading to both anxiety and depression, with Future Career Concerns being more of a factor in depression. Environmental factors such as Noise Level and Living Conditions also play a major role, and physical symptoms such as Breathing Problems and Blood Pressure contribute minimally but notably. Mental Health History and Blood Pressure were found to have the least influence on both conditions.

The Random Forest Classifier accurately predicted Anxiety and Depression in students, which revealed that Bullying, Sleep Quality, and Self-Esteem were all predictors for both disorders. Future Career Concerns was a significant predictor of Depression, while Breathing Problems and Noise Levels were significant predictors in the case of Anxiety. Even environmental variables like Noise Level and Living Conditions were significant predictors of student mental health. The Anxiety model was 75% accurate with balanced performance, although mildly skewed in favor of low anxiety, and the Depression model was excellent, with very high accuracy and only one false negative, indicating its strong capability in correctly identifying depression cases.

LightGBM (LGB)

It employed LightGBM (LGB) to predict Anxiety and Depression among students based on a number of psychological, physiological, social, environmental, and academic factors. LightGBM was particularly adept at handling large datasets and class imbalances with the capability of determining complex feature relationships at very high computation rates.

For Anxiety Prediction, self-esteem was the strongest predictor, followed by relations with teachers and students, future career concerns, bullying, social support, extracurricular activity, and sleep quality. Stress levels, history of mental health, and blood pressure were weaker. For Depression Prediction, self-esteem was again most important, followed by headaches, study load, study performance, and respiratory problems. Bullying and peer pressure also contributed a lot, whereas environmental factors like noise, safety, and basic needs were more important in depression than in anxiety.

In terms of model performance, the Anxiety model achieved 74% accuracy with recall values of 0.77 on low anxiety and 0.71 on high anxiety, while the Depression model did slightly better at 76% accuracy with recall values of 0.82 for low depression and 0.69 for high depression. The Depression model was more accurate and remembered more, classifying true positive cases more accurately, though both models were unbalanced in remembering high anxiety or depression cases. The Anxiety model had a higher false negative rate, while the Depression model had fewer false negatives but more false positives. From Figure 7, the AUC metrics of 0.87 for Anxiety and 0.88 for Depression indicate that both models have excellent ability to distinguish positive and negative cases. AUC values close to 1 indicate exceptional classification ability. The average precision (AP) score of 0.88 indicates a very good balance between precision (properly labeling positives) and recall (capturing as many positives as possible). These metrics show that the models are very capable of identifying anxiety and depression instances with a small number of false predictions.

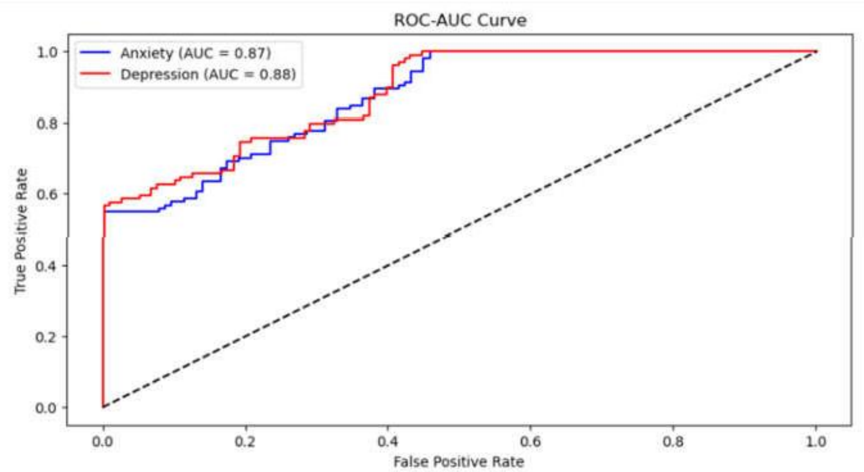

Figure 7. ROC – AUC Curve

Self-esteem was the strongest predictor for both Anxiety and Depression, suggesting that enhancing students' confidence and resilience through effectively designed interventions could be a strong way of conquering mental health disorders. Concerns regarding future careers and teacher-student relationships were particularly salient in Anxiety, emphasizing the function of academic guidance and mentorship programs to alleviate stress. Physical symptoms like headaches, sleep disturbances, and blood pressure were more directly linked to Depression, again supporting the need to address the mind-body interface in mental health care. Bullying, peer pressure, and social support were found to be key variables for both disorders, again highlighting the need for healthy social environments in schools.

KMeans combined with Agglomerative Clustering

This study used a hybrid clustering method that utilized both K-Means and Agglomerative Clustering in combining to cluster students according to their psychological, physiological, social, environmental, and academic considerations. This allowed pattern identification to identify clusters of students experiencing stress, anxiety, and depression.

The cluster analysis determined that the optimal number of clusters to categorize students based on their psychological, physiological, social, environmental, and academic features is k = 4. The Elbow Method also revealed that the Within-Cluster Sum of Squares (WCSS) curve became flat at k = 4, and Figure 8. Silhouette Score analysis also determined that k = 4 was optimal to differentiate clusters, with a steep decline in score for k > 5. Hierarchical Clustering Dendrogram also confirmed this choice as indicated in Figure 9 , revealing a steep increase in Euclidean distances at k = 4. K-Means performed slightly better than Hierarchical Clustering based on silhouette score, showing more defined clusters, although Davies-Bouldin Index for both methods was the same, reflecting the quality of clustering to be the same for both methods

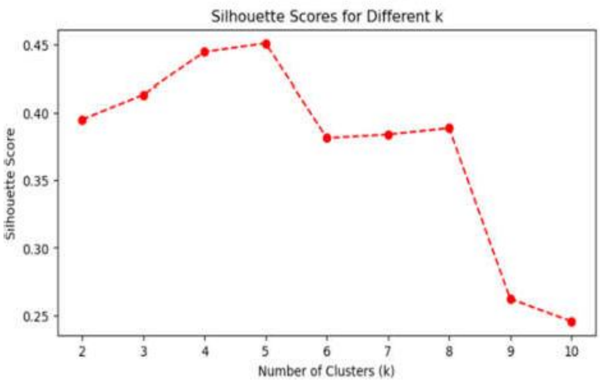

Figure 8. Silhouette Score

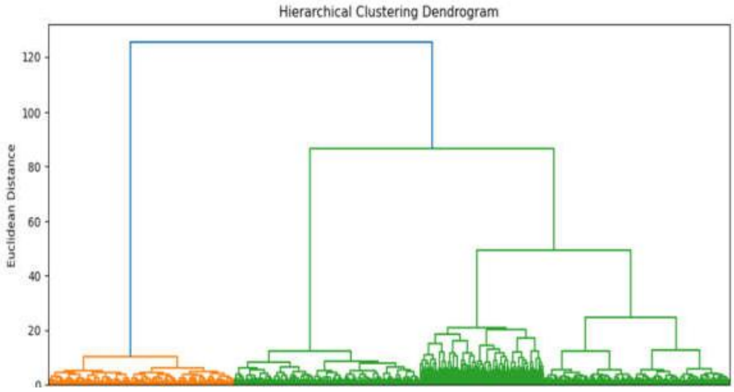

Figure 9. Hierarchical Clustering Dendrogram

The clustering analysis interpretations from Figure 10 identified four student groups according to their academic, mental health, and social profiles. Cluster 1, "High Anxiety & Depression – Academically Stressed," requires urgent attention, particularly for students with academic pressure, low self-esteem, and weak social support. Special interventions, such as mentorship programs, stress management workshops, and anti-bullying programs, would be beneficial. Cluster 0, "Moderate Well-Being with Mild Anxiety & Depression," suggests the utilization of preventative programs enhancing self-esteem, stress control, and academic assistance. Cluster 3, "High Well-Being & Academically Thriving," suggests students with minimal mental health concerns, and ongoing support to maintain their positive well-being is recommended, including access to leadership roles and academic involvement. Cluster 2, "Moderate Stress with Some Support," describes students who might benefit from additional social support, improved academic achievement, and mental health intervention to reduce levels of moderate depression and anxiety. These results underscore the importance of individualized mental health programs grounded in the specific needs of each cluster, which will allow institutions to effectively serve students in achieving successful academic and personal outcomes.

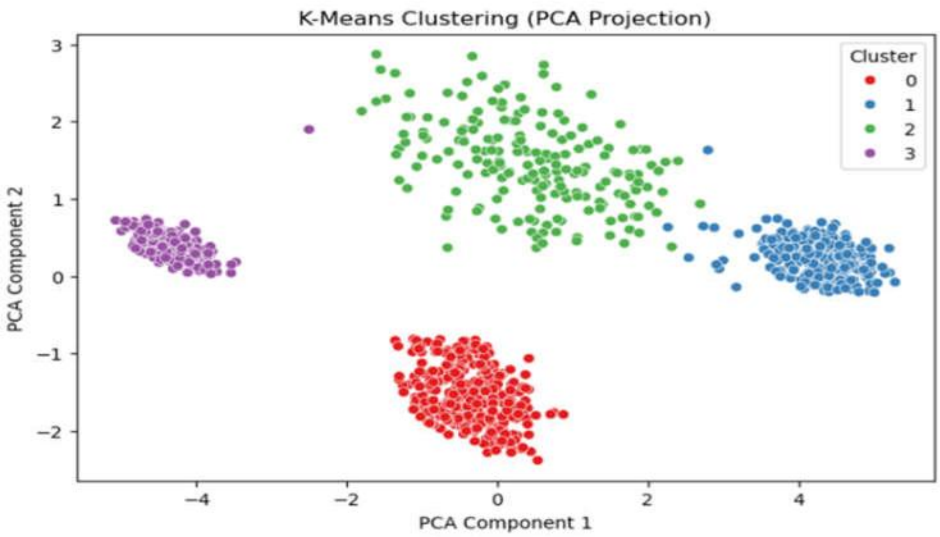

Figure 10.K-Means Clustering

5. Development of Mental Health Analyzer

The Mental Health Analyzer was developed as an interactive tool for the identification of potential mental health threats based on user inputs. The system specializes primarily in the rating of the likelihood of anxiety, stress, and depression using machine learning models trained on a set of 20 features known scientifically to be mental health related. Below are the steps that outline the implementation and development process:

As shown in Figure 11, the user begins by providing input regarding their symptoms based on the 20 features that were determined for mental health evaluation. Once the user provides their input, the data is analyzed through the trained RandomForestClassifier models for anxiety, stress, and depression. The models classify each condition (anxiety, stress, and depression) into high, moderate, or low.

Figure 11. Mental Health Analyzer

The Streamlit application works on the data interactively and, upon the click of the "Analyze" button, sends the input to the models. The models provide predictions and mark the severity of anxiety, stress, and depression. The results, along with recommendations based on the severity levels, are then displayed on the site, providing the user with useful recommendations to help themselves with their mental disorders. Figure 12 shows an example of the process ensuring a successful and informative experience for the user.

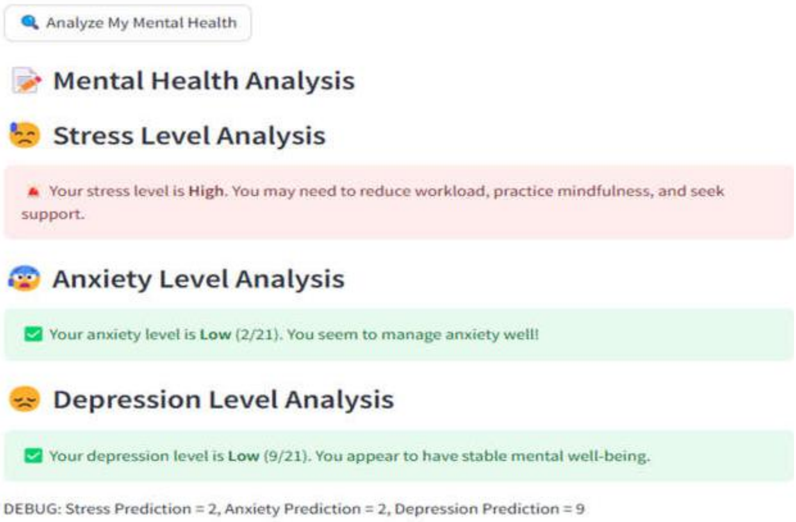

Figure 12. Mental Health Results (Sample)

V. Results and Analysis

The features of importance in this study revealed important factors impacting mental health, particularly anxiety and depression among students. Using the Random Forest Regressor, it was found that there were factors that were significant in both conditions. In anxiety, the most impactful factors were bullying, sleep quality, teacher-student relationships, future career concerns, and breathing problems. For depression, bullying was the most influential factor, followed by sleep quality, teacher-student relationships, and career opportunities. The strong association of sleep quality and bullying with both disorders indicate their central role in student mental health. The ML models trained using Random Forest Regressor and Random Forest Classifier were effective in predicting and classifying anxiety and depression.

The anxiety model had an MAE of 2.43, while the depression model worked with an MAE of 3.61, indicating greater variability in depression prediction. R^2 was 0.72 for anxiety, explaining 72% of the variance, and for depression, 0.63, explaining 63%. These results validate the models in terms of predicting significant risk factors. The KMeans combined with Agglomerative Clustering identified four clusters of students based on their mental health and academic characteristics.

Cluster 1 ("High Anxiety & Depression – Academically Stressed") presented the most mental health concerns, while Cluster 0 ("Moderate Well-Being with Mild Anxiety & Depression") and Cluster 2 ("Moderate Stress with Some Support") presented medium levels of anxiety and depression. Cluster 3 ("High Well-Being & Academically Thriving") evidenced minimal mental health concerns, highlighting the necessity for tailored interventions. Several visualization techniques, such as heatmaps, bar plots, and SHAP (Shapley Additive Explanations), were employed to interpret the model performances and feature importance. SHAP interpretation showed the contribution of each feature value towards the prediction of anxiety and depression, confirming the significant impact of sleep quality, bullying, and teacher-student relationships on mental health outcomes.

The most significant findings from the analysis were that bullying and sleep quality were both ranked consistently as the highest risk indicators for both depression and anxiety. Teacher-student relationships and future career concerns were also significant predictors with high correlations to both conditions. The Random Forest Classifier had high accuracy in the classification of anxiety and depression, with the depression model being slightly more accurate and having fewer false negatives. The clustering analysis demonstrated the need for personalized interventions based on the individual mental health profiles of the different student clusters. Comparing the results with earlier research, the findings are in line with earlier work, which also identifies bullying and sleep disturbance as important predictors of mental health issues in students. The strong correlation between teacher-student relationships and mental health supports findings from earlier research identifying the contribution of positive social support networks to mental health risk buffering.

The study has several limitations, Dataset bias may impact the generalizability of the findings since the data available might not be representative of all student groups, especially those with different geographical locations or socio-economic backgrounds. Also, the lack of real-time data limits the capacity to monitor students' mental health changes over a period of time. Moreover, the models were not able to include genetic factors, which may play a role in mental health outcomes but were not present in the available dataset. The findings have significant implications for mental health interventions. Students in Cluster 1 ("High Anxiety & Depression – Academically Stressed") require immediate intervention, with specialized programs such as mentorship programs, stress management workshops, and anti-bullying programs. Students in Cluster 0 and Cluster 2 would benefit from self-esteem-enhancing activities, sleep promotion programs, and additional academic support. For students in Cluster 3, ongoing leadership roles, academic involvement, and peer mentorship programs will work to sustain their high well-being.

This research emphasizes the need for individualized mental health programs according to student needs, with consideration of factors like bullying, sleep quality, and teacher-student relationships. The Mental Health Analyzer is a key tool developed in this study, which allows students to input symptoms and receive individualized analysis for anxiety, depression, and stress. It predicts mental health risk using trained Random Forest Classifier models and recommends interventions based on severity, including counseling or stress management. The analyzer is user-friendly and can track trends over time, enabling students to see their mental health and the effectiveness of interventions. It provides early detection to facilitate timely intervention and better mental health outcomes. This software then becomes a vital component of a school's mental health support system, ensuring that mental health issues do not interfere with the academic performance and well-being of students.

VI. Conclusion

In conclusion, the study puts into perspective the need for mental health programs to be individualized for students' needs, especially anxiety and depression. From the analysis, it was determined that bullying and sleep quality were always the greatest predictors of both conditions. Teacher-student relationships and future career aspirations also ranked high among predictors. The findings support earlier research, which highlights the enormous impact of these variables on the mental health outcomes of students. The Mental Health Analyzer developed in this research provides a useful tool for students to assess risks of their mental health through inputting symptoms and receiving personalized feedback. This program utilizes the Random Forest Classifier algorithm to detect threats of anxiety, stress, and depression and presents direct recommendations against such threats.

Among the salient findings of this analysis is that there exists a very strong correlation between stress, anxiety, and depression. Cluster analysis grouped students into four clusters based on their mental health characteristics, and stress preceded both anxiety and depression. Chronic stress also exacerbates the existing mental health issues, leading to anxiety. Anxiety can escalate to depression if not addressed, as stress, anxiety, and depression are interconnected in a vicious cycle. These findings emphasize intervention at an early level to break this cycle before it advances to more severe mental health issues.

The study also portrayed that the high-anxiety and depression students, especially Cluster 1 ("High Anxiety & Depression – Academically Stressed"), must be treated with intensive and prompt interventions. Mentorship, stress management, and anti-bullying initiatives are required to eliminate the impact of bullying and sleep disturbances that were outlined as the most salient risk factors. In addition, the findings necessitate ongoing support for Cluster 3 students ("High Well-Being & Academically Thriving") to maintain their well-being, with interventions in peer mentorship and leadership. Going forward, the inclusion of a mental health chatbot and actual student surveys would further enhance the Mental Health Analyzer.

Through the utilization of AI-aided tools for temporary mental health assessment, schools could monitor the improvement or deterioration of students' wellbeing over time and adopt more suitable interventions. In this study, valuable data were gained on student mental health dynamics, portraying the reality that addressing key factors of stress, bullying, and sleep quality significantly affects students' wellbeing and general academic performance.

VII. References

1. Reddy, V. E. N. K. A. T. A. S. H. I. V. A. "Mental health issues and challenges in India: A review." *International Journal of Social Sciences Management and Entrepreneurship (IJSSME)* 3.2 (2019).
2. Kumaraswamy, Narasappa. "Academic stress, anxiety and depression among college students: A brief review." *International review of social sciences and humanities* 5.1 (2013): 135-143.
3. Kalin, Ned H. "Anxiety, depression, and suicide in youth." *American journal of psychiatry* 178.4 (2021): 275-279.
4. Campbell, F., Blank, L., Cantrell, A., Baxter, S., Blackmore, C., Dixon, J., & Goyder, E. (2022). Factors that influence mental health of university and college students in the UK: a systematic review. *BMC public health*, *22*(1), 1778.
5. Fauzi, M. F., Anuar, T. S., Teh, L. K., Lim, W. F., James, R. J., Ahmad, R., ... & Salleh, M. Z. (2021). Stress, anxiety and depression among a cohort of health sciences undergraduate students: the prevalence and risk factors. *International journal of environmental research and public health*, *18*(6), 3269.
6. Saha, K., Yousuf, A., Boyd, R. L., Pennebaker, J. W., & De Choudhury, M. (2022). Social media discussions predict mental health consultations on college campuses. *Scientific reports*, *12*(1), 123.
7. Liu, X. Q., Guo, Y. X., Zhang, W. J., & Gao, W. J. (2022). Influencing factors, prediction and prevention of depression in college students: a literature review. *World journal of psychiatry*, *12*(7), 860.
8. Kassis, W., Vasiou, A., Govaris, C., Favre, C., Aksoy, D., & Graf, U. (2023). Social and individual factors predicting students' resilience: A multigroup structural equation model. *Education Sciences*, *14*(1), 15.
9. *Student Stress Factors: A Comprehensive analysis*. (2023, October 14). Kaggle. https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis/data
10. Mehrotra, R. (2023, April 24). Rise of depression amongst young adults in India. *Times of India Blog*. https://timesofindia.indiatimes.com/blogs/voices/rise-of-depression-amongst-young-adults-in-india/

11. *Depression and anxiety*. (2023, March 14). Centers for Disease Control and Prevention. https://www.cdc.gov/tobacco/campaign/tips/diseases/depression-anxiety.html

12. Contributors, E. (2023b, October 10). World Mental Health Day: 60-70 mn people in India suffer from common mental disorders; stigmatisation & f. *The Economic* https://economictimes.indiatimes.com/magazines/panache/world-mental-health-day-60-70-mn-people-in-india-suffer-from-common-mental-disorders-stigmatisation-financial-barriers-prevent-timely-treatment/articleshow/104289268.cms

13. World Health Organization: WHO. (2024, August 29). *Suicide*. https://www.who.int/news-room/fact-sheets/detail/suicide

14. *Student Stress Factors: A Comprehensive analysis*. (2023, October 14). Kaggle. https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis/data

15. World Health Organization: WHO & World Health Organization: WHO. (2023, March 31). *Depressive disorder (depression)*. https://www.who.int/news-room/fact-sheets/detail/depression

16. World Health Organization: WHO. (2022, June 8). *Mental disorders*. https://www.who.int/news-room/fact-sheets/detail/mental-disorders

17. Co Indicator details. (n.d.-b). https://www.who.int/data/gho/data/indicators/indicator-details/GHO/age-standardized-suicide-rates-(per-100-000-population)**py**

18. World Health Organization: WHO. (2022b, June 17). *Mental health*. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response