Depression Level Detection in Undergraduate Students: A Machine Learning Approach Using Academic and Psychological Factors*

Khandoker Wahiduzzaman Anik
Computer Science
BRAC UNIVERSITY
Dhaka, Bangladesh
khandoker.wahiduzzaman.anik@g.bracu.ac.bd

Mehrabul Islam
Computer Science
BRAC UNIVERSITY
Dhaka, Bangladesh
mehrabul.islam@g.bracu.ac.bd

***Abstract*—This study presents a machine learning approach for detecting depression levels among undergraduate students based on academic performance, psychological factors, and demographic characteristics. Using a comprehensive dataset collected from 1977 undergraduate students, we trained and evaluated multiple machine learning models including Logistic Regression, XGBoost, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and a Neural Network (MLP). Our best-performing model, XGBoost, achieved an accuracy of 94.95% and a weighted F1-score of 0.9495 in classifying six distinct depression severity categories. The 'Depression Value' feature, a composite score from a depression assessment, was identified as the most significant predictor. These findings demonstrate that the combination of academic metrics (CGPA), psychological assessments (anxiety, stress, and depression scores), and demographic information creates a robust framework for depression detection. These findings suggest the potential for implementing early screening tools in university settings to identify students at risk for depression, enabling timely intervention. The study contributes to the growing field of educational data mining for mental health applications, highlighting the effectiveness of machine learning in addressing critical student wellness challenges.**

***Index Terms*—Depression detection, machine learning, undergraduate students, mental health, educational data mining, XGBoost, psychological factors, academic performance.**

I. Introduction

Depression among undergraduate students represents a significant public health concern, with prevalence rates estimated between 10% and 85% across different studies and populations [1]. The university environment presents unique stressors that can trigger or exacerbate depressive symptoms, including academic pressure, financial concerns, social challenges, and career uncertainty [2]. These factors are particularly salient in competitive fields such as computer science and engineering, where high performance expectations and heavy workloads are common [3].

The consequences of untreated depression in student populations are severe, including impaired academic performance, increased dropout rates, deteriorating physical health, substance abuse, and in extreme cases, suicide [4]. Despite these serious implications, depression among students remains underdiagnosed and undertreated, with fewer than 25% of depressed students receiving adequate mental health care [5].

Early detection of depression is crucial for timely intervention and improved outcomes. Traditional detection methods rely on clinical interviews and standardized assessment tools, which require specialized resources and student initiative to seek help. These approaches often miss students who are reluctant to self-identify as needing mental health support due to stigma or lack of awareness [6].

Machine learning offers a promising alternative for depression detection by leveraging multiple data points that may be readily available in educational settings. Previous research has explored various predictors of student depression, including academic performance, psychological factors, and demographic characteristics. However, few studies have combined these dimensions into integrated predictive models that can accurately classify depression severity levels with high precision.

This study addresses this gap by developing and evaluating machine learning models for depression level detection in undergraduate students using a multidimensional approach. Specifically, we combine academic performance metrics (CGPA), psychological assessments (anxiety, stress, and depression scores), and demographic factors to predict depression severity across six categories: No Depression, Minimal Depression, Mild Depression, Moderate Depression, Moderately Severe Depression, and Severe Depression.

The primary contributions of this paper include:

- A comprehensive analysis of factors contributing to undergraduate depression using a dataset incorporating academic, psychological, and demographic dimensions.
- Development and rigorous evaluation of machine learning models for depression severity classification.
- Comparison of different algorithmic approaches, including Logistic Regression, XGBoost, SVM, KNN, and a Neural Network.
- Insights into the relative importance of different features in predicting depression, with a notable emphasis on the 'Depression Value' score.

- Recommendations for implementing data-driven early detection systems in university settings.

II. Methodology

A. Dataset Description

The dataset used in this study consists of information collected from 1977 undergraduate students primarily from the Computer Science and Engineering departments. The initial dataset contained 39 original variables, which were processed and feature-engineered, resulting in a final set of 11 input features and 1 target variable for the models. The dataset contains demographic information (age, gender, academic year), academic performance metrics (CGPA), financial status (scholarship/waiver), and comprehensive psychological assessment scores measuring anxiety, stress, and depression levels.

For depression assessment, we utilized a standardized instrument based on the Patient Health Questionnaire-9 (PHQ-9), which measures the frequency of nine depression symptoms. Each symptom is rated on a scale from 0 to 3, yielding a total 'Depression Value' (score) between 0 and 27. Anxiety was measured using a 7-item scale yielding an 'Anxiety Value', and stress was evaluated using a 10-item scale yielding a 'Stress Value', both focusing on academic-related contexts.

The 'Depression Value' scores were categorized into six severity levels for the target variable, `Depression_Target`. This variable was numerically encoded as follows:

- 0: No Depression (raw score 0-4)
- 1: Minimal Depression (raw score 5-9)
- 2: Mild Depression (raw score 10-14)
- 3: Moderate Depression (raw score 15-19)
- 4: Moderately Severe Depression (raw score 20-24)
- 5: Severe Depression (raw score 25-27)

The dataset exhibits an imbalance in the distribution of these depression levels: 'No Depression' (Class 0: 44 instances, 2.23%), 'Minimal Depression' (Class 1: 93 instances, 4.70%), 'Mild Depression' (Class 2: 408 instances, 20.64%), 'Moderate Depression' (Class 3: 449 instances, 22.71%), 'Moderately Severe Depression' (Class 4: 495 instances, 25.04%), and 'Severe Depression' (Class 5: 488 instances, 24.68%). This distribution is visualized in Fig. 1.

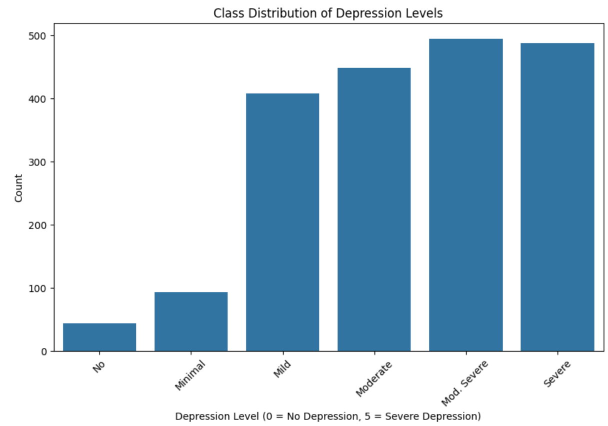

Fig. 1. Class Distribution of Depression Target Variable.

B. Data Preprocessing

The raw dataset underwent several preprocessing steps to prepare it for machine learning applications:

1) **Feature Selection and Engineering**: Based on domain knowledge and exploratory analysis, features representing academic performance, psychological states, and demographic factors were selected.
2) **Handling Categorical Variables**: Categorical variables were transformed using appropriate encoding methods:
   - Gender was one-hot encoded into three binary features: `2._Gender_Female`, `2._Gender_Male`, `2._Gender_Prefer_not_to_say`.
   - Scholarship status (`7._Did_you_receive_a_waiver_or_scholarship_`) was one-hot encoded into two binary features with `_No` and `_Yes` suffixes.
   - CGPA ranges were converted to `CGPA_numeric`, representing the midpoint of each range (e.g., 2.50-2.99 $\rightarrow$ 2.75).
   - Age ranges were converted to an approximate numeric value, `Age_numeric`.
   - Academic year was transformed to a numeric scale, `Year_numeric` (e.g., First Year $\rightarrow$ 1.0).
3) **Composite Score Utilization**: The pre-summed composite scores `Anxiety Value`, `Stress Value`, and `Depression Value` from the psychological assessments were used directly as numerical features.
4) **Target Variable Preparation**: As described above, the depression severity categories were encoded as integer values (0 to 5) for the `Depression_Target` variable.
5) **Data Splitting**: The dataset was split into training (80%) and testing (20%) sets, with stratification based on the `Depression_Target` to maintain class proportions in both splits. The test set comprised 396 instances.
6) **Addressing Class Imbalance**: The Synthetic Minority Over-sampling Technique (SMOTE) was applied exclusively to the training data to balance the class

distribution. Before SMOTE, the training data showed class counts such as Class 0: 35, Class 1: 75, Class 2: 326, Class 3: 359, Class 4: 396, and Class 5: 390. After SMOTE, each class in the training set was balanced to 396 instances. This is illustrated in Fig. 2.

7) **Data Augmentation**: Gaussian noise was added to the numerical features in the training data as a form of data augmentation to enhance model robustness.

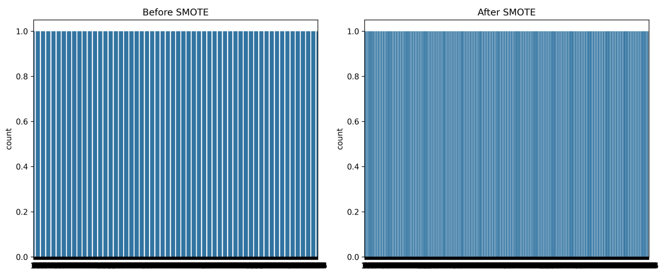

Fig. 2. Training Data Class Distribution Before and After SMOTE.

The final set of 11 input features fed into the models included: `Anxiety Value`, `Stress Value`, `Depression Value`, `CGPA_numeric`, `Age_numeric`, `Year_numeric`, `2._Gender_Female`, `2._Gender_Male`, `2._Gender_Prefer_not_to_say`, `7._Did_you_receive_a_waiver_or_scholarship_a` and `7._Did_you_receive_a_waiver_or_scholarshi`

C. Model Development

We implemented and compared five machine learning models:

1) **Logistic Regression**: A linear model used for multi-class classification.
2) **XGBoost**: A gradient boosting framework known for its high performance and efficiency.
3) **Support Vector Machine (SVM)**: A model that finds an optimal hyperplane to separate classes.
4) **K-Nearest Neighbors (KNN)**: A non-parametric, instance-based learning algorithm.
5) **Neural Network (MLP)**: A Multi-Layer Perceptron, a class of feedforward artificial neural network.

For each model, hyperparameter tuning was performed with cross-validation on the training data. As mentioned, SMOTE was applied to the training data to handle class imbalance.

D. Evaluation Metrics

Model performance was evaluated using the following metrics, primarily on the unseen test set:

1) **Accuracy**: The proportion of correctly classified instances.
2) **Weighted Precision**: Precision calculated for each class and then averaged, weighted by the number of true instances for each class.
3) **Weighted Recall**: Recall calculated for each class and then averaged, weighted by the number of true instances for each class.
4) **Weighted F1 Score**: The harmonic mean of weighted precision and recall, providing a balanced measure, especially useful for imbalanced datasets.
5) **Classification Report**: Per-class precision, recall, F1-score, and support to understand model behavior for each depression level.
6) **Confusion Matrix**: To visualize the performance of the classification model on each class.

III. Results

A. Model Performance Comparison

The performance evaluation of the machine learning models on the test set is summarized in Table I and visualized in Fig. 3.

TABLE I
COMPARISON OF MODEL PERFORMANCE METRICS ON TEST SET

Model	Accuracy	W. Precision	W. Recall	W. F1 Score
XGBoost	0.9495	0.9505	0.9495	0.9495
Neural Network (MLP)	0.9217	0.9225	0.9217	0.9214
Logistic Regression	0.9419	0.9423	0.9419	0.9418
SVM	0.8763	0.8787	0.8763	0.8767
KNN	0.6717	0.6724	0.6717	0.6682

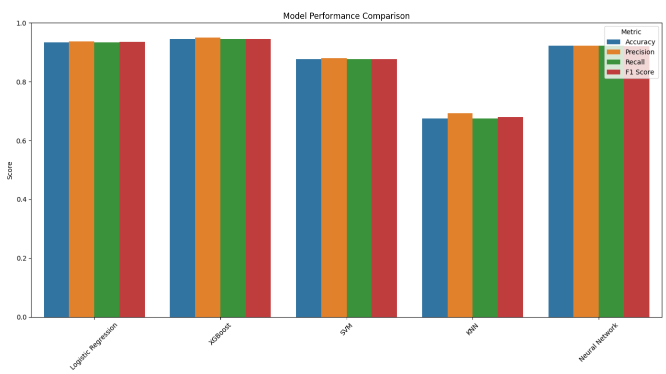

Fig. 3. Model Performance Comparison (Accuracy, Precision, Recall, F1 Score).

XGBoost emerged as the top-performing model, achieving an accuracy of 0.9495 and a weighted F1-score of 0.9495. Its confusion matrix (Fig. 4) shows excellent classification across all six depression levels. Specifically, it perfectly classified Class 4 (99 out of 99 instances) and Class 5 (98 out of 98 instances), with minor misclassifications in the other categories.

The Neural Network (MLP) also demonstrated strong performance with an accuracy of 0.9217 and a weighted F1-score of 0.9214. Its confusion matrix (Fig. 5) indicates good predictive power, with some misclassifications primarily between adjacent classes.

Interestingly, Logistic Regression provided robust results very close to XGBoost, with an accuracy of 0.9419 and a weighted F1-score of 0.9418. As seen in its confusion matrix

(Fig. 6), it performed consistently well across most classes but had slightly more misclassifications for the minority classes (Class 0 and 1).

SVM achieved an accuracy of 0.8763 and a weighted F1-score of 0.8767. Its confusion matrix (Fig. 7) shows more notable misclassifications, especially for Classes 0, 1, and 3.

KNN was the weakest performer, with an accuracy of 0.6717 and a weighted F1-score of 0.6682. The confusion matrix for KNN (Fig. 8) reveals considerable confusion between classes, particularly for Classes 2 and 3, which saw substantial misclassifications.

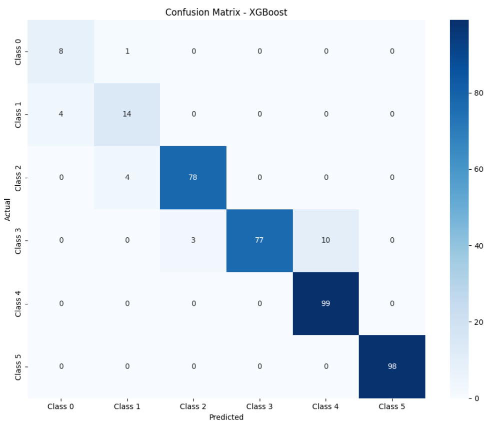

Fig. 4. Confusion Matrix - XGBoost.

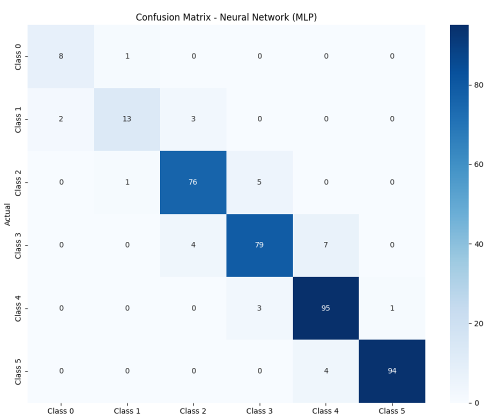

Fig. 5. Confusion Matrix - Neural Network (MLP).

B. Feature Importance

Feature importance analysis was conducted using the best-performing model, XGBoost, as shown in Fig. 9. The results overwhelmingly indicate that `Depression Value` (the raw score from the depression questionnaire) is the most influential predictor, with an importance score of approximately 0.75.

Other features contributed to the model to a much lesser extent. These included `Year_numeric` (0.013), `Stress Value` (0.013), `Anxiety Value` (0.013), `CGPA_numeric` (0.013), followed by

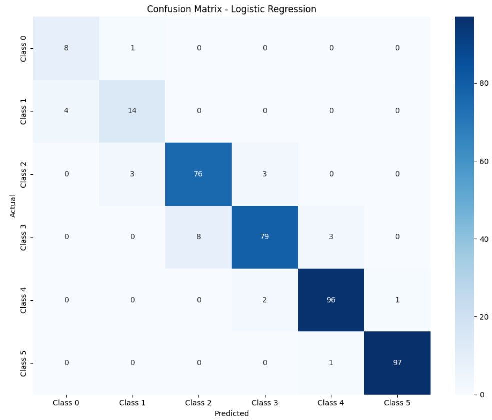

Fig. 6. Confusion Matrix - Logistic Regression.

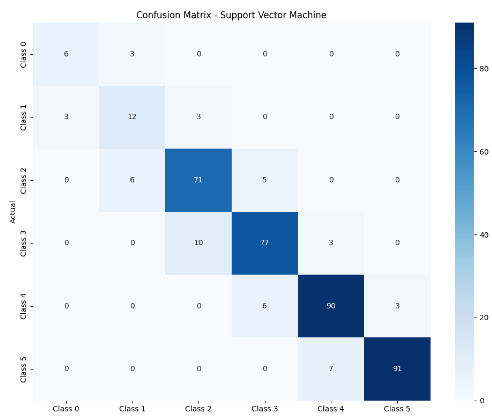

Fig. 7. Confusion Matrix - Support Vector Machine.

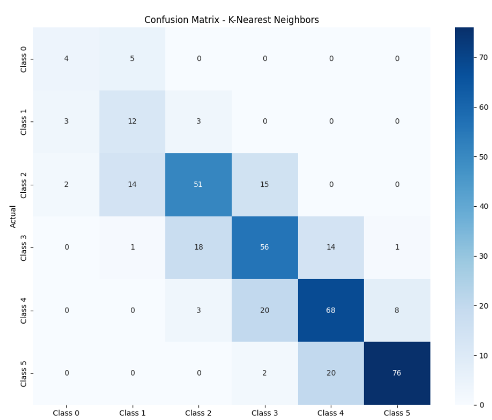

Fig. 8. Confusion Matrix - K-Nearest Neighbors.

7._Did_you_receive_a_waiver_or_scholarship_at_your_university?_No (0.012), Age_numeric (0.011), 2._Gender_Prefer_not_to_say (0.011), 2._Gender_Male (0.011), 2._Gender_Female (0.011), and 7._Did_you_receive_a_waiver_or_scholarship_at_your_university?_Yes (0.011). This distribution of feature importance suggests that while demographic and academic factors provide some predictive value, the direct assessment of depressive symptoms remains by far the most crucial indicator.

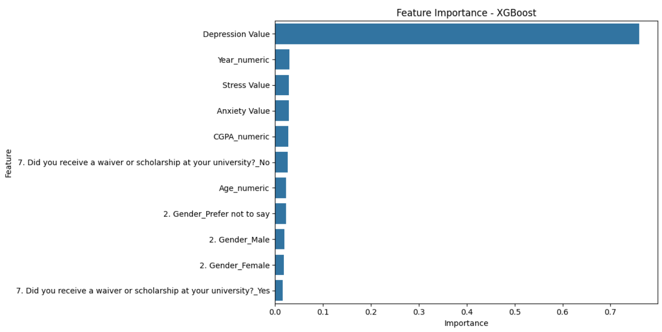

Fig. 9. Feature Importance - XGBoost.

IV. Discussion

A. Interpretation of Results

The high performance of XGBoost (94.95% accuracy) and the surprising strength of Logistic Regression (94.19% accuracy) in classifying depression severity levels demonstrates the strong predictive power of the selected features, especially after addressing class imbalance using SMOTE and augmenting data with Gaussian noise. The overwhelming importance of the `Depression Value` feature for the XGBoost model is logical, as it directly quantifies depressive symptomatology. However, the fact that other academic, demographic, and psychological (anxiety, stress) features still contribute, albeit modestly, suggests they provide auxiliary information that helps refine predictions, leading to the high accuracy.

The performance of Logistic Regression, nearly matching XGBoost, is noteworthy and indicates that even linear relationships, when combined with well-engineered features, can capture significant patterns in this type of data. The Neural Network, while powerful (92.17% accuracy), did not outperform the simpler Logistic Regression model, which may suggest that the relationship between the features and depression levels has strong linear components that don't necessarily benefit from the complex non-linear mappings of a neural network.

The lower performance of SVM (87.63% accuracy) and particularly KNN (67.17% accuracy) suggests that the feature space might have complexities or distributions (even after SMOTE) that these algorithms struggle with more than tree-based ensembles or linear models for this specific task. For KNN, the local neighborhood assumption might not hold as well across all depression severity boundaries, leading to significant confusion between adjacent classes as seen in its confusion matrix.

The confusion matrices reveal that most misclassifications occur between adjacent depression severity levels, which is understandable given the continuous nature of depression symptoms underlying our discrete categorization. This is clinically reasonable as the boundaries between, for instance, "Mild" and "Moderate" depression can be somewhat arbitrary.

B. Practical Implications

Our findings have several important practical implications:

1) **Early Detection Systems**: The high accuracy of models like XGBoost and even simpler models like Logistic Regression suggests the feasibility of developing automated screening tools. These tools could use data from brief psychological questionnaires (providing Anxiety, Stress, and Depression Values) along with readily available academic (CGPA, Year) and demographic data to identify students at risk.
2) **Resource Allocation**: Understanding the prevalence of different depression severity levels can help universities allocate mental health resources more effectively.
3) **Holistic Support**: The contribution of academic (`CGPA_numeric`, `Year_numeric`) and other psychological (`Anxiety Value`, `Stress Value`) factors, even if minor compared to `Depression Value`, supports the need for integrated mental health support that considers academic stress, anxiety, and overall student life.
4) **Financial Support Considerations**: The potential link between lack of scholarship support (indicated by one-hot encoded features) and depression, while a minor feature in the model, warrants further investigation in the broader context of student well-being. This aligns with literature on socioeconomic determinants of mental health.
5) **Model Selection for Implementation**: The strong performance of Logistic Regression suggests that in resource-constrained environments, simpler models might be sufficiently effective for depression screening, offering a balance between accuracy and computational efficiency.

C. Limitations and Future Work

Despite the strong results, this study has limitations:

1) **Sample Representativeness**: Our dataset primarily includes computer science and engineering students from one university, which may limit generalizability to other disciplines and institutions.
2) **Temporal Dynamics**: The cross-sectional nature of our data does not capture how depression levels may fluctuate throughout an academic term or year.
3) **Feature Dominance**: The overwhelming importance of the Depression Value feature suggests that our model might be largely replicating the existing depression screening tool rather than discovering novel patterns

or combinations of features that could provide early warning before clinical symptoms are evident.

4) **Feature Expansion**: Future work could consider incorporating additional predictors such as social support metrics, extracurricular involvement, sleep patterns, and substance use behaviors.
5) **Causal Relationships**: Our models identify associations but do not establish causality between academic factors and depression.
6) **Class Imbalance Consideration**: While we addressed class imbalance in training, the real-world distribution of depression severity is skewed, and model deployment should account for this when interpreting model outputs.

Future research directions include developing and validating real-time monitoring systems, conducting longitudinal studies to track depression trajectories, exploring alternative model architectures that might better capture complex interactions between features, and investigating the effectiveness of interventions based on machine learning predictions.

V. Conclusion

This study demonstrates the efficacy of machine learning approaches for detecting depression levels among undergraduate students using a combination of academic performance metrics, psychological assessments, and demographic factors. XGBoost achieved excellent classification accuracy (94.95%) in discriminating between six depression severity categories, with Logistic Regression performing similarly well (94.19%), after training on data balanced by SMOTE and augmented with Gaussian noise. The 'Depression Value' was the most dominant predictor, accounting for approximately 75% of feature importance in the XGBoost model.

The findings highlight the complex interplay between psychological states, academic performance, and demographic characteristics in undergraduate depression. The concerning prevalence of moderate to severe depression in our sample underscores the urgent need for effective detection and intervention strategies in university settings.

The practical implications of this research extend to potential applications in early warning systems, targeted support programs, and comprehensive mental health initiatives in higher education. By leveraging machine learning for depression detection, universities can move toward proactive approaches to student mental health, potentially improving academic outcomes and overall wellbeing. The strong performance of relatively simple models like Logistic Regression suggests that effective screening tools need not be computationally complex. Further validation on diverse datasets is essential to confirm the generalizability of these promising results, and future work should explore longitudinal approaches and expanded feature sets to enhance predictive capabilities and intervention strategies.

References

[1] A. K. Ibrahim, S. J. Kelly, C. E. Adams, and C. Glazebrook, "A systematic review of studies of depression prevalence in university students," *Journal of Psychiatric Research*, vol. 47, no. 3, pp. 391–400, 2013.

[2] R. Beiter, R. Nash, M. McCrady, D. Rhoades, M. Linscomb, M. Clarahan, and S. Sammut, "The prevalence and correlates of depression, anxiety, and stress in a sample of college students," *Journal of Affective Disorders*, vol. 173, pp. 90–96, 2015.

[3] J. R. Posselt and S. K. Lipson, "Competition, anxiety, and depression in the college classroom: Variations by student identity and field of study," *Journal of College Student Development*, vol. 57, no. 8, pp. 973–989, 2016.

[4] D. Eisenberg, E. Golberstein, and J. B. Hunt, "Mental health and academic success in college," *The B.E. Journal of Economic Analysis & Policy*, vol. 9, no. 1, 2009.

[5] C. Blanco, M. Okuda, C. Wright, D. S. Hasin, B. F. Grant, S. M. Liu, and M. Olfson, "Mental health of college students and their non-college-attending peers: Results from the national epidemiologic study on alcohol and related conditions," *Archives of General Psychiatry*, vol. 65, no. 12, pp. 1429–1437, 2008.

[6] D. Eisenberg, M. F. Downs, E. Golberstein, and K. Zivin, "Stigma and help seeking for mental health among college students," *Medical Care Research and Review*, vol. 66, no. 5, pp. 522–541, 2009.