

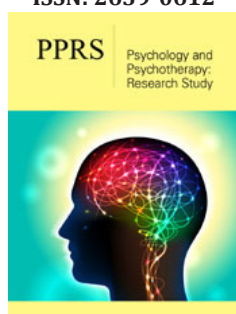
# Predicting Depression and Anxiety in Medical Students

Rija Ali<sup>1</sup>, Hammad Uallah<sup>1</sup>, Saad Ali<sup>1</sup> and Umair Arif<sup>2\*</sup>

<sup>1</sup>General practitioner at Shifa Medical Complex Jahanian District Khanewal, Pakistan

<sup>2</sup>Lecturer The University of Faisalabad, Faisalabad, Pakistan

ISSN: 2639-0612



**\*Corresponding author:** Umair Arif, General practitioner at Shifa Medical Complex Jahanian District Khanewal, Pakistan

**Submission:**  January 15, 2025

**Published:**  January 30, 2025

Volume 8 - Issue 5

**How to cite this article:** Rija Ali, Hammad Uallah, Saad Ali and Umair Arif\*. Predicting Depression and Anxiety in Medical Students. *Psychol Psychother Res Stud.* 8(5). PPRS. 000696. 2025. DOI: [10.31031/PPRS.2025.08.000696](https://doi.org/10.31031/PPRS.2025.08.000696)

**Copyright@** Umair Arif, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

## Abstract

Depression and anxiety are prevalent among medical students, impacting their academic performance and well-being. This study aims to predict mental health levels, focusing on depression and anxiety, among medical students using demographic variables such as age, sex, year of study, and health status. A cross-sectional dataset of 886 medical students was analyzed. Stepwise regression with backward selection was employed to identify significant predictors of depression, as measured by the CES-D scale. Independent variables included demographic and health-related factors. Multicollinearity was assessed using the Variance Inflation Factor (VIF), and model diagnostics were conducted using the Omnibus test, Durbin-Watson statistic, and Jarque-Bera test. The stepwise regression model explained 81.6% of the variance in depression scores (R-squared+0.816, adjusted R-squared+0.803). Anxiety (STAI\_T), Emotional Exhaustion (MBI\_EX), and health satisfaction emerged as significant predictors, with anxiety showing the strongest correlation ( $r+0.73$ ,  $p<0.001$ ). The model demonstrated a good fit, as indicated by low AIC (6007) and BIC (6050) values. Regression diagnostics showed slight deviations from normality in residuals but no significant multicollinearity or autocorrelation. Anxiety, emotional exhaustion, and health satisfaction are key predictors of depression among medical students. The findings emphasize the need for targeted mental health interventions in medical education. Addressing limitations such as residual normality can further improve predictive accuracy.

**Keywords:** Mental health; Medical students; Depression; Anxiety; Predictive modelling

## Introduction

Mental health issues among medical students are a growing concern worldwide [1]. The rigorous demands of medical education and personal, social, and financial stressors can significantly impact students' mental well-being [2]. This has led to a surge in studies focusing on the psychological challenges faced by medical students, with an emphasis on identifying the key predictors of mental health outcomes, including depression and anxiety. Addressing these issues is crucial not only for the personal welfare of students but also for ensuring their academic success and future ability to provide quality healthcare [3,4]. Depression and anxiety are among the most common mental health challenges reported by medical students [5]. Depression, often measured using the CES-D (Center for Epidemiologic Studies Depression Scale), reflects the presence and severity of depressive symptoms [6,7]. Anxiety, another significant mental health indicator, is commonly evaluated using tools like the STAI (State-Trait Anxiety Inventory). Both conditions are strongly linked to academic pressures, social isolation, lack of work-life balance, and health-related factors. Identifying and understanding the predictors of these conditions can guide the development of targeted interventions and mental health support systems for medical students [8,9].

Medical students represent a unique population due to their academic environment, which involves long study hours, exposure to emotionally taxing situations, and high expectations of performance [10]. Demographic variables such as age, gender, and cultural background play a critical role in shaping mental health outcomes. For example, gender-based differences in stress perception and coping mechanisms may influence depression and anxiety levels [11,12]. Similarly, academic factors such as the year of study and hours spent studying can

affect students' mental health, as higher academic years often bring increased pressure [13]. In addition, personal life factors, such as having a partner or holding a part-time job, may contribute to emotional resilience or additional stress [14].

Health satisfaction and lifestyle choices also significantly predict mental health outcomes [15]. Students who report higher satisfaction with their physical health tend to have better mental health. Moreover, engaging in psychotherapy or seeking mental health support can be a protective factor, helping students manage their stress and emotional challenges. Other psychological traits, such as empathy, emotional regulation, and burnout, as measured by tools like the MBI (Maslach Burnout Inventory), provide deeper insights into the emotional state of students [16-18]. Given the multi-dimensional nature of mental health, it is essential to use a comprehensive statistical approach to analyze how these diverse factors interact [19,20]. Stepwise regression is a powerful tool that enables the examination of relationships between a dependent variable, such as depression (CES-D score), and multiple independent variables [21]. This method helps identify the most significant predictors of mental health, shedding light on the relative importance of demographic, academic, and health-related factors.

This study aims to predict mental health outcomes, particularly depression, among medical students by analyzing demographic, academic, and health-related variables. By identifying key factors contributing to mental health challenges, the findings can guide effective interventions and policies to support students' academic and personal well-being. Focusing on the unique stressors faced by medical students, this research highlights the interplay between academic pressures, personal factors, and mental health. The insights gained extend beyond medical students, offering relevance to other high-stress academic and professional settings. By uncovering modifiable predictors of mental health, this study lays a foundation for future research and practical applications in mental health promotion, emphasizing evidence-based approaches to enhance well-being in high-stress populations.

## Methodology

### Study design and data collection

This study employs a quantitative, cross-sectional design to predict depression levels among medical students using a variety of demographic, academic, and health-related variables. The dataset, sourced from Kaggle, contains comprehensive information on the mental health, empathy, and burnout levels of medical students in Switzerland. Ethical standards were strictly adhered to throughout the study. No additional ethical review was required since the dataset is publicly available and anonymized. Participants in the original study provided informed consent before data collection, and the research adhered to ethical guidelines concerning data handling and participant confidentiality. Data were collected through a structured questionnaire incorporating validated scales to measure depression (CES-D), Anxiety (STAI), burnout (MBI), and Empathy (JSPE, QCAE). In addition to mental health

variables, demographic information such as age, gender, academic year, and partnership status were captured, along with health-related variables like satisfaction with health and participation in psychotherapy services.

### Preprocessing

To ensure data quality and consistency, preprocessing steps were applied [22], including imputation for missing values and handling of outliers to prevent skewed results. Categorical variables (e.g., gender and partnership status) were encoded for regression analysis, while continuous variables (e.g., age, study hours, and depression scores) were standardized for comparability. Diagnostic plots were generated to assess assumptions of linearity and normality. To evaluate the model's generalizability, the dataset was split into training and testing sets in an 80:20 ratio [23].

### Statistical analysis

The primary dependent variable in this analysis is the CES-D total score, which represents the depression levels among medical students. Independent variables (predictors) include demographic factors (age, sex, curriculum year, mother tongue, partnership status), academic factors (study hours, job status, year of study), and health-related variables (health satisfaction, psychotherapy consultation, empathy scores [JSPE, QCAE], and burnout scores [MBI: emotional exhaustion, cynicism, academic efficacy]). Multiple linear regression analysis was used to assess the relationships between these predictors and the CES-D score, to identify significant predictors of depression and quantify their effects. A stepwise regression approach was employed to select the most significant predictors, with variables included based on statistical significance. The final model was determined by evaluating p-values and R-squared values. The model's fit was assessed using R-squared, Adjusted R-squared, and p-values for individual predictors. All analyses were conducted using Python software.

### Experimental results

Out of all 886 participants, 68.4% were male and 31.6% were female (Figure 1). Out of 886 participants 160 belonged to the age group 15-19 years, 569 belonged to the age group 20-24, 131 belonged to the age group 25-29, 16 belonged to the age group 30-34 and 8 belonged to the age group 35-39 (Figure 2). The dataset reveals several insights into the demographic and lifestyle factors of the individuals. Regarding relationship status, more than half of the participants (56.32%) have a partner, while 43.68% are not in a relationship. Employment status shows that a majority (65.12%) of individuals are not in paid employment, with only 34.87% holding a paid job. In terms of mental health, a small fraction (22.46%) of individuals seeks psychiatric help, while the majority (77.53%) do not. When it comes to health satisfaction, a significant portion of the participants report being satisfied (45.37%) or very satisfied (25.28%) with their health, while 4.17% express very dissatisfaction. The distribution of students across different academic years shows that the first-year students are the largest group (27.65%), with the numbers declining gradually in higher years, suggesting that a larger proportion of individuals are still

early in their academic programs. Overall, the results indicate a trend where individuals are mostly satisfied with their health and relationships but show a low employment rate, with a smaller group seeking psychiatric assistance (Table 1).

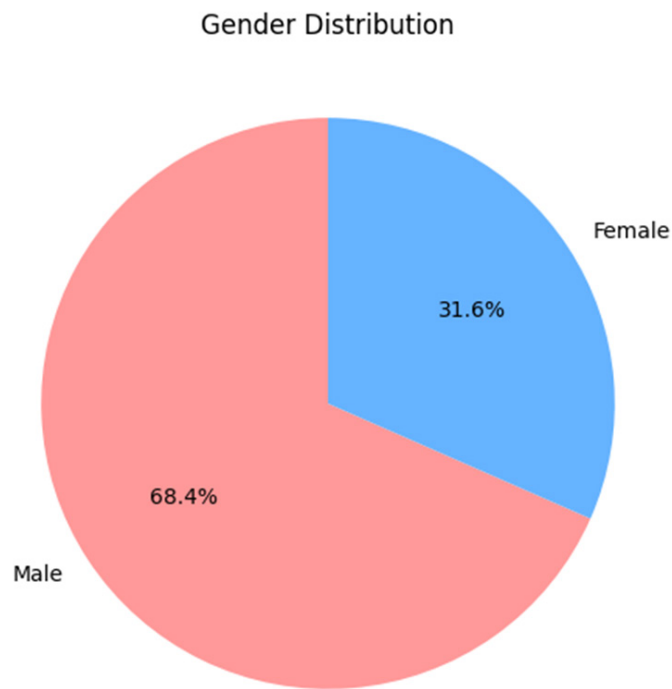


Figure 1: Gender distribution.

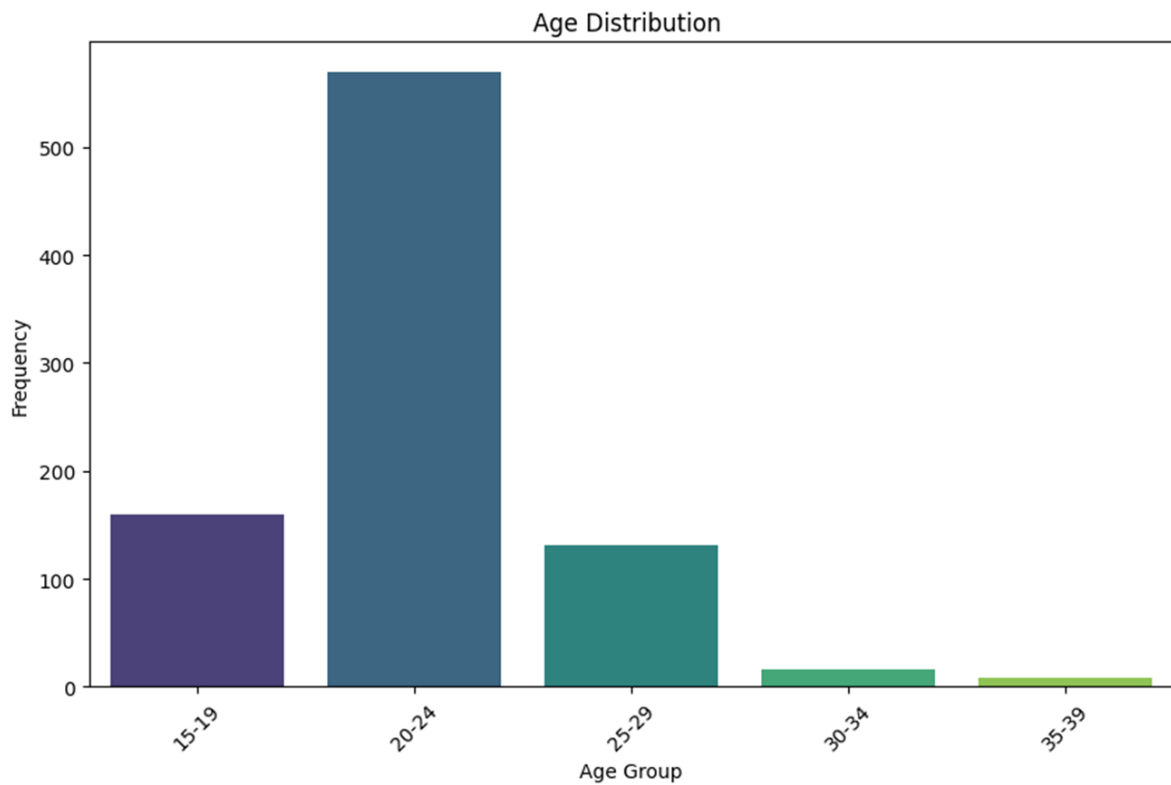


Figure 2: Age distribution.

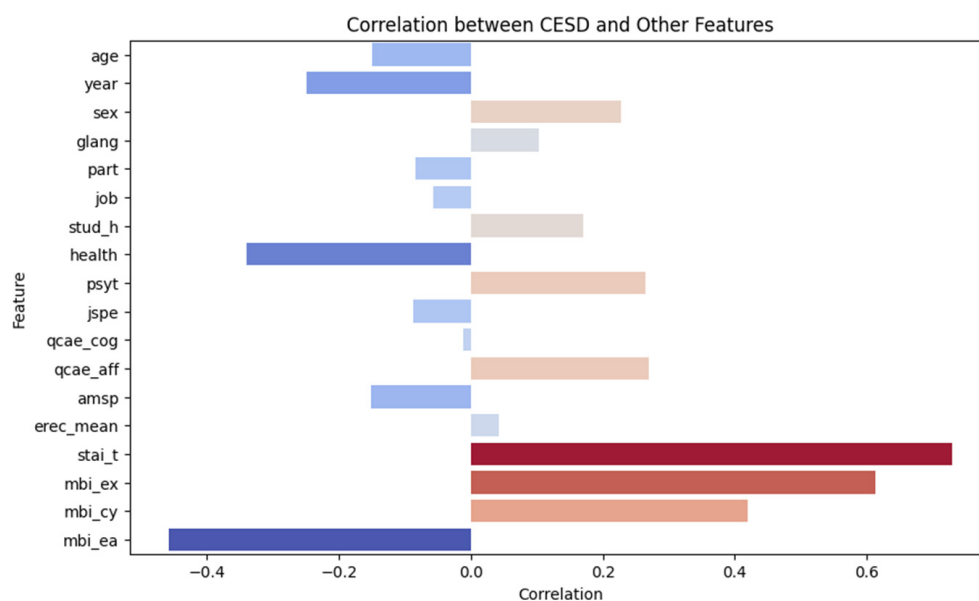
**Table 1:** Frequency distribution of demographic and lifestyle features.

Do you have a partner?	Freq		%		A Psychiatrist for your Health	Freq		%	
	Yes	No	499	56.32		Yes	No	199	22.46
Do you have a paid job?	Yes	387	43.68		Study Year	1	245	27.65	
	No	309	34.87			2	135	15.23	
How satisfied are you with your health?	Very dissatisfied	577	65.12			3	143	16.14	
	Dissatisfied	37	4.17			4	123	13.88	
	Neither	87	9.82			5	127	14.33	
	Satisfied	136	15.35			6	113	12.75	
	Very satisfied	402	45.37						
	Total	224	25.28						
	Total	886	100						

The correlation between CESD (depression) and various variables shows some notable patterns. State-Trait Anxiety (stai\_t) exhibits a strong positive correlation (0.73,  $p < 0.001$ ), suggesting that higher anxiety levels are associated with greater depression. Similarly, Emotional Exhaustion (mbi\_ex) has a moderate positive correlation (0.61,  $p < 0.001$ ), indicating that higher emotional exhaustion is linked to increased depression. Cynicism (mbi\_cy) also shows a moderate positive correlation (0.42,  $p < 0.001$ ), suggesting a relationship between higher cynicism and greater depression. In contrast, Health has a moderate negative correlation (-0.34,  $p < 0.001$ ), indicating that better health is associated with lower depression levels. Study Year (year) shows a moderate negative correlation (-0.25,  $p < 0.001$ ), suggesting that depression tends to decrease with increasing years of study. Psychiatric Treatment (psyt) is positively correlated (0.26,  $p < 0.001$ ), implying that those receiving psychiatric treatment are more likely to report higher depression levels. Finally, Cognitive-Affective Measures (qcae\_aff) exhibit a positive correlation (0.27,  $p < 0.001$ ), indicating a relationship between higher cognitive-affective issues and

depression.

These results highlight several key factors related to depression, underscoring the significance of anxiety, emotional exhaustion, health, and treatment in understanding depression levels (Figure 3). The table presents the Variance Inflation Factor (VIF) values for various features in the dataset, which measure the degree of multicollinearity among the independent variables. The constant term shows a very high VIF of 407.61, which is typically expected in regression models and indicates no direct collinearity but reflects the overall model structure. Age has a moderate VIF of 1.69, suggesting that it is not highly collinear with other features. Similarly, Sex (VIF+1.26), Glang (VIF+1.07), and Job (VIF+1.08) show relatively low multicollinearity, indicating they do not significantly overlap with other predictors in the dataset. However, some features exhibit higher VIF values. For example, Year (VIF+2.21), STAI\_T (VIF+1.97), and MBI\_EX (VIF+1.99) show moderate collinearity with other features, but they still remain within an acceptable range (typically below 5).

**Figure 3:** Correlation of features.

Features such as Stud\_h (VIF+1.42), Health (VIF+1.16), and QCAE\_Aff (VIF+1.46) also have low VIF values, suggesting minimal multicollinearity. On the higher end, MBI\_EA (VIF+1.95), QCAE\_Cog (VIF+1.39), and MPS (VIF+1.30) display VIF values just below 2, indicating a modest correlation with other variables. Overall, the dataset does not exhibit severe multicollinearity, as most features have VIF values well below 5, which would indicate problematic multicollinearity. The relatively high VIF for the constant term is expected due to its nature in the regression model. This analysis suggests that the multicollinearity in the dataset is not a significant issue and that the model should be able to perform reliably with these variables. (Table 2). In the stepwise regression using the backward selection process, several features were systematically removed from the model due to their high p-values, indicating that they were not statistically significant predictors of the target variable, *cesd* (depression). The features dropped included *job*, *amsp*, *erec\_mean*, *age*, *jspe*, *qcae\_aff*, *stud\_h*, *qcae\_cog*, *glang*, and *mbi\_ea*. The p-values of these variables ranged from 0.1136 to 0.9465, which were above the commonly used threshold of 0.05, suggesting that these variables did not contribute meaningfully to

explaining the variation in the dependent variable. The removal of these features helps simplify the model by focusing on the most influential predictors, thus improving its interpretability and potentially its predictive performance (Table 3). The results from the stepwise regression analysis indicate that a subset of independent variables significantly predict the CES-D depression score. The model explains 81.6% of the variance in the CES-D scores, as indicated by an R-squared of 0.816, with an adjusted R-squared of 0.803, suggesting that the selected features contribute meaningfully to the model. The F-statistic of 176.1, with a very low p-value of 1.18e-176, confirms that the regression model as a whole is highly significant. The variables retained in the model were selected through stepwise regression, removing features with high p-values, indicating that they did not contribute significantly to explaining the variability in CES-D scores. The low AIC (6007) and BIC (6050) values further suggest that the chosen model is well-fit, with minimal overfitting. Overall, the stepwise regression model presents a parsimonious yet robust predictor for CES-D, highlighting the importance of carefully selecting relevant features in multivariable analysis (Table 4).

**Table 2:** Variance Inflation Factor (VIF) for the Features.

No	Feature	VIF	No	Feature	VIF	No	Feature	VIF
1	const	408	8	stud_h	1.419088	15	erec_mean	1.086157
2	age	1.69	9	health	1.162814	16	stai_t	1.968893
3	year	2.21	10	psyt	1.1434	17	mbi_ex	1.987494
4	sex	1.26	11	jspe	1.315532	18	mbi_cy	1.837228
5	glang	1.07	12	qcae_cog	1.385918	19	mbi_ea	1.948829
6	part	1.07	13	qcae_aff	1.461311			
7	job	1.08	14	amsp	1.300699			

**Table 3:** Dropped out features.

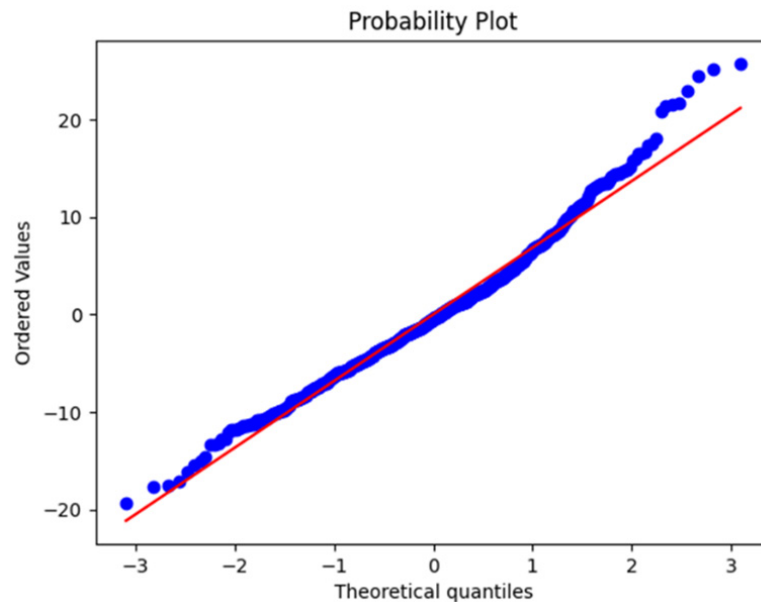
Features	P-values	Features	P-values
Job	0.9465	qcae_aff	0.6724
Amsp	0.8645	stud_h	0.4659
erec_mean	0.7951	qcae_cog	0.3938
age	0.6474	glang	0.1216
jspe	0.6635	mbi_ea	0.1136

**Table 4:** Stepwise regression model summary.

Variables	Coeff	St. Err	T	P-value	95% CI
const	-8.3205	1.8473	-4.50409	7.57E-06	[-11.9462, -4.69483]
year	-0.6491	0.1467	-4.4244	1.09E-05	[-0.93707, -0.36117]
sex	1.1829	0.5294	2.234225	0.02572	[0.143781, 2.222169]
part	-1.0475	0.4924	-2.12696	0.033702	[-2.01409, -0.08091]
health	-1.1132	0.2413	-4.61321	4.56E-06	[-1.58693, -0.63964]
psyt	1.3919	0.607	2.293142	0.022076	[0.2006, 2.58334]
stai_t	0.462	0.0252	18.26466	2.08E-63	[0.412416, 0.511721]
mbi_ex	0.4819	0.0621	7.752664	2.49E-14	[0.359952, 0.603983]
mbi_cy	0.2883	0.0631	4.567565	5.64E-06	[0.164437, 0.412229]
R-squared	adjusted R-squared	F-stat	p-value	AIC	BIC
0.816	0.803	176.1	1.18E-11	6007	6050

The regression model diagnostics provide several important insights into the model's assumptions and potential issues. The Omnibus test (68.262) with a p-value of 0.000 indicates that the residuals of the model are not normally distributed, suggesting a potential violation of the assumption of normality. The Jarque-Bera (JB) test confirms this, with a significant p-value (6.76e-21), indicating that the distribution of residuals deviates significantly from normality, especially in terms of skewness and kurtosis. The

skewness value of 0.626 suggests a slight positive skew in the residuals, while the kurtosis value of 3.974 is close to the ideal value of 3, indicating near-normal peakedness but with a tendency towards slightly heavier tails. The Durbin-Watson statistic of 2.065 indicates that there is little to no autocorrelation in the residuals, which is a good sign for the model's validity in time series or sequential data (Figure 4).



**Figure 4:** Probability plot.

## Conclusion

This study provides valuable insights into the factors influencing depression, as measured by the CES-D scale, in a sample of 886 individuals. The stepwise regression model, which explained 81.6% of the variance in depression scores, identified significant predictors such as anxiety, emotional exhaustion, and health satisfaction. The analysis revealed that anxiety and emotional exhaustion were strongly correlated with higher depression levels, while better health was associated with lower depression. Additionally, the study highlighted that the majority of participants are generally satisfied with their health, but a considerable proportion do not seek psychiatric treatment. The results suggest that mental health interventions focusing on anxiety and emotional exhaustion could be beneficial in addressing depression in this population. Moreover, the study demonstrated the importance of carefully selecting relevant predictors to improve model interpretability and predictive performance.

## Limitations

Despite the valuable insights provided by the regression model, there are several limitations to consider. First, the model's assumption of normality in residuals was violated, as evidenced by the Omnibus and Jarque-Bera tests. This suggests that the model may not fully capture the complexity of the data. Additionally, although the multicollinearity among variables was not a

significant issue, the possibility of omitted variable bias remains a concern. Finally, the study relied on self-reported measures, which can introduce biases such as social desirability or recall bias. Future studies could address these limitations by using more diverse samples, alternative data collection methods, and testing the model's robustness in different settings.

## References

1. Mirza AA, Mukhtiar B, Ghada M, Mohammed A, Abdulrahim A, et al. (2021) Depression and anxiety among medical students: A brief overview. *Advances in Medical Education and Practice* 12: 393-398.
2. Halperin, Matthew N, Sofia P, Jonathan N (2021) Prevalence of anxiety and depression among medical students during the Covid-19 pandemic: A cross-sectional study. *Journal of Medical Education and Curricular Development* 8: 2382120521991150.
3. Pokhrel NB, Khadayat R, Tulachan P (2020) Depression, anxiety, and burnout among medical students and residents of a medical school in Nepal: A cross-sectional study. *BMC Psychiatry* 20(1): 298.
4. Shawahna R (2020) Prevalence and factors associated with depressive and anxiety symptoms among Palestinian medical students. *BMC Psychiatry* 20(1): 244.
5. Capdevila G, Miguel J, Diego F, Mila G, Joaquín G (2021) Depression, anxiety, burnout and empathy among Spanish medical students. *PLoS One* 16(12): e0260359.
6. Ramadianto AS, Irmia K, Feranindhya A (2022) Symptoms of depression and anxiety in Indonesian medical students: association with coping strategy and resilience. *BMC Psychiatry* 22(1): 92.

7. Schwaab L, Nadja G, Hans C, Christoph N (2022) Climate change related depression, anxiety and stress symptoms perceived by medical students. *International Journal of Environmental Research and Public Health* 19(15): 9142.
8. Stanković M, Nešić M (2022) Association of internet addiction with depression, anxiety, stress, and the quality of sleep: Mediation analysis approach in Serbian medical students. *Current Research in Behavioral Sciences* 3: 100071.
9. Cimen ID, Tuncay M, Bülent C, Nur E (2022) Mental health of Turkish medical students during the COVID-19 pandemic. *International Journal of Social Psychiatry* 68(6): 1253-1262.
10. Haykal KA, Lara P, Aidan P, Karine F (2022) Medical student wellness assessment beyond anxiety and depression: A scoping review. *PLoS One* 17(10): e0276894.
11. Teshome BG (2022) Marginalized identities, mistreatment, discrimination, and burnout among US medical students: Cross sectional survey and retrospective cohort study. *BMJ* 376: e065984.
12. Jiang QH, Horta H, Yuen M (2022) International medical students' perspectives on factors affecting their academic success in China: A qualitative study. *BMC Medical Education* 22(1): 574.
13. Khazaie H, Javad Y, Jaffar A, Behzad M, Fakhreddin C (2023) Internet addiction status and related factors among medical students: a cross-sectional study in Western Iran. *Community Health Equity Research & Policy* 43(4): 347-356.
14. Huarcaya V, Claudia E, Diego C, Luis V, Andrés P (2023) Factors associated with mental health in Peruvian medical students during the COVID-19 pandemic: A multicentre quantitative study. *Revista Colombiana De Psiquiatría* 52(3): 236-244.
15. Senmar M (2023) Relationship between spiritual intelligence and lifestyle with life satisfaction among students of medical sciences. *BMC Medical Education* 23(1): 520.
16. Andraus GS (2023) Associations between lifestyle and sociodemographic factors in medical students: A cross-sectional study. *Journal of Lifestyle Medicine* 13(1): 73-82.
17. Sanchez D (2023) Sociodemographic aspects and healthy behaviors associated with perceived life satisfaction in health professionals. *Journal of Primary Care & Community Health* 14: 21501319221148332.
18. Pasarica M, Boring M, Lessans S (2022) Current practices in the instruction of lifestyle medicine in medical curricula. *Patient Education and Counseling* 105(2): 339-345.
19. Garg M (2024) Mental disturbance impacting wellness dimensions: Resources and open research directions. *Asian Journal of Psychiatry* 92: 103876.
20. Arslan G, Genç E (2022) Psychological maltreatment and college student mental wellbeing: A uni and multi-dimensional effect of positive perception. *Children and Youth Services Review* 134: 106371.
21. Çakmakkaya OS, Elif G, Ali M, Mehmet S, Öner S (2024) Factors affecting medical students' satisfaction with online learning: A regression analysis of a survey. *BMC Medical Education* 24(1): 11.
22. Özkan Y, Demirarslan M, Suner A (2024) Effect of data preprocessing on ensemble learning for classification in disease diagnosis. *Communications in Statistics-Simulation and Computation* 53(4): 1657-1677.
23. Albahra S (2023) Artificial intelligence and machine learning overview in pathology & laboratory medicine: A general review of data preprocessing and basic supervised concepts. *Semin Diagn Pathol* 40(2): 71-87.