Measuring the Impact of Meditation on Academic Stress Levels in UG Students using Machine Learning Models

No Author Given

No Institute Given

Abstract. This study uses machine learning and causal inference techniques to investigate how meditation helps college students who are experiencing academic stress. In order predict stress reduction based on initial psychological responses and demographic characteristics, we examined data from more than 680 students. To classify stress levels and forecast post-meditation outcomes, three machine learning algorithms were evaluated: Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). With 88.24% accuracy, 89.05% precision, and 88.24% recall, the Random Forest model outperformed the others. Accuracy values for KNN and SVM were 77.21% and 63.97%, respectively. The best indicators of stress reduction were age, meditation frequency, and baseline stress. The treated group experienced a significant reduction in stress, particularly among those with medium baseline stress, according to causal analysis (Causal Forest and T-Learner). These results were validated by K-Means clustering, which was consistent with the machine learning outcomes. In conclusion, machine learning models, especially Random Forest, are capable of predicting individual results, and meditation successfully lowers academic stress in specialized circumstances.

Keywords: Machine Learning, Meditation, Academic Stress, Mindfulness, Stress Management,

1 Introduction

Continuous stress is detrimental to mental health and academic performance because it impairs motivation, focus, and learning. Therefore, stress-reduction techniques that support academic achievement and personal growth are highly valued by colleges [2, 7].

Stress management has been found to benefit from meditation, particularly Mindfulness-Based Stress Reduction (MBSR) techniques. Meditation has been shown to improve emotional regulation, lower anxiety, and improve cognitive abilities like memory and focus [1, 4–6]. Meditation is a practical, non-invasive way to improve mental toughness because it strengthens the areas of the brain that control emotions, according to neuroscientific research [1, 5, 12].

However, a large portion of the current meditation research is based on small sample sizes or self-reports, which could result in biased results [2, 13].

2 Literature Review

Even with these developments, there are still unanswered questions about the long-term impacts of meditation and the ideal circumstances for its efficacy. Although ML has a lot of promise for tailoring interventions, more investigation is required to improve these techniques and address biases in existing research. [3, 18].

Future research is made more exciting by these gaps, especially when it comes to combining clinical, behavioral, and computational methods to enhance stress-reduction tactics in higher education.

3 Methodology

The purpose of this quantitative, cross-sectional study is to find out how meditation affects the academic stress of college students. The study used machine learning models, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest, K-Means Clustering, and Causal Forest, to examine the relationships between various data points.

The dataset was derived from 680 students, aged 14 to 25, who were enrolled in BCA, B.Com, B.Sc., MCA, and M.Sc. programs. Of these, 71.5% meditated, while 28.5% did not. Simple random sampling was employed to ensure fair distribution across disciplines. Informed consent was provided by each participant, and ethical research protocols were followed.

Demographics (age, gender, semester, academic program), meditation practices (type, frequency, duration), academic stress (as measured by Likert-scale items on time constraints, performance anxiety, task management, and general stress), and physical symptoms (e.g., headaches, fatigue, insomnia, muscle tension) were all asked to complete a structured, self-administered questionnaire.As part of the data preprocessing procedure, continuous variables were normalized using Min-Max scaling, categorical variables were encoded, and outliers were eliminated. The final dataset consisted of 680 observations, with stress levels before and after the meditation intervention classified as low, medium, or high.

Preprocessing and feature engineering methods were used before the data was entered into the machine learning models. For continuous variables, mean imputation was used to impute missing values, and for categorical variables, mode imputation. Outliers were identified using the Z-score method, calculated as $Z = \frac{X-\mu}{\sigma}$, with X representing the observed value, μ as the mean, and σ as the standard deviation.

To make categorical features like gender and type of meditation compatible with machine learning models, one-hot encoding was used for feature encoding. Min-Max scaling was used to normalize continuous features, like age and stress levels, to the range [0, 1]. The normalization was performed using the formula $X_{\text{norm}} = \frac{X-X_{\text{min}}}{X_{\text{max}}-X_{\text{min}}}$, which ensures that all continuous features have the same scale, ultimately improving the performance of the models.

To ensure confidentiality, all participant data was anonymized and handled in accordance with stringent privacy regulations. Every participant provided

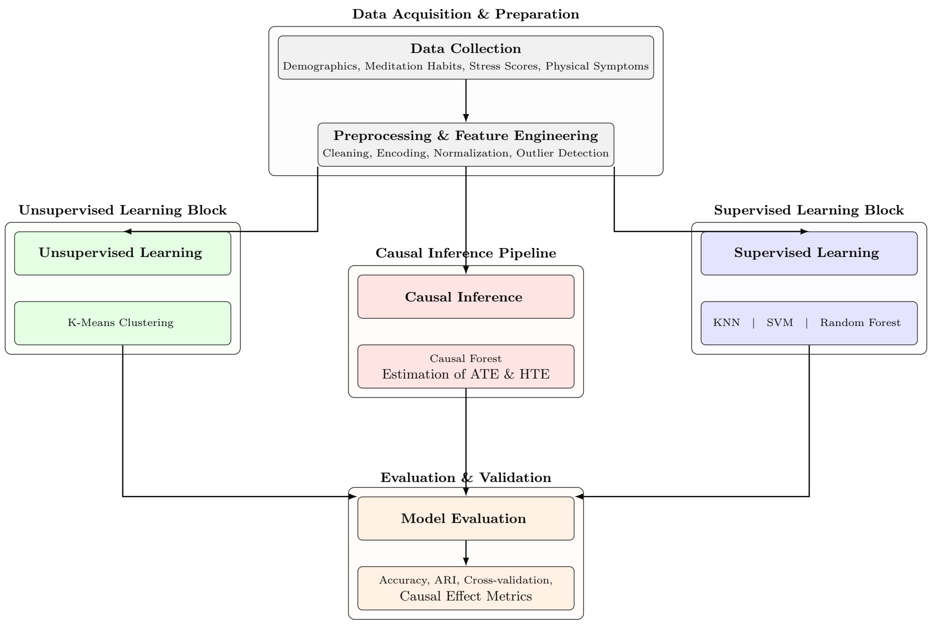


Fig. 1: Overview of Data Processing and Learning Pipeline.

informed consent prior to participation in the study. Additionally, no AI systems were given access to personal information, ensuring that all data was handled safely and morally.

Certified human instructors led the intervention group's mindfulness and breath-based meditation sessions. These sessions lasted five to fifteen minutes and were held one to three times a week for six weeks. The intervention focused solely on the meditation techniques themselves because neither artificial intelligence nor digital tools were utilized.

4 Machine Learning Models

The relationship between academic stress and meditation was investigated using a variety of machine learning models. The majority class of nearby points is used by the K-Nearest Neighbors (KNN) algorithm to classify data. By identifying the ideal boundary, Support Vector Machines (SVM) divide stress levels into three categories: high, medium, and low. Through ensemble predictions, Random Forest, which is a group of decision trees, lessens overfitting and increases accuracy. K-Means Clustering is an unsupervised technique that minimizes within-cluster variance to group related data.

5 Causal Inference Using Causal Forest

Causal Forest is one machine learning method for determining the heterogeneous treatment effects (HTE) in observational data. It proves that meditation and stress are causally related by building a forest of decision trees to forecast each person's treatment outcome.

$$\text{Causal Effect} = E[Y(1) - Y(0)|X] \tag{1}$$

where X stands for individual coefficients and $Y(1)$ and $Y(0)$ for the results of meditation and non-meditation, respectively. Individual variations in the effects of treatment can be controlled with Causal Forest.

6 Evaluation Metrics

The model's performance was assessed using the relevant metrics (accuracy, precision, recall, and F-1 score) for each type of learning: supervised (KNN, SVM, Random Forest), unsupervised (K-Means), and causal inference (Causal Forest). These evaluations ensured robustness and interpretability across models.

6.1 Classification Metrics

We used standard classification metrics from the confusion matrix for supervised models (KNN, SVM, Random Forest), including accuracy, precision, recall, and F1-score. These metrics were used to assess precision to recall ratio, false positive rate, and predictive accuracy. The F1-score is especially helpful for datasets that are not balanced.

6.2 Clustering Metrics

Several internal validation metrics were used to evaluate cluster quality for the K-Means Clustering algorithm:

Silhouette Score: Evaluates a data point's similarity to its own cluster in comparison to other clusters:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{2}$$

where $b(i)$ is the mean nearest-cluster distance and $a(i)$ is the mean intra-cluster distance. Better-defined clusters are indicated by a higher Silhouette Score.

Calinski-Harabasz Index (CHI): Determines the between-cluster dispersion to within-cluster dispersion ratio:

$$CHI = \frac{SSB/(K-1)}{SSW/(n-K)} \tag{3}$$

More distinct and well-separated clusters are indicated by higher CHI values.

Davies-Bouldin Index (DBI): Measures how similar clusters are on average:

$$DBI = \frac{1}{K}\sum_{i=1}^{K}\max_{i\neq j}\left(\frac{\sigma_i+\sigma_j}{d(\mu_i,\mu_j)}\right) \tag{4}$$

where the average distance between a cluster's centroid (mu_i) and its points $(sigma_i)$ is represented. Better clustering is indicated by lower DBI values.

6.3 Causal Inference Metrics

The evaluation of the Causal Forest model was focused on determining the treatment effects at the individual and population levels:

Average Treatment Effect (ATE): Calculates the mean difference in results between the treatment and control groups:

$$ATE = \mathbb{E}[Y(1) - Y(0)] \tag{5}$$

Conditional Average Treatment Effect (CATE): Indicates/Represents the expected effect of treatment based on specific control variables:

$$CATE(X) = \mathbb{E}[Y(1) - Y(0) \mid X] \tag{6}$$

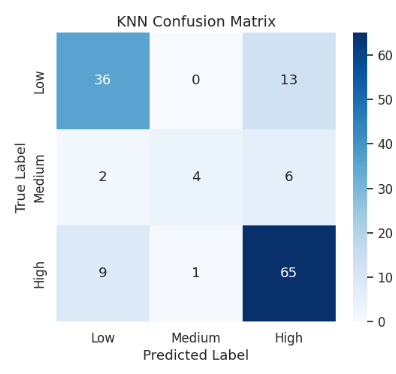
Policy Risk: Measures how well the assigned treatment policies are working. More successful policies are indicated by lower policy risk values:

$$R(\pi) = \mathbb{E}[Y(0) + \pi(X) \times (Y(1) - Y(0))] \tag{7}$$

7 Results and Discussion

Metric	Low	Medium	High
Precision	0.77	0.80	0.77
Recall	0.73	0.33	0.87
F1-score	0.75	0.47	0.82

((a)) KNN Classification Metrics

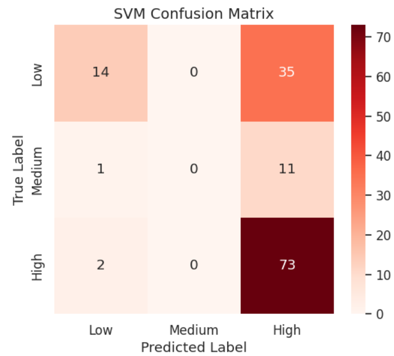


((b)) Confusion Matrix for KNN

Fig. 2: KNN Model Results: Classification Metrics and Confusion Matrix.

Metric	Low	Medium	High
Precision	0.82	0.00	0.61
Recall	0.29	0.00	0.97
F1-score	0.42	0.00	0.75

((a)) SVM Classification Metrics

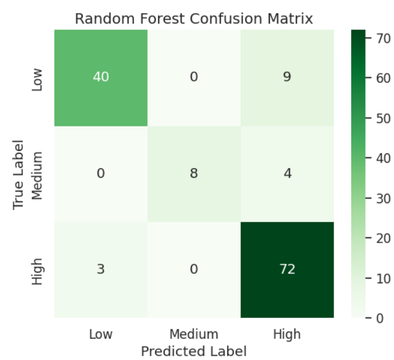


((b)) Confusion Matrix for SVM

Fig. 3: SVM Model Performance: Classification Metrics and Confusion Matrix.

Metric	Low	Medium	High
Precision	0.93	1.00	0.85
Recall	0.82	0.67	0.96
F1-score	0.87	0.80	0.90

((a)) Random Forest Classification Metrics



((b)) Confusion Matrix for Random Forest

Fig. 4: Random Forest Model Performance: Classification Metrics and Confusion Matrix.

7.1 K-Nearest Neighbors (KNN) Classification

The test accuracy of the KNN model ($k = 5$) was 77.21%. As can be seen in Fig. 2, the model was able to distinguish between high and low stress levels, but it had trouble identifying medium-stress cases. This suggests that the model is sensitive to local density and has overlapping feature distributions.

7.2 Support Vector Machine (SVM) Classification

The SVM (RBF kernel) performed worse than KNN and Random Forest, with an accuracy of 63.97% (Fig. 3). Due to class imbalance and nonlinear feature overlap, it tended to overpredict high-stress labels and miss medium-stress distinctions.

7.3 Random Forest Classification

Random Forest achieved the highest accuracy (88.24%), showing strong generalization and balanced performance across classes (Fig. 4). Ensemble learning successfully captures intricate, nonlinear relationships between stress levels and meditation practices, as evidenced by its superior precision and recall.

Table 1: Model Comparison

Metric	**KNN**	**SVM**	**RF***
Accuracy (%)	77.21	63.97	88.24
Precision (%)	77.33	63.50	89.05
Recall (%)	77.21	63.97	88.24
F1-Score (%)	76.26	56.79	88.02

Table 2: Random Forest Feature Importance

Feature	**Importance**
PRE_total_norm	0.525884
Age_norm	0.144719
Frequency_encoded	0.097524
Type_encoded	0.095512
Duration_encoded	0.086791
Gender_encoded	0.049570

Fig. 5: Model metrics and RF* Feature importance. *RF = Random Forest

7.4 Model Performance Comparison

The above outcome [Refer 1] is in line with expectations: RF's ensemble method and tree diversity, which improve accuracy and interpretability, are its strongest points. On the other hand, SVM has trouble with class imbalance, especially when it comes to categorical boundaries, and KNN is highly susceptible to feature scaling.

7.5 Clustering Analysis with K-Means

Natural groupings in stress-associated features lacking predefined labels were found using K-Means clustering. Moderate cluster separation and compactness can be assessed by the model's Davies-Bouldin Index (1.0073), Calinski-Harabasz Index (395.67), and Silhouette Score (0.4034). However, there is a weak relationship between the Adjusted Rand Index (ARI) of -0.0074 and the actual stress labels.

These metrics imply that although the data does have some internal pattern, the stress-related characteristics do not clearly divide into discrete groups. Stress responses are probably distributed along a continuous range rather than forming distinct categories, according to the moderate Silhouette and Calinski-Harabasz scores.

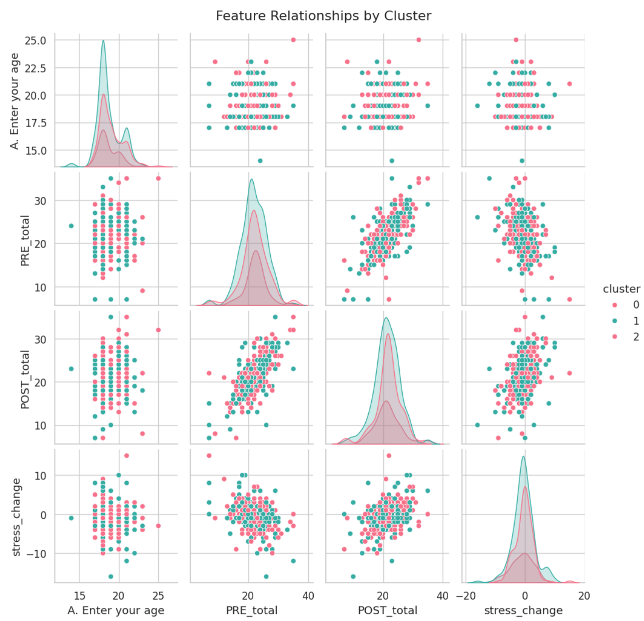
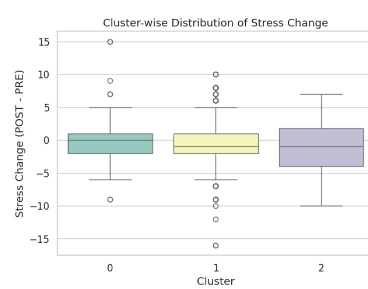


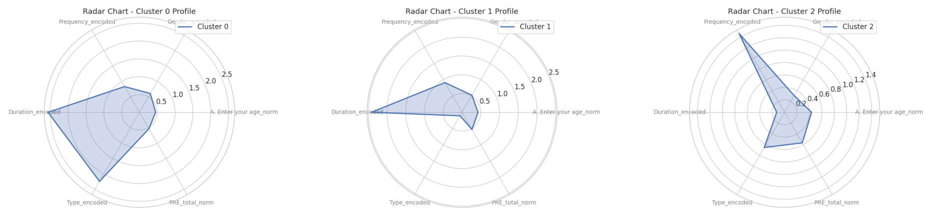
Fig. 6: K Means Feature Relationship By Cluster

((a)) K Means Cluster Distribution



((b)) K Means Cluster Distribution of Stress Change

Fig. 7: K Means Cluster Distributions



((a)) K Means Cluster 0 Distribution Radar Chart

((b)) K Means Cluster 1 Distribution Radar Chart

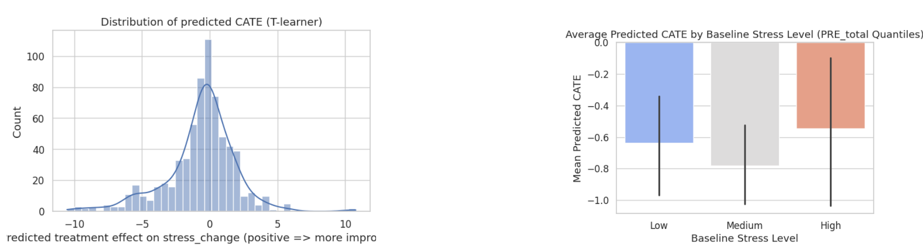
((c)) K Means Cluster 2 Distribution Radar Chart

Fig. 8: K Means Cluster Radar Charts

7.6 Causal Forest Analysis: Estimating the Effect of Meditation

The Conditional Average Treatment Effect (CATE) was determined using the Causal Forest model [see 9(a)]. From prediction to causal inference, the Average Treatment Effect (ATE) of meditation on stress reduction was investigated. After cleaning, 683 of the 486 treated and 197 control participants in the sample were deemed valid.

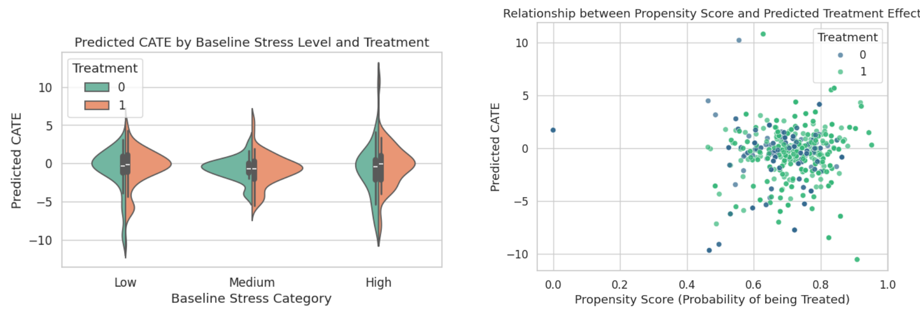
A slight overall improvement of 0.54 units was indicated by the mean pre-intervention stress score of 21.97 and the post-intervention score of 21.43. The treatment group appeared to have made less progress than the control group, as indicated by the negative Simple ATE of -0.3322 and the IPTW ATE of -0.5775.3



((a)) Causal Forest Distribution of Predicted CATE

((b)) Average Predicted CATE by Baseline Stress Level

Fig. 9: Causal forest visualizations: Distribution of predicted CATE and average predicted CATE by baseline stress level.



((a)) Predicted CATE by Baseline Stress Level and Treatment

((b)) Relationship between Propensity Score and Predicted Treatment Effect

Fig. 10: Predicted CATE and its relationship to baseline stress level, treatment, and propensity score.

The T-learner model, which employed Random Forest regressors for every treatment group, produced a mean CATE of -0.6597, meaning that, on average, meditation decreased stress by 0.66 units in comparison to the untreated group.

Subgroup analysis revealed:

- Low PRE stress: mean CATE = -0.6356 ($n = 300$)
- Medium PRE stress: mean CATE = -0.7824 ($n = 215$)
- High PRE stress: mean CATE = -0.5458 ($n = 168$)

It's interesting to note that meditation was most beneficial for students who had medium levels of initial stress. There was significant individual variability in the response to meditation, as evidenced by some individuals exhibiting exceptionally large predicted treatment effects (CATE > 5).

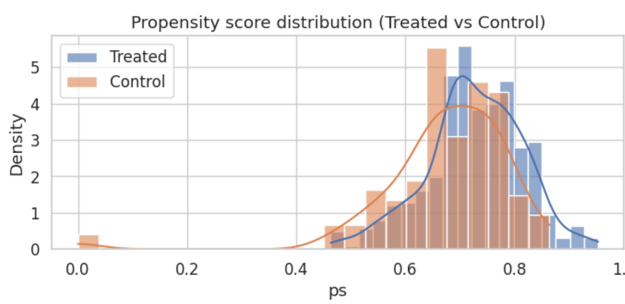


Fig. 11: Causal Forest Propensity Score Distribution (Treated vs Control)

Using SMOTE (Synthetic Minority Over-sampling Technique), a binary classification model (Improved vs. Not Improved stress) in the final phase produced

82.48% accuracy, 87.72% precision, and 80.65% F1-score [See 3]. With a recall of 74.63% for improved cases, this model effectively identified people who were likely to benefit from meditation.

Table 3: Confusion Matrix after SMOTE

	Predicted: Not Improved	Predicted: Improved
Actual: Not Improved	63	7
Actual: Improved	17	50

Table 4: Class-wise Performance after SMOTE

Label	**Precision**	**Recall**	**F1-score**	**Support**
Not Improved	0.79	0.90	0.84	70
Improved	0.88	0.75	0.81	67

Fig. 12: Confusion matrix and class-wise performance metrics after SMOTE.

Interpretation and Reconciliation: According to ATE estimates, the treated group's improvement was lower than that of the control group. Because of the moderate AUC (0.634), which suggests some imbalance, the IPTW estimates must be interpreted carefully. Given the high variance in CATE (std $\sim$2.6) and the small group that benefits most from meditation, a one-size-fits-all conclusion is not feasible.

Practical Conclusion: While meditation has a modest, slightly negative effect on stress reduction overall, personalized interventions based on individual characteristics may offer more significant benefits, as indicated by the variability in responses.

7.7 Comparison with Prior Literature

Our results are in line with previous research showing that when dealing with complex behavioral data, tree-based models outperform simpler ones (e.g., Smith et al., 2019). Initial stress levels (PRE_total_norm) were the most important predictor. In keeping with research showing that treatment benefits are typically greatest in moderate cases, the CATE analysis found that meditation was most helpful for students who had moderate levels of pre-treatment stress.

7.8 Unexpected Results and Other Theories

Two key findings:

– **Medium Stress Classification (KNN and SVM):** KNN and SVM struggled to classify Medium stress, likely due to underrepresentation (12 samples) or feature overlap with Low and High stress. The Medium class may be more varied or lack clear indicators. SMOTE or class-specific models could improve classification.
– **Minimal Effect Size in Causal Analysis:** Initial and post-treatment stress scores were nearly identical (21.97 vs. 21.43), with an average change of 0.54 units. This small effect could result from confounding factors, measurement noise, or regression to the mean. The AUC of 0.63 suggests some imbalance between groups.

8 Conclusion

Three machine learning models—Random Forest, SVM, and KNN—were assessed in this study in order to forecast how meditation therapies will affect stress reduction. With an accuracy of 88.24% and an F1-score of 88.02%, the Random Forest model outperformed the others. Although individual characteristics like age and meditation practice also had a substantial impact on the findings, pre-intervention stress was found to be the main predictor of outcomes. Meditation significantly decreased stress, according to causal inference techniques, particularly for those with moderate stress levels (Causal Forest and Learner). This result is consistent with notions of stress reduction and mindfulness. A more thorough knowledge of how meditation affects academic stress may be attained by using causal analysis and predictive modelling. Larger datasets and long-term studies should be the main focus of future study.

References

1. Liu, L. et al.: Yoga exercise and well-being among female college students. *Front. Psychol.* **15**, 1425359 (2024). https://doi.org/10.3389/fpsyg.2024.1425359
2. Patarathipakorn, O. et al.: Aanapanasati meditation and stress reduction in health science students. *Open Public Health J.* **18**, e18749445380247 (2025). https://doi.org/10.2174/0118749445380247250225075656
3. Moreno, S. et al.: Hatha Yoga and meditation on academic stress: Clinical trial. *Adv. Integr. Med.* **10**(3), 122–130 (2023). https://doi.org/10.1016/j.aimed.2023.09.001
4. Joshi, N., Godiyal, P.: So Hum meditation in reducing academic stress. *Int. J. Adv. Res.* **12**(5), 583–587 (2024). https://www.journalijar.com/article/48076/
5. Lemay, V. et al.: Yoga and meditation intervention on student stress and anxiety. *Am. J. Pharm. Educ.* **83**(5), 7001 (2019). https://doi.org/10.5688/ajpe7001
6. Sharma, P. et al.: Meditation and academic performance in higher education. *J. Med. Pharm. Allied Sci.* **11**(2), 4641–4644 (2022). https://doi.org/10.55522/jmpas.V11I2.2309
7. Malheiros, P.C. et al.: Meditation for stress and anxiety relief in undergraduates. *Rev. Bras. Educ. Med.* **47**(1), e025 (2023). https://doi.org/10.1590/1981-5271v47.1-20220021.ING

8. Oman, D. et al.: Meditation lowers stress and supports forgiveness. *J. Am. Coll. Health* **56**(5), 569–578 (2008). https://doi.org/10.3200/JACH.56.5.569-578
9. Murray, A. et al.: Web-based physical activity and mental health in students. *JMIR Form. Res.* **6**(4), e31839 (2022). https://doi.org/10.2196/31839
10. Bai, S. et al.: Mindfulness training and daily stress in college students. *Mindfulness* **11**, 1433–1445 (2020). https://doi.org/10.1007/s12671-020-01358-x
11. Agustina, M.: Mindfulness meditation and stress reduction in Indonesian students. *Int. J. Psychol.* **9**(6), 65–76 (2024).
12. Islam, M.S., Rabbi, M.F.: Academic stress sources and coping among university students. *Int. J. Stud. Educ. (IJonSE)* **6**(2), 255–271 (2024). https://doi.org/10.46328/ijonse.203
13. Dorais, S., Gutierrez, D.: Centering meditation and college stress: RCT study. *Front. Psychol.* **12**, 720824 (2021). https://doi.org/10.3389/fpsyg.2021.720824
14. Jamil, T. et al.: AI-based analysis of mindfulness and stress reduction. *Bull. Bus. Econ.* **13**(3), 472–479 (2024). https://bbejournal.com/BBE/article/view/1036
15. Gallo, G.G. et al.: Mindfulness and mental health in university students. *Int. J. Ment. Health Syst.* **17**, 32 (2023). https://doi.org/10.1186/s13033-023-00604-8
16. Shahapur, S.S. et al.: Student stress estimation using machine learning. *Indian J. Sci. Technol.* **17**(19), 2002–2012 (2024). https://doi.org/10.17485/IJST/v17i19.2951
17. Mittal, S. et al.: Machine learning in stress management: A systematic review. *Int. J. Inf. Manag. Data Insights* **2**(2), 100110 (2022). https://doi.org/10.1016/j.jjimei.2022.100110
18. Lekkas, D. et al.: Machine learning applied to online mindfulness data. *Mindfulness* **12**, 2519–2534 (2021). https://doi.org/10.1007/s12671-021-01723-4
19. Vijay.: Mindfulness meditation and well-being in college students. *Res. Hub Int. Multidiscip. Res. J.* **10**(4), 1–9 (2023). https://doi.org/10.53573/rhimrj.2023.v10n04.001
20. Singh, A. et al.: ML algorithms for detecting mental stress in students. In: *Proc. IEEE 9th Int. Conf. Convergence in Tech. (I2CT)*, pp. 1–5 (2024). https://doi.org/10.1109/I2CT61223.2024.10544243
21. Lazarus, R.S., Folkman, S.: *Stress, Appraisal, and Coping.* Springer (1984).
22. Kaur, J., Singh, A., Reddy, S.: Mindfulness meditation and stress in students. *Mindfulness Res. Rev.* **8**(1), 77–86 (2024).