Journal of Intelligent Computing System (JICS), Vol.1, Issue. 2, pp. 08-13, (May 2026)

A Machine Learning Approach to Classifying Stress Levels Using Psychological and Non-Psychological Features

Siddharth Kumar[1], Aaqib Abdullah[2], Maneet Singh[3*]

[1,2,3] Department of Information Technology, Gautam Buddha University Greater Noida, India
[1]siddharthk1500@gmail.com, [2]aaqibabdullah2006@gmail.com, [3]maneet.singh@gbu.ac.in

Abstract. Stress impacts students in a great number of ways, influencing not only how they feel but also their academic performance and the management of daily responsibilities. Measuring it, though, is difficult because most measures rely on self-reporting by students, which often lacks reliability. In this work, we investigate whether machine-learning methods can contribute to better stress level predictions based on data from a short online survey. The online survey consisted of basic demographic information, a few questions related to lifestyle, and all 21 items of the University Stress Scale (USS). It yielded a total of 110 completed responses. Four widely used classification models, namely Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost, were trained on three feature combinations to investigate the individual and combined contributions of psychological and non-psychological variables to the prediction. In the experiments conducted, the greatest contribution towards raising the quality of the prediction was made by the psychological features, while among the different models, the best results were obtained using the XGBoost model. The paper also analyses the comparison of the self-reported stress of students with their stress categories computed from the USS and found significant differences between the two, which points to the suspicion that personal judgment may not reflect the actual stress pattern. The results indicated in this paper show that a combination of structured psychological scales and machine-learning methods may provide a more reliable approach to understanding student stress.

Keywords: *Stress prediction, University Stress Scale, Psychological features, Behavioral features, Machine learning*

1. Introduction

Stress has become a very important issue among university students these days because they tackle academic demands, personal responsibilities and everyday challenges. Challenges when not handled properly lead to deterioration of focus, academic track, emotional and mental health. Research and studies conducted earlier show that stress accumulates over time and lead to more dangerous mental health issues. It leads us to having deeper understanding and measures to tackle them. [1], [2], [3]
However, despite considerable research in this area, a significant limitation was identified, mainly that most methods used to assess this area are based on self-report questionnaires. Although this method is simple and efficient, it is based on self-perception, which does not always correlate with actual psychological conditions [4]. Students may incorrectly perceive their levels of stress and might ignore changes in their emotional conditions. Students or adolescents may not know how they feel because they lack emotional self-awareness, which limits their ability to notice, identify, and label their emotions.[5]
To address this limitation, other methods were developed to provide more accurate and stable results. One such method is the University Stress Scale, which includes 21 questions that address all levels and types of stressors experienced by students [6]. This method was used in various studies and found to be more sensitive to changes in levels and types of stress factors. Other studies were conducted to identify that students do not experience stress from only one source, but rather how they cope with their emotions, how they react in critical situations, and how they are connected to certain personality traits such as resilience and alexithymia [7], [8]. Another factor that was identified is that students experience sudden changes in their levels of stress in response to unexpected events, such as university closures [1].
These results show the nature of stress and its need for both psychological and lifestyle analysis. We used the 110 responses that we collected in the survey with features categorized into three groups: first includes non-psychological i.e. demographic and lifestyle-related features, second was psychological i.e. 21 USS features and third was simply the combination of the first two feature sets. We used four different classification methods to calculate their relative effectiveness: Logistic Regression, Support Vector Machine, Random Forest and XGBoost. Beside developing a predictive model, our focus was also on comparing and understanding the impact of both psychological and non-psychological feature sets. Moreover, the study also aimed at the comparison of psychological test-based stress reports with the self-reported stress. Thus, to summarize, our study has the following objectives:

1. To develop a predictive model with the help of a machine learning algorithm that can predict stress level using psychological or non-psychological survey data.

2. To determine how combining both psychological and non-psychological features improve prediction results.
3. To compare the self-reported levels of stress, USS identified stress levels and categorized them into categories and determined how well the subjective self-assessment does in comparison to objective measurement

.

2. Related Work

The research on student stress can be viewed through different perspectives. It uses established tools and ways to measure, exploring both the contextual and individual influences [9-12]. In this field, University Stress Scale is a very important tool that was developed by Stallman and Hurst to measure multiple different types of stress factors that are common in university environments and their overall intensity [6]. It is very different from a self-rating model where it offers an evaluative framework of how students perceive academic, personal, financial and environmental pressures [6]. The contextual factors have been found to significantly affect stress expectations. For example, an instant sudden change in the learning environment, like campuses closing during the COVID-19 pandemic which led to a significant change in stress level reported by the university students and an increase in study-related stress levels following the 2020 campus closure and shifting to virtual learning was observed. [1].
Beside contextual factors, psychological traits also play a major role in how one reacts to stress. For example, the coping methods and levels of stress are linked to how students experience and deal with stress [7]. Psychological traits like resilience and emotional blindness which relate to the difficulty in identifying and describing emotions, are also found to affect university-related stress among female students [8]. Such stress eventually gives rise to other issues like anxiety and substance abuse [8]. Other studies have also observed that sensitivity to emotions and societal pressures lead to significant stress, anxiety and depression especially among female students [13]. A lot of people have also assessed how emotional health and adjustment patterns contribute to the way young adults deal with stressful situations [4]. Structured evidence tells that stress is a very common issue faced by students around the world and this perception may be caused by academic pressure, personal factors or any other reasons [3].
When combined together these studies show an underlying cause that students get stress from a lot of external and internal psychological conditions. More specifically, the prior work shows that both contextual stressors and internal psychological traits such as coping style , resilience and alexithymia shape distress [2,3,7,8]. These traditional assessment methods give us a very useful point of view but they often overlook subtle patterns that individuals might not recognize fully [1-3]. Moreover, the self-reporting tools utilized in the above studies have been observed to suffer from limitations which included factors like bias and failure to identify underlying conditions [4]. This shows that there is a need to incorporate advance automated methods like machine learning with the standard structured scales to create a more clear and thorough understanding of stress levels Recent studies that combine standardized psychological scale and data driven or machine learning approach have shown both hybrid method to capture more complex stress signature amd improve the predication or early identification of at risk student [14-16].

3. Methodology

3.1 Data Collection and Pre-processing

An online survey was conducted from 1st Nov 2025 to 10th Nov 2025. The dataset used in this study was collected through a self-administered online survey conducted among university students between 1st Nov 2025 and 10th Nov 2025 [17]. The survey consisted of several questions divided into four categories: (i) Generic information of a student like course, year and stream; (ii) Lifestyle-based questions, like number of sleeping hours per night, number of study hours, etc. (iii) 21 questions, one for each USS item; (iv) Self-reported stress-based questions. During data preparation, categorical variables were numerically encoded, and continuous features were standardized.

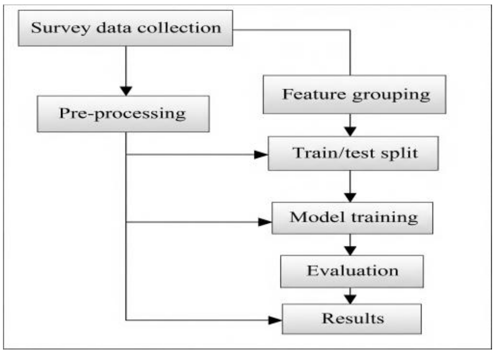

Fig. 1. Proposed architecture of the stress classification framework.

3.2 Model Training and Prediction

Because of the emphasis on the contribution of psychological versus non-psychological factors, features were grouped into three configurations. The first configuration corresponds to non-psychological features such as demographic and lifestyle variables. The second configuration incorporated responses to the 21 questions pertaining to USS-21 items, thereby focusing on psychological features. The third configuration combined the features from both first and second i.e. assessing non-psychological and psychological features together. As the study aims to test the capability of different feature configuration in predicting stress using a machine learning model, we labelled each model as follows:

- Model A (Non-Psychological Features): A machine learning model trained on demographic and lifestyle variables.
- Model B (Psychological Features): A machine learning model trained on a USS-21 questionnaire item.
- Full Model: A machine learning model trained on the combination of both non-psychological and psychological features.

The target variable to be predicted was the self-reported stress, which was measured on a 0–3 scale. During the preparation of the data, the categorical variables were numerically encoded, continuous features were standardized, and missing values imputed. Further, the data was divided in a ratio of 80:20 for training and testing. For each configuration of the features, four different supervised machine-learning models were implemented [18]: Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost. Parameters were filtered to obtain stable and reliable performance. Logistic regression gives us the linear relationship between the features and probabilities, SVM helps in identifying a hyperplane which gives us the most optimal result giving the maximum margin that separates classes. Random forest uses multiple decision trees to improve the generalization by ensemble learning where multiple models are combined to improve performance. XGBoost applied gradient boosting repeatedly to reduce errors and enhance the prediction accuracy. The performance of these models is reported using accuracy, weighted F1-score, and Matthews Correlation Coefficient (MCC) that offer total correct values, class balanced behavior and overall classification quality.

4. Result and Discussion

4.1 Model Performance on the basis of Full Feature Set

The first research objective aims to evaluate the predictive performance of the machine learning algorithms when they are trained on the combined set of psychological and non-psychological features. This approach is made to find out a more systematic and broad representation of student behavior, lifestyle and emotional states which will provide a broad basis to identify the stress levels.

The difference between the performance across the models in Table 1 show how the choice of algorithms majorly affects the stress prediction. Logistic Regression and SVM were struggling to achieve high accuracy, indicating their limited capability in the model overlapping and nonlinear stress patterns. Random Forest achieved a slightly better generalization due to ensemble averaging. However, XGBoost performed better than all the models with an accuracy of 0.695 and an MCC score of 0.557, showing its great ability to capture complex interactions within varying feature space [15]. These results ensure that combining the psychological and non-psychological variables make the stress prediction stronger especially when we use gradient boosting frameworks.

Table 1. Performance of ML Models Using the Full Feature Set

Model	Accuracy	MCC	Weighted F1 Score
Logistic Regression	0.434	0.128	0.381
SVM	0.434	0.185	0.432
Random Forest	0.521	0.294	0.515
XGBoost	0.695	0.557	0.694

4.2. Comparing Psychological vs Non-Psychological Features

The second research objective is focused on understanding how the predictive performance is different when the training model uses only non-psychological features, psychological features and the combined set. This comparison is critical to determine the predictive strength of USS-21 psychological indicators as compared to demographic or lifestyle variables.As seen in Figure 1, the Model B which uses only the USS-21 psychological indicators, gave the highest accuracy of all (0.739) and weighted F1-score too (0.739) but the MCC score was on the lower side (0.355). The higher weighted F1 score shows better precision or recall in the dominant classes while the lower MCC means a decline in the balanced prediction quality across all the classes.

However, when we compare the performance of a machine learning model which was trained on the psychological features with the one trained on the non-psychological features then we can see the psychological features one did better than the latter (Table 2).
To sum up we can say that the performance of a machine learning model is most useful when it is trained on both the psychological and non-psychological features. The non-psychological features based on lifestyle have been found to be related with the level of happiness in studies conducted earlier [19] which show a major negative correlation with recognized stress.

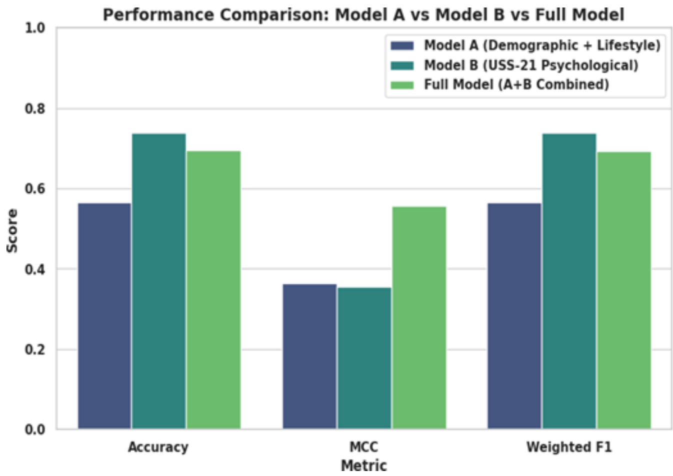

Fig. 2. Comparison of Accuracy, MCC, and F1-Score Across Model A, Model B, and the Full Model Using XGBoost

Table 2. XGBoost Performance Across Three Feature Configurations

Model	Accuracy	MCC	Weighted F1 Score
Model A (Non-Psychological)	0.565	0.363	0.564
Model B (USS-21 only)	0.739	0.355	0.739
Full Model	0.695	0.557	0.694

4.3 Self-Reported vs USS-Derived Stress

The third research objective examines the pattern between self-reported stress levels and USS-derived stress categories. This comparison helps in finding out whether the perception of a student's stress from their own view corresponds to a standard psychological measurement scale.

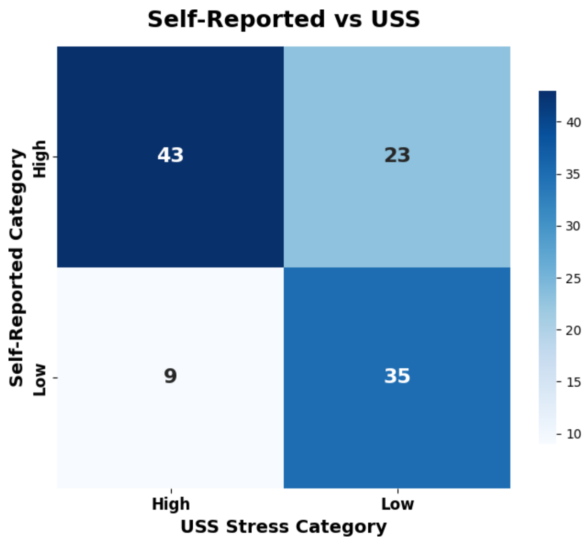

Fig. 3. Heatmap Comparing Self-Reported Stress and USS-Derived Stress Categories.

The heatmap analysis (Figure 3) shows that we have 43 students that were classified as having high stress in both the self-reports and USS scores while 35 students were classified as low stress all the time, showing the meaningful patterns between the two measures. However, some discrepancies came up where 23 students reported having high stress while having low USS scores and 9 students reported low stress levels while receiving high USS scores. These differences show the important psychological refinement as some students overestimate their stress due to emotional sensitivity, temporary situational factors or misinterpretation of their symptoms while the ones with high USS scores and low self-reported stress show denial, less and limited self-awareness or emotional suppression. Such inconsistencies could also happen due to the difference in individual interpretation of severity of stress as compared to the fixed clinical threshold used by standard scales [20]. While the USS uses a uniform cutoff score to categorize the stress levels, people's personal tolerance and perception of stress can also vary which can lead to differences between their subjective report and scale's classification.

5. Conclusion

In this paper, we sought to understand how different kinds of information-basic lifestyle habits, general background details, and the psychological items from the USS-can help in predicting how stressed students actually are. Although we expected that the former would matter, the results showed just how much stronger these features were compared to the non-psychological ones. Models with only demographic or lifestyle data often failed to capture the difference in levels, but once the data with the USS items was brought in, the performance noticeably improved. Among the four algorithms we tried, XGBoost consistently handled the data better than the others, especially when both features were combined.

Another important observation was the inconsistency between what the students said about their stress and what the scores on the USS suggested. Some students ranked themselves high ower than what the structured scale showed. Such discrepancies show that self-reported stress may not be reliable to be used in most cases, as psychological scales tend to capture patterns quite easily.

Although the sample size was very limited it suggests one thing that such a combination of psychological scale and machine learning models can give you a more clear picture of the stress variable as compared to if obtained by using either one alone. If more data and more detailed questions are taken this method can probably support early warning tools or feedback system to help the students understand their own stress levels much better and get help if required.

6. Acknowledgement

The authors express their sincere gratitude to Niraj Kumar, Nimesh Nirmal, and Anupama Sharma for their suggestions in designing the questions for the online survey, along with Tanya Singhania for promoting the survey among students.

7. Declarations

Conflict of Interest: The authors declare that they have no conflict of interest.

Funding: No external funding was received for this study.

Ethical Approval: The survey was conducted anonymously for academic research purposes.

Data Availability: The dataset used in this study is available from the corresponding author on reasonable request.

References

1. L. Von Keyserlingk, K. Yamaguchi-Pedroza, R. Arum, and J. S. Eccles, "Stress of university students before and after campus closure in response to COVID-19," *Journal of Community Psychology*, vol. 50, no. 1, pp. 285–301, 2022.
2. S. Viertiö, O. Kiviruusu, M. Piirtola, J. Kaprio, T. Korhonen, M. Marttunen, and J. Suvisaari, "Factors contributing to psychological distress in the working population, with a special reference to gender difference," *BMC Public Health*, vol. 21, no. 1, p. 611, 2021.
3. M. Gellisch, B. Olk, T. Schäfer, and B. Brand-Saberi, "Unraveling psychological burden: The interplay of socio-economic status, anxiety sensitivity, intolerance of uncertainty, and stress in first-year medical students," *BMC Medical Education*, vol. 24, no. 1, p. 945, 2024.
4. T. Razavi, "Self-report measures: An overview of concerns and limitations of questionnaire use in occupational stress research," 2001.
5. U. E. Rubab, N. Parveen, S. M. Jafari, and M. I. Yousuf, "Social and emotional self-awareness skills among students: A case study," *Qlantic Journal of Social Sciences and Humanities*, vol. 5, no. 1, pp. 336–343, 2024.
6. H. M. Stallman and C. P. Hurst, "The University Stress Scale: Measuring domains and extent of stress in university students," *Australian Psychologist*, vol. 51, no. 2, pp. 128–134, 2016.
7. M. Delara and R. L. Woodgate, "Psychological distress and its correlates among university students: A cross-sectional study," *Journal of Pediatric and Adolescent Gynecology*, vol. 28, no. 4, pp. 240–244, 2015.
8. M. Lyvers, N. Holloway, K. Needham, and F. A. Thorberg, "Resilience, alexithymia, and university stress in relation

to anxiety and problematic alcohol use among female university students," *Australian Journal of Psychology*, vol. 72, no. 1, pp. 59–67, 2020.

9. M. Lu, D. Li, and F. Xu, "Recognition of students' abnormal behaviors in English learning and analysis of psychological stress based on deep learning," *Frontiers in Psychology*, vol. 13, p. 1025304, 2022.
10. M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device: In laboratory and real life," in *Proc. 2016 ACM International Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 1185–1193.
11. S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of Health and Social Behavior*, pp. 385–396, 1983.
12. P. F. Lovibond and S. H. Lovibond, "The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories," *Behaviour Research and Therapy*, vol. 33, no. 3, pp. 335–343, 1995.
13. V. V. Shruti, "The role of gender in anxiety, stress, and depression among college students," *International Journal of Indian Psychology*, vol. 12, no. 2, pp. 4934–4938, 2024.
14. A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Proc. 2013 Humaine Association Conf. on Affective Computing and Intelligent Interaction*, 2013, pp. 671–676.
15. S. Gedam, S. Dutta, and R. Jha, "Analyzing mental stress in Indian students through advanced machine learning and wearable technologies," *Scientific Reports*, vol. 15, no. 1, p. 20610, 2025.
16. T. ShamsEldin, S. Gaber, S. Ansari, R. Elgohary, M. A. Shawky, M. Elbahnasawy, and M. Abdrabou, "Artificial intelligence for predicting depression anxiety and stress using psychometric data," *Scientific Reports*, vol. 15, no. 1, p. 37282, 2025.
17. S. Kumar, A. Abdullah, and M. Singh, "Student Stress Survey Dataset Containing Psychological and Non-Psychological Features," unpublished dataset, Gautam Buddha University, Greater Noida, India, Nov. 2025, available from the corresponding author on reasonable request.
18. R. Kumar, M. Singh, P. Singh, V. Parma, K. Ohla, S. B. Olsson, et al., "Leveraging machine learning and self-administered tests to predict Covid-19: An olfactory and gustatory dysfunction assessment through crowdsourced data in India," in *Proc. 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI)*, 2025, pp. 1–5.
19. A. A. Alhur, M. Alsahmmari, and M. A. Al-Khattab, "Can lifestyle habits predict happiness? An exploratory machine learning study using a visual data mining platform," *Journal of Pioneering Medical Sciences*, vol. 14, pp. 45–52, 2025.
20. A. C. Santee, K. Rnic, K. K. Chang, R. X. Chen, J. A. Hoffmeister, H. Liu, et al., "Risk and protective factors for stress generation: A meta-analytic review," *Clinical Psychology Review*, vol. 103, p. 102299, 2023.