9th International Research Conference of Uva Wellassa University, IRCUWU2025
"Smart Societies: Wellness, Innovation, and Sustainability in a Changing World"
10th & 11th December 2025 @ Uva Wellassa University, Badulla, Sri Lanka.

Paper ID: IRCUWU2025-

MACHINE LEARNING ANALYSIS OF STRESS FACTORS IN COLLEGE STUDENTS: PREDICTIVE MODELING AND FEATURE IMPORTANCE

INTRODUCTION

Stress among college students represents a growing concern, impacting academic performance, mental health, and overall well-being. Contemporary students face multifaceted pressures from academic demands, social dynamics, and personal challenges. Dataset comprising responses from 843 students aged 18–21, collected via Google Forms, captures 20 features across psychological, physiological, environmental, academic, and social domains. Variables include anxiety, self-esteem, sleep quality, and peer pressure, measured on a five-point Likert scale. Study employs machine learning to predict stress levels (Eustress, Distress, No Stress) and identify key contributing factors. Research aims to provide actionable insights for mitigating stress through targeted interventions.

RESEARCH PROBLEM(S)/ISSUE(S)

Rising stress levels among college students correlate with adverse outcomes, including reduced academic performance, mental health disorders, and diminished quality of life. Lack of clarity on dominant stress predictors complicates effective intervention design. Key issues include identifying which factors— psychological, physiological, or environmental—most significantly influence stress and how predictive models can leverage these to inform support strategies. Absence of robust, data-driven approaches to prioritize interventions underscores need for advanced analytical methods to uncover underlying patterns.

OBJECTIVE(S)

- Identify key predictors of stress among college students using feature selection techniques.
- Develop and compare machine learning models (Stacking Classifier, Voting 2 Classifier, Gradient Boosting, MultiLayer Perceptron) for stress level prediction.
- Achieve high predictive accuracy to support reliable identification of at-risk students.
- Provide actionable insights for designing targeted interventions to enhance student well-being

THEORETICAL CONSIDERATIONS AND EMPIRICAL EVIDENCE
Theoretical Considerations

Stress, conceptualized through Lazarus and Folkman's Transactional Model, arises from perceived imbalances between demands and resources. Psychological factors (e.g., anxiety, self-esteem) and physiological factors (e.g., sleep quality, blood pressure) interact with environmental and social stressors, influencing stress outcomes. Machine learning offers a framework to model these complex interactions, with ensemble methods like Stacking and Voting Classifiers enhancing predictive power by combining multiple algorithms.

Empirical Evidence

9th International Research Conference of Uva Wellassa University, IRCUWU2025
"Smart Societies: Wellness, Innovation, and Sustainability in a Changing World"
10th & 11th December 2025 @ Uva Wellassa University, Badulla, Sri Lanka.

Paper ID: IRCUWU2025-

Prior studies indicate strong correlations between stress and factors like sleep quality, self-esteem, and academic pressure. Feature selection techniques, such as permutation importance, have been effective in identifying critical predictors in health-related datasets. Ensemble models and neural networks have shown superior performance in classification tasks involving multifaceted data.

METHODOLOGY

Dataset includes 843 anonymized survey responses from college students aged 18–21, collected via Google Forms. Features span psychological (e.g., anxiety, depression), physiological (e.g., headache, blood pressure), environmental (e.g., noise level), academic (e.g., study load), and social (e.g., peer pressure) domains, rated on a five-point Likert scale. Target variable categorizes stress as Eustress, Distress, or No Stress. Dataset copied to preserve original data. Features reduced to six (self-esteem, mental health history, blood pressure, sleep quality, social support, stress level) based on Sequential Feature Selector (SFS) results. Data split into 70% training and 30% testing sets, with StandardScaler applied to normalize features for MultiLayer Perceptron.

Four models implemented:

- Gradient Boosting Classifier: 300 estimators, random state 44.
- Stacking Classifier: Combines Gradient Boosting with Logistic Regression (max iterations 400, C=0.9) as meta-model, 10-fold cross-validation.
- Voting Classifier: Integrates Logistic Regression, Stacking Classifier, and Gradient Boosting.
- MultiLayer Perceptron: Two hidden layers (16, 32 nodes), ReLU activation, Adam solver, 1000 iterations. SFS and permutation importance identify key features. RepeatedKFold (2 splits, 2 repeats) and 5-fold cross-validation evaluate model robustness.

RESULTS AND DISCUSSION

MultiLayer Perceptron achieves highest test accuracy (91.515%), followed by Stacking Classifier (91.212%) and Voting Classifier (90.909%). Cross-validation scores confirm robustness, with Stacking Classifier at 89.87% and Voting Classifier at 88.91%. Key predictors include blood pressure (0.21), sleep quality (0.21), and self-esteem (0.10), as identified by permutation importance. Findings align with prior research highlighting physiological and psychological factors as primary stress drivers (3; 5). High accuracy of the MultiLayer Perceptron suggests neural networks effectively capture complex interactions in stress data. Blood pressure and sleep quality's prominence indicates physiological health's critical role in stress management. Limitations include potential dataset bias and exclusion of some features post-selection, which may overlook minor contributors. Future research could incorporate longitudinal data for dynamic stress modeling.

CONCLUSION(S)

9th International Research Conference of Uva Wellassa University, IRCUWU2025
"Smart Societies: Wellness, Innovation, and Sustainability in a Changing World"
10th & 11th December 2025 @ Uva Wellassa University, Badulla, Sri Lanka.

Paper ID: IRCUWU2025-

Study successfully identifies blood pressure, sleep quality, and self-esteem as key stress predictors among college students. MultiLayer Perceptron outperforms other models, achieving 91.515% accuracy, with Stacking Classifier close behind. Results underscore need for interventions targeting physiological and psychological well-being, such as sleep improvement programs and self-esteem workshops. Machine learning proves effective in uncovering stress patterns, offering a scalable approach for student support systems. Future work should explore additional features and real-time data to enhance predictive models.

REFERENCES

[1] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[2] Dietterich, T. G. (2000). Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems, 1–15.

[3] Lund, H. G., et al. (2010). Sleep patterns and predictors of disturbed sleep in a large population of college students. Journal of Adolescent Health, 46(2), 124–132.

[4] Pascoe, M. C., et al. (2020). The impact of stress on students in secondary school and higher education. International Journal of Adolescence and Youth, 25(1), 104–112.

[5] Rosenberg, M. (1989). Society and the adolescent self-image. Wesleyan University Press.