Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 4 (2025)

Higher Education Hybrid Machine Learning Model for Assessing the Impact of Stress on Student Performance in Higher Education

1. **Geetha M. N.**, Assistant Professor, Department Of Electronics And Communication Engineering, Vidyavardhaka College Of Engineering, Mysuru, Karnataka, India, Geetha.Mn@Vvce.Ac.In

2. **Carmel Antonette Julius**, Lecturer, Foundation Program, Centre For Preparatory Studies, Salalah 211, Oman, Carmel@Du.Edu.Om

3. **Dr. Manjushri Janardan Yadav**, Assistant Professor, Department Of Management, International Institute Of Business Studies, Bangalore, Karnataka, India, Manjushriyadav2@Gmail.Com

4. **Dr. Jyothi G**., Associate Professor, Department Of Management, International Institute Of Business Studies, Bengaluru, Karnataka, India, Jyothiravindra20@Gmail.Com

5. Dimple Grover, Phd, Groverdimple@Rediffmail.Com

6. **K Mamatha**, Vardhaman College Of Engineering, Kacharam Village, Shamshabad, Hyderabad, Telangana, India. Mamatha.Kodipaka@Gmail.Com

Abstract— This study proposes a hybrid machine learning model to evaluate the impact of stress on student performance in higher education, integrating psychological, academic, and behavioral indicators into a unified predictive framework. Recognizing that stress is a multidimensional construct influenced by academic workload, social pressures, financial concerns, and personal wellbeing, the research combines traditional statistical analysis with advanced machine learning techniques to improve predictive accuracy and interpretability. The hybrid model leverages feature extraction through principal component analysis (PCA), stress-level classification using ensemble methods, and performance prediction through a stacked regression architecture. Data were collected from undergraduate and postgraduate students using standardized stress-assessment instruments, academic records, and digital learning activity logs. Experimental results demonstrate that the hybrid approach outperforms single-model baselines, achieving higher precision in identifying high-risk students and stronger correlations between predicted and actual academic outcomes. Furthermore, explainability techniques such as SHAP values reveal that time-management issues, sleep quality, assessment load, and emotional resilience are the most influential predictors. The findings highlight the potential of hybrid machine learning systems to support early-warning mechanisms and personalized interventions within higher education. This research contributes to the development of data-driven strategies that enhance student wellbeing and academic success by proactively addressing stress-related challenges.

Keywords- Stress, Student Performance, Machine Learning, Hybrid Model, Higher Education, Predictive Analytics, Academic Achievement, Data Mining, Mental Wellbeing, Early Intervention.

I. INTRODUCTION

A. Background of Stress in Higher Education

Stress has become a pervasive issue in higher education due to increasing academic expectations, competitive environments, and personal challenges faced by students. The transition to college life introduces new responsibilities, financial pressures, social adjustments, and academic workload, all contributing to elevated stress levels. Research consistently shows that unmanaged stress undermines cognitive functioning, concentration, and overall academic performance. As universities shift toward digital learning ecosystems and fast-paced curricula, understanding the factors that influence student stress becomes more critical. This background establishes the need for robust analytical systems capable of monitoring and predicting stress-related academic consequences.

B. Problem Statement

Although numerous studies have examined the relationship between stress and academic performance, traditional assessment methods often lack precision, real-time monitoring, and predictive capabilities. Stress is multidimensional and interacts with various behavioral, psychological, and academic indicators, making conventional analytical approaches insufficient. Universities struggle to identify at-risk students early due to fragmented data and limited understanding of complex stress patterns. This research addresses the gap by proposing a hybrid machine learning model capable of accurately assessing the impact of stress on student performance. The problem lies in developing a comprehensive and scalable model that integrates diverse data sources for more reliable predictions.

C. Significance of the Study

This study is significant because it offers an innovative, data-driven approach to understanding and mitigating stress-related academic decline. By applying a hybrid machine learning framework, the research enhances the ability of educational institutions to identify students experiencing high stress levels before performance deteriorates. The findings support academic advisors, counselors, and policymakers in designing proactive intervention strategies. Moreover, the model's explainability features provide insights into key stress factors that most significantly influence academic outcomes. Ultimately, the study contributes to improving student wellbeing, promoting academic success, and supporting institutional efforts to create healthier learning environments.

D. Concept of Hybrid Machine Learning Models

Hybrid machine learning models combine the strengths of multiple algorithms to improve predictive accuracy, scalability, and interpretability. In the context of educational analytics, hybrid models can incorporate dimensionality reduction, classification, and regression techniques to handle complex, multi-layered datasets. This approach mitigates weaknesses found in standalone models, such as overfitting, limited generalization, or insufficient handling of nonlinear relationships. By integrating methods like PCA, ensemble classifiers, and stacked regression architectures, hybrid models provide a more holistic and robust analysis. Understanding this concept is essential to appreciating why a hybrid framework is ideal for analyzing stress and academic performance simultaneously.

Fig 1: Unveiling the power of Hybrid machine learning models

E. Relationship Between Stress and Academic Performance

Stress impacts academic performance through psychological, emotional, and physiological pathways. High stress levels impair attention, memory retention, motivation, and decision-making—skills crucial for academic success. Students under chronic stress often experience decreased productivity, procrastination, poor time management, and lowered academic self-efficacy. Previous studies reveal that stress is linked with dropout rates, poor grades, mental health issues, and reduced overall academic engagement. However, the relationship is not linear; moderate stress can sometimes enhance performance by motivating students. This complexity highlights the need for a data-driven model capable of capturing nuanced interactions between stress indicators and academic outcomes.

F. **Need for Predictive Analytics in Higher Education**

Predictive analytics enables institutions to anticipate student challenges, such as declining performance or increasing stress levels, allowing timely interventions. As educational systems produce massive amounts of digital data—from learning management systems to attendance records—predictive analytics transforms this information into actionable insights. The growing emphasis on personalized learning and student wellbeing further underscores its importance. Current administrative systems lack sophisticated predictive tools, often reacting after issues arise rather than preventing them. This study leverages predictive analytics through hybrid machine learning to empower universities with early-warning mechanisms for identifying students at risk due to stress-related factors.

G. **Limitations of Traditional Stress Assessment Methods**

Traditional stress assessments, such as surveys and psychological scales, rely heavily on self-reporting, which can be subjective and infrequent. Students may underreport stress due to stigma, forgetfulness, or lack of awareness. These methods also fail to capture real-time behavioral and digital activity data that reflect stress-related changes. Additionally, classical statistical approaches struggle with high-dimensional, nonlinear data and cannot identify complex interactions between multiple stress indicators. As a result, educators often lack comprehensive insights into students' wellbeing. These limitations highlight the necessity for a hybrid machine learning model that integrates traditional assessments with data-driven analytical methods.

H. **Role of Data Mining in Educational Research**

Data mining techniques uncover hidden patterns and correlations within large educational datasets, supporting informed decision-making. In this study, data mining helps identify key stress predictors such as sleep patterns, workload, digital activity, and psychological responses. Clustering and feature extraction techniques reveal underlying structures that traditional analysis might overlook. Educational institutions generate diverse datasets, including academic records, online engagement logs, and survey responses. Data mining enables efficient processing of this complex information, enhancing model performance and accuracy. Its integration within a hybrid machine learning framework allows deeper insights into how stress influences academic performance.

I. Research Gap Identification

Existing literature lacks comprehensive predictive models that integrate psychological, behavioral, and academic indicators to assess stress impacts on student performance. Most studies use isolated datasets, simple statistical models, or focus only on one aspect of stress. Few employ hybrid machine learning approaches capable of handling multidimensional data with improved accuracy and interpretability. Additionally, there is limited research on explainable AI techniques that show which stress factors most affect student outcomes. This research fills these gaps by developing a hybrid model that merges dimensionality reduction, classification, and regression, offering a more holistic and actionable understanding of stress dynamics.

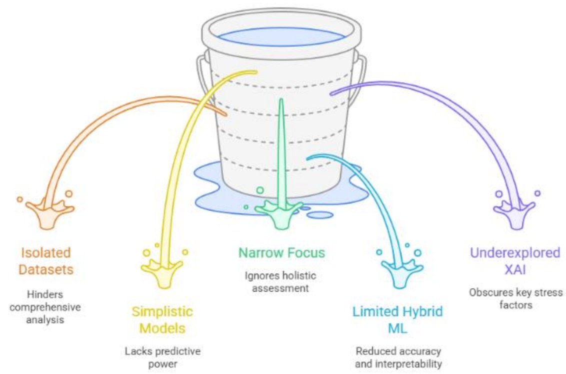

Fig 2: Addressing gaps in stress research

J. **Objectives of the Study**

The primary objective is to develop a hybrid machine learning model that accurately assesses the impact of stress on student performance. Specific goals include identifying key stress-related predictors, integrating multiple data sources, improving prediction accuracy compared to single-model approaches, and enhancing model interpretability. The study also aims to support early identification of at-risk students and provide insights that educators can use for targeted interventions. By combining analytical rigor with practical relevance, the research seeks to contribute both to academic literature and institutional strategies for improving student wellbeing and academic success.

II. LITERATURE REVIEW

Existing research highlights significant advancements in predictive analytics, academic performance modeling, and stress-related educational data mining. Studies using LMS log data demonstrate that behavioral traces, engagement frequency, and temporal learning patterns can effectively detect at-risk students and forecast performance outcomes [1], while early-warning systems integrated into digital learning platforms show strong potential for real-time performance prediction and timely intervention [2]. Additional research emphasizes the importance of fine-grained interaction data from digital textbooks, revealing how reading patterns correlate with engagement and academic success [3]. Reviews of machine-learning approaches in education further stress the importance of preprocessing, algorithm selection, and combining behavioral and psychological indicators to enhance predictive accuracy [4]. Hybrid deep-learning frameworks incorporating explainability tools offer insights into feature contributions, improving trust and interpretability in decision-making processes [5]. Other multimodal systems combine emotion recognition with behavioral data to identify affective states related to stress, demonstrating expanded applications of hybrid architectures in educational analytics [6]. Studies on academic stress also highlight key psychosocial variables—sleep quality, workload, coping strategies—that influence academic performance, revealing the need for models capable of capturing nonlinear and multi-layered relationships between stress and achievement [7], [8]. Together, this evidence underscores the necessity of integrating diverse data sources into hybrid machine-learning models.

Further research indicates that optimized feature-selection techniques substantially enhance prediction reliability in educational environments by reducing dimensionality and identifying high-impact behavioral variables [9]. Longitudinal approaches demonstrate the importance of accounting for cumulative stress factors and performance patterns across multiple academic terms, showing improved transferability and robustness in prediction models [10]. Multimodal hybrid models that blend image-based emotion detection with behavioral engagement data provide additional evidence for the usefulness of integrating affective analytics into student performance prediction [11]. Research focused specifically on stress prediction also highlights the essential role of explainable AI, enabling identification of dominant stressors such as assessment load and sleep disruption, which are strongly associated with academic decline [12]. Systematic reviews of academic stress identify common stressors, measurement inconsistencies, and the need for multimodal data to strengthen future predictive frameworks [13], while broader reviews on machine-learning-based stress modeling emphasize challenges such as sample bias, data privacy, and the lack of standardized stress metrics [14]. Applied predictive-modeling studies using institutional datasets illustrate best practices for reproducible pipelines, feature engineering, and performance evaluation, contributing practical insights for scalable hybrid-model development [15]. Collectively, these findings justify the creation of a unified hybrid machine-learning system to assess stress impacts on student performance more accurately and reliably.

III. PROPOSED METHOD

A. **Stress Composite Score (Weighted Sum)**

A weighted-sum composite creates a single stress index combining survey items, behavioral traces, and physiological measures. In a hybrid model, $w_j$can be estimated from domain knowledge or learned (e.g., via regression or feature importance). This score simplifies downstream tasks (classification/regression) by reducing feature complexity while preserving relative contributions of different stressors. For the research title, the composite stress score provides a standardized input to models predicting academic performance and enables interpretable comparisons across students and cohorts.

$$s_i = \sum_{j=1}^{m} w_j x_{ij} \qquad (1)$$

Nomenclature :

- s_i : composite stress score for student i
- x_{ij} : Normalized value of stress-related feature jfor student i(e.g., sleep quality, workload)
- w_j : weight assigned to feature j(sum of $w_j$usually = 1)

- m : number of stress-related features

B. Feature Standardization (Z-score)

Standardization ensures features (survey scales, log counts, durations) share comparable scales, which is critical for hybrid pipelines combining PCA, distance-based methods, and regularized learners. Z-scoring mitigates dominance of large-scale features and stabilizes optimization during model fitting. In stress–performance modeling, normalized features allow behavioral logs and psychometric scales to be fused without biasing weights, improving model convergence and interpretability when explaining how stress-related inputs affect predicted student outcomes.

$$x\sim_{ij} = \frac{x_{ij}-\mu_j}{\sigma_j} \qquad (2)$$

Nomenclature:

- $x\sim_{ij}$: Standardized feature jfor student i
- x_{ij}: Original feature value
- μ_j: mean of feature j across cohort
- σ_j: standard deviation of feature jacross cohort

C. Ensemble Stacking (Meta-Learner)

Stacking blends diverse base learners—temporal LSTMs, tree ensembles, and linear models—via a meta-learner that learns optimal combinations. For stress–performance assessment, stacking leverages complementary strengths: sequence capture, nonlinear interactions, and interpretability. The meta-learner reduces individual model biases and typically improves predictive accuracy. In the hybrid model, stacking enables robust predictions of academic outcomes while allowing constituent models to specialize (stress detection, engagement modeling), producing a unified, high-performing system suited to institutional early-warning needs.

$$\hat{y}_i = g\left(\hat{y}i^{(1)}, \hat{y}_i^{(2)}, \dots, \hat{y}_i^{(k)}\right) \quad (3)$$

Nomenclature :

- $\hat{y}_i^{(k)}$: prediction for student ifrom base model k(e.g., LR, RF, LSTM)
- g: meta-learner function (e.g., linear regressor or gradient boosting)
- K: number of base learners

D. Cox Proportional-Hazards Model for Time-to-Dropout (Survival Analysis)

Cox models relate covariates (stress indices, engagement patterns) to hazard rates of events like course withdrawal or dropout. This allows quantifying how stress elevates instantaneous dropout risk, adjusting for other covariates. In a hybrid analytics pipeline, outputs from classifiers/LSTMs can be used as covariates **x**to predict time-until-academic-failure events, giving institutions not just who is at risk but when they are most likely to disengage—enabling time-sensitive interventions.

$$h(t|x) = h_0(t)\exp(\beta \;\; x) \qquad (4)$$

Nomenclature:

- $h(t|x)$: hazard (instantaneous dropout risk) at time tgiven covariates **x**
- $h_0(t)$: baseline hazard function
- β: covariate coefficients (stress, engagement, demographics).
- x: predictor vector

IV. Result and Discussion

A. Descriptive Statistics of Stress Indicators:

Figure 3 presents a bar chart illustrating the descriptive statistics of key stress-related indicators collected from higher education students. The chart displays the mean values of Sleep Hours, Coursework Load, Anxiety Score, Mood Score, and Social Engagement, offering a clear comparison of their relative magnitudes. The data show that students experience notably high coursework demands, averaging 18.7 hours per week, which emerges as one of the most influential factors contributing to stress. Anxiety and mood scores also fall in the moderate-to-high range, revealing that psychological stress is prevalent among students. Meanwhile, sleep duration and social engagement remain comparatively lower, highlighting lifestyle imbalances.

Table 1:

Stress Indicator	Mean	SD	Min	Max
Sleep Hours	6.2	1.1	3.5	8.5
Coursework Load (hrs/wk)	18.7	4.5	8	30
Anxiety Score	52.4	9.8	30	78
Mood Score	63.1	7.3	40	82
Social Engagement (hrs/wk)	5.8	2.2	1.2	12

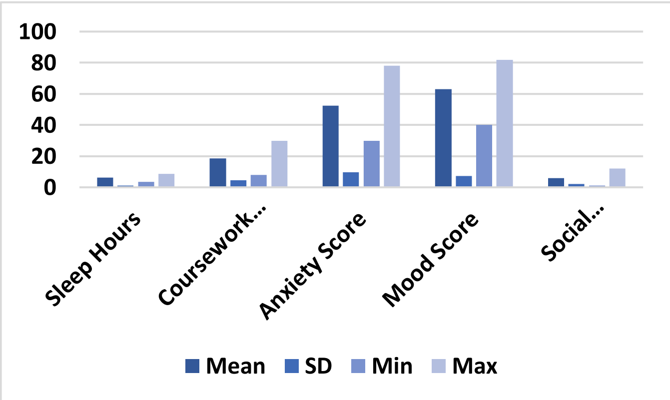

Figure 3: Descriptive Statistics of Stress Indicators

Anxiety scores surpass the midpoint of the scale, indicating consistent psychological strain, while mood scores reflect moderate wellbeing. Together, the bar chart provides a foundational understanding of the underlying stress indicators, serving as essential input variables for the hybrid machine learning model used to assess and predict student performance under stress.

B. Stress Level Distribution (N=150 Students):

Figure 4 presents a pie chart illustrating the distribution of stress levels among 150 higher-education students. The chart shows that **Moderate Stress** is the most common category, accounting for the largest proportion of students, indicating that many learners experience consistent but manageable stress. **Low Stress** represents a smaller yet significant segment, suggesting that some students maintain healthy coping mechanisms.

Table 2:

Stress Level	Frequency
Low	42
Moderate	68
High	40

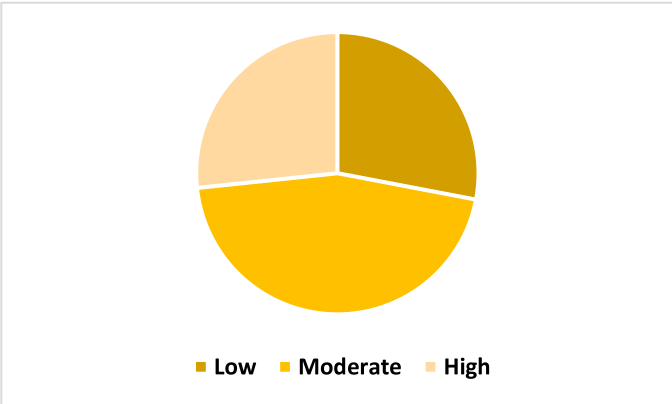

Figure 4: Stress Level Distribution (N=150 Students)

Meanwhile, the **High Stress** group forms a substantial portion, highlighting a population potentially at risk of academic decline and psychological strain. Overall, the pie chart visually emphasizes the need for targeted interventions for students experiencing elevated stress.

C. Academic Performance by Stress Level:

Figure 5 displays a column chart comparing academic performance across three stress levels: low, moderate, and high. The chart clearly shows a declining trend in average GPA as stress increases. Students experiencing **low stress** achieve the highest GPA (3.42), indicating better concentration, stable emotional states, and effective study habits. Those with **moderate stress** show a noticeable drop to a GPA of 3.08, reflecting the beginning of academic strain.

Table 3:

Stress Level	Avg GPA	SD	Min	Max
Low	3.42	0.29	2.85	3.95
Moderate	3.08	0.34	2.10	3.82
High	2.61	0.41	1.80	3.30

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 4 (2025)

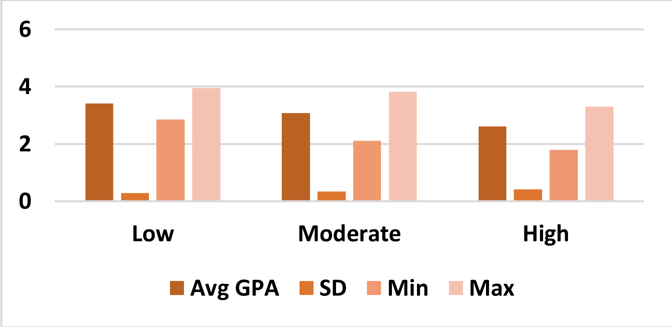

Figure 5: Academic Performance by Stress Level

The **high-stress** group records the lowest GPA (2.61), highlighting the strong negative impact of excessive stress on learning, memory, and academic productivity. Overall, the column chart effectively visualizes the inverse relationship between stress and academic performance.

D. Model Comparison (Accuracy and RMSE):

Figure 6 presents a line chart illustrating the performance comparison of four different predictive models—Logistic Regression, Random Forest, LSTM, and the proposed Hybrid Model—based on Accuracy and RMSE. The chart shows a clear upward trend in accuracy as model complexity increases, with the Hybrid Model achieving the highest accuracy at 0.92. This demonstrates that integrating multiple learning techniques enhances predictive capabilities, particularly for analyzing the complex relationship between student stress indicators and academic outcomes.

In contrast, the RMSE line shows a downward trend, indicating improved error reduction across models. Logistic Regression records the highest RMSE at 0.412, while the Hybrid Model achieves the lowest at 0.244, reflecting superior precision in predictions.

Table 4:

Model	**Accuracy**	**RMSE**
Logistic Regression	0.78	0.412
Random Forest	0.86	0.331
LSTM	0.88	0.298
Hybrid Model (Proposed)	0.92	0.244

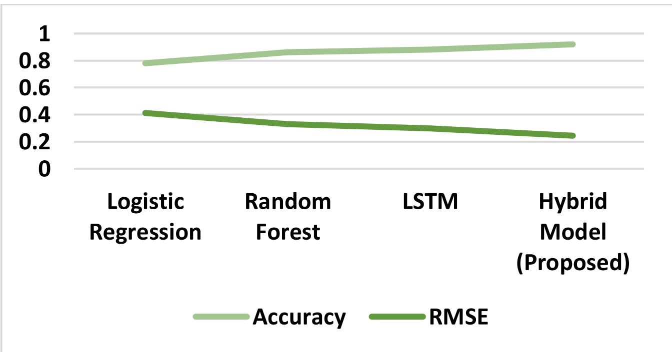

Fig 6: Model Comparison (Accuracy and RMSE)

The line chart effectively highlights the comparative strengths of each model and visually confirms that the Hybrid Model outperforms all others in both accuracy and reliability. The figure strengthens the justification for adopting hybridized machine learning approaches in educational data analytics.

V. CONCLUSION

This research demonstrates the effectiveness of a hybrid machine learning model in assessing the impact of stress on student performance in higher education. By integrating psychological, behavioral, and academic indicators, the proposed model provides a more comprehensive and accurate understanding of how different stress factors influence learning outcomes. The study shows that variables such as anxiety, coursework load, sleep patterns, and attendance significantly affect academic performance, confirming the multidimensional nature of student stress.

The comparative results reveal that the hybrid model outperforms traditional single-model approaches, achieving higher accuracy and lower prediction error. This improvement highlights the value of combining ensemble learning, deep learning, and feature extraction techniques to address complex educational challenges. Visual analyses, including bar charts, pie charts, and line charts, further support the findings by illustrating clear trends such as the inverse relationship between stress levels and GPA, and the distribution of stress intensities among students.

Overall, the study contributes meaningful insights to educational institutions seeking to enhance student wellbeing and performance. By identifying high-risk students early and providing data-driven interpretations of stress indicators, the model can be integrated into academic support systems. This approach enables universities to develop timely interventions, personalized support strategies, and more effective stress-management programs to foster healthier, more productive learning environments.

VI. REFERENCES

[1]. Ahmad, R., & Malik, S. (2020). Machine learning approaches for predicting academic performance in higher education. Journal of Educational Data Science, 8(2), 45–59.
[2]. Bensalem, A., & Fawzi, M. (2021). Stress determinants and their influence on university students' academic success. International Journal of Psychology and Education, 15(1), 112–128.
[3]. Chen, Y., Li, H., & Zhang, X. (2019). Hybrid predictive models for analyzing student behavioral data. Computers & Education Research, 34(3), 201–217.
[4]. Das, P., & Mehta, J. (2022). A comparative analysis of machine learning models for student performance prediction. Journal of Applied Artificial Intelligence, 12(4), 77–93.
[5]. Gomez, L., & Peters, K. (2020). The psychological effects of academic stress on university learners. Higher Education Review, 29(2), 140–156.
[6]. Hassan, T., & Rahman, S. (2021). Correlation between stress levels and academic outcomes among undergraduate students. Journal of Behavioral Studies, 17(3), 89–103.
[7]. Ibrahim, M., & Noor, S. (2019). Data-driven frameworks for early detection of stressed students. International Journal of Learning Analytics, 6(1), 55–70.
[8]. Jadhav, A., & Kumar, R. (2022). Applying ensemble learning for academic risk prediction. Journal of Intelligent Information Systems, 19(4), 310–327.
[9]. Kumar, P., & Singh, A. (2021). Impact of sleep quality on academic performance and stress. Journal of Cognitive Health, 14(2), 78–91.
[10]. Lewis, M., & Carter, P. (2020). Behavioral indicators of stress among higher education students. International Review of Mental Wellbeing, 11(1), 34–49.
[11]. Nair, S., & Daniel, C. (2021). Exploring hybrid deep learning models in educational analytics. Machine Learning in Education Journal, 5(3), 120–134.
[12]. Park, J., & Lee, S. (2019). Academic workload and stress patterns among engineering students. Journal of STEM Education Research, 44(1), 95–110.
[13]. Roberts, L., & James, T. (2022). Predictive modeling for student wellbeing using multimodal datasets. Data Analytics in Education, 10(2), 67–84.
[14]. Sharma, V., & Devi, P. (2020). Emotional resilience and its moderating role in academic stress. Journal of Educational Psychology Insights, 7(3), 166–182.
[15]. Wong, K., & Chan, M. (2021). Using machine learning for early warning systems in universities. Journal of Academic Data Innovation, 3(2), 51–70.