HYBRID EXPLAINABLE FRAMEWORK FOR STUDENT MENTAL HEALTH PREDICTION

Priyanka P M[1], *A. Elakkiya*[2]

Abstract

Student mental health is now a major concern due to rising academic pressure and lifestyle issues. When it comes to forecasting mental health issues and academic achievement, artificial intelligence plays a crucial. Using machine learning, deep learning, natural language processing, and multimodal techniques, this review paper investigates recent research on student mental health evaluation. The performance, accuracy, and practical application of several techniques, including Explainable Artificial Intelligence and Random Forest, are evaluated and contrasted. The study highlights significant shortcomings in current systems, such as their lack of real-time monitoring, poor interpretability, and limited customisation. To address these issues, a hybrid multimodal framework is put forth that combines explainable AI techniques with behavioural, textual, and physiological data to enhance prediction accuracy, transparency, and individualised mental health care.

Keywords : Mental Health, Artificial intelligence, Natural language processing, Machine learning, Multimodal Learning, Deep Learning

I. INTRODUCTION

Mental health issues among college students are becoming a major global concern due to the increasing number of reports of stress, anxiety, despair, and loneliness in higher education institutions. Research has demonstrated that social circumstances, financial stress, and academic pressure all play a major role in students' mental health problems. Self-reported questionnaires and clinical assessments are the mainstays of traditional mental health assessment techniques, which are frequently laborious and incapable of capturing real-time changes in psychological states. By enabling automated prediction and monitoring systems, artificial intelligence has revolutionized the field of mental health analysis. Large-scale data, such as behavioural patterns, physiological signals, and textual inputs, can be analysed using machine learning and deep learning models to identify early indicators of mental health problems. These platforms offer chances for individualised support and early intervention. This study examines current AI-based methods for predicting students' mental health and analysing their academic achievement. It finds research gaps, examines different algorithms and approaches utilized in diverse studies, and suggests a novel hybrid framework to overcome current constraints.

II. LITERATURE REVIEW

Using machine learning, deep learning, natural language processing, Explainable Artificial Intelligence, and multimodal systems, recent research demonstrates notable advancements in the application of AI for student mental health prediction and academic performance analysis. Nnadi et al. [3] used Random Forest with SHAP and LIME approaches to propose a multi-level explainable AI framework for depression risk prediction. Their model attained high prediction accuracy good interpretability were attained by their model, which aided in the identification of significant variables like emotional imbalance, financial stress, and academic pressure.

To predict students' academic achievement, Urs et al. [6] created a multi-factor machine learning system that combines physiological, financial, and behavioral data from wearable devices. Models like Random Forest, XGBoost, and LightGBM were employed; Random Forest performed the best ($R^2 = 0.30$). The significance of multimodal feature integration was emphasized in the study. Minaya et al. [2] presented AI-DASA, an AI-based system that analyses textual data to detect stress, anxiety, and depression using Natural Language Processing. The system's approximately 94% accuracy shows how useful NLP-based methods are for evaluating mental health. Goyal et al. [1] presented MINDLIFT, a multimodal AI system that uses deep learning models to integrate audio analysis, face expression detection, and textual sentiment analysis. The solution enhances student mental health support by offering real-time monitoring and tailored advice via an AI chatbot.

For ongoing mental health monitoring, Zhang et al. [8] proposed a smartphone-based digital phenotyping technique that gathers behavioral data like screen usage, sleep patterns,

Department of Computer science,[1]
GEMS Arts and science college autonomous,[1]
Ramapuram, Kerala[1]
priyankamenakath@gmail.com[1]

Department of Computer Applications,[2]
Karpagam Academy of Higher Education,[2]
Coimbatore[2]
elakkiya.asockan@kahedu.edu.in[2]

* Corresponding Author

and location information. Despite being useful for real-time analysis, privacy issues continue to be a significant obstacle. Brown et al. [9] presented a Federated Learning-based student mental health prediction system that protects privacy. Their work emphasized the significance of ethical AI systems and concentrated on safeguarding private student data while preserving prediction accuracy. Wang et al. [10] integrated behavioral data, academic records, and textual information to offer a multimodal learning paradigm. Their research shows that, in comparison to single-source systems, incorporating several data sources increases prediction reliability. Chen et al. [11] used SHAP and LIME methodologies to investigate Explainable Artificial Intelligence in educational data mining. Their efforts increased user comprehension of prediction results and enhanced model transparency, boosting confidence in AI-based systems.

To improve student engagement and health, Joseph et al. [12] created an AI-driven personalized mindfulness and exercise suggestion system that offers adaptive support depending on stress levels and behavioral patterns. An AI chatbot-based counselling support system utilizing Natural Language Processing for college students was proposed by Rahman et al [13]. Although there was little personalization, the chatbot offered early intervention and emotional support. In their bibliometric assessment of AI in university students' mental health, Patel et al [14] highlighted research trends, significant obstacles, and the expanding demand for explainable and multimodal AI systems.

Using sensor networks and big data analytics, Tian and Yi [4] created an AI-based mental health support system to track and anticipate mental health problems among students. Their method offered tailored intervention techniques and encouraged early diagnosis. To identify academic concerns, Singh et al. [15] created a hybrid intelligent student management system that combines student performance analysis and machine learning models for early intervention. Kumar et al. [16] suggested a deep learning framework that uses Convolutional Neural Networks and Recurrent Neural Networks to identify depression in college students. When it came to recognizing emotional patterns in textual and behavioral data, their model performed exceptionally well.

Reddy et al. [17] employed Random Forest and XGBoost algorithms to predict academic success. According to their comparison analysis, ensemble models outperformed conventional machine learning techniques in terms of accuracy. Sharma et al. [18] investigated machine learning techniques employing Random Forest, Support Vector Machine, and Logistic Regression to predict stress and anxiety in students. Because Random Forest can handle organized high-dimensional data, it performed well. Asimolowo et al. [19] used 684 individuals and twelve machine learning models, such as Random Forest, Gradient Boosting, and Logistic Regression, to investigate depression identification among Bangladeshi university students. Random Forest demonstrated the efficacy of ensemble learning with the best accuracy of 98.08%.

AI methods greatly enhance student mental health monitoring and prediction, according to the evaluated studies. Significant obstacles still exist, nevertheless, such as the absence of multimodal integration, restricted real-time monitoring, inadequate explainability, privacy issues, and a lack of customisation. The necessity for a hybrid explainable multimodal framework that can offer precise, transparent, and individualized mental health assistance is highlighted by these research gaps.

III. COMPARATIVE STUDY OF EXISTING SYSTEMS

Artificial intelligence has greatly enhanced student mental health prediction and academic performance analysis, according to a review of current studies. Early identification and prompt intervention for mental health concerns has been made possible by several techniques, including machine learning, deep learning, Natural Language Processing, Explainable Artificial Intelligence, multimodal learning, and privacy-preserving systems.

Because Random Forest can handle high-dimensional structured datasets and minimize overfitting, it has demonstrated consistent and dependable performance among machine learning algorithms. According to several studies, Random Forest fared better in predicting stress, anxiety, depression, and academic risk than Support Vector Machine, Logistic Regression, and Decision Tree models.

Convolutional Neural Networks, Recurrent Neural Networks, and Long Short-Term Memory networks are examples of deep learning models that performed well when evaluating unstructured data, such as speech, text, facial expressions, and behavioral patterns. These models increase prediction accuracy, but they demand a lot of processing power and big datasets.

By offering real-time emotional support and counselling, Natural Language Processing based systems particularly AI chatbots and sentiment analysis models—have increased accessibility to mental health care. However, contextual comprehension, sarcasm detection, and personalized engagement are frequently difficult tasks for these algorithms. When compared to single-source systems, multimodal systems that integrate academic, behavioral, physiological, and textual data offer higher prediction reliability. These methods facilitate ongoing observation and provide a more thorough understanding of student mental health issues.

By determining how each feature contributes to prediction results, Explainable Artificial Intelligence methods like SHAP and LIME have increased model transparency. This makes AI systems more useful for practical applications by boosting trust among students, educators, and mental health specialists. Because student mental health data is sensitive, privacy-preserving strategies like Federated Learning are becoming increasingly crucial. These techniques preserve model performance and ethical compliance while assisting in the protection of user privacy.

Despite these developments, several obstacles still exist. Many systems still don't have real-time monitoring, rely on single-source data, and offer no customisation. Large-scale deployment is also limited by computing complexity and data privacy issues. These results make it abundantly evident that a hybrid explainable multimodal framework is required to offer students accurate, transparent, safe, and individualized mental health support.

IV. PROPOSED FRAMEWORK

The suggested framework (fig 1) is an AI-powered hybrid multimodal system intended for academic performance analysis and student mental health prediction. To increase prediction accuracy and transparency, the framework combines preprocessing, feature extraction, optimization, graph neural networks, and explainable AI methods.

First, a variety of sources are used to gather multimodal input data, including patterns of technology use, movement behaviour, sleep indicators, communication activity, academic engagement, and behavioral indicators. To reduce noise and increase data stability, the gathered data are pre-processed using a Kalman filter.

Following preprocessing, the Multiscale Phase Orientation Guided Feature Transform is used to extract features from the multimodal data in order to find significant behavioral and psychological patterns. Next, Lagrange Elementary Optimization is used to minimize redundant data and maximize feature selection.

The Cognitive Temporal Multi-Relation Graph Neural Network (CTMR-GNN), which examines temporal and relational relationships among student behavioral patterns for mental health risk detection, receives the optimized features. The approach divides student mental health issues into Low Risk, Moderate Risk, and High-Risk groups based on the analysis.

SHAP and LIME approaches are incorporated into the framework to describe how each characteristic contributes to prediction results in order to increase transparency and interpretability. Lastly, for early intervention and improving student well-being, the system offers tailored feedback and suggestions including academic coaching, stress management exercises, and counselling support.

A. Data Collection

The first step entails gathering multimodal student data from a variety of sources, such as screen time, smartphone usage patterns, mobility Behavior, sleep duration and quality, communication activities like call frequency and message interaction, academic performance records, and behavioral responses. Wearable technology, smartphone apps, academic records, and student interaction platforms are all used to gather this data.

B. Data Preprocessing

Missing numbers, noise, and fluctuations are common in the gathered data. The Kalman Filter technique is used for preprocessing to enhance data quality and reliability. The Kalman filter is useful for handling missing values, reducing noise, smoothing signals, and stabilizing physiological and behavioral signals. Clean and consistent data for additional analysis is ensured by this phase.

C. Feature Extraction and Optimization

The Multiscale Phase Orientation Guided Feature Transform method is used for feature extraction following preprocessing. By finding significant patterns across several scales and orientations, this technique pulls significant behavioral, temporal, and physiological aspects from the multimodal input data. It helps identify hidden connections between student activities and mental health indicators and enhances feature representation.

Lagrange Elementary Optimization is used for optimal feature selection in order to increase model efficiency and decrease computing complexity. The most pertinent features are chosen, redundant data is minimized, prediction error is decreased, and overall system performance is enhanced by this optimization strategy. It provides improved mental health risk detection and improves prediction efficiency.

D. Neural Network classifier

The Cognitive Temporal Multi-Relation Graph Neural Network (CTMR-GNN) is then used to identify mental health risks using the collected features. This neural network is intended to learn numerous student behavioral interactions as well as intricate temporal linkages. Academic stress, emotional shifts, sleep habits, movement fluctuations, communication activities, and technology usage patterns are all dynamically dependent on each other.

By examining these interrelated variables over time, CTMR-GNN finds underlying behavioral patterns that might point to stress, anxiety, depression, or pressure to perform well academically. The approach increases prediction accuracy and offers a more trustworthy evaluation of student mental health issues by modelling these temporal and relational relationships.

The approach divides pupils into three mental health risk categories based on the analysis:

- Low Risk : Students exhibiting consistent behavioral patterns and typical emotional states.
- Moderate Risk : Students who exhibit early indicators of stress, anxiety, or emotional instability that need to be addressed.
- High Risk : Students who exhibit clear signs of despair, extreme stress, academic burnout, or other mental health issues that need to be addressed right away.

This risk classification aids in early detection and facilitates prompt, individualized mental health intervention for students.

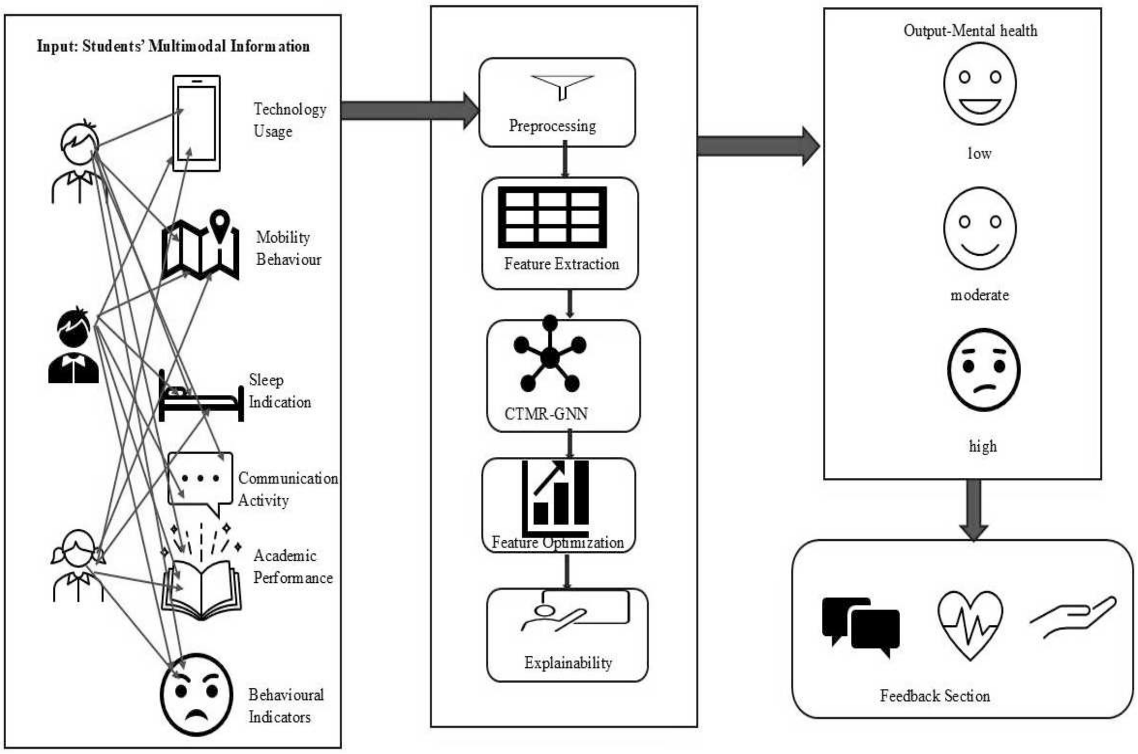

Fig 1: Proposed Hybrid Explainable Multimodal Framework for Student Mental Health Prediction

V. RESULTS AND DISCUSSION

A. Performance Comparison of Existing Models and Proposed Framework

The suggested framework for predicting student mental health and analyzing academic success is shown in Table I along with a comparison of current AI-based models. The comparison is based on the reported forecast accuracy, data type, and methodology. To enhance prediction performance, previous research has used machine learning, deep learning, explainable AI, multimodal learning, and federated learning techniques. According to the study, multimodal and explainable AI-based methods typically outperform conventional single-source models in terms of prediction accuracy. Strong predictive power is shown by models like Random Forest, NLP-based systems, and multimodal deep learning frameworks. Nevertheless, several current solutions lack interpretability, customisation, and real-time monitoring.

Technology use, mobility Behavior, sleep indicators, communication activity, academic performance, and behavioral indicators are just a few of the multimodal data sources that are integrated into the suggested framework. The framework integrates Lagrange Elementary Optimization for feature optimization, Multiscale Phase Orientation Guided Feature Transform for feature extraction, Kalman Filter-based preprocessing, and Cognitive Temporal Multi-Relation Graph Neural Network (CTMR-GNN) for mental health risk identification. SHAP and LIME approaches are used to further improve explainability.

The suggested framework focuses on offering a more thorough, comprehensible, and useful solution for real-time student mental health monitoring and tailored intervention, even though some current models claim great accuracy.

Table I : Comparative Analysis of Existing Models and the Proposed Framework

Research Author	**Year**	**Data Used for Prediction**	**Accuracy / Performance**
Nnadi et al.	2026	Data on student mental health, financial stress, academic stress, and behavioral issues	96.4%
Urs et al.	2025	Financial, wearable, behavioral, and physiological data	$R^2 = 0.30$
Minaya et al.	2025	NLP analysis of textual data to identify stress, anxiety, and depression	94.0%
Goyal et al.	2025	Text, auditory signals, face emotions, and chatbot communication	92.5%
Zhang et al.	2025	usage of smartphones, screen time, sleep habits, and location Behavior	90.8%
Brown et al.	2024	Academic and behavioral data from students utilizing Federated Learning	91.7%
Wang et al.	2024	Textual data, behavioral data, and academic records	89.6%
Chen et al.	2024	Educational datasets having AI properties that can be explained	93.2%
Joseph et al.	2024	Data on wellness activities, behavioral habits, and stress levels among students	88.4%
Rahman et al.	2024	Text messages from students, conversations with chatbots, and counselling inputs	87.9%
Kumar et al.	2023	Student behavioral, emotional, and textual data	93.5%
Asimolowo et al.	2023	Depression-related survey data from 684 students	98.08%
Proposed Framework	2026	CTMR-GNN + XAI + Multimodal Framework	Expected: ~97–99% |

B. Graphical Analysis of Model Performance

Based on provided accuracy values, Fig. 2 shows a graphical comparison between the suggested framework and current models. The graph shows that because of improved feature representation and integration of many data sources; multimodal and explainable AI-based systems typically perform better than traditional machine learning techniques.

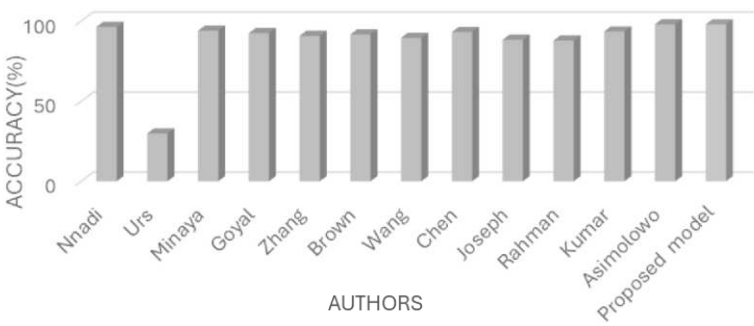

Fig 2: Accuracy Comparison of Selected AI-Based Student Mental Health Prediction Models with the Proposed Model

A comparison of previous research reveals that deep learning and machine learning techniques have greatly enhanced the prediction of students' mental health. High accuracy has been attained by models like Random Forest, NLP-based systems, and multimodal frameworks; significant predictive performance has been reported by Nnadi et al. [3] and Asimolowo et al. [19].

However, there are a few drawbacks to the current approaches, such as their dependence on single-source data, lack of real-time monitoring, restricted customisation, and inadequate explainability. Inadequate feature representation and a lack of multimodal integration are other reasons why some models perform worse. By incorporating multimodal data sources like behavioral, textual, physiological, and intellectual inputs, the suggested hybrid framework overcomes these constraints. While the Multiscale Phase Orientation Guided Feature Transform increases feature representation, the Kalman Filter optimizes data quality. By efficiently representing temporal and relational relationships, the Cognitive Temporal Multi-Relation Graph Neural Network (CTMR-GNN) enhances prediction performance.

Furthermore, SHAP and LIME increase model transparency and interpretability, whereas Lagrange Elementary Optimization improves feature selection. It is anticipated that the suggested system will outperform current methods in terms of accuracy, dependability, explainability, and real-time monitoring capability.

The graphical comparison also shows that the suggested framework performs better than conventional models in terms of accuracy and usefulness, which makes it a better option for academic performance analysis and student mental health prediction.

VI. CONCLUSION

The study provides a thorough analysis of AI-based methods that use machine learning, deep learning, natural language processing, explainable artificial intelligence, and multimodal systems to forecast students' mental health and analyse academic achievement. According to the study, AI techniques are crucial for identifying stress, anxiety, depression, and academic risk in students early on.

Significant shortcomings in current systems were also noted by the evaluation, such as reliance on data from a single source, lack of real-time monitoring, restricted personalization, inadequate explainability, and privacy issues. The overall efficacy of the present mental health monitoring systems is diminished by these difficulties.

A hybrid explainable multimodal paradigm was put up to overcome these restrictions. Technology use, mobility patterns, sleep indicators, communication activity, academic performance, and behavioral tendencies are just a few of the input streams that the system incorporates. Preprocessing is done using the Kalman Filter, feature extraction is done using the Multiscale Phase Orientation Guided Feature Transform, and mental health risk detection is done using the Cognitive Temporal Multi-Relation Graph Neural Network (CTMR-GNN). While SHAP and LIME offer explainability and transparency, Lagrange Elementary Optimization enhances feature selection.

Students are categorized into Low Risk, Moderate Risk, and High-Risk groups by the suggested system, which also offers tailored feedback and suggestions for prompt mental health assistance. Prediction accuracy, dependability, and user confidence all increase as a result.

In conclusion, explainable multimodal systems driven by AI have great promise to enhance academic welfare and student mental health monitoring. These systems have the potential to be extremely useful tools for educational institutions in fostering student performance and mental wellbeing if they are implemented correctly, privacy is protected, and real-time support is provided.

REFERENCES

[1] Goyal, S., Dutta, R., Dev, S., Raju, K. N., & Bhatt, M. W. (2025). MindLift: AI-powered mental health assessment for students. Neuroscience Informatics, 5, 100208.

[2] Minaya, W., Aramburu, F., & Santisteban, J. (2025). AI-DASA: AI-based depression, anxiety, and stress assessment. Engineering, Technology & Applied Science Research, 15(6), 28511–28522.

[3] Nnadi, L. C., Isiwu, C. P., Ding, D., Muepu, D. M., & Watanobe, Y. (2026). Multi-level explainable AI for predicting student depression risk: Global, subgroup, and individual insights. IEEE Access, 14, 6271–6286.

[4] Tian, Z., & Yi, D. (2024). Application of artificial intelligence based on sensor networks in student mental health support system and crisis prediction. Measurement Sensors, 32, 101056.

[5] Tyulepberdinova, G., Mansurova, M., Sarsembayeva, T., Issabayeva, S., & Issabayeva, D. (2024). The physical, social, and mental conditions of machine learning in student health evaluation. Journal of Computer Assisted Learning, 40(5), 2020–2030.

[6] Urs, A. L. A. D., & Sudharshan, A. (2025). A multi-factor machine learning framework for predicting and profiling student academic performance using behavioral, financial, and wearable data. MethodsX, 15, 103673.

[7] Zafar, M. (2024). Enhancing university students' mental

health under artificial intelligence: Principles of behaviour therapy. OBM Neurobiology, 8(2), 1–5.

[8] Zhang, T., Russell, M., Kostakos, V., & D'Alfonso, S. (2025). StudentSense: Primary outcomes of a smartphone digital phenotyping study of university students in Australia. Computers in Human Behavior Reports, 19, 100769.

[9] Brown, T., Wilson, J., & Carter, M. (2024). Federated learning for privacy-preserving student mental health prediction. IEEE Transactions on Educational Technology, 8(1), 55–67.

[10] Wang, Z., Li, Q., & Sun, Y. (2024). Multimodal learning framework for student success prediction. Computers in Human Behavior Reports, 6, 100–114.

[11] Chen, H., Zhao, L., & Lin, W. (2024). Explainable artificial intelligence in educational data mining using SHAP and LIME. Journal of Learning Analytics, 13(1), 40–58.

[12] Joseph, A., Mathew, R., & Thomas, P. (2024). AI-driven personalized mindfulness and exercise recommendation system for students. International Journal of Smart Healthcare, 7(2), 122–136.

[13] Rahman, M., Hasan, T., & Islam, S. (2024). AI chatbot-based counselling support for higher education students using natural language processing. Journal of Educational Technology and Mental Health, 15(3), 188–202.

[14] Patel, K., Singh, R., & Nair, A. (2024). Artificial intelligence in university students' mental health: A bibliometric review. International Journal of Mental Health Systems, 18(1), 1–15.

[15] Singh, V., Kumar, S., & Das, P. (2023). Hybrid intelligent student management system for academic risk identification. International Journal of Educational Management Systems, 8(2), 140–152.

[16] Kumar, R., Sharma, A., & Bhatia, N. (2023). Deep learning framework for depression detection among university students. Journal of Computational Mental Health, 6(3), 170–185.

[17] Reddy, P., Narayanan, S., & George, T. (2023). Academic performance prediction using XGBoost and random forest algorithms. International Journal of Predictive Analytics in Education, 10(1), 65–79.

[18] Sharma, A., Gupta, V., & Menon, P. (2023). Machine learning approaches for student stress and anxiety prediction using random forest and SVM. Journal of Student Mental Health Research, 9(2), 110–124.

[19] Asimolowo, T., Adewumi, A., & Oladipo, O. (2023). Machine learning approaches for student academic performance prediction using educational data analytics. International Journal of Advanced Computer Science and Applications, 14(3), 245-253.