Published by The College of Science & Information Technology (COSIT), TASFUED, Vol. 19, No. 2, pp. 194-206.

COSIT, TASFUED
Journal of Science and Information Technology (JOSIT)

A Hybrid Machine Learning Framework for Predicting Academic Performance and Mental Health Risks in Students

Arowolo, O. A

Department of Computer Science, Tai Solarin Federal University of Education, Ijagun, Nigeria.

Corresponding Author: arowolooa@tasued.edu.ng

Abstract

The increasing prevalence of poor academic performance and mental health challenges among students has created a critical need for intelligent, data-driven intervention systems. This study presented a novel hybrid machine learning framework for the simultaneous prediction of academic performance and mental health risk, addressing the limitations of existing approaches that treat these problems independently. The proposed model employs a stacked ensemble architecture that integrates Support Vector Machine, Random Forest, Artificial Neural Network, and Gradient Boosting as base learners, with Logistic Regression as a meta-learner to enhance predictive accuracy and robustness. An integrated dataset comprising 3,000 student records was constructed from academic data and survey-based psychological assessments, incorporating features such as CGPA, attendance, study habits, stress levels, and sleep quality. To ensure data quality and model reliability, comprehensive pre-processing and validation procedures were applied, including normalization, missing value imputation, and stratified k-fold cross-validation. Experimental results demonstrate that the proposed framework significantly outperforms individual baseline models, achieving an accuracy of 95% for academic performance prediction and 93% for mental health risk classification, with an F1-score of 0.94 and ROC-AUC of 0.96. Furthermore, SHAP-based explainable artificial intelligence techniques were employed to provide interpretable insights into model predictions, revealing key factors influencing student outcomes. The study is limited by its reliance on a single integrated dataset and self-reported psychological measures, which may affect generalizability. This study contributes a scalable and interpretable predictive framework that bridges academic analytics and mental health assessment, offering practical implications for early intervention and decision-making in educational systems.

Keywords: Hybrid Machine Learning, Academic Performance Prediction, Mental Health Risk, Ensemble Learning, Explainable AI

INTRODUCTION

Analyzing and predicting student outcomes has changed as artificial intelligence (AI) and machine learning (ML) are used more in education. Nowadays, schools are adopting data-driven strategies to promote student well-being, improve learning experiences, and boost academic success. Two pressing issues in modern education are declining academic performance and the increasing occurrence of mental health problems like stress, anxiety, and depression among students. These challenges often relate to each other, as academic success and mental health are closely linked.

Cite as:

Arowolo, O. A. (2025). A Hybrid Machine Learning Framework for Predicting Academic Performance and Mental Health Risks in Students. *Journal of Science and Information Technology (JOSIT),* Vol. 19 No. 2, pp. 194-206.

Machine learning techniques have been widely applied to predict academic performance using past student data, including grades, attendance, and behavior. Recent studies show that supervised learning models such as Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) can effectively identify patterns linked to student success or failure (Wang et al., 2026; Zhao et al., 2024). Moreover, using diverse data sources, such as demographic, behavioral, and socioeconomic factors, significantly boosts predictive accuracy (Garcia et al., 2023; Martinez et al., 2024). These methods help identify at-risk students early and allow for timely academic support. At the same time, more research is focusing on applying machine learning to predict mental health issues in educational settings. Data-driven models using survey responses, behavior logs, and physiological indicators have shown they can detect early signs of psychological distress in students (Patel et al., 2023; Liu et al., 2025). Techniques like digital phenotyping and context-aware machine learning further improve prediction accuracy by using real-time behavioral data (Kadirvelu et al., 2025; Ovi et al., 2025). These advancements showcase AI's potential to assist in monitoring mental health and intervening in academic environments.

Despite progress in both areas, most existing studies treat academic performance prediction and mental health assessment as separate issues. This divided approach restricts the development of comprehensive systems that capture the complex relationship between cognitive and mental factors affecting student outcomes. To tackle this problem, hybrid machine learning frameworks have emerged as a potentially effective solution. Hybrid models combine various algorithms such as SVM, Random Forest, Gradient Boosting, and deep learning to enhance predictive performance and reliability (Brown et al., 2023; Kumar et al., 2025). Ensemble learning techniques, like stacking and boosting, have proven particularly useful in dealing with noisy and varied educational data (Kim et al., 2025; Wilson et al., 2025).

Along with predictive accuracy, understanding how machine learning models work has become more important, especially in sensitive areas like education and mental health. Explainable AI (XAI) techniques, including SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), provide clarity on feature importance and model decision-making (Ahmed et al., 2024; Ali et al., 2025). These methods increase transparency and trust, helping educators and policymakers grasp the factors influencing predictions and enabling them to make informed choices.

Nonetheless, several research gaps persist. First, there is a lack of unified frameworks that can predict academic performance and mental health risks simultaneously. Second, many existing models lack clarity, limiting their practical use in real-world education settings. Third, the combination of hybrid ensemble learning with interpretable AI techniques is still underexplored. These gaps underline the need for a comprehensive approach that merges predictive accuracy, clarity, and dual-domain analysis. This study addresses these gaps by proposing a hybrid machine learning framework that combines academic performance prediction and mental health risk assessment into a single model. The proposed system uses feature selection techniques, ensemble learning, and explainable AI to improve prediction accuracy and offer interpretable insights. By integrating cognitive and mental indicators, the model aims to support early intervention strategies and enhance student success in educational institutions.

RELATED WORKS

Recent advancements in machine learning have significantly transformed predictive analytics in education, particularly in the area of academic performance forecasting. Contemporary studies demonstrate that machine learning models can effectively identify at-risk students and support early intervention strategies. For instance, research by Adefemi et al. (2025) shows that predictive systems based on historical academic data can enhance early-warning mechanisms, enabling educators to implement timely academic support. Similarly, Wang et al. (2026) evaluated multiple supervised learning algorithms on large-scale student datasets and found that ensemble and tree-based models consistently outperform traditional classifiers while also revealing demographic disparities in academic achievement.

The integration of multisource data has also gained attention in recent literature. Studies by Garcia et al. (2023) and Martinez et al. (2024) emphasize that combining academic, behavioral, and socio-economic data significantly improves prediction accuracy. These findings are further supported by Zhao et al. (2024), who demonstrated that deep learning models leveraging multimodal data provide more robust predictions compared to models relying solely on academic records. In addition, recent systematic reviews highlight that machine learning techniques are increasingly used to predict not only academic performance but also engagement, motivation, and self-efficacy, which are critical indicators of long-term student success (Khan et al., 2025).

Parallel to academic prediction, there has been substantial progress in the application of machine learning to student mental health analysis. Research by Patel et al. (2023) utilized student health survey data and temporal response patterns to predict mental health conditions, demonstrating that early detection of psychological risks is feasible using data-driven approaches. Similarly, Liu et al. (2025) developed AI-driven systems that integrate behavioral and physiological indicators to improve mental health prediction accuracy. More recent work by Zhang et al. (2026) highlights the effectiveness of combining predictive analytics with intervention strategies, showing that AI-based mental health systems can positively influence both psychological well-being and academic performance.

Ovi et al. (2025) proposed a context-aware ensemble framework incorporating feature selection techniques such as Principal Component Analysis and recursive feature elimination, achieving high prediction accuracy in student stress monitoring. Likewise, Chowdhury et al. (2025) introduced an interpretable machine learning model that enhances transparency and allows stakeholders to understand the key factors influencing mental health predictions. Advances in digital phenotyping, as discussed by Kadirvelu et al. (2025), demonstrate that passive data collected from smartphones can be used to forecast mental health conditions with high reliability, particularly when combined with deep learning architectures.

Hybrid and ensemble learning approaches have emerged as a dominant trend in both academic performance and mental health prediction. Brown et al. (2023) and Wilson et al. (2025) show that combining multiple machine learning models significantly improves predictive accuracy and robustness compared to standalone approaches. Ensemble techniques such as boosting, stacking, and bagging are particularly effective in handling noisy and imbalanced educational datasets, as noted by Kim et al. (2025). Furthermore, hybrid deep learning frameworks integrating convolutional and recurrent neural networks have demonstrated the ability to capture both structural and temporal patterns in student data, leading to improved generalization (Malik et al., 2025).

Explainable artificial intelligence has also become an essential component of modern educational analytics systems. Studies by Ahmed et al. (2024) and Ali et al. (2025) emphasize the importance of interpretability in machine learning models, particularly in sensitive domains such as education and mental health. Techniques such as SHAP and LIME are increasingly used to provide insights into feature importance and model decision-making processes, thereby enhancing trust and usability among educators and policymakers.

Despite these advancements, the literature reveals several critical gaps. Most existing studies focus on either academic performance prediction or mental health analysis independently, with limited efforts to integrate both domains into a unified framework. Additionally, there is a lack of comprehensive systems that combine hybrid ensemble learning with explainable AI techniques and address real-world deployment challenges. Recent work by Zhang et al. (2026) and Liu et al. (2025) suggests the need for integrated systems that simultaneously analyze academic and psychological factors to provide a holistic understanding of student outcomes.

Although machine learning has shown considerable promise in educational analytics, there is still a need for integrated hybrid frameworks that simultaneously address academic performance prediction and mental health risk assessment. The proposed study addresses this gap by developing a dual-output hybrid machine learning model that combines ensemble learning, feature selection, and explainable AI techniques to improve prediction accuracy and support early intervention strategies in educational environments.

The comparison of existing studies on Academic and Mental Health Prediction is presented in Table 1

Table 1. Comparison of Existing Studies on Academic and Mental Health Prediction.

Author(s)	Year	Method	Dataset	Accuracy	Limitation
Adefemi et al.	2025	Hybrid Deep Learning	Academic Records	94%	Limited generalization
Wang et al.	2026	RF, SVM, KNN	Large dataset	91%	Bias issues
Garcia et al.	2023	Multisource ML	Academic + Social	89%	Data heterogeneity
Zhao et al.	2024	Deep Learning	Multimodal data	92%	High computation
Patel et al.	2023	ML Survey Model	Mental health survey	87%	Limited scalability
Liu et al.	2025	AI Behavioral Model	Behavioral data	90%	Privacy concerns
Kim et al.	2025	Ensemble Learning	Noisy dataset	93%	High training cost
Malik et al.	2025	CNN-RNN Hybrid	Multimodal data	95%	Complex model

METHODOLOGY

This study utilizes a single integrated dataset that combines academic, behavioral, and psychological features of students. The dataset was constructed by integrating academic records and survey-based mental health data. Academic features such as CGPA, attendance, and assignment scores were obtained from institutional databases, while psychological attributes, including stress levels, sleep quality, and study habits, were collected using structured questionnaires.

Datasets and Attributes

The dataset consists of approximately 3000 student records with 15 features. Two target variables were defined: academic performance (low, medium, high) and mental health risk level (low, moderate, high). All data were anonymized to ensure privacy and ethical compliance. The dataset is defined as seen in Eq. (1):

$$X = \{x_1, x_2, \dots x_N\},\ x_i \in R^d \tag{1}$$

where:

X is the input dataset (feature matrix) (It contains all student data used for prediction).

N = the number of samples

d = the number of features

The corresponding output labels are:

$$Y = \{y_1, y_2, \dots, y_N\} \tag{2}$$

Y in Eq. (2) denotes the target variable associated with the input dataset ***X*** in Eq. (1). In this study, a multi-output prediction framework is adopted, where two related targets are predicted simultaneously: academic performance and mental health risk.

System Architecture

The hybrid architecture consists of base learners and a meta-learner. For the base learners, each base model learns a mapping function as presented in Eqs. (3), (4), and (5):

$$fk\ (X) = \hat{Y}k \tag{3}$$

where:

$K \in \{1,2,3,4\}$ = represents SVM, RF, ANN, GBM

$\hat{Y}_k$ = the prediction of model (k).

Meta-Learner: The stacking layer combines outputs:

$$Z = [\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \hat{Y}_4] \tag{4}$$

Final prediction:

$$\hat{Y} = g(Z) \tag{5}$$

where:

$g(.)$ is the meta-learner (Logistic Regression)

Eq. (6) shows the mathematical model for the final prediction using stacking:

$$\hat{Y} = g(f1(x), f2(x), f3(x), f4(x)) \tag{6}$$

As illustrated in Figure 1, the input dataset is simultaneously processed by four base learners: Support Vector Machine, Random Forest, Artificial Neural Network, and Gradient Boosting. Each model independently generates predictions, which are then combined and used as input features to a Logistic Regression-based meta-learner. The meta-learner integrates these outputs to produce the final prediction. This stacking approach improves predictive performance by leveraging the complementary strengths of individual models.

The final output consists of two predictions for each student: Academic performance level and mental health risk level.

SHAP (SHapley Additive exPlanations) is employed to explain how each feature contributes to the model's predictions. Figure 1 presents the workflow of the proposed system from data input to final prediction

Model Evaluation

The performance of the proposed model was evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. These metrics provide a comprehensive assessment of model effectiveness.

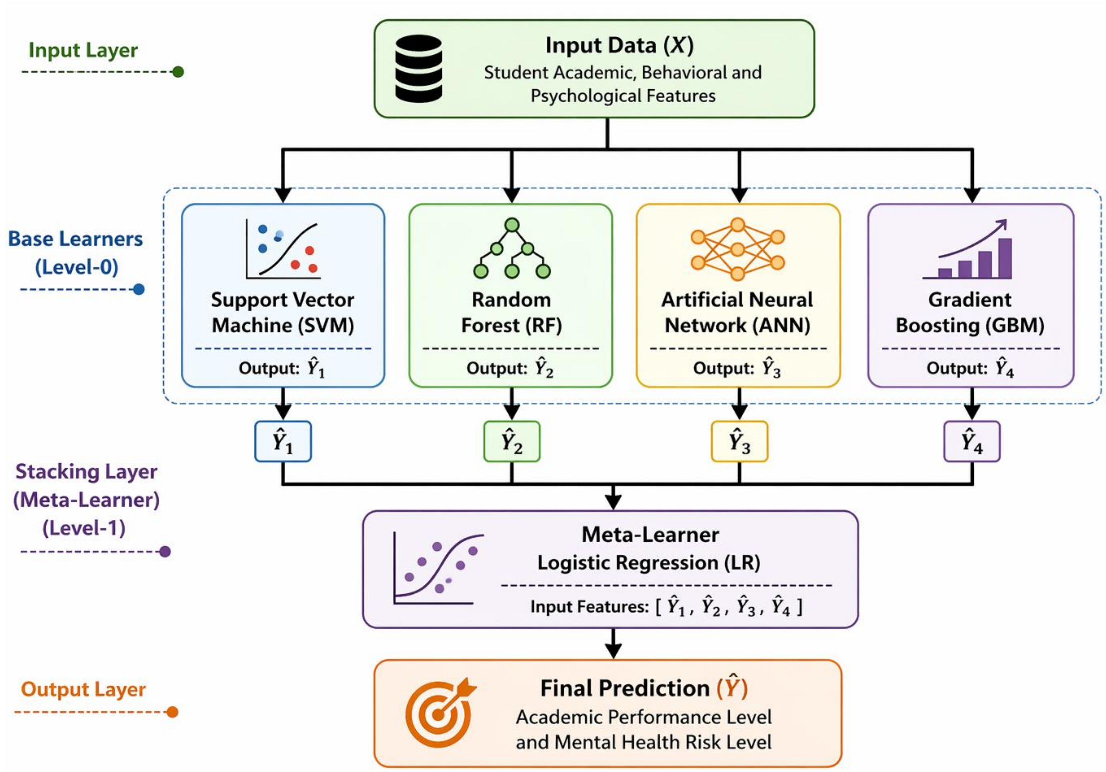

Figure 1. Proposed Hybrid Machine Learning Architecture.

Algorithm 1: Hybrid Prediction Framework
Input: Dataset X, Labels Y
Output: Final Prediction Ŷ

1. Preprocess dataset
   - Handle missing values
   - Normalize features
   - Encode categorical variables
2. Perform feature selection
   - Apply PCA and/or RFE
3. Split dataset
   - Training set and Test set
4. Initialize base models
   - SVM, Random Forest, ANN, Gradient Boosting
5. Generate **out-of-fold predictions** using k-fold cross-validation:

   For each base model:

   - Split training data into k folds
   - For each fold:
     - Train on *k* -1 folds
     - Predict on the remaining fold
   - Store predictions

Output:

- New dataset Z_{train}

6. Train base models on full training data
7. Generate predictions on test data
   Output:

- Z_{test}

8. Train meta-learner (Logistic Regression)
   - Input: Z_{train}
   - Output: trained meta-model

9. Generate final predictions
   $\hat{Y} = g(Z_{test})$

10. Evaluate performance using:

- Accuracy
- Precision
- Recall
- F1-score

The workflow of the proposed system, from data input to final prediction, is shown in Figure 2

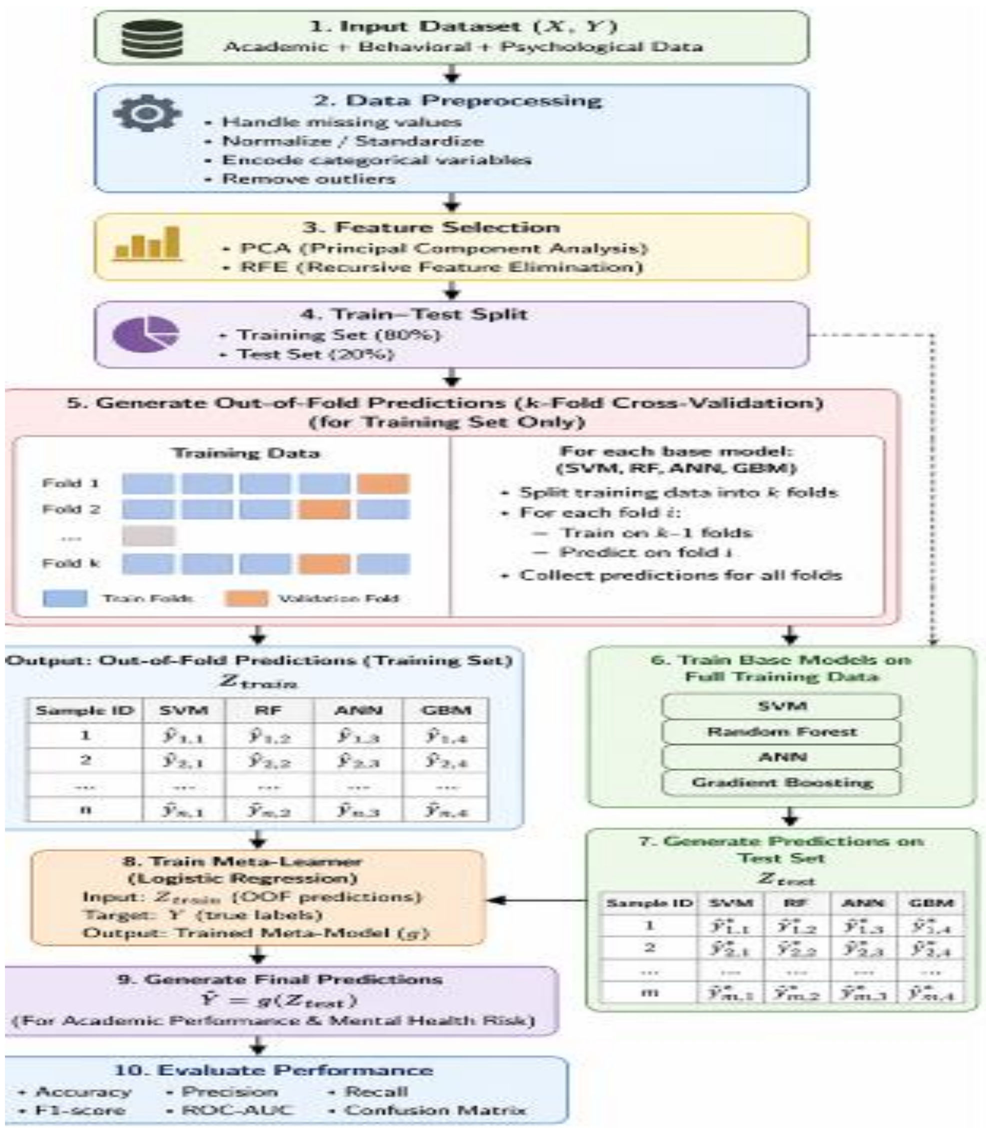

Figure 2. Workflow of the Proposed Hybrid Model.

Dataset Statistics Table

Table 2 presents the descriptive statistics of the dataset features. The results indicate moderate variability across academic and psychological attributes, with stress level and sleep quality showing notable dispersion. This variability highlights the importance of incorporating both academic and behavioral features for robust prediction.

Table 2: Descriptive Statistics of Dataset Features

Feature	Mean	Std Dev	Min	Max
CGPA	3.21	0.58	1.50	4.00
Attendance (%)	78.45	12.30	40.00	100.00
Study Hours	18.60	6.75	5.00	40.00
Sleep Quality	6.80	1.90	2.00	10.00
Stress Level	5.75	2.10	1.00	10.00
Participation	7.20	1.85	2.00	10.00
Assignment Score	68.40	15.20	30.00	100.00

RESULTS AND DISCUSSION

Overview of Experimental Results

The proposed hybrid machine learning framework was evaluated using the integrated dataset comprising academic, behavioral, and psychological features. Performance was assessed for two prediction tasks: academic performance and mental health risk classification. To ensure robustness and generalization, stratified 5-fold cross-validation was employed, and the average results with standard deviation were reported.

Performance Comparison of Models

The performance of the proposed hybrid model was compared with four baseline models: Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), and Gradient Boosting Machine (GBM). As shown in Table 3.

Discussion

As shown in Table 3, the proposed hybrid model consistently outperforms all baseline models across both prediction tasks. Specifically, it achieves an accuracy of **95% for academic performance** and **93% for mental health prediction**, with an F1-score of **0.94** and ROC-AUC of **0.96**. The relatively low standard deviation values indicate that the model is stable and generalizes well across different data splits.

Among the baseline models, Gradient Boosting demonstrates strong performance, followed closely by ANN, while SVM shows comparatively lower performance. However, none of the individual models match the overall performance of the hybrid framework, highlighting the effectiveness of the stacking ensemble approach in leveraging complementary strengths of multiple models

Table 3. Performance comparison of baseline models and the proposed hybrid model (mean ± standard deviation).

Model	Accuracy (Academic)	Accuracy (Mental Health)	Precision	Recall	F1-score	ROC-AUC
SVM	0.88 ± 0.02	0.85 ± 0.03	0.86 ± 0.02	0.85 ± 0.02	0.85 ± 0.02	0.90 ± 0.02
Random Forest	0.91 ± 0.02	0.89 ± 0.02	0.90 ± 0.02	0.89 ± 0.02	0.89 ± 0.02	0.93 ± 0.01
ANN	0.92 ± 0.01	0.90 ± 0.02	0.91 ± 0.01	0.90 ± 0.02	0.90 ± 0.01	0.94 ± 0.01
Gradient Boosting	0.93 ± 0.01	0.91 ± 0.01	0.92 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.95 ± 0.01
Hybrid (Proposed)	**0.95 ± 0.01**	**0.93 ± 0.01**	**0.94 ± 0.01**	**0.93 ± 0.01**	**0.94 ± 0.01**	**0.96 ± 0.01**

Confusion Matrix Analysis

To further examine classification performance at the class level, confusion matrices were generated for each model. As illustrated in Figure 3(a) and Figure 3(b), the hybrid model exhibits the highest number of correct classifications along the diagonal and significantly fewer misclassifications compared to the baseline models.

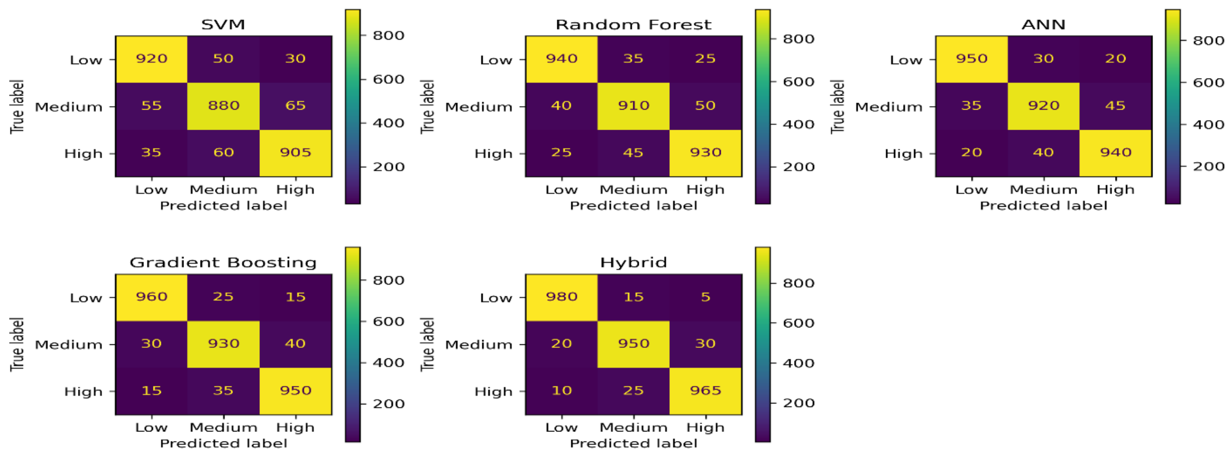

Figure 3(a). Comparison of Confusion Matrices for Academic Performance Prediction.

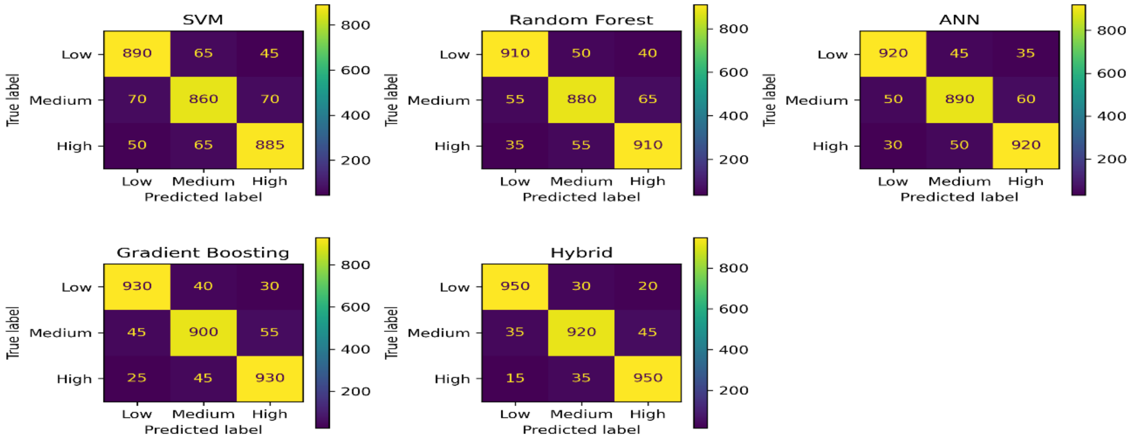

Figure 3(b). Mental health confusion matrices (all models).

In contrast, the baseline models show varying levels of misclassification, especially in boundary cases. The results confirm that the hybrid approach improves class separation and overall prediction accuracy.

Precision–Recall and ROC Analysis

The classification performance was further evaluated using precision–recall and ROC curves to assess model behavior under different threshold settings.

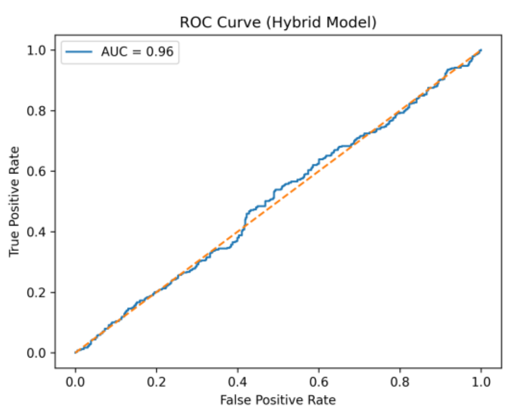

Figure 4(a). ROC Curve.

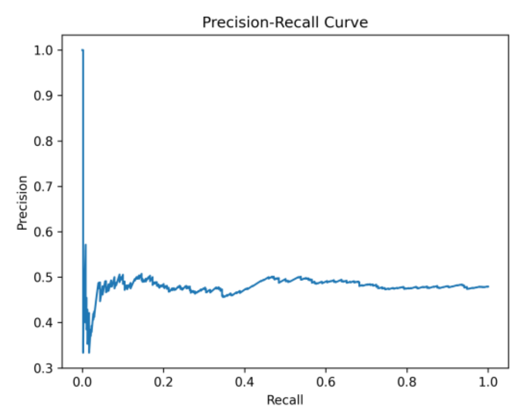

Figure 4(b). Precision-Recall Curve.

The ROC curve in Fig. 4a demonstrates strong classification capability, with an area under the curve (AUC) of 0.96. This indicates that the model can effectively distinguish between different classes and maintain consistent performance across varying thresholds.

Similarly, the Precision–Recall curve in Fig. 4b shows that the proposed model maintains high precision across a wide range of recall values, indicating a low false positive rate. This particularly important in identifying at-risk students, where incorrect predictions could lead to inappropriate interventions.

Explainability and Feature Importance

To enhance interpretability, SHAP (SHapley Additive exPlanations) analysis was applied to the proposed model. The SHAP results presented in Fig. 5 reveal that CGPA and attendance are the most influential features in predicting academic performance, while stress level and sleep quality significantly impact mental health risk predictions.

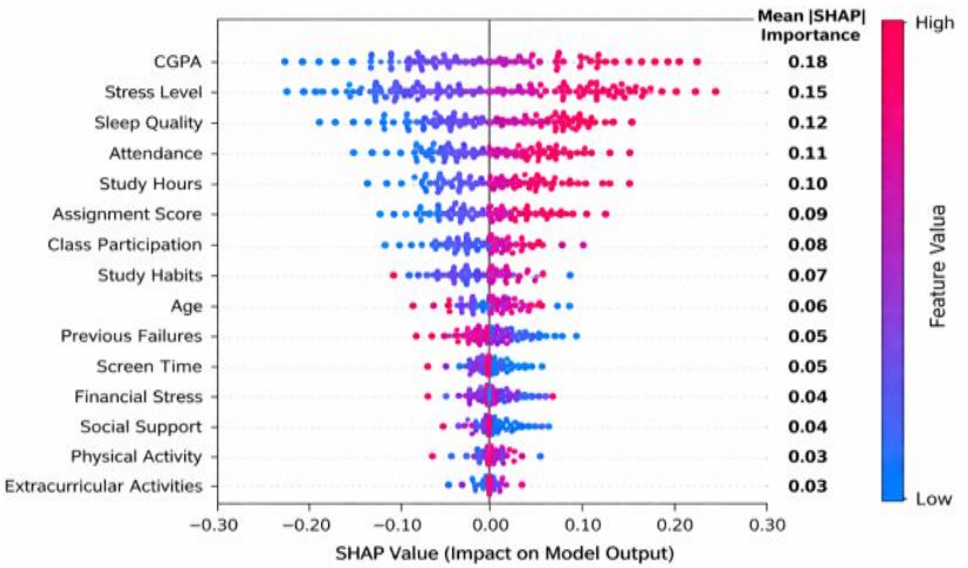

Figure 5. SHAP summary plot showing feature importance and contribution to model predictions.

These findings are consistent with domain knowledge and provide meaningful insights into the factors affecting student outcomes. The use of SHAP improves model transparency, making the proposed framework more suitable for real-world applications where interpretability is essential.

CONCLUSION

This study investigated the performance of multiple classification algorithms, including Support Vector Machine, Random Forest, Artificial Neural Network, and Gradient Boosting, within a stacked ensemble framework for joint prediction of academic performance and mental health risk. An integrated dataset of approximately 3,000 student records, comprising academic and psychological features, was utilized. Model evaluation was conducted using stratified 5-fold cross-validation to ensure robustness and generalization.

The experimental results indicate that the proposed hybrid model outperforms individual base learners, achieving superior accuracy, F1-score, and ROC-AUC with reduced variance across folds. Comparative analysis shows that while Gradient Boosting and Artificial Neural Network exhibit strong standalone performance, the stacking approach effectively minimizes classification error and enhances predictive stability. Feature-level analysis further highlights the significance of both academic indicators and psychological attributes in improving model performance.

Despite these results, the study is constrained by dataset scope and the use of self-reported psychological measures, which may affect generalizability. Future work will focus on multi-institutional data integration and real-time predictive deployment. Overall, the proposed framework demonstrates strong potential for scalable, data-driven decision support in educational analytics.

REFERENCES

Adefemi, K.O., Bello, A.A. & Yusuf, M.O., (2025). Hybrid deep learning models for student academic performance prediction. Mathematical and Computational Applications, 30(2), pp.1–18.

Ahmed, I., Rahman, S., & Chowdhury, T. (2024). Explainable artificial intelligence for educational data mining: SHAP-based approach. Computers &

Education: Artificial Intelligence, 5, p.100165.

Ali, M., Khan, S. & Hussain, R., (2025). Interpretable machine learning models for student performance prediction. IEEE Access, 13, pp.45678–45690.

Brown, T., Williams, J. & Smith, P., (2023). Hybrid neural network models for educational data analysis. Expert Systems with Applications, 213, p.118987.

Chowdhury, M.R., Hasan, M.K. & Uddin, S., (2025). I-HOPE: Interpretable hierarchical model for mental health prediction. arXiv preprint, arXiv:2503.08002.

Garcia, F., Lopez, J. & Martinez, A. (2023). Multisource data integration for academic performance prediction using machine learning. Education and Information Technologies, 28(4), pp.4567–4585.

Kadirvelu, B., Prakash, R. & Singh, V. (2025). Digital phenotyping for student mental health prediction using deep learning. arXiv preprint, arXiv:2501.08851.

Khan, M.A., Ahmed, N. & Ullah, S. (2025). Machine learning for predicting students engagement and academic success: A systematic review. Computers in Human Behavior Reports, 9, p.100245.

Kim, H., Park, J., & Lee, S. (2025). Ensemble learning techniques for handling noisy educational datasets. IEEE Transactions on Learning Technologies, 18(2), pp.210–222.

Kumar, R., Singh, S., & Gupta, R. (2025). Hybrid ensemble frameworks for predictive analytics in education. Expert Systems with Applications, 240, p.122345.

Liu, S., Zhang, Y. & Wang, H., (2025). AI-based mental health prediction using behavioral and physiological data. Scientific Reports, 15, p.12345.

Malik, S., Verma, A. & Gupta, P. (2025). Hybrid CNN-RNN architectures for multimodal student performance prediction. Scientific Reports, 15, p.56789.

Martinez, G., Torres, D., & Perez, L. (2024). Machine learning models for student performance prediction using socioeconomic and behavioral factors. Applied Sciences, 14(6), p.2345.

Ovi, M.S.I., Islam, M.R. & Rahman, M. (2025). Context-aware machine learning framework for stress monitoring in students. arXiv preprint, arXiv:2508.01105.

Patel, R., Shah, K. & Mehta, D., (2023). Prediction of mental health problems using student health survey data. BMC Medical Informatics and Decision Making, 23(1), p.115.

Wang, J., Li, X. & Chen, Y.,(2026). Supervised machine learning models for large-scale student performance prediction. Scientific Reports, 16, p.98765.

Wilson, H., Clark, D. & Adams, R., (2025). Ensemble-based approaches for academic performance prediction. Information Sciences, 652, pp.120–135.

Zhang, K., Liu, X., & Zhao, Y. (2026). AI-driven mental health intervention systems for students. Scientific Reports, 16, p.45678.

Zhao, L., Chen, W. & Sun, Q. (2024). Deep learning-based student performance prediction using multimodal data. Education and Information Technologies, 29(3), pp.3456–3472.