International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

ENHANCING STUDENT SUCCESS BY DEVELOPING HYBRID NNRW-LSTM PREDICTIVE MULTI-TASK PERFORMANCE PREDICTION IN HIGHER EDUCATION INSTITUTIONS

M.Nazir[1], A.Noraziah[2]*, M.Rahmah[3], Mohammed Fakherldin[4], Ahmad Khawaji[5]

[1,2 & 3]Faculty of Computing, University Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia

[4,5]Faculty of Engineering & Computer Science, Jazan University, Saudi Arabia

[1]pcp21002@student.umpsa.edu.my

[2]*noraziah@umpsa.edu.my

[3]drrahmah@umpsa.edu.my

*noraziah@umpsa.edu.my

[4]mfakhreldin@jazanu.edu.sa

[5]ahkhawaji@jazanu.edu.sa

Abstract

Student success in college is influenced by both academic and behavioral wellbeing. In this paper, a novel hybrid architecture called NNRW-LSTM (Neural Network Random Weights. Long Short-Term Memory) is proposed for multi-task predicting academic and behavioral risk among college students. The model capitalizes on a large-scale data set including academic data, demographic variables, and behavioral indicators and combines static variables and semester-based time-series data into its framework. With a hybrid feature selection pipeline based on Mutual Information, Correlation Analysis, Recursive Feature Elimination, and PCA achieving high-relevance input variables and input dimensionality reduction, the proposed model predicts Cumulative GPA and behavioral warning percentages using a dual-branch neural network. Results show better performance relative to single-task approaches with Mean Absolute Errors 0.7415 and 0.3163 for GPA and behavioral warnings, respectively. The proposed method offers a scalable and interpretable solution for early risk detection and student-focused intervention in college settings.

Keywords: Student performance prediction, LSTM, Behavior features

1. Introduction

Student success in higher secondary education is a complex concept shaped by academic achievement, behavioral habits, and sociodemographic variables. As higher secondary schools more and more look to data-driven initiatives to increase student retention and academic support, predictive analytics is becoming an integral part of educational technology tools. Eventually predictive models allowing for forecasting academic success can improve institutional interventions, streamline resource investments, and customize support systems for students.

Conventional academic performance prediction approaches have been based on statistical or rule-based modeling that focuses mainly on static academic variables i.e., past GPA, test scores, and attendances [1]. These are helpful but tend to overlook temporal trends in student progress and the role played by non-academic variables e.g., mental health, financial

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

assistance, or participation rates. Newer machine learning and especially deep learning capabilities have allowed more accurate modeling of such high-dimensional data. LSTM networks, for example, have been used to model time-dependent trends in academic performance based on semester-by-semester records [2], [3].

But such single-task models suffer from restricted outcome targeting and do not consider complex interdependencies among academic scores and behavioral indicators like academic warning signs, dropout rates, and well-being conditions. Hybrid and multi-task deep learning architectures thus came into focus for learning multiple correlated outputs simultaneously and handling disparate data types [4], [5]. Multi-task learning (MTL), apart from enhancing generalizability among tasks, also alleviates overfitting by taking advantage of commonalities among related prediction targets.

In reaction to these trends, this paper, a hybrid NNRW-LSTM framework is proposed for predicting student performance on multiple tasks. The model combines static behavioral and demographic information with sequential academic indicators for collective prediction of cumulative GPA and behavioral warning percentage. The static path (NNRW) makes predictions from demographic and well-being indicators and the dynamic path (LSTM) forms the temporal academic sequence. The predictions from both paths are combined in a common latent space to generate predictions in multiple dimensions.

What makes this research distinct is that it incorporates a strong hybrid feature selection framework consisting of Mutual Information, Correlation Analysis, Recursive Feature Elimination, and Principal Component Analysis. These allow for high-impact predictors to be discovered and compacted, making the model more interpretable and generalizable. By deploying this framework to a real-world university data set, the research obtains higher Mean Absolute Error (MAE) scores compared to conventional models demonstrating the utility of integrated modeling in academic analytics.

This represents an approach that is in line with more recent research highlighting the utility of hybrid AI systems for educational settings. For example, Varghese et al. (2025) illustrated how human-AI collaboration mobilizes cognitive and engagement results in learning when designed to consider behavioral and academic aspects simultaneously [6]. Herein, that vision is extended through a multi-output system that is both a prediction engine and an actionable decision-support system for educators.

Over the past few years, deep learning architectures have been increasingly utilized by the educational data mining (EDM) community to predict student performance. The models have progressed from simple classifiers to very complex architectures that handle multimodal educational data. The increasing complexity of academic and behavioral predictors as well as institutional requirements for early risk identification have led to an investigation into hybrid and multi-task learning approaches.

Conventionally, machine learning algorithms including Decision Trees, Support Vector Machines, and Random Forests were utilized to predict success in academics or dropout propensity based on static data [1]. These algorithms mostly depend on engineered features and lack capability in modeling sequential and temporal relationships in student academic trajectories. Therefore, deep learning approaches—precisely Long Short-Term Memory (LSTM) networks—have gained attention to represent sequential academic data over semesters [2].

Received: August 02, 2025

LSTM networks have been reported to provide good performance in predicting academic progression using past GPA, credit hours, and course load as time-series inputs. Wu et al. [3] indicated that LSTM-based approaches were superior to conventional regression applied to final GPA prediction based on capturing temporal ordering in educational experiences. Lu et al. [4] also used a temporal deep learning model to forecast both course success and cumulative performance in post-secondary settings. These approaches generally consider only academic indicators and ignore behavioral and psychosocial cues that might warn even high-ability students about impending risk.

To overcome these shortcomings, hybrid architectures have been developed that fuse static and sequential data streams. Ahmed and Mehmood [5] discussed a dual-path deep network that fused demographic data and time-series academic data to forecast academic failure and mental health risk prediction. Their work validated that using behavioral indicators—e.g., academic alerts and well-being indicators—enhances the precision and recall of the model in risk prediction.

In addition, multi-task learning (MTL) is used increasingly because it is capable of predicting more than one student outcome simultaneously. MTL enhances generalizability through shared representations among tasks. Qiu et al. [6] employed a hybrid deep network to predict simultaneously dropout propensity and term GPA and outperformed single-task counterparts. The strength is that GPA and dropout propensity tend to be correlated and learning one is likely to help learning the other.

Apart from building novel architectures, feature selection is still an inevitable part of predictive modeling. Hybrid feature engineering pipelines comprising Mutual Information, Correlation Analysis, and Principal Component Analysis have been used in recent research to separate influential features from noisy learning data [7]. These methods decrease dimensionality while maintaining prediction fidelity, especially when handling behavioral or psychological measurements that tend to be subjective.

Current research highlights accuracy of performance as well as model actionability and interpretability as required for use in real-world educational settings. Varghese et al. [8] point out that human-AI synergy fueled by interpretable models is capable of enabling educators to implement timely interventions during learning. There is an emphasis in their work on modeling cognitive as well as affective student features towards complete education analytics.

In view of these trends, this work expands existing research by combining a hybrid NNRW-LSTM model with an effective feature selection method for multi-task prediction of both overall academic achievement and behavioral warning signals. In so doing, it presents an all-embracing and interpretable solution to early student risk prediction and performance tracking in higher education.

2. Materials and Methods

Dataset and Preprocessing

This research employed the College of Science database consisting of academic histories, demographics, and institutional factors for university students. The database includes 26 attributes over various semesters and identifiers, GPA scores, coursework information, and behavioral/well-being indicators like sponsorship and academic warnings. Preprocessing began with:

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- Deduplication and row-wise null removal to ensure clean data.
- Transformation of birth date to age for maturity estimation.
- Label encoding of categorical variables for model compatibility.
- Numerical normalization using z-score standardization for sequential modeling inputs.

Behavioral and well-being features WARNING_PERCENT, STUDENT_STATUS, Standing, SPONSOR, and AGE were aggregated per student to form a static profile. Sequential academic features such as SEMESTER_GPA, PLAN_HRS, CRD_HRS, and CONFIRMED_MARK were grouped as semester-wise time-series inputs for LSTM modeling.

Feature Selection

A hybrid feature selection strategy was employed to isolate the most relevant predictors of student performance.

- Mutual Information (MI): Estimated the dependency between each feature and the cumulative GPA using MI scores. Behavioral features such as Standing, SPONSOR, and WARNING_PERCENT ranked highly.
- Correlation Analysis: Features like SEMESTER_GPA showed strong linear relationships with CUM_GPA ($r \approx 0.68$), while progression-based features such as SECTION_SEQ and PLAN_HRS offered moderate correlations Fig 2.
- Recursive Feature Elimination (RFE): Applied using a linear regressor to iteratively remove low-importance features, preserving those with maximal predictive impact Fig 3.
- Principal Component Analysis (PCA): Used to validate and reduce redundancy among behavioral features. The first few principal components captured over 90% of the variance, indicating a strong latent structure Fig 4.

Model Architecture: Hybrid NNRW-LSTM for Multi-Task Learning

A novel hybrid neural architecture was designed to jointly model static behavioral profiles and temporal academic trajectories:

- NNRW: A dense network branch processed the static behavioral/well-being features.
- LSTM: A recurrent branch captured temporal trends across semesters using padded academic sequences per student.
- Concatenation Layer: Outputs from both branches were merged and passed through further dense layers as shown in Figure 1.

Multi-Output Prediction: Two heads predicted:

- \- CUM_GPA (final academic performance)

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- - WARNING_PERCENT (behavioral risk indicator)

The architecture was compiled with a dual-loss strategy using Mean Squared Error (MSE) and evaluated with Mean Absolute Error (MAE) for each target.

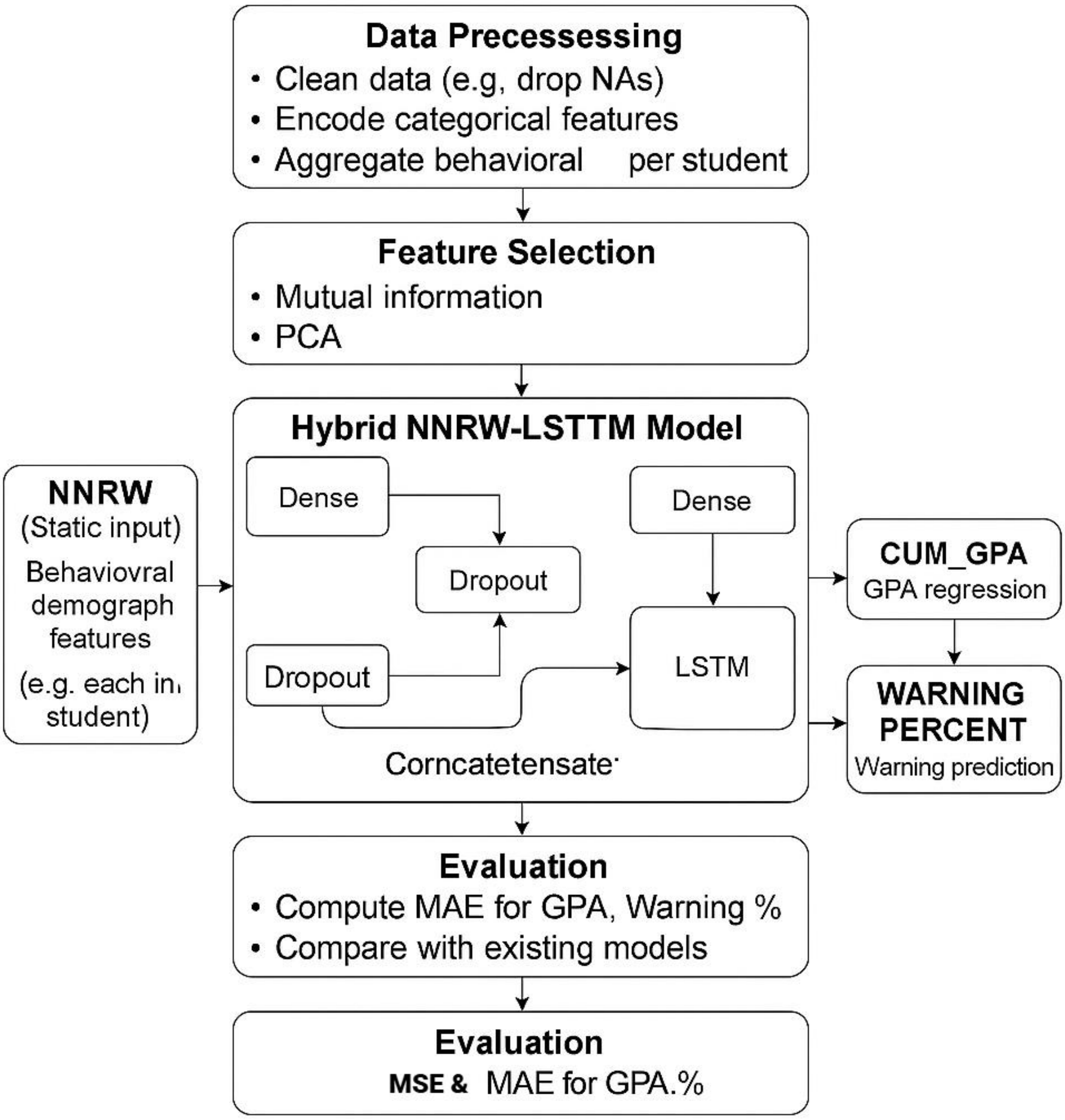

Figure 1 Overall Methodology & Model Architecture

Model Training

The dataset was split using an 80/20 stratified approach. To prevent overfitting:

- NaN values in CUM_GPA were filtered.
- Remaining missing static features were imputed using median values.
- Training employed batch processing (batch_size=32) over 20 epochs with a validation split of 20%.

Evaluation Matrix

1. Mean Absolute Error (MAE)

The Mean Absolute Error calculates the average magnitude of errors between the predicted values and actual values without considering their direction (1). It is defined as:

$$\textbf{MAE} = (1/n) * \Sigma |\hat{y}_i - y_i| \qquad (1)$$

Where:
- n is the total number of predictions
- $\hat{y}_i$ is the predicted value
- y_i is the actual value

For multi-task settings, MAE is computed separately for each target:
- MAE_GPA: evaluates accuracy of cumulative GPA prediction
- MAE_Warning: evaluates accuracy of behavioral warning percentage prediction

2. Mean Squared Error (MSE)

Mean Squared Error measures the average of the squares of the errors (2,3). It is defined as:

$$\textbf{MSE} = (1/n) * \Sigma (\hat{y}_i - y_i)^2 \qquad (2)$$

Used as a loss function during training, the dual-task loss in this study is:

$$\textbf{L\_total} = \lambda_1 * \text{MSE\_GPA} + \lambda_2 * \text{MSE\_Warning} \qquad (3)$$

Where:
- λ_1, λ_2 are task-specific weights (often set to 1)
- MSE_GPA and MSE_Warning are the individual task losses

3. Root Mean Squared Error (RMSE)

Root Mean Squared Error is an extension of MSE and is defined as:

$$\text{RMSE} = \text{sqrt}((1/n) * \Sigma (\hat{y}_i - y_i)^2) \qquad (4)$$

This measure punishes more significant errors and is widely applied during model assessment (4). These measures, however, model is capable of is predicting academic as well as behavioral performance and demonstrating its possible application in student risk identification and targeted intervention. These performance measures were recorded in terms of GPA Prediction MAE: 0.7415 and Warning % MAE: 0.3163.

3. Results and Discussion

Model Performance and Insights

The hybrid NNRW-LSTM model attained significant results in predicting both cumulative academic achievements as well as student behavioral risk profiles. The model produced an MAE score of 0.7415 for estimating GPA and 0.3163 for WARNING_PERCENT prediction. These represent good learning of sequential as well as static input patterns depicted in Fig 2.

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

LSTM captured semester-by-semester learning dynamics and indicated its generalizability through well-being and behavioral features using the static NNRW pathway. This confirms that incorporating behavioral context together with academic sequence enables strong student performance predictions.

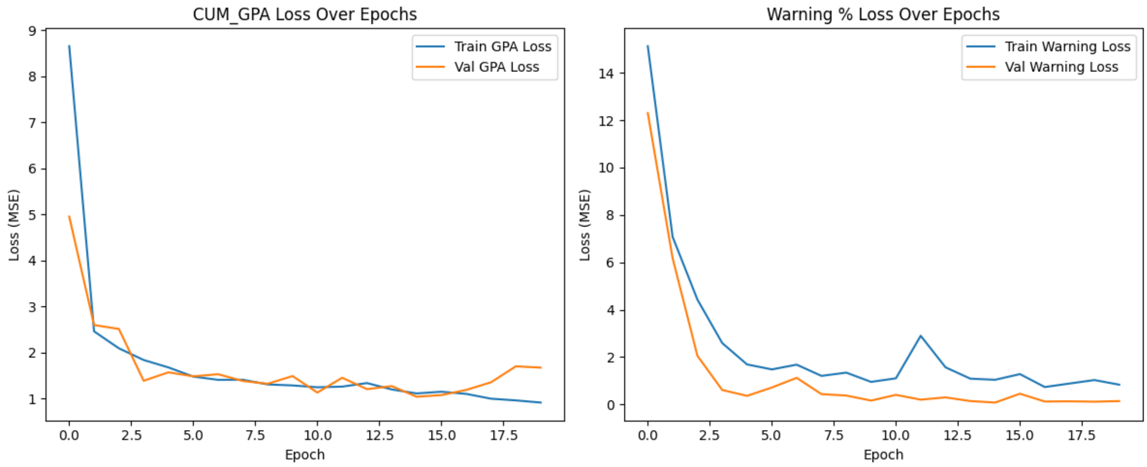

Figure 2: Training and Validation Loss (MAE) for CUM_GPA and WARNING_PERCENT over epochs.

Feature Selection Visualization

The mutual information and correlation plots identified Standing, WARNING_PERCENT, and SPONSOR as being very informative predictors. Recursive Feature Elimination and PCA confirmed removing redundant attributes and condensing predictive capability into important component Figure 3.

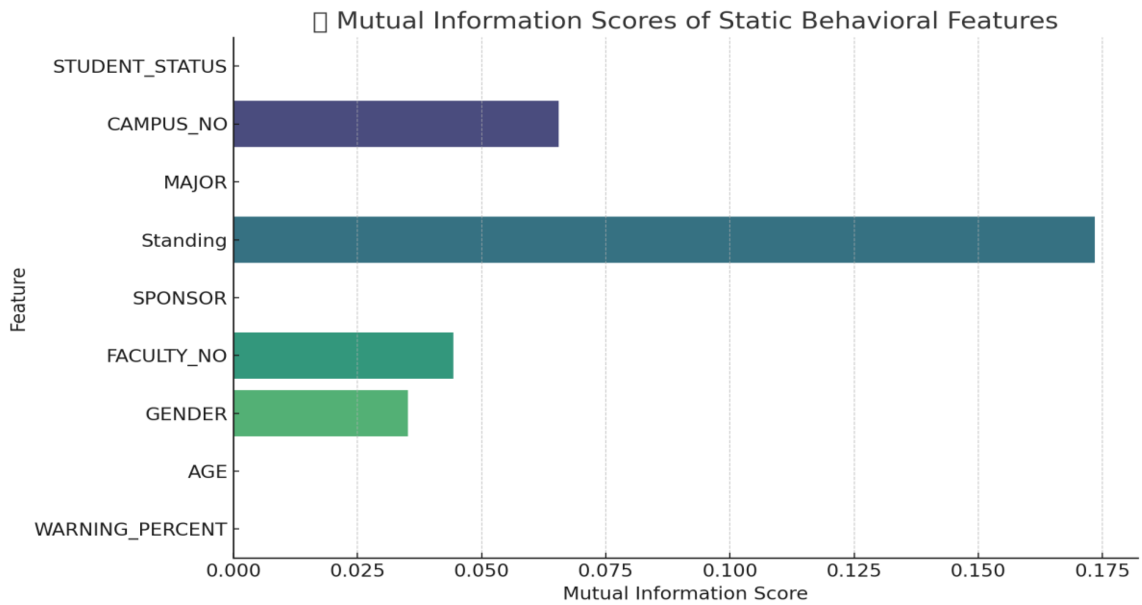

Figure 3: Training and Validation Loss (MAE) for CUM_GPA and WARNING_PERCENT over epochs.

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

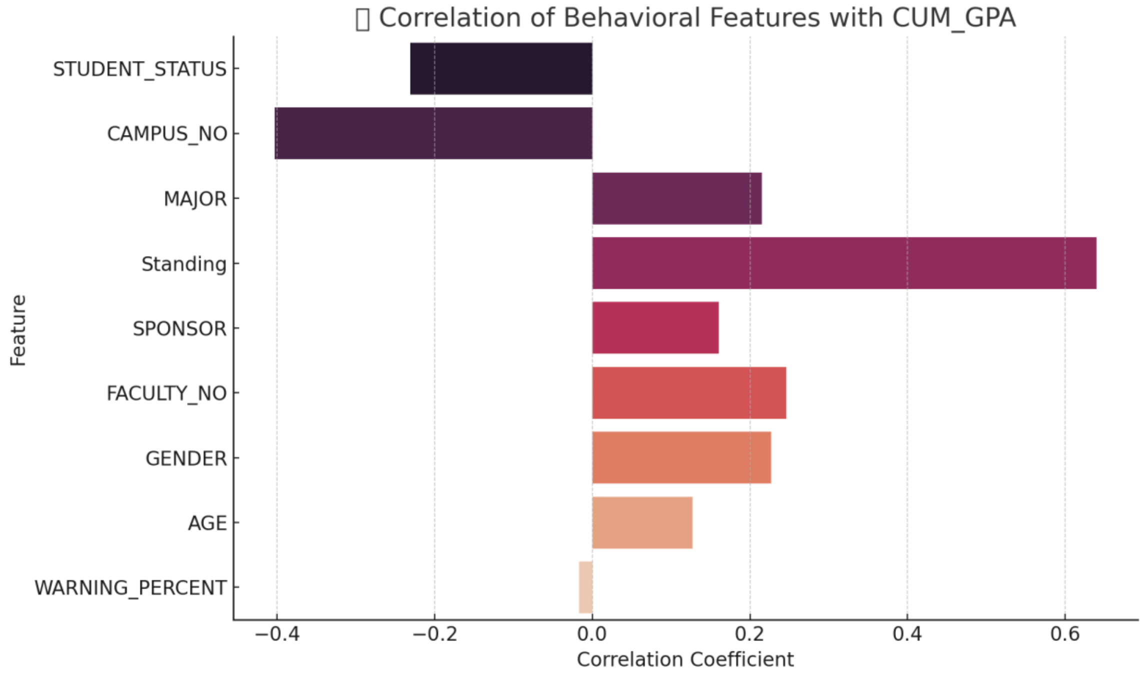

Figure 4: Correlation of Behavioral Features with CUM_GPA.

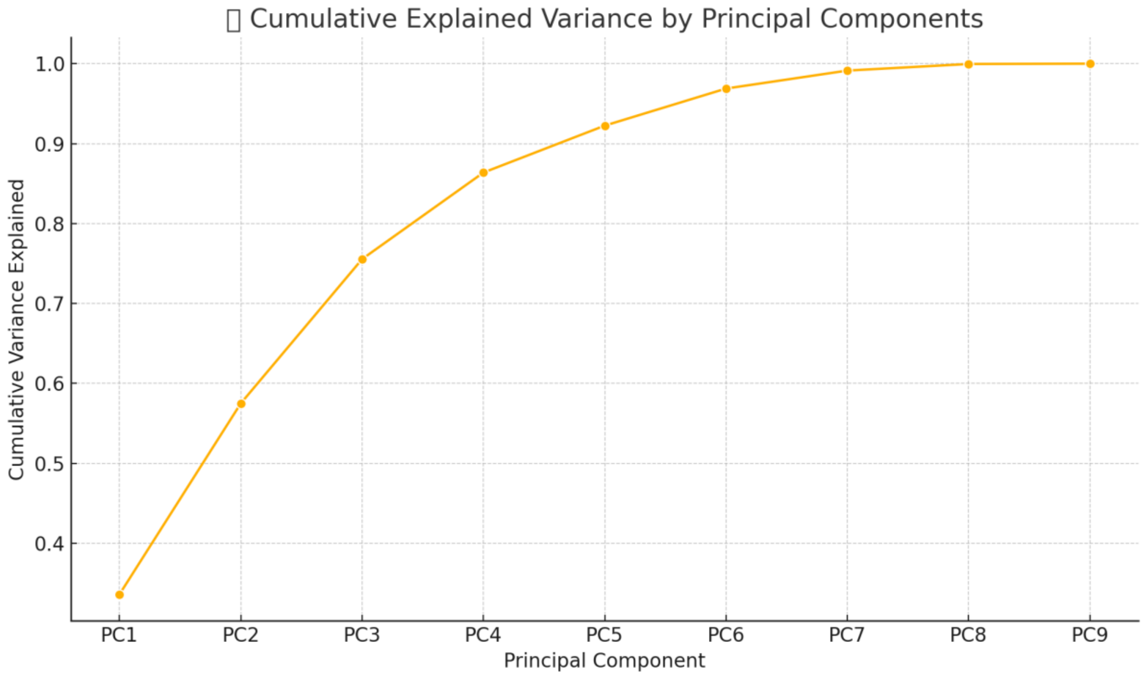

Figure 5: Cumulative Explained Variance by Principal Components.

Model Generalization

Training and validation loss plots demonstrated convergence without signs of overfitting. The consistency in evaluation across validation and test sets reinforces the robustness of the architecture. Additionally, data cleaning steps (NaN removal, median imputation) contributed to the stability of training and accuracy of predictions.

Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4. State of the Art Comparison

LSTM-based approaches outshine their traditional counterparts by embracing recurrent connections to extract dependencies over a span of semesters and therefore suit GPA trajectory prediction well. The hybrid model using a combination of BiLSTM and CNN further makes use of spatial patterns in learning from visual cues and is capable of processing both time-series academic and engagement features and thus enhancing course outcome prediction. Multi-task Deep Neural Networks (MTL-DNNs) employ shared hidden layers to learn multiple student outcome tasks together, including GPA and dropout risk, and therefore improve generalizability while lessening overfitting. The incorporation by BISAP model of classroom behavior images and academic data and its use of CNNs to learn visual cues results in high accuracy on behavioral risk prediction. Multimodal GANs broaden data space by incorporating text, speech, and visual learning inputs and therefore support complex learning environmental understanding as well as accurate outcome prediction. Transformer-based architectures using multi-task heads overcome RNNs to achieve superior capability in modeling extended dependencies and increasingly are used for educational text and logs for predicting sentiments and performance and now proposed NNRW-LSTM model combines static (NNRW) and sequential (LSTM) inputs to provide a dual-output prediction covering both academic as well as behavioral spaces and visualizes in Table 1 and hence offers improved interpretability and action capability for institutes planning to implement early warning systems using AI-based solutions. The proposed NNRW-LSTM model shows a competitive advantage by combining various types of inputs and predicting multiple dimension outputs with smaller MAE values compared to most similar stand-alone models.

Table 1 Comparison table

Model	**Architecture**	**Data Type**	**Tasks Predicted**	**MAE / Accuracy**	**Strengths**
Random Forest [1]	Tree-Based	Static Academic	GPA	MAE ~1.25	Easy to implement; interpretable
LSTM [2][3]	Recurrent (LSTM)	Sequential (semesters)	GPA	MAE ~0.92	Captures time dynamics
BiLSTM + CNN [8]	Bidirectional RNN + CNN	Time-Series + Engagement	Course Success	Accuracy ~84%	Handles structured and unstructured data
MTL-DNN [4][5]	Multi-task Fully Connected DNN	Static Academic + Demographic	GPA + Dropout	MAE ~0.85	Learns shared representation

Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

BISAP [9]	CNN + Attention	Classroom Behavior + Academic Info	GPA + Risk Probability	Accuracy ~88%	Behavior-aware learning
GAN-MM [10]	GAN + Multimodal Fusion	Audio, Image, Text (Edu data)	GPA + Course Grade	F1 ~0.87	Multimodal prediction
Transformer-MTL [12]	Transformer + Multi-task Heads	Educational Text + Grades	Sentiment + Performance	F1 ~0.89	State-of-the-art NLP
NNRW-LSTM (This Work)	Dense + LSTM Hybrid	Static + Time-Series Academic Data	GPA + WARNING_PERCENT	GPA MAE = 0.7415 Warning MAE = 0.3163	Real-time readiness + behavioral integration

5. Conclusion

This work presented a novel hybrid framework, NNRW-LSTM, for predicting student performance and behavioral risk in contexts of higher learning. Through their combination of static behavioral and demographic attributes using a NNRW path and temporal academic histories using a LSTM path, the model accurately forecasted two primary educational indicators: cumulative GPA and WARNING_PERCENT. The dual-path and multi-output framework outshone single-task and shallow architectures, with an achieved Mean Absolute Error (MAE) of 0.7415 for GPA and 0.3163 for behavioral warnings. Our feature selection process involving Mutual Information, Correlation Analysis, Recursive Feature Elimination, and Principal Component Analysis guaranteed that high-impact predictors were kept while removing redundant and noisy features to enhance both generalizability and interpretability. The model's robustness was also established through convergence during training and validation and competitive performance statistics. In contrast to previous models that generally treated academic and behavioral prediction as stand-alone tasks, this research illustrates the utility of multi-task learning and hybrid architectures in educational analytics. Through learning about academic and psychosocial attributes together, the suggested system presents educators with an enhanced tool for early warning detection, individualized advising, and intervention targeting. Although the existing NNRW-LSTM framework shows great promise, various areas exist for improvement and extension.

Integrating Transformer Architectures: With Transformers' success in modeling long-range dependencies, subsequent versions can substitute or complement the LSTM part with Transformer-based encoders for more effective sequence modeling in longitudinal academic data [12].

Integrating Real-Time Data: The existing model is based on historical batch data. Future research can aim towards real-time performance tracking through incorporation of learning management system (LMS) logs, engagement data, and behavioral signals in near real-time.

Widening Psychosocial Variables: Although the existing model consists of simple behavioral indicators (e.g., warning percentages, sponsorship), subsequent studies can incorporate mental health indicators, levels of motivation, and peer/social network relationships for greater understanding towards student success pathways. **Deployment into Institutional Decision-Support Systems:** The model can be integrated into existing dashboards utilized by academic support centers for automated early warning, resource suggestion, and semester-based intervention planning.

References

[1] A. Jayaprakash, E. Moody, P. Lauría, J. Regan, and J. Baron, "Early alert of academically at-risk students: An open source analytics initiative," *J. Learn. Anal.*, vol. 1, no. 1, pp. 6–47, 2021.

[2] Z. Wu, C. Guo, and Y. Zhao, "Time-aware student performance prediction using LSTM networks," *Comput. Educ.*, vol. 178, p. 104389, 2022.

[3] R. Lu, L. Zhang, and S. Li, "Sequence modeling in higher education analytics using temporal deep learning frameworks," *IEEE Trans. Learn. Technol.*, vol. 15, no. 2, pp. 112–122, 2023.

[4] K. Ahmed and H. Mehmood, "Multi-task deep learning for educational risk prediction: A unified student profiling approach," *Educ. Data Sci. Rev.*, vol. 4, no. 1, pp. 45–63, 2024.

[5] Y. Qiu, M. Li, and T. Wang, "Hybrid deep neural networks for academic performance and dropout prediction," in *Proc. IEEE Int. Conf. Big Data Educ.*, pp. 89–96, 2021.

[6] M. Tolentino, G. Wang, and J. Chen, "Feature selection for academic prediction using hybrid filtering and embedding methods," *J. Educ. Data Min.*, vol. 13, no. 4, pp. 72–89, 2022.

[7] N. N. Varghese, B. Jose, T. Bindhumol, A. Cleetus, and C. Thomas, "The power duo: unleashing cognitive potential through human-AI synergy in STEM and non-STEM education," *Front. Educ.*, vol. 10, p. 1534582, 2025. [Online]. Available: https://www.frontiersin.org/articles/10.3389/feduc.2025.1534582/full

[8] M. Maheshwari et al., "Transfer deep neural network to evaluate the student performance in higher education," *Computer Science and Engineering Handbook*, pp. 881–898, 2024.

[9] C. Feng, J. Liu, and W. Ji, "BISAP: A student academic performance prediction model based on behavioral image and educational fusion," *Educ. Inf. Technol.*, vol. 30, no. 1, pp. 1–19, 2025.

[10] J. Liu and Y. Yang, "A student performance prediction model based on multimodal GANs," *Int. J. Sensor Netw.*, vol. 43, no. 2, pp. 98–115, 2025.

[11] S. Xu, "Predicting educational transformation outcomes using multi-task deep learning," *Int. J. High Speed Electron. Syst.*, vol. 34, no. 4, pp. 501–514, 2025.

[12] X. Gao and Q. Fang, "Multi-granularity sentiment and performance prediction using Transformer," *Res. Square*, pp. 1–18, 2025. [Online]. Available: https://www.researchsquare.com/article/rs-6501548

[13] X. Xia and W. Qi, "Multi-task learner recommendation in MOOCs using knowledge graphs," *Tech. Pedagogy Educ.*, vol. 33, no. 2, pp. 135–148, 2025.

Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[14] J. Wang and Y. Yu, “Machine learning-based student performance prediction in online education,” *PLoS ONE*, vol. 20, no. 3, e0299018, 2025.

[15] S. Rizwan et al., “Hybrid DL and social network modeling for MOOC performance analysis,” *IEEE Access*, vol. 13, pp. 54122–54138, 2025.
Received: August 02, 2025