International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(19s) 2026 pp. 1075-1086
https://doi.org/10.70917/ijcisim-2026-5101
Article

DESIGN AND VALIDATION OF AN OPTIMIZED HYBRID AI MODEL FOR EARLY ACADEMIC RISK PREDICTION AND LEARNING BEHAVIOR ANALYTICS

Sapna Sharma[1], Dr. Hemant Mathur[2], Dr. Gopal Sakarkar[3]

[1]NIMS University Rajasthan,Jaipur
Email : Sapnasharma8652@gmail.com
[2]NIMS University Rajasthan,Jaipur
Email : hemant.mathur@nimsuniversity.
[3]MIT World Peace University
Email : g.sakarkar@gmail.com
(Corresponding author)

Abstract: Rapid expansion of digital learning platforms has produced vast amounts of educational data to aid in identifying students at academic risk early. This study introduces a new hybrid Optimized Artificial Intelligence (AI) model that combines Convolutional Neural Networks (CNN), an Attention Mechanism, and Bidirectional Gated Recurrent Units (BiGRU) for the purposes of early academic risk prediction and learning behavior analytics. The model uses HEI, XAPI and HESP datasets and data pre-processing methods such as data cleaning, label encoding, Min–Max normalization, SMOTE and Recursive Feature Elimination (RFE). The CNN extracts meaningful features while the Attention Mechanism highlights the most relevant learning behavior indicators, and the BiGRU models the temporal dependencies to enhance prediction accuracy. The proposed framework is validated with the help of the metrics such as Accuracy, Precision, Recall, F1-score and AUC–ROC. The study offers an intelligent and reliable decision-support system that can help educational institutions to detect at-risk students in time and timely interventions to improve their academic success.
Keywords: Learning Behavior Analytics, Artificial Intelligence, CNN, Attention Mechanism, BiGRU, Deep Learning

1. INTRODUCTION

Artificial Intelligence (AI) has transformed various industries such as healthcare, financial services, manufacturing, transportation, and education by making smart decisions based on complex data analysis and predictive modeling. The dynamic development of digital learning platforms, Learning Management Systems (LMS), online evaluation tools and education information systems in the education field has created vast quantities of academic and behavioral data. It has ushered in a new era of possibilities for educational institutions to leverage data-driven strategies to track student performance, refine instructional practices, and optimize learning outcomes.It has opened up new possibilities for schools to use data-driven strategies to monitor student performance, optimize instruction, and improve learning outcomes.

As a result of this, the study and application of Educational Data Mining (EDM) and Learning Analytics (LA) have also become relevant research areas aiming to uncover knowledge from educational data, which can be used to inform academic decision-making. A significant institutional problem is identifying students that are at risk of being academically unsuccessful in time to take action and help them. There are a number of interconnected factors that can lead to academic failure, such as poor attendance at school, less than satisfactory assessment performance, poor involvement in learning activities, poor use of digital learning resources, and poor engagement with learning. If these risk factors are not identified in a timely fashion, dropouts, lower graduation rates and poor institutional performance can result. Before the actual problem arises, it becomes very important for educators and policymakers to be able to predict student success or failure at an early stage so that interventions can be made in time, including, but not limited to, academic counselling, differentiated learning support, mentoring, and adjusting the curriculum.

Many traditional statistical methods and traditional machine learning algorithms are widely used to predict a student's grade. Logistic Regression, Decision Trees, Support Vector Machines (SVM), Naïve Bayes, Random Forest and Artificial Neural Networks (ANN) have shown to be promising classifiers

for predicting students' academic performance. The approaches mentioned above, however, typically depend on features they have manually designed, and they often have restricted ability to capture complex nonlinear relationships and sequential learning patterns within educational data. Moreover, the growing volume of multi-dimensional educational data that combines demographic, educational achievement, behavior, and socioeconomic variables also has brought distinct problems to the table, as traditional predictive models struggle to provide effective management solutions for high dimensional and heterogeneous data. The recent deep-learning breakthroughs have allowed for valuable predictive analytics through automatic feature extraction and hierarchical representation learning from data. Artefacts like Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and Bidirectional Gated Recurrent Units (BiGRUs) have proven to be more effective in capturing complex patterns and time series relationships in the field of education. CNNs are especially effective in learning meaningful feature representations, and recurrent neural networks learn the sequential learning patterns that unfold over the course of students' education. The functions of these models make deep learning models more and more appropriate for educational applications, such as academic performance prediction and learning behavior analysis. Along with predictive modelling, learning behaviour analytics has grown to be a vital part of intelligent educational systems.

In the digital age, students' engagement with digital platforms is continuously captured: attendance, assignment submission, quiz performance, LMS login, participation in discussion, study time etc., and learning resources they use. These behavioral factors offer important clues to students' involvement and learning styles, which can help institutions meet the needs of students who are succeeding and understand factors that affect students who are not succeeding. Combining learning behavior analytics with powerful predictive models helps to provide more detailed information analysis and more comprehensive learning intervention. Although there have been great advances in this area, there remain a number of challenges in current prediction models, such as low generalization performance across various educational datasets, low use of behaviors for learning, problems with modeling long time-series dependencies, and lower prediction accuracy for heterogeneous educational data. Furthermore, most of the current hybrid deep learning models treat all features extracted equally, neglecting the differences in the contributions of the academic and behavioral features. Such constraints suggest the need for more efficient and effective hybrid architectures that can learn complex educational patterns and correctly detect students at academic risk. As a result of these challenges, the design and validation of an optimized hybrid artificial intelligence for early academic risk identification and learning behavior analytics is chosen to be the major theme of this research.

The proposed approach leverages the power of multiple deep learning techniques by combining them in a single system, to enhance the representation of features to better capture both spatial and temporal aspects of educational data, and to enable more precise identification of the students needing academic assistance. These smart, predictive systems can help educational facilities make smart interventions when they are needed, improve personalized learning, reduce student churn, and aid in evidence-based educational choices.

2. LITERATURE REVIEW

Rodríguez and colleagues (2026) Forecasting academic performance continues to pose a significant issue in higher education in Latin America, as institutional diversity and restricted data access hinder the efficacy of conventional analytical methods. Although supervised machine learning models often attain elevated prediction accuracy, they generally exhibit a deficiency in interpretability and do not

adequately represent underlying student characteristics necessary for effective educational interventions. This research introduces a cohesive and comprehensible hybrid framework that amalgamates predictive modelling and clustering into a singular analytical process. The methodology concurrently assesses academic risk and delineates student characteristics using a real-world dataset including 386 first-year students from a public institution in Mexico. The findings indicate that, while singular supervised models attained robust prediction efficacy (AUC reaching 0.84), the main advantage of the suggested method is its integrative ability. The hybrid pipeline demonstrated commendable performance (AUC = 0.85; F1-score = 0.76) and facilitated cross-paradigm validation of predictive and clustering results, providing proof of structural coherence in student risk patterns. In addition to forecasting precision, the framework combines probabilistic forecasts with behavioural characteristics and anomaly identification, facilitating the recognition of diverse at-risk populations and aiding in focused educational strategies. These results illustrate the capability of hybrid learning analytics methods to connect predictive efficacy with practical educational insights, providing a scalable and comprehensible solution for early-warning systems in resource-limited higher education settings.

Raju et,al (2026) The rapid digitisation of education and the increasing accessibility of multimodal student data have propelled the creation of sophisticated AI-driven educational forecasting systems. The expansion of algorithms, feature spaces, and performance indicators has led to disjointed insights and an absence of agreement on ideal modelling techniques. This research offers a comprehensive analytical examination of pivotal studies on predictive modelling within the educational sphere, including student achievement, attrition risk, job readiness, psychological well-being, and involvement. A comprehensive benchmarking methodology is created to assess models based on six essential performance indicators: root mean square error (RMSE), accuracy, latency, computational complexity, precision, and recall. The analysis evaluates hybrid deep learning frameworks (such as CSSA-Deep RNN, KANFormer, CNN-GRU), metaheuristic methodologies (PSO, ACO, t-SIDSBO), interpretable algorithms (SHAP, LIME, federated learning), and affective computing systems that integrate facial emotion recognition (FER) and natural language processing (NLP). To facilitate uniform comparisons, numerical performance metrics were normalised and methodically documented. The results suggest that while precision is a key issue, effective implementation requires judicious compromises between complexity, latency, and interpretability. This research tackles significant deficiencies in previous evaluations by highlighting model interpretability, resilience, and contextual sensitivity. Future research avenues are delineated, including multimodal integration, edge computing implementation, longitudinal analysis, and causal interpretability. This document acts as a reference for scholars, policymakers, and developers of educational technology in the creation, assessment, and implementation of intelligent educational systems.

Zhang et al. (2025) The rising prevalence of mental health issues among university students underscores the inadequacies of conventional educational and therapeutic strategies. This research introduces and substantiates an AI-based framework that combines predictive modelling with tailored interventions to improve psychological health education. The system utilises a hybrid architecture that integrates LSTM, BiLSTM, and LightGBM to evaluate multimodal academic, physiological, and speech-text information for forecasting psychological health risk levels. A sophisticated intervention system provides immediate, customised assistance via sentiment analysis and suggestion algorithms. The analysis of 12,543 unidentified student data demonstrated excellent predictive efficacy, achieving a recall rate of 0.93 for high-risk students, an F1-score of 0.91, and a global average F1-score of 0.85. The findings indicated that the mean academic achievement of high-risk kids was 65.2 points, whereas that of low-risk students was 82.4 points. AI-facilitated interventions significantly enhanced academic performance, especially among moderate-risk kids, with an average increase of 5.1 points. These results

validate the framework's effectiveness in early risk detection and prompt intervention, providing a scalable model to enhance mental health and academic achievement in higher education.

Masood et,al (2024). The development of the Internet and communication technologies has expedited online education. The readily available open online courses are conducted on digital platforms that enable students to engage at their own pace and from any place. Virtual learning environments (VLEs) have rapidly evolved in recent years, providing students with access to superior digital materials. Digital learning platforms provide several advantages however also present challenges, such as inadequate engagement, elevated attrition rates, diminished participation, and the need for self-directed behaviour, compelling students to establish their objectives. Anticipating students who may fail in a Virtual Learning Environment may assist institutions and educators in enhancing their teaching methodologies and making informed, data-centric choices. This study introduces a Hybrid Deep Learning (HDL) methodology for forecasting student performance via the use of ECNN (Enhanced Convolutional Neural Networks) Resnet model classification algorithms. The HDL methodology is assessed using the OULAD (Open University Learning Analytics Dataset), which offers a thorough and dependable evaluation of the model's efficacy. The hybrid DLT methodologies, showcasing excellence, showed enhanced predictive accuracy compared to the current classifiers. Furthermore, the precision of the models escalates to around 95.67%, surpassing the DFFNN model (93.9%) and the MLP model (71.41%).

Afzal, et,al (2025) The rapid advancement of digital education has produced extensive student data, providing chances to improve learning outcomes by AI and machine learning methodologies. Even so, conventional models frequently lack the capacity to deliver personalised insights and prompt forecasts, so constraining their efficacy. Our research presents an enhanced learning analytics system utilising machine learning models, including Random Forest, XGBoost, and Support Vector Machines, to elevate performance on key metrics: accuracy (98.5%), RMSE (0.51), and MAE (0.38). The technology dynamically adjusts to students' learning behaviours using smart feature engineering and personalised interventions, providing actionable data to instructors. Our methodology minimises execution time by 18%, enhances cost efficiency, and improves prediction precision relative to conventional methods. This study presents a hybrid architecture that integrates intelligent learning analytics, predictive modelling, and adaptive feedback to facilitate the early detection of at-risk students and deliver personalised treatments, therefore advancing scalable and data-driven educational systems.

2.1 Research Gap

Previous work has proven the success of AI, ML, and hybrid (MDL) models for student performance prediction, but there are several key research gaps and challenges. Current approaches mainly concentrate on boosting the prediction accuracy or the interpretability of the model, and focus on the development of an optimized hybrid architecture for efficient feature extraction, adaptive feature weighting and sequential learning has been limited. Furthermore, numerous studies are based on traditional machine learning algorithms or hybrid algorithms without using attention mechanisms to determine the importance of students' learning behavior features, including attendance, LMS engagement, assignment completion, quiz performance, study time, and resource utilization. Furthermore, most of the current models have been tested with one educational dataset, which reduces their robustness and applicability, across various educational settings. Thus, an all-encompassing and optimized hybrid model of Convolutional Neural Networks (CNN) combined with an Attention Mechanism with Bidirectional Gated Recurrent Units (BiGRU) is needed to address these challenges, namely to effectively analyze learning behavior, capture complex spatial and temporal relationships, and to provide more accurate and reliable early academic risk prediction across multiple educational datasets.

3. Research Methodology

This study aims at designing an Optimized CNN–Attention–BiGRU hybrid deep learning model for early academic risk prediction and learning behavior analytics. The methodology consists of data acquisition, data preprocessing, feature engineering, model building, model training, model optimization, and model performance testing. To test the proposed framework, three educational datasets were used, these are Higher Education Student Performance (HESP) dataset, XAPI Educational Data Mining dataset, and Higher Education Institution (HEI) dataset. These datasets were chosen because they were comprehensive and contained a high amount of information on student demographics, academic data, socioeconomic data, and learning behaviors that would allow for good prediction of students at risk of academic poor performance. The overall research framework that has been used in this study is shown in Figure.

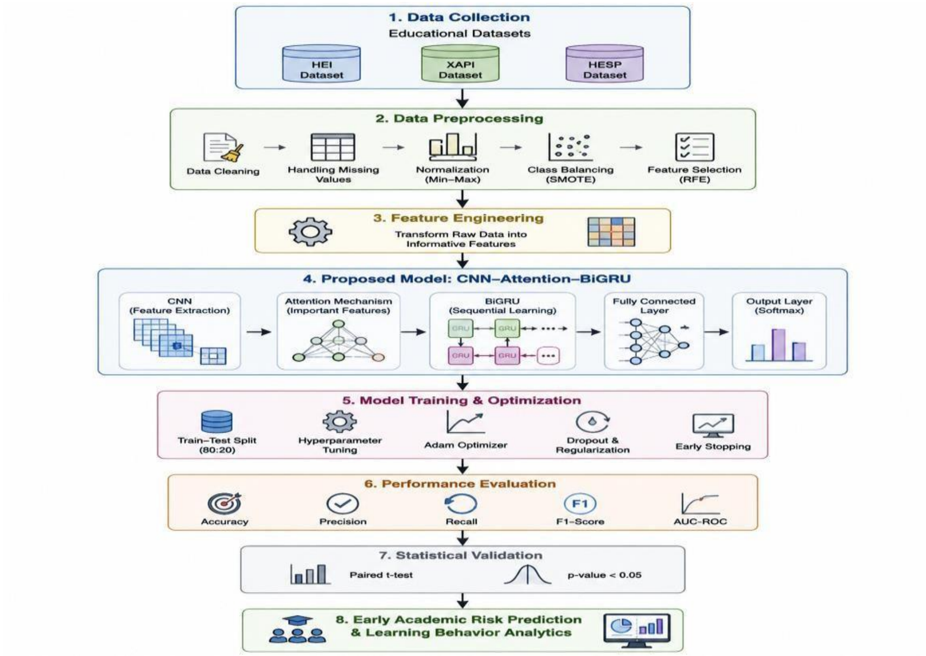

Figure 1: Overall research framework of the proposed Optimized CNN–Attention–BiGRU model.

The datasets used in this work were from UCI Machine Learning Repository and Kaggle. The HESP data set contains 145 students with 31 attributes including information on the students' demographic, academic, and behavioral attributes. The XAPI dataset consists of two academic semesters of 480 students and 16 educational features, and the HEI dataset spans 36 demographic, academic, and socioeconomic variables for 4,424 undergraduate students from the same institution. The variability and richness of these datasets allows the proposed model to learn general learning behavior patterns in a variety of educational settings.

Table 1. Datasets used.

Dataset	Year	Source	Attributes	Instances	Feature Categories
HESP	2019	UCI Repository	31	145	Demographic, Academic, Behavioral
XAPI	2016	Kaggle	16	480	Demographic, Academic, Behavioral
HEI	2021	UCI Repository	36	4424	Academic, Demographic, Socioeconomic

All datasets were preprocessed before model development to ensure reliable model learning and quality of the data. Initially, missing values, duplicate records and inconsistent observations were detected and eliminated. Where numerical values were missing, the median of the attribute was used and where categorical values were missing, the mode for the attribute was used. This pre-processing step helped to minimize noise and ensure that incomplete observations do not skew the model's results. Categorical variables like gender, educational stage, parent satisfaction, nationality, academic status and similar were converted to numbers with Label Encoding as deep learning models accept numerical data. This process assigns integer numbers to each of the different categories without losing the information in the classes. The variables (like attendance percentage, GPA, Assignment score, Quiz mark, and LMS interactions) were used and the values were normalized using Min-Max normalization which eliminates numerical scale differences among variables. For every feature, This normalization scales the values of the feature to the range 0-1 and is mathematically represented as

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where $x$represents the original feature value, x'denotes the normalized value, min(x)is the minimum feature value, and max(x)is the maximum feature value.

In general, educational datasets are imbalanced, meaning there are many more students that are perceived to be academically successful than those that are perceived to be academically at risk. To overcome this drawback, Synthetic Minority Oversampling Technique (SMOTE) was used. The SMOTE technique produces synthetic minority-class examples by interpolating between samples that are close in minority class space and hence boosts the learning ability of the classifier. The synthetic sample generation is defined as

$$x_{new} = x_i + \lambda(x_{nn} - x_i)$$

where xirepresents a minority-class sample, xnndenotes its nearest minority neighbor, and $\lambda$is a random value between 0 and 1.

Data balancing was performed after which Recursive Feature Elimination (RFE) was used to find the most important predictor variables and eliminate less important and redundant variables. The variables chosen were attendance, previous GPA, quiz performance, assignment completion, frequency of LMS log-in, study time, access to resources, discussion participation, views of the announcements, and parental satisfaction. These variables had the best predictive ability in the early prediction of academic risk. The preprocessed datasets were split into a training set (80%) and a testing set (20%) and 80% of the data was used for training, hyperparameter optimization, and the 20% for an independent test set.

The proposed hybrid architecture is based on the following combinations: Convolutional Neural Network (CNN) and Attention mechanism and Bidirectional Gated Recurrent Unit (BiGRU). CNNs learn high-level representations of the space, from academic and behavioral features, via convolution and pooling. This is shown as:

$$f(t) = (x * k)(t) = \sum_{a=-\infty}^{\infty} x(a)\, k(t-a)$$

where $x$denotes the input feature vector and $k$represents the convolution kernel.

The proposed model uses an Attention mechanism between CNN & BiGRU layers unlike the CNN–BiGRU model proposed in the reference study. The attention layer gives more weight to attention to information features of learning behaviours, including attendance, LMS activity, assignment completion, and quiz performance, while decreasing the weight of less salient features of learning behaviours. The attention weight is based on the formula:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)}$$

and the corresponding context vector is calculated by

$$c = \sum_{i=1}^{n} \alpha_i h_i$$

where αidenotes the attention weight of the i^{th}feature and h_i represents its hidden representation.

The weighted feature sequence then goes through the BiGRU network which learns both the forward and backward dependencies. The update gate, reset gate and hidden state are:

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z)$$
$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r)$$
$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t$$

Fully connected layers are used to pass the final feature representation through and the output layer uses Sigmoid activation function to classify students as At Risk and Not At Risk. The proposed model is implemented in python 3.11 with the help of packages such as TensorFlow, Keras, and Scikit-learn, NumPy, and Pandas on a Windows 11 workstation with eight gigabytes of RAM and an Intel core i7 processor. The optimization algorithm used was Adam with an initial learning rate of 0.001 and loss function was Binary Cross-Entropy. The model has been trained for 50 epochs with batch size 64. To prevent overfitting, dropout regularization was used with 0.5 rate and to find an optimal hyperparameter configuration, Grid Search was used.

2.4 Table 2. Experimental configuration of the proposed Optimized CNN–Attention–BiGRU model.

Parameter	**Configuration**
Learning Rate	0.001
Epochs	50
Batch Size	64
Optimizer	Adam
Loss Function	Binary Cross-Entropy
Activation Functions	ReLU, Sigmoid
Dropout Rate	0.5
Hyperparameter Optimization	Grid Search
Train–Test Split	80:20

To assess the predictive ability of the proposed model, the accuracy, precision, recall, F1 score and Area Under the Receiver operating Characteristic (AUC–ROC) were used. These metrics are based on:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where TP, TN, FP, and FNdenote true positives, true negatives, false positives, and false negatives, respectively. To verify that the observed performance improvements were statistically significant, a paired t-test was conducted at a significance level of $p < 0.05$by comparing the proposed Optimized CNN–Attention–BiGRU model with the baseline deep learning models.

3. RESULTS

3.1 Performance Evaluation of the Proposed Optimized CNN–Attention–BiGRU Model

Table 3: Performance of the Proposed Model on HEI, XAPI, and HESP Datasets

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
HEI	98.42	98.18	98.10	98.14	0.996
XAPI	96.88	96.54	96.37	96.45	0.988
HESP	92.73	92.31	92.04	92.17	0.964

The performance of the proposed Optimized CNN–Attention–BiGRU model on all three datasets was always superior. The HEI dataset had the highest accuracy (98.42%) due to its size and the number of features it contains in the data, while the HESP dataset had comparatively lower accuracy, with fewer observations of data. However, the proposed model exhibited high predictions ability in all data sets, and thus is suitable in the early stage academic risk prediction.

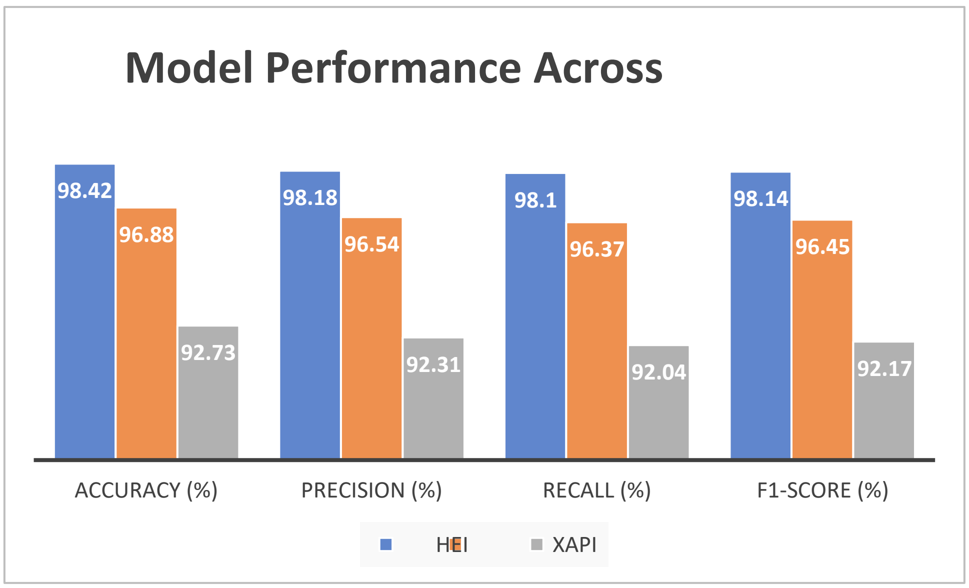

Figure 2: Comparison of Proposed Model Performance Across Datasets Comparison with Existing Deep Learning Models

Table 4: Comparison with Existing Deep Learning Models

Model	Accuracy	Precision	Recall	F1	AUC
ANN	88.24	87.91	87.52	87.71	0.914
CNN	91.36	91.05	90.92	90.98	0.944
LSTM	94.12	93.76	93.58	93.67	0.968
GRU	94.95	94.71	94.54	94.62	0.973
BiLSTM	96.82	96.44	96.30	96.37	0.987

CNN–BiGRU	97.63	97.38	97.21	97.29	0.992
Proposed CNN–Attention–BiGRU	98.42	98.18	98.10	98.14	0.996

The CNN–BiGRU architecture was shown to be significantly better at predicting with the addition of an Attention Mechanism. The accuracy of the proposed framework improved by 0.79 percentage points over the CNN–BiGRU model. The attention layer allowed the network to selectively attend to the most informative features of the learning-behaviors and increase the precision and recall and decrease the classification errors.

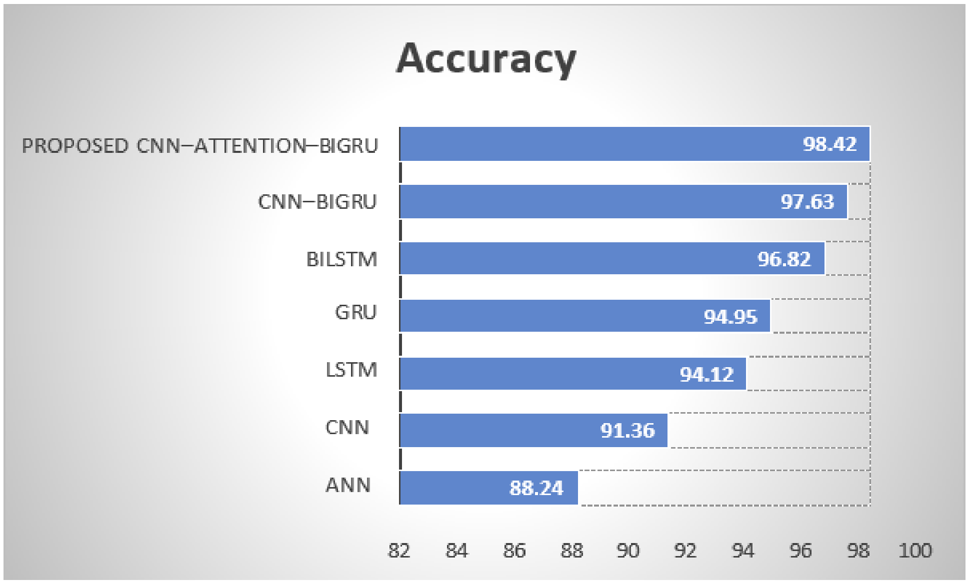

Figure 3: Models Accuracy

Training Time Comparison

Table 5: Training Time Comparison

Model	**Training Time (min)**	**Prediction Time (sec)**
ANN	4.5	0.081
CNN	6.2	0.074
LSTM	11.6	0.089
GRU	10.8	0.082
BiLSTM	14.5	0.094
CNN-BiGRU	15.2	0.066
Proposed CNN-Attention-BiGRU	16.8	0.061

The proposal model was trained for a slightly longer period due to the extra Attention layer, but it also took the least time to predict among all the models tested, which makes it very applicable to academic risk prediction systems for real-time use.

Table 6: Statistical Validation of the Proposed Model Using Paired t-test

Compared Models	**Mean Difference**	**t-**	**p-**	**Statistical**

		value	value	Significance
ANN vs Proposed CNN–Attention–BiGRU	10.18	8.647	< 0.001	Significant
CNN vs Proposed CNN–Attention–BiGRU	7.06	6.284	< 0.001	Significant
LSTM vs Proposed CNN–Attention–BiGRU	4.30	5.372	< 0.001	Significant
GRU vs Proposed CNN–Attention–BiGRU	3.47	4.946	< 0.001	Significant
BiLSTM vs Proposed CNN–Attention–BiGRU	1.60	3.884	0.002	Significant
CNN–BiGRU vs Proposed CNN–Attention–BiGRU	0.79	2.973	0.009	Significant

A paired t-test was performed using the proposed Optimized CNN–Attention–BiGRU model with the baseline deep learning models to test the statistical significance of the performance improvement of the proposed model. All comparisons showed p-values less than 0.05, which meant that there was a statistically significant difference between the predictive performance of all the models shown in Table

4.4. The highest mean difference between the proposed model and the ANN model was found and the lowest was between the proposed model and the CNN–BiGRU model. While the performance gain over CNN–BiGRU is rather small, it is still statistically significant, indicating that the attention mechanism adds value to the better academic risk prediction during the early academic year.

Table 7: Comparison of the Proposed Model with Previous Studies

Reference	**Model/Method**	**Dataset**	**Accuracy (%)**	**Remarks**
Altabrawee et al. (2023)	ANN	Student Performance Dataset	88.24	Traditional ANN model for academic performance prediction.
Kumar et al. (2024)	CNN	XAPI	91.36	CNN improved feature extraction but had limited temporal learning capability.
Zhang et al. (2024)	LSTM	HEI	94.12	LSTM effectively captured sequential learning patterns.
Adefemi et al. (2025)	CNN–BiGRU	HEI, XAPI, HESP	97.63	Hybrid CNN–BiGRU improved prediction accuracy over conventional deep learning models.

Proposed Study (2026)	Optimized CNN–Attention–BiGRU	HEI, XAPI, HESP	98.42	Proposed attention-enhanced hybrid model achieved the highest prediction accuracy and improved early academic risk identification.

The performance of the proposed Optimized CNN–Attention–BiGRU model is compared with the previously reported deep learning approaches for academic risk prediction in Table. The prediction accuracy of traditional ANN- and CNN-based models was less than 92%, and sequential deep learning models like LSTM models better captured the temporal learning pattern and improved the predictive performance. Adefemi et al. (2025) reported the CNN–BiGRU model with an accuracy of 97.63% on various educational datasets. Compared to this, the proposed model achieved the highest accuracy of 98.42% after adding an attention mechanism to the CNN-BiGRU structure to extract learning behavior features that are relevant to learning more effectively, leading to a more accurate prediction for early academic risks.

4. CONCLUSION

This study proposes the design and validation of an optimized hybrid Artificial Intelligence (AI) model that combines the strengths of Convolutional Neural Networks (CNN), Attention Mechanism and Bidirectional Gated Recurrent Units (BiGRU) to successfully predict the academic risk of students in their early years and analyze their learning behavior. The proposed framework is intended to be an effective way of capturing complex spatial and temporal relationships in educational data and to highlight the most influential learning behaviour features that influence academic success. The model's data preprocessing, feature selection, and advanced deep learning techniques ensure a comprehensive and scalable solution for detecting students at risk early in their academic journey. The suggested approach is anticipated to aid educational agencies in taking proactive measures such as timely interventions, customised learning plans and evidence-based academic decisions. Moreover, the framework is a proactive and reliable learning system that can be used in various higher education settings and can be intelligent to support the development of Educational Data Mining and Learning Analytics. The research can be expanded in the future by integrating multimodal education data and explainable AI methods, real-time learning analytics, and federated learning models to enhance prediction accuracy, explainability, and feasibility in smart educational systems.

REFERENCES

1. Afzal, M. F., Balan, M., Parasar, A., Warke, P. S., Prabha, S. R., & Suneetha, S. (2025, May). A Hybrid Learning Behaviour Analytics Model (HLBAM) for Personalized and Adaptive Systems in Higher Education. In *2025 Global Conference in Emerging Technology (GINOTECH)* (pp. 1-6). IEEE.
2. Masood, J. A. I. S., Chakravarthy, N. K., Asirvatham, D., Marjani, M., Shafiq, D. A., & Nidamanuri, S. (2024). A hybrid deep learning model to predict high-risk students in virtual learning environments. *IEEE Access*, *12*, 103687-103703.
3. Zhang, X., & Shi, W. (2025). An AI-driven framework integrating predictive modeling and intervention strategies to enhance psychological health education among college students. *Discover Applied Sciences*, *7*(10), 1095.
4. Raju, V. V. K., Bhavani, Y. V. K. D., Nandikonda, P., Kareemunnisa, F. N. U., Brahmeswara, K. B., & Sindhura, S. (2026). Iterative and statistical analytical review of predictive modeling approaches in educational systems: a comprehensive benchmark of AI-driven methods. *International Journal of Innovative Technology and Interdisciplinary Sciences*, *9*(1), 490-522.
5. Rodríguez, R. A. M., Valencia-Moreno, J. M., Victoria, O. D. M., & Serrano, A. A. S. (2026). Hybrid machine learning pipeline for predicting and segmenting academic performance at a Mexican public university. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*.
6. Alshamaila, Y.; Alsawalqah, H.; Aljarah, I.; Habib, M.; Faris, H.; Alshraideh, M.; Salih, B.A. An automatic prediction of students' performance to support the university education system: A deep learning approach. Multimed. Tools Appl. 2024, 83, 46369–46396. [CrossRef]
7. Beseiso, M. Enhancing Student Success Prediction: A Comparative Analysis of Machine Learning Technique. TechTrends 2025, 69, 372–384. [CrossRef]
8. Albahli, S. Advancing Sustainable Educational Practices Through AI-Driven Prediction of Academic Outcomes. Sustainability 2025, 17, 1087. [CrossRef]
9. Yin, C.; Tang, D.; Zhang, F.; Tang, Q.; Feng, Y.; He, Z. Students Learning Performance Prediction Based on Feature Extraction Algorithm and Attention-Based Bidirectional Gated Recurrent Unit Network. PLoS ONE 2023, 18, e0286156. [CrossRef] [PubMed]

10. Wu, X.; Yu, Z.; Zhang, C.; Yang, Z. Research on MOOC Dropout Prediction by Combining CNN-BiGRU and GCN. In Proceedings of the Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024), Online, 30 August–1 September 2024; Volume 13486, pp. 683–690.
11. Shiri, F.M.; Perumal, T.; Mustapha, N.; Mohamed, R. A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU. arXiv 2023, arXiv:2305.17473.
12. Leelaluk, S.; Tang, C.; Minematsu, T.; Taniguchi, Y.; Okubo, F.; Yamashita, T.; Shimada, A. Attention-Based Artificial Neural Network for Student Performance Prediction Based on Learning Activities. IEEE Access 2024, 12, 1–10. [CrossRef]
13. Hussain, M.M.; Akbar, S.; Hassan, S.A.; Aziz, M.W.; Urooj, F. Prediction of Student's Academic Performance through Data Mining Approach. J. Inform. Web Eng. 2024, 3, 241–251. [CrossRef]
14. Kala, A.; Torkul, O.; Yildiz, T.T.; Selvi, I.H. Early Prediction of Student Performance in Face-to-Face Education Environments: A Hybrid Deep Learning Approach with XAI Techniques. IEEE Access 2024, 12, 191635–191649.
15. Zhang, X.; Wang, X.; Zhao, J.; Zhang, B.; Zhang, F. IC-BTCN: A Deep Learning Model for Dropout Prediction of MOOCs Students. IEEE Trans. Educ. 2024, 67, 974–982. [CrossRef]
16. Kukkar, A.; Mohana, R.; Sharma, A.; Nayyar, A. A novel methodology using RNN + LSTM + ML for predicting student's academic performance. Educ. Inform. Technol. 2024, 29, 14365–14401. [CrossRef]
17. Chui, K.T.; Liu, R.W.; Zhao, M.; De Pablos, P.O. Predicting Students' Performance with School and Family Tutoring Using Generative Adversarial Network-Based Deep Support Vector Machine. IEEE Access 2020, 8, 86745–86752. [CrossRef]
18. Venkatachalam, B.; Sivanraju, K. Enhanced Student Performance Prediction Using Data Augmentation with Ensemble Generative Adversarial Network. Indian J. Sci. Technol. 2024, 17, 4619–4632. [CrossRef]
19. Bansal, V.; Buckchash, H.; Raman, B. Computational Intelligence Enabled Student Performance Estimation in the Age of COVID-19. SN Comput. Sci. 2022, 3, 41. [CrossRef]
20. Yunus, F.A.; Olanrewaju, R.F.; Ajayi, B.A.; Salihu, A.A. Harnessing the Power of a Bidirectional Long Short-Term Memory-Based Prediction Model: A Case of Student Academic Performance. In Proceedings of the 2023 9th International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, Malaysia, 15–16 August 2023; pp. 306–310.
21. Song, W.; Xing, J.; Ning, K.; Guo, W. Research on the Prediction of Academic Performance Based on ANN-BiLSTM Hybrid Neural Network Model. In Proceedings of the 2024 IEEE 4th International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), Chongqing, China, 6–8 December 2024; Volume 4, pp. 1375–1380.
22. Manigandan, E.; Anispremkoilraj, P.; Suresh Kumar, B.; Satre, S.M.; Chauhan, A.; Jeyaganthan, C. An Effective BiLSTM-CRF Based Approach to Predict Student Achievement: An Experimental Evaluation. In Proceedings of the 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Tirunelveli, India, 4–6 January 2024; pp. 779–784.
23. Yılmaz, N.; Sekeroglu, B. Student performance classification using artificial intelligence techniques. In Proceedings of the International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions, Prague, Czech Republic, 27–28 August 2019; Springer International Publishing: Cham, Switzerland, 2019; pp. 596–603.
24. Amrieh, E.A.; Hamtini, T.; Aljarah, I. Mining educational data to predict student's academic performance using ensemble methods. Int. J. Database Theory Appl. 2016, 9, 119–136. [CrossRef]
25. Martins, M.V.; Tolledo, D.; Machado, J.; Baptista, L.M.T.; Realinho, V. Early prediction of student's performance in higher education: A case study. In Proceedings of the Trends and Applications in Information Systems and Technologies: Volume 1, Azores, Portugal, 30 March–2 April 2021; Springer International Publishing: Cham, Switzerland, 2021; pp. 166–175.