Article

Hybrid Deep Learning Models for Predicting Student Academic Performance

Kuburat Oyeranti Adefemi *, Murimo Bethel Mutanga ![ORCID] **and Vikash Jugoo**

Department of Information and Communication Technology, Mangosuthu University of Technology, Umlazi, Durban 4026, South Africa; mutangamb@mut.ac.za (M.B.M.); vikash@mut.ac.za (V.J.)
* Correspondence: adefemi.kuburat@mut.ac.za or adefemioyeranti@gmail.com

Abstract: Educational data mining (EDM) is instrumental in the early detection of students at risk of academic underperformance, enabling timely and targeted interventions. Given that many undergraduate students face challenges leading to high failure and dropout rates, utilizing EDM to analyze student data becomes crucial. By predicting academic success and identifying at-risk individuals, EDM provides a data-driven approach to enhance student performance. However, accurately predicting student performance is challenging, as it depends on multiple factors, including academic history, behavioral patterns, and health-related metrics. This study aims to bridge this gap by proposing a deep learning model to predict student academic performance with greater accuracy. The approach combines a convolutional neural network (CNN) and a bidirectional gated recurrent unit (BiGRU) network to enhance predictive capabilities. To improve the model's performance, we address key data preprocessing challenges, including handling missing data, addressing class imbalance, and selecting relevant features. Additionally, we incorporate optimization techniques to fine-tune hyperparameters to determine the best model architecture. Using key performance metrics such as accuracy, precision, recall, and F-score, our experimental results show that our proposed model achieves improved prediction accuracy of 97.48%, 90.90%, and 95.97% across the three datasets.

Keywords: artificial intelligence; data mining; educational data mining; machine learning; student academic prediction

1. Introduction

Education serves as the cornerstone of national development and personal success, driving institutions to continuously enhance the quality of the learning experience [1]. Student academic performance is a key indicator of an institution's success [2]. However, many students struggle with high failure and dropout rates, which pose significant challenges for educators and administrators. In response to these issues, educational data mining (EDM) has emerged as a powerful approach for analyzing student data and predicting academic outcomes [3,4]. Through the application of data mining techniques, statistical models, and machine learning algorithms, EDM enables institutions to identify at-risk students, optimize educational resources, and implement targeted interventions to support student success [5–7]. Despite these advancements, accurately predicting academic performance remains a challenging task due to factors such as course difficulty, grading policies, and individual learning behaviors [8,9]. Various data-driven approaches have been explored to enhance student performance prediction, including traditional machine learning [4,10–20]. In recent years, the advent of deep learning (DL) has significantly transformed the field of

EDM by enabling the automatic extraction of complex, high-level features directly from raw educational data [21–23]. Convolutional neural networks (CNNs), initially developed for image recognition tasks, have recently demonstrated strong performance in analyzing structured educational data due to their ability to capture localized feature interactions [24,25]. Similarly, recurrent neural networks (RNNs), particularly gated recurrent units (GRUs) and their bidirectional variants (BiGRUs), have proven effective in modeling temporal dependencies in sequential data, including student academic histories and learning patterns [26,27]. Compared to more complex architectures like BiLSTMs, BiGRUs often achieve similar predictive performance while using fewer parameters and require less training time. This makes them especially suitable for medium-sized educational datasets with limited computational resources [28]. Building on these strengths, hybrid deep learning models have gained increasing attention for their ability to integrate complementary architectures. Specifically, combining CNNs for feature extraction with RNNs for temporal modeling has shown promise in capturing both spatial and sequential patterns in complex multidimensional datasets [27,28]. Such hybrid approaches are particularly relevant in educational contexts, where student data often exhibit structured relationships and time-dependent progressions. The accurate prediction of academic performance enables timely interventions and the provision of personalized support, which are essential for improving student outcomes. However, despite the potential of AI-driven models, several limitations continue to hinder their widespread application in academic settings. Traditional machine learning models often rely on manual feature engineering, which can limit their ability to generalize effectively to unseen data. In contrast, deep neural networks offer the advantage of automatic feature extraction, enabling them to learn complex patterns directly from raw data. However, these models typically require large and diverse datasets to achieve reliable and unbiased performance. Unfortunately, such datasets are not always readily available in educational research, especially within smaller institutions or for course-specific analyses. Furthermore, educational datasets frequently suffer from challenges such as missing values and class imbalance, both of which can negatively impact the predictive accuracy and robustness of the learning models. To address these challenges, this study proposes a novel hybrid deep learning model that combines the strength of CNNs with the temporal modeling capabilities of BiGRUs for efficient student academic prediction. The key contributions of this research are as follows:

- We developed a deep learning model that combines a convolutional neural network (CNN) and a recurrent neural network (RNN). This method uses the strengths of both neural networks to improve prediction accuracy.
- Our work demonstrates the effective hybridization of a convolutional neural network with the bidirectional gated recurrent unit (BiGRU) to improve model performance. This unique hybridization is specifically designed to improve the accuracy of student academic prediction.
- We improved academic dataset quality by handling missing values, using the synthetic minority over-sampling technique (SMOTE) to handle class imbalance, selecting useful features to increase model accuracy, and also incorporating advanced regularization techniques.
- We conducted experiments with other baseline models to validate our model's superiority and compare it with other high-rated models in the literature.

The rest of this paper is structured as follows: Section 2 presents the related works. Section 3 presents the methodology. Section 4 presents the results and discussion. Section 5 concludes the study.

2. Related Works

In recent years, the education sector has experienced a significant increase in the application of data mining techniques, particularly within the context of EDM. One of the notable areas of research within the EDM is the analysis and prediction of student academic performance. Several models have been proposed in the literature to predict academic success. The work of Leelaluk et al. [29] is notable. The authors proposed an Attention-Based Artificial Neural Network (Attn-ANN) to identify at-risk students. They evaluated their model on a university dataset and reported promising results, achieving 89.5% accuracy, an F-score of 83.4%, and an AUC of 92.8%. However, their study was limited by the small and imbalanced dataset, which could affect the model's generalizability. Hussain et al. [30] proposed the Levenberg–Marquardt Algorithm to predict student performance. The proposed model achieved an accuracy of 88.6%. However, the issue of class imbalance remained unaddressed. Nabil et al. [21] presented a deep artificial neural network integrated with several resampling approaches, such as SMOTE, ROS, ADASYN, and SMOTE-ENN, to predict students' academic performance. The data obtained from a public university was used to evaluate the models. The model was compared with various ML methods, including DT, RF, gradient boosting (GB), LR, SVM, and KNN algorithms. The results revealed that DNN outperformed other algorithms, with an accuracy of 89%, an F1-score of 89%, and a sensitivity of 89%. Ahmed [2] proposed a machine learning model to predict student outcomes in higher education. The authors combined SVM, DT, and NB with K-means clustering that uses Davies' Bouldin method to identify key features affecting student performance. The proposed model achieved 96% accuracy for SVM, 93.4% for DT, and 83.3% for Naive Bayes. Roy and Farid [8] proposed a new adaptive feature selection algorithm (AFSA) for predicting student performance. They used four different academic datasets to evaluate the performance of LR, KNN, SVM, NB, and DT. The DT algorithm achieved an accuracy of 75%. Kala et al. [31] proposed a hybrid method that combines particle swarm optimization (PSO) and deep neural networks for predicting student academic performance. They used the widely known XAPI-academic dataset, which covers various courses. The performance of the method was compared with various machine learning algorithms, including LR, DT, SVM, ANN, RF, and KNN, as well as a standalone DNN. According to the author, the proposed model achieved a 63% accuracy rate. However, the proposed model's accuracy is lower but performs 6% better than the traditional algorithms. Zhang et al. [32] proposed a hybrid model that combined image convolutional and bi-directional temporal convolutional network (IC-BTCN) to predict student dropout in massive open online courses (MOOCs). The proposed model achieved an accuracy of 89.3%, a precision of 96.5%, a recall of 90.5%, and an F1 score of 93.4%. Kukkar et al. [33] proposed an RNN and LSTM model to predict student performance. The proposed approach was evaluated using the OULAD dataset. The authors also combined their proposed models with RF, SVM, NB, and DT. The proposed model RNN + LSTM + RF algorithm obtained 97% accuracy, outperforming RNN + LSTM + SVM, RNN + LSTM + NB, and RNN + LSTM + DT models, which achieved 90.67%, 86.45%, and 84.42%, respectively.

Chui et al. [34] developed a hybrid model, ICGAN-DSVM, which combines an improved conditional generative adversarial network with a deep support vector machine for student academic prediction. The model achieved 0.968 specificity, 0.971 sensitivity, and an AUC of 0.954. Similarly, Venkatachalam and Sivanraju [35] proposed an ensemble generative adversarial network (EGAN), which blends Divergence GAN (DivGAN) and Success-aware GAN (SucGAN) to enhance student academic prediction accuracy. Their model achieved 94.71% accuracy and an RMSE of 0.0529, outperforming several baseline deep learning models.

Bansal et al. [36] proposed a student performance prediction using LSTM, GRU, and latent space models (such as variational autoencoders) combined with traditional machine learning to find hidden patterns in student data. It was reported that variational autoencoders combined with MLP achieved the highest R^2 score of 0.867.

Yunus et al. [37] proposed a predictive framework that combines improved feature selection techniques with a BiLSTM model to predict students' academic performance. Their model achieved an accuracy of 90.16%, with precision, recall, and F-score values of 86.16%, 90%, and 90%, respectively. Song et al. [38] proposed an ANN-BiLSTM approach that integrates artificial neural networks for feature extraction with BiLSTM for sequential learning for student academic performance prediction. Evaluated using the OULAD dataset, this model outperformed several baselines, including RNN, GRU, and ANN-LSTM, achieving an accuracy of 73%, a precision of 79%, a recall of 56%, and an F-score of 65%. Manigandan et al. [39] developed a BiLSTM-CRF hybrid model to predict student success based on historical academic performance. The model achieved 90.6% accuracy, 92.7% precision, 89.7% recall, and 92.4% specificity. Finally, Yousafzai et al. [22] incorporate an attention mechanism into a BiLSTM framework. Their attention-BiLSTM model achieved a predictive performance with an accuracy of 90.16% and precision, recall, and F-score of 0.90.

Existing studies have employed various traditional machine learning (ML) and deep learning (DL) techniques for predicting student academic performance. Traditional ML models such as SVM, ANN, and extra trees have demonstrated reasonable accuracy; however, they often struggle to capture the long-range and temporal dependencies within student data. This limitation can lead to suboptimal performance, particularly when modeling complex educational patterns over time. To address sequential learning, RNNs have been employed in various studies. While RNNs can process time-series data, they are prone to the vanishing gradient problem, which restricts their ability to retain long-term dependencies. LSTM networks improve this issue by incorporating gating mechanisms that preserve relevant information over longer sequences. Nevertheless, LSTM models operate in a unidirectional manner, potentially overlooking future context that is often critical in educational prediction tasks.

Bidirectional LSTM (BiLSTM) networks extend the LSTM architecture by capturing both forward and backward dependencies and have been used effectively in prior research for tasks such as predicting dropout, student success, and behavioral trends. However, BiLSTM models often require significant computational resources and longer training times due to their dual-gated structure, which can limit their scalability and practicality, especially for medium-sized datasets or time-constrained applications. To overcome these limitations, we propose a hybrid model that integrates convolutional neural networks (CNNs) with bidirectional gated recurrent units (BiGRUs). CNNs are responsible for extracting localized and hierarchical feature representations from the input data, while the BiGRU models temporal dependencies in both directions, similar to BiLSTM but with fewer parameters and reduced training complexity. This results in faster convergence and improves computational efficiency.

3. Materials and Methods

This section presents the methodology employed in this study, detailing the key steps involved. It includes a description of the datasets, data preprocessing approach, the proposed method, experimental setup, baseline models, and performance metrics used to assess the model's effectiveness. A summary of the study's framework is presented in Figure 1.
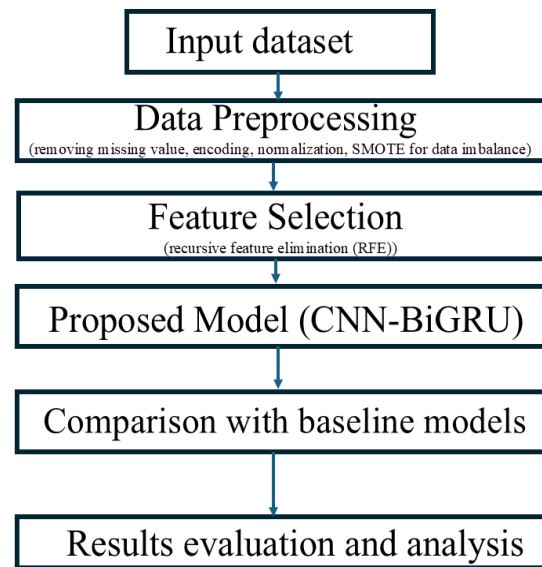
Figure 1. Flowchart of the proposed CNN-BIGRU model.

3.1. Datasets Description

The publicly available academic dataset from the UCI and Kaggle repositories was used to predict student performance. The dataset was selected because it offers a comprehensive collection of student academic records, including grades, attendance, and study behavior. These datasets are suitable for deep learning approaches, making the trained model more generalizable to real-world applications.

3.1.1. HESP Dataset

The higher education student performance dataset (HESP) from UC Irvine was collected from students in the faculty of engineering and educational sciences in 2019. The dataset consists of 145 instances and 31 attributes that cover demographic, academic, and behavioral features [40].

3.1.2. XAPI Dataset

The XAP -educational mining dataset from Kaggle has 480 student records and 16 features. The features are divided into three categories: demographic, academic, and behavioral. The dataset was collected over two semesters in 2016, with 245 records from the first semester and 235 from the second semester [41].

3.1.3. HEI Dataset

This higher education institution dataset contains information about students enrolled in various undergraduate degrees, such as their demographics, socioeconomic characteristics, as well as their academic success at the end of the two semesters. The dataset consists of 4424 instances with 36 attributes [42]. Table 1 summarizes the dataset used.

Table 1. Summary of the dataset used.

Dataset	Year	Source	Attributes	Instances	Features
HESP [40]	2019	UCI	31	145	Demographic, Academic, and Behavior
XAPI [41]	2016	Kaggle	16	480	Demographic, Academic, and Behavior
HEI [42]	2021	UCI	36	4424	Academic Path, Demographic and Social Economic Factors

3.2. Data Preprocessing

Data preprocessing helps to improve the quality of datasets for accurate prediction. The steps are briefly discussed below.

3.2.1. Missing Data

Real-world datasets in academia frequently encounter the issue of missing data, which could be due to typographical errors. In this study, we analyzed the dataset to find missing data and addressed this issue by eliminating entire records. Deleting the features improves the model's confidentiality, and it ensures that only complete and reliable data are used for training.

3.2.2. Data Encoding

Encoding converts non-numerical data to numerical data. In this study, we used a label encoder to convert the dataset's categorical variables to numerical variables, allowing the data to be used for algorithm training.

3.2.3. Normalization of Data

Normalization standardizes or scales data to a common range or format. Data normalizations help to prevent numerical instability and remove bias towards variables with higher values. In this study, Min-Max normalization techniques were employed to scale the features to a range between 0 and 1. Normalization can be expressed mathematically as Equation (1):

$$x' = \frac{x - min\,(x)}{max\,(x) - min\,(x)} \tag{1}$$

where x is the original value, x' is the normalized value, $min\,(x)$ is the minimum value in the dataset, and $max\,(x)$ is the maximum value of the feature.

3.2.4. Data Imbalance

Data imbalance is an important phase in data preprocessing as it can have a major impact on the model's performance. In this study, we employed the synthetic minority oversampling technique (SMOTE) approach to address the issue of data imbalance. It works by picking samples from the minority class, finding their nearest neighbors, and creating new similar samples.

3.2.5. Feature Selection

The feature selection is an important stage where the most relevant features from the given dataset are selected. The recursive feature elimination (RFE) technique was employed in this study. This technique starts with all features and gradually eliminates the ones that have the least impact on the performance of the model. This approach reduces the complexity of the model, enhances performance, and reduces the risk of overfitting. More information on RFE can be found in [8].

3.2.6. Data Splitting

We partition the dataset into training and testing, using an 80:20 ratio. The training set was specifically utilized for training and hyperparameter tuning. The test set evaluates the model's performance and generalizability.

3.3. Deep Learning Models Description

3.3.1. Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) are deep learning models used for extracting meaningful features through convolutional operations, making them effective for process-

ing and analyzing high-dimensional data. CNN consists of multiple layers, including a convolutional layer, which uses convolutional filters known as kernels to generate feature maps that capture spatial hierarchies and local patterns within the dataset [28,43]. The pooling layer, which reduces the dimensionality of these feature maps using operations such as max or average pooling, thereby simplifying feature representation and optimizing computational efficiency [43]. Convolutional and pooling layers often use the ReLU activation function to classify inputs efficiently. The CNN process involves a fully connected layer that uses a Softmax activation function to integrate extracted features and make predictions. This activation function is typically applied in the final layer for classification tasks [43]. The process can be mathematically expressed in Equation (2)–(4):

$$f(t) = (x * k)(t) = \sum_{a=-\infty}^{\infty} x(a) \cdot k(t - a) \tag{2}$$

$$f(x) \max (0, x) \tag{3}$$

$$p(y = j | z) = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}} \tag{4}$$

where $x * k$ is the convolution operation, $x(a)$ is the input data, $k(t - a)$ is the kernel function sliding over the input data at each position t, (j | z) is the input to the Softmax for class j, and k is the number of classes.

3.3.2. Bidirectional Gated Recurrent Units (BiGRUs)

The gated recurrent units (GRUs) is a recurrent neural network with two gates: a reset and update gate [44]. The update gate helps to determine the amount of previous data that needs to be transmitted in the future. It also determines which data needs to be discarded or used [45]. A reset gate determines the amount of previous knowledge to be forgotten. GRUs are slightly faster to train due to their fewer tensor operations [Dey, R]. BiGRU is an extension of GRU; it combines two GRUs, which process sequences in both forward and backward directions. BiGRU captures information from both past and future time steps, providing a more comprehensive understanding of the sequence. The outputs from the forward and backward GRUs are then concatenated to create a bidirectional contextual representation of the input sequence. BiGRU improves model learning, faster and less computationally expensive. The operation of the BiGRU can be mathematically expressed as follows in Equation (5)–(8):

$$u_t = \sigma(W_u * [h_{t-1}, x_t] + b_u) \tag{5}$$

$$r_t = \sigma(W_r * [h_{t-1}, x_t] + b_r) \tag{6}$$

$$\widetilde{ht} = \tanh (W_h * [r_t * h_{t-1}, x_t] + b \tag{7}$$

$$ht = (1 - z_t) * h_{t-1} + z_t * \widetilde{ht} \tag{8}$$

where u_t represents the update gate, r_t denotes the reset gate, ht indicates the hidden state, x_t is the input at time step t, $\widetilde{ht}$ signifies candidate state at time t, W_u, W_r, and W_h are the weight matrices, σ represents sigmoid function, tanh serves as an activation function, and b_u and b_r are associated biases.

3.4. Baseline Methods

This subsection presents the deep learning techniques used as baseline models to compare the performance of our proposed model. The deep learning algorithms were selected

because of these reasons: automatic feature extraction, sequential dependencies, computational efficiency, and so on. These techniques were briefly discussed in the subsection:

3.4.1. Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) are designed to mimic human brain behavior. The ANN architecture consists of an input, hidden, and output layer. These layers perform various transformations on inputs, producing outputs based on the training process. The training process involves repeatedly feeding input data to the neural network, fine-tuning weights and bias, and using mathematical functions to map the input dataset.

3.4.2. Long Short-Term Memory (LSTM) Network

The long short-term memory (LSTM) network is a fully connected neural network that controls information flow between layers using a gating mechanism composed of an input gate, output gate, forget gate, and memory cell [46]. Bidirectional long short-term memory (BiLSTM) is an extension of the LSTM. It consists of two LSTM layers, one that processes data forward from past to future and the other that processes data backward from future to past.

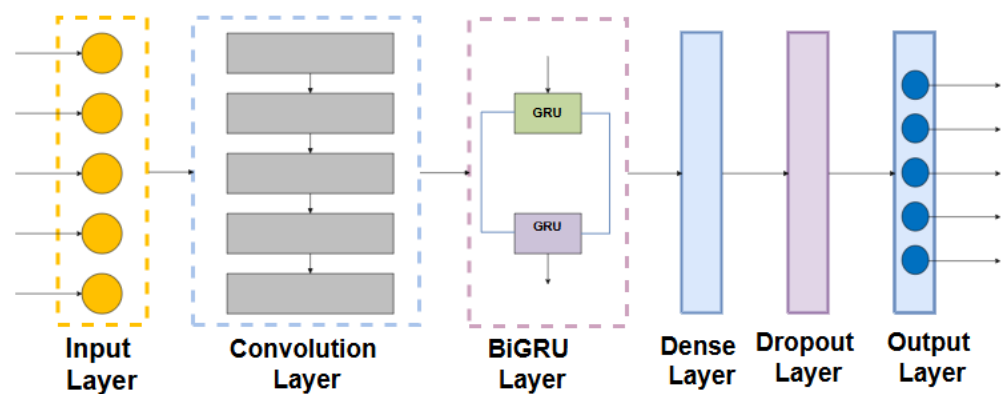### 3.5. Proposed CNN-BiGRU Model Architecture

This study proposed the CNN-BiGRU model, which combines the strengths of convolutional neural networks and bidirectional gated recurrent units to address the limitations of existing models in predicting student academic success. The primary objective of this study is to develop a binary classification model using deep learning techniques to identify at-risk students with greater accuracy, by outputting probabilities that indicate whether a student is at risk or not at risk. The model aims to support early intervention efforts as it allows educators and institutions to implement support strategies, potentially preventing academic failure. The CNN-BiGRU model consists of an input layer, a CNN layer, a BiGRU layer, fully connected (dense) layers, a dropout layer, and an output layer, as shown in Figure 2. The input layer takes in the preprocessed datasets and serves as the entry point to the predictive model. It ensures that the data is formatted appropriately for feature extraction and learning. The convolutional layer extracts high-level features from the input data. It consists of a convolutional layer, a ReLU activation function, and a max pooling layer as discussed in the previous section. This layer identifies important patterns before passing them to the recurrent layer. After the CNN extracts relevant features, the BiGRU layer processes the feature sequences in both forward and backward directions. This dual-layered approach helps us to capture both past and future dependencies, making the model more effective in handling sequential relationships. The fully connected layer transforms CNN-BiGRU features into a meaningful dense representation that can be used for classification. To prevent overfitting, we employed a dropout layer. It randomly deactivates a fraction of neurons, forcing the remaining ones to adapt and generalize better to unseen data during data training. The output layer is the final decision-making layer that uses the sigmoid activation function to output the probability score representing the likelihood of a student achieving academic performance between 0 and 1. This layer is responsible for classifying students as at risk or not at risk.

3.6. Experimental Setup

The experiments were conducted in a Python (version 3.11.9) environment on a Windows 11 PC with an Intel (R) Core TM i7 1135G7 processor running at 2.40 GHz and 8 GB of RAM. We trained the proposed model on the training data. For model optimization, we employed the Adam optimizer, and a binary cross-entropy was used as the loss function. We employed a grid search approach to carefully select the learning rate. The model was

trained for 50 epochs and a batch size of 64. To improve model performance, we carried out extensive hyperparameter tuning for each model. The best hyperparameter value for the ANN is 64 neurons and a learning rate of 0.001, after we tried with 32, 64, and 96 neurons and learning rates of 0.001 and 0.0001. Similarly, with CNN, we increase the number of filters to 32, 64, and 96, with kernel sizes of 3, and learning rates of 0.01 and 0.001. The most effective CNN is 64 filters, a kernel size of 3, and a learning rate of 0.01. For sequential models, LSTM, BiLSTM, and GRU, we tuned the number of units to 16, 32, 64, and 96 and learning rates to 0.01 and 0.001. The ideal hyperparameters were as follows: 16 LSTM units with a learning rate of 0.001, 32 BiLSTM units with a learning rate of 0.001, and 16 GRUs with a learning rate of 0.001. Similarly, the CNN-BiGRU model achieved the best performance with 32 BiGRUs and a learning rate of 0.001. Table 2 summarized the model settings (parameters and hyperparameters) used. We evaluated the performance of our models using standard performance metrics to assess their real-world effectiveness.

Figure 2. Architecture of the proposed CNN-BiGRU model.

Table 2. CNN-BiGRU model parameters and hyperparameters.

Parameters	Configuration/Value
Learning rate	0.001
Number of epochs	50
Batch sizes	64
Activation function	ReLU and Softmax
Loss function	Categorical cross-entropy
Optimization algorithm	Adam optimizer
Hyperparameter optimization	Grid search
Regularization techniques	Dropout technique
Dropout rate	0.5

3.7. Performance Evaluation

To evaluate the proposed CNN-BiGRU and the baseline model performance, metrics such as accuracy, precision, sensitivity, and F-score are employed, which are expressed mathematically in Equations (9) to (12), respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Precision = \frac{TP}{FP + TP} \tag{10}$$

$$Recall = \frac{TP}{FN + TP} \tag{11}$$

$$F - score = 2 \times \frac{(precision \times recall)}{(precision + recall)} \tag{12}$$

where True positives (TP) represent students correctly predicted as good; False Positives (FP) refer to students incorrectly predicted as good but who actually belong to the weak category; False negatives (FN) predict students who are actually good but were predicted as weak.

## 4. Results and Discussion

This section covers our results and discussion from various experiments.

4.1. Proposed CNN-BiGRU Model Results

The performance of the proposed CNN-BiGRU model was evaluated using three datasets: HESP, XAPI, and HEI. Table 3 and Figure 3 summarize the results on test data, comparing the model's effectiveness across the datasets. Using the HESP dataset, the proposed model achieved an accuracy of 90.90%, a precision of 90.77%, a 90.76% recall value, and an F-score of 90.76%. Using the XAPI dataset, the proposed model yielded an accuracy of 95.97%, with a precision of 95%, a recall of 95%, and an F-score of 95%. Similarly, on the HEI dataset, the proposed model achieved the highest accuracy of 97.48%, precision of 97.12%, 96.95% recall value, and F-score of 97.03%. As shown in the figure, it can be seen that the proposed model performed best on the HEI dataset. This could be attributed to the size of the dataset and diverse attributes, which appear to improve generalization. These results indicate that our proposed model is highly effective in predicting student academic performance, consistently achieving over 95% across all metrics.

Table 3. Results of the proposed model and baseline models.

Dataset	Model	Accuracy	Precision	Recall	F-Score
HEI	CNN	90.49	90.77	90.70	90.73
	GRU	92.97	92.59	92.53	92.55
	ANN	86.61	86.52	86.31	86.41
	LSTM	92.36	92.47	92.44	92.45
	BiLSTM	96.24	96.19	96.19	96.19
	CNN-BiGRU	97.48	97.12	96.95	97.03
HESP	CNN	81.67	81.79	81.54	81.66
	GRU	84.74	84.70	84.46	84.57
	ANN	80.79	80.72	80.26	80.48
	LSTM	84.12	84.06	84.09	84.07
	BiLSTM	85.28	84.63	84.52	84.57
	CNN-BiGRU	90.90	90.77	90.76	90.76
XAPI	CNN	90.92	90.15	90.43	90.28
	GRU	93.98	92.71	92.18	92.44
	ANN	81.25	81.09	80.05	80.59
	LSTM	93.80	92.07	92.07	92.07
	BiLSTM	95.00	94.91	94.69	94.68
	CNN-BiGRU	95.97	95.00	95.00	95.00
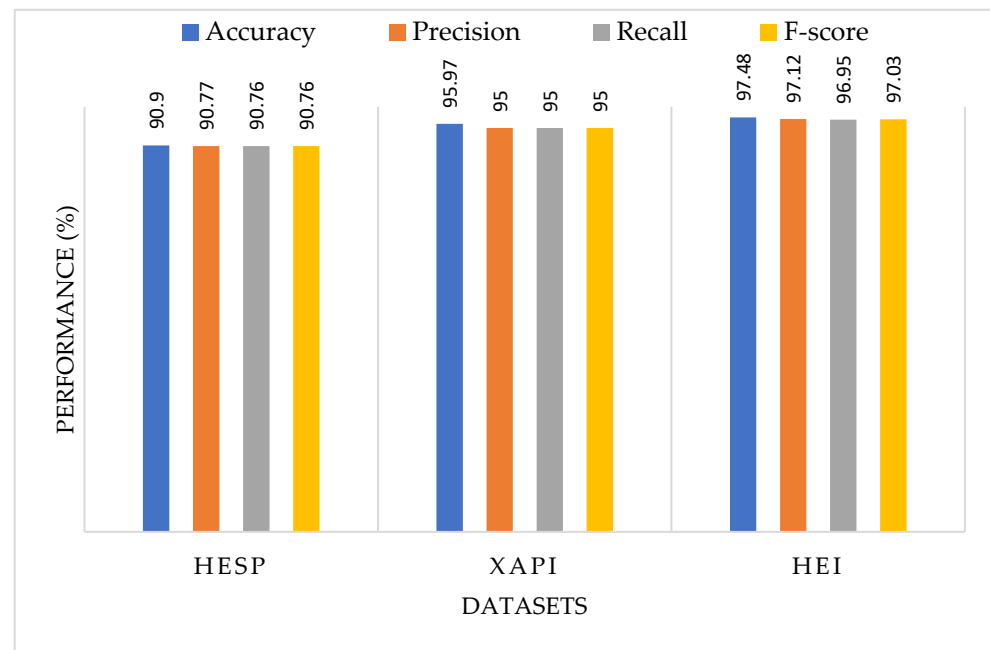
Figure 3. CNN-BiGRU model results across all the datasets.

4.2. Performance Comparison of the Proposed Model with the Baseline Models

The comparison of the proposed CNN-BiGRU model and baseline models (CNN, GRU, ANN, LSTM, and BiLSTM) is presented in Table 3. Using the HEI dataset, as shown in Figure 4, the CNN-BiGRU model achieved the highest performance across all metrics with 97.48% accuracy, 97.12% precision, 96.95% recall, and 97.03% F-score. The second best model is the BiLSTM model, which achieved 96.24% accuracy, 96.19% precision, 96.19% recall, and 96.10% F-score, slightly lower than the proposed model. The GRU model achieved an accuracy of 92.97%, 92.59% precision, 92.55% recall, and 92.55% F-score, and LSTM achieved 92.36% accuracy, 92.47% precision, 92.44% recall, and 92.45% F-score. It can be seen that these models also performed well, but their slightly lower F-scores suggest they may not generalize as effectively as CNN-BiGRU. Compared to the standalone CNN, the sequential models outperformed the CNN method by 2 to 6% accuracy and approximately 6% across precision, recall, and F-score. ANN performance was very low. It is evident that incorporating both convolutional layers and bidirectional gated recurrent units (BiGRUs) enhances the ability to capture spatial and sequential dependencies in student performance data.

Using the HESP dataset, Table 3 and Figure 5 illustrate the performance of ANN, CNN, GRU, LSTM, and BiLSTM models when modeled individually, as well as the results from the proposed CNN-BiGRU model. The CNN-BiGRU model demonstrated superior performance with 90.90% accuracy, 90.77% precision, 90.76% recall, and an F-score of 90.76%. Furthermore, the BiLSTM model attained an accuracy of 85.28%, which is the second best performing model. The GRU and LSTM also achieved a better performance, achieving an accuracy of 84.74% and 84.12%, respectively, compared to CNN and ANN, which yielded an accuracy of 81.67% and 80.79%, respectively, also in terms of precision, recall, and F-score.

Using the XAPI dataset, as shown in Figure 6 and Table 3, CNN-BiGRU again outperformed all models, achieving 95.97% accuracy, 95.00% precision, 95.00% recall, and an F-score of 95.00%. BiLSTM attained 95% accuracy, 94.91% precision, 94.69% recall, and 94.68% F-score. The GRU model achieved 93.98% accuracy, 92.71% precision, 92.18% recall, and 92.44% F-score. LSTM model yielded an accuracy of 93.80%, precision of 92.07%, recall

value of 92.07%, and F-score of 92.07%. CNN achieved an accuracy of 90.92%, 90.15% precision, 90.43% recall, and 90.28% F-score. The marginal increase in accuracy and F-score suggests that CNN-BiGRU effectively balances precision and recall, reducing both false positives and false negatives. The ANN model had the lowest performance across all the metrics. The model achieved an accuracy of 81.25%, 81.09% precision, 80.05% recall, and 80.59% F-score.
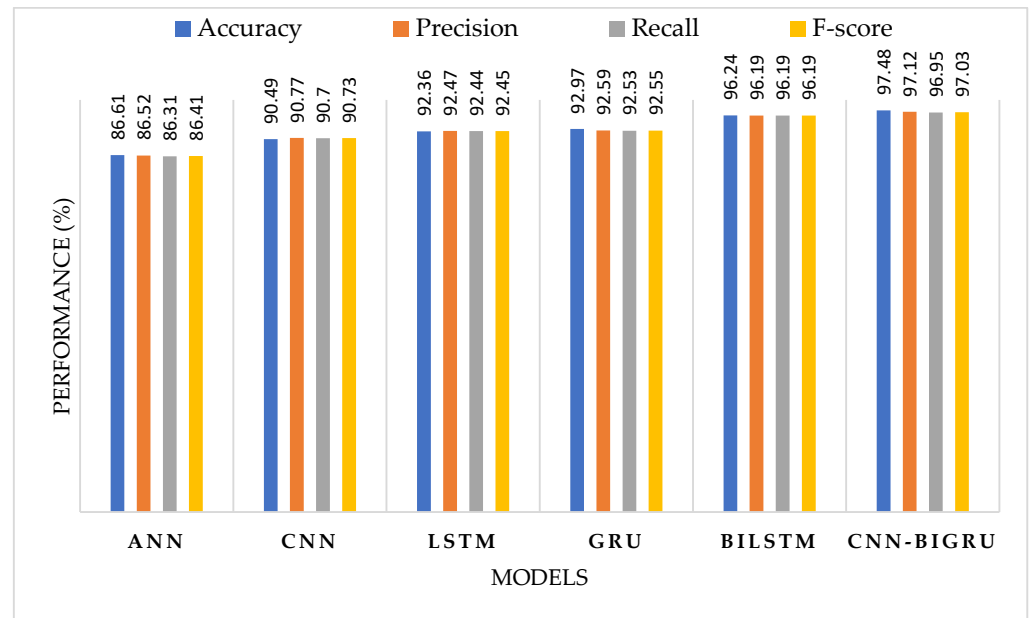

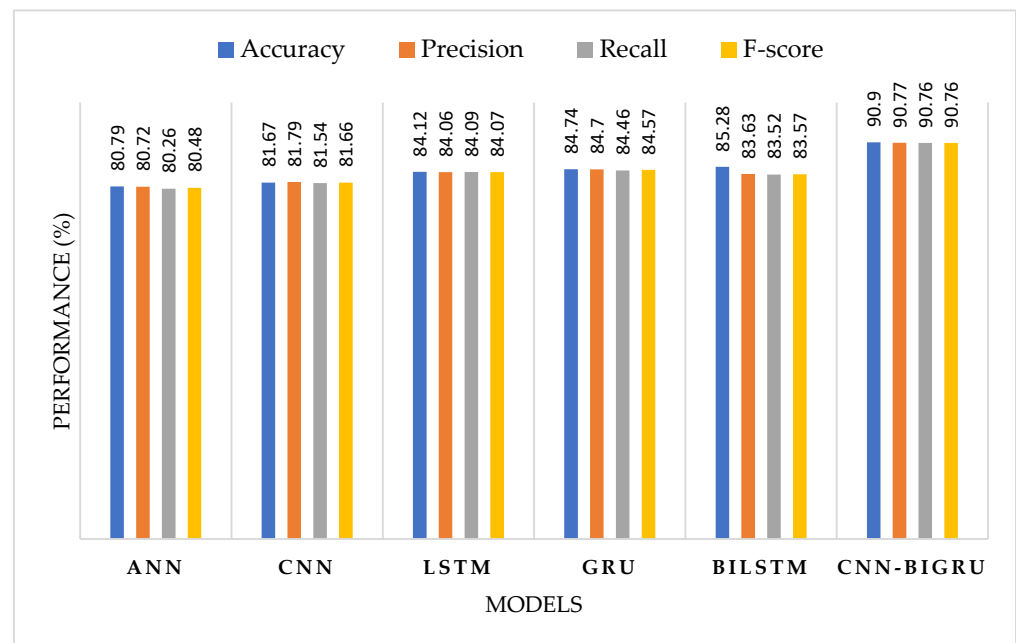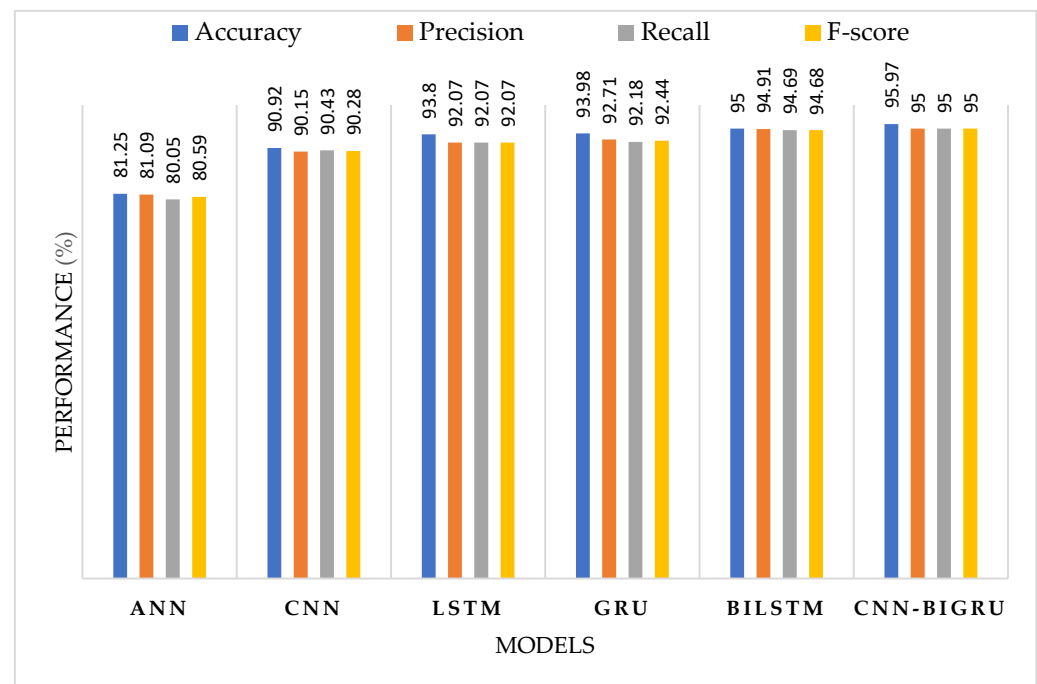**Figure 4.** CNN-BiGRU model result using the HEI dataset.

Figure 5. CNN-BiGRU model results using HES dataset.

It is evident that the CNN-BiGRU model consistently outperformed other models across all datasets, achieving the highest accuracy, precision, recall, and F-scores. Notably, CNN-BiGRU maintained high recall values, ensuring minimal false negatives, an essential factor when predicting students at risk of poor performance. The superior performance of CNN-BiGRU can be attributed to the complementary strengths of its components. The

CNN layer extracts local and spatial features (e.g., patterns in assessment scores or learning behaviors), which are particularly useful in educational data where certain feature interactions may indicate performance trends. These features are then passed to the BiGRU layer, which effectively captures sequential dependencies in both forward and backward directions, enhancing temporal understanding. This bidirectional processing likely contributed to more accurate predictions by considering both past and future context within the input sequence. While GRU and LSTM models showed similar performance, BiLSTM generally performed slightly better due to its bidirectional nature, which, as with BiGRU, helps capture full context in student progression data. The sequential models (LSTM, GRU, BiLSTM, and BiGRU) consistently outperformed CNN and ANN standalone models, suggesting that temporal dependencies are critical in modeling student academic behavior. ANN recorded the lowest performance across datasets. This result is expected since simple feedforward networks lack memory mechanisms to model time-dependent patterns, making them less suitable for educational datasets with sequential or time-sensitive inputs. In terms of dataset size, the HEI dataset (4424 instances) yielded the best overall results with an accuracy of 97.48%. This suggests that larger datasets enable better generalization and model training.

Figure 6. CNN-BiGRU model results using the XAPI dataset.

4.3. Proposed Models Training and Prediction Time

This section presents the computational efficiency of the proposed CNN-BiGRU model compared to the baseline models. As shown in Table 4, the CNN-BiGRU model demonstrated faster prediction time, achieving 0.06 s, outperforming ANN and CNN which achieved 0.08 s, as well as LSTM, GRU, and BiLSTM (0.09, 0.08, and 0.10 s). ANN and CNN have the fastest training times, LSTM takes longer due to its sequential processing ability, GRU is slightly faster than LSTM due to its simpler architecture, and BiLSTM exhibits the longest training time because of its bidirectional nature. CNN-BiGRU demonstrates a moderate training time with the fastest prediction time, making it suitable for real-time applications. The efficiency of CNN-BiGRU can be attributed to the combination of convolutional and recurrent layers. This combination allows CNN-BiGRU to process student

data more efficiently. The reported 0.06 s prediction time suggests that CNN-BiGRU can be effective for real-time identifying at-risk students.

Table 4. Model training and prediction time.

Model	Training Time (sec)	Prediction Time (sec)
ANN	0.11	0.08
CNN	0.12	0.08
LSTM	0.17	0.09
GRU	0.13	0.08
BiLSTM	0.19	0.10
CNN-BiGRU	0.16	0.06

4.4. Proposed Model Decision Interpretation

To interpret the prediction of the proposed method in this study, we employed a paired T-test to evaluate the effectiveness of the proposed hybrid CNN-BiGRU model compared to several baseline models (CNN, GRU, ANN, LSTM, and BiLSTM) on three datasets: HEI, HESP, and XAPI, as shown in Table 5. The T-test is a statistical method that helps assess whether there are significant differences between the models' performances, making it ideal for understanding the impact of our proposed model over others. The T-test results for the HEI dataset revealed a statistically significant improvement in the performance of the Hybrid CNN-BiGRU model compared to the baseline models (T-statistic = 3.63, p = 0.0221). This indicates that the proposed model outperforms the baseline models with high confidence. For the HESP dataset, the results were even more pronounced, with a highly significant improvement in the CNN-BiGRU model's performance over the baseline models (T-statistic = 8.58, p = 0.0010). The p-value is well below the 0.05 threshold, suggesting a strong and statistically meaningful difference in the model's ability to predict student performance. The XAPI dataset also presented a statistically significant difference in the performance of the CNN-BiGRU model compared to the baseline models (T-statistic = 3.38, p = 0.01).

Table 5. Paired T-test interpretations.

Dataset	T-Statistic	p-Value	Significance
HEI	3.63	0.0221	Significant ($p < 0.05$)
HESP	8.58	0.0010	Highly Significant ($p < 0.05$)
XAPI	3.38	0.01	Significant ($p < 0.05$)

Although the baseline models, particularly the BiLSTM, also showed a strong performance but the results indicate that the hybrid CNN-BiGRU model performs exceptionally well, where it demonstrates a statistically significant improvement over baseline models. This emphasizes the effectiveness of combining CNNs and BiGRU for tasks involving complex, sequential data.

4.5. Performance Comparison with Previous Studies

In Table 6, our proposed hybrid CNN-BiGRU model demonstrates significant improvements over previous approaches across various datasets. Leelaluk et al. [29] achieved 89.5% accuracy and an F-score of 83.4%; our model attains 97.48%, marking an improvement of 7.9%. Compared to Hussain et al. [30], which achieved an accuracy of 88.6%, our model attains 97.48%, marking an improvement of 8.88%. Similarly, our model outperforms Nabil

et al. [21] by 8.48%, as they reported 89% accuracy. In contrast to Ahmed, whose model achieved 96%, our method yields a modest but meaningful improvement of 1.48%. More improvement is observed from Roy and Farid [8], whose model achieved an accuracy of 75%, leading to a 22.48% increase with our approach. Furthermore, we outperform Zhang et al. [32], which achieved 89.3%, showing a gain of 8.18%. Compared to Kukkar et al. [33], whose model attained 97%, our approach shows a slight 0.48% increase. Beyond accuracy, other comprehensive evaluation metrics such as precision, recall, and F-score also provide deeper insights into our model's predictive capability. Future work could aim to improve performance for more diverse datasets. In summary, our proposed CNN-BiGRU model outperforms previous studies in terms of its performance, offering a more accurate and robust solution for student performance prediction.

Table 6. Performance comparison with previous studies.

Articles	Year	Model	Dataset	Accuracy	Precision	Recall	F-Score
[2]	2024	TMLs	University Data	96.03	94.18	98.43	-
[8]	2024	TMLs	XAPI, SSP, HESP, & Western-OC2-Lab Data	75	76	75	74
[21]	2021	DNN	XAPI & University Data	89	-	89	89
[29]	2024	ATTN-ANN	Kyushu University Data	89.5	-	-	83.4
[30]	2024	LMA	University Data	88.6	96.3	89.6	93.3
[32]	2024	IC-BTCN	KDD Cup 2015	89.3	96.5	90.5	93.4
[33]	2024	RNN & LSTM	OULAD	96.78	90.86	95.00	92.89
Our Work		**CNN-BiGRU**		**97.48**	**97.12**	**96.95**	**97.03**

5. Conclusions

In this study, we proposed a hybrid deep learning model that combined the convolution neural network and recurrent neural network (CNN-BiGRU) model for predicting student academic performance by identifying at-risk students. To improve model accuracy, we addressed key data preprocessing challenges, including handling missing values, mitigating class imbalance, and selecting relevant features. Our model was evaluated using accuracy, precision, recall, and F-score. Specifically, our CNN-BiLSTM model achieved an accuracy of 97.48%, 90.90%, and 95.97% across the three datasets. Moreover, additional experiments were carried out on the datasets to compare the effectiveness of the proposed hybrid model using various deep learning methods as baseline methods. According to experimental results, the proposed model outperformed all the baseline models (ANN, CNN standalone, LSTM, GRU standalone, and BiLSTM). These findings highlight the potential of data-driven models, especially the hybridization of deep learning models in educational data mining. In future work, we aim to explore model interpretability approaches such as SHAP, LIME, attention mechanisms, or other explainable AI methods to better evaluate the effectiveness of the model. Furthermore, we will consider exploring huge datasets to further enhance predictive accuracy.

Author Contributions: Conceptualization, K.O.A. and M.B.M.; methodology, K.O.A.; software, K.O.A.; validation, K.O.A., M.B.M. and V.J.; formal analysis, K.O.A., M.B.M. and V.J.; investigation, K.O.A.; resources, M.B.M.; data curation, K.O.A.; writing—original draft preparation, K.O.A.; writing—review and editing, K.O.A. and M.B.M.; visualization, K.O.A., M.B.M. and V.J.; supervision, M.B.M.; project administration, M.B.M. and V.J.; funding acquisition, M.B.M. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Publicly open-source datasets were analyzed in this study. These data can be accessed at https://www.kaggle.com/datasets (accessed on 16 January 2025) and https://archive.ics.uci.edu/datasets (accessed on 16 January 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AFSA	Adaptive Feature Selection Algorithm
ANN	Artificial Neural Network
Bi-GRU	Bidirectional Gated Recurrent Unit
CNN	Convolutional Neural Network
DT	Decision Tree
DNN	Deep Neural Network
EDM	Educational data mining
GAN	Generative Adversarial Network
GRU	Gated Recurrent Unit
KNN	K-Nearest Neighbour
LR	Logistic Regression
PSO	Particle Swarm Optimization
RF	Random Forest
SSP	Secondary Student Performance
SVM	Support Vector Machine
SMOTE	Synthetic Minority Over-sampling Technique

References

1. Ramaphosa, K.I.M.; Zuva, T.; Kwuimi, R. Educational Data Mining to Improve Learner Performance in Gauteng Primary Schools. In Proceedings of the 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 6–7 August 2018; IEEE: New York, NY, USA, 2018; pp. 1–6.
2. Ahmed, E. Student performance prediction using machine learning algorithms. *Appl. Comput. Intell. Soft Comput.* **2024**, *2024*, 4067721. [CrossRef]
3. Pelima, L.R.; Sukmana, Y.; Rosmansyah, Y. Predicting university student graduation using academic performance and machine learning: A systematic literature review. *IEEE Access* **2024**, *12*, 23451–23465. [CrossRef]
4. Bellaj, M.; Dahmane, A.B.; Boudra, S.; Sefian, M.L. Educational Data Mining: Employing Machine Learning Techniques and Hyperparameter Optimization to Improve Students' Academic Performance. *Int. J. Online Biomed. Eng.* **2024**, *20*, 3. [CrossRef]
5. Pecuchova, J.; Drlik, M. Enhancing the Early Student Dropout Prediction Model Through Clustering Analysis of Students' Digital Traces. *IEEE Access* **2024**, *12*, 159336–159367. [CrossRef]
6. Almaghrabi, H.; Soh, B.; Li, A.; Alsolbi, I. SoK: The Impact of Educational Data Mining on Organisational Administration. *Information* **2024**, *15*, 738. [CrossRef]
7. Kok, C.L.; Ho, C.K.; Chen, L.; Koh, Y.Y.; Tian, B. A Novel Predictive Modeling for Student Attrition Utilizing Machine Learning and Sustainable Big Data Analytics. *Appl. Sci.* **2024**, *14*, 9633. [CrossRef]
8. Roy, K.; Farid, D.M. An adaptive feature selection algorithm for student performance prediction. *IEEE Access* **2024**, *12*, 75577–75598. [CrossRef]
9. Bognár, L. Predicting Student Attrition in University Courses. In *Machine Learning in Educational Sciences: Approaches, Applications and Advances*; Springer Nature: Singapore, 2024; pp. 129–157.
10. Cheng, B.; Liu, Y.; Jia, Y. Evaluation of students' performance during the academic period using the XG-Boost Classifier-Enhanced AEO hybrid model. *Expert Syst. Appl.* **2024**, *238*, 122136. [CrossRef]
11. Yağcı, M. Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learn. Environ.* **2022**, *9*, 11. [CrossRef]
12. Baashar, Y.; Alkawsi, G.; Mustafa, A.; Alkahtani, A.A.; Alsariera, Y.A.; Ali, A.Q.; Hashim, W.; Tiong, S.K. Toward predicting student's academic performance using artificial neural networks (ANNs). *Appl. Sci.* **2022**, *12*, 1289. [CrossRef]

13. Pallathadka, H.; Wenda, A.; Ramirez-Asís, E.; Asís-López, M.; Flores-Albornoz, J.; Phasinam, K. Classification and prediction of student performance data using various machine learning algorithms. *Mater. Today Proc.* **2023**, *80*, 3782–3785. [CrossRef]

14. Monteverde-Suárez, D.; González-Flores, P.; Santos-Solórzano, R.; García-Minjares, M.; Zavala-Sierra, I.; de la Luz, V.L.; Sánchez-Mendiola, M. Predicting students' academic progress and related attributes in first-year medical students: An analysis with artificial neural networks and Naïve Bayes. *BMC Med. Educ.* **2024**, *24*, 74. [CrossRef] [PubMed]

15. Tirumanadham, N.K.M.K.; Thaiyalnayaki, S.; SriRam, M. Evaluating boosting algorithms for academic performance prediction in E-learning environments. In Proceedings of the 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bangalore, India, 24–25 January 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–8.

16. Alsalem, G.M.; Sarhan, N.; Hammad, M.; Zawaideh, B. Predicting Students' Performance using Machine Learning Classifiers. In Proceedings of the 2024 25th International Arab Conference on Information Technology (ACIT), Zarqa, Jordan, 10–12 December 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–5.

17. Rabelo, A.M.; Zárate, L.E. A model for predicting dropout of higher education students. *Data Sci. Manag.* **2025**, *8*, 72–85. [CrossRef]

18. AlShaikh-Hasan, M.; Ghinea, G. Evaluating the Impact of Multi-Layer Data on Machine Learning Classifiers for Predicting Student Academic Performance. In Proceedings of the 2024 25th International Arab Conference on Information Technology (ACIT), Zarqa, Jordan, 10–12 December 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.

19. Chen, M.; Liu, Z. Predicting performance of students by optimizing tree components of random forest using genetic algorithm. *Heliyon* **2024**, *10*, e12262. [CrossRef]

20. Adu-Twum, H.T.; Sarfo, E.A.; Nartey, E.; Adesola Adetunji, A.; Ayannusi, A.O.; Walugembe, T.A. Role of Advanced Data Analytics in Higher Education: Using Machine Learning Models to Predict Student Success. *J. Data Sci. Artif. Intell.* **2024**, *3*, 1.

21. Nabil, A.; Seyam, M.; Abou-Elfetouh, A. Prediction of students' academic performance based on courses' grades using deep neural networks. *IEEE Access* **2021**, *9*, 140731–140746. [CrossRef]

22. Yousafzai, B.K.; Khan, S.A.; Rahman, T.; Khan, I.; Ullah, I.; Ur Rehman, A.; Baz, M.; Hamam, H.; Cheikhrouhou, O. Student-performulator: Student academic performance using hybrid deep neural network. *Sustainability* **2021**, *13*, 9775. [CrossRef]

23. 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]

24. Beseiso, M. Enhancing Student Success Prediction: A Comparative Analysis of Machine Learning Technique. *TechTrends* **2025**, *69*, 372–384. [CrossRef]

25. Albahli, S. Advancing Sustainable Educational Practices Through AI-Driven Prediction of Academic Outcomes. *Sustainability* **2025**, *17*, 1087. [CrossRef]

26. 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]

27. 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.

28. 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.

29. 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]

30. 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]

31. 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.

32. 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]

33. 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]

34. 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]

35. 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]

36. 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]

37. 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.

38. 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.

39. 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.

40. 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.

41. 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]

42. 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.

43. Mustapha, M.T.; Ozsahin, I.; Ozsahin, D.U. Convolution neural network and deep learning. In *Artificial Intelligence and Image Processing in Medical Imaging*; Academic Press: Cambridge, MA, USA, 2024; pp. 21–50.

44. Sekar, A. Performance Analysis: LSTMs, GRUs, Single & Bidirectional RNNs in Classification & Regression Problems. Available online: https://www.researchgate.net/publication/381654032_Performance_Analysis_LSTMs_GRUs_Single_Bidirectional_RNNs_in_Classification_Regression_Problems (accessed on 1 April 2025).

45. Dey, R.; Salem, F.M. Gate-variants of gated recurrent unit (GRU) neural networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1143–1146.

46. Adefemi Alimi, K.O.; Ouahada, K.; Abu-Mahfouz, A.M.; Rimer, S.; Alimi, O.A. Refined LSTM Based Intrusion Detection for Denial-of-Service Attack in Internet of Things. *J. Sens. Actuator Netw.* **2022**, *11*, 32. [CrossRef]