International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

NOVEL STUDY OF DEEP LEARNING ALGORITHMS FOR STUDENTS' PERFORMANCE PREDICTION

M. Kannan, [2]Dr. S. Albert Antony Raj, [3]Dr. K. R. Ananthapadmanaban

[1]Department of Computer Applications, College of Science,

SRM Institute of Science and Technology,

Kattankulathur, Tamil Nadu

km1200@srmist.edu.in

ORCID ID: 0000-0002-6907-3086

[2]Department of Computer Science and Applications, College of Science,

SRM Institute of Science and Technology,

Kattankulathur, Tamil Nadu

alberts@srmist.edu.in

ORCID ID:

[3]Professor and Head, Department of Computer Science and Application, SRM Arts and Science, Kattankulathur, Chennai, Tamil Nadu

toananths@gmail.com

ORCID ID: 0000-0001-5430-3355

Abstract

Higher education refers to that phase of education that starts after schooling. Universities offer graduate courses in multiple disciplines, and students select courses that interest them. Both the government and private sectors run universities in India. These institutions certify the students based on their performance in the academic sessions. These universities compete with each other in providing quality education that opens new avenues for further studies or employment globally. Due to the intense competition in the education and employment sectors, designing a robust assessment system for students' performance has become imperative. Given the vast majority of the student population and their potential in various domains, there is a need for a computer-aided algorithm that can analyse their performance with the highest possible degree of accuracy. The existing methods that used machine learning algorithms to assess student performance gave less accuracy and could not handle massive datasets. This paper presents a deep learning methodology capable of managing diverse datasets, ranging from minimal to extensive, incorporating a Deep Artificial Neural Network (DANN), a Convolutional Neural Network (CNN), and Bidirectional Long Short-Term Memory (BiLSTM) algorithms, along with hybrid deep learning architectures, including DANN-BiLSTM and CNN-BiLSTM. This

Received: August 09, 2025

approach aims to assess student performance, known as the student performance grading system.

The research shows that the proposed Students Performance Grading system gave accurate results on various factors of training and testing compared to the existing ones. The results show that the MSE of CNN is 0.0227, the MSE of Deep ANN is 0.0219, and the MSE of Bi-LSTM is 0.0233. Additionally, the MSEs are 0.0211 for DANN-Bi-LSTM and 0.0215 for CNN-BiLSTM. The accuracies are 96.74% for DANN, 96.42% for CNN, and 95.80% for Bi-LSTM. and DANN-Bi-LSTM attained an accuracy of 99.31%, while CNN-Bi-LSTM achieved an accuracy of 97.30%. This research identified that DANN-Bi-LSTM is the most effective algorithm for forecasting student success.

. Keywords: Deep Learning, DANN, CNN, Bi-LSTM, hybrid deep learning architectures, Students' Performance prediction

Introduction

COVID-19 has changed the way educational institutions operate. As learning and teaching transition online, considering these factors is significant for extrapolating learner performance. Conversely, assessing student performance is a critical issue in education, as it influences intervention strategies and the success of both educators and learners. Statistical methods that have been around for a long time, as well as more recent deep learning techniques, both offer interesting options for addressing this difficulty. A crucial aspect that significantly influenced the sparsity of student action data was the anticipated accuracy of the model.

Academic performance can be evaluated from several perspectives. The extent to which a student has learned and understands the course material. Students' effort and dedication to their studies include attendance, organisation, extracurricular activities and time management. A student’s behaviour in the classroom consists of participation, cooperation, and attention to the teacher and peers. Examination results, quiz scores, completed online courses, and other evaluations indicate a student's comprehension and grasp of the course content. A student's academic performance summary is often derived from a synthesis of the variables above. A student's ability to interact with others, manage emotions, and develop positive relationships can significantly impact their academic success.

Various deep learning (DL) methodologies, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory (Bi-LSTM) networks, have been employed in machine learning to predict student performance. These algorithms use sequential student action data to forecast future performance behaviour. Challenges associated with the scarcity of student engagement data, characterised by minimal interaction with course materials, disparities in demographic, academic, and socioeconomic factors, may distort predictive models assessing student success. This study considers the following features for the experiment: Student ID, age, gender, GPA for the last semester, attendance, study hours per week, MOOC participation, online course hours, spoken tutorial participation, extracurricular activity score, test scores, and current GPA.

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

The primary goal of this research is to evaluate the effectiveness of deep learning algorithms in predicting student performance while addressing the challenges associated with data sparsity and class imbalance. Many models were tested using the real-time dataset to determine the optimal modelling approach for addressing data sparsity and class imbalance in educational datasets. Furthermore, the effectiveness of CNN, ANN, Bi-LSTM, CNN-BiLSTM, and ANN-BiLSTM in addressing difficulties in predicting student performance was evaluated using MSE, Accuracy, Precision, Recall, and F1-Score.

Literature Review

This paper examines various literature, detailing significant studies that employed deep learning methods for predicting student performance and analysis. The combination of advanced learning algorithms, data mining, and extensive learning models yields a comprehensive framework for performance prediction and the development of a complex model for future iterations. The model is being experimented with in Assam on three school students using AI technologies powered by deep learning algorithms and data mining techniques. The results provided more accurate statistical data about the students and a clearer view of them. The results were compared with the original data, providing better performance metrics S. Hussain et al., 2019). Numerous literature evaluations have been conducted on these ideas, and the viability of machine learning methods in assessing performance has also been seen. The studies indicate that these algorithms enhance the assessment of student performance and capabilities, offering deeper insights (B. Albreiki et al., 2021).

With the development of various technologies, the education sector has adopted them and is considering implementing a virtual environment that helps automate and efficiently implement the teaching system. However, during the introduction process, the efficiency and its effectiveness need to be checked. Hence, various research works are carried out to check the feasibility through various machine learning techniques and compare their Performance (A. Rivas et al. (2019), H. Pallothadka et al. (2021), Y.A. Alsariera (2022)). The performance evaluation procedure using machine learning and deep learning algorithms has also been tested in several research projects. The linear regression model is suggested to yield the necessary outcomes. The results provided a more accurate evaluation of student performance, and the error in the proposed model is assessed through the average mean error. The same model is then repeated using both supervised and unsupervised methods, with training and implementation directly on the model (A.O. Oyedeji et al., 2020). Many data mining operations are automated utilising machine learning and deep learning methods, such as deep neural networks, decision trees, random forests, support vector machines, k-nearest neighbours, and other deep learning methodologies. In comparison to machine learning algorithms, deep learning algorithms offer superior data classification and sufficient capacity for data mining. (Vijayalakshmi, V., & Venkatachalapathy, K. (2019), M. Yagci (2022), AS. Hashim et al. (2020), A. Nabil et al. (2021). It was also utilised on various networking platforms, which provided improved efficiency. The GridNet model, proposed in conjunction with binary LSTM models, provided teachers with a more accurate evaluation of student performance and a timely review of educational progress. The model was experimented with by Udacity students, compared to the logistic regression model, which provided the necessary prediction with better accuracy (BH Kim et al., 2018). In some predictive models, supervised machine learning algorithms, such as the XGBoost algorithm, produced better classification and prediction of student performance analysis. When compared to other boosting algorithms, such as AdaBoost and random forest algorithms, it provided better results for student performance analysis (A. Asselman et al., 2021; K.M. Hasib et al., 2022; Amjad et al., 2022).

This paper, Kumar et al. (2021), develops a Model using single and ensemble classifiers. Data preparation removes duplicates and extraneous data, improving precision. The initial classification phase employs Naive Bayes, while the subsequent phase utilises ensemble classifiers, including Boosting, Stacking, and Random Forest. The performance analysis shows the model's classification accuracy and efficiency. Siddiqa et al. (2022) suggested an updated progressive approach for feature subset selection that incorporates neuro-fuzzy categorisation for educational data mining. The Chaotic Whale Optimisation Algorithm (CWOA) identifies relevant feature groups to improve classification, followed by Neuro-Fuzzy Classification (NFC).

In accordance with the peculiarities of MOOC learning, a study by Wen et al. (2020) introduces a feature matrix for local learning behaviour correlations and a CNN model for performance prediction, demonstrating enhanced accuracy with early predictions. Lin et al. (2020) combine CNN and LSTM networks for continuous facial emotional pattern recognition to analyse academic emotions, thereby improving the ability of e-learning systems to understand and motivate students. Experimental findings indicate a 12% enhancement in meticulousness compared to the utilisation of CNN in isolation.

Farissi et al. (2020) propose using a Genetic Algorithm (GA) for feature selection, combined with ensemble classification, to predict academic performance. GA attains optimal feature selection for enhanced accuracy, assessed using AUC on Kaggle data, demonstrating remarkable predictive outcomes. Turabieh et al. (2021) utilise a dynamic controller and KNN algorithm for feature selection and clustering, employing several classifiers, including LRNN, and obtaining 92% accuracy with the improved combination of HHO and LRNN.

Yang et al. (2021) train lightweight CNN models by mimicking the learning trajectory of a teacher model (TM). Landmarks from the TM stages guide the student model (SM), thereby increasing prediction accuracy on datasets such as CIFAR-10 and ImageNet-10. A different study [21] examines the application of deep learning for forecasting academic performance and risk characteristics in children, utilising deep neural networks, decision trees, and random forests with an accuracy of 89%. Diverse resampling techniques tackle skewed datasets.
Nabil et al. (2021) present a comprehensive reasoning evaluation framework that enhances traditional models through deep learning to assess students' proficiency in abilities and problem-solving, taking into account inadvertent errors and random guesses. Empirical studies on actual datasets demonstrate the model's efficacy.

Research Gab

- Most researchers use standard features, such as personal information, geographic information, and students' marks, for the experiment.

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

- Many experiments follow machine learning algorithms.
- Accuracy still needs improvement.
- Failed to capture extracurricular activities like MOOC courses, learning hours, online courses, interests, etc.

Data Collection

This project involves gathering data from students at SRM Arts and Science College in Tamil Nadu, utilising several data sources, including student archives, course constituents, online knowledge platforms, and feedback forms. These records may encompass numerical and category factors as follows,

Table 1. Features of Data

Features	Datatype
Student_ID	Numerical
Age	Numerical
Gender	Categorical
GPA_Last_Semester	Numerical
Attendance	Numerical
Study_Hours_Per_Week	Numerical
MOOC_Participation	Numerical
Online_Courses_Hours	Numerical
Spoken_Tutorial_Participation	Numerical
Extracurricular_Activity_Score	Numerical
Test_Scores	Numerical
GPA_Current	Numerical

It's a deep learning analysis that doesn't require any previous preprocessing techniques. At the same time, the feature 'gender' was converted to label encoding, where male is represented by zero and female is represented by 1 for students' performance prediction.

Methodology

The education system has generated vast amounts of data related to student performance, including test scores, attendance records, and other key metrics. However, manually analysing massive amounts of students' data is time-consuming and tedious. Consequently, an automated system is required to analyse data, discover trends, and forecast future results. This study aims to utilise deep learning algorithms to analyse student performance data and predict their future academic outcomes. The system uses a dataset of students' activity data to develop a model that predicts grades, assesses performance, and provides insights into factors that enhance academic achievement. The objective is to create a system that delivers actionable data to instructors and proprietors, enabling them to make educated decisions to improve student

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

outcomes. The following figure shows the experimental deep learning algorithms for the students' performance system.

"Students' Performance Analysis Using Deep Learning Algorithm" aims to develop a predictive model utilising deep learning algorithms to assess students' performance and forecast their future academic outcomes. This model employs deep learning algorithms to analyse extensive datasets and discern patterns that can predict future performance. To thoroughly analyse a student's educational performance, the model takes into account a variety of parameters, including the student's past academic performance, online course study time, the kind of online course, test score, and external factors. The purpose is to identify factors that influence student performance and deliver actionable insights to improve teaching and learning processes. Ultimately, our study aims to improve student learning outcomes and empower educators to make informed, data-driven decisions about teaching and learning strategies.

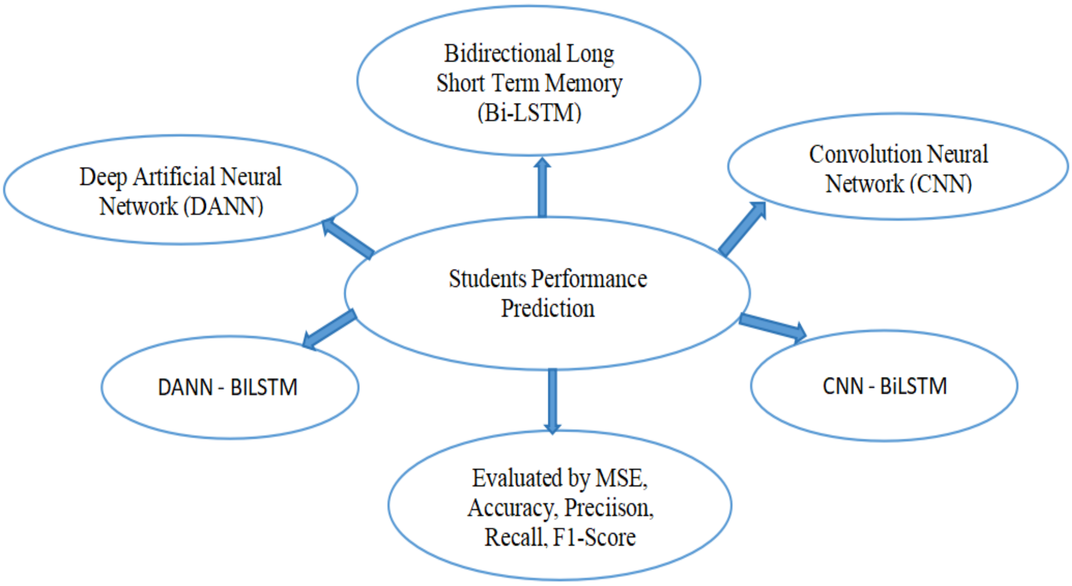

Figure 1. Schematic Workflow

Deep Artificial Neural Network:

DANN is an upgraded model of an Artificial Neural Network constructed with numerous hidden layers to capture complex patterns. It collects information from the input layer, two or more hidden layers detect complex and sensitive patterns in the data, and the output layer generates the final prediction. Each feature in the input data is sent to the neurons of the first hidden layer. Each feature has a weight, which is multiplied and added to the bias, and the final value is passed to the next layer. The activation function utilises this weighted sum to capture the intricate and complex patterns. The ADM optimiser updated layer weights. Here, the ReLU activation function is used, and this process is repeated for every forward pass through the layers. Backpropagation is also performed until the final solution is obtained. The entire linked layer finally produces the prediction of the students' performance system. The hyperparameter tuning was carried out to improve the prediction. The final prediction and actual prediction were evaluated by loss and accuracy values. Figure 2 illustrates the layer erection of DANN.

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 128)	1536
dense_2 (Dense)	(None, 64)	8256
dense_3 (Dense)	(None, 32)	2080
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 1)	17

Total params: 12,417
Trainable params: 12,417
Non-trainable params: 0

Figure 2. Layer structure of DANN

Convolution Neural Network (CNN)

Convolutional Neural Networks (CNNs) are extensively employed in computer vision and geographical data analysis. They convert intricate inputs into more manageable elements through a sequence of processes, effectively capturing spatial and temporal connections in images. CNN architectures demonstrate superior performance with image data due to their reduced parameters and reusable weights. They utilise shared-weight neural networks and convolutional procedures to partition input into two-dimensional feature maps. This work employs ID CNN to handle time series data, while adhering to the exact configuration of layers and activation functions to achieve the objective. This structure is built by an input layer to fetch the given features, and an ID Convolution slide is applied over the features to identify strong relationships. The pooling layer reduces the unwanted features and focuses on the dominant features. Upon traversing several convolutional and pooling layers, the output layer is pooled to transform the feature maps into a one-dimensional vector. Fully connected or dense layers receive ID vectors to generate the final prediction. The advantage of IDCNN is that it captures the least significant to the most essential features. The following figure shows the layer structure of IDCNN.

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 4, 64)	192
max_pooling1d_1 (MaxPooling1D)	(None, 2, 64)	0
flatten_1 (Flatten)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8256
dense_2 (Dense)	(None, 32)	2080
dense_3 (Dense)	(None, 1)	33

Total params: 10,561
Trainable params: 10,561
Non-trainable params: 0

Figure 3. 1DCNN layer Architecture

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Bidirectional Long Short-Term Memory

BiLSTM is identical to Long Short-Term Memory, but it processes data in both forward and backwards directions. It contains 100 memory cells or neurons to process input vectors. It processes data in both directions, thereby capturing past data and training it with current data. The following merged data is passed to the flatten layer, and the 2D output of BiLSTM is converted into an ID vector, which eases the process of consecutive layers. The final layer considers the production from the previous layers and performs the prediction.

Forward and backwards LSTMs have separate weights for each, resulting in more parameters than an LSTM, and they change according to the process. BiLSTM has no trainable parameters, but the units can be changed.

```
Layer (type)              Output Shape            Param #
=================================================================
bidirectional_1 (BiLSTM)      (None, timesteps, 200)     161600
_________________________________________________________________
flatten_1 (Flatten)          (None, 200 * timesteps)    0
_________________________________________________________________
dense_1 (Dense)              (None, 3)             timesteps * 3
=================================================================
Total params: 161,603 + (timesteps * 3)
Trainable params: 161,603 + (timesteps * 3)
Non-trainable params: 0
=================================================================
```

Figure 4. Layer details of BiLSTM

Hybrid DANN-BILSTM

This hybrid structure combines the feature learning capability of ANN and the sequential processing capability of BiLSTM, which captures the relationship between features and sequential dependencies in the data. It evaluates the characteristics. Student_ID, Age, Gender, Last_Semester_GPA, Attendance, Weekly_Study_Hours, MOOC_Participation, Online_Course_Hours, Spoken_Tutorial_Participation, Extracurricular_Activity_Score, Test_Scores, Current_GPA, it is a synthesis of sequential and non-sequential characteristics.

Hybrid algorithm steps

1. Input layer 1 - Deep ANN processed the non-sequential features

2. Input layer 2 - BiLSTM processed the sequential features

3. Multiple fully connected layers worked for DANN

4. Long-term dependencies observed by BiLSTM

5. DANN and BiLSTM outputs were combined

6. Combined output processed by dense layer

7. The output layer predicts the final prediction

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

8. The outcome was assessed using MSE loss value, Exactness, Meticulousness, Reminiscence, and F1-Score.

The following figure presents the layer structure of DANN-BiLSTM.

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 256)	2048
dense_2 (Dense)	(None, 128)	32896
dense_3 (Dense)	(None, 64)	8256
bidirectional_1 (Bidirectional)	(None, 200)	160800
concatenate_1 (Concatenate)	(None, 264)	0
dense_4 (Dense)	(None, 64)	16960
dense_5 (Dense)	(None, 32)	2080
dense_6 (Dense)	(None, 3)	99

Total params: 220,139
Trainable params: 220,139
Non-trainable params: 0

Figure 5. DANN-BiLSTM Layers

Hybrid CNN-BiLSTM

CNN extracts the features, and BiLSTM captures long-term dependencies. To extract essential configurations from the provided data, an ID convolutional layer with 64 filters, three kernel sizes, and a ReLU activation function is utilised. Max pooling helps to reduce the dimensionality of the data. The next 1D Convolution layer, with 126 features, identifies the significant features, and in the subsequent pooling layer, the feature map was downsampled. At the same time, BiLSTM with 200 units, which included 100 units for the forward layer and 100 units for the backwards layer. Here, CNN extracts the features, and BiLSTM processes the dependencies. After completing both layer processes, the outputs were combined and passed to the final layer.

1. Input layer 1 - CNN processed the non-sequential features
2. Input layer 2 - BiLSTM processed the sequential features
3. Multiple fully connected layers worked for CNN
4. Long-term dependencies observed by BiLSTM
5. CNN and BiLSTM outputs were combined
6. Combined output processed by dense layer

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

7. The output layer predicts the final prediction

8. Several metrics, including MSE loss value, Accuracy, Precision, Recall, and F1-Score, were utilised to assess the outcome.

The following figure presents the layer structure of CNN-BiLSTM.

```
=================================================================
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_1 (Conv1D)            (None, time_steps-2, 64)   448
_________________________________________________________________
max_pooling1d_1 (MaxPooling1D)   (None, (time_steps-2)//2, 64)  0
_________________________________________________________________
conv1d_2 (Conv1D)            (None, (time_steps-2)//2-2, 128)  24704
_________________________________________________________________
max_pooling1d_2 (MaxPooling1D)   (None, ((time_steps-2)//2-2)//2, 128)  0
_________________________________________________________________
bidirectional_1 (Bidirectional)  (None, 200)            184800
_________________________________________________________________
dense_1 (Dense)              (None, 64)                12864
_________________________________________________________________
dense_2 (Dense)              (None, 32)                2080
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 33
=================================================================
Total params: 224,929
Trainable params: 224,929
Non-trainable params: 0
=================================================================
```

Figure 6. CNN-BiLSTM Layers

Evaluation Metric

In this study, a deep learning-based technique is employed to evaluate a student's performance accurately. Throughout history, several research projects have been conducted in this field. This paper describes the implementation of a deep learning model designed to enhance the model's prediction accuracy. Performance metrics values are used to evaluate both the student's and the suggested model's performance. The student's performance is assessed using a range of factors. It is also necessary to determine the value of the performance metrics to predict the model's efficiency. Several metrics, including Mean Squared Error, precision, recall, accuracy, and f1-score value, are used to achieve this goal.

Mean Squared Error

It determines the average difference between the numbers reported and those projected. This is one way that it can be expressed:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \acute{y}_i)^2 \text{ --------(1)}$$

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

n - Number of data points

y_i - Actual Value

$\acute{y}_i$ - Predicted Value

Precision

This metric is used to evaluate the actually predicted positive value from the total number of positively predicted values. It can be calculated as follows.

$$Precision = \frac{TP}{TP+FP} \text{---------(2)}$$

Recall

The total number of incorrect positive values calculates this metric. It is represented as follows.

$$Recall = \frac{TP}{TP+FN} \text{---------(3)}$$

F1-Score

It helps to find the class imbalance by using precision and recall. It's computed as follows.

$$F1Score = 2 \times \frac{Preciison \times Recall}{Preciison + Recall} \text{---------(4)}$$

Result and discussion

This section provides a summary of the results of the algorithm experimented with for predicting pupil performance. To illustrate the descriptive character of the feature, the figure is presented.

	Student_ID	Age	Gender	GPA_Last_Semester	Attendance	Study_Hours_Per_Week	MOOC_Participation	Online_Courses_Hours
0	1	24	Female	3.48	51	38.6	4	7.6
1	2	21	Female	3.04	86	28.1	3	45.6
2	3	22	Female	2.14	56	23.3	4	44.6
3	4	24	Female	2.74	75	5.8	4	32.7
4	5	20	Male	3.84	75	19.5	1	33.6

Spoken_Tutorial_Participation	Extracurricular_Activity_Score	Test_Scores	GPA_Current
0	3.48	70.32	3.77
1	6.57	82.16	2.91
0	0.16	84.75	3.54
0	0.64	55.21	3.69
0	1.00	92.54	3.42

Figure 7. Descriptive nature of the Dataset

EDA analysis shows the correlation between

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The following table and figure show the MSE values of DANN, CNN, BiLSTM, DANN-BiLSTM, and CNN-BiLSTM.

Table 1. MSE values for deep learning models

Models	MSE
DANN	0.0219
CNN	0.0227
BiLSTM	0.0233
DANN-BiLSTM	0.0211
CNN-BiLSTM	0.0215

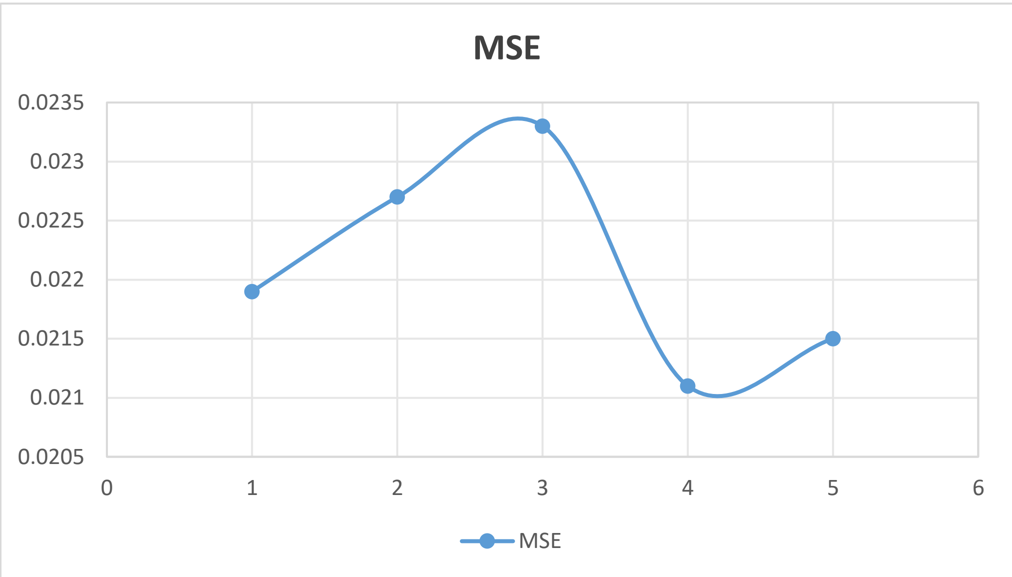

Figure 8. MSE values for deep learning models

From the above result, it's observed that a Low MSE value of 0.0211 was attained for the DANN-BiLSTM model, which is the best model, as it exhibits a low MSE value.

The following table and figure present the performance of deep learning architectures and are mentioned in terms of (%)

Table 2. Model evaluation

Models	Accuracy	Precision	Recall	F1-Score
DANN	96.74	95.78	95.73	95.83
CNN	96.42	94.61	94.63	94.69
BiLSTM	95.80	93.71	93.02	94.91

Received: August 09, 2025

Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

DANN-BiLSTM	99.31	98.83	98.80	98.94
CNN-BiLSTM	97.30	96.87	96.14	96.90

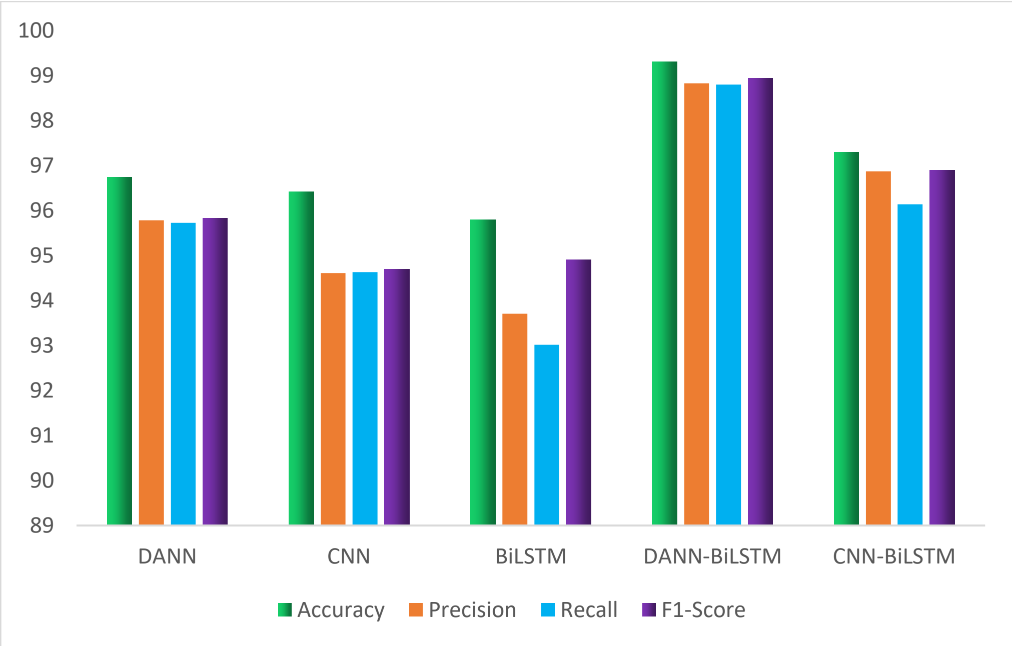

Figure 9. Model Evaluation

According to the above figure, this research found that DANN-Bi-LSTM achieved an accuracy of 99.19%, precision of 98.83%, recall of 98.80%, and F1-score of 98.94%. The following figures illustrate the training and validation of deep learning architectures, among which DANN-BiLSTM achieved high accuracy and minimal loss.

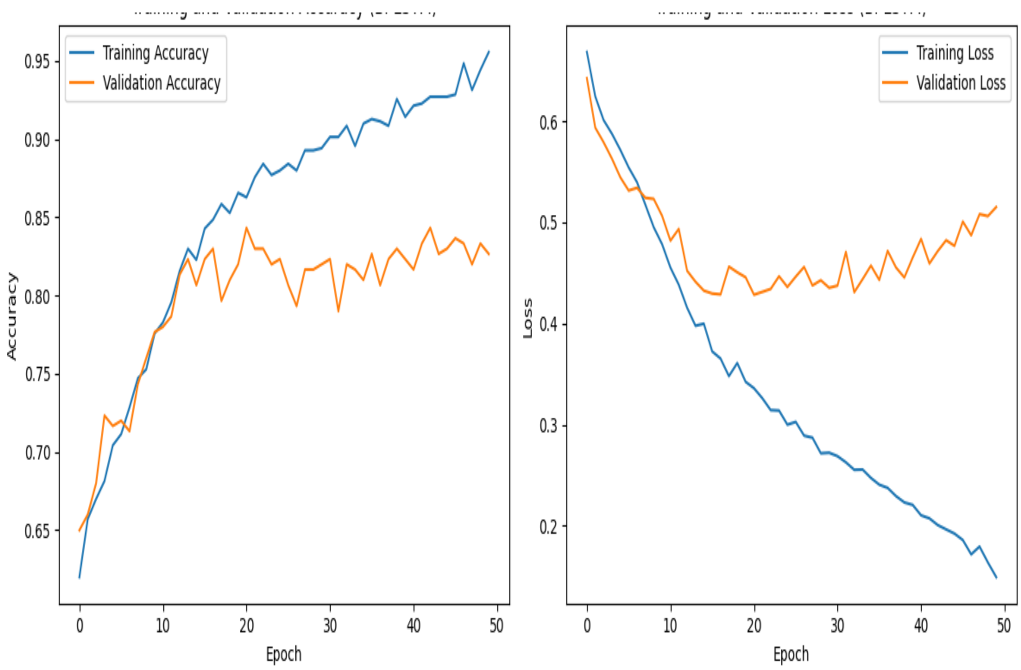

Figure 10. Training and validation of DANN

Received: August 09, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

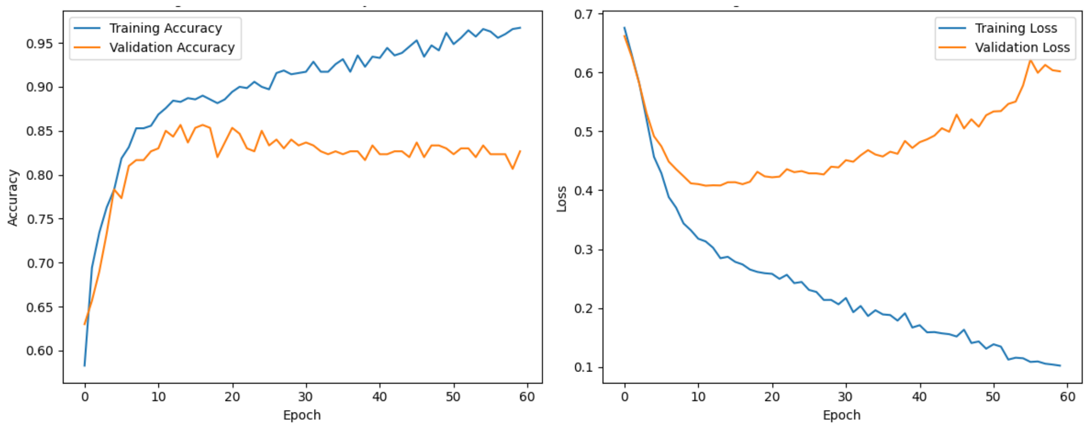

Figure 11. Training and validation of CNN

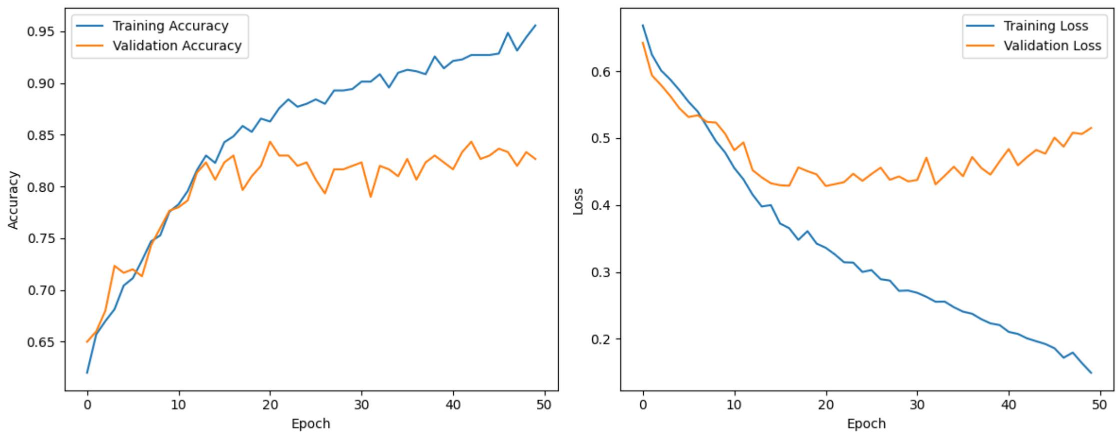

Figure 12. Training and validation of BiLSTM

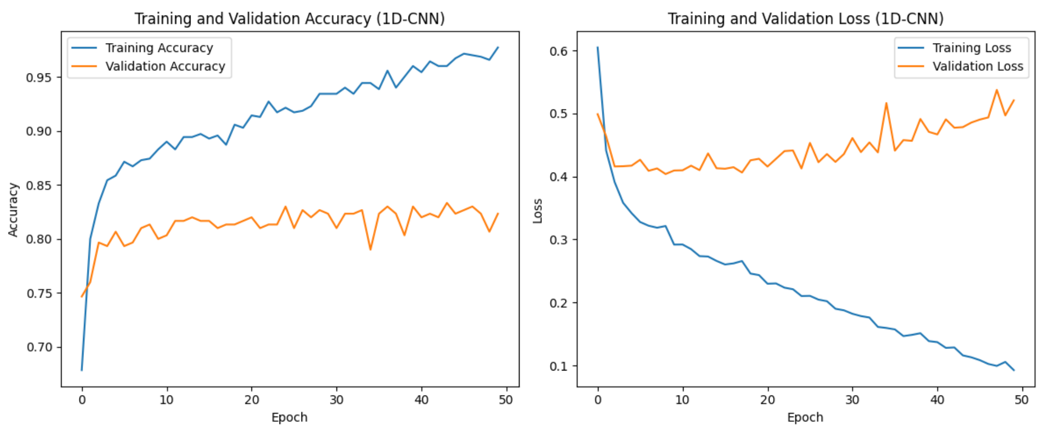

Figure 13. Training and validation of CNN-BiLSTM

Received: August 09, 2025

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

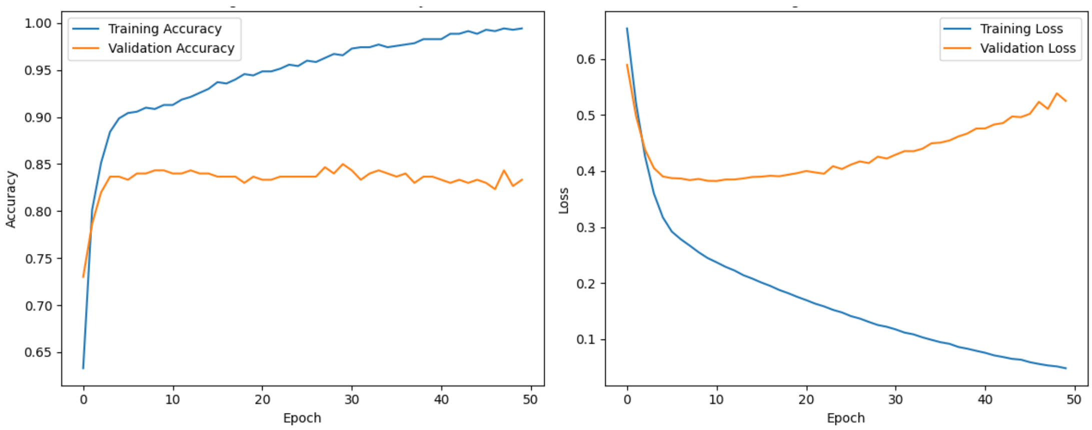

Figure 14. Training and validation of DANN -BiLSTM

Findings:

- The features MOOC_Participation, Online_Courses_Hours, and Spoken_Tutorial_Participation support the grading score of students, which shows high accuracy in prediction.
- These features create a high impact on students' grades.
- DANN-BiLSTM is a more effective model for predicting students' performance.

Conclusion

The purpose of this work is to investigate the challenging task of predicting student performance, considering difficulties such as data scarcity and class imbalance. Initially, deep learning techniques were tested using real-time datasets, and the results demonstrated that deep learning algorithms performed exceptionally well. This research considers the models DANN, CNN, BiLSTM, DANN-BiLSTM, and CNN-BiLSTM, which yield promising results for DANN-BiLSTM. For the purpose of predicting students' performance, the hybrid algorithm DANN-BiLSTM achieved a higher accuracy of 99.31%, precision of 98.83%, recall of 98.80%, and F1 score of 98.94% when compared to the other five techniques. DANN-Bi-LSTM achieves a low MSE value of 0.0211. This research finds that online courses, MOOC Courses, and spoken tutorials create a higher impact on students' GPA scores. This result helps educational institutions develop programs for at-risk students, enhance students' extracurricular activities, and improve students' online learning capabilities. It can be used by teachers, school administrators, policymakers, and other stakeholders to inform decision-making related to education and enhance teaching and learning practices. The primary objective of this research is to improve student learning outcomes and assist educators in making data-informed decisions to improve teaching and learning methodologies.

References

1. Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S. (2019). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review*, *52*, 381-407.

Received: August 09, 2025

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

2. Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student performance prediction using machine learning techniques. *Education Sciences*, *11*(9), 552.

3. Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Rodríguez, S., & Corchado, J. M. (2019). The analysis of students' performance is based on machine learning techniques. In *Learning Technology for Education Challenges: 8th International Workshop, LTEC 2019, Zamora, Spain, July 15–18, 2019, Proceedings 8* (pp. 428-438). Springer International Publishing.

4. Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development.

5. Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. *Computational Intelligence and Neuroscience*, *2022*.

6. Oyedeji, A. O., Okunowo, W. O., Osuntoki, A. A., Olabode, T. B., & Ayo-Folorunso, F. (2020). Insecticidal and biochemical activity of essential oil from Citrus sinensis peel and constituents on Callosobruchus maculatus and Sitophilus zeamais: Pesticide *Biochemistry and Physiology*, *168*, 104643.

7. Vijayalakshmi, V., & Venkatachalapathy, K. (2019). Comparison of predicting students' performance using machine learning algorithms. *International Journal of Intelligent Systemsand Applications*, *11*(12), 34.

8. Yağcı, M. (2022). Educational data mining: predicting students' academic performance usingmachine learning algorithms. *Smart Learning Environments*, *9*(1), 11.

9. Hashim, A. S., Awadh, W. A., & Hamoud, A. K. (2020, November). Student performance prediction model based on supervised machine learning algorithms. In *IOP Conference Series: Materials Science and Engineering* (Vol. 928, No. 3, p. 032019). IOP Publishing.

10. Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of students' academic performance based on courses' grades using deep neural networks. *IEEE Access*, *9*, 140731- 140746.

11. Asselman, A., Khaldi, M., & Aammou, S. (2021). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, 1-20.

12. Hasib, K. M., Rahman, F., Hasnat, R., & Alam, M. G. R. (2022, January). A machine learning and explainable AI approach for predicting secondary school student performance. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0399-0405). IEEE.

13. Amjad, S., Younas, M., Anwar, M., Shaheen, Q., Shiraz, M., & Gani, A. (2022). Data mining techniques to analyse the impact of social media on the academic performance of high school students. *Wireless Communications and Mobile Computing*, *2022*.

14. Xiong, S. Y., Gasim, E. F. M., Ying, C. X., Wah, K. K., & Ha, L. M. (2022). A Proposed Hybrid CNN- RNN Architecture for Student Performance Prediction. *International Journal of Intelligent Systems and Applications in Engineering*, *10*(3), 347-355.

15. E. S. V. Kumar, S. A. A. Balamurugan and S. Sasikala, "Multi-tier student performance evaluation model (MTSPEM) with integrated classification techniques for educational decision making", *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 1796-1808, Mar. 2021.

16. A. Siddiqa, S. A. Z. Naqvi, M. Ahsan, A. Ditta, H. Alquhayz, M. A. Khan, et al., "An improved evolutionary algorithm for data mining and knowledge discovery", *Comput. Mater. Continua*, vol. 71, no. 1, pp. 1233-1247, 2022.

17. Y. Wen, Y. Tian, B. Wen, Q. Zhou, G. Cai and S. Liu, "Consideration of the local correlation of learning behaviours to predict dropouts from MOOCs", *Tsinghua Sci. Technol.*, vol. 25, no. 3, pp. 336-347, Jun. 2020.

18. S. Y. Lin, C. M. Wu, S. L. Chen and T. L. Lin, "Continuous facial emotion recognition method based on deep learning of academic emotions", *Sens. Mater.*, vol. 32, no. 10, pp. 3243-3259, Oct. 2020.

19. A. Farissi, H. Mohamed Dahlan and Samsuryadi, "Genetic algorithm-based feature selection with ensemble methods for student academic performance prediction", *J. Phys. Conf. Ser.*, vol. 1500, no. 1, Apr. 2020.

20. H. Turabieh, S. A. Azwari, M. Rokaya, W. Alosaimi, A. Alharbi, W. Alhakami, et al., "Enhanced Harris hawks optimisation as a feature selection for the prediction of student performance", *Computing*, vol. 7, pp. 1417-1438, Jan. 2021.

21. H. B. Ma, S. Y. Yang, D. Z. Feng and L. C. Jiao, "Progressive mimic learning: A new perspective to train lightweight CNN models", *Neurocomputing*, vol. 456, pp. 220-231, Oct. 2021.

22. Thamilarasi V., Roselin. R., Automatic Classification and Accuracy by Deep Learning Using CNN Methods in Lung Chest X-Ray Image, IOP Conf. Ser.: Mater. Sci. Eng. 1055 012099.(2020).

23. A. Nabil, M. Seyam and A. Abou-Elfetouh, "Prediction of students' academic performance based on courses' grades using deep neural networks", *IEEE Access*, vol. 9, pp. 140731-140746, 2021.

24. L. Gao, Z. Zhao, C. Li, J. Zhao and Q. Zeng, "Deep cognitive diagnosis model for predicting students' performance", *Future Gener. Comput. Syst.*, vol. 126, pp. 252-262, Jan. 2022.

25. P. Mishra, A. Biancolillo, J. M. Roger, F. Marini and D. N. Rutledge, "New data preprocessing trends based on an ensemble of multiple preprocessing techniques", *TrAC Trends Anal. Chem.*, vol. 132, Nov. 2020.

26. M. Kannan. (2024). Enhancing Student Performance Prediction Using Deep Belief Networks with Ant Lion Optimisation. International Journal of Intelligent Systems and

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

Applications in Engineering, 12(21s), 4505.

27. Marjani M, Mahdianpari M, Mohammadimanesh F. CNN-BiLSTM: A Novel Deep Learning Model for Near-Real-Time Daily Wildfire Spread Prediction. Remote Sensing. 2024; 16(8):1467. https://doi.org/10.3390/rs16081467

28. Alshaikh, Fatema & Hewahi, Nabil. (2024). Convolutional Neural Network for Predicting Student Academic Performance in an Intelligent Tutoring System. International Journal of Computing and Digital Systems. 15. 239-258. 10.12785/ijcds/150119.

29. V. Thamilarasi, “Artificial Intelligence-Driven Smart Scenic Management: Automated Decision Making and Optimisation”, Int J Intell Syst Appl Eng, vol. 12, no. 4, pp. 2731 – , Jun. 2024.