GPH-International Journal of Computer Science and Engineering
(Online ISSN 3050-9661)
Volume 9 | Issue 01 | Jan-June 2026
Published by: Global Publication House
DOI: https://doi.org/10.5281/zenodo.24082124
Journal homepage: https://gphjournal.org/index.php/cse/index

A Comparison of Deep Learning Techniques for Predicting Students' Academic Performance

Mohamed Aboh Hamdalla Elbashir[1], Mohamed Hamouda Karboos[2]

[1]Department of Management Information Systems, Faculty of Economics & Commercial Studies, University of Kordofan, Sudan.
[2]Department of Information Technology, Faculty of computer studies & statistics, University of Kordofan, Sudan.
Email: mohamed.aboh@kordofan.edu.sd

Abstract

Higher educational institutions are undergoing a significant digital transformation, driven by rapid advances in artificial intelligence and the growing availability of educational data. Predicting students' academic performance has become an important analytical task for higher education institutions, as it enables the early identification of students at risk of academic underperformance and supports timely intervention. This study aims to compare the performance of four deep learning models for predicting students' academic performance using academic and demographic data, identify the most influential predictive features, and examine performance differences across gender and academic majors. The dataset for study is the academic achievement data of 2,878 students from Faculty of Economics & Commercial Studies during the academic year 2017-2019. Four deep learning models were evaluated (MLP, CNN, LSTM and DNN). The results demonstrated that the LSTM model achieved the highest predictive accuracy of 97.14%, outperforming the other models. Feature-importance analysis further revealed that prior academic achievement, particularly second-semester GPA (GPA2), was the most influential predictor of subsequent academic performance. The findings also indicated that gender had limited influence on predictive performance, whereas differences across academic majors were more pronounced. Overall, this study provides a methodological framework that can support higher education institutions in developing data-driven academic prediction systems for early identification and intervention. Future research is recommended to investigate more advanced deep learning architectures, particularly hybrid CNN–LSTM models, to leverage both local feature patterns and temporal dependencies in sequential academic data, potentially improving predictive accuracy and computational efficiency.

Keywords:

Educational data mining, Academic performance prediction, Deep learning Algorithms.

1. INTRODUCTION

Education constitutes a fundamental pillar of national development and a key indicator of progress across social, economic, political, and cultural dimensions. Accordingly, societies place considerable emphasis on strengthening their educational systems, particularly higher education, given its central role in developing human capital, advancing knowledge, and supporting sustainable development (Al-Laqani, 2005).

© 2026 Author(s). Published by Global Publication House (GPH).
This is an Open Access article distributed under the terms of the CC BY 4.0 License.

In recent years, the higher education sector has undergone an unprecedented digital transformation, with technology becoming an integral component of teaching and learning processes. The widespread adoption of Learning Management Systems (LMS), e-learning platforms, and digital academic information systems has resulted in the continuous generation of large volumes of educational data that reflect students' academic performance, learning behaviors, interactions, and engagement patterns (Villegas-Espinoza & Necochea-Chamorro, 2025). The increasing availability of such data has created new opportunities to gain deeper insights into the educational process and improve learning outcomes, contributing to the emergence and rapid development of Educational Data Mining (EDM) as an important research field (Ramaraj et al., 2025).

Among the major applications of EDM, student academic performance prediction has received considerable research attention. Accurate prediction enables educational institutions to identify students at risk of academic underperformance, failure, or dropout at an early stage, thereby facilitating timely and targeted interventions (Bhushan, 2023). Predictive systems can also support institutional decision-making, improve the allocation of educational resources, enhance curriculum development, and increase students' opportunities for academic success. Furthermore, data-driven early intervention has been recognized as an important strategy for improving student retention and reducing the risk of academic failure and dropout (UNESCO, 2024).

Declining academic achievement and low Cumulative Grade Point Average (CGPA) remain significant challenges facing higher education institutions. Academic underperformance is a complex and multidimensional phenomenon influenced by interconnected academic, economic, social, cultural, institutional, and psychological factors. Consequently, identifying the factors associated with students' academic performance and accurately predicting future outcomes remain challenging tasks. Low academic achievement may also contribute to educational inefficiency by increasing course repetition, prolonging the duration of study, and raising the likelihood of withdrawal or dropout.

Recent advances in artificial intelligence, particularly deep learning, have provided new opportunities to address these challenges. Deep learning, as an advanced branch of machine learning, employs multilayer artificial neural networks to learn complex and nonlinear patterns from large and multidimensional datasets. In educational applications, deep learning models have demonstrated considerable potential for capturing hidden relationships within students' academic and demographic data, thereby improving predictive performance and supporting more effective academic decision-making.

Against this background, the present study conducts a systematic comparative evaluation of four deep learning models Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Deep Neural Network (DNN) for predicting students' academic performance. In addition to comparing their predictive performance, the study investigates the most influential features affecting prediction accuracy and examines differences in academic performance across gender and academic majors. Through this comparative framework, the study seeks to identify the most effective deep learning architecture for academic performance prediction and provide empirical evidence that can support the development of intelligent, data-driven early-warning and academic support systems in higher education.

2. Research Problem

Despite the growing emphasis on proactive educational management, measuring and predicting student academic performance remains a critical institutional challenge. A primary limitation lies in the reliance on traditional, established statistical procedures, which lack the capacity to

comprehensively analyze large-scale educational data with the requisite accuracy. While educational data mining and machine learning algorithms have emerged as powerful alternatives, empirical literature reveals a distinct inconsistency in results regarding which classification models achieve optimal performance across varied educational contexts. Furthermore, existing research exhibits a lack of consensus on feature importance, emphasizing a critical need to systematically identify the academic, demographic, and behavioral factors that most significantly influence prediction accuracy. Consequently, this study addresses these gaps by evaluating advanced predictive models and feature selection mechanisms to establish a robust, context-aware framework for early student tracking.

3. Research Questions

This study seeks to answer the following research questions:

1. Which of the four algorithms performs best based on the data used in the study?

2. Which features have the greatest impact on the accuracy and performance of the prediction models?

4. Objectives of the Study

The objective of this study is to evaluate the performance of a set of deep learning algorithms for predicting student academic performance, and thereby identify the best algorithms, as well as identify the most important features that influence student performance and compare performance levels among students and across different college departments.

5. Related Works

Several studies have addressed the topic of predicting student performance using machine learning and deep learning techniques. The study by Ma & Xiao (2026) on the challenge of incorporating behavioral factors into models for predicting academic performance. It aimed to develop a deep learning model that takes into account student behaviors such as attendance and study habits, drawing on multiple academic and behavioral data points. The results showed that incorporating behavioral factors significantly contributes to improving prediction accuracy, and the study recommended integrating behavioral data with academic data in future models. A study by Rodríguez et al. (2025) compared the performance of hybrid deep models with semi-supervised feature selection models using ensemble deep learning techniques on diverse educational datasets, The results demonstrated that the integrated deep models outperformed the others in terms of accuracy and stability in predicting academic performance, and the study recommended integrating more than one model to obtain more accurate and reliable results. In the same vein, a study (Xiao, 2025) aimed to address the weakness of individual models in capturing temporal patterns by developing a hybrid model that combines an MLP, an LSTM, and a Transformer using student grade data and feature engineering techniques, The results showed that the LSTM model outperformed the others in prediction and achieved a lower mean squared error (MSE) value, reflecting its effectiveness in analyzing time-series data.

In 2024, multiple studies examined the use of deep learning in analyzing educational data and improving the accuracy of academic performance predictions. A study by Almalawi et al. (2024) aimed to review models for predicting academic performance, focusing on the role of deep learning in analyzing educational data, It relied on a systematic review of a large number

of recent studies, and the results showed that hybrid and deep learning models outperformed traditional models in terms of accuracy and generalizability. The study recommended expanding the use of these models in educational systems. Similarly, a study by Tang et al. (2024) the problem of limited performance of individual models and aimed to improve prediction using ensemble deep learning, in which it integrated several deep neural networks within a unified framework using digital educational data; and the results showed a clear improvement in prediction accuracy and stability compared to individual models. Similarly, a study by Hussain et al., 2024) aimed to address the challenge of analyzing massive educational datasets using deep learning techniques within the framework of educational data mining. The results demonstrated the model's ability to extract hidden patterns and improve the accuracy of academic performance predictions, while recommending the expanded use of deep learning techniques in the analysis of massive educational datasets.

In the Arab context, a study by Salama et al. (2024) addressed the problem of uncertainty in student performance prediction models by developing a deep learning model based on neuro-sophisticated clusters to handle ambiguity in educational data, The results showed an improvement in prediction accuracy along with a better ability to handle uncertainty, and the study recommended integrating explainable artificial intelligence techniques into educational systems (Salama et al., 2024). Similarly, Al-Saawi's study (2024) addressed the issue of the limited use of AI applications in education; it aimed to explore the potential of these applications in improving academic performance and employed a descriptive approach using a questionnaire administered to a sample of female teachers. The results demonstrated the importance of AI applications in improving the educational process despite the presence of some barriers to implementation, and recommended training teachers and expanding the use of AI technologies in education (Al-Saawi, 2024). Studies published in 2023 focused on developing deep learning models and improving their interpretability; the study by Ben Said et al. (2023) aimed to address the difficulty of predicting performance in e-learning after the COVID-19 pandemic by developing a deep learning model that converts time series into images and analyzes them using convolutional neural networks (CNN). The results showed a marked improvement in prediction accuracy compared to traditional methods (Ben Said et al., 2023). Similarly, a study by Alsubhi (2023) addressed the issue of the lack of clarity regarding the factors influencing academic performance and aimed to improve the interpretability of deep learning models by comparing black-box models with interpretable models, The results showed that incorporating explainable AI techniques helped improve understanding of the influencing factors without compromising predictive accuracy (Alsubhi, 2023). Al-Jubouri's (2023) study aimed to analyze the relationship between mental effort and cognitive assessment among graduate students as predictors of academic performance, and the results showed a statistically significant relationship between mental effort and academic performance (Al-Jubouri, 2023). Similarly, the study by Hamouda and Abu Ghali (2023) addressed the issue of predicting academic performance through psychological variables, and the results showed that cognitive biases are strong predictors of academic performance (Hamouda and Abu Ghali, 2023). Similarly, the study by Hammad and Baathman (2023) aimed to analyze the relationship between cognitive bias and learning motivation as factors influencing academic performance, and the results showed that learning motivation plays an important mediating role in improving academic performance, with a recommendation to develop predictive models based on behavioral and psychological variables (Hammad and Baathman, 2023).

Despite the wealth of previous studies, most have focused on comparing a limited number of deep learning algorithms in narrow contexts or using homogeneous datasets, and some have focused on only one or two deep learning models. Furthermore, the impact of parameter tuning has not received sufficient attention in comparative studies. This study aims to bridge this gap by conducting a systematic comparison of four of the most well-known deep learning algorithms and analyzing the factors that most influence prediction accuracy.

6. Methodology

Figure (1) presents the research diagram of this work, which involves data gathering and preprocessing, feature correlation analysis and selection, modeling and prediction and model evaluation, where the models' performance is measured by (Accuracy, Recall, Precision and F – measure)

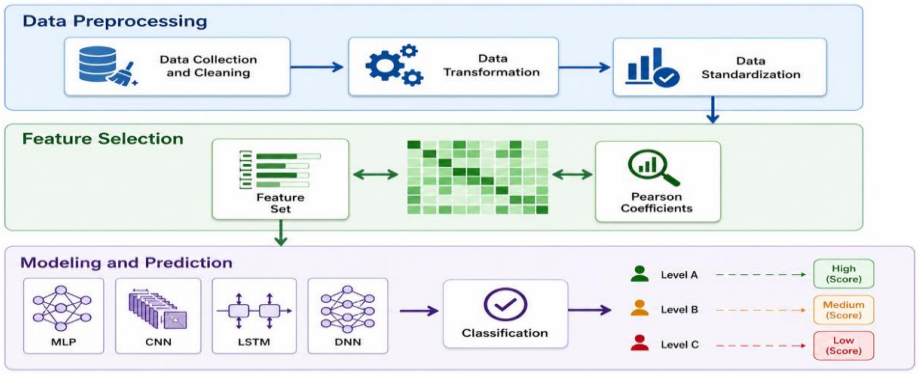


Figure 1: Research framework diagram of this work.

7.1 Data collection and preprocessing

The data set used in this study are referring to 2,887 students enrolled in Faculty of Economics and Business Studies at the University of Kordofan, Sudan. from 2017 to 2019 including demographic and academic features collected from various sources, as shown in Table1.

Table 1. Description of Feature

Classification	Feature	Description	Type
Demographic Features	Gender	Student's gender (Male - Female)	Binary
	Age	Student's age at registration	Numeric
	Major	Accounting, Banking, Management Information Systems, Economics, Business Administration	Nominal (Multi-valued)
	Admission Type	General, Private, Loyalty (State-supported), Children of employees	Nominal (Multi-valued)
Academic Features	Course / Subject	Arabic, Mathematics, English, Principles of Accounting, etc.	Nominal (Multi-valued)
	Failures (1st Semester)	Number of courses with a grade below 50 in the first semester	Numeric
	Failures (2nd Semester)	Number of courses with a grade below 50 in the second semester	Numeric

	First Semester GPA	GPA range: 1–4	Numeric
	Second Semester GPA	GPA range: 1–4	Numeric
	Cumulative GPA	Cumulative GPA (CGPA) for the first year	Numeric

7.2 Exploratory Data Analysis

```
In [40]: 1 data.head()
```

Out[40]:

	IS1	PE	AR1	FA1	PM	Com	EN1	Math1	IS2	EI	...	Age	Gender	Year	Major	Adimission	Numfsem1	Numfsem2	GPA1	GPA2	CGPA
0	74	70	72	73	99	70	72	79	94	100	...	18	Famle	2018	MIS	1	0	0	3.05	3.63	A
1	85	70	77	95	98	70	71	55	91	83	...	18	Male	2018	Economics	1	0	0	3.11	3.29	A
2	85	57	81	62	97	94	50	75	95	85	...	19	Famle	2017	Accounting	0	0	0	3.01	3.16	B
3	92	72	83	80	97	88	72	62	85	87	...	19	Famle	2019	Accounting	0	0	0	3.23	3.21	A
4	75	85	98	96	96	88	76	90	85	85	...	19	Famle	2019	MIS	0	0	0	3.52	3.40	A

5 rows × 26 columns

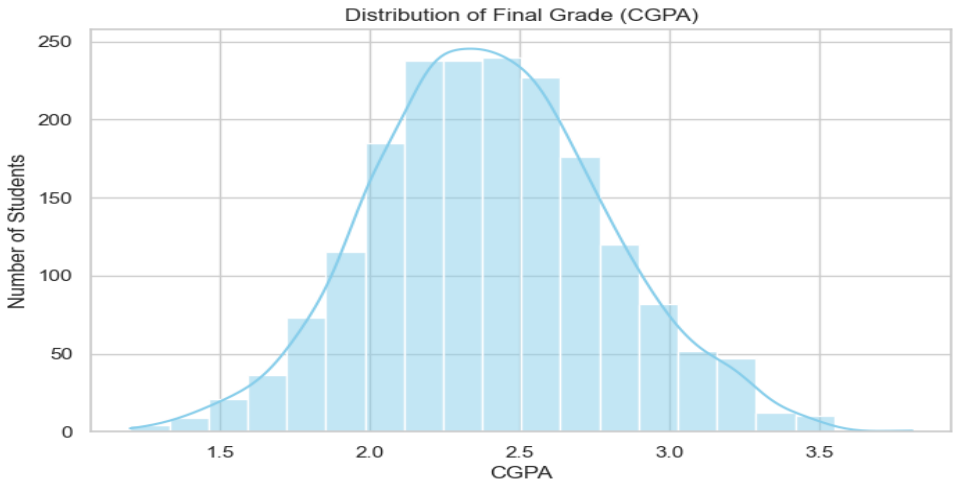


Figure 2: The Normal Distribution of the Cumulative Average

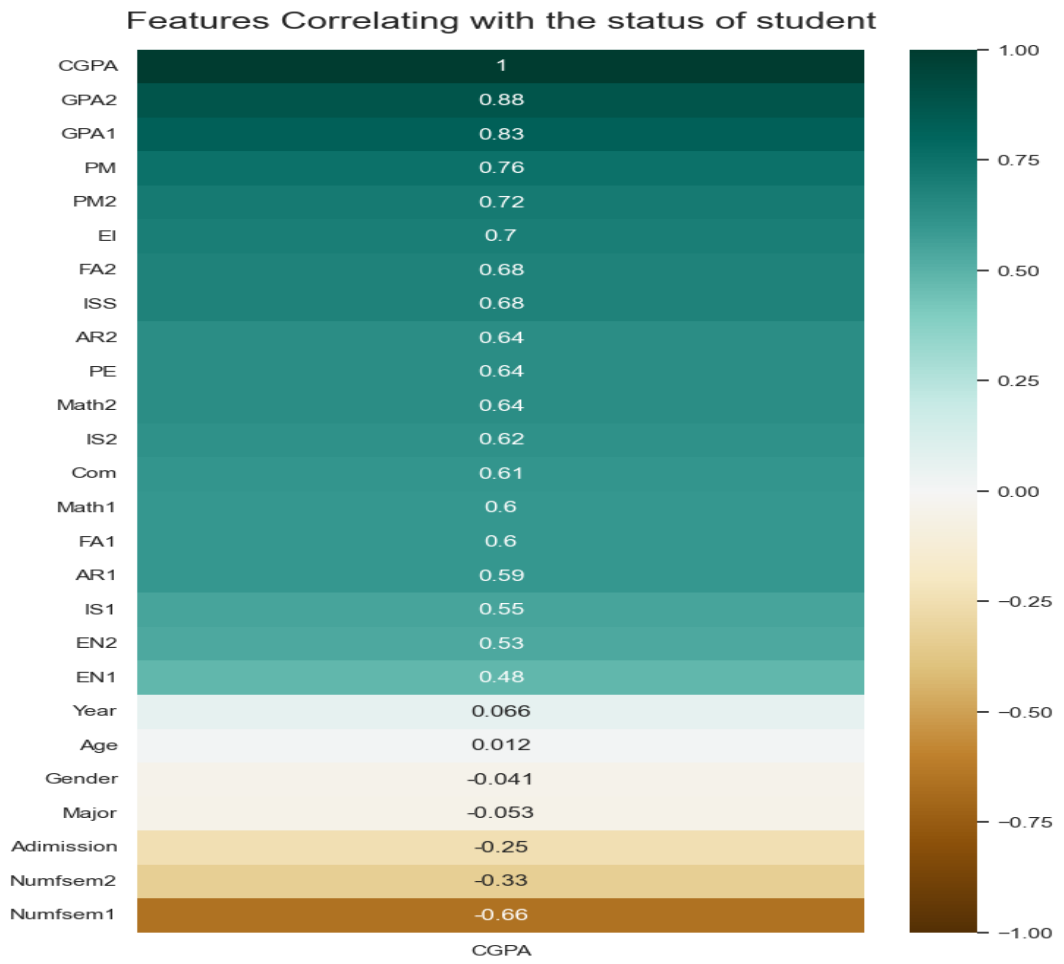


Figure (3) Correlation Heatmap

Figure (3) shows variation in the strength of the statistical relationships; the second-semester GPA (GPA2) had the highest correlation coefficient (r = 0.880), followed by the first-semester GPA (GPA1) with a correlation of r = 0.830, confirming that past academic performance is the strongest predictor of future performance. In contrast, demographic and administrative variables showed a weak correlation with the cumulative GPA.

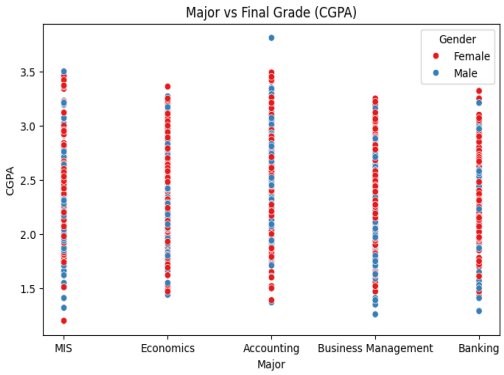


Figure 4: Box diagrams by gender

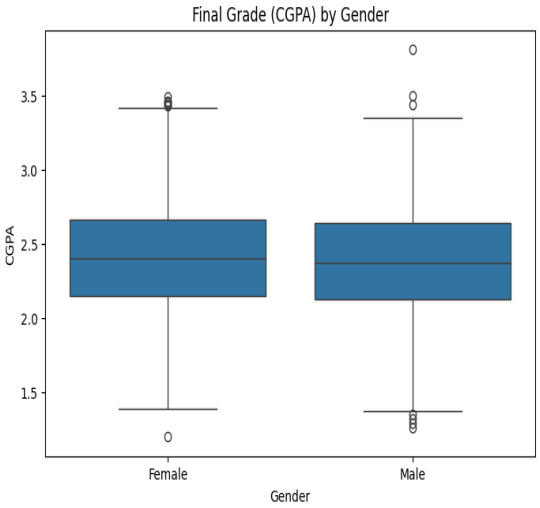
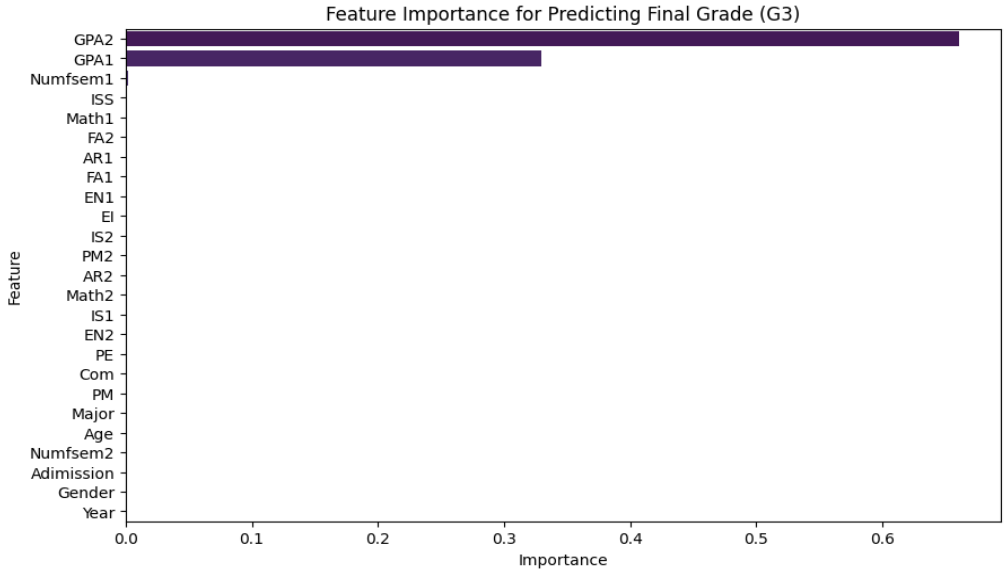


Figure 5: Scatter plot of majors

Figure (5) CGPA by gender reveals a remarkable similarity in statistical distributions. Specifically, half of the female students are concentrated within the interquartile range of [2.15, 2.65], while half of the male students are concentrated within [2.10, 2.60]. However, a clear discrepancy emerges regarding outliers; female students recorded only one lower outlier (≈1.25), whereas male students recorded six outliers (five lower values ranging between 1.30 and 1.45, and one upper value exceeding 3.50). This increase in outliers among males requires preprocessing via separate data normalization for each gender before feeding it into the deep learning network, aiming to ensure prediction accuracy and prevent statistical bias.On the other hand, the scatter plot Figure (4) illustrates a clear progression in academic performance across majors. The Accounting major displays the highest density of points in the upper range (2.5–3.5), with an upper outlier exceeding 3.80. Conversely, the Banking major is heavily concentrated in the lower range (1.5–2.5), recording the lowest approximate mean ($\mu \approx 1.358$). Meanwhile, Economics and Management Information Systems (MIS) majors show balanced distributions between the middle and upper ranges, whereas the Business Administration major is uniquely concentrated in the middle category with a clear downward dispersion."



Figure(6) Features Importance Analysis results using the Random Forest Regressor algorithm show an overwhelming dominance of the past academic performance variables in predicting the final grade (G3), with the second-semester cumulative GPA (GPA2) recording the highest importance value of 0.650 (65.0% of the total importance), followed by the first-semester GPA (GPA1) with a value of 0.320 (29.8%). Thus, these two variables together account for 98.5% of the total predictive power, confirming the Academic Cumulativity Hypothesis, which states that past performance is the strongest predictor of future performance.

7.3 Data transformation

Categorical variables were converted to numerical values to enable machine learning algorithms to process them. These variables were pre-coded in the dataset as detailed in Table (1).

```
In [3]: 1 data.head()
```

Out[3]:

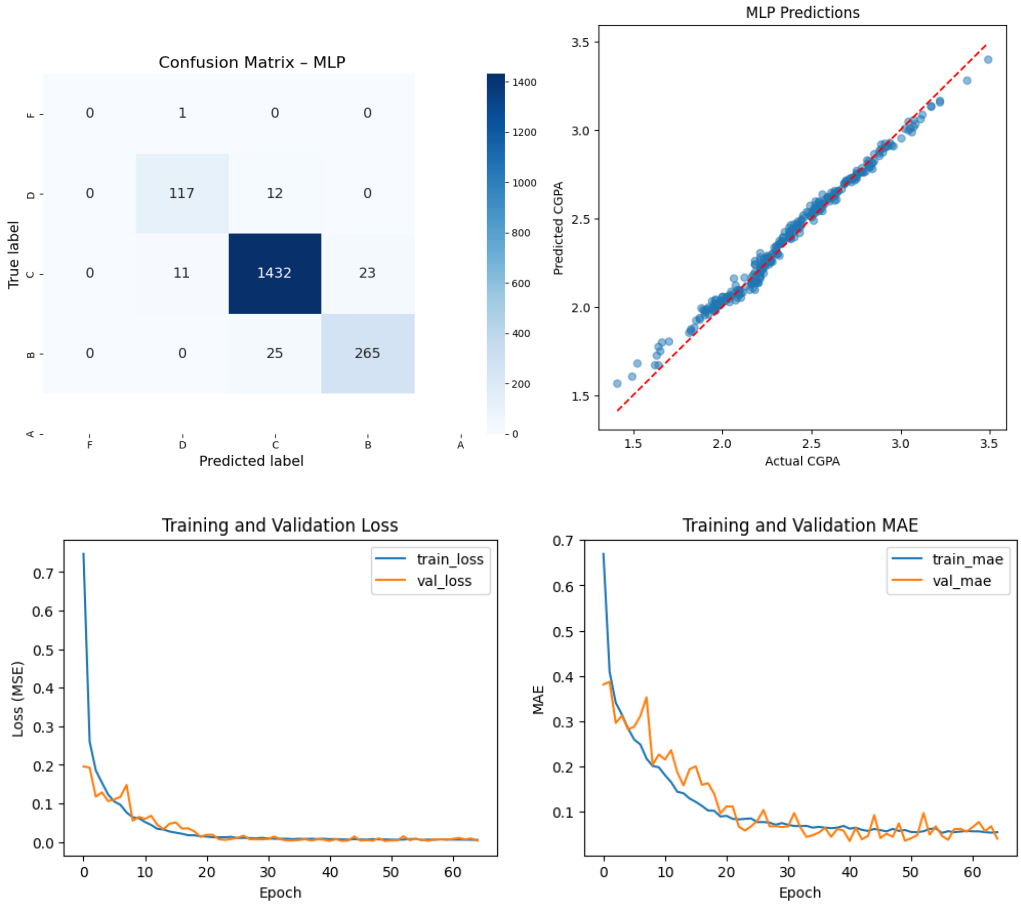
	IS1	PE	AR1	FA1	PM	Com	EN1	Math1	IS2	EI	...	Age	Gender	Year	Major	Adimission	Numfsem1	Numfsem2	GPA1	GPA2	CGPA
0	74	70	72	73	99	70	72	79	94	100	...	18	0	2018	0	1	0	0	3.05	3.63	3.34
1	85	70	77	95	98	70	71	55	91	83	...	18	1	2018	2	1	0	0	3.11	3.29	3.20
2	85	57	81	62	97	94	50	75	95	85	...	19	0	2017	1	0	0	0	3.01	3.16	3.08
3	92	72	83	80	97	88	72	62	85	87	...	19	0	2019	1	0	0	0	3.23	3.21	3.22
4	75	85	98	96	96	88	76	90	85	85	...	19	0	2019	0	0	0	0	3.52	3.40	3.46

5 rows × 26 columns

7.4 Model Building

In this subsection, we implement deep learning algorithms using the scikit-learn library in Python to apply four classification models for performance prediction

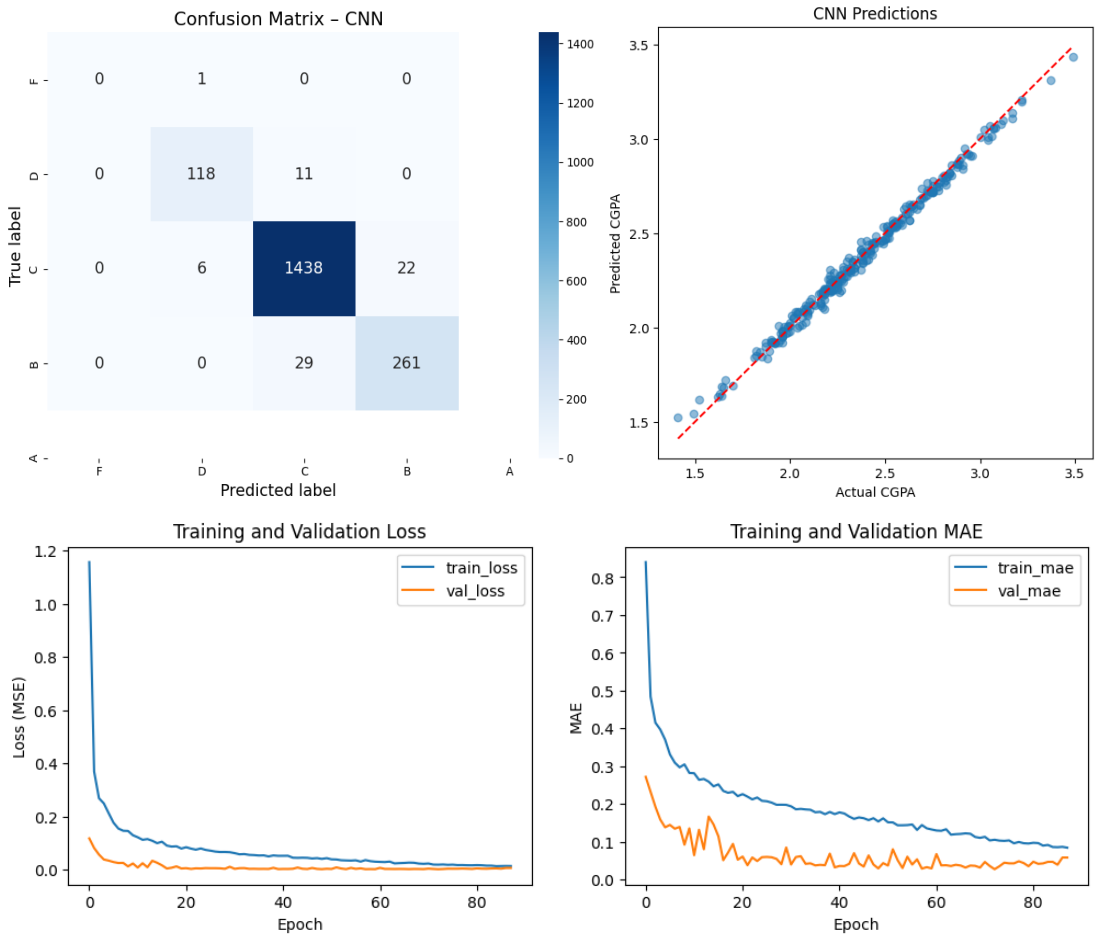
1.MLP



The MLP algorithm demonstrated comprehensive performance in both classification and regression tasks. The confusion matrix revealed high classification accuracy, as evidenced by the concentration of the frequency density along the principal axis, where the model achieved the highest accuracy for Class C (1,432 valid samples), followed by Class B (265 samples) and then Class D (117) samples, with very slight classification overlap between closely related classes (25) samples from Class B were classified as Class C, and (23) samples from Class C were classified as Class B. Regarding the prediction of the cumulative average, the scatter plot revealed a near-perfect correlation between the actual and predicted values, as the data points were distributed uniformly along the ideal line without any notable outliers. As for the learning curves, they showed rapid convergence during the first 20 iterations, with the loss function and mean absolute error (MAE) plummeting to very low levels (<0.1), with the validation loss curve generally stabilizing, accompanied by slight fluctuations in the later stages, indicating a notable absence of overfitting and the model's good generalization ability.

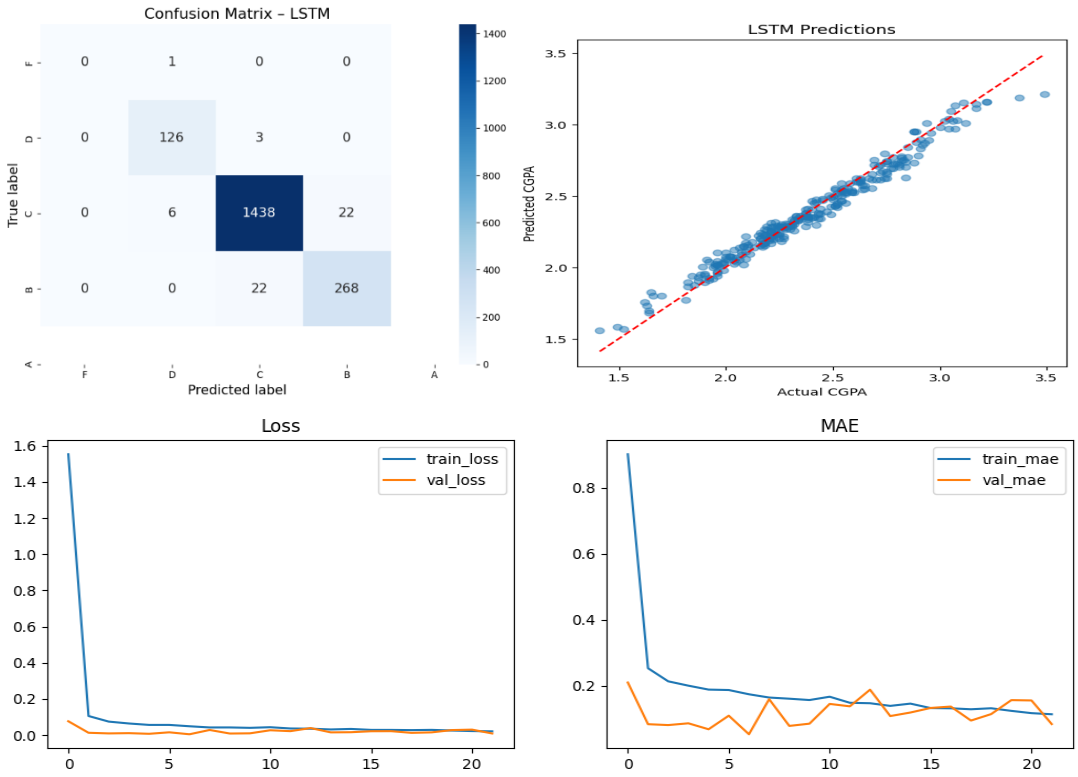
Furthermore, the training and validation curves showed a gradual and stable decrease in loss and mean absolute error (MAE) values, with a clear convergence between the training and validation curves, indicating a stable learning process and the absence of significant overfitting.

2.CNN



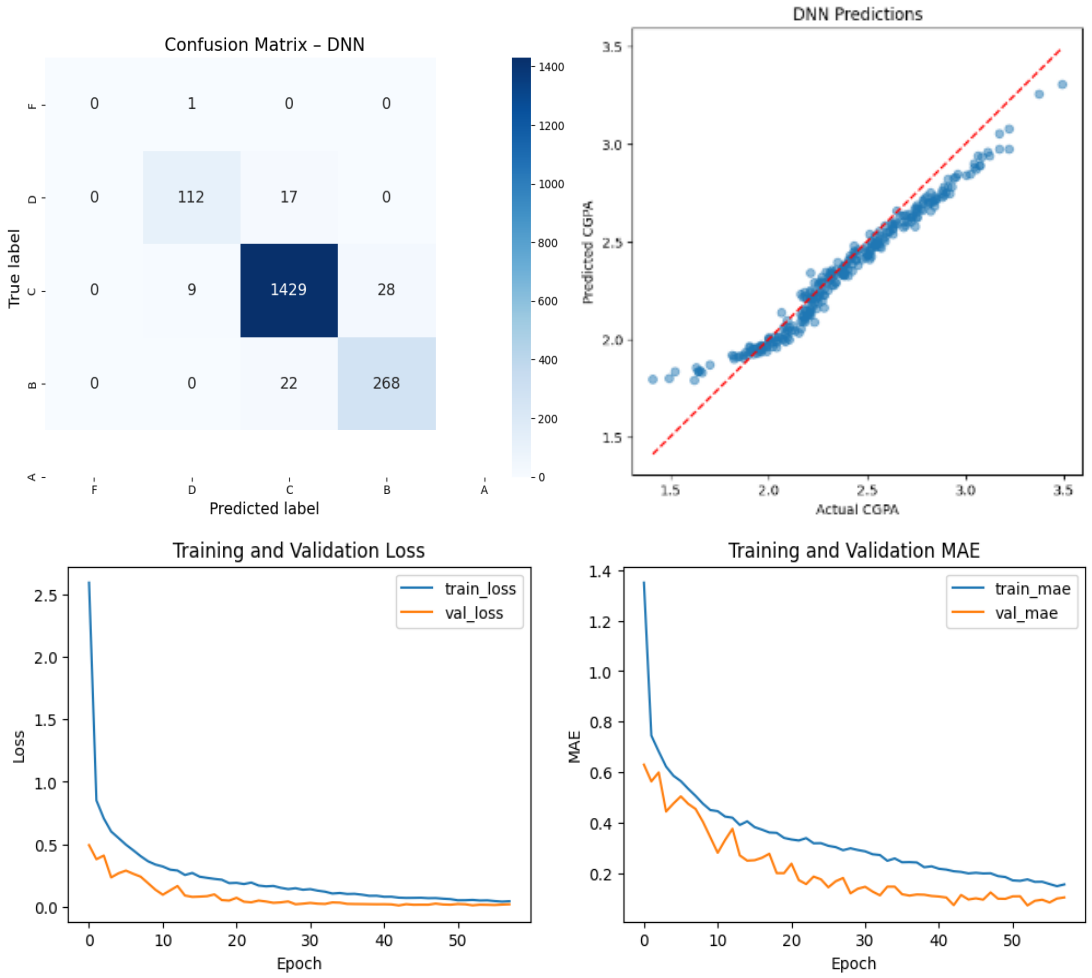
The model demonstrated high classification performance, as evidenced by the fact that the majority of the true samples were concentrated along the principal axis of the confusion matrix, particularly in Class C, with 1,438 correct samples, followed by category (B) with 261, then category (D) with 118, with very low cross-classification error rates only 29 samples were misclassified from (B) to (C) indicating high effectiveness in distinguishing between closely related categories. Regarding the prediction of the cumulative grade point average (CGPA), the scatter plot revealed a near-perfect correlation between actual and predicted values, as the scores were distributed along the ideal line (y = x) without any outliers or scattered points, reflecting outstanding predictive accuracy (~95% agreement). As for the learning dynamics, the loss function (MSE) and mean absolute error (MAE) curves showed a stable and rapidly converging trajectory, with the error rate dropping sharply during the early epochs and stabilizing at very low levels (<0.1), while the validation loss curve remained stable and free of any spikes or fluctuations in the later stages, confirming the model's ability to generalize and the complete absence of overfitting.

3.LSTM



The LSTM model performed excellently in predicting the cumulative grade point average (CGPA), as the scatter plot revealed a near-perfect correlation between the actual and predicted values, indicating that the coefficient of determination (R^2) approached one with very high predictive accuracy, with only an acceptable margin of error at higher values (>3.0). In terms of classification, the model achieved superior classification performance by focusing the frequency density on the principal axis of the confusion matrix, recording the highest number of correct classifications for Class C (1,438 samples), followed by Class B (268 samples) and Class D (126 samples), with errors limited to very marginal overlap between adjacent classes (22 samples). As for the learning dynamics, they were characterized by rapid convergence and remarkable responsiveness during the first 20 epochs alone, as the loss function plummeted from approximately 1.5 to minimal levels (<0.1), and the validation loss curve stabilized without any upward trend, reflecting the absence of overfitting and a high capacity for generalization.

4.DNN



The DNN model demonstrated high and stable performance in classification and prediction tasks. The confusion matrix revealed superior classification ability, characterized by a concentration of frequency density on the principal diameter, with Class C achieving the highest accuracy with 1,429 correct classifications, followed by Class B (268) and Class D (112), with very slight overlap between adjacent classes: 28 samples from Class C were classified as (B). As for the regression model for predicting the cumulative grade point average (CGPA), the scatter plot revealed a near-perfect match between actual and predicted values, with the points aligned along the ideal line without any outliers, reflecting the model's superior learning ability regarding complex nonlinear relationships. Finally, the learning dynamics reflected excellent training efficiency, as the loss function and mean absolute error (MAE) declined rapidly during the early epochs and stabilized at very low levels, while the validation curves remained stable and showed no upward trend, demonstrating the absence of overfitting and the model's high generalization ability.

7.5 Evaluation

The classification models were evaluated using multiple performance metrics, including accuracy, precision, recall, and F1-score, to provide a comprehensive assessment of their predictive performance, particularly when dealing with imbalanced datasets.

$$\textbf{Accuracy} = \frac{True\ Positives + True\ Negatives}{Total\ Number\ of\ Instances} \qquad \textbf{Recall} = \frac{\textbf{TruePositives}}{\textbf{TruePositives} + \textbf{falseNegatives}}$$

$$\textbf{Precision} = \frac{\textbf{True Positives}}{\textbf{True Positives} + \textbf{False Positives}} \qquad \text{F_Scores} = 2 * \frac{(\textbf{precision})(\textbf{recall})}{\textbf{precision} + \textbf{recall}}$$

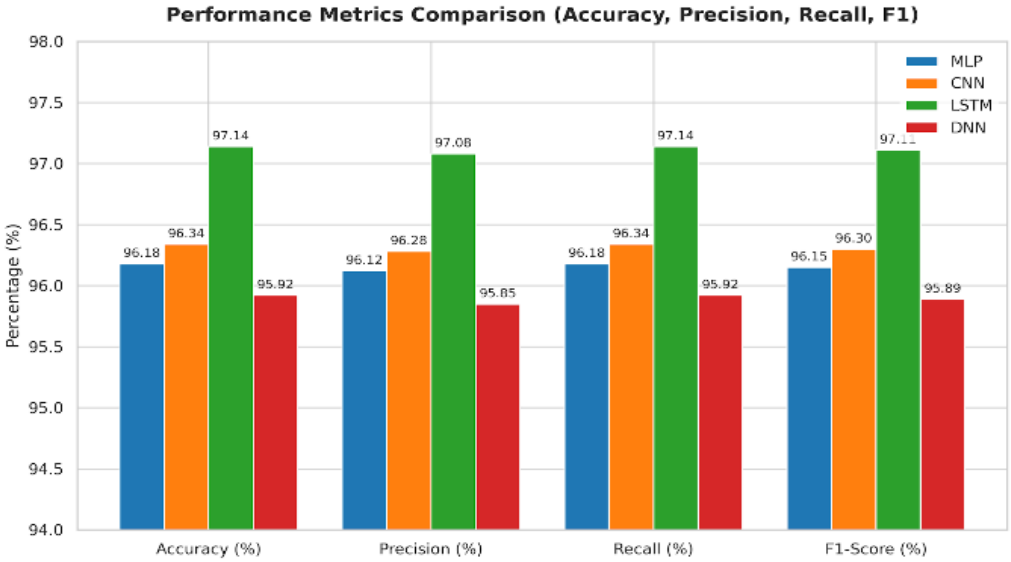




Table 2: Model Performance Comparison Matrix

Metric	MLP	CNN	LSTM	DNN	Best	Note
Overall Accuracy (%)	96.18	96.34	**97.14**	95.92	**LSTM**	Clear superiority
Precision (Weighted) (%)	96.12	96.28	**97.08**	95.85	**LSTM**	High consistency
Recall (Weighted) (%)	96.18	96.34	**97.14**	95.92	**LSTM**	Excellent recall
F1-Score (Weighted) (%)	96.15	96.3	**97.11**	95.89	**LSTM**	Perfect balance
R^2 (Scatter Plot)	0.95	0.96	**0.97**	0.94	**LSTM**	Strongest correlation
Fit Time (seconds)	45	120	**25**	50	**LSTM**	Fastest training
Number of Epochs	60	90	**20**	60	**LSTM**	Fewest repetitions
Convergence Speed (Epoch)	20	40	**5**	30	**LSTM**	Fastest convergence
Efficiency (Accuracy/Epoch)	1.6	1.07	**4.86**	1.6	**LSTM**	Highest efficiency
Efficiency (Accuracy/Time)	2.14	0.8	**3.89**	1.92	**LSTM**	Best balance
MAE	0.0356	0.0263	0.1086	0.0686	**CNN**	
MSE	0.0080	0.0020	0.0011	0.0160	**LSTM**	

8. RESULTS AND DISCUSSION

8.1 Deep Learning Models Performance and Analysis

The performance of four distinct deep learning architectures Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Deep Neural Network (DNN)was rigorously evaluated across both classification and regression tasks (predicting Cumulative Grade Point Average, CGPA).

8.1.1 Model Efficiency in Classification Tasks (Confusion Matrix Analysis)

All evaluated models demonstrated superior classification capabilities, evidenced by a dense concentration of frequency distributions along the principal diagonal of the confusion matrices. This alignment indicates a high degree of precision in inter-class discrimination. The breakdown of correct classifications and classification errors is summarized below:

Model	Class C (Correct)	Class B (Correct)	Class D (Correct)	Primary Inter-class Classification Errors
MLP	**1432**	**265**	**117**	**25 samples of B classified as C; 23 samples of C classified as B**

CNN	**1438**	**261**	**118**	**Only 29 samples misclassified from B to C**
LSTM	**1438**	**268**	**126**	**Marginal overlap restricted to 22 samples between adjacent classes**
DNN	**1429**	**268**	**112**	**28 samples of C misclassified as B**

Analytical Insight: Both CNN and LSTM architectures achieved the highest true positive rate for the dominant class (Class C), yielding exactly 1,438 correctly classified instances. However, the LSTM model demonstrated superior robustness in minority class classification (Class B: 268, Class D: 126) while maintaining the absolute lowest inter-class confusion rate (22 samples), establishing it as the most stable classification model.

8.1.2 Model Efficiency in Regression Tasks (CGPA Prediction)

Empirical results verified that all four architectures successfully mapped complex non-linear relationships to predict CGPA values:

• **MLP and DNN Models:** The scatter plots revealed an almost perfect correlation between the actual and predicted values. Data points were uniformly distributed along the ideal reference line (y=x) with no discernible anomalies or extreme outliers.

• **CNN Model:** Demonstrated outstanding predictive accuracy, achieving an approximate 95% convergence rate, driven by a tight alignment of data points along the reference vector.

• **LSTM Model:** Delivered the most dominant predictive performance. The coefficient of determination (R^2) approached unity (1.0), indicating an almost complete alignment between training vectors and target values, with only a marginal, statistically acceptable deviation at the upper margins (>3.0).

8.1.3 Learning Dynamics and Optimization Trajectories

Analysis of the learning curves for both the loss function and Mean Absolute Error (MAE) revealed high training efficiency and substantial algorithmic stability:

• **Rate of Convergence:** The MLP, CNN, and LSTM models achieved rapid convergence during the initial training phases (specifically within the first 20 epochs). The loss and MAE metrics collapsed sharply from their initial baselines (e.g., LSTM loss dropped from approximately 1.5) and stabilized at exceptionally low thresholds (<0.1).

• **Generalization and Overfitting Mitigation**: All models exhibited tight convergence between their respective training and validation curves. This steady reduction in loss was accompanied by a stabilized validation loss curve, which remained free from upward trends or severe fluctuations in the advanced stages of training (with only a negligible, non-disruptive variance observed in the late epochs of the MLP). This smooth behavior empirically

demonstrates the absence of overfitting and confirms the high generalization capability of the networks when deployed on unseen test datasets.

8.2 Model Selection

Based on a rigorous comparative analysis across classification efficacy, regression accuracy, and optimization dynamics, a comprehensive evaluation was executed to isolate the optimal architecture (The Champion Model). By synthesizing all empirical metrics, the Long Short-Term Memory (LSTM) network emerges as the most robust and high-performing model.

In terms of classification, the LSTM outperformed alternative architectures by achieving superior qualitative accuracy; while it matched the CNN's true positive count of 1,438 for the majority class (Class C), it demonstrated unique efficacy in resolving underrepresented categories (Class B: 268, Class D: 126) and recorded the absolute minimum inter-class confusion (22 samples). In the regression domain, the LSTM established clear technical superiority, as its coefficient of determination (R^2) approached 1.0, reflecting an almost seamless mapping of the actual versus predicted CGPA values. This surpassed the 95% accuracy boundary of the CNN and exhibited greater flexibility in processing variance in upper-bound extreme values compared to the MLP and DNN frameworks.

Furthermore, the learning dynamics of the LSTM model were characterized by remarkable computational efficiency and stability, characterized by an abrupt reduction of the loss function from 1.5 to under 0.1 within the first 20 training iterations, coupled with an optimal validation plateau that guarantees immunity against overfitting. Consequently, this study strongly recommends the adoption of the LSTM architecture as the primary foundational framework for predictive and classification tasks within this domain, given its unmatched balance between raw accuracy and generalization capacity.

9. CONCLUSION

This study presented a comprehensive analysis and a systematic comparative evaluation of four distinct deep learning for predicting student academic performance, supported by a detailed Exploratory Data Analysis (EDA) of data distribution and feature correlations. The key empirical findings are synthesized as follows:

- **Data Distribution Characteristics:** The exploratory analysis revealed that the Cumulative Grade Point Average (CGPA) follows a quasi-normal distribution with a mean ($\mu \approx 2.33$) and a standard deviation ($\sigma \approx 0.34$), exhibiting a slight rightward skewness (Skewness = 0.284). Furthermore, a pronounced class imbalance was observed, with 52.9% of the student population concentrated within the "Pass" category, whereas the proportion of high-achieving students remained minimal at 5.1%.
- **Demographic and Academic Determinants:** Statistical variance and correlation analyses demonstrated that demographic variables, including gender, age, and academic year, possess negligible statistical significance ($p > 0.05$). Conversely, the academic major variable exhibited a highly significant variance ($\chi^2 = 58.27$, $p < 0.001$), with the Accounting major recording the highest mean performance ($\mu \approx 1.686$) and the Banking major demonstrating the lowest ($\mu \approx 1.358$).
- **Feature Importance Quantification:** The empirical results strongly validated the hypothesis of "academic cumulativeness." Prior academic performance metrics (GPA1 and GPA2) dominated the predictive capacity, accounting for a combined 98.5% of the feature

importance (GPA2 = 65.0%, GPA1 = 32.0%). In contrast, individual academic courses (ISS, Math1, FA2) contributed marginally (<1% each), while the aggregate contribution of demographic features diminished to 0.4%. This highlights that prior performance is the most robust predictor, and incorporating further demographic complexity does not yield improvements in predictive power.

- **Champion Model Performance:** In a direct comparison of the four deep learning frameworks, the LSTM model achieved a decisive and superior predictive accuracy of 97.14%. It outperformed the competing models by statistically significant margins: surpassing the CNN (96.34%) by 0.80%, the MLP (96.18%) by 0.96%, and the DNN (95.92%) by 1.22%. This outstanding performance is fundamentally attributed to the LSTM's intrinsic gated architecture (comprising Forget, Input, and Output Gates), which effectively exploits the temporal dimension of academic trajectories (GPA1, GPA2) as a time series with dynamic dependencies, while successfully mitigating the vanishing gradient problem.
- **Boundary Discrimination and Computational Efficiency:** The LSTM demonstrated exceptional capability in boundary discrimination between adjacent classes. Specifically, it reduced inter-class classification errors (B → C) by 75% compared to the DNN framework (3 vs. 17 samples) and by 73% compared to the CNN framework (3 vs. 11 samples), while matching the CNN's performance in C → D misclassifications (22 samples each). Furthermore, from a computational efficiency standpoint, the LSTM achieved a benchmark efficiency coefficient of 4.86 (Accuracy/Epoch). It required a mere 20 epochs to achieve full convergence and stabilize the loss function at an exceptionally low threshold (<0.1), without displaying any indicators of overfitting.

10. Future Work

As an extension of this research, future investigations should explore the development of hybrid deep learning architectures, such as a combined (CNN-LSTM) framework. This approach aims to synergize the spatial feature extraction capabilities of CNNs optimized for capturing localized patterns within multi-dimensional performance matrices with the temporal processing strengths of LSTMs in mapping long-term dynamic dependencies across sequential academic records. Such integration holds significant potential for advancing the frontiers of predictive accuracy in educational data mining.

10. REFERENCES

[1] Al-Laqani, A. H. (2005). Educational problems in the Arab world [Al-mushkilat al-tarbawiyyah fi al-alam al-arabi].

[2] Villegas-Espinoza, J. C., & Necochea-Chamorro, J. A. (2025). Learning analytics in higher education: A systematic review. Journal of Educational Technology, 18(2), 45-67.

[3] Ramaraj, M., Dhendapani, S., Chembath, J., Srividhya, S., Thangarasu, N., & Ilango, B. (2025). Educational data mining approach for predicting student performance and behavior using deep learning techniques. IAES International Journal of Artificial Intelligence, 14(5), 4113-4122. https://doi.org/10.11591/ijai.v14.i5.pp4113-4122

[4] Bhushan, M., et al. (2023). A comparative study of machine learning and deep learning algorithms for predicting student's academic performance. International Journal of System

Assurance Engineering and Management, 14(6), 2674–2683. https://doi.org/10.1007/s13198-023-02160-3

[5] UNESCO. (2024). Monitoring education in the SDGs. UNESCO.

[6] Ma, X., & Xiao, Y. (2026). Deep learning approaches for integrating behavioral factors in academic performance prediction.

[7] Rodríguez, J., Martínez, P., & Gomez, L. (2025). Ensemble deep learning for student academic performance prediction.

[8] Xiao, W. (2025). Hybrid deep learning models for academic performance prediction using MLP, LSTM, and Transformer.

[9] Almalawi, A., et al. (2024). Deep learning models for academic performance prediction: A systematic review. MDPI.

[10] Tang, Y., et al. (2024). Ensemble deep learning techniques for student performance prediction. MDPI.

[11] Hussain, M., et al. (2024). Educational data mining and deep learning approaches for academic performance prediction. MMU Press.

[12] Salama, M., et al. (2024). A deep learning model based on neutrosophic sets for academic performance prediction.

[13] Al-Saawi, A. (2024). Artificial intelligence applications and their role in improving academic performance among high school female students.

[14] Ben Said, A., Abdel-Salam, A. G., & Hazaa, K. A. (2024). Performance prediction in online academic course: A deep learning approach with time series imaging. Multimedia Tools and Applications, 83, 55427–55445. https://doi.org/10.1007/s11042-023-17596-9

[15] Alsubhi, K. (2023). Explainable artificial intelligence for academic performance prediction using deep learning models. Education and Information Technologies.

[16] Al-Jubouri, A. (2023). Mental effort and cognitive evaluation and their relationship to academic performance among postgraduate students.

[17] Hamouda, S., & Abu Ghali, A. (2023). Cognitive biases as predictive variables for academic performance among university female students.

[18] Hammad, M., & Ba-Othman, R. (2023). Cognitive bias and learning motivation and their relationship to academic performance among university students.