International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. V, No. 04, pp. 24-30, April

Student Academic Performance Prediction Using Hybrid Mamba-Transformer Network with Adaptive Attention Learning

Ramya. R[4]

[4]Assistant Professor, Department of Electronics and Communication Engineering,
Arunachala College of Engineering for Women, Tamil Nadu, India.

ABSTRACT: The prediction of student performance has become an essential component for online educational platforms, as it improves student engagement in learning, supports adaptive learning strategies and enables timely academic interventions. The diverse data sources provide valuable insights into student behaviour and academic progress, making virtual learning environments a rich and promising field for educational research and predictive analytics. Each learning activity record consists of two categories of features: student behavioural features and exercise-related features. In this analysis, it is acknowledged that each factor has a distinct effect on student performance and combination of activity with behaviour among students features. It is crucial for enhancing intelligence prediction reliability among students for providing better options in their employments. In this study, raw educational data is transformed to structured and deep machine learning format through pre-processing unit that helps in achieving robustness and better accuracy, while data visualization paves way in selecting patterns and variables relationships before training the model. Then the processed data is splitted for training, validation and testing which in turn selection of data takes place and fed into classifier model for performance prediction. Mamba transformer network is employed to improve quality of image data, using multi-layer processing recognition accuracy is improved. Use of deep learning, which is classified under artificial intelligence (AI), is crucial to the field of performance prediction of intelligent students in universities. Therefore, this study develops a Mamba transformer model along with adaptive attention mechanism to find out performance of students based on activity-related data and student behaviours, which performs with better accuracy of 97.3%, precision of 97% and its F1 score value is 96% than previous integrative models.

Keywords: Artificial intelligence (AI), Hybrid Mamba-Transformer Network, Adaptive Attention Learning, Student performance prediction.

1. Introduction

In our day-to-day life, education plays a major role but determining students performance in their academics is a challenge because of pandemic situation faced during past years that led a practical knowledge crisis among students and unemployment. Predicting performance of student has become a fundamental task in advancement of online learning platforms, as it enables accurate forecasting of students' ability to correctly answer future assessment questions based on their learning progress and historical performance [1-3]. The rapid expansion of Internet-based educational resources has led to the widespread emergence of online learning platforms. These platforms offer students access to diverse learning resources, encouraging active participation in learning process and facilitating the development of a more comprehensive and well-structured knowledge base. Performance of students in their academic is widely recognized as a crucial benchmark that helps in noticing quality of education provided by higher education institution. Recognizing students who are likely to have poor success in their academics should be mitigated immediately, so that in future impacts could be reduced by providing better education [4-6]. Since there is a need of larger models and increase in dataset, transformer model is required along with learning techniques. A neural network such as Convolutional Neural network (CNN), Long Short-Term Memory (LSTM) along with transformer model is proposed in [7], that improves learning of discrete features to predict student performance but processing data requires more storage and it is less suitable for constrained environments with computational complexity. A novel Dynamic Key-Value Memory Networks Knowledge Tracing (DKVMN-KAPS) model is proposed in [8], that predicts students future using analysis of question-answering interaction. Moreover, this model has high computational complexity with data quality sensitivity. Deep learning methods like Support Vector Machines, Random forests have limited amount to capture contextual relationship among students academics and self-attention mechanism has limited capability of extracting information related to large-scale educational dataset.. Hence, in this work, a novel method that integrates transformer network model and attention learning mechanism for better outcomes in predicting students performance is proposed [9, 10]

2. Recent Works

Sukrit Leelaluk et al [2024] [11] have proposed a novel method using Attention-Based Artificial Neural Network (Attn-ANN) in predicting academic performance of students. Attn-ANN model employs attention mechanisms on both temporal and feature dimensions to assess the importance of lectures and learning activities, facilitating accurate predictions and interpretable insights through attention weight visualization and it supports in predicting students at-risk. Despite the advantages, practical deployment and scalability under resource-constrained environments are not obtained in this approach.

Yupei Zhang et al [2023] [12] have considered a multi-source sparse attention convolutional neural network (MsaCNN**)** that provides information about grades in students course for actual formulation. Results obtained from real-world university datasets shows that MsaCNN surpasses traditional methods in predictive performance and offers meaningful interpretations of student performance by learned relationships among courses. Although this approach faces scalability issues, lack of personalization validation, limited generalizability and limited factors are considered.

Daozong Sun et al [2023] [13] have proposed a method to analyse students academic performance in universities using attention mechanism and multi-feature fusion. This method uses historical academic grades from different perspective such as student and courses, depending on different courses or even among students. The accuracy rate of this model is about 72.5%, which is more when compared to traditional machine learning techniques. On the other hand, interpretability among instructors or administrators are not easy, model depends on historical grades and faces scalability issues.

Han Wan et al [2026] [14] have proposed a blended learning methodology to predict learning activities of students using multi model dataset. This work undergoes with dual stage training model that depends on transformer for temporal attention based multimodal fusion, which provides better prediction than unimodal technique. It helps in improving precise application of instructional interventions by enriching academic outcomes of students. Data dependency, variations in educational settings and privacy concerns further limits generalizability and scalability of this approach.

Yu Liu et al [2023] [15] have proposed a method based on attention mechanism and deep learning (MCAG) for predicting academic performance with more accuracy than other baseline methods. It captures relationships among various factors and identifies the factors that impacts outcome of results. This approach identifies correlation among variables and it does not provide necessary information of factors that affects outcomes which in turn limits interpretability for educators.

The benefaction of this paper is listed below:

a) A novel structure combine Mamba-transformer network with adaptive attention learning is proposed for predicting intelligent students in their academics.

b) Data pre-processing takes place to convert raw educational data to deep learning language.

c) Classifier model performs with greater accuracy, precision, recall and F1 measure rate than other baseline models.

The methodology of proposed work is explained in the following section.

3. Proposed Work Explanation

In this methodology, raw data is provided as input to the processing unit in which one-hot encoder acts as preprocessor that helps in converting obtained variables in categories into binary columns (0s and 1s). Block diagram of the methodology is shown in Figure 1.

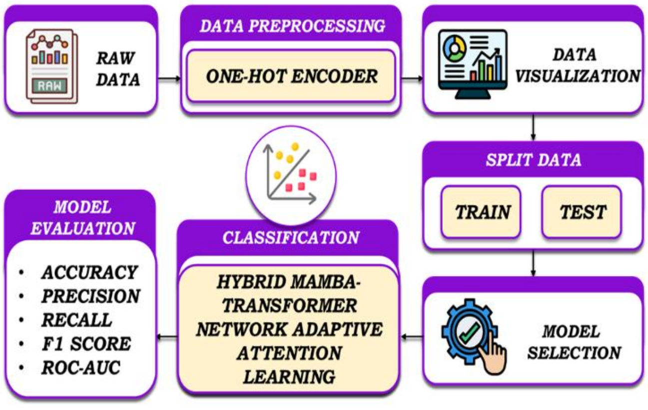

Figure 1: *Block diagram of proposed work*

After preprocessing the data, removal of noise and identifying missing values of data takes place in data visualization unit. Further data are splitted for training and testing which in turn output is subjected for selection process, then classification of data is performed. The hybrid classifier model combined with deep learning is explained as follows.

3.1 Hybrid Mamba-transformer Network Adaptive attention learning

Computational capabilities of self-attention learning mechanism is complex that produce hierarchical structures are limited. Hence, hybrid model is employed to classify input dataset using transformer network along with adaptive-attention learning techniques, that aids in enhancement of outcomes with better accuracy than other baseline models. Mamba transformer structure is built on state space models (SSMs), that are class of sequence models stimulated with deep learning techniques for enhancing performance of classifiers. In SSMs, input sequence $x(t)$ is mapped to output sequence $y(t)$ with latent state $z(t)$, which has high dimension governed by equations (1) & (2) are given below:

$$z(t) = Az(t-1) + Ax(t) \tag{1}$$

$$y(t) = Cz(t) \tag{2}$$

Where learnable parameter matrices is denotes as A, B and C. This sequence is performed in two divisions:

Discretization: Using discretization process such as zero-order hold, continuous time parameters (Δ, A, B) are converted into discrete components (A, B) as governed in equations (3) and (4),

$$A = \exp(\Delta A) \tag{3}$$

$$B = (\Delta A)^{-1}(exp(\Delta A) - 1)B \tag{4}$$

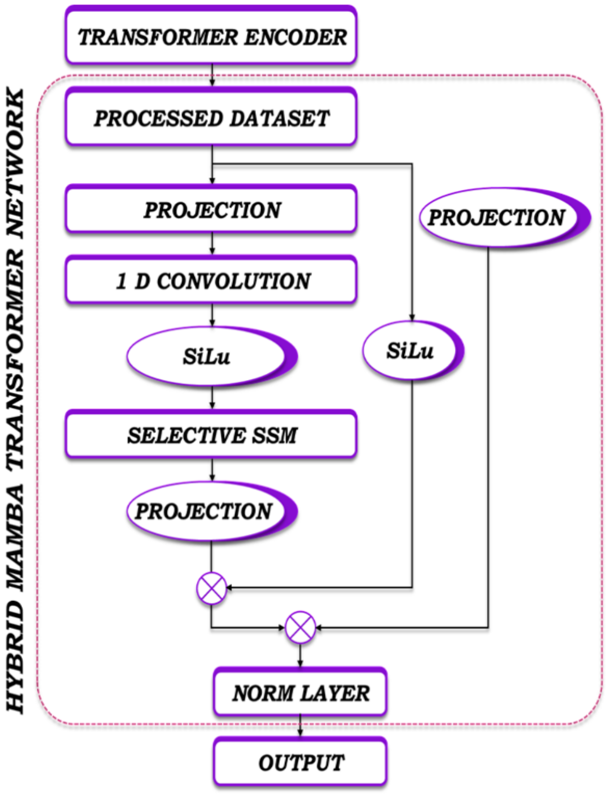

Figure 2: *Structural flow chart of hybrid Mamba transformer network*

Update of state and output: Updating state vector $z(t)$ that depends input $x(t)$and previous state $z(t-1)$. Output $y(t)$ is obtained by projection of state through C as equation (2). Flow chart of proposed network is shown in Figure 2.

In Mamba block, processed dataset input enters into deeper projection network and 1D convolutional layer, which is then processed by SSMs that detects state transition matrix A, input-dependent matrices B and C, finally output matrix D. Then output is passed to projection layer to yield result of mamba block and classified model is validated to provide expected performance metrics.

3.2 Mathematical Expressions and Symbols

In this section, the mathematical equations are provided for better evaluation of model classification of prediction. Higher value of accuracy, defined as ratio of correctly classified data to summation of samples in dataset, provides perfection in predicting performance. Precision denotes ratio of positive samples that are perfectly categorized to summation of positive samples predicted by model. F1-Measure based on recall and precision. Equation (5) to (8) represents accuracy, precision, recall and F1 measure value respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (5)$$

$$Precision = \frac{\mathrm{TP}}{TP+FP} \quad (6)$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (7)$$

$$F1 - Measure = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (8)$$

Where TN denotes samples which are truly negative but not positive, TP indicates data that are truly positive and specified as positive, FN represents data that are truly positive but are not properly specified as positive, FP indicates samples that are truly negative but carelessly denoted as positive.

4. Results and Discussion

In this portion, performance predicting proposed model is signified through experiments. Access of dataset is provided using link https://www.kaggle.com/datasets/mahwiz/students-dropout-and-academic-success-dataset. The dataset shape contains 4424 rows and 37 columns. Figure 3 to 6 provides detail about performance of classifiers.

Details about students distribution in various categories is depicted in Figure 3. In target distribution 3(a), count is calculated among dropout, graduate and enrolled students. In 3(b), admission grade distribution is plotted between frequency and admission grade of students, scholarship holder distribution is shown in 3(c) in which students who are receiving and not getting scholarship are indicated, age at enrolment density distribution is depicted in 3(d) that details about density vs age at enrolment.

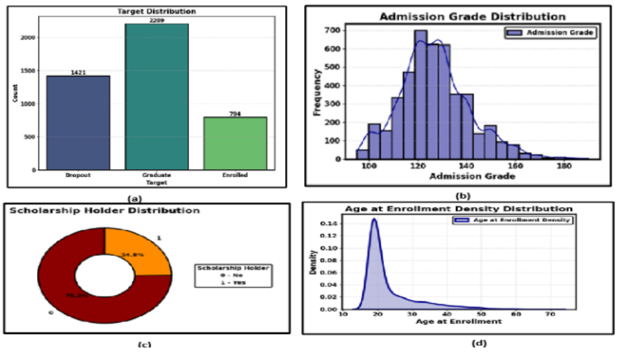

Figure 3: *Comparison of (a) Target distribution (b)Admission Grade distribution (c) Scholarship holder distribution (d) Age at enrolment density distribution*

In Figure 4(a), economic indicators like GDP, inflation rate and unemployment rates are indicated based on cumulative probability. In Figure 4(b) gender distribution such as 35.2% male and 64.8% female are indicated. Admission grade vs age is indicated in 4(c) based on performance of students. Top 10 course in universities are listed in Figure 4(d) based on students enrolment.

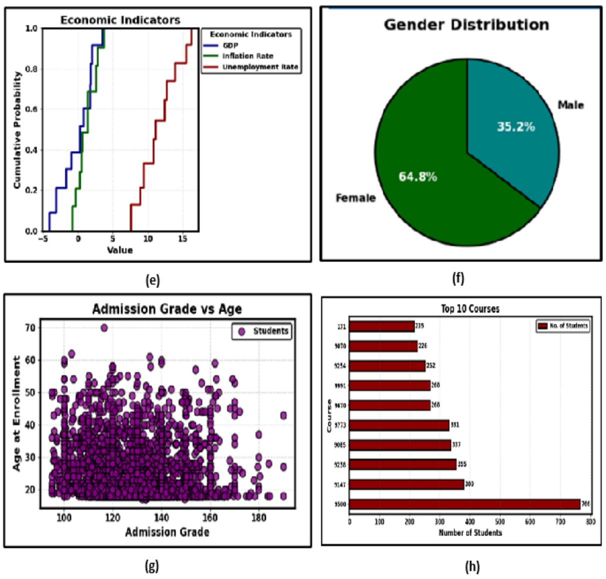

Figure 4: *Comparison on (a) Economic indicators (b) Gender distribution (c) Admission grade vs age (d) Top 10 courses*

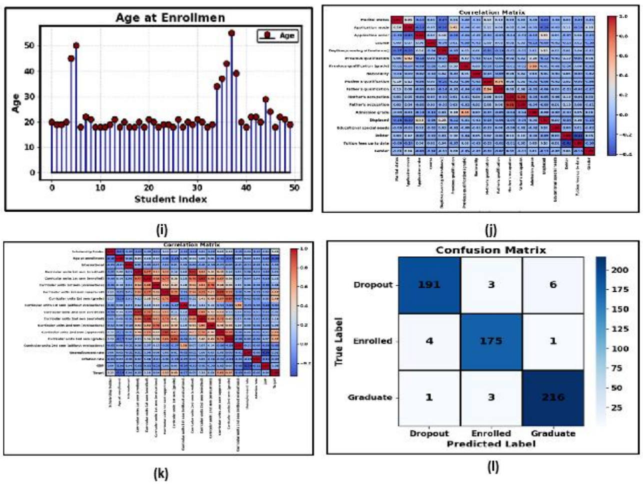

Figure 5: *Comparison on (a) Age at enrolment (b) Correlation matrix (c) Correlation matrix (d) confusion matrix*

Age at enrolment is plotted in Figure 5(a) based on age vs students index, correlation matrix are shown in Figure 5(b) and 5(c) based on various factors are represented, confusion matrix is shown in Figure 5(d) which is based on true label vs predicted label is indicated effectively.

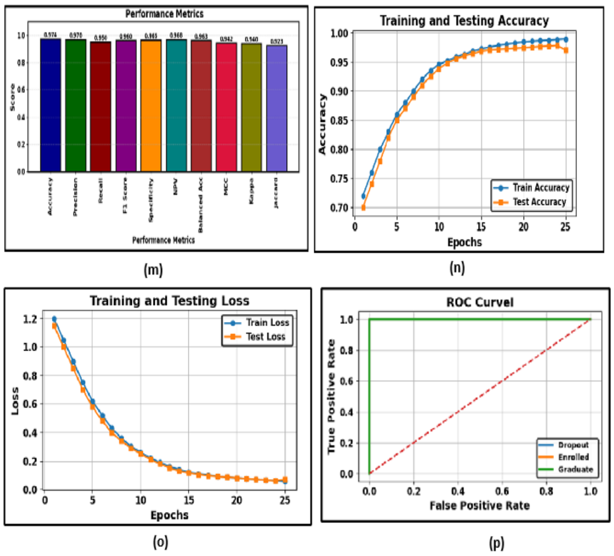

Figure 6: *Comparison on (a) performance metrics (b) Training and testing accuracy (c) training and testing loss (d) ROC curve*

In Figure 6(a), model performance metrics is shown that gives details about accuracy of 0.9735, precision of 0.97, recall value as 0.95, F1 score value as 0.96, specificity as 0.965, NPV value of 0.9680, balanced accuracy of 0.9625, Mattews coefficient is 0.942, cohens kappa value is 0.94 and Jaccard score is 0.9230. Training and testing accuracy curve is shown in Figure 6(b), accuracy increase in our training model but there is no increase after 15 epochs in test accuracy. The training and testing loss curve is shown in Figure 6(c), that states classifier model decreases over epochs. ROC curve is shown in Figure 6(d) that details about true positive rate and false positive rate of dropout, enrolled and graduate students.

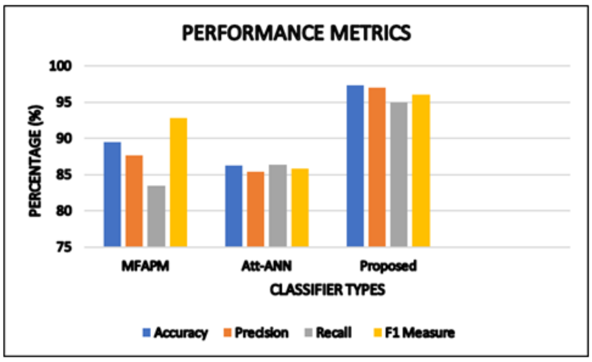

Figure 7: *Comparison chart on performance metrics*

The performance of classifier is analysed with conventional models such as MFAPM and Att-ANN model with proposed model and compared its accuracy, precision, recall and F1 measure values and a graph is plotted as shown in Figure 7.

5. Conclusion

The proposed framework estimates students' performance and anticipate learning outcomes, guides lectures and administrators for improving their educational quality in universities to provide good employments among students. In this methodology, an innovative hybrid Mamba transformer network with adaptive attention learning is proposed for better classification of dataset that aids enhancement in accuracy than any other baseline techniques. This model especially finds out performance of students that provides detailed knowledge of results than single form of evaluating students. This type of integration help management directors and educators to take immediate action in data-informed decision taking actions and provide better opportunities in students career. Extended work can be proposed in future by integrating explainable Artificial Intelligence(XAI) to provide details of students without interpretabilities and predict factors that contribute students outcomes. Effective usage of this classifier mitigates risk of mislabelling that leads to destruction in performance of model and hence classifier performs with enhanced accuracy of 0.9735, precision with 0.97, recall value of 0.95 and F1 measure with 0.96 rates than previous integrative technique.

References

1. Mfowabo Maphosa; Wesley Doorsamy; Babu S. Paul, Year: 2023, "Student Performance Pa tterns in Engineering at the University of Johannesburg: An Exploratory Data Analysis", in IEEE Access, Vol: 11, pp. 48977-48987.
2. Hongkun Qiu; Haojie Zhu; Xiaoyong Hu; Yajie Wang; Yue Xu; Wenlong Li, Year: 2026, "Predicting Student Performance Using a Cost-Sensitive Deep Learning Network With Sparse Attention Mechanism", in IEEE Transactions on Education, Vol: 69, No: 1, pp. 10-18.
3. Zhonghua Zhan; Tongping Shen; Jiaming Wang, Year: 2025, "Predicting Student Performance Based on Knowledge Characteristics and Learning Ability", in IEEE Access, Vol: 13, pp. 26105-26115.
4. Essa Alhazmi; Abdullah Sheneamer, Year: 2023, "Early Predicting of Students Performance in Higher Education", in IEEE Access, Vol: 11, pp. 27579-27589.
5. Ziling Chen; Gang Cen; Ying Wei; Zifei Li, Year: 2023, "Student Performance Prediction Approach Based on Educational Data Mining", in IEEE Access, Vol: 11, pp. 131260-131272.
6. Jin Li; Shu Li; Yuan Zhao; Longjiang Guo; Fei Hao; Meirui Ren; Keqin Li, Year: 2023, "Predicting dropouts before enrollments in

MOOCs: An explainable and selfsupervised model", IEEE Transactions on Services Computing, Vol: 16, No: 6, pp. 4154–4167.

7. Nesren Farhah; Adnan Muhammad; Ahmed Abdullah Alqarni; M. Irfan Uddin; Theyazn HH Aldhyani, Year: 2025, "AI-Driven Innovation Using Multimodal and Personalized Adaptive Education for Students With Special Needs", in IEEE Access, Vol: 13, pp. 69790-69820.
8. Wei Zhang; Zhongwei Gong; Peihua Luo; Zhixin Li, Year: 2024, "DKVMN-KAPS: Dynamic Key-Value Memory Networks Knowledge Tracing With Students' Knowledge-Absorption Ability and Problem-Solving Ability", in IEEE Access, Vol: 12, pp. 55146-55156.
9. Ike Fibriani; Eko Mulyanto Yuniarno; Ronny Mardiyanto; Mauridhi Hery Purnomo, Year: 2024, "ViTMa: A Novel Hybrid Vision Transformer and Mamba for Kinship Recognition in Indonesian Facial Micro-Expressions", in IEEE Access, Vol: 12, pp. 164002-164017.
10. Yu Feng; Minghui Li; Mingjie Chen; Benning Song; Dongjun Yu, Year: 2026, "An Efficient Ensemble Mamba Architecture for IMU Data Processing in the FAST Feed Cabin", in IEEE Sensors Journal, Vol: 26, No: 7, pp. 10589-10598.
11. Sukrit Leelaluk; Cheng Tang; Tsubasa Minematsu; Yuta Taniguchi; Fumiya Okubo; Takayoshi Yamashita; Atsushi Shimada, Year: 2024, "Attention-Based Artificial Neural Network for Student Performance Prediction Based on Learning Activities", in IEEE Access, Vol: 12, pp. 100659-100675.
12. Yupei Zhang; Rui An; Shuhui Liu; Jiaqi Cui; Xuequn Shang, Year: 2023, "Predicting and Understanding Student Learning Performance Using Multi-Source Sparse Attention Convolutional Neural Networks", in IEEE Transactions on Big Data, Vol: 9, No: 1, pp. 118-132.
13. Daozong Sun; Rongxin Luo; Qi Guo; Jiaxing Xie; Hongshan Liu; Shilei Lyu; Xiuyun Xue; Zhen Li; Shuran Song, Year: 2023, "A University Student Performance Prediction Model and Experiment Based on Multi-Feature Fusion and Attention Mechanism", in IEEE Access, Vol: 11, pp. 112307-112319.
14. Han Wan; Shiyang Yue; Mengying Li; Xin Luo; Yaofeng Hu; Baoliang Che; Jingyuan Wang, Year: 2026, "Integrating Blended Learning Behaviors via Multimodal Fusion for Student Performance Prediction", in IEEE Transactions on Learning Technologies, Vol: 19, pp. 87-104.
15. Yu Liu; Yanchuan Hui; Dongxu Hou; Xiao Liu, Year: 2023, "A Novel Student Achievement Prediction Method Based on Deep Learning and Attention Mechanism", in IEEE Access, Vol: 11, pp. 87245-87255.