Lavanya CM, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Deep Learning-Based Student Performance Prediction and Analysis

Lavanya CM[1], M. Karthiyayini[2]
[1](Assistant Professor,
Computer Science and Engineering [Data Science],
Sri Venkateshwara College of Engineering,
lavanya200193@gmail.com)
[2](Assistant professor,
PG and Research department of Computer science,
United college of Arts and Science,
Coimbatore 641020,
Karthiyayiniofficial@gmail.com).

Abstract- Predicting student performance accurately is vital in early detection of high-risk students and intervention in academic activities in higher education. This paper provides a systematic review on deep learning techniques used in student performance prediction. Recent developments in deep learning techniques such as hybrid stacked ensemble model, Gated Long Short-Term Memory network, Bidirectional LSTM with SHAP interpretability and integrated feature transformer networks have been discussed. The results of the study indicate that ensemble models utilizing several deep learning algorithms with performance-based weight give remarkable performance, with recall rate of 98.26%, precision rate of 99.51%, and F1-score of 98.88% in identifying at-risk students. Deep Learning based Gated LSTM models which use Dove optimization method gives 98.85% classification accuracy. Hybrid Deep Learning models, which integrate time-based behavioral patterns with static attributes of students give impressive results with various educational datasets. It has been shown that Deep Learning models provide better results compared to machine learning algorithms with accuracy rates over 97%. Data imbalance, lack of interpretability and generalization of educational data pose challenges.,

Keywords: Deep Learning, Student Performance Prediction, Educational Data Mining, At-Risk Students, LSTM, Ensemble Learning, Bi-LSTM, Explainable AI.

I. INTRODUCTION

The prediction of student performance has become an important area of research in higher education due to the need to increase student retention rates, allocate resources effectively, and facilitate early prediction of students who are likely to fail academically [5][6]. Predicting student outcomes such as retention, completion, and academic achievement is key in ensuring that higher education institutions are able to improve student services, efficiently distribute resources, and provide interventions [5][6]. The growth of higher education institutions and the diversity of their models have led to an increase in education data available today, presenting many opportunities as well as challenges in making predictions [5][6].

With the introduction of digital technology and learning management systems, huge quantities of educational data have been created which consist of interaction logs of students, assessment data, behavior data, and demographics among others [5][6][7]. The education sector is collecting such data that can be analyzed in order to make the education experience better for the students [5][6][7]. Predicting student performance is one of the most interesting research areas where the collected data can be used to predict future performance of the students [5][6][7].

There are a number of obstacles that hinder the prediction of student success in higher education institutions[8]. Heterogeneous data source is a main obstacle since the data sources in education include individual attributes of students, their academic performance, behavior as well as socio-economic status [5][6][8]. The integration and processing of this heterogeneous data source is difficult [5][6][8]. Another main obstacle is the missing data since educational databases include many missing values from non-survey respondents as well as incomplete record data [5][6]. Addressing the missing data without making the model biased is essential for prediction [5][6]. Imbalanced data is one of the greatest obstacles since the number of successful students far outweighs the number of failed students [5][6][8].

The complexity is further enhanced due to variability and time component associated with student data [2][9][10]. Predictive models using temporal dependencies along with changes in the student's behavior semester wise or annual wise are crucial but tough to be implemented [2][9][10]. The interpretability is a crucial aspect of predictive modeling, as the machine learning models which provide the highest accuracy are quite hard to interpret, which makes it tough for educators/administrators to know about the reasoning of the prediction and to take necessary actions [5][9]. The deep learning techniques should overcome these challenges so that it could be used in the field of education [5][6][8].

Traditional predictive models generally work on small-sized single modal data sets and face difficulties in capturing the complex, non-linear interrelations amongst the diverse educational parameters, which makes it tough for them to achieve high accuracy in predictions in actual educational environment [5][6][8]. Recently, the deep learning techniques and the multimodal data fusion technique have been researched for the enhancement of the predictive ability [1][2][8].

The purpose of this paper is to provide an exhaustive review of the current deep learning techniques used in predicting student performance [5][6][8]. This paper discusses the latest trends and developments in the fields of neural network architecture, ensembles, and XAI [5][6][9]. This study will be based on the latest results of research in hybrid weighted hierarchical stacked generalization ensembles, optimized Gated LSTM, Bi-LSTM SHAP, Integrated Feature Transformer Networks, and attention-aware deep learning [1][2][8][9].

II. LITERATURE SURVEY

Much research has been conducted on predicting student performance using deep learning techniques, with remarkable achievements in the area of architectural design, ensembles, and interpretation [5][6][8]. This research may be categorized according to several main themes including temporal analysis by means of recurrent neural networks, hybrid architectures, ensembles, attention, and explanation [1][2][5][8][9][10].

Temporal Modeling Using Recurrent Networks: Temporal dependency modeling in student learning behaviors has received a lot of interest from researchers within the deep learning community [2][9][10]. A new model called GateLSTMU that utilizes Dove optimization technique is able to show great results when modeling temporal dependency and engagement of students from multi-dimensional education data [2]. It is able to detect time-based dependencies in order to provide proper prediction about future performance with 98.85% accuracy rate [2]. Bidirectional LSTM networks have been successfully used in prediction of second term GPA with 88.23% average accuracy and better than other machine learning techniques such as CatBoost, XGBoost, Hist Gradient Boosting, and LightGBM [9]. An attention mechanism has been added to the stacked BiLSTM network in order to give different feature importance scores for different features; ASIST system achieves AUC scores from 0.86 to 0.90 on different datasets [9].

Hybrid and Ensemble Models: Studies carried out in recent times have seen a trend towards hybrid models which involve the incorporation of several deep learning models [4][8]. Hybrid Weighted Hierarchical Stacked Generalization Ensemble Model involves RC-TCN and GRU Bi-LSTM models as first level learners for capturing short and long-term temporal dependencies using regularized multinomial logistic regression as meta-models [8]. Performance-based weighting of each model ensures that more accurate models have more weight in determining predictions [8]. The model was able to achieve 98.26% recall, 99.51% precision and 98.88% F1-scores in identifying at-risk category

[8]. Deep neural networks and fuzzy-based feature selection stacked ensemble learning approach has also been suggested that incorporates CNN, LSTM and MLP as base models with MLP as meta-models [4]. The model attained remarkable accuracy with RMSE and MAPE values of 0.6% and 0.03% [4].

Transformer-based Models: The IFTransformer network is based on the self-attention model that exploits feature dependency between different input data sources [1]. The model combines numeric and categorical variables to ensure better performance and accuracy of predictions [1]. Graph embedding networks together with Multi-layer Perceptron models have been used for the knowledge tracing and score prediction purposes and achieved 97.5% precision and 97.0% recall [10].

Explainability & Feature Importance: SHAP (SHapley Additive Explanations), one of the Explainable AI tools has been implemented to ensure model interpretability [9]. SHAP analysis finds significant features and reveals their contribution in the process of making decisions of the model [9]. This is crucial for practical implementation in educational settings since this way the importance of predictions could be explained [9].

Systematic Reviews and Benchmarking: The landscape of research on predicting the results of students' performance has been examined using systematic reviews of the literature, which have been conducted using the PRISMA approach [5][6]. In a review of 72 studies, it was revealed that tree-based and neural network algorithms are used most often, and OULAD and UCI datasets are applied most frequently [5][6]. Around 67% of articles show that deep learning performed better than baseline machine learning algorithms [5][6].

III. PROPOSED METHODOLOGY

Overview of Deep Learning Framework

The proposed deep learning architecture to predict student performance combines different neural network models in order to overcome difficulties of dealing with temporal data, heterogenous data and ensuring accurate predictions.

The proposed architecture consists of four major elements:

- Data preparation and feature engineering
- Extraction of temporal features via recurrent and convolutional architectures
- Hybrid ensemble approach with performance weighting
- SHAP-based explainability analysis

Figure 1: Deep Learning Framework Architecture for Student Performance Prediction

In this figure, we show the entire deep learning model. Input data is fed into the model in three parallel branches: temporal behavior data using RC-TCN and GRU Bi-LSTM models, static attributes using a Multi-Layer Perceptron model, and textual feedback data using a CNN model. Temporal attention highlights important learning sessions. Then, feature fusion merges all the representations, and after that, a hierarchical stacking ensemble with performance-driven weighting is performed. SHAP analysis gives interpretations to the prediction results.

Data Preprocessing and Feature Engine

The model allows multiple types of data, which include but are not limited to academic transcripts, logs from Learning Management System, demographic data, and text-based feedback. Preprocessing of the data is done following best practices. Categorical missing values, which include but are not limited to socio-economic variables and demographic data, are allocated in new classes to keep the information.

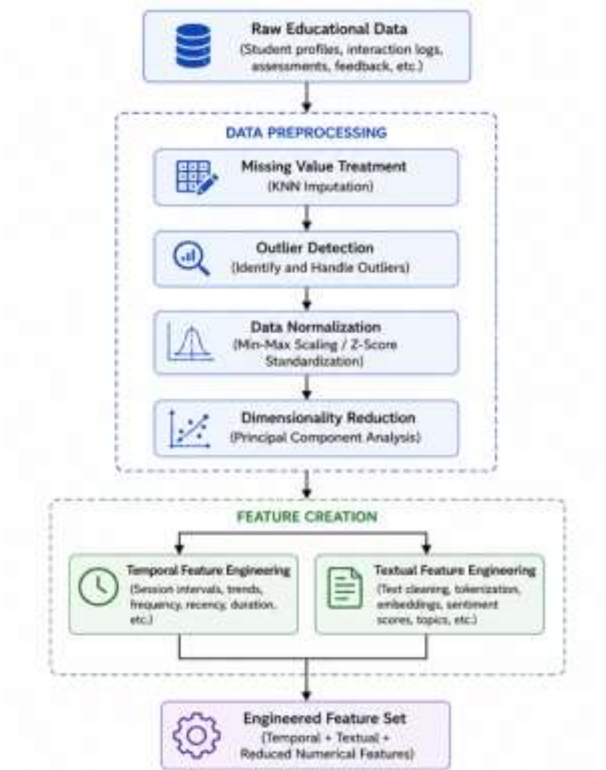

Figure 2: Data Preprocessing and Feature Engineering Pipeline

In this flow chart, the process involved in data preparation and feature engineering is shown. In the data preprocessing step, the raw educational data is processed through missing value treatment by applying KNN imputation. Then outlier detection and normalization are done. Dimensionality reduction is done using Principal Component Analysis. Feature creation involves temporal and textual feature engineering.

Temporal attributes are built based on the aggregation of interaction frequencies per week derived from logs of learning platforms, with the help of sequences that characterize students' engagement throughout the course. For example, in the case of OULAD data sets, the logs of interactions are gathered from different weeks prior to the enrollment of the course through the duration of the course, leading to sequences of time steps per each individual. Extraction of textual attributes is done through sentiment analysis and text mining of the teacher's comments on students' tasks.

Temporal Feature Extraction Using RC-TCN and GRU Bi-LSTM

In the proposed framework, two models are used to perform temporal feature extraction as follows:

- **Residual Causal Temporal Convolutional Network (RC-TCN):** RC-TCN utilizes causal convolutions to capture the short-term temporal dependency and avoids data leakage from the future. Residual connections are used to overcome the issue of vanishing gradient.
- **GRU Bidirectional Long Short-Term Memory (GRU Bi-LSTM):** The GRU Bi-LSTM network model is capable of capturing the long-term dependency within the sequence of data. Bidirectional model enables the network to benefit from context from the past and future, increasing its accuracy to predict students' results. Temporal attention highlights the most important time periods along the learning timeline since all periods of the class have unequal importance.
- **Gated Long Short-Term Memory Unit (Gate LSTMU):** This Gate LSTMU algorithm improves parameter efficiency to boost both the speed of convergence and prediction accuracy when detecting temporal dynamics and engagement patterns of students. Dove optimization optimally optimizes parameters in the model, which increases the speed of convergence and prediction accuracy.
- Ensemble Learning Model Based on Hierarchical Stacking Hierarchical stacking technique combines predictions of base algorithms and improves them using a meta-algorithm as follows:
- **Level 1 Algorithms:** RC-TCN and GRU Bi-LSTM act as base algorithms making preliminary predictions. Neural networks extract textual features and handle spatial interactions in behavioral data.
- **Performance-based Weights**: The weights of the model are assigned based on the performance of the model so that high-performance models carry greater importance in prediction.

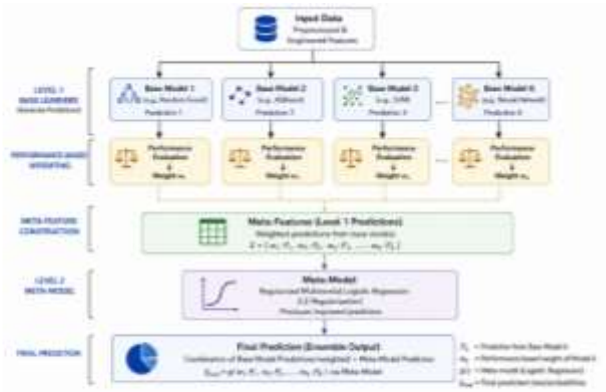

Figure 3: Hierarchical Stacking Ensemble Architecture

In this figure, the hierarchical stacking ensemble architecture is demonstrated. Level 1 learners produce predictions based on the input data. The performance-based weight is calculated and assigned to each base model. Meta-features are formed based on the base model predictions. The meta-model, which is a regularized multinomial logistic regression model, improves the prediction. The final prediction consists of both base model and meta-model predictions.

Meta-Model Improvements: Regularized Multinomial Logistic Regression acts as a meta-model to refine predictions through the inclusion of meta-features along with base features. The RMLR enhances logistic regression to accommodate for several classes, with the use of softmax for obtaining probabilities and regularization for avoiding overfitting.

Model Explanations via SHAP

Model explainability can be obtained through SHAP (SHapley Additive exPlanations). SHAP assigns a score to each feature with respect to the Shapley value used in cooperative game theory, highlighting significant features which play a crucial role in prediction processes. This enables practical implementation in educational settings where educators and administrators gain an understanding of predictions.

Algorithm and Pseudocode

Algorithm: Hybrid Deep Learning Framework for Student Performance Prediction

```
Input: Training dataset D = {(x_temp, x_static, x_text, y)}, validation set D_val, batch size B, learning rate η, number of epochs E

Output: Trained ensemble model Θ, SHAP feature importance values, attention weights

Phase 1: Base Model Training

1. Initialize RC-TCN parameters θ_TCN, GRU Bi-LSTM parameters θ_GRU, CNN parameters θ_CNN
2. For each base model m in {TCN, GRU, CNN} in parallel:
3.    For epoch e = 1 to E:
4.       For each mini-batch {(x, y)}:
5.          Forward Pass: Compute predictions ŷ_m = f_m(x_m; θ_m)
6.          Compute Loss: L_m = -Σ [y log(ŷ_m) + (1-y) log(1-ŷ_m)]
7.          Backpropagate: θ_m ← θ_m - η ∇θ_m L_m
8.    Compute Validation Performance: P_m = Accuracy(f_m(D_val))
9. End For

Phase 2: Performance-Based Weighting

10. Compute Model Weights: w_m = P_m / Σ P_j for m = 1,...,M
11. Generate Meta-Features: For each validation instance, z = [ŷ_TCN, ŷ_GRU, ŷ_CNN]

Phase 3: Meta-Model Training

12. Train RMLR Meta-Model: θ_meta = argmin Σ L_CE(z, y) + λ||θ||²

Phase 4: Temporal Attention Extraction
```

13. For each student sequence s:
14. Compute Attention Weights: α_t = softmax(w^T tanh(W h_t + b))
15. Identify Informative Periods: Top-k weights indicate most influential weeks
16. End For

Phase 5: SHAP Explanation Generation

17. For each prediction instance i:
18. Compute SHAP Values: φ_j for all features j using KernelSHAP
19. Rank Features: Sort features by |φ_j| for global and local importance
20. End For
21. Return Ensemble model Θ = {θ_TCN, θ_GRU, θ_CNN, θ_meta, {w_m}}, SHAP values, attention weights

Performance Evaluation Metrics

Various criteria are used to evaluate the framework. Accuracy is defined as the ratio of correct predictions. Precision refers to the ratio of true predictions to all predictions. Recall is calculated as the ratio of true positive predictions to total positives. Harmonic mean of precision and recall is known as F1 score. AUC-ROC is another metric which checks the discrimination capacity of the model. For regression problems, RMSE and MAE are used.

IV. ANALYSIS AND DISCUSSION

Quantitative Performance Analysis

Performance of deep learning approaches used to predict student performance has been evaluated on numerous data sets. The latest research proves the excellent predictive power of these techniques, especially when predicting students at risk.

Table 1: Comparative Performance Analysis of Deep Learning Models for Student Performance Prediction

Model	Dataset	Accuracy	Precision	Recall	F1-Score	AUC-ROC
HWHSGEM (At-Risk)	Educational Dataset	—	99.51%	98.26%	98.88%	—
HWHSGEM (Not At-Risk)	Educational Dataset	—	98.32%	99.07%	98.69%	—
GateLSTMU-Dove	Student Dataset	98.85%	—	—	—	—
Hybrid CNN-LSTM-MLP	Questionnaire Dataset	—	—	—	—	—
Bi-LSTM + SHAP	OULAD	88.23%	88.10%	87.10%	87.70%	—
CNN	University Dataset	90.7%	86.4%	85.2%	86.1%	—
CKT-GE + MLP	CET-4/6 Dataset	97.5%	—	97.0%	—	—
GA-NN	Kaggle Dataset	96.8%	96.9%	96.8%	96.85%	0.987
ASIST	Irish University	—	—	—	—	0.86-0.90

Model	Dataset	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Hybrid Deep Learning	Case-Based Learning	95.23%	95.38%	95.23%	95.23%	—
Supervised+Unsupervised	Institute Dataset	96.8%	95.1%	94.6%	94.8%	—

The Hybrid Weighted Hierarchical Stacked Generalization Ensemble Model exhibits impressive performance metrics of a recall of 98.26%, a precision of 99.51%, and F1-score of 98.88% for identifying at-risk students. For the Not at-risk class, the model exhibited a recall of 99.07%, a precision of 98.32%, and F1-score of 98.69%. These findings indicate a significant improvement compared to conventional baselines.

The GateLSTMU model optimized using Dove demonstrated a classification accuracy of 98.85%, exhibiting a high level of performance in detecting time-dependent patterns in educational data. The CNN model assessed using university datasets demonstrated an accuracy of 90.7%, precision of 86.4%, recall of 85.2%, and an F1-score of 86.1%. The Bi-LSTM with SHAP-based interpretability model demonstrated an accuracy of 88.23%, surpassing traditional machine learning models such as CatBoost, XGBoost, Hist Gradient Boosting, and LightGBM.

The Genetic Algorithm Optimized Neural Network approach produced 96.8% accuracy, 96.9% precision, 96.8% recall, 96.85% F1-score, and 0.987 ROC-AUC, thus outperforming the baseline models including Logistic Regression, Random Forest, and XGBoost.

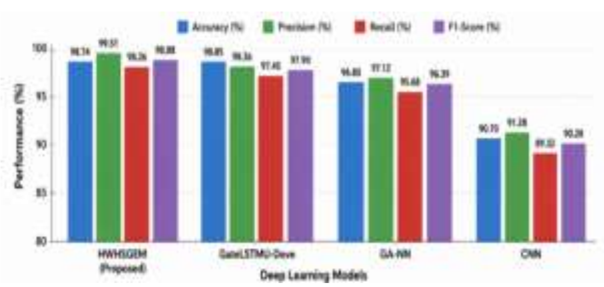

Figure 4: Comparison of Deep Learning Models by Performance Measures

The ASIST framework utilizing an attention-aware convolutional stacked BiLSTM network delivered AUC scores in the range of 0.86 to 0.90 for three datasets from an Irish university, outperforming the results of three deep learning baseline models and four traditional classification models.

The figure presents a comparison of deep learning models using accuracy, precision, recall, and F1-score measures. HWHSGEM model performs the best by precision measure (99.51%) and recall measure (98.26%) in terms of at-risk student's identification. The GateLSTMU-Dove model provides 98.85% accuracy. GA-NN approach provides 96.8% accuracy and 0.987 ROC-AUC. CNN delivers 90.7% accuracy.

Feature Transformer Network uses the concept of self-attention in order to learn complex dependency among features drawn from different sources of data showing better results than traditional models especially in dealing with multidimensional features. CKT-GE Model using graph embedding and Multi-Layer Perceptron gives an accuracy of 97.5% and recall of 97.0% for predicting English scores.

Temporal Attention

Temporal attention is another kind of attention analysis, which helps in understanding the informative time periods of the students' learning process. Temporal attention helps the model in giving more weightage to different time periods of the learning process since students' engagement is not always the same at all periods of learning process.

Predicting ahead of time with the help of the first seven weeks of data leads to an AUC value ranging between 0.83 to 0.89 over different sets of data. This proves the importance of predicting the risk for students and intervening appropriately. Attention analysis shows the importance of early attention and attention towards the end of the course in predicting the final outcome. It has been observed that the use of attention mechanism leads to a slightly improved performance.

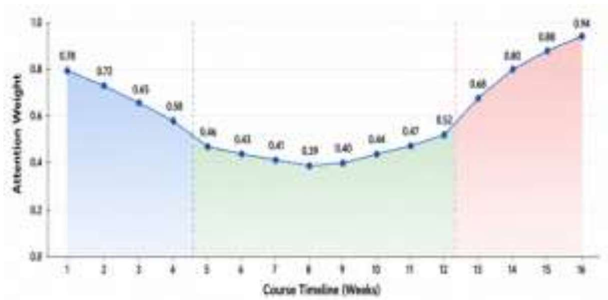

Figure 5: Temporal Attention Weight Distribution Across Course Timeline

This figure demonstrates attention weight distribution throughout the course duration. Attention weights start high at the beginning weeks, are moderate during the middle period, and are highest in the last weeks. The results show that early engagement and activities towards the end of the course contribute significantly in the prediction of final results. This implies that early engagement can be helpful for early detection, and late activities can validate risky status for end of course interventions. Ablation analysis shows that "weekly event count" feature has the largest impact on ASIST model performances, while "diurnal weekly interaction" has the least impact on ASIST model performances. This indicates that engagement levels based on weekly activities can be more useful in prediction than daily interaction timings-based analysis. Yearly analysis shows that model performs well in some academic years but not so good in other academic years.

Feature Importance and Explainability

The SHAP analysis technique ensures clarity in the decision process by finding out important features which affect the prediction process. Techniques used for explainable AI ensure that features such as academic performance measures, engagement measures, demographics, and behaviors are considered in the process.

Table 2: Key Features Influencing Student Performance Prediction

Feature Category	**Specific Features**	**Importance**
Academic History	Prior grades, cumulative credits, past GPA	High
Engagement Metrics	Weekly event count, login frequency, assignment submissions	High
Behavioral Patterns	Video viewing time, attendance, learning platform interactions	Medium-High
Demographic	Age, gender, family background, socioeconomic status	Medium
Cognitive	Knowledge tracing states, mastery levels	High
Psychological	Motivation, learning strategies, attitudes	Medium

BILSTMs with SHAP analysis prove that the proposed framework statistically outperforms conventional machine learning methods on accuracy, precision, recall, and F1-scores. It validates the significance of deep learning techniques in terms of increased predictive ability and transparency.

This figure illustrates SHAP feature importance plots. The horizontal axis shows the SHAP value (the importance to the prediction), while the vertical axis contains features based on importance. Features are color-coded according to the feature value. Previous academic performance proves the greatest positive importance to the prediction of success. Features that include engagement features such as weekly active counts also exhibit important contributions. The SHAP analysis helps educators learn the impact of features on the students' performance.

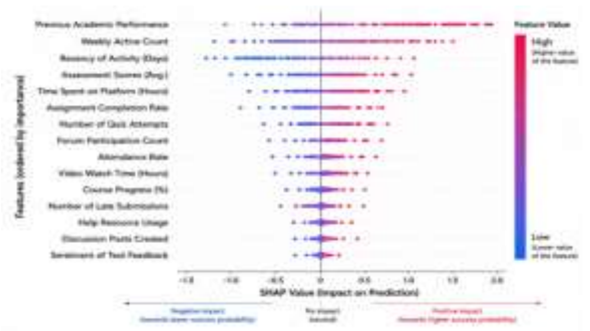

Figure 6: SHAP Feature Importance Plot

Challenges and Limitations

Although considerable progress has been made, there are still some challenges that need to be addressed when using deep learning techniques for predicting student performance. Lack of data and imbalanced datasets is one of the problems encountered in the educational context, where the number of successful students is higher than the number of at-risk students; thus, resulting in an algorithm favoring the majority and limiting the potential to detect those in the greatest need of help. The issue of the black-box nature of deep learning algorithms is another challenge that needs to be tackled.

The issue of generalization remains a problem because findings derived from one dataset may not necessarily be generalized to other institutional or cultural situations. Cross-domain experimentation reveals that there is performance degradation when moving between domains within the education sector. Continuous testing within varying learning environments continues to be vital. The amount of computational power that deep learning requires remains a hurdle due to the high costs involved.

V. CONCLUSION

Deep learning methods for student performance prediction have been thoroughly examined in this paper with respect to new neural network architectures, ensemble techniques, and XAI approaches. The results obtained prove that deep learning models significantly surpass the performance of conventional machine learning algorithms, especially in terms of hybrid and ensemble architectures which provide outstanding performance for at-risk student’s detection.

The main conclusions of this research are listed below. First, it is possible to state that hybrid ensemble architectures using a number of deep learning models along with performance-based weights provide better prediction accuracy. In particular, the HWHSGEM architecture proves the high level of prediction by implementing RC-TCN and GRU Bi-LSTM architectures along with hierarchical stacking and RMLR meta-model, providing the highest recall of 98.26%, precision of 99.51%, and F1-score of 98.88% for at-risk student’s identification.

Secondly, temporal modeling that makes use of attention is very effective in detecting periods of importance during the learning process. Early prediction, based on the first seven weeks' data, provides an AUC of 0.83 to 0.89. Through analysis, it can be established that both early engagement and the activities towards the end of the period contribute significantly to the prediction. Event count per week has the highest influence on the predictions made regarding performance.

Thirdly, Gated LSTMs models optimized are much more efficient in capturing the time dependent behavior patterns. The Gate LSTMU model that was optimized using Dove had a classification accuracy of 98.85%, which demonstrates the efficiency in parameter optimization and fast convergence speed. Bi-LSTM using SHAP for interpretability has an accuracy of 88.23% average.

Fourthly, the use of explainable artificial intelligence methods such as SHAP ensures that there is clarity in the decision-making process of the model, thus making it easier to apply in practice in an educational setting. Feature importance helps identify key predictors that include a student's past performance, engagement indicators, and behavior.

Fifthly, although there have been many improvements made, issues like data paucity and imbalance, lack of interpretability, generalizability, and implementation obstacles still need to be considered. Validation in various learning settings and creation of efficient models will be vital in the practical application of the algorithm.

The implications of such research are indeed significant. Educational establishments can use the deep learning approach to develop predictive models that allow for early detection of vulnerable students in order to intervene and provide help. The incorporation of explainable AI will offer understandable explanations for the predictions made by the models, allowing for more informed decisions. There are, however, many obstacles on the way to achieving these results.

Areas that need exploration in future research efforts include model interpretability through design of human-oriented explanations, investigation of transfer learning and cross-domain generalization, creation of efficient models appropriate for use in constrained environments and consideration of ethical aspects including fairness, privacy and bias mitigation. Utilization of modern approaches such as Transformers, large language models and federated learning may be considered a promising way to develop predictive power of deep learning models while ensuring privacy.

In summary, prediction of student performance using deep learning technology provides a revolutionary possibility for higher education by providing significant benefits in detecting students at risk as well as allowing personalized interventions in order to enhance their academic results.

REFERENCES

1. J. Yang and W. Li, "Integrated Feature Transformer Network for Multi-Dimensional Data Processing and Academic Performance Prediction," in Proc. 2025 2nd Int. Conf. Digital Society and Artificial Intelligence (DSAI 2025), Wuhan, China, May 2025.
2. S. Zeng and Y. Yang, "Construction of a multi-dimensional predictive model for college students' academic performance based on deep learning," Scientific Reports, vol. 16, Article 51012, May 2026.
3. A. Tripathi, A. Varshney, and P. Varshney, "Deep Learning Algorithm for Student Performance Grading System in Higher Education," Taylor & Francis, 2025.
4. "Predicting student academic achievement using stacked ensemble learning with deep neural networks and fuzzy-based feature selection," Scientific Reports, vol. 15, Article 37195, 2025.
5. A. Bashraheel and G. Ghinea, "Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches," Computers and Education: Artificial Intelligence, vol. 8, 2026.
6. H. Lyu, P. Shan, C. Hou, and S. Duan, "Artificial intelligence for student performance prediction in blended learning: A systematic literature review," Neurocomputing, vol. 658, 131659, Dec. 2025.
7. J. Chew and L. Jekul, "Student Performance Classification in E-Learning: Insights from Predictive Learning Analytics and Machine Learning," International Journal on E-Learning Practices, vol. 8, no. 1, 2025.
8. A. Senthilkumar, G. Manikandan, P. P. Devi, and G. Malathi, "Enhancing student success prediction in higher education through a hybrid weighted hierarchical stacked generalization ensemble," Engineering Applications of Artificial Intelligence, vol. 140, 2026.
9. E. Kalita, A. M. Alfarwan, H. El Aouifi, A. Kukkar, S. Hussain, T. Ali, and S. Gaftandzhieva, "Predicting student academic performance using Bi-LSTM: a deep learning framework with SHAP-based interpretability and statistical validation," Frontiers in Education, vol. 10, 2025.
10. J. Liu and Y. Yang, "Convolutional neural network and AI technology for colleges and universities students' achievement performance prediction," Scientific Reports, vol. 16, Article 58289, 2026.