ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

Conference Proceedings

Open Access

A Machine Learning-Based Framework for Predicting and Improving Student Outcomes Using Big Educational Data (Approved by ICITET 2024)

Srikanth Reddy Vangala[1]*, Ram Mohan Polam[2], Bhavana Kamarthapu[3], Ajay Babu Kakani[4], Sri Krishna Kireeti Nandiraju[5] and Sandeep Kumar Chundru[6]

[1]University of Bridgeport, USA

[2]University of Illinois at Springfield, USA

[3]University Fairleigh Dickinson, USA

[4]University Wright State, USA

[5]University of Illinois at Springfield, USA

[6]University of Central Missouri, USA

ABSTRACT

Evaluating student academic achievements through prediction tools serves as a vital resource for both teaching staff and students seeking better educational methods. Effective predictive techniques enable students to study in targeted areas that use forecast outcomes while effective analytical approaches help instructors develop proper educational materials. A deep learning-based approach uses educational data from three colleges in Assam, India to forecast student academic performance in this research. A Gated Recurrent Unit (GRU) neural network served as the proposed method for detecting temporal patterns and dependencies within student data. The proposed model outperformed traditional approaches with Artificial Neural Networks (ANN) and Decision Trees (DT) where it delivered an accuracy rate at 99.70% and precision at 98.60% and recall at 96.30% with F1-score at 97.40%. The robustness and generalization capability of the GRU model is substantiated through evaluation using confusion matrix alongside accuracy and loss curve metrics. Deep learning analytics shows great promise for educational applications because this research delivers critical information about preventive actions and academic achievement enhancements. The findings reveal that the GRU model delivers exceptional capacity for identifying at-risk students early while enabling data-driven educational interventions to create better academic results.

***Corresponding author**

Srikanth Reddy Vangala, University of Bridgeport, USA.

Received: August 16, 2024; **Accepted:** August 20, 2024; **Published:** August 28, 2024

Keywords: Education System, Student Performance Prediction, Academic Performance, Educational Data, Machine Learning, Big Data, Educational Data

Introduction

Education functions as the essential foundation for contemporary economic growth because it directs national productivity while powering innovation and competitive advantages in global markets. The transition to knowledge-based economies now relies on three vital educational elements which determine how well countries can prosper in the future. The education sector transforms into an adaptive industry which needs to distribute knowledge while adapting to technological progress and changing workplace requirements. Educational institutions face mounting pressure to improve their learning effectiveness while enhancing outcomes for their students [1]. The school's ability to achieve both student achievements and enrollment metrics functions as vital markers for measuring institutional performance and academic success. Many educational institutions that receive public funding struggle with high dropout rates and prolonged graduation times along with erratic academic results that drive up educational expenses. Student outcome enhancement represents a strategic requirement for institutions if they wish to stay relevant and secure long-term financial strength [2].

A proactive approach to this issue demands advanced evaluation methods that incorporate data-driven techniques for both performance forecasting and improvement [3]. Modern education resulted in institutional systems gathering massive amounts of data about grades while also collecting records of attendance and more demographic data, and course activity data [4]. Educational data large enough to provide rich insights presently sits untouched, although it provides significant potential to guide intervention decisions along with strategic planning initiatives [5]. Educational institutions use AI and ML to properly analyze and execute decisions based on this data. Through these technologies, institutions gain access to sophisticated analytical techniques that enable them to create complex models while performing academic risk prediction and adapting learning approaches to individual students [6]. Educational institutions find ML particularly effective in their settings because it reveals concealed patterns within student actions and historical academic progress. Educational staff use predictive analytics to take early action, which helps student success alongside stronger institutional accomplishments.

Motivation and Contributions of the Study

The goal of this research focuses on enhancing performance prediction precision through advanced DL methods so that it becomes possible to identify students who need assistance before they show signs of decline and create data-driven educational solutions. Student performance prediction models based on traditional approaches face limitations when dealing with substantial unbalanced dataset information, alongside the inability to recognize sequential student records. The research implements the GRU model to solve previous predictive performance issues while supporting better academic planning tools and methods for student success. The key contributions are as:

- Collected and analyzed a large-scale educational dataset from three colleges in Assam, India.
- Addressed missing data, class imbalance (using SMOTE), and performed data standardization for improved model performance.
- Categorical data was successfully converted using one-hot encoding.
- Implemented a GRU model tailored for sequential student performance data.
- Analyze model effectiveness using f1-score, recall, accuracy, and precision.

Organization of the paper

The paper is structured as follows: Sections II provide a review of existing research on student performance prediction. Section III describes the proposed ML-based framework, model implementation, and methodology. Section IV presents experimental results, compares different models. Finally, Section V wraps up the work and offers some avenues for additional investigation to enhance the prediction of student outcomes.

Literature Review

This section presents an analysis of prior studies focused on detecting cybersecurity threats in IoT environments using ML and DL methodologies.

Rimadana, et al. use data on time management skills from the TSQ to predict students' academic achievement using a number of ML models. Five separate ML models were constructed using TSQ data to forecast pupils' academic achievement. Additionally, the same methodology is used to forecast students' English proficiency as a comparison. Because of this, the LSVM model can use TSQ data to predict 80% accuracy in academic achievement and 84% accuracy in English for students [7].

Hasan et al. the goal is to use AI to help the learner avoid the terrible outcome that was anticipated. By forecasting the final exam score, a student might be aware of his or her future performance and be informed of what steps to take to prevent poor outcomes. Students and educators would benefit from this study with the greatest precision of 94.88% [8].

Dubey and Mani analyze the potential applications of supervised machine learning models to predict high school students' employability for part-time employment with local businesses. Additionally, compare the effectiveness of the trained models used in this study. The employability of high school pupils may be accurately predicted by local firms, according to empirical findings. With up to 93% accuracy, the trained prediction models perform better with bigger datasets [9].

Pereira, et al. through the use of log data collected from an online judge, they have developed a set of successful attributes that are linked to the student's grade and applied to a database including 486 Grade 1 students. they leveraged this set of features in their machine learning pipelines, which included an evolutionary algorithm with an automated approach with hyperparameter-tuning and random search. They had a 75.55% accuracy rate in predicting the kids' final grades using data from only the first two weeks [10].

Jain and Solanki investigated the performance of four machine learning algorithms on educational datasets used to forecast student progress in advance. Although there is a wealth of research on predicting student achievement, their work is different from previous studies in the following ways: (i) Their prediction is not limited to the binary classification of pass or fail; rather, they have used a multiclass classification where students are divided into three classes: poor, average, and good performing students; (ii) They have developed a more accurate model with 95% accuracy and less execution time by using feature extraction, data preprocessing, and fine-tuning algorithm parameters, which is not a focus of many researchers; (iv) They have used a correlation heatmap to examine the relationship between various attributes [11].

Casuat and Festijo carried out a case study using 27,000 data points, containing 3000 observations and 9 characteristics, which included the students' GPA, OJT student performance rating, and assessment outcomes from simulated job interviews. OJT courses were taken by the students between 2015 and 2018. The accuracy, precision, and recall metrics from the performance matrix, as well as the F1-score and support measures, were used to evaluate the three methods. With an accuracy value of 91.22% throughout the testing, SVM fared better than other learning algorithms, including DT 85% and RF 84% [12].

Kaunang and Rotikan uses the data mining approach to offer a model for predicting computer science students' academic achievement. Questionnaires including information on the information was gathered using the pupils' demographics, previous GPA, and family history. Students' data is subjected to two data mining models (DT and RF) in order to provide the optimum model for predicting their academic success. The study's conclusion, which obtained the maximum accuracy value of 66.9%, indicates that DT is the superior model when compared to RF [13].

Al-Shehri et al. (2017) the authors developed two computational models to estimate student grades in their final examination. Student performance forecasts enable educational institutions to make decisions before difficulties occur through protective measures and immediate interventions, and appropriate student selection. predicted student grades using the SVM and KNN algorithms on the dataset, then assessed the prediction accuracy for improved assessment. SVM demonstrated marginally superior results compared to K-Nearest Neighbor because it generated correlation coefficients of 0.96 and 0.95, respectively [14].

Table 1 Compares Several Machine Learning Techniques for Predicting Student Performance, Emphasizing Methods, Datasets, Accuracy Attained, and Constraints with Regard to Data Scope and Generalizability.

Table 1: Summary of literature study for Student Performance Prediction using Machine learning

Author	Methodology	Data	Performance	Limitation
Rimadana et al.	Machine Learning (Linear Support Vector Machine)	Time Structure Questionnaire (TSQ) data	80% accuracy in predicting academic performance, 84% accuracy for English performance	Limited to Time Structure Questionnaire data; may not generalize to other learning factors
Hasan et al.	Artificial Intelligence for predictive analysis	Historical academic data, final exam marks	94.88% accuracy in predicting final exam marks	Predicted results may not fully capture real-time academic progress
Dubey and Mani	Supervised Machine Learning (Various models)	Data on high school students and local businesses	93% accuracy in predicting employability	Limited to local businesses and part-time job matching
Pereira et al.	Machine Learning with Evolutionary Algorithm & Random Search for hyperparameter tuning	Log data from an online judge (486 CS1 students)	75.55% accuracy predicting final grades	Data only from the first two weeks of the course; may not generalize to other subjects
Jain and Solanki	Machine Learning (Multiple algorithms)	Educational dataset (demographic, performance data)	95% accuracy with multiclass classification	Focuses on multiclass classification; execution time optimization may limit scalability
Casuat and Festijo	Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM)	Student data on mock job interviews, OJT performance, GPA	91.22% accuracy with SVM, DT (85%), RF (84%)	Limited to specific educational and job evaluation data
Kaunang and Rotikan	Data Mining (Decision Tree, Random Forest)	Student demographics, GPA, family background information	66.9% accuracy with Decision Tree	Low accuracy for Decision Tree compared to other models; data limited to Computer Science students
Al-Shehri et al.	K-Nearest Neighbor (KNN) and Support Vector Machine (SVM)	Dataset from University of Minho, Portugal (395 students)	SVM: 0.96 correlation, KNN: 0.95 correlation	Limited to a single subject (Math); may not generalize across other subjects

Methodology

This section details how a predictive student performance model is developed by employing ML techniques according to comprehensive methods. The research process starts with obtaining educational data from an open-source database. Data preprocessing maintains dataset clarity through extensive tasks that handle missing values and drop unneeded columns while converting categorical variables into one-hot encoding, then balances class distributions with SMOTE and standardizes measurements, and divides the data into training and Test portions having 80:20 ratios. Researchers implemented a GRU neural network model for student performance prediction because it combines long-term data dependency detection with computational effectiveness. In addition to visual aids like confusion matrices, accuracy, and loss curves, the model's performance was assessed using classification measures including accuracy, precision, recall, F1-score, and binary cross-entropy loss. The flowchart of the methodology is shown below in Figure 1.

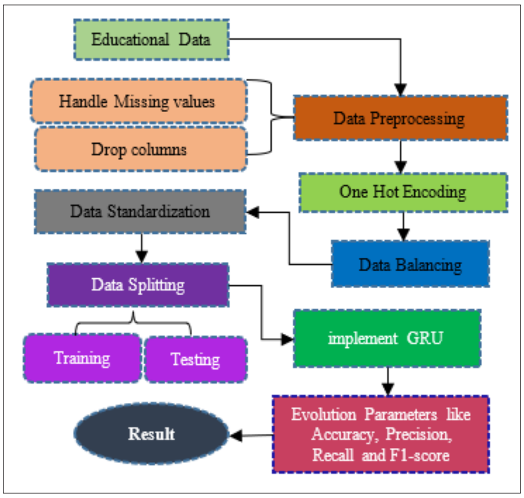

Figure 1: Flowchart for Student Performance

The following steps of a flowchart are briefly explained in below:

Data Collection

The educational dataset was gathered from three distinct Assamese institutions. These three institutions were Doomdooma College, Duliajan College, and Digboi College. 10140 records with 10 characteristics were found. A few values were absent from the datasets. No consideration was given to the missing data. The internal assessment percentages and the linear projection of different variables in relation to student outcomes show patterns, clusters, and outliers. Below are the visual insights of the data:

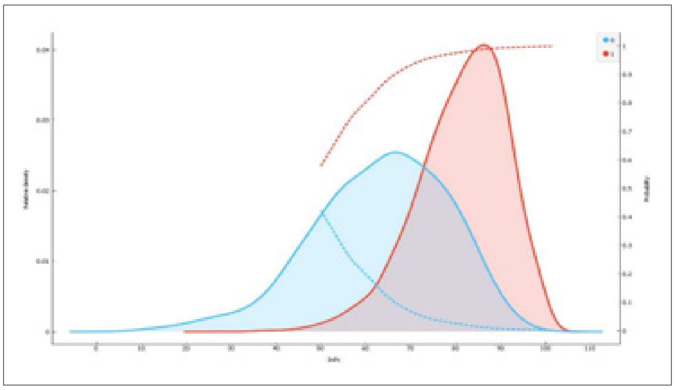

Figure 2: Distributions of Internal Assessment Percentage

The distribution of students' internal assessment percentages, segmented by result categories, is presented in Figure 2. The blue curve, representing one category, peaks around 70% with a broader distribution, while the red curve, corresponding to higher-performing students, peaks sharply near 85%, indicating stronger performance concentration among successful students.

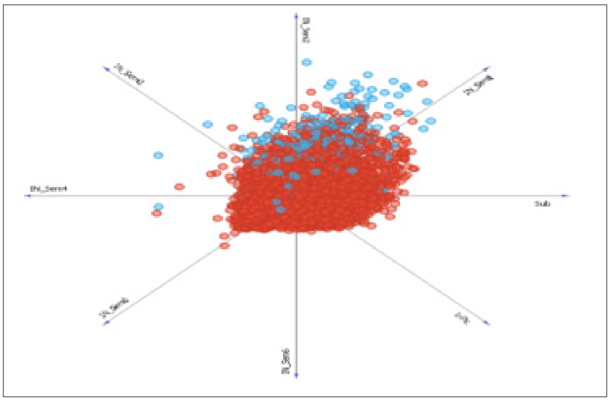

Figure 3: Linear Projection of All the Features

It displays the linear projection of every dataset characteristic in relation to student outcomes, highlighting performance trends across multiple attributes in Figure 3. The radial layout reveals a dense cluster of data points with a few outliers, suggesting that most students exhibit similar patterns, while some deviate significantly.

Data Preprocessing

Data Preparing the dataset before using classification algorithms requires pre-processing [15]. It should be noted that the job has a direct impact on the outcome because of the calibre and dependability of the information that is provided. To get rid of any anomalies, this activity involves carefully analyzing the variables and their related values. The following are crucial terms for further processing:

- **Handle Missing Values:** Depending on the proportion and importance of missing data, different strategies were applied. The missing values were imputed with mean and median.
- **Drop Columns:** Remove and discard any extraneous dataset columns. To remove the columns, use the dropna() function.

One-Hot Encoding for Data Labeling

In order to transform certain categorical data types, one-hot encoding is required since machine learning algorithms rely on numerical data to operate. Every determined category would have its corresponding binary value set to one for existing categories but all non-existing categories would receive zero value through this method. The method allows reliable data representation through this approach yet it provides no information about the order of categories between different groups.

Class Balancing

The issue of class imbalance occurs when there are significantly less instances in one class than there are in another or other classes. With 22.03% of students failing the course in the dataset that was gathered, there is a significant issue with class imbalance. In order to balance unbalanced datasets, Rather of reproducing existing samples, SMOTE generates new ones specifically for the minority class.

Data Standardization

The preprocessing method of data standardization transforms features into units with average values equal to zero and variance values equal to one. In Equation (1), a standardization formula is used where observation x represents the data's mean and standard deviation.

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

The standardized value Z results from dividing feature X by its mean μ against its standard deviation σ. ML model performance improves when features receive equal weighting during training as a result of normalization.

Data Splitting

A pre-development split separated the dataset into groups for testing and training 80% of the data was used for training, while the remaining 20% was used to evaluate the models' generalization and prediction capabilities.

Proposed Gated Recurrent Unit (GRU)

The GRU operates as a special RNN that fixes the gradient vanishing issue present in standard RNNs. The GRU functions similarly to LSTM cells while requiring fewer computational parameters which leads to more efficient processing. The GRU incorporates two key gates which form its structure:

- **Update Gate:** The system determines which historical data should be transferred into the upcoming computational interval.
- **Reset Gate:** This factor controls the level of prior memory information that should be neglected or forgotten.

These gating mechanisms enable the model to determine which information will stay as it analyzes sequence data. The GRU achieves effective long-term dependency identification and pattern discovery in sequential data through this method which surpasses standard RNN functionality. Mathematically, the GRU model can be expressed as Equation (2 to 5):

Update Gate:

$$z_t = \sigma \left(W_z \cdot [h_{t-1}, x_t] + b_z\right) \quad (2)$$

Reset Gate:

$$\sigma \left(W_r \cdot [h_{t-1}, x_t] + b_r\right) \quad (3)$$

Candidate Memory:

$$\tilde{h_t} = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t] + b_h) \quad (4)$$

Final Hidden State:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h_t} \quad (5)$$

At timestamp t the hidden state measurement is h_t and the sigmoid activation function is denoted by σ and the hyperbolic tangent function by tanh. W_z, W_r, W_h and b_z, b_r, b_h are learnable parameters. The GRU's gating mechanism lets it effectively learn and process sequential educational data that serves as a foundation for predictive modeling tasks, including student performance prediction.

Model Evaluation

Model evaluation for classification models assesses the decision accuracy of classifiers through assessments of their classification decisions [15]. The enumerated counts in this table receive the identification as a confusion matrix. A plot of a NxN matrix's categorization is called a confusion matrix, and it may be found in Table II.

Table 2: Confusion Matrix

	Predicted	
Actual	**T**	**F**
T	TP	FN
F	FP	TN

In a confusion matrix, TP refers to the instances TN are instances in which the classifier accurately predicts the negative class, whereas the positive class is properly predicted in these instances. Wrong predictions of a positive result for a negative occurrence are known as FP, whereas wrong predictions of a negative outcome for a positive instance are known as FN. These four metrics are essential for assessing the performance of categorization models. It is possible to represent the assessment metrics using these formulas:

Accuracy

Accuracy serves as a gauge for model's quality. It is assumed that if model performs brilliantly, it will be closer to 1 in Equation (6).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Precision

Precision is the quantity of significant items chosen. Specifically, the proportion of anticipated values that are really accurately predicted. For example, look at Equation (7): It stands for the probability that a positive sample will be correctly classified.

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

Recall

If accuracy is more like one, then the expectations are increasingly exact. The number of pertinent things selected is indicated by recall. The model's ability to detect assaults is measured by the recall detection rate, as shown in Equation (8).

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1-Score

Accuracy is measured by the F-score is really the average of accuracy and recall. The F1 Measure, a more comprehensive evaluation metric than accuracy, is shown in Equation (9):

$$F1 - Score = \frac{2*Precision*Recall}{(Precision+Recall)} \quad (9)$$

Loss Function

The loss function guides model optimization during training, ensuring convergence towards minimal error. For binary classification, the Loss is employed, defined as Equation (10):

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y \,.\log(y_i) + (1 - y_i).\log(1 - \hat{y}_i)] \quad (10)$$

where: N is the total number of samples, $y_i \in \{0,1\}$ is the true label for the i[th] instance, $y\hat{}i \in [0, 1]$ is the predicted probability of being a fraudulent credit card item. This loss penalizes the model more heavily when it makes confident but incorrect predictions, thus promoting accurate probabilistic outputs.

Result Analysis and Discussion

An experimental PC with a Core™ i5-8250U CPU operating at 1.8 GHz, 12 GB of RAM, and Windows 10 Professional 64-bit was used to conduct and assess this study. The DL and ML models for predicting student performance were implemented using Python 3. A thorough assessment of the suggested GRU model on a sizable educational dataset is one of the outcomes. Additionally, visual representations such as a confusion matrix, an accuracy curve, and the predictive power and resilience of the model are further confirmed using a loss curve. As shown in Table III, the GRU model achieved remarkable results. GRU demonstrated exceptional precision by achieving 99.70% accuracy which proved its overall stability. With a precision of 98.60% and a recall of 96.30%, the model effectively balanced correct predictions and relevant identifications. The F1-score value of 97.40% validates how well the GRU model improves student performance prediction models.

Table 3: Results of the Gru on the Educational Dataset for Student Performance Prediction

Evaluation Matrix	Gated Recurrent Unit (GRU)
Accuracy	99.70
Precision	98.60
Recall	96.30
F1-Score	97.40

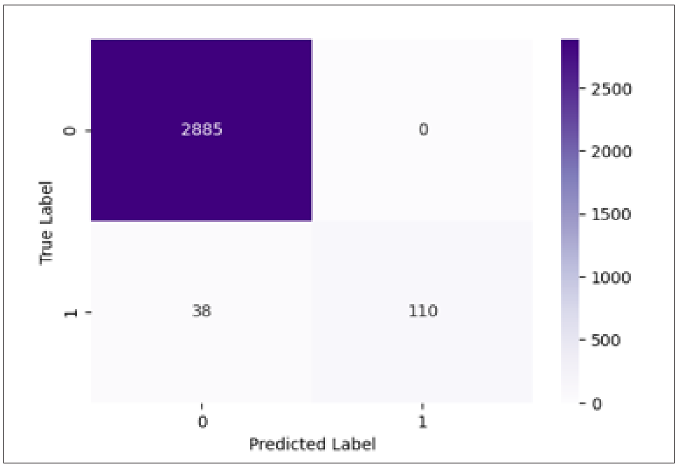

Figure 3: Confusion Metrix of GRU

The predictive capability of the GRU model for analyzing large-scale educational datasets appears strong according to the confusion matrix shown in Figure 4. Through these predictions, the model demonstrated accuracy by classifying 2885 true negatives and 110 true positives alongside 38 false negatives with no false positives detected. The high accuracy of the classification indicates that the GRU model shows promise for enabling strategic data approaches that enhance educational results among students.

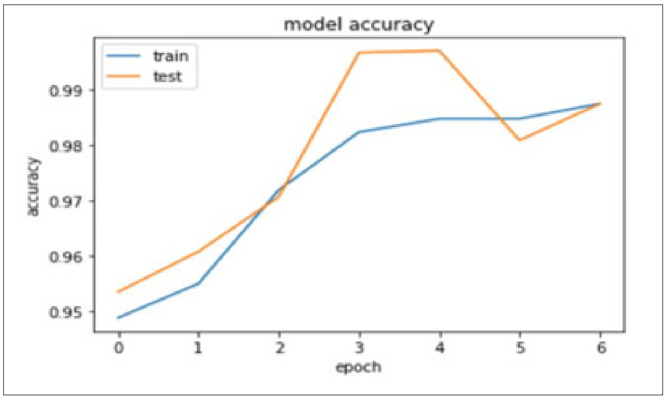

Figure 4: Accuracy Curve of the GRU

The Figure 4 accuracy curve reveals how the GRU model optimizes training and validation accuracy to reach a 99% test accuracy by the third epoch. The quick learning speed, combined with low overfitting, points to the model's excellent generalization capacity for big educational datasets. Such performance validates the suitability of GRU architectures for predictive modeling tasks aimed at optimizing student performance outcomes.

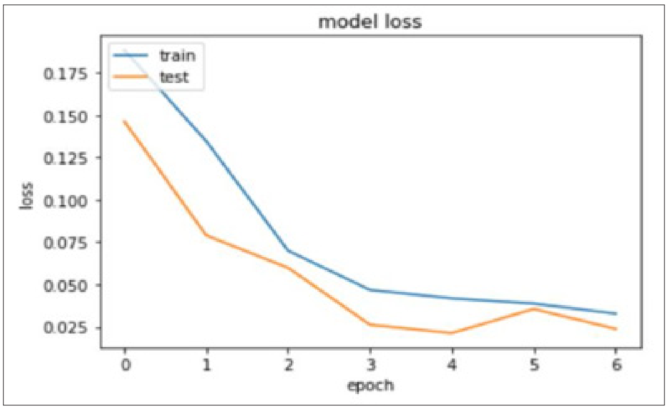

Figure 5: Loss Curve of the GRU

Figure 5 shows the GRU model's training and testing loss progression over epochs. Initially, the training loss decreases sharply from approximately 0.18 to 0.07 by the second epoch, while the testing loss similarly drops from around 0.15 to 0.06. By the third epoch, the testing loss further reduces to nearly 0.025, and the training loss stabilizes around 0.045. Beyond the third epoch, both curves exhibit minimal fluctuations, suggesting the model has effectively converged. The close correspondence between training and testing loss curves indicates robust generalization capability with no substantial signs of overfitting.

Comparative Analysis

The Comparative Analysis section evaluates Table IV displays the suggested GRU model's predictive performance in comparison to traditional models of 77.92%, and an F1-score of 78.47%. In comparison, the DT model reported the lowest performance, with all four metrics recorded at 62.26%. Analyses revealed that the DT model generated the most deficient outcomes by achieving 62.26% across each metric. Research findings verify that DL and, particularly, GRU demonstrate superior ability in accurately predicting student performance through analysis of extensive educational data.

Table 4: Comparative Analysis between Propose and Existing Model Performance for Predicting Student Outcomes

Models	Accuracy	Precision	Recall	F1-Score
ANN[16]	77.04	79.19	77.92	78.47
DT[17]	62.26	62.26	62.26	62.26
GRU	99.70	98.60	96.30	97.40

In educational data mining, the proposed GRU-based model delivers considerable value through its accurate and effective student performance prediction method. Through its sequential data processing and class imbalance management capabilities, the model provides early detection of failing students for institutions to intervene before it is too late. The system leads to enhanced academic results along with optimized budget distribution and data-based choices within educational settings.

Conclusion and Future Scope

Educational institution exists to provide their student body with the best possible academic knowledge and educational achievements. It is necessary to correctly identify and intervene with pupils who need extra help in order to achieve this educational goal. The study shows how the GRU model predicts student outcomes by efficiently using extensive educational data from three Assam Indian institutions. With a 99.70% accuracy rate, 98.60% precision, 96.30% recall, and a 97.40% F1-score, the Gated Recurrent Unit model outperformed conventional machine learning techniques. The results of the experiment demonstrate that this model remains stable when handling imbalanced data and identifying complex patterns in time-dependent sequences. The model exhibits powerful generalization proficiency according to results obtained through confusion matrix analysis, together with loss/accuracy curves. One major limitation emerged from using dataset information solely from three colleges and limiting the study's ability to become widely applicable. Future research should expand by adding multiple features to the model alongside broadening the dataset scope to increase prediction performance through hybrid modeling approaches. The development of real-time prediction systems for dynamic academic support offers promising potential to increase system practicality [16-41].

References

1. Harvey JL, SAP Kumar (2019) A Practical Model for Educators to Predict Student Performance in K-12 Education using Machine Learning. IEEE Symp Ser Comput Intell SSCI 3004-3011.
2. Gatsheni BN, Katambwa ON (2018) The Design of Predictive Model for the Academic Performance of Students at University Based on Machine Learning. J Electr Eng 6: 4.
3. Acharya A, Sinha D (2014) Early Prediction of Students Performance using Machine Learning Techniques. Int J Comput Appl 107: 37-43.
4. Hamoud A, Hashim AS, Awadh WA (2018) Predicting student performance in higher education institutions using decision tree analysis. Int. J. Interact. Multimed. Artif. Intel 5 : 26–31.
5. Iqbal MS, Luo B, (2019) Prediction of educational institution using predictive analytic techniques. Educ Inf Technol.
6. Kaur A (2018) Machine Learning Approach to Predict Student Academic Performance. Int. J. Res. Appl. Sci. Eng. Technol.
7. Rimadana MR, Kusumawardani SS, Santosa PI, Erwianda MSF (2019) Predicting Student Academic Performance using Machine Learning and Time Management Skill Data. International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) 511-515.
8. Hasan HMR, Rabby AKMS A, Islam MT, Hossain SA (2019) Machine Learning Algorithm for Student's Performance Prediction. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) 1-7.
9. Dubey A, Mani M (2019) Using Machine Learning to Predict High School Student Employability – A Case Study. IEEE International Conference on Data Science and Advanced Analytics (DSAA) 604-605.
10. Pereira FD, Oliveira EHT, Fernandes D, Cristea A (2019) Early Performance Prediction for CS1 Course Students using a Combination of Machine Learning and an Evolutionary Algorithm. IEEE 19th International Conference on Advanced Learning Technologies (ICALT) 183-184.
11. Jain A, Solanki S (2019) An Efficient Approach for Multiclass Student Performance Prediction based upon Machine Learning. International Conference on Communication and Electronics Systems (ICCES) 1457-1462.
12. Casuat CD, ED Festijo (2019) Predicting Students' Employability using Machine Learning Approach. IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS) 1-5.
13. Kaunang FJ, Rotikan (2018) Students' Academic Performance Prediction using Data Mining. Third International Conference on Informatics and Computing (ICIC) 1-5.
14. H Al-Shehri (2017) Student performance prediction using Support Vector Machine and K-Nearest Neighbor. IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE) 1-4.
15. Mueen A, Zafar B, Manzoor U (2016) Modeling and predicting students' academic performance using data mining techniques. Int J Mod Educ Comput Sci 8: 36-42.
16. Altabrawee H, Ali OAJ, Ajmi SQ (2019) Predicting Students' Performance Using Machine Learning Techniques, J. Univ. Babylon Pure Appl Sci 27: 194-205.
17. Francis BK, Babu SS (2019) Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. J Med Syst 43: 162.
18. Kho and Vea (2017) Credit card fraud detection based on transaction behavior. in TENCON 2017 - 2017 IEEE Region 10 Conference, IEEE 1880-884.
19. Katnapally N, Chinta PCR, Routhu KK, Velaga V, Bodepudi V, et al. (2021) Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights. American Journal of Computing and Engineering 4: 35-51.
20. Karaka LM (2021) Optimising Product Enhancements Strategic Approaches to Managing Complexity. SSRN.
21. Chinta PCR, Karaka LM Agentic AI and reinforcement learning: Towards more autonomous and adaptive ai systems. Journal for Educators, Teachers and Trainers https://jett.labosfor.com/index.php/jett/article/view/2699.
22. Boppana SB, Moore CS, Bodepudi V Jha, KM Maka SR, Sadaram G (2021) AI And ML Applications In Big Data Analytics: Transforming ERP Security Models For Modern Enterprises 2197-5523.
23. Chinta PCR, Katnapally N, Ja K, Bodepudi V, Babu S, et al. (2022) Exploring the role of neural networks in big data-driven ERP systems for proactive cybersecurity management. Kurdish Studies.
24. Chinta PCR (2022) Enhancing Supply Chain Efficiency and Performance Through ERP Optimisation Strategies. Journal of Artificial Intelligence & Cloud Computing 1: 47310-47363.
25. Sadaram G, Sakuru M, Karaka L M, Reddy MS, Bodepudi V, et al. (2022) Internet of Things (IoT) Cybersecurity Enhancement through Artificial Intelligence: A Study on Intrusion Detection Systems. Universal Library of Engineering Technology.
26. Moore C (2023) AI-powered big data and ERP systems for autonomous detection of cybersecurity vulnerabilities. Nanotechnology Perceptions 19: 46-64.
27. Chinta PCR (2023) The Art of Business Analysis in Information Management Projects: Best Practices and Insights 10.
28. Chinta PCR (2023) Leveraging Machine Learning Techniques for Predictive Analysis in Merger and Acquisition (M&A). Journal of Artificial Intelligence and Big Data, 3: 10-31586.
29. Krishna Madhav J, Varun B, Niharika K, Srinivasa Rao M, Laxmana Murthy K (2023) Optimising Sales Forecasts in ERP Systems Using Machine Learning and Predictive Analytics. J Contemp Edu Theo Artific Intel: JCETAI 104.
30. Maka SR (2023) Understanding the Fundamentals of Digital Transformation in Financial Services: Drivers and Strategic Insights SSRN.
31. Routhu KishanKumar, Katnapally, Niharika, Sakuru, Manikanth (2023) Machine Learning for Cyber Defense: A Comparative Analysis of Supervised and Unsupervised Learning Approaches. Journal for ReAttach Therapy and

Developmental Diversities 6.
32. Chinta Purna Chandra Rao, Moore Chethan Sriharsha (2023) Cloud-Based AI and Big Data Analytics for Real-Time Business Decision-Making 36: 96-123.
33. Krishna Madhav J, Varun B, Niharika K, Srinivasa Rao M, Laxmana Murthy K (2023) Optimising Sales Forecasts in ERP Systems Using Machine Learning and Predictive Analytics. J Contemp Edu Theo Artific Intel: JCETAI 104.
34. Bodepudi V (2023) Understanding the Fundamentals of Digital Transformation in Financial Services: Drivers and Strategic Insights. Journal of Artificial Intelligence and Big Data. 3:10-31586.
35. Jha KM, Bodepudi V, Boppana SB, Katnapally N, Maka SR, et al. (2023) Deep Learning-Enabled Big Data Analytics for Cybersecurity Threat Detection in ERP Ecosystems.
36. Kuraku S, Kalla D, Samaah, F, Smith N (2023) Cultivating proactive cybersecurity culture among IT professional to combat evolving threats. International Journal of Electrical, Electronics and Computers 8.
37. Kalla D, Smith N, Samaah F, Polimetla K (2022) Enhancing Early Diagnosis: Machine Learning Applications in Diabetes Prediction. Journal of Artificial Intelligence & Cloud Computing. 191: 2-7.
38. Kuraku DS, Kalla D (2023) Impact of phishing on users with different online browsing hours and spending habits. International Journal of Advanced Research in Computer and Communication Engineering 12.
39. Kalla D, Kuraku S (2023) Phishing website url's detection using nlp and machine learning techniques. Journal of Artificial Intelligence 5: 145.
40. Kuraku DS, Kalla D, Samaah F (2022) Navigating the link between internet user attitudes and cybersecurity awareness in the era of phishing challenges. International Advanced Research Journal in Science, Engineering and Technology 9: 116-124.
41. Kuraku DS, Kalla D, Smith N, Samaah F (2023) Exploring How User Behavior Shapes Cybersecurity Awareness in the Face of Phishing Attacks. International Journal of Computer Trends and Technology 71: 74-79.