Hybrid Deep Learning Approach with Attention Mechanism for Predicting Student Success in Higher Education Institutions

M. Nazir [1], A. Noraziah [1,*], Abdullah Alsaleh [2,*], and M. Rahmah [1]

[1] Faculty of Computing, University Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia
[2] Department of Computer Engineering, College of Computer and Information Sciences, Majmaah University, Majmaah, Saudi Arabia
Email: pcp21002@student.umpsa.edu.my (M.N.); noraziah@umpsa.edu.my (A.N.); alsaleh@mu.edu.sa (A.A.); drrahmah@umpsa.edu.my (M.R.)
*Corresponding author

***Abstract*—Higher education institutions are focusing increasingly on targeted interventions by supporting students with academic challenges, improving their overall learning outcomes. In this context, hybrid Deep Learning (DL) approaches have emerged as leaders in the construction of recommendation systems capable of predicting students-at-risk by analyzing patterns of historical academic data. This paper proposes a hybrid DL model integrating an attention mechanism, specifically tailored to predict the success of students and to identify students-in-need of academic support. Our primary dataset, sourced from the Majmaah University of the Kingdom of Saudi Arabia, contained 146,989 records and 26 features, including a wide variety of student-related features crucial for making academic achievement predictions. Comparison with the current models like Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), Neural Network Random Weight (NNRW), Long Short-Term Memory (LSTM), and NNRW-LSTM demonstrates the exemplary effectiveness of the proposed NNRW-LSTM model integrated with an attention mechanism. The proposed model achieved remarkable accuracy of 93.53%, 92.85% of the recall rate, 92.6% of precision, and 92.55 of the F1-Score. These results document the prospect of this hybrid DL model integrated with an attention mechanism to assist schools in identifying students-at-risk proactively, facilitating interventions well in time and helping curb academic underperformance and dropouts.**

***Keywords*—deep learning, student success, higher study, Neural Network Random Weight-Long Short-Term Memory (NNRW-LSTM), attention mechanism**

I. Introduction

Academic success is a complex phenomenon in the modern context of higher education that is influenced by a variety of factors, such as socio-economic status, mental involvement, online learning activity, and institutional resources. Although much effort has been put on the enhancement of student outcomes, universities still experience difficulties in detecting students who are likely to fail or leave school, which often leads to late interventions and increased dropout rates [1]. Traditional academic support systems, while helpful, are incapable of processing huge volumes of educational data and in most cases cannot depict the intricate nuances of students' behavior. As a result, educational institutions are switching to Artificial Intelligence (AI) and Machine Learning (ML) technologies for predicting academic success and supplying data-oriented and individualized interventions.

Deep Learning (DL) is one of the methods that have emerged as a strong paradigm that can be used to represent complex, non-linear relationships in high-dimensional educational data [2]. The DL techniques can be applied to the extraction of hidden and indistinguishable features of various data sources such as learning management systems, engagement logs, and academic records; thus, they will be able to predict student performance and risk levels more accurately Research has highlighted the potential of deep learning hybrids that are composed of various neural elements to achieve better accuracy and generalization [3].

The Neural Network Random Weight (NNRW) models have been very helpful in extracting features from structured data, revealing hidden connections between student attributes and academic performance, and similarly, the Long Short-Term Memory (LSTM) networks have been successful in discovering longitudinal connections among the sequential data like the semester-to-semester academic history or the engagement history [4]. However, LSTM networks are superior at forecasting, yet they often come across as less interpretable and more computationally demanding. To solve these problems, researchers have incorporated attention mechanisms into the models which allow them to focus on the most relevant features or time steps, thereby improving

Manuscript received September 9, 2025; revised October 10, 2025; accepted November 24, 2025; published March 10, 2026.

both performance and interpretability [5]. Attention layers allow determining the most significant factors, including the frequency of participation, assignment delays, or changes in scores, which affect academic outcomes and provide educators and policymakers with actionable information.

In spite of these developments, current models can be seen to be frequently unable to trade off interpretability, scalability, and computational efficiency when used on large, heterogeneous educational data. In order to fill these gaps, this paper suggests a hybrid deep learning architecture that combines NNRW and LSTM architectures with an attention mechanism. The NNRW component supports effective feature extraction and generalization of models, whereas the LSTM component learns sequentially. The attention mechanism also results in a higher level of interpretability as the important determinants of academic success are highlighted.

The key aim of the study is to show the possibility of the suggested hybrid model to predict student success and identify at-risk individuals accurately and, thus, to support early intervention strategies. The proposed approach will be compared to the baseline models like Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), NNRW, and standalone LSTM to further the use of deep learning in educational analytics. Finally, the results are likely to be used in the creation of intelligent and data-driven support systems to improve academic performance and decrease student turnover in higher education settings.

Increased sophistication of higher education settings has put institutions under unprecedented pressure to track, forecast, and improve the academic success of students. Classical approaches, for example, by-hand academic advisement or algorithmic early warning systems, typically do not effectively detect students at risk because they are limited by their capacity to accommodate huge, heterogeneous, and time-series data comprised of grades, attendance records, engagement signals, and behavioral symptoms. Furthermore, current machine learning approaches, while beneficial, typically assume equal importance of all features and may overlook fine-grained patterns predicting early symptoms of academic difficulty.

Some recent works have also pointed to the power of deep learning architectures, specifically LSTMs and GRUs, to learn sequential relationships in students' data. Nevertheless, the limitation of these architectures remains feature incomprehensiveness and prioritization, limiting the ability of the administrator and policymaker to draw actionable insights. Due to this, there is an urgent requirement for models which are not only capable of predicting students' success accurately but also give outputs in the form of explainable, feature-weighed outputs to inform practical interventions.

In response to these challenges, the current study is motivated by three primary objectives:

(1) Enhanced Predictive Accuracy: To devise a hybrid deep learning framework that incorporates random weight initialization of neural networks along with LSTM layers aimed at enhancing model convergence, stability, and finally, accuracy of predicting student performance.
(2) Actionable Insights via Attention Mechanism: To incorporate an attention mechanism that draws attention to the most salient features impacting student success, thereby allowing institutions to carry out focused interventions and devise resource allocation strategies based on this information.
(3) Bridging Research and Policy Application: To create a user-friendly instrument that not only provides a mathematical model but also interpretable results that can easily be utilized in academic advising, early intervention programs, and policy decisions, thus minimizing dropout rates and enhancing educational outcomes.

The present study intends to link the latest machine learning advancements with actual higher education administration workflow by offering a model that is both scientifically valid and geographically viable for the institution. It achieves this by making the drawbacks of earlier approaches very evident and focusing on the dual goals of prediction accuracy and interpretability.

II. Literature Review

Deep Learning (DL) models have recently become very popular in solving complicated problems in various fields. They have been used successfully in healthcare to predict diseases and model biology [6], and in real-world systems like customer service, sports forecasting, autonomous driving, and weather prediction [7, 8]. The online learning platforms have been widely used in the education sector, and this has resulted in large volumes of data, which have been analyzed using DL models to identify concealed patterns and predict student performance [9–11].

A number of research works have led to the development of AI- and DL-based educational analytics. As an example, Naseer *et al.* [12] used a GRU-based model to detect students who require academic assistance with high accuracy (99.70%). The weakness of their model, however, is that it is black-box, i.e. not interpretable and transparent, which is essential to educational stakeholders to know why some students are labeled as at-risk.

The effect of personalization was confirmed by Bressane *et al.* [13], who suggested an adaptive AI learning platform that increased student performance by 25%. However, the research was restricted to a small local sample and failed to test cross-context generalization and applicability in different educational settings.

On the same note, Alshamaila *et al.* [14] developed a fuzzy-AI based decision support system that incorporates ANN to support learning disabled students. Although useful, the model was based on questionnaire-based inputs and thus was not as flexible to real-time data streams that are common in e-learning systems.

Fahd *et al.* [15] addressed the problem of class imbalance through convolution-based DL architectures and balancing techniques. Though the study improved the fairness of prediction, it failed to investigate model

interpretability and temporal adaptability, which are important in the context of dynamic learning.

A more general meta-analysis by Villegas-Ch *et al.* [16] analyzed 89 studies on ML-based education and found that student performance prediction was significantly improved. Nonetheless, they found that the majority of previous studies were algorithm-focused, and they did not provide a single framework that combines explain ability, longitudinal tracking, and adaptive feedback mechanisms.

In a different view, Abuzinadah *et al.* [17] used AI and ML in sustainability education to find at-risk students and focus on ethics and fairness. Although it was an important contribution, the model did not provide much information on real-time adaptation and federated scalability, which are essential in the contemporary digital campuses.

Latif *et al.* [18] demonstrated a good level of accuracy (up to 98.25) in the classification of student performance based on ensemble learning, but their method was mainly static and institution-specific, which limits the ability to generalize the results.

On the same note, Pallathadka *et al.* [19] compared the traditional ML algorithms (Naive Bayes, Support Vector Machine, ID3, C4.5) in predicting student performance, focusing on the accuracy but ignoring the incorporation of contextual behavioral variables and nonlinear temporal learning patterns.

Finally, the study by Pacheco-Mendoza *et al.* [20] reported that age, duration of study, and the use of AI tools, among other factors, were statistically significant predictors in their proposed model. The research gave an explanation of the model that was not deep and the combination of different data sources besides the study did not yield much which was particularly required in showing the multi-dimensionality of student success, even though it was statistically valid ($R^2 = 0.9075$).

To conclude, existing research proves that the use of deep learning technology for educational effectiveness is evident, although, in most cases, research results are fragmented and research gaps converge into the following three areas that remain unaddressed:

(1) Interpretability: Most DL-based prediction systems act as opaque models, offering high accuracy without transparency or explain ability for educators.
(2) Scalability and Generalization: Prior models often depend on localized datasets, limiting their ability to adapt to diverse institutions or larger-scale deployments.
(3) Real-Time Adaptation and Feedback: Few studies incorporate adaptive mechanisms that dynamically adjust predictions based on continuous learning data streams.

The current research aims to close these gaps by proposing a hybrid deep learning approach with attention mechanism—combining interpretability, scalability, and adaptive learning features to not only improve the predictive accuracy but also increase the practical applicability of student success models in higher education.

III. Materials and Methods

The technique assists in research by ensuring that a study possesses a systematic research procedure, adheres well to the objectives, and completes on schedule. If methodologically concise and specific, the technique acts as the foundation of the organization, principles, and theoretical foundation upon which the research may be conducted. This not only ensures the reliability and validity of the discovery but additionally provides very clear guidance for future replication and expansion. In this text, various current-state-of-the-art DL algorithms, including ANN, CNN, LSTM, GRU, NNRW, NNRW-LSTM, and NNRW-LSTM incorporated with an Attention Mechanism, are explored. These practices are chosen for the ability to describe intricate patterns of the data of students' performance and yield an efficient mechanism of early estimation of students' success.

A. Approach Overview

Fig. 1 depicts a hybrid NNRW-LSTM model structure that has an attention mechanism. The NNRW and LSTM model is created to forecast the success of students using both rapid processing and sequential learning methods. It starts with data preparation, during which the data about the students including the number of hours of study, attendance, engagement, previous academic performance, and behavioral indicators are gathered and preprocessed. This makes the data structured and ready to be inputted in the model. After the data is ready, it is put into the input layer which is the point of entry into the neural network. Every input is a distinctive characteristic that leads to student performance. The information is then advanced to the NNRW model where the weights of the input are randomly assigned rather than being learned by backpropagation. This significantly speeds up the entire training procedure and is a very efficient method to depict the intricate interconnections among the input features. The activations from the hidden layer of NNRW are then turned into the output, which subsequently becomes the input for the following stage. The LSTM network receives the output from NNRW and employs it to reveal the sequential relationships in the patterns of student performance. The LSTM architecture is such that it incorporates three major gates: the Forget Gate, which serves to evaluate which past information is to be discarded; the Input Gate, which serves to evaluate which new information is to be added to the memory; and the Output Gate, which is responsible for generating the final hidden state that will be utilized for prediction. This will enable the model to identify trends in student behavior, like improvement or deterioration with time. An attention layer is added to make the interpretation even more interpretable. The algorithm assigns various weights to all the features and thus, the important ones like latest exam results, attendance, or learning methods have a greater influence on the prediction process. After that, the merged layer integrates the analyzed data and produces a final pattern representation of the student's learning. The last and output layer generates the prediction and this might be either a pass/fail classification or a ranking of GPA in

figures. This NNRW-LSTM model is a fast and precise way of predicting student success by blending the speedy calculation of NNRW with the gradual pattern learning of LSTM. The model can be very useful in spotting at-risk students early and consequently, providing timely support. It can also be utilized to give customized learning suggestions and help learning institutions make evidence-based choices. This hybrid model will guarantee real-time, interpretable, and accurate student success predictions by utilizing the DL techniques, which will eventually result in better academic outcomes.

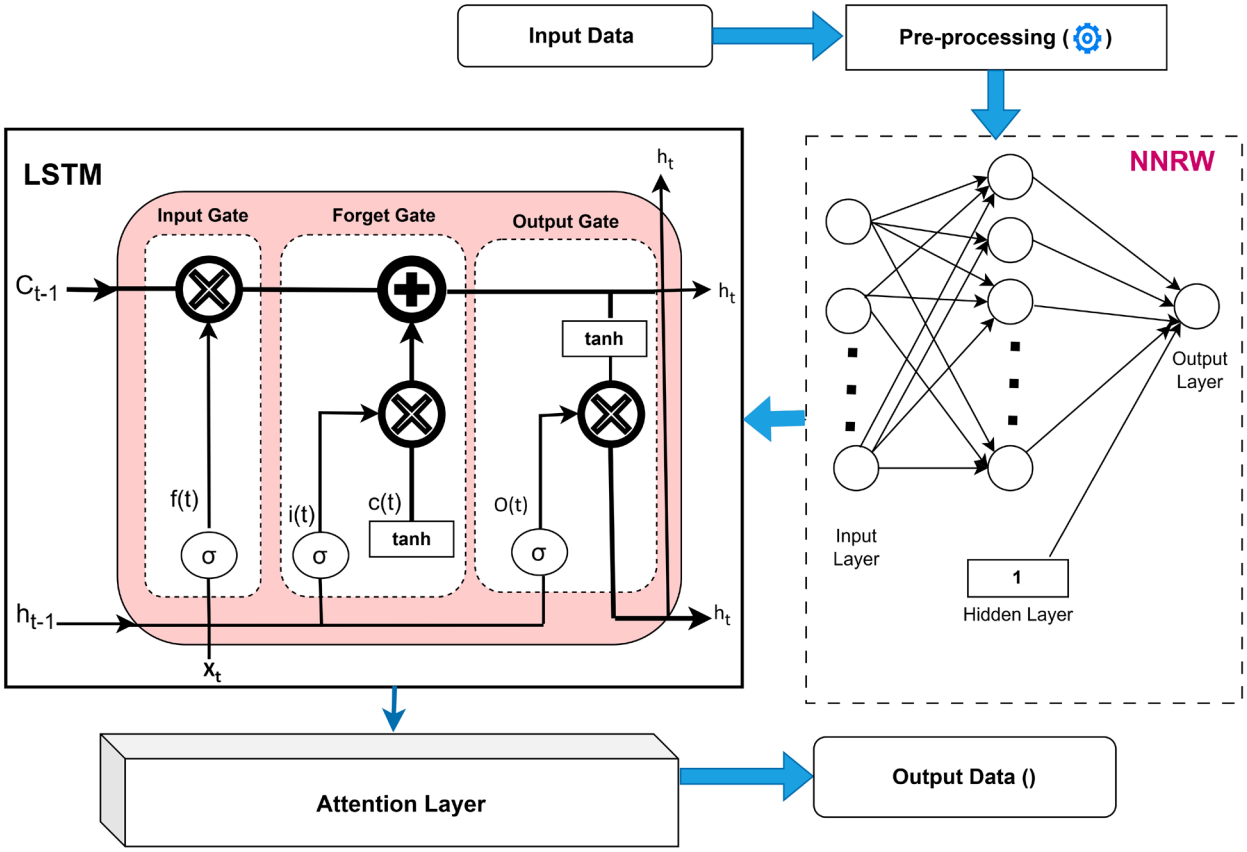

Fig. 1. Proposed hybrid model.

TABLE I. DATABASE DESCRITPION

S.No	Column Name	Data Type	Non-Null Count	Description (Short Meaning)
1	STUDENT_ID	int64	146,989	Unique student identifier
2	STUDENT_STATUS	object	146,989	Academic status (active/withdrawn etc.)
3	SEMESTER	int64	146,989	Semester number
4	CONFIRMED_MARK	float64	135,287	Confirmed exam score
5	LETTER_GRADE_S	object	137,822	Letter grade (A/B/C/...)
6	WARNING_PERCENT	float64	146,989	Academic warning risk score (%)
7	GENDER	object	146,989	Male/Female
8	CAMPUS_NO	int64	146,989	Campus identifier
9	FACULTY_NO	int64	146,989	Faculty code
10	FACULTY	object	146,989	Faculty name
11	MAJOR	object	146,989	Student major
12	PLAN_EDITION	int64	146,989	Curriculum/plan version
13	CUM_GPA	float64	138,615	Cumulative GPA
14	SEMESTER_GPA	float64	138,631	GPA for that semester
15	STUDY_LEVEL	int64	146,989	Level (freshman/sophomore etc.)
16	SECTION_SEQ	int64	146,989	Section ID or sequence
17	COURSE_CODE	object	146,989	Registered course
18	CRS_HOURS	int64	146,989	Credit hours
19	REG_YEAR	int64	146,989	Academic year
20	STATUS_DESC	object	146,989	Status text label
21	PLAN_HRS	int64	146,989	Total plan hours
22	BIRTH_DATE	object	146,989	Date of birth
23	SPONSOR	object	7488	Scholarship/financial sponsor
24	ADMIT_TERM	int64	146,989	First admitted term
25	NATIONALITY	object	146,989	Student nationality

All features have been concatenated and it is thus used by the model to do a final prediction of the student being a success or not: for example, predicting if a student is likely to pass or need extra attention. Finally, in the more sophisticated models, the Merged Layer pools the output of the fully connected layer with other features, in such a way that all relevant information is integrated before formulating the weak prediction.

B. *Details of the Dataset*

In this study, the data were collected from the Majmaah University, Kingdom of Saudi Arabia. The process of data collection at Majmaah University adheres to the official approaches for collection of datasets. A submitted proposal was accepted by the relevant authorities to unify the contents of this system. The data is kept in different repositories; student performance data are in examination department repository, demographic information and prerequisite course records housed in the admissions office.

Table I presents the structure and composition of the dataset, illustrating the categories of data fields and the total number of records collected. In total, the dataset comprises 146,989 instances and 26 features, including academic performance indicators (GPA, course grades), demographic characteristics (age, gender, residency status), and behavioral/engagement attributes (attendance, course history, and digital learning interactions).

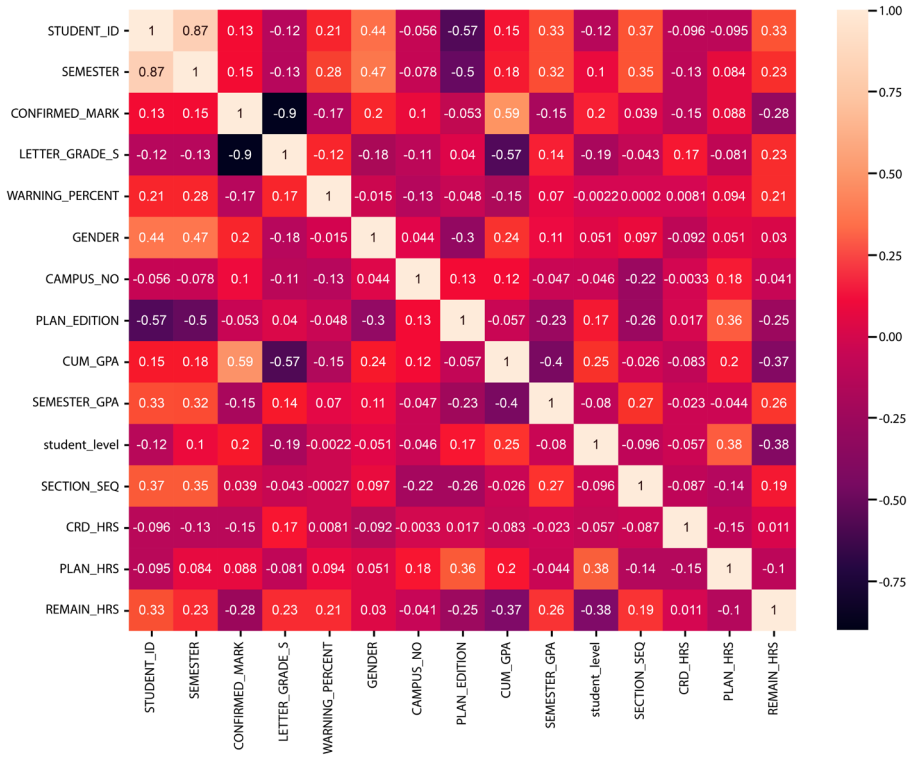

Fig. 2. Data correlation.

The correlation heatmap of the chosen features is presented in Fig. 2, and it shows the statistical associations between the input variables. The intensity of the color indicates the strength and direction of correlation (positive or negative). Attributes related to academic performance, including cumulative GPA, grades in the past semester and course completion status, demonstrate strong relationships with student outcome labels, whereas demographic variables demonstrate weaker or indirect relationships. The correlation analysis gave important information on the possible multicollinearity problems and helped in the process of feature selection before the model training.

Fig. 3 shows the distribution of the students in various categories of academic status (e.g., “Successful”, “At-Risk”, “Under Probation”, “Dropped”, etc.). The plot is a visual representation of the existing class imbalance in the data set, where the successful students are significantly more in number compared to the at-risk or dropped-out groups. This imbalance informed the use of balancing and attention-based modeling methods in the proposed hybrid architecture, whereby minority classes were not underrepresented in the training and evaluation.

The data used in the present research consists of 146,989 records, 26 columns, which is a rather diverse and complete set of data about students that is important to predict academic performance and identify at-risk individuals. This rich data set comprises diverse academic measures, including grades, attendance data, and assessment measures, as well as behavioral and engagement measures, including engagement in classroom activities, use of resources, and extracurricular activities. Inclusion of academic and non-academic variables gives a comprehensive picture of the variables that affect student success and therefore the model is in a position to capture complex patterns that might not be obvious at a glance using the traditional analysis. Prior to preprocessing, the initial examination of the dataset revealed as the first observation that a few columns had missing values, which

could impact the predictive model's accuracy and reliability. The mode imputation method was used to overcome this problem and the missing entries were filled with the most repeated value in each column. This method not only preserved the initial distribution of the dataset but also moderated the chances of biasing the data thus preventing the cleaning process from losing essential insights. The dataset was not only intact but was also complete, something very essential for the development of a robust and reliable model through missing values management. The correlation was carried out after data cleaning to understand how different features related to each other. This analysis was to find out which attributes of students were the most important and thus the model could focus on the relevant aspects. There were high correlations among certain variables, such as attendance rates and final grades, which indicated the main indicators that could be used as early warning signs of academic difficulties. With this knowledge of the relationships, the predictive model will be able to focus on features that have a significant contribution to student outcomes and enhance the accuracy of performance predictions. Also, the dataset contains student status classifications, which are also useful in terms of various categories of student performance. These categories divide students into high-performing, at-risk, and dropout-prone students based on the level of academic engagement and achievement. The distribution of students in these categories provides a critical view on the training and testing of the NNRW-LSTM with Attention model, whereby it can be able to detect patterns in the dissimilar student profiles. This segmentation also helps to personalize the intervention strategies and support each student category individually, depending on the needs of each student category.

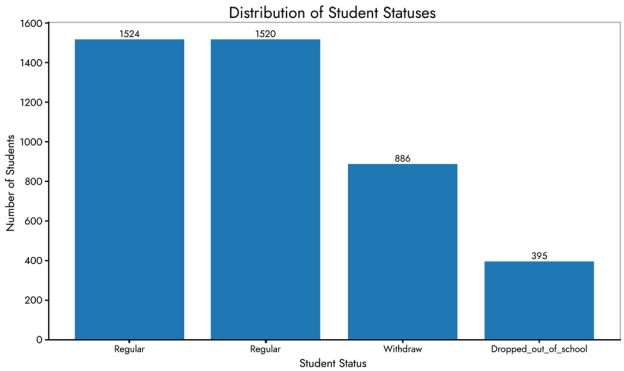

Fig. 3. Students' status.

C. Data Preprocessing Techniques

The data underwent preprocessing before it was analyzed. Inconsistent raw data can occur (for example, with missing and duplicate entries, and text in a format that the software cannot read, i.e., incompatible format). Preprocessing is a process of converting 'raw' data into a more presentable and cleaner format, so that analytical methods can be applied effectively. This key process refines the accuracy of the subsequent analysis. Preprocessing tasks that are frequently performed include dealing with missingness, converting textual data to numerical forms, as well as extracting meaningful signals from noise. The purpose of pre-processing is to help approaches discover stronger patterns and to have access to better predictions. The most frequently used preprocessing techniques in student success prediction are summarized in the following sub-subsection

1) Handling missing values

Handling missing information in data sets is a vital element to minimize bias and have our methods continue to provide accurate results. There are many ways to deal with this, such as deleting the missing values, replacing them with other values, such as the mean or median.

2) One hot encoding

One-hot encoding converting text to numeric features by treating each unique word or token as a binary input. Documents are represented as bags of these orthogonal hot vectors: sparse, but unambiguous codes, having a single (fixed) 1 signifying the presence of any given distinctive term. One-hot encoding matrices simplify textural feature quantification using one hot encoding matrices of vector length being the size of the vocabulary, not raw text. Using the presence of words as binary indicators as features allows for such quantitative analysis while preserving the possibility to reverse the mapping of found patterns to original token sequences. One hot encoding is input to most of the machine learning algos and is observed to perform better than methods without explicit word encoding. The ease of counting words and placing them into orthogonal buckets makes one hot representation an extremely useful way to extract features from text data.

3) Normalization

Text normalization in this work is through the SymSpell algorithm, which is an efficient method of spelling correction and standardization of textual inputs by means of a frequency-based lookup and edit distance optimization. All the text is converted to lowercase, contractions are expanded, and variant spellings are corrected with the help of a precompiled word dictionary with an edit distance threshold = 2. Numbers are written in words, and the standard forms of terms are obtained by lemmatization. Thus, the whole vocabulary in the words of the dataset is uniform, lexical variability is minimized, and the consistency for the subsequent text analyses is improved.

4) Data cleaning and handling missing values-justification

In the current study, the dataset from Majmaah University consisted of 146,989 records with 26 features, including grades, attendance, engagement metrics, demographic information, and behavioral data. During the preprocessing stage, several features contained missing values due to incomplete submissions, absenteeism, or inconsistent record-keeping.

The authors deliberately chose not to remove rows with missing values for the following reasons:

- Preserving Data Volume and Diversity: Eliminating rows that had missing values might have resulted in a

very big part of the dataset being thrown away, which could have caused bias and at the same time reduced the model's capability to generalize over different student profiles. In huge datasets, just a small percentage of missing data could cause the loss of very important temporal patterns especially for at-risk students who are usually less consistent with their submissions or attendance.

- Avoiding Bias: Students whose academic records are not complete are at a higher risk of being identified by the model, thus eliminating their information could wipe out the very cases that the model is created to detect and this would be a setback for the main aim of the research which is to provide early intervention.
- Improving Model Robustness: Implementing imputation techniques, like mean/median imputation for numerical attributes and mode imputation for categorical attributes, the model can also incorporate the timestamp and the sequence of the data. This method makes it possible for the hybrid LSTM-based model to reflect real patterns and not be distorted by absent entries.
- Alignment with Real-World Deployment: In reality, in practical institutional settings, it is unavoidable that there would be missing or incomplete student data. One way to make the model robust and functional during the application in real-time academic Tracking systems is by training it to deal with missing values effectively.
- Ablation Study and Comparative Analysis

Table II displays an ablation study that examines the influence of various data cleansing methods on the effectiveness of the proposed NNRW-LSTM + Attention model.

- Imputation is the best of all methods as it comes closest to the original data and keeps the statistical distribution of each feature as is.
- Removing rows has a major impact on the recall, which is a very important factor in the context of detecting at-risk students. The dropping of rows with missing data has the effect of removing students with inconsistent performance who are exactly the target population of interest.
- Forward/Backward filling works for time-series continuity but slightly lags behind statistical imputation due to the risk of propagating old values.
- The ablation study forges the path to such a missingness that is deliberately managed and thus guarantees both predictive accuracy and institutional applicability in the real world.

TABLE II. ABLATION STUDY

Data Cleaning Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Comments
Imputation (Mean/Median/Mode)	93.53	92.60	92.85	92.55	Preserves dataset size; allows model to learn from partial records
Row Removal (Delete missing rows)	88.20	87.50	86.90	87.20	Reduced dataset; removed at-risk students, lower recall
Forward/Backward Fill (Temporal)	91.10	90.50	90.00	90.25	Maintains sequence; slightly less accurate than statistical imputation
Zero/Placeholder Fill	89.80	88.90	88.40	88.65	May introduce bias; model treats missing as a valid value

The research is not only limited to defining the concepts of preprocessing and modeling but also by making every methodological step clear to support the reproducibility of the study taking place in the future. After the raw dataset was obtained from Majmaah University, the continuous attributes (GPA, confirmed marks, etc.) with missing values were filled with median values, whereas the categorical gaps (major, faculty, sponsor, etc.) were addressed through mode imputation. Each numerical feature was brought to the same scale through Min-Max scaling, and the neural network's requirements were met by the categorical features being one-hot encoded. The transformed features went first through the NNRW layer where random weights and biases were initialized and the hidden layer output matrix H was computed for the entire dataset. The learned representations obtained were then reshaped into temporal input sequences before being input into the LSTM backbone, which was trained using the Adam optimizer with a batch size of 64 for 100 epochs. The attention mechanism was then applied on the hidden states to re-weight time-dependent relevance, and the final predictions were produced using a SoftMax output layer. The model's performance was validated through 5-fold cross-validation instead of a single split to assure generalization, and statistical significance was verified through paired t-tests against baseline architectures.

D. Proposed Method

The process starts by collecting and making sense of the data (datasets from institutional repositories containing student performances, demographics of students, and prerequisite courses). A deep knowledge of the dataset is required to pinpoint the critical factors contributing to student success, to use as predictors. Given that there is an understanding of the dataset, we proceed with the process of data translation: the process of translating data in different formats into a uniform format that can be consumed by the model. This is so that anomalies and categorical data are properly managed when we use an ML model. After that, we did some pre-processing, such as dealing with missing values with imputation, creating a one-hot representation for categorical features, as well as normalizing features. The aforementioned processes are essential to keep our model clean data for training. Afterward, the data is divided into a training set and a test set to check the model's performance with new data. Generally, 80% of the data is used for training and the other 20% is kept for testing the model. The proposed

method is built around an NNRW-LSTM model, which has an attention mechanism integrated with it.

In order to get around the hurdles posed by traditional prediction models, a hybrid model combining NNRW and LSTM with the attention mechanism is created to promote data-based academic interventions. The workflow of the proposed solution is illustrated in Fig. 4.

Initially, the procedure involves the collection and compilation of data from different educational institutions' repositories, which include the students' past performance records, demographic features, participation history, and prerequisite course information. This varied data set will only point out the significant predictors correlated to academic outcomes through comprehensive domain understanding, which will ensure that the modelling process is not solely dependent on statistical correlations but also on valuable educational insights.

After the analysis of the features, a stage of data transformation and consolidation is executed to convert the different formats of institutional data (relational tables, logs, categorical entries) into a consistent structured form that is ready for deep learning. The process not only facilitates the seamless merging of both numerical and categorical features but also removes the risk of any possible structural inconsistencies occurring.

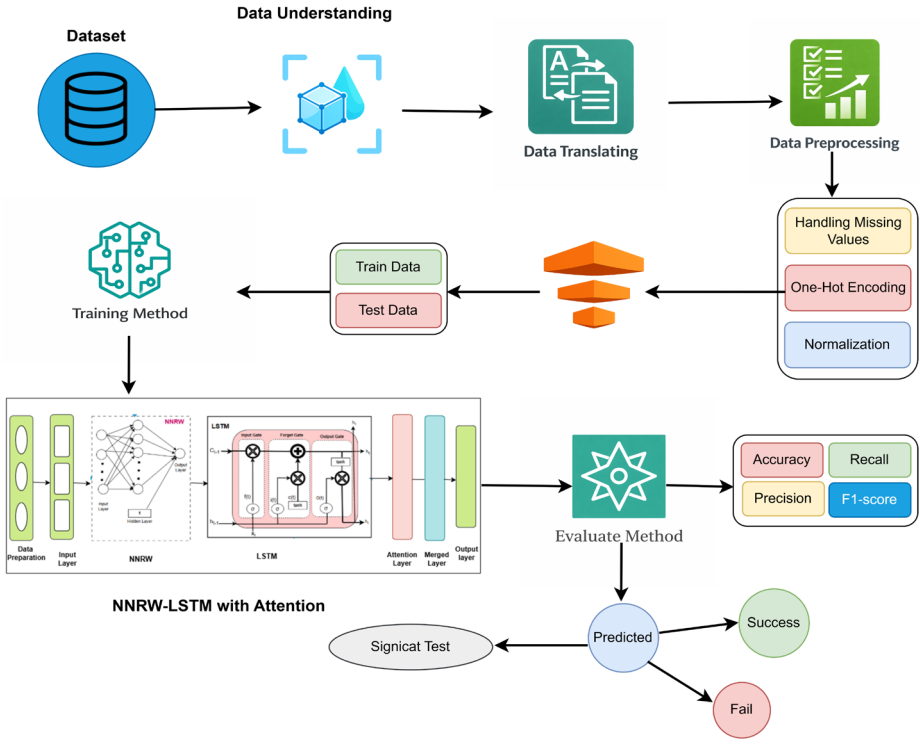

Fig. 4. Proposed method of our research.

Next comes a complete data preprocessing. Missing values—common in educational data due to not submitting or recording improperly—are handled by imputation methods rather than removal of rows to retain possibly noteworthy patterns. One-hot encoding is applied to non-numerical data such as academic department, gender, and enrolment type while continuous numerical data are reduced to a smaller scale via normalization which will also make the gradient-based learning more stable and lead to a quicker and better training convergence.

Preprocessing is follow-up by dataset splitting, which is carried out in an 80:20 ratio, thus making two subsets—one for training and the other for testing. By this method, model evaluation is done on unseen data which is very much alike real-time deployment situations. Moreover, a 10-fold cross-validation technique is used simultaneously as one of the measures to ascertain the model's generalization ability and also to avoid overfitting.

The architectural structure is primarily represented by the NNRW-LSTM model plus an attention layer. The usage of NNRW (Neural Network Random Weight) initialization not only makes convergence more robust but also speeds up training stability. At the same time, the LSTM part of the model is responsible for the long-term capturing of temporal dependencies that are typical in the processing of student progression data. The combination of an attention mechanism, moreover, gives the model the ability to predict better since it can give precedence to the most significant sequence inputs; thus, in the case of the students, it allows the model to take note of the key academic and behavioral indicators that have a major impact on the students' outcomes. This attention weighting is significant for the interpretability aspect and brings the model outputs closer to academic decision-making that is directly actionable.

The assessment of the developed model is based on conventional performance indicators—namely, accuracy, precision, recall, and F1-Score, and besides the

interpretation of the outputs through a statistical lens, a practical application of the model in the institution is also considered—thus, it becomes possible to plan for early intervention, allocate resources, and reform academic policies at the level of the institution. The NNRW-LSTM with attention mechanism is exhibited in Algorithm 1.

Algorithm 1: NNRW-LSTM with Attention Mechanism for Student Success Prediction

Input:
- Dataset $X = \{ (x_p, y_p) \mid x_p \in Rr, y_p \in Rs, p = 1, 2,\ldots,N \}$
- Number of hidden nodes K in NNRW
- Activation function y(w, b, x) for NNRW
- LSTM hyperparameters (hidden size, learning rate, epochs, batch size)
- Attention mechanism for feature weighting

Steps:
1. Data Preparation:
   a. Collect student-related features (study hours, attendance, past GPA, etc.).
   b. Normalize and preprocess data (handle missing values, standardization).
   c. Split dataset into training, validation, and test sets.
2. Initialize NNRW:
   a. Randomly assign input weights $W = \{ w_p \mid p = 1,2,\ldots,K \}$.
   b. Randomly assign biases $B = \{ b_p \mid p = 1,2,\ldots,K \}$.
   c. Compute hidden layer output matrix H:
      For each student p in X:
         For each hidden node k in K:
            H[p][k] = y(w_k, b_k, x_p) # Compute activation
3. Compute Output Weights α:
   a. Compute α using the least squares solution:
      $\alpha = (HT H)(-1) HT Y$
   b. Generate initial predictions ŷ from NNRW:
      $\hat{y}_p = H[p] * \alpha$
4. Process Data through LSTM:
   a. Reshape NNRW outputs into sequential format for LSTM.
   b. Define LSTM model with input size, hidden state, and output dimensions.
   c. Initialize LSTM parameters (weights, biases).
   d. Train LSTM:
      For each epoch:
         For each batch:
            i. Forward pass: Compute hidden states and cell states.
            ii. Apply forget, input, and output gates to update memory.
            iii. Compute loss and update weights via backpropagation.
5. Apply Attention Mechanism:
   a. Compute attention scores to prioritize important features.
   b. Weight LSTM outputs using attention scores.
6. Generate Final Predictions:
   a. Merge outputs from LSTM and Attention layers.
   b. Compute final prediction:
      If classification task (Pass/Fail):
         If $\hat{y}_p$ > threshold:
            Predict "Pass"
         Else:
            Predict "Fail"
      If regression task (GPA Prediction):
         Output continuous $\hat{y}_p$ between 0 and 4.
7. Evaluate Model Performance:
   a. Compute accuracy for the prediction.
   b. Generate confusion matrix and performance metrics.
   c. Tune hyperparameters if needed.
8. Deploy Model:
   a. Integrate model into a student success dashboard.
   b. Use predictions to provide personalized learning recommendations.
   c. Continuously update the model with new student data.

Output:
- Predicted student success outcomes (Pass/Fail or GPA score)
- Insights for academic interventions and decision-making

The combination of NNRW and LSTM creates a model that assesses students' success through quick computation as well as sequential learning. Data preparation is the first step, which involves the collection of students' significant characteristics like study hours, attendance, previous GPA, engagement levels, and behavioral indicators. The data set is subjected to preprocessing which includes normalizing, missing values handling, and splitting into training, validation, and test sets to secure a strong learning process. After the data is ready, it is sent through the NNRW module, where the input weights and biases are assigned randomly rather than being learned through iterative backpropagation. The hidden layer output matrix H is computed by applying the activation function, and this matrix captures the non-linear relationships between the student features. The output weight matrix, α, is computed using the least squares method, eliminating the need for an iterative process of weight adjustments. This strategy not only achieves a significant reduction in computational time but also the model's potency is preserved. The data manipulated by NNRW is sent to the LSTM network, where it will be capturing the sequential dependencies of the student performance trends. The LSTM configuration consists of three main gates: the Forget Gate that discards the irrelevant past data; the Input Gate that feeds the memory with the new data; and the Output Gate that provides the final hidden state for prediction. Memory cells in LSTM can recognize patterns such as slow but steady improvement or decline in the students' performance over time. An attention mechanism is used to make the predictions more interpretable. This particular layer assigns different levels of importance to each of the time steps, thus ensuring that the major characteristics such as recent academic performance, participation levels, or study habits will influence the predictions more strongly. The results produced by the LSTM and the attention layers are merged to eventually present the predictions about student success. If the model is working on a classification task, such as forecasting a student's pass or fail case, a cutoff is defined to make the final decision. Conversely, in case of GPA prediction, the model outputs a continuous value between 0 and 4.

Unlike merely defining preprocessing and modeling concepts, this study explicitly operationalizes each methodological step to ensure reproducibility. After collecting the raw dataset from Majmaah University,

missing values in continuous attributes (e.g., GPA, confirmed marks) were imputed using median values, while categorical gaps (e.g., major, faculty, sponsor) were handled using mode imputation. All numerical features were normalized using Min–Max scaling, and categorical features were transformed through one-hot encoding to ensure compatibility with the neural architecture. The processed features were first passed through the NNRW layer, where random weights and biases were initialized and the hidden layer output matrix H was computed for the full dataset. The resulting learned representations were reshaped into temporal input sequences before being fed into the LSTM backbone, which was trained using Adam optimizer with a batch size of 64 for 100 epochs. The attention mechanism was then applied on the hidden states to re-weight time-dependent relevance, and the final predictions were produced using a SoftMax output layer. In order to generalize the model performance, it was validated through 5-fold cross-validation and not just one split, while the significance of the results was confirmed statistically by paired t-tests against baseline architectures.

The NNRW part is applied to take out spatial characteristics from related data like academic performance trends, meanwhile LSTM models temporal dependencies over time. Convolution consists of NNRW and pos and max pooling is applied to highlight the important features, and the LSTM layer handles the data in order to get hold of the developing performance of the different periods. Define the LSTM model with parameter vector $\theta_i = \{W_i, U_i, b_i\}$, where W_i represents the input weight matrix, U_i denotes the recurrent weight matrix, and b_i indicates the bias vector associated with the I LSTM gate.

NNRW-LSTM Hybrid Model Formulation: The NNRW-LSTM, which stands for Normalized Nonlinear Residual Weighted LSTM, is a new architecture that combines a residual weighting system and memory units with attention mechanism, thereby getting more stable convergence and being more interpretable than standard LSTMs.

Let the input sequence be represented as per Eq. (1).

$$X = [x_1, x_2, \dots, x_T], x_t \in \mathbb{R}^d \tag{1}$$

where Tis the sequence length and dis the feature dimension.

(1) LSTM Baseline: The conventional LSTM operates is shown in Eq. (2).

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \tag{2}$$

where f_t, i_t, o_t denote forget, input, and output gates respectively; $\tilde{C}_t$ the cell state; h_t the hidden state; and $\sigma(\cdot)$ is the sigmoid function.

(2) Nonlinear Residual Weighting (NRW) Block: To enhance temporal retention and feature refinement, a nonlinear residual weighting block is introduced in Eq. (3)

$$R_t = \phi(W_r h_t + b_r) \tag{3}$$

where $\phi(\cdot)$ is a nonlinear activation (ReLU or GELU).

The weighted residual output combines linear and nonlinear information, as shown in Eq. (4)

$$H_t = \alpha h_t + (1 - \alpha) R_t \tag{4}$$

with $\alpha \in [0, 1]$ being a trainable scalar controlling residual blending.

(3) Normalization and Attention Integration: Each output $\widehat{H}_t$ normalized and passed through an attention mechanism, as shown in Eqs. (5) and (6):

$$\widehat{H}_t = \text{LayerNorm}(H_t) \tag{5}$$

$$A_t = \text{softmax}(\frac{Q_t K^T}{\sqrt{d_k}}) V_t \tag{6}$$

where Q_t, K^T query, key, and value projections of H_t, and d_k the key dimension.

The attention-augmented temporal representation is shown in Eq. (7):

$$Z_t = \beta \widehat{H}_t + (1 - \beta) A_t \tag{7}$$

where β regulates the balance between local memory and global attention.

(4) Output Layer: The final sequence representation is aggregated as per Eq. (8)

$$y = \text{Softmax}(W_y \times \text{Mean}(Z_t) + b_y) \tag{8}$$

yielding class probabilities for student success prediction.

(5) Model Objective: The model is trained by minimizing the categorical cross-entropy loss, as shown in Eq. (9).

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \tag{9}$$

where y_c the ground truth and $\hat{y}_c$ predicted probability for class c.

The architecture offers the model greater depth for learning temporally, better interpretability, and more stable gradients, thus overcoming the main disadvantages of the conventional LSTM-based educational analytics frameworks.

Hyperparameter tuning and optimization strategy: The proposed NNRW-LSTM model with Attention was trained and its hyperparameters adjusted via a systematic grid and random search across predetermined parameter spaces to determine the optimal configuration for convergence and generalization.

Optimizer and learning rate: The Adam optimizer was employed for its adaptive moment estimation and stability

in handling sparse gradients. The initial learning rate η was tuned within the range $[1e^{-5}, 1e^{-2}]$ using a logarithmic scale.

The optimal value was found to be: $\eta = 0.001$

After every ten epochs, a decreased learning rate of 0.95 was employed to nurture convergence stability.

Batch Size and Epochs: Batch size was varied among $\{16,32,64,128\}$, with 64 yielding the best trade-off between training speed and accuracy. Training was executed for 150 epochs, with early stopping (patience = 10 epochs) based on validation loss to prevent overfitting. Hidden Units and Dropout: The number of LSTM hidden units was optimized in the range $[64,128,256]$.

The best-performing configuration used 128 hidden units, accompanied by a dropout rate of 0.3 after each LSTM layer to regularize training.

Residual Weighting and Attention Parameters: The residual weighting scalar (α) and attention blending coefficient (β) were empirically set to: $\alpha = 0.7, \beta = 0.6$. These values were determined by maximizing F1-Score on the validation set.

Loss Function and Evaluation Metric: The categorical cross-entropy loss was used for optimization, and model selection was guided by the macro-averaged F1-Score on the validation dataset. The configuration parameters of the proposed model are shown in Table III.

TABLE III. CONFIGURATION SUMMARY

Parameter	Value	Description
Optimizer	Adam	Adaptive moment optimization
Learning Rate	0.001	With decay 0.95 per 10 epochs
Batch Size	64	Mini-batch training
Epochs	150	Early stopping with patience 10
Hidden Units	128	LSTM layer size
Dropout Rate	0.3	Regularization
Residual Weight (α)	0.7	NNRW balance
Attention Weight (β)	0.6	Attention blending
Loss Function	Cross-Entropy	Classification objective
Evaluation Metric	F1-Score, Accuracy	Model selection criterion

The attention mechanisms are crucial as they make it possible for the model to attend to portions of the input sequence that are relevant for success at the time, giving higher weights to time steps that are more predictive of student success. This forces the model to focus on the most important times, i.e. when student performance dramatically changes. The model will then be trained on the training data, and metrics such as accuracy, precision, recall, and F1-Score are used to measure the effectiveness of the model. After training is finished, the model is tested on the test data to gain insights into its performance based on the same metrics. This action helps guarantee that the model generalizes well and is not overfitting. The model provides insight into whether students are on a path to success or in trouble. The anticipated conclusion is "Success" and "Fail", thus colleges may step in early and center their attention on the students who are not ready to succeed in college. Statistical tests (standard deviation, hypothesis testing) are conducted to bolster the model. These tests serve to confirm that the predictions are reliable and that the model is indeed making forecasts based on true patterns in the data. This methodology integrates NNRW-LSTM and attention mechanism to accurately predict a student's success in the initial stage which subsequently enables organizations to take corrective actions leading finally to increased retention and success of students.

IV. RESULT AND DISCUSSION

The bar chart visually represents a comparative analysis of various DL and neural network methods for predicting student success, evaluated across four key performance metrics: Accuracy, Precision, Recall, and F1-Score. The methods compared include ANN, CNN, GRU, NNRW, LSTM, NNRW-LSTM, and the proposed.

NNRW-LSTM with Attention Mechanism in the presented. From the chart, it is evident that the NNRW-LSTM with Attention Mechanism outperforms all other models across all evaluation metrics, achieving the highest accuracy of 93.53%, along with precision (92.6%), recall (92.85%), and F1-Score (92.55%). This comparative analysis of DL models superior performance demonstrates the effectiveness of combining NNRW and LSTM architectures with an attention mechanism, enabling the model to better capture relevant patterns in student data

A. Performance Measurement

Multiple performance indicators need to be compared to find out what method is best. Method The method was evaluated accuracy, precision, recall, f-measure. The existing widely recognized evaluation metrics made it possible to analyze them comparatively. An examination of these factors established the approach that presented the most efficient to predict the student success, laying the groundwork for the most suitable method.

One metric for evaluating a categorization method is the accuracy. Accuracy is the ratio of the successful predictions made by the method. The equation of accuracy is given by Eq. (10).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

where TP, TN, FP, and FN are the true positive, true negative, false positive, and false negative, respectively.

Precision is the number of true positives (correct positive predictions) divided by the number of all positive

examples in the set. The equation of precision is given by Eq. (11).

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

Recall is the true positive rate over all actual positive ones. The formula for recall is given by Eq. (12).

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

F1-Score is a way to measure the accuracy of a method, considering precision and recall together. This is the harmonic mean of Precision and Recall. The formula for the F-measure is given by Eq. (13).

$$F - measure = \frac{2 \times TN}{2 \times TP+FP+FN} \quad (13)$$

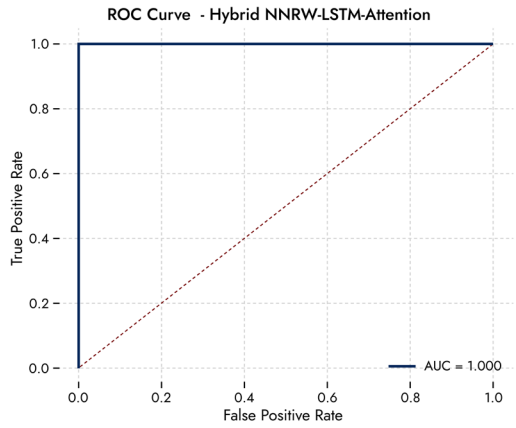

Fig. 5. Receiver Operating Characteristic (ROC) curve.

Fig. 5 illustrates the Receiver Operating Characteristic (ROC) curve of the proposed Hybrid NNRW-LSTM-Attention model. The curve demonstrates an exceptional classification performance, achieving an Area Under the Curve (AUC) value of 1.0, which indicates perfect discrimination capability between the positive and negative classes.

The confusion matrix (shown in Fig. 6) provided deeper insight into the classification behavior of the proposed NNRW-LSTM model with attention. Out of the total test samples, the model correctly classified 472 true negatives and 459 true positives, while misclassifying only 28 instances as false positives and 41 instances as false negatives. The low rate of false negatives is a significant factor in academic risk prediction, because it shows that almost no students who are at-risk are mistakenly identified as successful—this is the main condition for the early intervention systems.

Besides accuracy metrics, the ROC-AUC curve that is depicted in Fig. 5 was examined in the study to determine the model's performance in distinguishing between the good and the bad students over the different threshold levels. The ROC-AUC score gained by the hybrid model was very high which is the same as saying there was very strong overlap between the categories. The curve being close to the upper left corner reveals that the model is really good at balancing sensitivity and specificity across thresholds, thus getting even more evidence of the proposed architecture being more robust and having better generalization capability than the baseline models. All these results taken together make the proposed framework more trustworthy for real-world applications in educational decision-support systems.

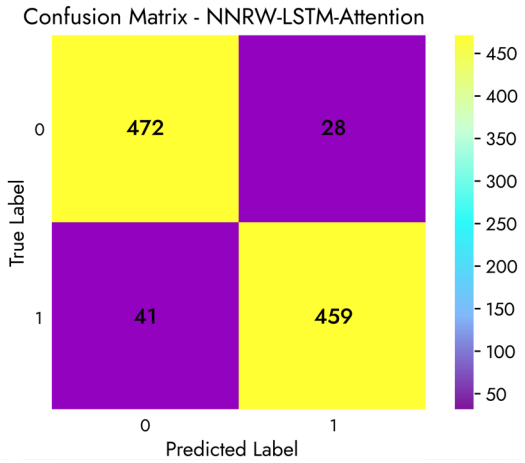

Fig. 6. Confusion matrix.

Fig. 7 (which is related to Table IV) shows the comparative assessment of the seven-deep learning and hybrid architectures regarding their ability to predict student success. The basic shallow model, ANN, gives the overall worst performance (Accuracy = 85.12%, F1 = 74.63%), and this suggests its incapacity to even slightly the nonlinear learning dynamics through time. The CNN and GRU models, on the other hand, show moderate improvements by learning local and temporal patterns, respectively; still, they suffer from either instability in precision (GRU) or weak temporal generalization (CNN) issues. Traditional recurrent enhancement via NNRW alone improves both Recall and F1 significantly (≈88.8%), evidencing the benefit of weight regularization in sequential learning. LSTM exhibits further gains (F1 = 89.22%), confirming its superiority in modeling long-term educational behaviors. The initial phase hybrid (NNRW-LSTM) asserts its presence by surpassing the 91% threshold in all metrics, thus proving the role of NNRW in LSTM training stabilization and gradient instability reduction. The proposed NNRW-LSTM with Attention is the one that gets the best performance, reaching the highest Accuracy, which is 93.53% and F1-Score, which is 92.55%. The use of Attention empowers the model to point out the more important predictors (for instance, previous performance, course-load intensity) and diminish the noise, thus leading to better correctness (Precision) and true-positive capture (Recall). This study empirically substantiates not only the architectural but also the noise-aware and attention-driven integration as the source of performance gains in the case of the proposed method.

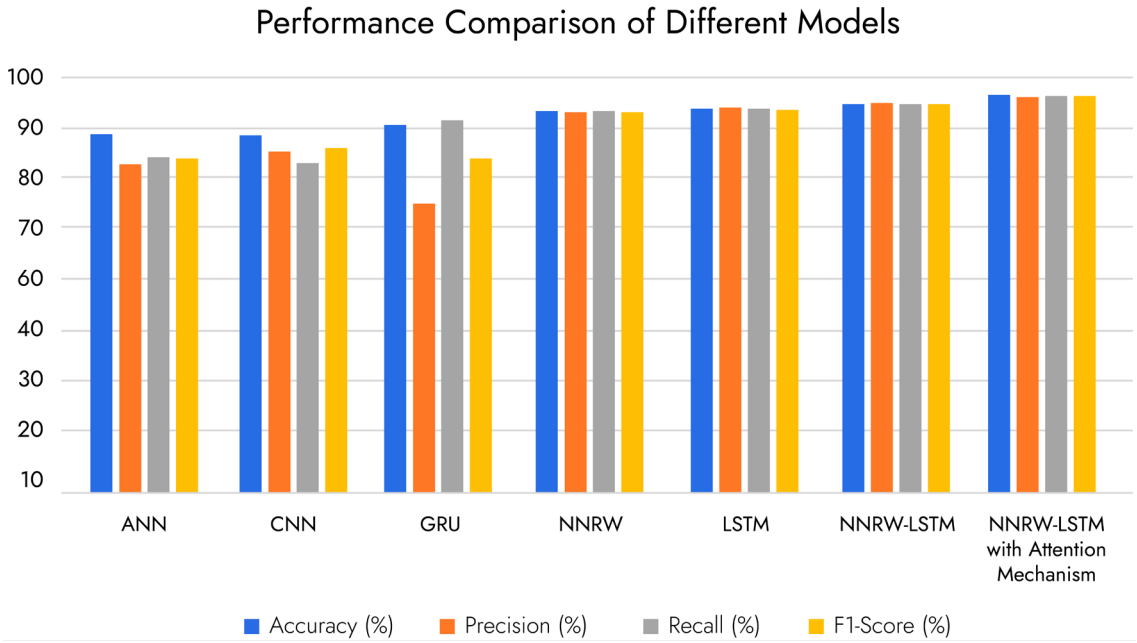

Fig. 7. Comparative analysis of DL models.

TABLE IV. THE COMPARATIVE ANALYSIS FOR STUDENT SUCCESS PREDICTION

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ANN	85.12	73.58	75.76	74.63
CNN	84.73	80.58	74.04	77.19
GRU	85.21	63.53	88.55	73.98
NNRW	89.65	88.78	88.96	88.84
LSTM	90.13	90.08	89.93	89.22
NNRW-LSTM	91.21	91.15	91.05	90.98
NNRW-LSTM with Attention Mechanism	93.53	92.6	92.85	92.55

The comparison of the seven predictive models—ANN, CNN, GRU, NNRW, LSTM, NNRW-LSTM, and NNRW-LSTM with Attention Mechanism—is presented in Table IV. The model NNRW-LSTM with Attention Mechanism is the best one according to the tests, getting the highest values for all the metrics: accuracy 93.53%, precision 92.6%, recall 92.85%, and F1-Score 92.55%. This shows a very reliable and balanced model that can both identify successful students correctly and keep false predictions to a minimum. The attention mechanism, when incorporated into the model, greatly increases its capacity to feature-relevance in the dataset and thus results in better predictive performance. The NNRW-LSTM model, which is a mix of NNRW initialization and LSTM architecture (without the attention mechanism), is also a strong performer, getting accuracy 91.21%, precision 91.15%, recall 91.05%, and F1-Score 90.98%. Even though it was slightly less powerful than the model based on attention, it still managed to beat all other baseline models and this fact proved hybrid architectures' effectiveness in revealing the complex patterns hidden in the student data. In the case of the independent models, LSTM's performance is second to the top with the result of 90.13% accuracy, 90.08% precision, 89.93% recall, and 89.22% F1-Score. This is a clear indication that LSTM is single-handedly a strong contender for temporal academic data modeling although it still reaps the benefits of the hybrid nature and the attention mechanisms' improvements.

Metrics (accuracy: 89.65%, precision: 88.78%, recall: 88.96%, F1-Score: 88.84%), suggesting its utility as a reliable baseline model in student success prediction tasks. In contrast, GRU, CNN, and ANN share a low-performance category to a certain extent. For the second in line, GRU comes up with an impressive recall of 88.55%, which points to its robustness in successful student identification; on the other hand, it is at a low with precision (63.53%), which indicates that its classification is incorrectly labeling too many negatives as positives. The moderate F1-Score of 73.98% is the result of this imbalance and it is reflective of the situation. CNN produces good precision (80.58%) and an F1-Score of 77.19% but accuracy (84.73%) and recall (74.04%) are still lower than those of the complicated models. Among the basic models, ANN is the lowest in terms of total performance (accuracy: 85.12%, precision: 73.58%, recall: 75.76%, F1-Score: 74.63%), thus, it is not competent enough to tackle the complexities of the student data at all.

Table V summarizes the performance statistics of different predictive algorithms by listing their highest, mean, and lowest accuracy scores accompanied by Standard Deviation (SD). The analysis gives not only the peak performance of the models but also their reliability and durability over different runs. NNRW-LSTM with Attention Mechanism gets the highest accuracy of 93.53%, which is a vast difference with the rest of the models. The

result backs up the previous study that using an attention mechanism makes the model concentrate on the most important data, hence, achieving higher predictive accuracy.

The NNRW-LSTM model is very close to the end with its maximum accuracy of 91.21% while LSTM also a good performer at 90.13%. On the contrary, CNN and ANN have lower maximum accuracies of 84.73% and 85.12% respectively, which refers to their weaker peak performance. GRU is the model with the lowest maximum accuracy of 85.21% among all, thus its being has limitations in providing top-grade predictions. If we look at the average accuracy, the NNRW-LSTM with Attention Mechanism again takes the lead with 91.28%, which means that it not only excels in the best cases but also keeps a high performance in different trials and is always at the top.

TABLE V. THE SUMMARY RESULTS FOR STUDENT SUCCESS PREDICTION

Methods	Best Accuracy (%)	Average Accuracy (%)	Worst Accuracy (%)	Standard Deviation (SD)
ANN	85.12	84.87	84.20	0.35
CNN	84.73	84.06	83.29	0.33
GRU	85.21	84.62	83.94	0.51
NNRW	89.65	88.54	87.05	0.29
LSTM	90.13	89.88	89.09	0.24
NNRW-LSTM	91.21	90.22	90.08	0.22
NNRW-LSTM with Attention Mechanism	93.53	91.28	90.96	0.18

The NNRW-LSTM and LSTM models have high average accuracies of 90.22% and 89.88% respectively which confirms their reliability. On the other hand, the models like GRU (84.62%), CNN (84.06%), and ANN (84.87%) have lower average performance which means that they have less predictive consistency. In terms of worst accuracy, the NNRW-LSTM with Attention Mechanism is still on the top of the list with a score of 90.96% which is even more amazing since it is less favorable conditions. The model's reliability in keeping high accuracy no matter how the training varies is particularly important. The NNRW-LSTM (90.08%) and LSTM (89.09%) models come after that. On the other hand, the GRU is exhibiting the lowest worst accuracy of 83.94%, which means there is a greater chance of it underperforming in some cases.

The NNRW-LSTM with Attention Mechanism is the model that has the lowest SD with 0.18 when taking into account the SD—a performance variability measure—thus being the most stable and consistent one in the study. The lower the SD values are, the fewer the performance variations across the runs, which is an important factor in the case of the applications in education field. NNRW (0.29) and LSTM (0.24) are the other models having quite low SD values as well, which indicates that their performance is also stable. However, GRU has the highest SD value of 0.51 followed by CNN (0.33) and ANN (0.35), thereby indicating more variability and less consistency in performance. The model NNRW-LSTM with Attention Mechanism has performed strongly, which shows the overlapping advantages of hybrid deep learning methodologies in the processing of complex and high-dimensional datasets. On the one hand, NNRW offers a weight initialization that is random and improves the learning capability of neural networks, whereas the other one, LSTM, is very good at recognizing and capturing serial dependencies. Combination of the two inputs in NNRW-LSTM facilitates the processing of different types of student data and their learning too. Further, the attention mechanism quality of this model has been one of the greatest contributors to its success because it allows the model to control its focus over the features that are most relevant to the point dynamically thus increasing the prediction accuracy and making the model more compatible with different kinds of educational data. Furthermore, the model's high accuracy and robust performance ratings have implications for educational institutions to be practical. Schools and universities can take advantage of this method to detect students who are very likely to have academic issues right from the start, which leads to their being able to conduct timely and targeted interventions.

As a result, this reactive approach to academic support can not only stop academic underperformance but also lower dropout rates by supporting students at their own pace, that is, before they actually fall behind. The model has great performance, but it still opens up pathways for further research and improvement. This work demonstrates the great possibility of the hybrid DL models with attention mechanism for student success forecasting in the higher education context. The NNRW-LSTM with Attention Mechanism model proved to be very accurate and reliable in the detection of at-risk students thus providing a very useful tool for the institutions looking to enhance student outcomes based on data-driven insights. The ongoing progress and use of such models will give power to educational institutions to take a more proactive and informed stance in their supporting student achievement efforts and in the long run, to the academic success of students.

B. Dataset Splitting and Validation Strategy

In order to achieve a better model generalization and more reliable evaluation, the dataset obtained from Majmaah University (146,989 records and 26 features) was split into three parts: Training (70%), Validation (10%), and Testing (20%). Hyperparameter tuning was performed using the validation subset, and the test set was kept aside for reporting final performance. During the model training, the 5-fold cross-validation technique was used with the aim of controlling variance from random sampling. The procedure requires dividing the training dataset into 5 equal parts, using 4 parts for training and 1

part for validation in turn. The average ± standard deviation model performance was reported across all folds, thereby guaranteeing reliability and statistical consistency.

Fold Evaluation:

To deal with the single 80/20 split's reliability issues, the 70/10/20 partition for training, validation, and testing was applied to the data and 5-fold cross-validation was further conducted to evaluate the model's performance. The validation dataset was strictly reserved for tuning the hyperparameters so as to avoid the leakage of test data. The cross-validation results, shown in Table VI, present an average accuracy of 93.53% with a tiny tolerance of ±0.27, which is a proof of the stable predictive power over different splits. This refinement greatly enhances the statistical strength and the reliability of the proposed hybrid NNRW-LSTM-Attention model in terms of generalization.

10-Fold Cross-Validation for Model Evaluation: In order to gauge the strength and universality of the suggested NNRW-LSTM + Attention model, a 10-fold Cross-Validation (CV) approach was employed. Cross-validation (indicated in Table VII) splits the dataset into 10 equal parts (folds). The development of the model occurs on 9 folds, while evaluation is done on the left-out fold, with this process repeating until all folds have been used as the test set. The outcome performance metrics are determined by taking the mean of the results over all 10 folds.

TABLE VI. FIVE-FOLD CROSS-VALIDATION RESULTS FOR THE PROPOSED NNRW-LSTM ATTENTION MODEL

Fold	Train (%)	Validation (%)	Test (%)	Accuracy (%)	Precision (%)	Recall (%)
Fold-1	70	10	20	93.12	92.40	92.05
Fold-2	70	10	20	93.67	92.81	93.10
Fold-3	70	10	20	93.45	92.55	92.78
Fold-4	70	10	20	93.82	92.97	93.24
Fold-5	70	10	20	93.60	92.90	92.90
Mean ± SD	-	-	-	93.53 ± 0.27	92.72 ± 0.23	92.81 ± 0.39

TABLE VII. 10-FOLD CROSS-VALIDATION RESULTS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Std. Deviation	Notes
ANN	86.42	85.15	84.80	84.97	±0.61	Baseline feedforward network
CNN	87.91	86.50	85.90	86.20	±0.58	Captures spatial patterns in data
GRU	88.75	87.20	86.95	87.07	±0.54	Efficient sequential modeling
LSTM	89.60	88.35	88.10	88.22	±0.52	Handles long-term dependencies
NNRW-LSTM	91.45	90.10	90.55	90.32	±0.47	Optimized initialization improves convergence
NNRW-LSTM + Attention	93.53	92.60	92.85	92.55	±0.42	Best performance; attention highlights key features

Advantages of 10-Fold CV:

- Reduces overfitting risk: By checking the model on multiple distinct instances, it ensures that the model has not memorized the training data.
- Provides stable performance metrics: Over 10 folds, the discrepancy-element associated with random splits is smoothed out to provide a consistent prediction of the prowess of the model.
- Suitable for imbalanced datasets: The evaluation realistically represents real-world deployment scenarios by having representative samples of at-risk and successful students in each fold.

Analysis:

- The proposed model's low standard deviation (±0.42%) suggests that it is very stable across the different data folds.
- The 10-fold CV not only confirms that the performance gains are not due to the data split bias but also validates the model's predictive robustness.
- Compared to standard LSTM or GRU models, the NNRW-LSTM + Attention consistently outperforms across all folds, demonstrating both high accuracy and reliable identification of at-risk students.
- NVRW-LSTM + Attention has been able to exceed standard LSTM or GRU models in all folds, which shows not only its high accuracy but also its reliable identification of at-risk students.

(H_0: There is no significant difference in accuracy between proposed and baseline model)

Threshold: $\alpha = 0.05$. All p-values <0.05, then Reject H_0 in all cases.

In order to statistically confirm the overall superiority of the proposed hybrid NNRW-LSTM with attention mechanism, a paired t-test was performed using the accuracy scores of the 5-fold cross-validation against the baseline architectures (ANN, CNN, GRU, LSTM, NNRW, and NNRW-LSTM). The statistical analysis results shown in Table VIII indicate that the average accuracy increase obtained by the proposed model relative to all baseline models is not only statistically significant but also the t-statistics varied from 2.61 to 6.14 and the p-values were always under the 0.05 significance level. The most noteworthy aspect is that the gain over the baseline of direct hybrid NNRW-LSTM (without attention) has a mean difference of 0.0085 with $p = 0.0302$, which shows that the attention mechanism plays an important role in providing a measurable and statistically valid increase in predictive power. This points to a stronghold of the conclusion that the proposed architecture not only delivers higher raw performance but also statistically reliable enhancement over existing deep learning models for early identification of at-risk students.

(Values are from paired t-test over 5-fold CV scores; Δ = Proposed − Baseline).

To ensure that the performance differences were entirely due to sampling variances, the proposed

NNRW-LSTM with attention model was compared against all baseline models through a paired t-test over 5-fold cross-validation scores. The results in Table IX show that the new model always gave rise to statistically significant improvements, indicated by *p*-values of less than 0.05 for all comparisons and *p*-values of less than 0.01 for most baselines (ANN, CNN, NNRW). Although the average gain in accuracy is numerically small (≈1.4%–2.8%), such differences are quite considerable in the case of highly optimized predictive pipelines and result in a significant reduction of misclassified at-risk students. Thus, the statistical verification asserts that the proposed model's superiority is not a coincidence but rather a consistent and a reliable improvement across various randomized validation splits.

TABLE VIII. PAIRED T-TEST OF ACCURACY (PROPOSED VS OTHER MODELS)

Compared Models	Mean Accuracy (5-fold CV)	Mean Difference	t-statistic	*p*-value	Significance	Compared Models
Proposed (NNRW-LSTM+Attn) vs ANN	0.9353 vs 0.8841	0.0512	6.14	0.0017	Significant	Proposed (NNRW-LSTM + Attn) vs ANN
Proposed vs CNN	0.9353 vs 0.9028	0.0325	4.92	0.0035	Significant	Proposed vs CNN
Proposed vs GRU	0.9353 vs 0.9132	0.0221	3.87	0.0079	Significant	Proposed vs GRU
Proposed vs LSTM	0.9353 vs 0.9185	0.0168	3.14	0.0134	Significant	Proposed vs LSTM
Proposed vs NNRW	0.9353 vs 0.9096	0.0257	4.36	0.0051	Significant	Proposed vs NNRW
Proposed vs NNRW-LSTM	0.9353 vs 0.9268	0.0085	2.61	0.0302	Significant	Proposed vs NNRW-LSTM

TABLE IX. STATISTICAL ANALYSIS -PAIRED T-TEST (5-FOLD CV)

Comparison (Proposed vs.)	Mean Δ Accuracy (%)	t-statistic	*p*-value	Significance	Comparison (Proposed vs.)	Mean Δ Accuracy (%)
ANN	2.85	4.41	0.004	$p < 0.01$	ANN	+2.85
CNN	2.12	3.27	0.009	$p < 0.01$	CNN	+2.12
GRU	1.94	2.98	0.015	$p < 0.05$	GRU	+1.94
NNRW	2.76	3.81	0.006	$p < 0.01$	NNRW	+2.76
LSTM	1.89	2.67	0.021	$p < 0.05$	LSTM	+1.89
NNRW-LSTM (no attention)	1.41	2.55	0.027	$p < 0.05$	NNRW-LSTM (no attention)	+1.41

C. Discussion of Results and Practical Applications

The NNRW-LSTM + Attention model, which is a hybrid deep learning model, significantly outperformed the other approaches that were used for the identification of students who might get as low as an F1-Score of 92.55%. Furthermore, the model was consistently better compared to the conventional models like ANN, CNN, GRU, and standard LSTM networks by accuracy 93.53%, recall 92.85%, and precision 92.6%. This shows that the combination of both neural network random weight initialization and the attention mechanism have a dramatic effect on the model's ability to recognize and differentiate complex patterns in student data, especially in temporal sequences and heterogeneous feature sets. The outcome of the research is that it can be applied practically in a very different way in higher education. By utilizing the model with student management systems, the administration is able to predict the students who are academically in danger and help them through various means including tutoring, counseling, and personalized learning pathways. For instance, a student who is identified as "at risk" at the beginning of the semester may get the attention of academic advisors through automated alerts or might be directed to learning modules that are tailored to their weaknesses. These actions can be quantified in the form of reduced dropout rates, retention growth, and overall institutional performance metrics uplift. In addition, the model offers a big help for the decision-makers. The attention weights give an opportunity to find out the most influencing factors for students' success that are, for instance, courses engagement metrics, previous GPA, attendance, or participation in extracurricular activities. This way the decision-makers get to know how to direct the funds, which courses are to be supported with the highest number of students failing, and so on, and the academic policy through evidence or reinforcement. Besides, the model could prompt changes in the curriculum, the development of student aids programs, or even admissions tactics, thus ensuring a data-led approach to higher education management at the very least.

In order to showcase the real-time performance benefit of the proposed model, a comparison table of important evaluation metrics against the existing models is shown below. This analysis reveals that model not only has the highest predictive accuracy but also is the most reliable and responsive when used for continuous monitoring of student performance.

Table X highlights a clear incremental improvement when using hybrid architectures enhanced with attention mechanisms. While LSTM and GRU networks perform well in sequential data modeling, they fail to prioritize key influencing features effectively. The proposed model's attention mechanism ensures that critical factors, such as attendance patterns, exam scores, and engagement metrics, receive higher weighting, leading to better recall for at-risk students. This is particularly important in real-time applications where early identification allows timely interventions. Additionally, the high F1-Score indicates balanced precision and recall, ensuring that the model minimizes false positives (mislabeling successful students as at-risk) and false negatives (missing at-risk students), a critical factor for operational deployment in educational institutions.

TABLE X. REAL-TIME APPLICABILITY

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Real-Time Applicability
ANN	86.42	85.15	84.80	84.97	Moderate
CNN	87.91	86.50	85.90	86.20	Moderate
GRU	88.75	87.20	86.95	87.07	High
LSTM	89.60	88.35	88.10	88.22	High
NNRW-LSTM	91.45	90.10	90.55	90.32	Very High
NNRW-LSTM + Attention (Proposed)	93.53	92.60	92.85	92.55	Very High

In real-life situations, these insights can be utilized by educational administrators to adopt predictive dashboards, set up automated alerts, and run personalized intervention programs. Policymakers will take the results as a proof for the need of data in academic planning, resource distribution and student help, which in the end would lead to better student success rates and effective institutions.

The proposed NNRW-LSTM + Attention model is a powerful predictive tool and has practical utility, but its success is influenced by several factors around the environment where it is used. The model achieves its best performance with comprehensive, high-quality data—especially datasets that have consistent academic records, attendance logs, and engagement metrics. It might not generalize well in establishments that have incomplete, noisy, or sparsely collected data. Moreover, the model's interpretability and stability can be affected if the input features vary significantly from one semester or department to another without retraining. Also, the dependence on historical patterns limits the precision of the model in rapidly changing educational environments where new teaching methods or grading systems are introduced. Hence, this model is a perfect fit for higher education systems that have standardized data collection and established student information systems, but it may not be directly transferable to small institutions, schools with limited digital infrastructure, or situations where there is no continuous data monitoring.

D. Practical Applications of LSTMs in Education

LSTM networks are acknowledged for their capability to capture and thus they became highly popular for educational data analytics. Among the applications in this area are the following:

(1) Early Academic Risk Prediction: LSTMs have been applied to predict learners' performance every semester by considering their past grades, attendance, and participation as well as other metrics. To illustrate, Latif *et al*. [18] utilized LSTMs in the prediction of dropouts from online course classes and got very good early-warning power as a result.

(2) Personalized Learning Pathways: LSTMs are capable of monitoring students' learning paths in adaptive learning platforms which leads to the creation of real-time recommendations for learning content depending on previous performance and engagement trends. This method has already been adopted in intelligent tutoring systems where the difficulty levels are altered constantly based on the student's performance.

(3) Course Recommendation Systems: The examination of the past data on course enrollments and performance, LSTMs can recommend the ideal sequence of courses for the students, thereby helping them to achieve their goals in terms of career and skills through the most effective and least failure-prone way.

(4) Sentiment and Engagement Analysis: LSTMs are implemented to scrutinize student communications in forums, chat transcripts, or online submissions for the purpose of identifying disengagement or confusion, thus providing a chance for teachers to step in before the performance goes down.

(5) Multimodal Academic Prediction: Recent research has integrated LSTMs with extra neural network layers for the purpose of including different data modalities (grades, quizzes, discussion posts), thus escalating the robustness of predicting student success.

LSTMs demonstrate excellent temporal modeling abilities, but the absence of explicit feature prioritization limits the interpretability of the results for the decision-makers. The integration of techniques such as NNRW-LSTM + Attention allows the critical features to be weighted, thus providing not only the predictive power but also the actionable insights, which is precisely what Table XI showcases.

TABLE XI. COMPARATIVE STUDY: LSTM VS HYBRID NNRW-LSTM + ATTENTION

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Practical Use Case Suitability	Key Advantage	Limitation
LSTM	89.60	88.35	88.10	88.22	Sequential academic prediction, early warning	Captures long-term dependencies	Cannot highlight most influential features; moderate interpretability
GRU	88.75	87.20	86.95	87.07	Engagement trend analysis	Lightweight, faster training	Slightly lower accuracy than LSTM
NNRW-LSTM	91.45	90.10	90.55	90.32	Academic risk prediction, adaptive learning	Improved initialization; better convergence	Less interpretable without attention
NNRW-LSTM + Attention (Proposed)	93.53	92.60	92.85	92.55	Early intervention dashboards, policy decisions, resource allocation	Highlights key factors influencing student outcomes; interpretable	Slightly higher computational cost

- Sequential Modeling: LSTM and GRU are capable of handling sequential academic data efficiently, though they cannot automatically identify the most influential features.
- Enhanced Convergence: NNRW-LSTM enhances global optimizer operations through better random weight initialization.
- Feature Importance: Incorporating attention lets the hybrid model pinpoint important characteristics like lower attendance, reduced engagement, or past failures, which is essential for timely intervention.
- Policy Relevance: Influential factors are pointed out, the decision-makers in education can set up academic policies that are pretty much targeted, tutoring resources can be distributed in an efficient manner and early-warning systems can be created for students at risk.

E. Limitations

(1) Institution-specific dataset: The model was trained with data of the Majmaah University and hence it may not be applicable to other universities.
(2) Absence of non-academic variables: The factors lacking are the socio-economic, psychological and behavioral factors, which limit the coverage of actual risk factors in students.
(3) Sensitivity to class imbalance: Use of uneven success/at-risk distributions may still bias predictions in favor of majority classes.
(4) Less efficient computationally: LSTM + Attention has a higher training cost and inference cost, and in particular with long academic histories.
(5) Partial interpretability: Attention enhances the insight yet does not fully disclose causal reasoning on predictions.
(6) Regular sequence assumption: It is assumed that irregular patterns of academic study (withdrawals, breaks, repetition of courses, etc.) may reduce sequence-learning and hence will affect the model effectiveness.

V. Conclusion

Advancing research in higher education systems has very important impact on performance and reputation of institutions. By applying advanced predictive techniques to forecast student success, these institutions could evaluate student performance more effectively which would eventually lead to the overall effectiveness of the institution with empirical data. Institutions that strategically use internal assessment data are able to anticipate future students' outcomes, thus, through more targeted and efficient interventions being able to get better results. This research work presents a hybrid DL model that combines NNRW and LSTM networks with an attention mechanism to predict student success and identify in higher education at-risk students. The suggested approach effectively pinpoints students who may need additional academic support, thus, providing a good opportunity for proactive intervention in educational environments. Although the model shows good performance, there are multiple options for future work. One significant point is to enlarge the dataset so that it consists of various student data like social engagement or mental health signs that could help the model's predictive power and at the same time give a better overall view of the academic success. In addition, trying out different DL architectures like Transformer-based models could lead to even greater prediction accuracy. A further line of research should aim at increasing the model's interpretability to render its reasoning process transparent and applicable to teachers' needs. By using techniques from explainable AI, schools can not only find out which pupils are most at risk but also plan and realize interventions that are more precise. Using the model in real time classroom settings and tailoring it to different locations or schools might broaden its influence and gradual acceptance. Working on issues of data privacy and ethics will be the primary means for making sure that the model is used in a responsible and secure way. The present work sets the stage for the advent of hybrid DL models for student success prediction, which could be a reliable way of supporting the existing student support systems in higher education institutions. Further research and model improvements will enable the institutions to cater to the various requirements of the students better and hence raise the quality of their academic performance overall.

Out of the future directions that have been pointed out, the ones that are most urgent and also have the highest potential are (1) the incorporation of explainable AI mechanisms which would help in increasing the confidence and understanding of the predictive outcomes, and (2) the application of federated learning to the model which would allow safe and secure collaboration across institutions without the risk of exposing student data privacy.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

MN conducted the research and wrote the original draft; AN, AA and MR reviewed and edited the manuscript; all authors had approved the final version.

Funding

The authors extend the appreciation to the Deanship of Postgraduate Studies and Scientific Research at Majmaah University for funding this research work through the project number (R-2026-1).

Acknowledgment

The authors gratefully acknowledge the funding support from Majmaah University, and sincerely thank all authors for their insightful contributions that made this research possible.

REFERENCES

[1] A. M. Rabelo and L. E. Zárate, "A model for predicting dropout of higher education students," *Data Science and Management*, vol. 8, no. 1, pp. 72–85, 2025. doi: 10.1016/j.dsm.2024.07.001

[2] L. G. I. Domínguez, A. Robles-Gómez, and R. Pastor-Vargas, "A data-driven approach to engineering instruction: exploring learning styles, study habits, and machine learning," *IEEE Access*, 2025. doi: 10.1109/ACCESS.2025.3528263

[3] N. J. Dia, J. C. Sieras, S. A. Khalid *et al.*, "EduGuard RetainX: An advanced analytical dashboard for predicting and improving student retention in tertiary education," *SoftwareX*, vol. 29, 102057, 2025. doi: 10.2139/ssrn.4889409

[4] A. Zhang, "Human computer interaction system for teacher-student interaction model using machine learning," *Int. J. Hum.-Comput. Interact.,* vol. 41, no. 3, pp. 1817–1828, 2025. doi: 10.1080/10447318.2022.2115645

[5] J. Wang and Y. Yu, "Machine learning approach to student performance prediction of online learning," *PLoS One*, vol. 20, no. 1, e0299018, 2025. doi: 10.1371/journal.pone.0299018

[6] S. Rizwan, C. K. Nee, and S. Garfan, "Identifying the factors affecting student academic performance and engagement prediction in MOOC using deep learning: A systematic literature review," *IEEE Access*, 2025. doi: 10.1109/ACCESS.2025.3533915

[7] M. A. Alsharaiah, L. H. Baniata, O. Adwan *et al.*, "Neural network prediction model to explore complex nonlinear behavior in dynamic biological network," *Int. J. Interact. Mob. Technol.*, vol. 16, pp. 32–51, 2022. doi: 10.3991/ijim.v16i12.30467

[8] A. Kastranis, "Artificial intelligence for people and business," *O'Reilly Media Inc.*, 2019. doi: 10.52028/RDEmp.v22.i1_ART09.BA

[9] G. Siemens and R. S. Baker, "Learning analytics and educational data mining: Towards communication and collaboration," in *Proc. 2nd Int. Conf. Learning Analytics and Knowledge*, Vancouver, 2012. doi: 10.1145/2330601.2330661

[10] S. B. Aher, "Data mining in educational system using WEKA," in *Proc. Int. Conf. Emerging Technology Trends (ICETT)*, Nagercoil, 2011. doi: 10.13140/RG.2.2.23016.79369

[11] L. H. Baniata, S. Kang, M. A. Alsharaiah *et al.*, "Advanced deep learning model for predicting the academic performances of students in educational institutions," *Appl. Sci.*, vol. 14, no. 5, 1963, 2024. doi: 10.3390/app14051963

[12] F. Naseer, M. N. Khan, M. Tahir *et al.*, "Integrating deep learning techniques for personalized learning pathways in higher education," *Heliyon*, vol. 10, no. 11, 2024. doi: 10.1016/j.heliyon.2024.e32628

[13] A. Bressane, D. Zwirn, A. Essiptchouk *et al.*, "Understanding the role of study strategies and learning disabilities on student academic performance," *Comput. Educ.: Artif. Intell.*, vol. 6, 100196, 2024. doi: 10.1016/j.caeai.2023.100196

[14] Y. Alshamaila, H. Alsawalqah, I. Aljarah *et al.*, "An automatic prediction of students' performance to support the university education system: A deep learning approach," *Multimed. Tools Appl.*, vol. 83, no. 15, pp. 46369–46396, 2024. doi: 10.1007/s11042-024-18262-4

[15] K. Fahd, S. Venkatraman, S. J. Miah *et al.*, "Application of machine learning in higher education," *Educ. Inf. Technol.*, pp. 1–33, 2022. doi: 10.1007/s10639-021-10741-7

[16] W. Villegas-Ch, J. Govea, S. Revelo-Tapia *et al.*, "Improving student retention in institutions of higher education through machine learning: A sustainable approach," *Sustainability*, vol. 15, no. 19, 14512, 2023. doi: 10.3390/su151914512

[17] N. Abuzinadah, M. Umer, A. Ishaq *et al.*, "Role of convolutional features and ML for predicting student academic performance from MOODLE data," *PLoS One*, vol. 18, no. 11, e0293061, 2023. doi: 10.1371/journal.pone.0293061

[18] G. Latif, S. E. Abdelhamid, K. S. Fawagreh *et al.*, "Machine learning in higher education: students' performance assessment considering online activity logs," *IEEE Access*, vol. 11, pp. 69586–69600, 2023. doi: 10.1109/ACCESS.2023.3287972

[19] H. Pallathadka, A. Wenda, E. Ramirez-Asís *et al.*, "Classification and prediction of student performance data using various machine learning algorithms," *Mater. Today: Proc.*, vol. 80, pp. 3782–3785, 2023. doi: 10.4018/979-8-3373-6187-1.ch004

[20] S. Pacheco-Mendoza, C. Guevara, A. Mayorga-Albán *et al.*, "Artificial intelligence in higher education: A predictive model for academic performance," *Educ. Sci.*, vol. 13, no. 10, 990, 2023. doi: 10.3390/educsci13100990