






Article

# Parallel Attention-Driven Model for Student Performance Evaluation

Deborah Olaniyan<sup>1</sup>, Julius Olaniyan<sup>1,\*</sup>, Ibidun Christiana Obagbuwa<sup>2</sup>, Bukohwo Michael Esiefarienrhe<sup>3</sup>  
and Olorunfemi Paul Bernard<sup>4</sup>

<sup>1</sup> Department of Computer Science, Bowen University, Iwo 232101, Osun, Nigeria; deborah.olaniyan@bowen.edu.ng

<sup>2</sup> Department of Computer Science, Sol-Plaatje University, Kimberley 8301, South Africa; ibidun.obagbuwa@spu.ac.za

<sup>3</sup> Department of Computer Science, North-West University, Mafikeng X2046, South Africa; 25840525@nwu.ac.za

<sup>4</sup> Department of Computer Science, Auchi Polytechnic, Auchi 312101, Edo, Nigeria; bernard.femi@auchipoly.edu.ng

\* Correspondence: julius.olaniyan@bowen.edu.ng

**Abstract:** This study presents the development and evaluation of a Multi-Task Long Short-Term Memory (LSTM) model with an attention mechanism for predicting students' academic performance. The research is motivated by the need for efficient tools to enhance student assessment and support tailored educational interventions. The model tackles two tasks: predicting overall performance (total score) as a regression task and classifying performance levels (remarks) as a classification task. By handling both tasks simultaneously, it improves computational efficiency and resource utilization. The dataset includes metrics such as Continuous Assessment, Practical Skills, Presentation Quality, Attendance, and Participation. The model achieved strong results, with a Mean Absolute Error (MAE) of 0.0249, Mean Squared Error (MSE) of 0.0012, and Root Mean Squared Error (RMSE) of 0.0346 for the regression task. For the classification task, it achieved perfect scores with an accuracy, precision, recall, and F1 score of 1.0. The attention mechanism enhanced performance by focusing on the most relevant features. This study demonstrates the effectiveness of the Multi-Task LSTM model with an attention mechanism in educational data analysis, offering a reliable and efficient tool for predicting student performance.

**Keywords:** e-learning; student; performance; multi-task; deep learning; attention mechanism



**Citation:** Olaniyan, D.; Olaniyan, J.; Obagbuwa, I.C.; Esiefarienrhe, B.M.; Bernard, O.P. Parallel Attention-Driven Model for Student Performance Evaluation. *Computers* **2024**, *13*, 242. <https://doi.org/10.3390/computers13090242>

Academic Editor: Antonio Sarasa Cabezuelo

Received: 6 September 2024

Revised: 14 September 2024

Accepted: 19 September 2024

Published: 23 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Traditional student evaluation practices offer valuable insights into academic achievement and learning progress, providing a structured means to measure how well students understand course material and achieve learning objectives [1]. These methods, such as standardized testing, summative assessments, grading, and multiple-choice tests, are designed to assess a wide range of cognitive skills, from basic recall to higher-order thinking [2].

However, despite their utility, these traditional practices have long relied on methods that are ethically and intellectually questionable [3]. A significant concern is that these practices often do not allow students to explain their answers [4]. For instance, multiple-choice tests and other standardized assessments typically offer limited opportunities for students to demonstrate their reasoning or the process by which they arrived at their answers [5]. This can lead to an incomplete understanding of a student's true capabilities and knowledge, as these formats prioritize the final answer over the thought process.

Moreover, traditional evaluation methods frequently fail to accommodate the diverse physiological strengths and weaknesses of students [6]. Standardized tests, for instance, are often designed with a one-size-fits-all approach that does not consider individual differences in learning styles, cognitive processing speeds, or test-taking abilities [7]. Students

with different learning needs, such as those with learning disabilities or attention disorders, may find these assessments particularly challenging and unreflective of their true potential [8].

Student performance accountability should not rest solely on the students themselves. Instead, assessing learning outcomes and guiding instructional practices should involve a broader range of metrics, which are often overlooked [9]. Student evaluation extends beyond mere testing and intellectual performance; it should not rely predominantly on exam scores as the primary measure of academic achievement. Although exams offer valuable insights into a student's grasp of the course material, they often fall short in capturing the multifaceted nature of learning and development [10].

For physiological differences, time-honored practices often overlook the importance of providing a holistic view of student development [11]. Behavioral and practical skills, critical thinking, creativity, and collaborative abilities are crucial aspects of learning that are not adequately captured by conventional testing methods. As a result, students who excel in these areas but struggle with traditional exams may not receive the recognition they deserve for their overall capabilities and contributions [12].

Although traditional student evaluation practices provide essential insights into academic performance, their limitations in addressing ethical and intellectual concerns, along with their failure to consider the diverse needs and strengths of all students, highlight the need for a more comprehensive and inclusive approach to assessment [13]. This approach should allow for greater flexibility, accommodate individual differences, and emphasize a broader range of skills and competencies to truly reflect a student's overall learning and development [14].

Incorporating a broader range of metrics to assess student performance such as Behavioral Factors, Practical Skills, Class Participation, and Engagement has received some attention from researchers [15]. However, these should be acknowledged as essential components of the learning process that contribute significantly to a student's overall success [16].

Despite this recognition, incorporating diverse performance metrics into a cohesive evaluation framework remains a significant challenge. Existing evaluation models often lack the flexibility to accommodate multiple types of data and struggle to provide interpretable insights into the factors driving evaluation outcomes [17]. As a result, educators face difficulty in understanding the variations in student performance and tailoring their instructional strategies accordingly [18].

To address these challenges, this study proposes the development of an innovative holistic student evaluation model for e-learning environments. This model aims to incorporate a diverse array of performance metrics, including exams, behavior assessments, practical assignments, presentations, class attendance, participation, and continuous assessments. By leveraging advanced deep learning techniques, such as LSTM-based Multi-Task learning and attention mechanisms, the proposed model seeks to capture the complex relationships among various evaluation metrics, providing a more comprehensive understanding of student performance.

This paper makes the following key contributions:

1. **Development of a Holistic Evaluation Model:** A novel student evaluation model that integrates a wide range of metrics beyond exams, including behavioral assessments, practical assignments, attendance, and participation.
2. **Application of Multi-Task Learning:** The proposed model utilizes LSTM-based Multi-Task learning, simultaneously addressing regression (total score prediction) and classification (performance category) tasks, thus optimizing computational efficiency.
3. **Attention Mechanism:** An attention mechanism is also introduced in order to render the model more focused on relevant features, enhancing both the accuracy and interpretability of its predictions.
4. **Extensive Performance Analysis:** In this context, the proposed model is evaluated using a generated dataset that simulates detailed student performance metrics, demon-

strating its ability to capture complex relationships across various evaluation criteria and provide valuable insights into overall student performance.

5. Potential for Wide Application: The suggested method has great potential for use in many educational areas as a powerful tool for the comprehensive assessment of students in e-learning.

The paper is organized as follows: Section 2 presents a comprehensive review of the related literature and empirical findings. In Section 3, the methodology is discussed, outlining the research design, model specification, material used, and data processing techniques. Section 4 focuses on the evaluation metrics and presents the experimental findings. Finally, Section 5 offers a detailed discussion of the results, conclusions, and recommendations for future research.

## 2. Related Works

The integration of technology in education has transformed the landscape of teaching and learning, leading to the emergence of innovative approaches to student assessment, prediction, and recommendation. As digital platforms become increasingly prevalent in educational settings, researchers are exploring novel methods to enhance learning outcomes, personalize instruction, and improve student engagement. This literature review examines recent studies that investigate various aspects of e-learning, including a cognitive classification of text, prediction of student behavior, performance analysis, and course recommendation. Recently, Sebbaq [19] focused on the cognitive classification of text in e-learning materials, employing Bloom's taxonomy as a framework. The study introduces MTBERT-Attention, a model combining Multi-Task learning (MTL), BERT, and a co-attention mechanism to enhance generalization capacity and data augmentation. Comprehensive testing demonstrates the model's superior performance and explainability compared to baseline models. Furthermore, Liu et al. [20] address the prediction of student behavior in e-learning environments. They propose a variant of Long-Short Term Memory (LSTM) and a soft-attention mechanism to model heterogeneous behaviors and make multiple predictions simultaneously. Experimental results have validated the effectiveness of the proposed model in predicting student behaviors and improving academic outcomes. Additionally, Xie, 2021 [21], focuses on predicting student performance in online education using demographic data and click-stream interactions. The study introduces an Attention-based Multi-layer LSTM (AML) model, which combines demographic and click-stream data for comprehensive analysis. Experimental results demonstrate improved prediction accuracy and F1 score compared to baseline methods.

He et al. [22] explore Knowledge Tracing (KT) in e-learning platforms, proposing Multi-Task Attentive Knowledge Tracing (MAKT) to improve prediction accuracy. The study introduces Bi-task Attentive Knowledge Tracing (BAKT) and Tri-task Attentive Knowledge Tracing (TAKT) models, which jointly learn hint-taking, attempt-making, and response prediction tasks. Experimental results show that MAKT outperforms existing KT methods, indicating promising applications of Multi-Task learning in KT. And more recently, Su et al. [23] investigated cross-type recommendation in Self-Directed Learning Systems (SDLSs), proposing the Multi-Task Information Enhancement Recommendation (MIER) Model. The study integrates resource representation and recommendation tasks using an attention mechanism and knowledge graph. Experimental results demonstrate the superior performance of the MIER model in predicting concepts and recommending exercises compared to existing methods. Ren et al. [24] focus on course recommendation in online education platforms, proposing a deep course recommendation model with multimodal feature extraction. The study utilizes LSTM and attention mechanisms to fuse course video, audio, and textual data, supplemented with user demographic information and feedback. Experimental results show that the proposed model achieves significantly higher AUC scores compared to similar algorithms, providing accurate course recommendations for users.

Conclusively, the current literature on e-learning reveals significant gaps that justify the need for designing a holistic student evaluation model using an LSTM Multi-Task attention-based deep learning approach. Existing studies predominantly focus on isolated aspects of e-learning, such as cognitive classification, behavior prediction, performance analysis, and course recommendation, without integrating multiple performance metrics for a comprehensive evaluation. There is also a lack of interdisciplinary integration of advanced techniques like Multi-Task learning and attention mechanisms, which could enhance model robustness. A holistic evaluation model that incorporates these aspects would significantly improve the effectiveness and user experience of e-learning platforms.

### 3. Materials and Methods

#### 3.1. Materials

##### 3.1.1. Proposed Architecture

Figure 1 illustrates the architectural framework for the Multi-Task LSTM with an attention mechanism (MLSTM-AM) model for accurate prediction of students' academic performance.

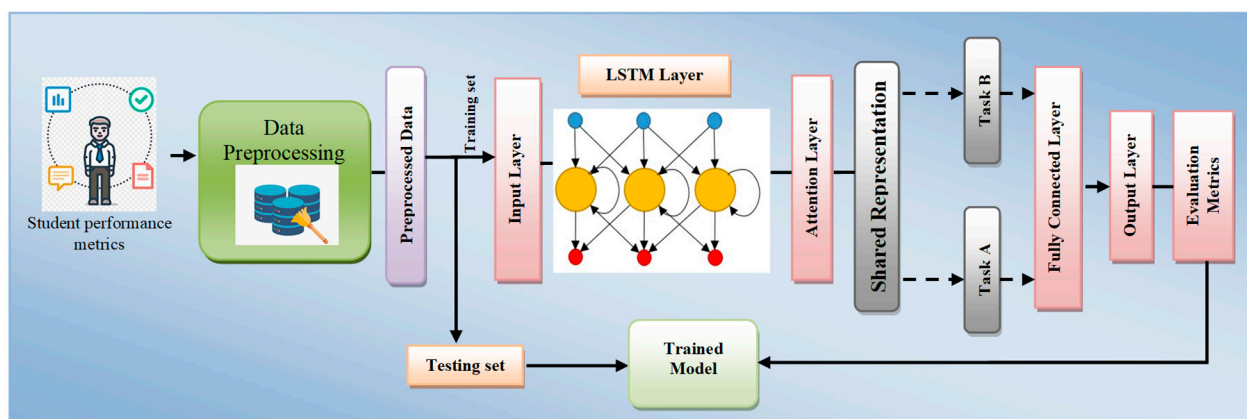


Figure 1. Proposed framework.

The proposed framework depicts the different components of the model, starting with the dataset, data pre-processing, model design and training, and validation.

##### 3.1.2. Dataset Creation

The typical student performance dataset traditionally revolves around two key variables: Continuous Assessment (CA) and Examination (Exam) scores. These metrics have long been the sole basis for evaluating student's academic achievement. However, this research introduces a paradigm shift by expanding the dimensions of this performance evaluation framework. By integrating additional variables such as Attendance, Demeanor, Practical Skills, Class Participation, and Presentation Quality [25], this study pioneers a more comprehensive approach to assessing student success.

The dataset needed for this kind of new model is not readily available anywhere, hence the need to create a new dataset that captures the above-mentioned variables. To achieve this, a suitable mathematical formulation was utilized, which was then translated into a computer algorithm for creating such a higher dimensional dataset, as shown in this section.

Let  $x_1, x_2, x_3, \dots, x_7$  be Attendance, Practical, Demeanor, Presentation, Participation in class, Continuous Assessment, and Examination, respectively, where  $x_1, x_2, x_3, \dots, x_6$  can take on 10 distinct values (i.e., 10% each), while  $x_7$  can take on 40 distinct values (i.e., 40% only). The total score, which is the sum of all these components, must not exceed 100%, the maximum score a student can achieve in a given course.

i.e.,

$$\text{score} = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7$$

which can also be written as

$$score = \sum_{i=1}^n x_i \forall i = 1, 2, \dots, n \text{ and } n = 7$$

i.e.,  $score \leq 100\%$

To find the total number of combinations, multiply the number of possibilities for each variable:

$$\begin{aligned} \text{Total combinations} &= (\text{Number of possibilities for } x_1) \times (\text{Number of possibilities for } x_2) \times (\text{Number of possibilities for } x_3) \times (\text{Number of possibilities for } x_4) \times (\text{Number of possibilities for } x_5) \times (\text{Number of possibilities for } x_6) \times (\text{Number of possibilities for } x_7) \\ &= 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 40 \\ &= 1,000,000 \times 40 \\ &= 40,000,000 \end{aligned}$$

So, the total number of combinations is 40,000,000. In other words, the total number of data points would be 40,000,000, i.e., the new dataset containing all the above features would have 40,000,000 records, which is enough to train the proposed model. This formulation was implemented in Python, and the code snippet is displayed in Algorithm 1.

---

**Algorithm 1:** Pseudocode for data pre-processing

---

Input: combinations of values for  $x_1$  to  $x_6$ , then appends  $x_7$  to each combination.

Output: CSV file ("resultPredictionDataset.csv")

---

```
import itertools
import csv
# Define the range of values for each variable
range_values = range(11) # Values for x1 to x6 (0 through 10)
range_x7 = range(41) # Values for x7 (0 through 40)
# Generate all combinations of the variables x1 to x6
combinations = list(itertools.product(range_values, repeat=6))
# Append x7 to each combination
combinations_with_x7 = [(c + (x7,)) for c in combinations for x7 in range_x7]
# Calculate the total for each combination (sum of x1 to x7)
combinations_with_total = [(c + (sum(c),)) for c in combinations_with_x7]
# Specify the file name
file_name = "resultPredictionDataset.csv"
# Write combinations with total to CSV file
withopen(file_name, 'w', newline='') as csvfile:
    csvwriter = csv.writer(csvfile)
    # Write the header row
    csvwriter.writerow(["x1", "x2", "x3", "x4", "x5", "x6", "x7", "total"])
    # Write the data rows
    csvwriter.writerows(combinations_with_total)
print(f"Final Dataset Generated Successfully to {file_name}")
```

---

### 3.1.3. Dataset Description

The dataset generated by Algorithm, which has been published in the Kaggle repository (<https://www.kaggle.com/datasets/olaniyanjulius/student-academic-performance-dataset>) assessed on 4 September 2024, contains a collection of 40,000,000 records, each detailing various aspects important for evaluating student academic performance. These records include Continuous Assessment (CA), Practical Skills proficiency, Demeanor, Presentation Quality, Attendance records, Participation in class, and Examination results. The final two columns represent the overall performance (total score) and the performance class (remarks), which range from 1 to 5, indicating different levels of student achievement.

To provide clarity and facilitate analysis, each feature is assigned a corresponding variable;  $x_1$  corresponds to CA,  $x_2$  corresponds to Practical proficiency,  $x_3$  corresponds to

Demeanor,  $x_4$  corresponds to Presentation Quality,  $x_5$  corresponds to Attendance,  $x_6$  corresponds to Participation in class, and finally,  $x_7$  corresponds to Examination results. This structured framework not only simplifies data interpretation but also lays the groundwork for comprehensive analysis and insights into student academic performance.

The CA variable evaluates ongoing performance through assignments, quizzes, and tests, providing insight into consistent engagement and mastery of material. The Practical Skills variable assesses hands-on proficiency in applying theoretical knowledge through lab work and projects [26]. Demeanor focuses on punctuality, attentiveness, and overall conduct, reflecting social and emotional intelligence [27]. Presentation Quality is evaluated through the clarity and effectiveness of student presentations, highlighting communication skills [28]. Attendance records quantify commitment and consistency in attending classes. Participation in class measures active engagement and contribution during discussions and activities [29,30]. Together, these variables offer a comprehensive framework for assessing student strengths, areas for improvement, and overall academic progress. Figure 2 presents the structure of the dataset generated in this research.

	A	B	C	D	E	F	G	H	I
1	x1	x2	x3	x4	x5	x6	x7	total	remarks
2	0	1	1	6	8	4	35	55	3
3	0	0	4	5	2	6	0	17	1
4	0	0	8	9	3	7	0	27	1
5	0	0	9	0	9	8	25	51	3
6	0	0	4	6	6	7	30	53	3
7	0	0	0	7	6	1	34	48	2
8	0	0	7	1	10	4	17	39	1
9	0	0	2	4	5	3	25	39	1
10	0	0	6	0	5	2	17	30	1
11	0	0	3	5	9	9	33	59	3
12	0	1	1	6	2	9	21	40	2
13	0	0	9	1	4	2	24	40	2
14	0	0	1	5	1	8	9	24	1
15	0	0	4	2	10	1	2	19	1
16	0	0	4	9	2	7	3	25	1
17	0	0	9	0	1	10	10	30	1
18	0	0	5	4	6	8	6	29	1
19	0	0	7	5	8	6	16	42	2
20	0	1	0	5	6	10	30	52	3
21	0	0	4	2	1	10	26	43	2
22	0	0	6	5	6	7	14	38	1
23	0	1	0	6	5	6	32	50	3
24	0	0	5	7	5	9	37	63	4
25	0	0	0	2	2	5	14	23	1
26	0	0	6	2	0	8	4	20	1
27	0	0	1	4	1	4	37	47	2
28	0	0	5	10	9	6	23	53	3
29	0	0	6	7	6	9	25	53	3
30	0	1	1	10	6	5	7	30	1
31	0	0	6	6	3	8	28	51	3
32	0	0	6	4	8	2	23	43	2

Figure 2. Dataset sample.

### 3.1.4. Data Analysis

Given the generated dataset, this study aims to explore the differences in student performance across various classes. To achieve this, an Analysis of Variance (ANOVA) was employed to determine if there were statistically significant differences in the performance scores among the different classes. To ensure the validity of the ANOVA results, normality

was first checked using Q-Q plots, shown in Figure 3, and homogeneity of variance was assessed using Levene’s test, as depicted in Tables 1 and 2.

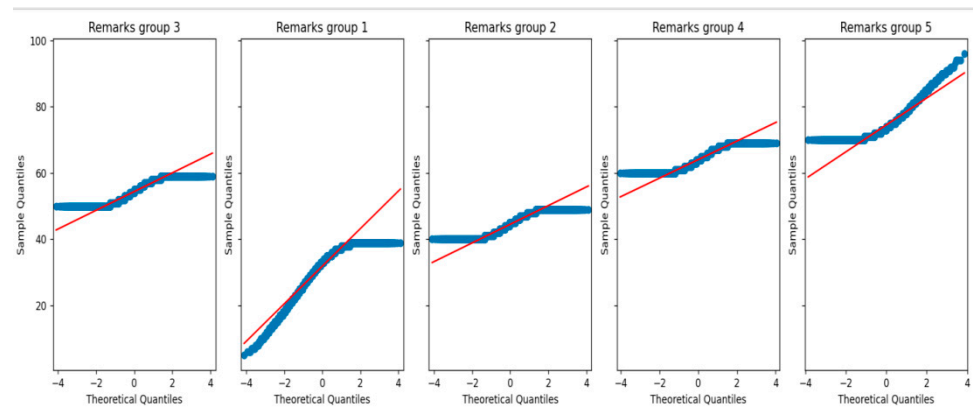


Figure 3. Q-Q plots.

Table 1. ANOVA results.

Levene’s Test: $p$ -Value = 0.0				
Source	sum_sq	df	F	PR (>F)
C (remarks)	$3.701384 \times 10^7$	4.0	624,662.877938	0.0
Residual	$2.962631 \times 10^6$	199,995	NaN	NaN

Multiple comparison of means—Tukey HSD, FWER = 0.05.

Table 2. Tukey’s HSD multiple comparisons of means test.

Group1	Group2	Meandiff	$p$ -Adj	Lower	Upper	Reject
1	2	12.6824	0.0	12.6151	12.7497	True
1	3	22.5543	0.0	22.4871	22.6215	True
1	4	32.2145	0.0	32.1428	32.2863	True
1	5	42.6735	0.0	42.5821	42.765	True
2	3	9.8719	0.0	9.8036	9.9402	True
2	4	19.5321	0.0	19.4593	19.6049	True
2	5	29.9911	0.0	29.8998	30.0834	True
3	4	9.6602	0.0	9.5876	9.7329	True
3	5	20.1192	0.0	20.027	20.2114	True
4	5	10.459	0.0	10.3634	10.5546	True

The Q-Q plots indicated that the normality assumption was reasonably met, while Levene’s test revealed a  $p$ -value of 0.0, indicating significant differences in variances across groups.

As shown in Tables 1 and 2, the analysis of variance (ANOVA) performed on the dataset revealed significant differences in student performance across different classes. Levene’s test for homogeneity of variances produced a  $p$ -value of 0.0, indicating that the variances among the groups were significantly different. Although this result suggests a violation of the homogeneity of variance assumption, it was expected, given the context of the dataset, which comprises different classes of student performance.

The ANOVA results as depicted in Table 1 further support the presence of significant differences between groups. The F-statistic was calculated to be 624,662.88 with a corresponding  $p$ -value of 0.0. This extremely low  $p$ -value allows us to reject the null hypothesis and conclude that there are significant differences in the mean performance scores among the different classes of students as expected in this context.

Following the ANOVA, Tukey’s HSD post hoc test as shown in Table 2 was conducted to identify which specific groups differed significantly from each other. The results showed

that all pairwise comparisons between the groups were significant, with each group displaying a mean difference that was statistically significant ( $p\text{-adj} < 0.05$ ) as expected. This indicates that the performance of students in each class is distinctively different from the others.

In summary, the analysis demonstrates clear and significant differences in student performance across different classes. Although the homogeneity of variance assumption was not met, the context of varying student performance classes justifies these significant differences. The results align with expectations and provide a robust indication of distinct performance levels among the different classes.

### 3.1.5. Data Preprocessing

As illustrated in Algorithm 2, the dataset is first loaded from a CSV file, ensuring that all subsequent operations are based on the complete dataset. Following this, relevant features and target variables are extracted. The features include various metrics such as 'x1', 'x2', 'x3', 'x4', 'x5', 'x6', and 'x7', while the target variables consist of the 'total' score and 'remarks'. The 'remarks' target variable is adjusted to zero-based indexing in order to align the target values with typical numerical representations used in model training. The feature set is then reshaped to fit the input requirements of a Long Short-Term Memory (LSTM) network. This reshaping transforms the data into the format required for LSTM input, with dimensions corresponding to samples, timesteps, and features [31]. Then, the dataset is divided into training and testing subsets, allowing the model to be trained on one portion of the data and evaluated on a separate, independent portion to assess its performance effectively. Finally, the input shape for the LSTM network is defined based on the reshaped data, which is important for setting up the network architecture correctly.

---

#### Algorithm 2: Data Preprocessing

---

**Input:** X: Features, y\_total: Total score, y\_remarks: Performance levels

**Output:** Input shape of the data for LSTM

---

```
function dataPreprocessing(dataset)
    # Step 1: Load the dataset from the CSV file
    X, y_total, y_remarks = extractFeaturesAndTargets(dataset)
    # Step 2: Adjust y_remarks for zero-based indexing (if needed)
    y_remarks = adjustZeroBasedIndex(y_remarks)
    # Step 3: Reshape the input features X to fit the LSTM input format
    # (samples, timesteps, features) X_resaped = reshapeForLSTM(X)
    # Step 4: Split the dataset into training and testing sets X_train, X_test, y_total_train,
    y_total_test, y_remarks_train, y_remarks_test = splitData(X_resaped, y_total, y_remarks)
    # Step 5: Define the input shape for the LSTM network based on the reshaped data
    input_shape = defineInputShape(X_resaped)
    # Return processed datasets and input shape
    return X_train, X_test, y_total_train, y_total_test, y_remarks_train, y_remarks_test,
    input_shape
end function
```

---

## 3.2. Method

### 3.2.1. The Model

The proposed model integrates Multi-Task LSTM with an attention mechanism to enhance the prediction of students' academic performance. As detailed in Algorithm 3, the model addresses two dependent variables: the total and remarks variables. Predicting the total variable is a regression task, while predicting the remarks variable is a classification task. Traditionally, handling these tasks would involve splitting the dataset into two and training them separately, which is time-consuming and resource-intensive [32]. Therefore, a Multi-Task LSTM model is employed to manage both tasks concurrently. Additionally, an attention mechanism is utilized to identify and extract the most relevant features



from the dataset for the Multi-Task LSTM model [33]. The integration of these models is mathematically presented in this section.

### 3.2.2. LSTM Layer

For sequence  $S = [s_1 s_2 \dots s_T]$  with input features  $S_t \in R^d$ , the input features  $S_t$  represent a vector of dimension  $d$  at each time step  $t$ , LSTM layer computes forget gate, input gate, cell state update, cell state, output gate, and hidden state as illustrated in Equations (1)–(6), respectively.

$$f_G = \text{SAF}(W_f * [h_{t-1}, n_t] + b_f) \quad (1)$$

$$i_G = \text{SAF}(W_i * [h_{t-1}, n_t] + b_i) \quad (2)$$

$$Z_t = \tan h(W_z * [h_{t-1}, n_t] + b_z) \quad (3)$$

$$C_t = f_G * C_{t-1} + i_G * Z_t \quad (4)$$

$$o_G = \text{SAF}(W_o * [h_{t-1}, n_t] + b_o) \quad (5)$$

$$h_G = o_G * \tan h(C_t) \quad (6)$$

where *SAF* is the Sigmoid Activation Function;  $f_G$  is the forget gate;  $W_f$  is the weight matrix for the forget gate, Concatenation of the previous hidden state  $[h_{t-1}]$  and the current input  $n_t$ ;  $b_f$  is the bias term for the forget gate;  $i_G$  is the input gate;  $W_i$  is the weight matrix for the input gate; and  $b_i$  is the bias term for the input gate.  $Z_t$  is the candidate cell gate,  $W_z$  is the weight matrix for the candidate cell gate,  $b_z$  is the bias term for the candidate cell state,  $C_t$  is the cell state update,  $C_{t-1}$  is the previous cell state,  $o_G$  is the output gate,  $h_G$  is the hidden state, and  $\tan h$  is the hyperbolic tangent function.

### 3.2.3. Attention Mechanism

The attention mechanism is used to dynamically compute a **context vector**  $C_t$ , which is based on the sequence of hidden states  $H = [h_1, h_2, \dots, h_T]$ . The context vector  $C_t$  is derived by taking a weighted sum of these hidden states, where each hidden state  $h_t$  contributes to the final context depending on its importance, as defined by its corresponding attention score  $A_t$ . This process is described in Equation (7):

$$C_t = \sum_{t=1}^T A_t * h_t \quad (7)$$

Attention score ( $A_t$ ) is computed as depicted in Equation (8).

$$A_t = \frac{\exp(e_t)}{\sum_{j=1}^T \exp(e_j)} \quad (8)$$

where  $e_t$  is a raw score that reflects the relevance of the hidden state  $h_t$  to the current context.  $\sum_{j=1}^T \exp(e_j)$  is the sum of the exponentials of all raw scores from time steps 1 to  $T$ , ensuring that the attention scores are a proper probability distribution over the sequence.  $T$  defines how many time steps are in the sequence, and  $j$  is used to compute the normalized attention score by summing over all time steps

### 3.2.4. Regression Analysis

For the regression task, the output denoted by  $y_t$  is given in Equation (9).

$$y_t = W_r * h_r + b_r \quad (9)$$

where  $y_t$  is the predicted output at time step  $t$ .  $W_r$  is the weight matrix for the output layer,  $h_r$  is the hidden state from the previous layer or the current hidden representation. And  $b_r$  is the bias term for the output layer.

$$h_r = \text{ReLU}(W_{rh} * h + b_{rh}) \quad (10)$$

where  $h_r$  is the hidden state at time  $t$ ,  $W_{rh}$  is the weight matrix for the hidden layer.  $h$  represents the input or hidden state from the previous layer.  $b_{rh}$  is the bias term for the hidden layer.  $\text{ReLU}$  is the activation function, defined as  $\text{ReLU} = \max(0, x)$ , which introduces non-linearity.

### 3.2.5. Classification Analysis

For the classification task, the output layer denoted by  $y_o$  is given in Equation (11).

$$y_o = \text{softmax}(W_o * h_o + b_o) \quad (11)$$

where  $y_o$  is the predicted output for the classification task,  $W_o$  is the weight matrix for the output layer,  $h_o$  is the hidden state from the previous layer or hidden representation and  $b_o$  is the bias term for the output layer.  $\text{softmax}$  is an activation function often used for multi-class classification. It converts the raw output scores into probabilities that sum to 1, making it suitable for categorical prediction.

$$h_c = \text{ReLU}(W_c * h + b_c) \quad (12)$$

The Equation (12) transforms the input  $h$  by applying a weight matrix  $W_c$  and bias  $b_c$ , followed by passing the result through the ReLU activation function. This creates a non-linear hidden representation  $h_c$  used for further processing in the network. Therefore, the integration is modeled as in Equations (13)–(16).

$$\text{LSTM Layer Output} : H = \text{LSTM}(X) \quad (13)$$

$$\text{Attention Output} : c = \text{Attention}(H) \quad (14)$$

$$\text{Regression Output} : y_{total} = W_o * \text{ReLU}(W_r * H + b_o) + b_r \quad (15)$$

$$\text{Classification Output} : y_{class} = W_o * \text{softmax}(W_o * \text{ReLU}(W_c * H + b_c) + b_o) \quad (16)$$

where  $H$  is the output sequence from the LSTM layer, and  $W, b$  are the respective weights and biases for each task.

### 3.2.6. Combined Regression and Classification

Then, both regression and classification losses are computed as illustrated in Equations (17) and (18), respectively.

$$\text{Regression Loss} : \text{Loss}_{reg} = \frac{1}{N} \sum_{i=1}^N (y_{total} - \hat{y}_{total})^2 \quad (17)$$

$$\text{Classification Loss} : \text{Loss}_{class} = -\frac{1}{N} \sum_{i=1}^N y_{class} \log(\hat{y}_{class}) \quad (18)$$

where  $\hat{y}$  denotes the predicted values.

## 4. Results and Discussions

This section presents the results and findings of the experimental study focused on developing the Multi-Task LSTM model with an attention mechanism for predicting students' academic performance conducted in this research.

#### 4.1. Performance Evaluation Metrics

The performance of the proposed model is assessed using a variety of metrics tailored to both the regression and classification tasks. These metrics offer a detailed evaluation of the model's effectiveness in predicting students' academic performance.

##### 4.1.1. Regression Task

For the regression task, which involves predicting the 'total' variable, three primary metrics are employed. The Mean Absolute Error (MAE) measures the average magnitude of errors in the predictions, indicating how closely the predicted values align with the actual values [34]. The MAE for the regression task is 0.0249, reflecting a low average prediction error. The Mean Squared Error (MSE) calculates the average of the squared differences between predicted and actual values, with a value of 0.0012. This metric penalizes larger errors more significantly and indicates that the model maintains a low average squared error. Additionally, the Root Mean Squared Error (RMSE) provides a measure of the average magnitude of prediction errors in the same units as the target variable [35]. With an RMSE of 0.0346, the model demonstrates a minimal average prediction error, underscoring its regression accuracy.

##### 4.1.2. Classification Task

For the classification task, which involves predicting the 'remarks' variable, several performance metrics are considered. Accuracy measures the proportion of correctly classified instances out of the total instances [36], and the model achieves a perfect accuracy of 1.0, signifying flawless classification performance. Precision reflects the proportion of true positive predictions among all positive predictions made by the model [37], and a precision of 1.0 indicates that every positive classification was correct. Recall, which measures the proportion of actual positive instances correctly identified by the model [38], also achieves a perfect score of 1.0, demonstrating the model's ability to identify all actual positive instances. Finally, the F1 Score, the harmonic mean of precision and recall, balances these two aspects of classification performance [39]. An F1 Score of 1.0 highlights the model's ideal precision and recall, illustrating its overall effectiveness in classification. The visual depiction of the model's performance metrics is illustrated in Figure 4 and Table 3.

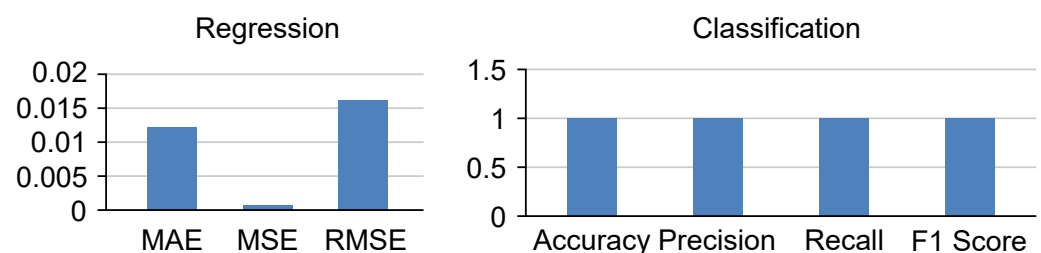


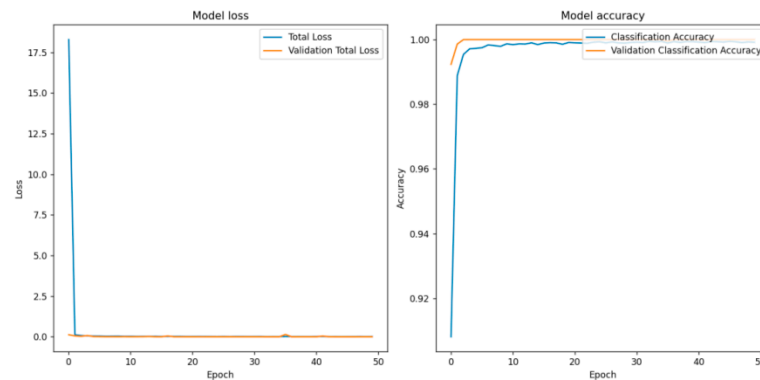
Figure 4. MLST-AM performance plot.

Table 3. Performance metrics of the Multi-Task LSTM model.

Metric	Value
Mean Absolute Error (MAE)	0.012
Mean Squared Error (MSE)	0.000254
Root Mean Squared Error (RMSE)	0.01594
Accuracy	1.0 (100%)
Precision	1.0
Recall	1.0
F1 Score	1.0

#### 4.2. Training and Evaluation Results

To further evaluate the model's performance, both the training and evaluation accuracies and losses were plotted for visualization, as presented in Figure 5.



**Figure 5.** Training and validation accuracy and loss.

##### 4.2.1. Regression Task Result

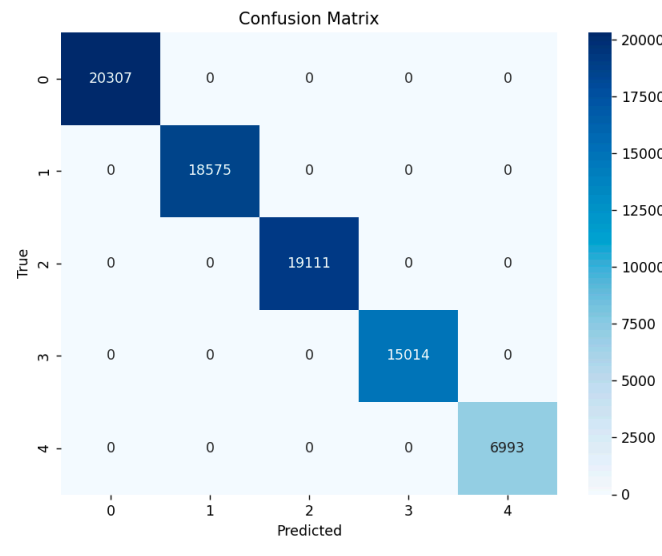
Over the course of 50 epochs, the training and validation metrics of the MLSTM-AM model demonstrated substantial improvement. The training and validation process of the proposed model demonstrates its robust performance in both regression and classification tasks. For the regression task, the model achieved a Mean Absolute Error (MAE) of 0.012, indicating a low average magnitude of prediction errors, while the Mean Squared Error (MSE) of 0.000254 and the Root Mean Squared Error (RMSE) of 0.01594 further confirm the model's precision, with minimal differences between predicted and actual values. These metrics collectively highlight the model's capability to accurately predict continuous outcomes, ensuring reliable performance in the regression task.

##### 4.2.2. Classification Task Result

For the classification task, the model exhibits exceptional accuracy, precision, recall, and F1 score, all achieving a perfect value of 1.0. This indicates that the model correctly classifies all instances without any errors. The high classification accuracy reflects the model's ability to distinguish between different classes effectively, while the precision and recall values demonstrate its proficiency in identifying true positives and minimizing false positives and negatives. The perfect F1 score balances these metrics, reinforcing the model's overall robustness in handling classification tasks. These results show the effectiveness of the training and validation process, ensuring the model's reliability and accuracy in predicting both continuous and categorical outcomes.

##### 4.2.3. Confusion Matrix

The performance of the model in the evaluation of the academic performance of students is based on five categories: Fail = 0, Pass = 1, Good = 2, Very Good = 3, and Excellent = 4. Figure 6 illustrates a confusion matrix that further shows the performance of the model. From this matrix, correctly classified students are 20,307 as Fail, 18,575 as Pass, 19,111 as Good, 15,014 as Very Good, and 6993 as Excellent, while misclassifications do not occur for any category. That means that at each level, the model has correctly classified students without mistakes. The perfect classification of all categories proves the model's appropriateness for assessing and distinguishing various levels of academic performance and affords reliability to it as a tool in the study of academic evaluation.



**Figure 6.** Confusion Matrix for the Classification.

#### 4.3. Comparative Analysis

As illustrated in Table 4, this section presents the comparison of the performance of the proposed model (MLSTM-AM) with several other recent e-learning studies. The comparison is based on Focus Area, Techniques Used, Metrics, and Gaps Addressed since they were all trained on the different dataset and performance metrics. The comparison models shown in Table 4 show several significant improvements. Existing models such as [19–21,23,24] MTBERT-Attention, Liu et al.’s best model, Xie et al. Multi-Task Attentive Knowledge Tracing (Su et al.’s MIER Model), and Ren et al.’s deep recommendation model offer significant insights in their respective area; they place emphases on particular text classification or behavior prediction or course recommendations, but neither of these leverages multiple performance metrics at once.

**Table 4.** Performance comparison.

Author	Focus Area	Techniques Used	Metrics	Gaps	Proposed Model
[19]	Cognitive classification of text	Multi-Task BERT (MTBERT-Attention) with co-attention mechanism	Superior performance and explainability in text classification	Focuses on text classification only, lacks holistic student evaluation	Integrates multiple performance metrics, captures complex relationships
[20]	Prediction of student behavior	LSTM with soft-attention mechanism	Effective in predicting student behaviors and improving academic outcomes	Does not consider holistic student performance, limited to behavior prediction	Uses LSTM with Multi-Task learning for both regression and classification
[21]	Predicting student performance	Attention-based Multi-layer LSTM (AML)	Improved prediction accuracy and F1 score using demographic and clickstream data	Limited to performance prediction, lacks comprehensive metric integration	Combines various metrics for a complete evaluation of student performance
[22]	Knowledge Tracing (KT)	Multi-Task Attentive Knowledge Tracing (MAKT)	Improved prediction accuracy in KT tasks	Focuses on KT, does not address real-time feedback or holistic evaluation	Provides real-time feedback, integrates multiple metrics for holistic evaluation

Table 4. Cont.

Author	Focus Area	Techniques Used	Metrics	Gaps	Proposed Model
[23]	Cross-type recommendation in SDLS	Multi-Task Information Enhancement Recommendation (MIER) Model with attention and knowledge graph	Superior performance in concept prediction and exercise recommendation	Limited to recommendation systems, does not provide holistic student evaluation	Utilizes attention mechanisms for comprehensive evaluation of multiple student metrics
[24]	Course recommendation	Deep course recommendation model with LSTM and Attention	Higher AUC scores in course recommendations	Focuses on course recommendations, lacks integration of diverse metrics	Integrates multimodal data for comprehensive student performance evaluation

The proposed model eliminates the shortfalls by employing an LSTM-based Multi-Task learning technique, encroached upon with attention mechanisms to offer a more in-depth analysis of student achievement. It integrates various metrics, offers real-time feedback, and demonstrates high precision in both regression and classification tasks, thereby enhancing overall student evaluation in e-learning environments.

## 5. Conclusions and Outlook

In this study, a Multi-Task LSTM model with an attention mechanism was proposed to predict student academic performance effectively. The model addressed both regression and classification tasks, predicting the ‘total’ score as a continuous variable and the ‘remarks’ as a categorical variable. This approach allowed for the efficient use of computational resources and time by handling both tasks concurrently.

The performance metrics demonstrated the model’s high accuracy and low error rates. For the regression task, the model achieved a Mean Absolute Error (MAE) of 0.012, a Mean Squared Error (MSE) of 0.000254, and a Root Mean Squared Error (RMSE) of 0.01594, indicating precise and reliable predictions. For the classification task, the model reached perfect scores across all metrics, with an accuracy, precision, recall, and F1 score of 1.0. These results highlight the model’s robustness and effectiveness in both predicting continuous outcomes and classifying categorical data.

Overall, the integration of a Multi-Task LSTM model with an attention mechanism proved to be a powerful approach for predicting student performance. The model’s ability to accurately predict both types of outcomes showcases its potential for broader applications in educational data analysis and other fields requiring multi-task learning capabilities. Future work could explore further enhancements and applications of this model to continue improving predictive accuracy and efficiency in diverse contexts.

**Author Contributions:** Conceptualization, D.O. and J.O.; methodology, D.O. and J.O.; software, J.O.; validation, I.C.O., D.O. and J.O.; formal analysis, O.P.B.; investigation, O.P.B.; resources, O.P.B.; data curation, J.O.; writing—original draft preparation, D.O.; writing—review and editing, I.C.O. and J.O.; supervision, D.O.; project administration, I.C.O.; funding acquisition, B.M.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding. And the APC was funded by Prof. Bukohwo Michael Esiefarienrhe.

**Data Availability Statement:** The data is available at <https://www.kaggle.com/datasets/olaniyanjulius/student-academic-performance-dataset> (accessed on 22 September 2024).

**Acknowledgments:** The acknowledgment section in this research paper expresses gratitude to various individuals and entities whose contributions were vital to the study’s completion.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Suskie, L. *Assessing Student Learning: A Common Sense Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
2. Scully, D. Constructing multiple-choice items to measure higher-order thinking. *Pract. Assess. Res. Eval.* **2019**, *22*, 4.
3. Morley, J.; Floridi, L.; Kinsey, L.; Elhalal, A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci. Eng. Ethics* **2020**, *26*, 2141–2168. [[CrossRef](#)] [[PubMed](#)]
4. Wei, L. Transformative pedagogy for inclusion and social justice through translanguaging, co-learning, and transpositioning. *Lang. Teach.* **2024**, *57*, 203–214. [[CrossRef](#)]
5. Say, R.; Visentin, D.; Saunders, A.; Atherton, I.; Carr, A.; King, C. Where less is more: Limited feedback in formative online multiple-choice tests improves student self-regulation. *J. Comput. Assist. Learn.* **2024**, *40*, 89–103. [[CrossRef](#)]
6. Li, S.; Zhang, X.; Li, Y.; Gao, W.; Xiao, F.; Xu, Y. A comprehensive review of impact assessment of indoor thermal environment on work and cognitive performance-Combined physiological measurements and machine learning. *J. Build. Eng.* **2023**, *71*, 106417. [[CrossRef](#)]
7. Ramsey, M.C.; Bowling, N.A. Building a bigger toolbox: The construct validity of existing and proposed measures of careless responding to cognitive ability tests. *Organ. Res. Methods* **2024**, 10944281231223127. [[CrossRef](#)]
8. Maki, P.L. *Assessing for Learning: Building a Sustainable Commitment across the Institution*; Routledge: London, UK, 2023.
9. Geletu, G.M.; Mihiretie, D.M. Professional accountability and responsibility of learning communities of practice in professional development versus curriculum practice in classrooms: Possibilities and pathways. *Int. J. Educ. Res. Open* **2023**, *4*, 100223. [[CrossRef](#)]
10. Mohan, R. *Measurement, Evaluation and Assessment in Education*; PHI Learning Pvt. Ltd.: Dehli, India, 2023.
11. Yuksel, P.; Bailey, J. Designing a Holistic Syllabus: A Blueprint for Student Motivation, Learning Efficacy, and Mental Health Engagement. In *Innovative Instructional Design Methods and Tools for Improved Teaching*; IGI Global: Hershey, PA, USA, 2024; pp. 92–108.
12. Thornhill-Miller, B.; Camarda, A.; Mercier, M.; Burkhardt, J.M.; Morisseau, T.; Bourgeois-Bougrine, S.; Vinchon, F.; El Hayek, S.; Augereau-Landais, M.; Mourey, F.; et al. Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *J. Intell.* **2023**, *11*, 54. [[CrossRef](#)]
13. Zughoul, O.; Momani, F.; Almasri, O.H.; Zaidan, A.A.; Zaidan, B.B.; Alsalem, M.A.; Albahri, O.S.; Albahri, A.S.; Hashim, M. Comprehensive insights into the criteria of student performance in various educational domains. *IEEE Access* **2018**, *6*, 73245–73264. [[CrossRef](#)]
14. AlAfnan, M.A.; Dishari, S. ESD goals and soft skills competencies through constructivist approaches to teaching: An integrative review. *J. Educ. Learn. (EduLearn)* **2024**, *18*, 708–718. [[CrossRef](#)]
15. Wong, Z.Y.; Liem, G.A.D. Student engagement: Current state of the construct, conceptual refinement, and future research directions. *Educ. Psychol. Rev.* **2022**, *34*, 107–138. [[CrossRef](#)]
16. Al-Adwan, A.S.; Albelbisi, N.A.; Hujran, O.; Al-Rahmi, W.M.; Alkhalifah, A. Developing a holistic success model for sustainable e-learning: A structural equation modeling approach. *Sustainability* **2021**, *13*, 9453. [[CrossRef](#)]
17. Zaffar, M.; Garg, S.; Milford, M.; Kooij, J.; Flynn, D.; McDonald-Maier, K.; Ehsan, S. Vpr-bench: An open-source visual place recognition evaluation framework with quantifiable viewpoint and appearance change. *Int. J. Comput. Vis.* **2021**, *129*, 2136–2174. [[CrossRef](#)]
18. Goodwin, B.; Rouleau, K.; Abila, C.; Baptiste, K.; Gibson, T.; Kimball, M. *The New Classroom Instruction That Works: The Best Research-Based Strategies for Increasing Student Achievement*; ASCD: Arlington, VA, USA, 2022.
19. Sebbaq, H. MTBERT-Attention: An Explainable BERT Model based on Multi-Task Learning for Cognitive Text Classification. *Sci. Afr.* **2023**, *21*, e01799. [[CrossRef](#)]
20. Liu, H.; Zhu, Y.; Zang, T.; Xu, Y.; Yu, J.; Tang, F. Jointly modeling heterogeneous student behaviors and interactions among multiple prediction tasks. *ACM Trans. Knowl. Discov. Data (TKDD)* **2021**, *16*, 1–24. [[CrossRef](#)]
21. Xie, Y. Student performance prediction via attention-based multi-layer long-short term memory. *J. Comput. Commun.* **2021**, *9*, 61–79. [[CrossRef](#)]
22. He, L.; Li, X.; Wang, P.; Tang, J.; Wang, T. Integrating fine-grained attention into multi-task learning for knowledge tracing. *World Wide Web* **2023**, *26*, 3347–3372. [[CrossRef](#)]
23. Su, Y.; Yang, X.; Lu, J.; Liu, Y.; Han, Z.; Shen, S.; Huang, Z.; Liu, Q. Multi-task Information Enhancement Recommendation model for educational Self-Directed Learning System. *Expert Syst. Appl.* **2024**, *252*, 124073. [[CrossRef](#)]
24. Ren, X.; Yang, W.; Jiang, X.; Jin, G.; Yu, Y. A deep learning framework for multimodal course recommendation based on LSTM+ attention. *Sustainability* **2022**, *14*, 2907. [[CrossRef](#)]
25. Hamidi, H.; Hejran, A.B.; Sarwari, A.; Edigeevna, S.G. The Effect of Outcome Based Education on Behavior of Students. *Eur. J. Theor. Appl. Sci.* **2024**, *2*, 764–773. [[CrossRef](#)]
26. Anoling, K.M.; Abella CR, G.; Cagatao PP, S.; Bautista, R.G. Critical Perspectives, Theoretical Foundations, Practical Teaching, Technology Integration, Assessment and Feedback, and Hands-on Practices in Science Education. *Am. J. Educ. Res.* **2024**, *12*, 20–27. [[CrossRef](#)]
27. Wicaksono, W.A.; Arifin, I.; Sumarsono, R.B. Implementing a Pesantren-Based Curriculum and Learning Approach to Foster Students' Emotional Intelligence. *Munaddhomah J. Manaj. Pendidik. Islam* **2024**, *5*, 207–221. [[CrossRef](#)]

28. Rencewigg, R.; Joseph, N.P. Enhancing presentation skills: A comparative study of students' performance in virtual and physical classrooms. *Multidiscip. Rev.* **2024**, *7*, 2024156. [[CrossRef](#)]
29. Smith, J. Attending School: A Qualitative Study Exploring Principals' Strategies for Enhancing Attendance. Doctoral Dissertation, Trident University International, Cypress, CA, USA, 2024.
30. Leino, R.K.; Gardner, M.R.; Cartwright, T.; Döring, A.K. Engagement in a virtual learning environment predicts academic achievement in research methods modules: A longitudinal study combining behavioral and self-reported data. *Scholarsh. Teach. Learn. Psychol.* **2024**, *10*, 149. [[CrossRef](#)]
31. Liu, X.; Zhou, J. Short-term wind power forecasting based on multivariate/multi-step LSTM with temporal feature attention mechanism. *Appl. Soft Comput.* **2024**, *150*, 111050. [[CrossRef](#)]
32. Araf, I.; Idri, A.; Chair, I. Cost-sensitive learning for imbalanced medical data: A review. *Artif. Intell. Rev.* **2024**, *57*, 80. [[CrossRef](#)]
33. Ashraf, A.; Naw, N.M.; Shahzad, T.; Aamir, M.; Khan, M.A.; Ouahada, K. Dimension Reduction using Dual-Featured Auto-encoder for the Histological Classification of Human Lungs Tissues. *IEEE Access* **2024**, *12*, 104165–104176. [[CrossRef](#)]
34. Lin, Y.; Wang, D.; Jiang, T.; Kang, A. Assessing Objective Functions in Streamflow Prediction Model Training Based on the Naïve Method. *Water* **2024**, *16*, 777. [[CrossRef](#)]
35. Khan, M.; Anwar, W.; Rasheed, M.; Najeh, T.; Gamil, Y.; Farooq, F. Forecasting the strength of graphene nanoparticles-reinforced cementitious composites using ensemble learning algorithms. *Results Eng.* **2024**, *21*, 101837. [[CrossRef](#)]
36. Chowdhury, M.S. Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting. *Environ. Chall.* **2024**, *14*, 100800. [[CrossRef](#)]
37. Yu, C.; Jin, Y.; Xing, Q.; Zhang, Y.; Guo, S.; Meng, S. Advanced user credit risk prediction model using lightgbm, xgboost and tabnet with smoteenn. *arXiv* **2024**, arXiv:2408.03497.
38. Ge, W.; Coelho, L.M.; Donahue, M.A.; Rice, H.J.; Blacker, D.; Hsu, J.; Newhouse, J.P.; Hernandez-Diaz, S.; Haneuse, S.; Westover, M.B.; et al. Automated identification of fall-related injuries in unstructured clinical notes. *Am. J. Epidemiol.* **2024**, kwae240. [[CrossRef](#)] [[PubMed](#)]
39. Peretz, O.; Koren, M.; Koren, O. Naive Bayes classifier—An ensemble procedure for recall and precision enrichment. *Eng. Appl. Artif. Intell.* **2024**, *136*, 108972. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.