

Article

Deep Learning for Predicting Attrition Rate in Open and Distance Learning (ODL) Institutions

Juliana Ngozi Ndunagu ¹, David Opeoluwa Oyewola ² , Farida Shehu Garki ¹ , Jude Chukwuma Onyeakazi ³ , Christiana Uchenna Ezeanya ¹  and Elochukwu Ukwandu ^{4,*} 

¹ Department of Computer Science and Information Technology, Faculty of Sciences, National Open University of Nigeria, Plot 91, Abuja 900108, Nigeria; jndunagu@noun.edu.ng (J.N.N.); fgarki@noun.edu.ng (F.S.G.); cezeanya@noun.edu.ng (C.U.E.)

² Department of Mathematics and Statistics, Faculty of Science, Federal University Kashere, PMB 0182, Gombe 760001, Nigeria; davidakaprof01@yahoo.com

³ Directorate of General Studies, Federal University of Technology, PMB 1526, Owerri 460114, Nigeria; jude.onyeakazi@futo.edu.ng

⁴ Department of Applied Computing, Cardiff School of Technologies, Cardiff Metropolitan University, 200 Western Avenue, Cardiff CF5 2YB, UK

* Correspondence: eaukwandu@cardiffmet.ac.uk

Abstract: Student enrollment is a vital aspect of educational institutions, encompassing active, registered and graduate students. All the same, some students fail to engage with their studies after admission and drop out along the line; this is known as attrition. The student attrition rate is acknowledged as the most complicated and significant problem facing educational systems and is caused by institutional and non-institutional challenges. In this study, the researchers utilized a dataset obtained from the National Open University of Nigeria (NOUN) from 2012 to 2022, which included comprehensive information about students enrolled in various programs at the university who were inactive and had dropped out. The researchers used deep learning techniques, such as the Long Short-Term Memory (LSTM) model and compared their performance with the One-Dimensional Convolutional Neural Network (1DCNN) model. The results of this study revealed that the LSTM model achieved overall accuracy of 57.29% on the training data, while the 1DCNN model exhibited lower accuracy of 49.91% on the training data. The LSTM indicated a superior correct classification rate compared to the 1DCNN model.

Keywords: long short-term memory; attrition rate; active and inactive student; one-dimensional convolutional neural network; student enrollment; deep learning



Citation: Ndunagu, J.N.; Oyewola, D.O.; Garki, F.S.; Onyeakazi, J.C.; Ezeanya, C.U.; Ukwandu, E. Deep Learning for Predicting Attrition Rate in Open and Distance Learning (ODL) Institutions. *Computers* **2024**, *13*, 229. <https://doi.org/10.3390/computers13090229>

Academic Editors: Yorghos Voutos, Akrivi Krouska, Christos Troussas, Phivos Mylonas and Cleo Sgouropoulou

Received: 20 June 2024

Revised: 6 September 2024

Accepted: 6 September 2024

Published: 11 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

Several authors [1–5] have used machine learning in the prediction of the student attrition rate. Due to the increase in the datasets accumulated in educational systems, deep learning has become a better tool in predicting the student attrition rate [6]. Deep learning (DL) is a subset of Artificial Intelligence (AI) and machine learning (ML) and is increasingly seen as a key technology of the Fourth Industrial Revolution (4IR, or Industry 4.0) [7]. Artificial Neural Networks (ANNs) are the source of DL technology, which has gained popularity in the computing community due to its ability to learn from data. DL has surpassed well-known ML techniques in a few fields, including cybersecurity, natural language processing, bioinformatics, infrared image detectors [8], robotics and control [9], medical information processing [10] and educational systems [7,8]. DL technology, which grew out of ANN, has become important in computing because it can learn from data. The ability to learn enormous volumes of data is one of the benefits of DL [11]. DL is portrayed as an overriding concept or learning approach that affects the entire educational system with respect to student attrition. Student attrition can have different definitions for

different authors; according to Pierrakeas et al. [12], “An attrition is considered as a student who for any reason other than death leaves school before graduation without transferring to another school”. Ashour [13] contends that attrition is associated with circumstances in which students drop out of school or are unable to finish their course of study. Del Bonifro et al. [14] proposed that the term attrition is defined as an increase in the total number of students who started their studies but are unable to finish. Aldowah et al. [15] defined attrition as the total number of students enrolled in a course less the total number of students who finished it.

The student attrition rate is acknowledged as the most complicated and significant problem facing educational systems [16]. The educational system in this perspective comprises conventional and non-conventional institutions. Non-conventional institutions, also known as open and distance learning (ODL), are the concern of this study. Akmal [2] and Park [17] revealed that the average rate of attrition in an ODL system is between 10 and 20% lower than that in conventional institutions and that only 50% of students enrolled in an ODL complete their studies; moreover, Sun et al. [18] affirmed that there is a much higher chance of an attrition rate from non-conventional than from conventional programs. Many studies have affirmed that one of the major problems of ODL institutions is the high rate of student attrition, and it remains a major concern for many educational institutions [3,19–32]. The student attrition rate can be classified into three categories, namely circumstantial, status and cohorts [33]. Hassan [33] defined circumstantial attrition as the kind of attrition which results from all the circumstances or scenarios that are present in the student’s surroundings at a given time and push the student—who might or might not be a dropout—to leave school.

This study presents a model for predicting the student attrition rate at the National Open University of Nigeria (NOUN), an ODL institution with over 500,000 enrolled students. About 22,000 academic records of students who could not finish their program from 2012 to 2022 were analyzed and used for learning. The model was implemented using a variety of DL methods, such as the One-Dimensional Convolutional Neural Network (1DCNN) model and the Long Short-Term Memory Network (LSTM) model, a type of Recurrent Neural Network (RNN). The performance of each model was compared through experiments.

2. Related Works

Deep Neural Networks (DNNs) constitute a subfield of machine learning. This is a model of a network consisting of neurons with multiple parameters and layers between the input and output. DL uses neural network topologies as its basis, and, consequently, these networks are known as DNNs [34]. It provides autonomous learning of the characteristics and their hierarchical representation at multiple levels. In contrast to conventional machine learning approaches, this robustness is due to DL’s powerful process. DL’s whole architecture is used for feature extraction and modification; the early layers perform rudimentary processing of incoming data or learn simple features, and the output is sent to the upper layers, which are responsible for learning complicated features. Hence, DL is suited for handling large datasets and greater complexity.

DL models have an advantage over traditional machine learning models due to the increased number of learning layers and a higher level of abstraction. One more reason for this advantage is that there is direct data-driven learning for all model components. DL is quite data-hungry, given that it includes representation learning [35]. A well-behaved performance model for DL requires an enormous quantity of data, and it requires more data than traditional machine learning algorithms. Machine learning may be utilized with as few as 1000 data points, but DL often needs more data points [36,37]. See Figure 1, which shows a typical architecture of a DL workflow. A summary of other related works is shown in Table 1.

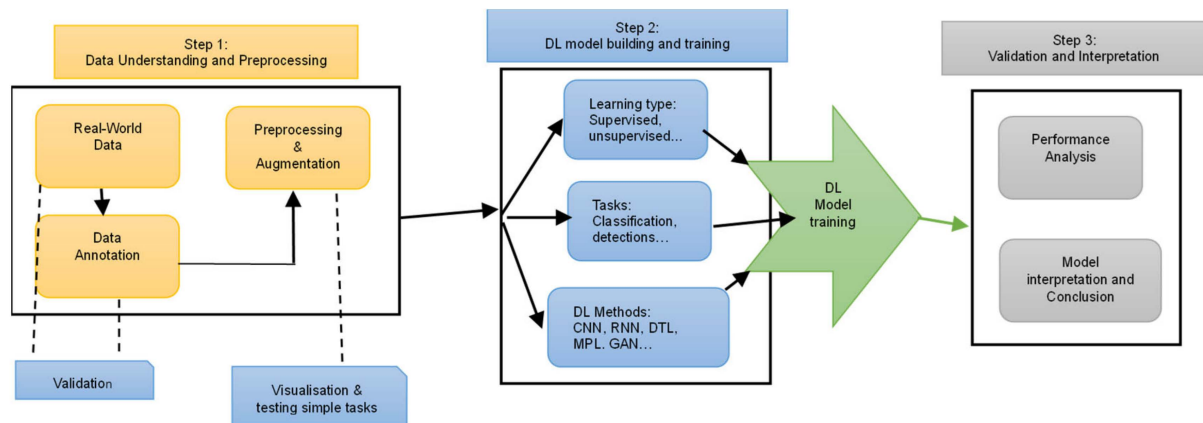


Figure 1. A typical DL workflow to solve real-world problems [7].

Table 1. Summary of related work on student attrition.

S/N	Author	Objective of Study	Classification Techniques Studied	Recommendation of Classification Techniques by Author(s)
1	[38]	This research work demonstrates the effectiveness of this approach in predicting student performance, and ablation feature analysis is conducted to gain insights into the underlying factors that contribute to performance prediction.	Graph Neural Network (GNN) and Convolutional Neural Networks (CNNs),	Graph Neural Network (GNN)
2	[39]	This study used machine learning model in predicting student dropout rates in the Business Informatics BSc course at the Faculty of Finance and Accounting of Budapest Business School using data extracted from the administration system.	Boosted Decision tree	Boosted Decision tree as best suited for predicting attrition rate.
3	[40]	To find how effective the instructor in the higher education systems is, a group of machine and deep learning algorithms were applied to predict instructor performance in higher education systems.	Machine and Deep Learning Techniques	Deep Learning Techniques
4	[41]	The aim of this study was to utilize machine and deep learning models to predict employee attrition with a high accuracy, furthermore, to identify the most influential factors affecting employee attrition.	Machine Learning and Deep Learning Techniques	Deep Learning
5	[42]	Implemented a method to predict student attrition in the upper years of a physiotherapy program with 23.58% males and 17.39% females' population in the attrition student group.	KNN and boosted Decision tree	KNN
6	[43]	The proposed deep neural network model outperforms existing machine learning methods in terms of accuracy, achieving up to 85.4% accuracy	Machine Learning and Deep Neural Network	Deep Neural Network—DNN

Table 1. Cont.

S/N	Author	Objective of Study	Classification Techniques Studied	Recommendation of Classification Techniques by Author(s)
7	[44]	The main goal of this paper is to explore the efficiency of deep learning in the field of EDM, especially in predicting students' academic performance, to identify students at risk of failure	A deep neural network (DNN), decision tree, random forest, gradient boosting, logistic regression, support vector classifier, and K-nearest neighbor	Deep Neural Network—DNN
8	[45]	This study predicts student dropout in two Chilean universities using machine learning models. It focused on finding out variables that trigger first-year engineering student probability of dropout.	KNN, SVM, Decision tree, Random Forest, Gradient-boosting decision tree, Naive Bayes, Logistic regression and neural network.	Gradient-boosting decision trees reports the best model.
9	[46]	This study was focused on students at Abu Dhabi School of Management (ADSM) in the UAE that are at the risk of dropping out.	Decision tree	The use of decision tree has high significance in predicting students at risk of dropping out.
10	[18]	The study mainly analyze the trends of feature processing and the model design in dropout prediction, respectively	Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Graph Neural Networks (GNN) and other deep learning models	Recurrent Neural Networks (RNN)
11	[1]	The model has diverse features which can be utilized to assess how new students will perform and thus contributes to early prediction of student retention and dropout.	BLSTM and CRF deep learning techniques	BLSTM and CRF deep learning techniques.
12	[47]	This study aims to predict students' performance in a specific course as it is continuously running, using the statistic personal biographical information and sequential behavior data with VLE.	simple RNN, GRU, and LSTM.	GRU and simple RNN
13	[48]	This study aimed at predicting student dropout at the Karlsruhe Institute of Technology (KIT)	logistic regressions and decision trees	Decision trees produced slightly better results than logistic regressions
14	[31]	Deep learning algorithms could be applied directly on raw input data, and this could spare the most time-consuming process of feature engineering	SVM, LOGREG or MLP and RNN—Deep learning Technique	RNN
15	[49]	This study explored the relationship between attrition and entrance examination, student place of origin and grades up to the point of abandonment of the major.		
16	[29]	this study proposes to use the deep learning algorithm to construct the dropout prediction model and further produce the predicted individual student dropout probability	K-nearest neighbors (KNN), support vector machines (SVM), decision tree and Deep Learning techniques	Deep Learning techniques

3. Methodology

Data collection was conducted between April 2023 and September 2023 among a population of students who dropped out of their programs in NOUN using a structured questionnaire. A sample size of 22,000 inactive enrolled students was used, and they were contacted through e-mail, phone and personal social media contacts to collect data on the reasons for their inactive status from NOUN between 2012 and 2022. However, only 2000 responses were received, which is 9.1% of the target population. The low percentage of responses could be attributed to emails and short message service (SMS) rebounds and abandoned email addresses. The responses from the dataset were divided into two (2), namely institutional challenges and non-institutional challenges. These datasets went through the preprocessing stages.

3.1. Data Cleaning

The data collected from the sample size underwent a rigorous cleaning process to eliminate errors, inconsistencies and missing values. This pivotal step was essential to ensure the accuracy and reliability of the dataset for subsequent analysis. The cleaning processes involved several intricate procedures, including inputting missing data, rectifying data errors and ensuring data consistency across various sources. Inputting missing data involved filling in any gaps or null values in the dataset with appropriate estimates or values derived from surrounding data points. This helped to maintain the integrity of the dataset and prevent the loss of valuable information. Additionally, correcting data errors entailed identifying and rectifying any inaccuracies or anomalies in the data, such as typographical errors or outliers. This was crucial in ensuring the validity of the data and avoiding any misleading conclusions drawn from the analysis. Overall, the meticulous cleaning processes undertaken on the collected data from the NOUN were instrumental to the high-quality dataset for analysis. It laid the foundation for accurate and reliable insights into factors influencing student attrition rates, ultimately contributing to the advancement of learning analytics in the educational domain.

3.2. Feature Selection

This study focused on predicting student attrition rate at the NOUN, where the target variable (y) represents attrition, or whether a student will drop out or not, while predictor variable (x) encompasses a wide range of factors that could potentially influence a student's decision to drop out, including the perceived difficulty of course materials, frustration with information dissemination processes, lack of social networking opportunities, poor academic performance, communication challenges within families, financial constraints, health issues and other relevant factors.

3.3. Train–Test Set

In the process of building a predictive model, it is essential to split the data into training and testing sets to assess the model's performance accurately. This step ensures that the model's effectiveness is evaluated on data it has not seen during training, helping to gauge its ability and generalize new and unseen data. Typically, this involves allocating approximately 80% of the dataset to the training set, where the model learns patterns and relationships from the data. The remaining 20% is then reserved for the testing set, allowing for an independent evaluation of the model's performance on unseen data. This division helps to prevent overfitting, where the model memorizes the training data but fails to generalize well to new instances. By striking a balance between training and testing data, reliable insights into the model's predictive capabilities can be obtained and used to make informed decisions about its deployment.

The first dataset, which is the institutional challenges, comprises attrition, difficult course materials, frustrating information (NOUN), no social networking, poor academic performance and inadequate communication, while the second dataset, which is non-institutional challenges, comprises attrition, family challenges, financial reasons, sickness

and others, as shown in Tables 2 and 3. The classification basis for Tables 2 and 3 is binary, with attrition classified as either ‘Yes’ (indicating attrition) or ‘No’ (indicating no attrition).

Table 2. Description of institutional challenges in the National Open University of Nigeria.

Column	Data Type
attrition	object
Difficulty-Course-materials	object
Frustration-information-NOUN	object
No-Social-Networking	object
Poor-Academic-Performance	object
Inadequate-Communication	object

Table 3. Description of non-institutional challenges in the National Open University of Nigeria.

Column	Data Type
attrition	object
Family-Challenges	object
Financial-Reasons	object
Sickness	object
Others	object

3.4. Deep Learning Models

3.4.1. Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) structure is a type of RNN [50–53] that is used to solve the issue of vanishing gradients and capture dependencies that last over time in sequential data. It works especially well for operations like machine translation, speech recognition and natural language processing that require sequential or dependent-upon-time information. Networks using LSTM are made up of memory cells with long-term information manipulation and maintenance capabilities. An input gate, a forget gate and an output gate are the three main parts of every memory cell. The data that enter, exit and are contained within the memory cell are managed by these gates. Based on the current input and the previously hidden state, the input gate calculates the amount of new information added to the cell state. Based on the current input and the previous hidden state, the forget gate determines which data from the cell state are to be retained or discarded. Based on the updated cell state and the current input, the output gate controls the cell’s output. The fact that the structure of LSTM mitigates the vanishing gradient problem, a training challenge for deep neural networks, relevant data can be retained over long sequences. By allowing the network to learn when to retain and discard information, the LSTM gates help to preserve contextual data and keep valuable trends from being lost.

3.4.2. One-Dimensional Convolutional Neural Network (1DCNN)

1DCNN is a kind of neural network structure that is frequently used for processing and analyzing data in sequence. 1DCNN is made to deal with one-dimensional sequences, like historical data or text data, as opposed to traditional Convolutional Neural Networks (CNNs), which work with two-dimensional image data [54–58]. The key component of a 1DCNN is the convolutional layer. The convolution operation involves applying a set of filters to the input sequence, extracting local patterns or features. These filters, also known as kernels, slide across the input sequence and perform element-wise multiplications and summations. The resulting output is a feature map that represents the learned features at different positions along the sequence. In a 1DCNN, multiple convolutional filters can be

applied simultaneously, each capturing different patterns or features within the sequence. This allows the network to learn complex patterns and hierarchies of features at different levels of abstraction. After the convolutional layers, other typical layers such as pooling, dropout and fully connected layers can be added to further process the features and extract higher-level representations. By down sampling the feature maps, pooling layers minimize their dimensionality while maintaining the most crucial data. By arbitrarily turning off a portion of neurons while training, dropout layers aid in preventing overfitting. After combining the acquired attributes, layers that are fully linked make forecasts using the final representations.

3.5. Proposed Model

The researchers conducted a comprehensive investigation on the factors influencing students' attrition rate at the NOUN using DL techniques. Specifically, this study aims at comparing the performance of two leading DL architectures, namely LSTM and 1DCNN, with the objective of determining the most effective model in predicting and understanding student attrition patterns. The LSTM and 1DCNN were employed as a robust tool to delve into the intricate web of factors contributing to the student attrition rate to identify the most effective tool in predicting it. This strategic choice ensured a focused and nuanced exploration of temporal dependencies, sequential patterns and contextual features that play a pivotal role in students' decisions to discontinue their studies at NOUN. The goal of this research is to provide actionable insights that inform the development of targeted interventions and support mechanisms, fostering a more inclusive and supportive learning environment. By employing the best-performing DL model, we aspire to contribute significantly to the ongoing efforts aimed at enhancing student success and retention in the distinctive landscape of open and distance learning at NOUN.

4. Results and Discussion

In Figure 2, the pie chart shows that 93.5% of students are ready to come back to NOUN if recalled, while only 6.5% are not ready to come back and finish their program, even if recalled. This suggests that most reasons for NOUN student's attrition are circumstantial and agree with [33], who stated that circumstantial attrition results from all the circumstances or scenarios that are present in their surroundings at a given time and push the student—who might or might not be a dropout—to leave school.

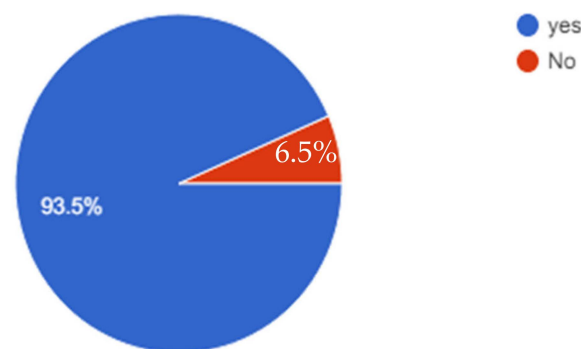


Figure 2. Students' readiness to complete their study.

Institutional Challenge Dataset—Course Materials

Figure 3 shows the difficulty in understanding NOUN course materials. Thus, 58.9% of the students find it difficult to understand NOUN course materials, while 41.1% of students do not find it difficult.

Figure 4 shows the percentage of students who abandoned the program due to frustration and delay in obtaining information from NOUN. Further, 77.3% indicates the percentage of students who were frustrated, while 22.7% comprised students who abandoned their program but not because of frustration.

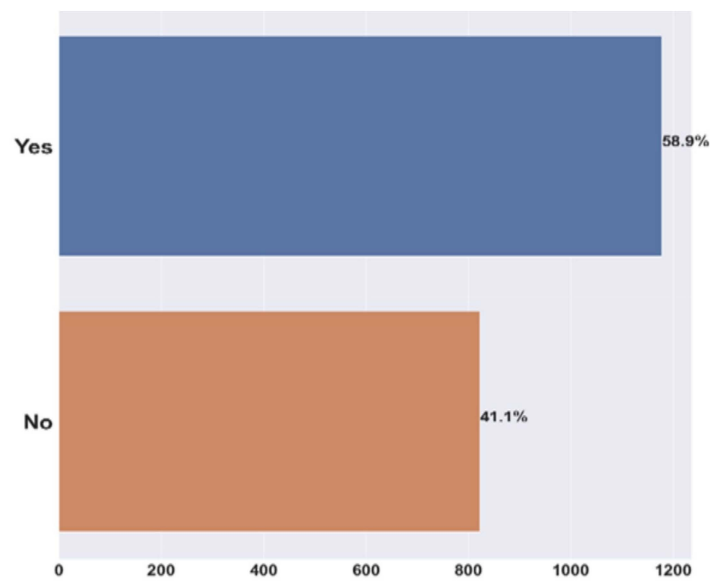


Figure 3. Difficulty in understanding the course material.

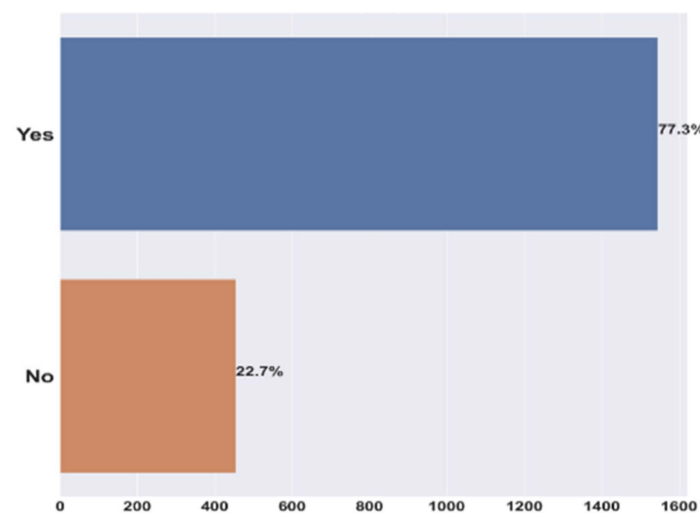


Figure 4. Institutional challenge dataset: frustration in accessing information from NOUN.

Figure 5 shows that 51.7% abandoned their study due to lack of the use of a social network platform to communicate with their departmental course mates, while 48.3% was not due to lack of the use of a social network platform. This suggests that social networking platforms are a popular way for students to stay connected with each other and share information.

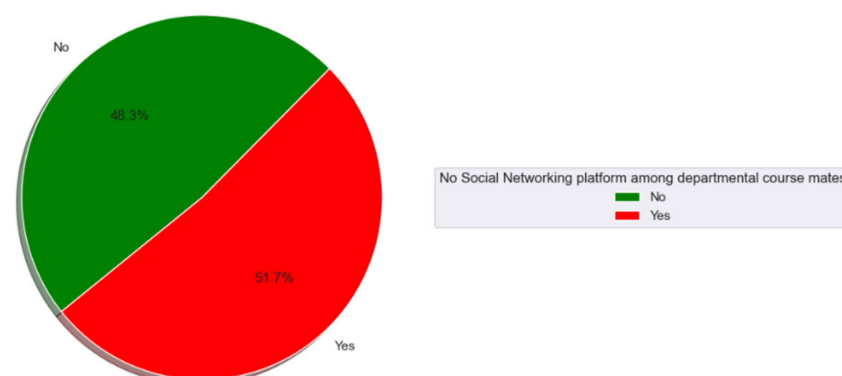


Figure 5. Institutional challenge dataset—use of social network platforms.

Figure 6 shows that 48.8% of the students abandoned their studies due to poor academic performance, while 51.2% of the students were not due to poor academic performance.

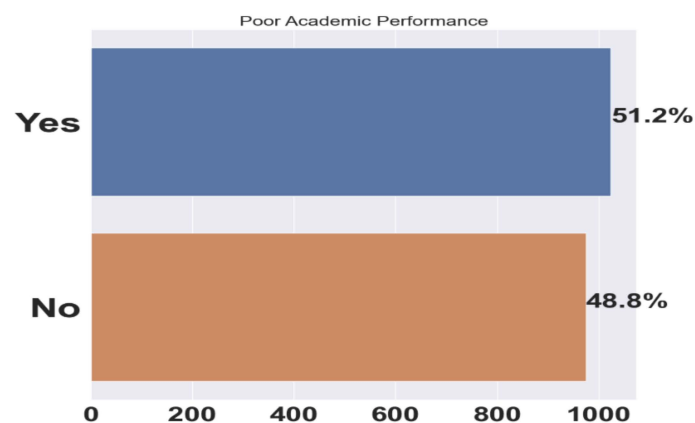


Figure 6. Institutional challenge dataset—academic performance.

Figure 7 indicates that 61.5% of the students who abandoned their program did so due to inadequate communication skills with the university, while 38.5% of NOUN students who abandoned their programs did not do so due to inadequate communication skill with the institution.

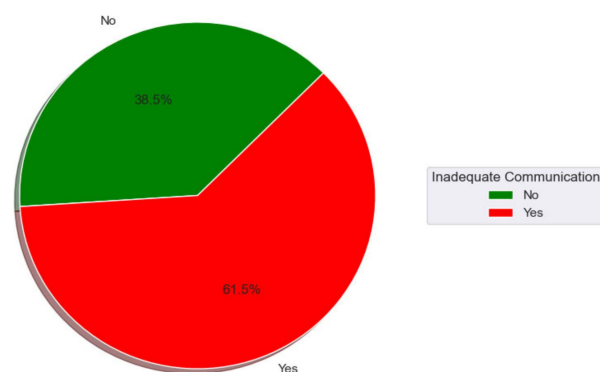


Figure 7. Institutional challenge dataset—inadequate communication.

Figure 8 shows that 47.8% of the students failed to complete their studies in NOUN due to family challenges, while 52.2% of NOUN students did not complete their studies due to other reasons.

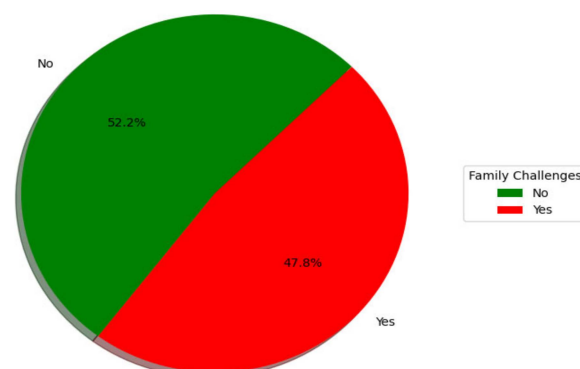


Figure 8. Non-institutional challenge dataset—family challenges.

Figure 9 indicates that 58.2% of NOUN students abandoned their studies due to financial reasons, while 41.8% was due to other reasons.

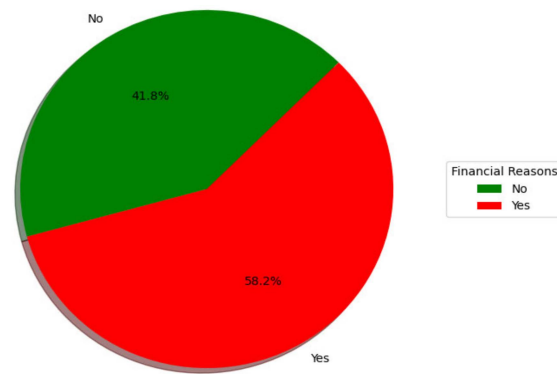


Figure 9. Non-institutional challenge dataset—financial reasons.

Figure 10 indicates that 44.9% of NOUN students failed to complete their studies on health grounds, while 55.1% was due to other reasons.

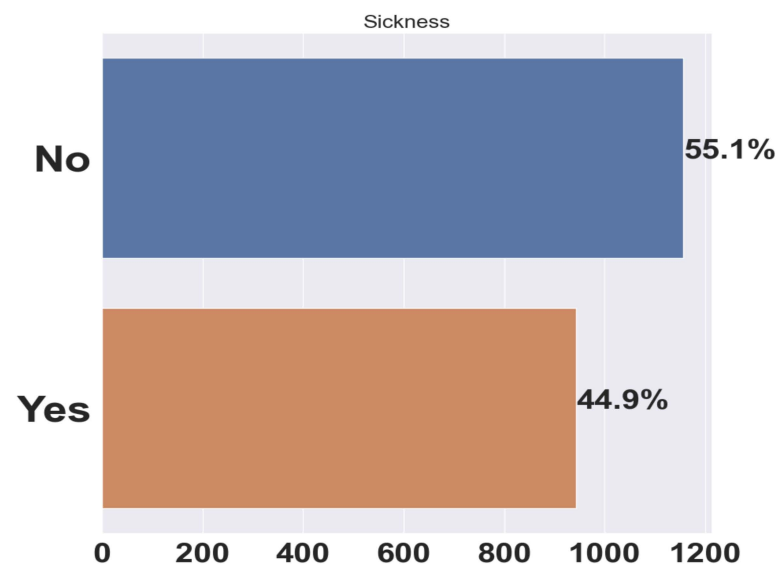


Figure 10. Non-institutional dataset—sickness.

Figure 11 shows a correlation matrix of the factors affecting students' attrition. The Pearson correlation coefficient was reflected on every pair of factors. There are three possible correlation coefficients: -1 for a perfect negative correlation, 0 for no correlation, and 1 for a perfect positive correlation. The strongest positive correlation is between poor academic performance and no social networking ($r = 0.14$). This means that students who had poor academic performance were more likely to drop out of the course due to social networks. Other significant positive correlations include difficult course materials and no social networking ($r = 0.11$). This means that students who found the course materials difficult or who were frustrated with not knowing about social networking provided by the university were more likely to abandon their studies.

Figure 12 shows a correlation matrix between different reasons for attrition for non-institutional challenges. The values in the cells represent the correlation coefficient between two reasons. A positive correlation coefficient indicates that two reasons are positively correlated, meaning that if one reason increases, the other reason is also likely to increase. A negative correlation coefficient indicates that two reasons are negatively correlated, meaning that if one reason increases, the other reason is likely to decrease. The strongest positive correlation is between financial reasons and others (0.012). This means that if a student is facing financial reasons for attrition, they are also likely to be facing other reasons for attrition. The strongest negative correlation is between attrition and family challenges (-0.0005). This means that if a student is facing family challenges, they are less likely to be facing attrition.

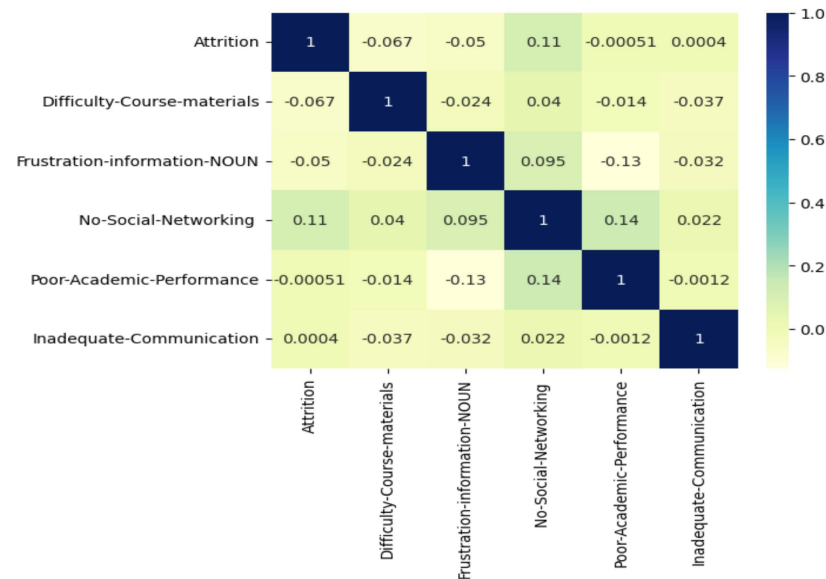


Figure 11. Correlation matrix between different factors that may affect student attrition in the institutional challenge dataset.



Figure 12. Correlation matrix between different factors that may affect student attrition in non-institutional challenge dataset.

The LSTM model's accuracy during training and validation is displayed in Figure 13. The percentage of correctly classified examples in the training set is known as the training accuracy, and the percentage of correctly classified examples in the validation set is known as the validation accuracy. The model is trained on a dataset of 80%, with a validation set of 20%. When the model understands the fundamental characteristics of the data, the training correctness rises quickly. The validation accuracy rises as well but more slowly, because the validation set is a more complex dataset than the training set. The model eventually reaches a point where the training accuracy and validation accuracy are both about 57%. This means that the model can correctly classify about 57% of the examples in both the training and validation sets. Additionally, the figure demonstrates that both the training and validation accuracies are greater than 55%, indicating that the model can accurately classify over 55% of the attrition dataset in both the training and validation sets. This is a positive outcome, demonstrating the model's capacity to learn the characteristics of the data and generate predictions.

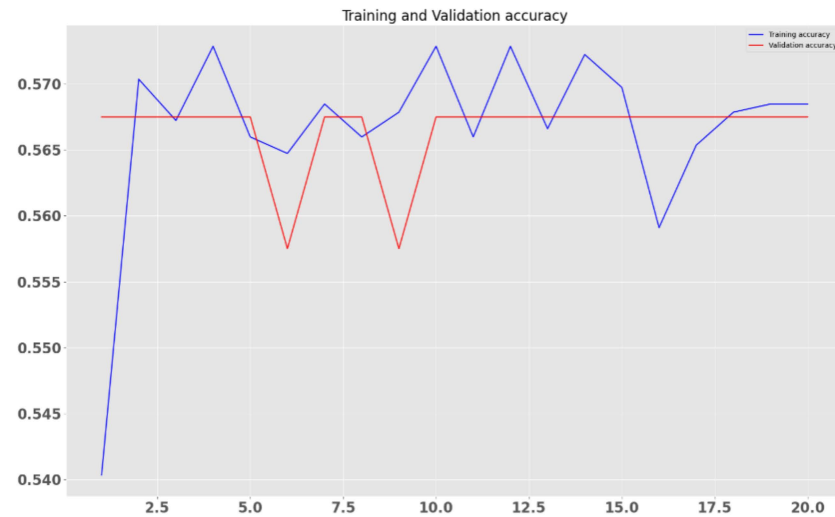


Figure 13. Training and validation accuracy of LSTM model in the institutional challenge dataset.

The LSTM model's training and validation losses are depicted in Figure 14. The model's loss in training data is referred to as the training loss, and the model's loss in validation data is referred to as the validation loss. Because the model can be trained from the training data, the training loss is usually smaller than the validation loss. As the total amount of epochs grows in this figure, the training loss and validation loss decline. This suggests that the model is gaining knowledge from the data and improving its forecasting abilities.



Figure 14. Training and validation loss of LSTM model in the institutional challenge dataset.

The CNN model's reliability during training and validation is displayed in Figure 15. The predictive ability of the model on training data is known as training accuracy, and the model's accuracy on validation data is known as validation accuracy. Because the model can learn the training data more quickly than the validation data, the training accuracy is usually higher than the validation accuracy. The generalization gap is the difference in accuracy between training and validation data. The success rate for training in this study is approximately 50.5%, and the validation accuracy is approximately 49.9%. This implies that while the model is highly adept at learning the training set, it struggles to make strong generalizations to untested data. The absence of information or overfitting is possibly because of this.



Figure 15. Training and validation accuracy of CNN model in the institutional challenge dataset.

Figure 16 shows the training and validation loss of a CNN model. “Training loss” describes the model’s loss in training data, whereas “validation loss” refers to the model’s loss in validation data. Given that the model can be trained using the training data, the training loss is usually smaller than the validation loss. In Figure 16, as the number of epochs increases, both the validation loss and the training loss decline. This shows that the model is not overfitting and instead learning from the data. The model reaches its lowest validation loss at around 12 epochs. After this point, the validation loss starts to increase, which indicates that the model is starting to overfit to the training data.



Figure 16. Training and validation loss of CNN model in the institutional challenge dataset.

The LSTM model’s accuracy during training and validation is displayed in Figure 17. The model’s accuracy on training data is known as training accuracy, and the model’s accuracy on validation data is known as validation accuracy. Because the model can learn the training data more quickly than the validation data, the training accuracy is usually higher than the validation accuracy. Since the validation accuracy is unaffected by the model’s overfitting to the training set, it provides a more accurate assessment of the model’s performance. About 51% of the training and 46% of the validation accuracy are found in this study.

The LSTM model’s training and validation losses are shown in Figure 18. The model’s loss on training data is referred to as the training loss, and the model’s loss on validation data is referred to as the validation loss. The goal of training a neural network model is to minimize the training loss and the validation loss. Figure 18 illustrates how, as the number of epochs rises, both the training loss and the validation loss decline. This indicates that while the model continues to be trained, it is also learning. Since the training data are only a portion of the validation data, it is expected that the training loss remains lower than the validation loss. The percentage shown in Figure 18 is the validation accuracy, which is the accuracy of the model

on the validation data. The validation accuracy increases as the number of epochs increases, which means that the model is learning to classify the data correctly.

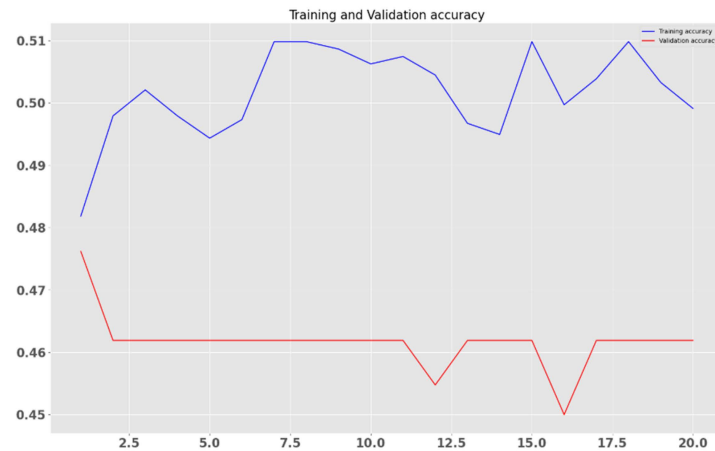


Figure 17. Training and validation accuracy of LSTM model in the non-institutional challenge dataset.



Figure 18. Training and validation loss of LSTM model in the non-institutional challenge dataset.

The CNN model's accuracy during training and validation is displayed in Figure 19. The model's accuracy on training data is known as training accuracy, and the model's accuracy on validation data is known as validation accuracy. Because the validation data are usually harder to analyze than the training data, the validation accuracy is typically lower than the training accuracy. About 50.5% the training and 49.9% of validation accuracy are found in this study.

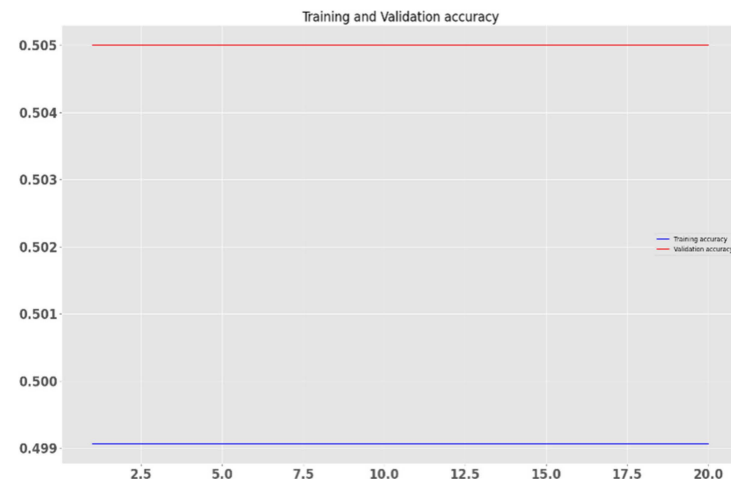


Figure 19. Training and validation accuracy of CNN model in the non-institutional challenge dataset.

The training and validation loss of a CNN model is displayed in Figure 20. The model's loss on training data is referred to as the training loss, and the model's loss on validation data is referred to as the validation loss. Because the model can be trained from the training data, the training loss is usually smaller than the validation loss. In this study, as the number of epochs rises, both the training loss and validation loss decline. This suggests that the model is effectively generalizing to freshly collect data and learning from the data.

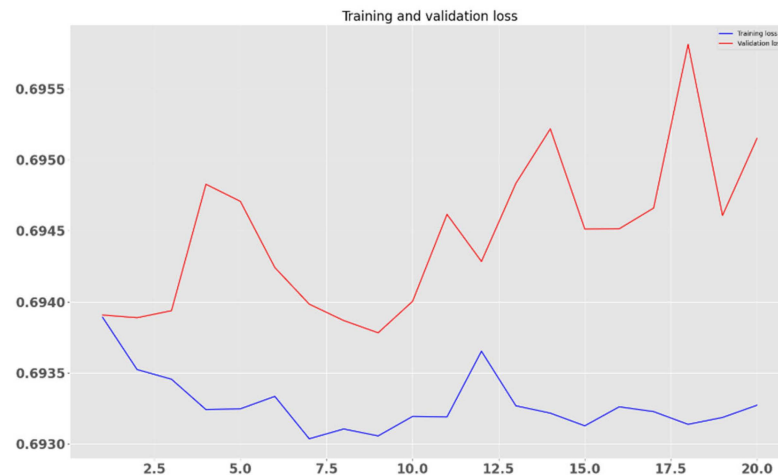


Figure 20. Training and validation loss of CNN model in the non-institutional challenge dataset.

In Table 4, the training and test accuracies of the LSTM and CNN models in the institutional challenge dataset are presented. The LSTM model exhibits a training accuracy of 57.29% and a test accuracy of 56.75%, indicating its ability to predict labels in both seen and unseen data. In comparison, the CNN model achieves a training accuracy of 49.91% and a test accuracy of 50.50%. Notably, the LSTM model outperforms the CNN model on both training and test datasets, showcasing its superior learning and generalization capabilities in this specific context. This agrees with [32], who demonstrated that LSTM performs better than the other model. Consideration of dataset characteristics and further evaluation metrics can provide a nuanced understanding of model performance.

Table 4. Training and test accuracy of LSTM and CNN in the institutional challenge dataset.

MODEL	Training (%)	Test (%)
LSTM	57.29	56.75
CNN	49.91	50.50

Table 5 presents the training and test loss values for the LSTM and CNN models in the institutional challenge dataset. The LSTM model achieved a training loss of 0.6765 and a test loss of 0.6852, while the CNN model showed slightly lower loss values with a training loss of 0.6730 and a test loss of 0.6782. Lower loss values indicate reduced discrepancies between predicted and actual labels, showcasing effective error minimization by both models during training and evaluation. The marginal superiority of the CNN model in minimizing losses suggests its slightly enhanced performance in handling the dataset's intricacies. These loss metrics contribute valuable insights to the models' efficacy and generalization capabilities on the institutional challenge dataset.

Table 5. Training and test loss of LSTM and CNN in the institutional challenge dataset.

MODEL	Training (%)	Test (%)
LSTM	0.6765	0.6852
CNN	0.6730	0.6782

In Table 6, the training and test accuracies of the LSTM and CNN models in the non-institutional challenge dataset are presented. Both models exhibit identical training and test accuracy values, with the LSTM and CNN models achieving 50.98% accuracy during training and 46.19% accuracy on the test dataset. This symmetry in performance suggests that, in the context of the non-institutional challenge dataset, neither model outperforms the other, yielding similar predictive capabilities. The comparable accuracies indicate that both the LSTM and CNN models have similar effectiveness in learning patterns and making predictions on unseen data, emphasizing the importance of dataset characteristics and the suitability of model architectures in achieving satisfactory performance in this specific non-institutional challenge scenario.

Table 6. Training and test accuracy of LSTM and CNN in the non-institutional challenge dataset.

MODEL	Training (%)	Test (%)
LSTM	50.98	46.19
CNN	50.98	46.19

In Table 7, the training and test loss values for the LSTM and CNN models in the non-institutional challenge dataset are presented. Both models exhibit closely aligned loss metrics, with the LSTM model showing a training loss of 0.6928 and a test loss of 0.6953, while the CNN model records a training loss of 0.6929 and a test loss of 0.6951. The minimal disparity in loss values underscores the similar performance of the two models in minimizing errors during both training and evaluation on the non-institutional challenge dataset. These comparable loss metrics indicate that neither the LSTM nor CNN model demonstrates a significant advantage in error reduction, highlighting a balance in their capacity to capture patterns and make predictions on the non-institutional challenge dataset. The marginal differences in loss values emphasize the nuanced comparison of these models' efficacy in handling the specific characteristics of the non-institutional challenge scenario.

Table 7. Training and test loss of LSTM and CNN in the non-institutional challenge dataset.

MODEL	Training (%)	Test (%)
LSTM	0.6928	0.6953
CNN	0.6929	0.6951

5. Conclusions

In conclusion, this study investigated the factors contributing to attrition within two distinct datasets: institutional challenges and non-institutional challenges. The findings shed light on the predictive performance of the LSTM and CNN models in both institutional and non-institutional challenge scenarios. Notably, in the institutional challenge dataset, the LSTM model exhibited superior accuracy compared to the CNN model, while in the non-institutional challenge dataset, both models demonstrated comparable accuracy and loss values. These insights emphasize the nuanced interplay between various factors contributing to attrition and underscore the importance of tailoring predictive models to the specific challenges inherent in institutional and non-institutional contexts. The authors believe that LSTM outperformed 1DCNN in this study because LSTMs are specifically designed to handle sequential data, such as time-series data, speech, text or video. They can learn long-term dependencies and patterns in data. Furthermore, LSTMs can also learn dependencies that span hundreds or even thousands of time steps, making them ideal for tasks like language modeling or speech recognition. LSTMs are designed to mitigate the vanishing gradient problem, which can hinder training in traditional RNNs.

5.1. Limitations of This Study

The generalizability of the findings may be constrained by the specific characteristics of the datasets and the challenges considered. The reliance on deep learning models implies that the predictive accuracy is contingent upon the quality and representativeness of the training data. Furthermore, the absence of temporal dynamics in the dataset may limit the models' ability to capture changes in attrition patterns over time. Additionally, while the models show predictive potential, their interpretability remains a challenge, and further research is warranted to enhance the transparency of deep learning models in educational contexts. Despite these limitations, this study contributes valuable insights into attrition prediction, paving the way for more refined models and interventions aimed at mitigating attrition challenges in diverse educational settings.

5.2. Recommendation for Further Studies

Further research and refinement of models may provide deeper insights into addressing attrition challenges in diverse educational settings. In considering future research directions, this study lays the foundation for further exploration into attrition factors within educational contexts. First, expanding the scope of the datasets by incorporating additional variables and collecting data from a broader range of institutions could enhance the models' predictive capabilities. Additionally, examining the interplay between institutional and non-institutional challenges in attrition prediction could provide a more comprehensive understanding of student outcomes. Future research endeavors might also involve the incorporation of qualitative data, such as using interviews or focus groups, to capture nuanced aspects contributing to attrition that quantitative measures alone may not fully capture.

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

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. This is because the primary data were deposited in the institutional internal data repository, and their release is subject to approval by the requisite institutional authority.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Uliyan, D.; Aljaloud, A.S.; Alkhalil, A.; Al Amer, H.S.; Mohamed, M.A.E.A.; Alogali, A.F.M. Deep Learning Model to Predict Students Retention Using BLSTM and CRF. *IEEE Access* **2021**, *9*, 135550–135558. [\[CrossRef\]](#)
2. Akmal, A. Predicting Dropout on E-Learning Using Machine Learning. *J. Appl. Data Sci.* **2020**, *1*, 29–34. [\[CrossRef\]](#)
3. Alamri, A.; Sun, Z.; Cristea, A.I.; Senthilnathan, G.; Shi, L.; Stewart, C. Is mooc learning different for dropouts? a visually driven, multi-granularity explanatory ml approach. In *International Conference on Intelligent Tutoring Systems*; Springer: Cham, Switzerland, 2020; pp. 353–363.
4. Lainjo, B. Mitigating Academic Institution Dropout Rates with Predictive Analytics Algorithms. *Int. J. Educ. Teach. Soc. Sci.* **2023**, *3*, 29–49. [\[CrossRef\]](#)
5. López-Pernas, S.; Kleimola, R.; Väisänen, S.; Hirsto, L. Early detection of dropout factors in vocational education: A large-scale case study from Finland. In *Proceedings of the Finnish Learning Analytics and Artificial Intelligence in Education Conference (FLAIEC22)*, Joensuu, Finland, 29–30 September 2022; pp. 44–50.
6. Akour, M.; Alsghaier, H.; Al Qasem, O. The effectiveness of using deep learning algorithms in predicting students' achievements. *Indones. J. Electr. Eng. Comput. Sci.* **2020**, *19*, 387–393. [\[CrossRef\]](#)
7. Taye, M.M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers* **2023**, *12*, 91. [\[CrossRef\]](#)
8. Zhang, X.; Cao, Z. A framework of an intelligent education system for higher education based on deep learning. *Int. J. Emerg. Technol. Learn. (Online)* **2021**, *16*, 233. [\[CrossRef\]](#)

9. Aggarwal, K.; Singh, S.K.; Chopra, M.; Kumar, S.; Colace, F. Deep learning in robotics for strengthening industry 4.0.: Opportunities, challenges and future directions. In *Robotics and AI for Cybersecurity and Critical Infrastructure in Smart Cities*; Springer: Cham, Switzerland, 2022; pp. 1–19.
10. Suganyadevi, S.; Seethalakshmi, V.; Balasamy, K. A review on deep learning in medical image analysis. *Int. J. Multimed. Inf. Retr.* **2022**, *11*, 19–38. [\[CrossRef\]](#)
11. Dargan, S.; Kumar, M.; Ayyagari, M.R.; Kumar, G. A survey of deep learning and its applications: A new paradigm to machine learning. *Arch. Comput. Methods Eng.* **2020**, *27*, 1071–1092. [\[CrossRef\]](#)
12. Pierrakeas, C.; Koutsonikos, G.; Lipitakis, A.D.; Kotsiantis, S.; Xenos, M.; Gravvanis, G.A. The variability of the reasons for student dropout in distance learning and the prediction of dropout-prone students. In *Machine Learning Paradigms: Advances in Learning Analytics*; The Variability of the Reasons for Student Dropout in Distance Learning and the Prediction of Dropout-Prone Students; Springer: Cham, Switzerland, 2020; pp. 91–111.
13. Ashour, S. Analysis of the attrition phenomenon through the lens of university dropouts in the United Arab Emirates. *J. Appl. Res. High. Educ.* **2020**, *12*, 357–374. [\[CrossRef\]](#)
14. Del Bonifro, F.; Gabrielli, M.; Lisanti, G.; Zingaro, S.P. Student dropout prediction. In Proceedings of the Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, 6–10 July 2020; Proceedings, Part I 21. Springer International Publishing: New York, NY, USA, 2020; pp. 129–140. [\[CrossRef\]](#)
15. Aldowah, H.; Al-Samarraie, H.; Alzahrani, A.I.; Alalwan, N. Factors affecting student dropout in MOOCs: A cause and effect decision-making model. *J. Comput. High. Educ.* **2020**, *32*, 429–454. [\[CrossRef\]](#)
16. Kim, D.; Kim, S. Sustainable education: Analyzing the determinants of university student dropout by nonlinear panel data models. *Sustainability* **2018**, *10*, 954. [\[CrossRef\]](#)
17. Park, C. Development of prediction model to improve dropout of cyber university. *J. Korea Acad.-Ind. Coop. Soc.* **2020**, *21*, 380–390.
18. Sun, Z.; Harit, A.; Yu, J.; Cristea, A.I.; Shi, L. A Brief Survey of Deep Learning Approaches for Learning Analytics on MOOCs. In Proceedings of the Intelligent Tutoring Systems: 17th International Conference, ITS 2021, Virtual Event, 7–11 June 2021; pp. 1–11.
19. Dake, D.K.; Buabeng-Andoh, C. Using Machine Learning Techniques to Predict Learner Drop-out Rate in Higher Educational Institutions. Edited by Robin Singh Bhadoria. *Mob. Inf. Syst.* **2022**, *2022*, 2670562. [\[CrossRef\]](#)
20. Yilmaz, B.A.; Karataş, S. Why do open and distance education students drop out? Views from various stakeholders. *Int. J. Educ. Technol. High. Educ.* **2022**, *19*, 28. [\[CrossRef\]](#)
21. Queiroga, E.M.; Lopes, J.L.; Kappel, K.; Aguiar, M.; Araújo, R.M.; Munoz, R.; Villarroel, R.; Cechinel, C. A Learning Analytics Approach to Identify Students at Risk of Dropout: A Case Study with a Technical Distance Education Course. *Appl. Sci.* **2020**, *10*, 3998. [\[CrossRef\]](#)
22. Obaido, G.; Agbo, F.J.; Alvarado, C.; Oyelere, S.S. Analysis of attrition studies within the computer sciences. *IEEE Access* **2023**, *11*, 53736–53748. [\[CrossRef\]](#)
23. Stephenson, C.; Miller, A.D.; Alvarado, C.; Barker, L.; Barr, V.; Camp, T.; Frieze, C.; Lewis, C.; Mindell, E.C.; Limbird, L.; et al. *Retention in Computer Science Undergraduate Programs in the Us: Data Challenges and Promising Interventions*; ACM: New York, NY, USA, 2018.
24. Mubarak, A.A.; Cao, H.; Hezam, I.M. Deep Analytic Model for Student Dropout Prediction in Massive Open Online Courses. *Comput. Electr. Eng.* **2021**, *93*, 107271. [\[CrossRef\]](#)
25. Bañeres, D.; Rodríguez, M.E.; Guerrero-Roldán, A.E.; Karadeniz, A. An early warning system to detect at-risk students in online higher education. *Appl. Sci.* **2020**, *10*, 4427. [\[CrossRef\]](#)
26. Isidro, C.; Carro, R.M.; Ortigosa, A. Dropout detection in MOOCs: An exploratory analysis. In Proceedings of the 2018 International Symposium on Computers in Education (SIIE), Cádiz, Spain, 19–21 September 2018; pp. 1–6. [\[CrossRef\]](#)
27. Raga, R.C.; Raga, J.D. Early prediction of student performance in blended learning courses using deep neural networks. In Proceedings of the 2019 International Symposium on Educational Technology (ISET), Hradec Králové, Czech Republic, 2–4 July 2019; pp. 39–43.
28. Ndunagu, J.; Ukwandu, E.; Ezema, M.; Olebara, C.; Ugwu, M.; Igwe, J.; Oyewola, D.; Fapohunda, S. Future of remote learning: The virtual laboratory perspective. *Univ. Ib. J. Sci. Log. ICT Res.* **2022**, *8*, 55–68.
29. Xing, W.; Du, D. Dropout prediction in MOOCs: Using deep learning for personalized intervention. *J. Educ. Comput. Res.* **2019**, *57*, 547–570. [\[CrossRef\]](#)
30. Cristea, A.I.; Alamri, A.; Kayama, M.; Stewart, C.; Alsheri, M.; Shi, L. Earliest predictor of dropout in moocs: A longitudinal study of futurelearn courses. In *Information Systems Development: Designing Digitalization*; Association for Information Systems: Atlanta, GA, USA, 2018.
31. Körösi, G.; Farkas, R. Mooc performance prediction by deep learning from raw clickstream data. In *International Conference on Advances in Computing and Data Sciences*; Springer: Singapore, 2020; pp. 474–485.
32. Li, K.C.; Wong, B.T. Factors related to student persistence in open universities: Changes over the years. *Int. Rev. Res. Open Distrib. Learn.* **2019**, *20*, 132–151.
33. Hassan, W.; Thalh, N.P.; Shar, M.M.; Aqeel, C.; Soomro, U.; Fatima, S.S. Causes of Students' Dropout at Elementary Level in Pakistan. *Palarch's J. Archaeol. Egypt/Egyptol.* **2020**, *17*, 4308–4325.
34. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117.
35. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* **2014**, *4*, 3320–3328.
36. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)

37. Bengio, Y. Learning Deep Architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–127. [\[CrossRef\]](#)
38. Kannan, K.R.; Abarna, K.T.M.; Vairachilai, S. Graph Neural Networks for Predicting Student Performance: A Deep Learning Approach for Academic Success Forecasting. *Int. J. Intell. Syst. Appl. Eng.* **2024**, *12*, 228–232.
39. Fauszt, T.; Erdélyi, K.; Dobák, D.; Bognár, L.; Kovács, E. Design of a Machine Learning Model to Predict Student Attrition. *Int. J. Emerg. Technol. Learn.* **2023**, *18*, 184. [\[CrossRef\]](#)
40. Abunasser, B.S.; AL-Hiealy, M.R.; Barhoom, A.M.; Almasri, A.R.; Abu-Naser, S.S. Prediction of Instructor Performance using Machine and Deep Learning Techniques. (*IJACSA*) *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 78–83. [\[CrossRef\]](#)
41. Arqawi, S.A.; Rumman, M.A.; Zitawi, E.A.; Rabaya, A.H.; Sadaqa, A.S.; Abunasser, B.S.; Abu-Naser, S.S. Predicting Employee Attrition and Performance Using Deep Learning. *J. Theor. Appl. Inf. Technol.* **2022**, *100*, 6526–6536.
42. Barramuño, M.; Meza-Narváez, C.; Gálvez-García, G. Prediction of student attrition risk using machine learning. *J. Appl. Res. High. Educ.* **2022**, *14*, 974–986. [\[CrossRef\]](#)
43. Lokhande, D.B.; Chavan, T.A.; Patil, D.P. Deep neural network in prediction of student performance. *Gradiva Rev. J.* **2022**, *8*, 811–816.
44. Nabil, A.; Seyam, M.; Abou-Elfetouh, A. Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. *IEEE Access* **2021**, *9*, 140731–140746. [\[CrossRef\]](#)
45. Opazo, D.; Moreno, S.; Álvarez-Miranda, E.; Pereira, J. Analysis of first-year university student dropout through machine learning models: A comparison between universities. *Mathematics* **2021**, *9*, 2599. [\[CrossRef\]](#)
46. Shilbayeh, S.; Abonamah, A. Predicting student enrollments and attrition patterns in higher educational institutions using machine learning. *Int. Arab J. Inf. Technol.* **2021**, *18*, 562–567.
47. He, Y.; Chen, R.; Li, X.; Hao, C.; Liu, S.; Zhang, G.; Jiang, B. Online At-Risk Student Identification Using RNN-GRU Joint Neural Networks. *Information* **2020**, *11*, 474. [\[CrossRef\]](#)
48. Kemper, L.; Vorhoff, G.; Wigger, B.U. Predicting student dropout: A machine learning approach. *Eur. J. High. Educ.* **2020**, *10*, 28–47. [\[CrossRef\]](#)
49. Aguilar-Gonzalez, S.; Palafox, L. Prediction of Student Attrition Using Machine Learning. In *Advances in Soft Computing, Proceedings of the 18th Mexican International Conference on Artificial Intelligence, MICAI 2019, Xalapa, Mexico, 27 October–2 November 2019*; Proceedings 18; Springer International Publishing: Cham, Switzerland, 2019; pp. 212–222.
50. Oyewola, D.O.; Dada, E.G.; Omotehinwa, T.O.; Emebo, O.; Oluwagbemi, O.O. Application of Deep Learning Techniques and Bayesian Optimization with Tree Parzen Estimator in the Classification of Supply Chain Pricing Datasets of Health Medications. *Appl. Sci.* **2022**, *12*, 10166. [\[CrossRef\]](#)
51. Oyewola, D.O.; Oladimeji, L.A.; Julius, S.O.; Kachalla, L.B.; Dada, E.G. Optimizing Sentiment Analysis of Nigerian 2023 Presidential Election Using Two-Stage Residual Long Short Term Memory. *Heliyon* **2023**, *9*, e14836. [\[CrossRef\]](#)
52. Sabri, M.; El Bouhdidi, J.; Chkouri, M.Y. *A Proposal for a Deep Learning Model to Enhance Student Guidance and Reduce Dropout BT—Artificial Intelligence and Industrial Applications*; Masrour, T., El Hassani, I., Cherrafi, A., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 158–165.
53. Moghar, A.; Hamiche, M. Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Comput. Sci.* **2020**, *170*, 1168–1173. [\[CrossRef\]](#)
54. Qazi, E.U.; Almorjan, A.; Zia, T. A One-Dimensional Convolutional Neural Network (1D-CNN) Based Deep Learning System for Network Intrusion Detection. *Appl. Sci.* **2022**, *12*, 7986. [\[CrossRef\]](#)
55. Wu, C.; Jiang, P.; Ding, C.; Feng, F.; Chen, T. Intelligent Fault Diagnosis of Rotating Machinery Based on One-Dimensional Convolutional Neural Network. *Comput. Ind.* **2019**, *108*, 53–61. [\[CrossRef\]](#)
56. Harbola, S.; Coors, V. One Dimensional Convolutional Neural Network Architectures for Wind Prediction. *Energy Convers. Manag.* **2019**, *195*, 70–75. [\[CrossRef\]](#)
57. Sharma, Y.; Singh, B.K. One-Dimensional Convolutional Neural Network and Hybrid Deep-Learning Paradigm for Classification of Specific Language Impaired Children Using Their Speech. *Comput. Methods Programs Biomed.* **2022**, *213*, 106487. [\[CrossRef\]](#)
58. Moitra, D.; Mandal, R.K. Classification of Non-Small Cell Lung Cancer Using One-Dimensional Convolutional Neural Network. *Expert Syst. Appl.* **2020**, *159*, 113564. [\[CrossRef\]](#)

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.