JOURNAL OF STATISTICAL SCIENCES & DATA ANALYTICS
Department of Statistics, Ahmadu Bello University, Zaria, Nigeria | www.jossda.abu.edu.ng www.jossda.com.ng

Early Prediction of Student Dropout Risk from Learning Management System Interaction Sequences using Recurrent Deep Learning Models

T. A. Ogunleye[1*], M. M. Mundu[2] and N. O. Afolabi[3]

[1]*Department of Mathematics and Statistics, Kampala International University, Uganda.*
[2]*Department of Elect. Telecommunications and Computer Engineering, Kampala International Univ., Uganda*
[3]*Department of Statistics, Ogun State Institute of Technology, Igbesa, Ogun State, Nigeria*

Corresponding Author:
T. A. Ogunleye, Department of Mathematics and Statistics, Kampala International University, Uganda. timothy.ogunleye@kiu.ac.ug

Abstract

Given the fluidity of engagement in online and blended learning, early identification of students who are at risk of dropping out remains a major challenge. The study aims to explore the effectiveness of recurrent deep learning models in predicting students' dropout using sequential interaction data from the LMS. An interaction dataset from a public LMS system with approximately 100,000 activity logs from students was used to model their temporal learning behaviours. After pre-processing data, feature engineering, sequence construction, and normalisation, two RNN models were employed and evaluated based on accuracy, precision, recall, F1-score and ROC-AUC performance metrics. The classification accuracy for both models was around 60%, with similar results. Their predictive performance, though, was still modest, with ROC-AUC around 0.50 and recall less than 25%, which means that there was not much discriminatory power between at-risk and retained students. The feature importance analysis showed that the interactions in LMS were weak predictors of dropout, meaning that the information from behavioural engagement alone is limited in predicting dropouts. The results highlight the need for the development of more powerful predictive models using multimodal ensembles which combine the LMS interaction sequences with academic, demographic, and behavioural data. This study offers empirical proof of the challenges of sequence-based RNNs for educational risk prediction and lays the groundwork for building more powerful, explainable and intelligent early warning systems for timely intervention and student retention in digital learning environments.

Keywords: GRU; learning analytics; LMS; LSTM; RNN; student dropout prediction.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
Ogunleye *et al.* 2026 **DOI**: 10.XXXXX/jossda.2026.1.1.014

1. Introduction

Dropout is one of the most prominent issues facing higher education in general, especially in the era of online/blended learning. Early school dropouts negatively impact student achievement, institutional effectiveness and overall socio-economic development, and lower the return on investment for both public and private investments in education (Marcolino *et al.*, 2025). Therefore, targeting early identification of students at risk of dropping out has become a strategic focus for schools looking to put in place rapid and targeted interventions (Pan *et al.*, 2024; El Hajoui *et al.*, 2025). A new wave of Learning Management Systems (LMSs) such as Moodle, Canvas and Blackboard have provided unprecedented possibilities for this, generating vast amounts of behavioural data that document the interactions that learners have with digital learning environments (Romero and Ventura, 2020). An LMS interaction log offers detailed over-time data, such as how often students log in, which resources they view, how much time they spend on the forum, when they submit assignments, and the time they spend on the LMS course (Wen & Hu, 2023; El Hajoui *et al.*, 2025; Marcolino *et al.*, 2025; Osemwegie and Amadin, 2023; Ohkawauchi and Tanaka, 2023).

With the advent of Educational Data Mining (EDM) and learning analytics, it is possible to see how far the analysis of these behavioural data has progressed, combining various statistical, machine learning and Artificial Intelligence techniques to predict the educational outcomes of the learners (Romero & Ventura, 2020). While the logistic regression, decision tree, support vector machine, random forest, and other machine learning models have shown promising performance in predicting student learning behaviors, they rely on aggregated or engineered features, neglecting the sequential and temporal aspects of students' learning behaviours (Siemens and Baker, 2012; Sokolova and Lapalme, 2009; Wen and Hu, 2023; El Hajoui *et al.*, 2025; Gokul *et al.*, 2025; Nti and Ramanayake, 2025). Because student dropout tends to be a gradual process over a period of time, rather than a sudden one, it is important that analytical models incorporate learning of student behavioural trajectories over time (Tinto, 1993; Xing & Du, 2019; Laili *et al.*, 2024). They were thus developed as solutions to capture long-term temporal dependencies that are not always identified by static models, and to retain historical information and model sequential data, as recurrent neural networks (RNNs), such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, are specially designed for these tasks (Hochreiter and Schmidhuber, 1997; Chung *et al.*, 2014; Hu *et al.*, 2024; Putra *et al.*, 2025; Gokul *et al.*, 2025; Marcolino *et al.*, 2025; Ogunleye, 2025).

Recent research has shown that LMS interaction data can be leveraged to improve the predictive performance of a recurrent deep learning model by using attention layers to determine the most informative behavioural patterns (Prenkaj *et al.*, 2020; Segura *et al.*, 2022; Ohkawauchi and Tanaka, 2023; Kumar *et al.*, 2025). The most practical applications of these models are the ability to detect at-risk learners at an early stage in a course that allows institutions to provide personalised feedback, targeted tutoring, adaptive learning pathways, counselling and other interventions promptly before a learner becomes disengaged and the process is irreversible (Early *et al.*, 2017; Dawson *et al.*, 2017; Pineda-Pertuz *et al.*, 2022; Hu *et al.*, 2024; Hidayat *et al.*, 2026). But there are still some critical research gaps. Previous research often uses aggregated interaction features instead of sequence modelling, which is

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.014

limited in its scope of research on the timing of reliable prediction, or which has various issues with class imbalance, missing or noisy interaction data, model interpretability, and generalisability of results across different educational contexts (Karabacak and Yaslan, 2023; Jamshidi *et al.*, 2023; Hu *et al.*, 2024; Kumar *et al.*, 2025; Laili *et al.*, 2024; Sokolova and Lapalme, 2009; Siemens and Baker, 2012; Romero and Ventura, 2020; Osemwegie and Amadin, 2023; Cho *et al.*, 2024a; Won *et al.*, 2023; Ogunleye, 2025; Fakunle *et al.*, 2025).

In this context, the aim of the current study is to develop and assess LSTM- and GRU-based recurrent deep learning models for predicting the dropout rate of students based on their interaction sequence in the LMS. Specifically, the study investigates and systemises the sequential interaction data to develop meaningful temporal representations of learner engagement, develops recurrent neural network models for identifying students at risk of dropping out, evaluates the models' predictions according to well-known classification metrics, and analyses interaction patterns in the behavior which best explain the dropout risk. The study's innovative contributions are intended to provide a sound methodological base for how to build an early warning system that can support proactive institutional interventions, enhance student retention, and integrate AI into educational data mining and learning analytics, addressing the limitations of current approaches and extending them with sequence-aware modelling.

2. Materials and Methods

2.1. Model Development

Two prominent types of recurrent deep learning architectures, namely Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, both state-of-the-art architectures, were used to develop the predictive framework. They were chosen due to their success in modelling sequential learning behaviours (Krüger *et al.*, 2023; Núñez *et al.*, 2024) and in capturing long term temporal dependencies in learning data regarding educational interactions (Lykourentzou *et al.*, 2009; Bayer *et al.*, 2012; Barber and Sharkey, 2012; Cho *et al.*, 2014b; Márquez-Vera *et al.*, 2016; Mduma *et al.*, 2019; Kim *et al.*, 2023; Putra *et al.*, 2025). Recurrent architectures maintain historical information by means of a recurrent memory mechanism, which makes them ideally suited to be applied to the analysis of ordered sequences of interactions in the Learning Management System (LMS), and to find early behavioural patterns that are closely linked to student dropout (Chawla *et al.*, 2002; Kizilcec *et al.*, 2013; Whitehill *et al.*, 2015; Akçapınar *et al.*, 2019; Coussement *et al.*, 2020). The same set of pre-processed input sequences was used for both models, making significant differences in performance.

2.2. Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) network is a variant of the RNN network that features a memory cell and three gates that control information flow over time (Hochreiter and Schmidhuber, 1997; Hu *et al.*, 2024; Albugami *et al.*, 2024; Kumar *et al.*, 2025). These gates allow the network to keep the relevant historical learning behaviours while ignoring irrelevant information, so as to help overcome the vanishing-gradient problem, which, of course, is typical of long sequential data.

The forget gate controls the amount of previous memory to be retained:

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.014

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{1}$$

where x_t denotes the current interaction vector, h_{t-1} is the hidden state from the previous time step, W_f and U_f are learnable weight matrices, b_f is the bias vector, and $\sigma(.)$ denotes the sigmoidal activation function. The input gate determines the quantity of new information getting into the memory cell,

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{2}$$

The candidate memory state is calculated as:

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)\,. \tag{3}$$

The cell state is then updated based upon

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t \tag{4}$$

The output gate is obtained as:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{5}$$

The propagation of the hidden representation to the next time step is:

$$h_t = o_t \odot \tanh(C_t). \tag{6}$$

The gating mechanism allows the LSTM network to remember informative engagement behaviours over long periods of learning, which is especially useful for modelling a gradual change in behaviours leading to students dropping out.

2.3. Gated Recurrent Unit (GRU)

A computationally efficient alternative to the LSTM architecture, the Gated Recurrent Unit (GRU) lends an extra element of control to the flow of historical information by replacing the forget and input gates with a single update gate, and adds a reset gate to control how much information is remembered (Chung *et al.*, 2014; Karabacak and Yaslan, 2023). This simplified structure has a much lower computational complexity without sacrificing a significant amount of prediction accuracy.

The update gate is defined as follows.

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{7}$$

where (z_t) defines how much historical information will be kept.

The reset gate is given by

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{8}$$

and which controls the amount of memory that the candidate hidden state draws on from previous memory.

The candidate hidden representation is obtained by calculating:

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.014

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \tag{9}$$

the last cover will be considered as

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h_t. \tag{10}$$

The GRU is smaller than the LSTM, and therefore typically trains in less time and has a lower memory cost, without sacrificing predictive ability for sequential data from the education domain. By comparing LSTM and GRU, insights can be gained into the accuracy-compatibility dilemma for predicting early student dropout.

2.4. Model Training

Both of the models were trained with a supervised binary classification method, using the target variable defined as

$$y_i = \begin{cases} 1, & \text{student at risk of dropout,} \\ 0, & \text{student retained.} \end{cases} \tag{11}$$

Parameters of the network were optimised using the Binary Cross-Entropy (BCE) loss function given by

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{12}$$

where N denotes the total number of student sequences, y_i represents the observed class label, and $\hat{y}_i$ denotes the predicted dropout probability. The optimisations of the parameters were done by Backpropagation Through Time (BPTT) with the Adam optimiser. To decrease overfitting, dropout regularisation was used in hidden layers as

$$h_t^{(\text{drop})} = h_t \odot m, \qquad m_j \sim \text{Bernoulli}(1 - p), \tag{13}$$

where p is the dropout probability and (m) is a random binary mask.

Finally, the model performances were assessed with the help of accuracy, precision, recall, F1-Score, and ROC-AUC to comprehensively observe the predictive effectiveness and discriminative power of the LSTM and GRU models in predicting student dropouts at an early stage.

3. Results and Discussion

To gain insights into behavioural differences between the students who finally completed the courses and the students who dropped out, exploratory analysis was carried out. The student interactions collected from the Learning Management System (LMS) included clicks, submissions, uploads, downloads, comments, scrolling interactions, views of learning materials, time spent in the course and course enrolment properties. The variables were then converted into sequential input for the recurrent neural network models.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.014

3.1. Exploratory Data Analysis

The distribution of LMS interaction type across the students, whether they completed or did not complete the activities, is shown in Figure 1. There are clear behaviour differences between active and dropout students. The students who finished their courses had significantly more frequent assignment submissions, participation in quizzes and a longer duration of click activity than the ones who left the course. In contrast, however, students who dropped out showed significantly lower levels of intensity of involvement in almost all types of interaction.

3.1.1. The Distributions of the Features by Dropout Status

The difference between the two groups was the largest in terms of the relationship between their behaviours and assessment, such as in the submission of assignments and interaction with quizzes, which suggests that assessment-related activities are better behavioural indicators of persistence than passive learning activities. Conversely, access to viewing documents and access to videos had fairly comparable distributions for both groups, suggesting that viewing documents alone does not offer much discriminatory information with respect to dropout.

Engagement in time was also significantly different between the two groups. The engagement of students on weekdays was relatively stable for active students, but irregular for dropout students, with some evening and/or weekend usage. Sporadic participation could be a sign of lack of commitment to course activities and of a lack of academic persistence (Figure 1). The overall results indicate that the differences in behaviour are significant between active and dropout students, and support empirical modelling of the interaction histories using recurrent neural networks.

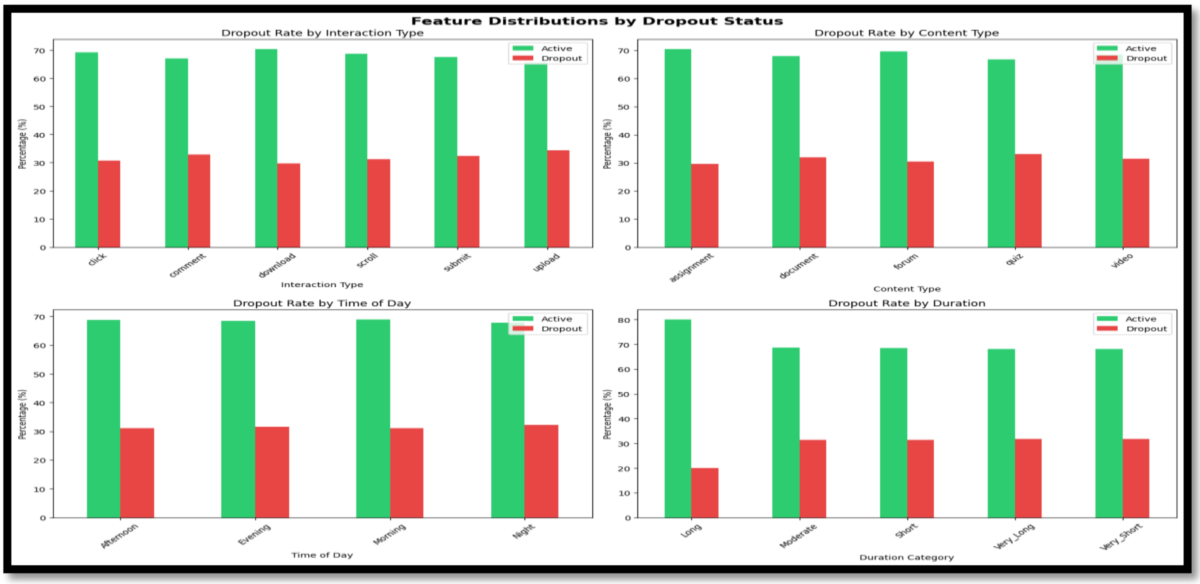

Figure 1. Feature Distributions by Dropout Status

3.1.2. Student Engagement Patterns

The number of LMS interactions students were able to earn across all courses was consistently significantly greater than the number of interactions earned by students who

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.014

ultimately dropped out. The distributions partially overlap, but there is a clear difference in central tendency and dispersion between the groups of highly engaged learners and students at risk of dropping out, suggesting that the students at risk of dropping out interact significantly less with the LMS than do the highly engaged learners.

The overlap between both distributions does exist, however, indicating that interaction frequency is not always a good predictor of dropout status. There were relatively low numbers of interactions from some active students and a moderate number of interactions from other dropout students before they dropped out. Therefore, it is important to consider more contextual information in addition to a simple interaction frequency to accurately identify at-risk students in the predictive models.

This observation underscored the fact that the phenomenon of student dropping out is complex, multidimensional and not only depends on the level of engagement but also on the consistency of the behaviours, academic performance, characteristics of the learners and external contextual factors.

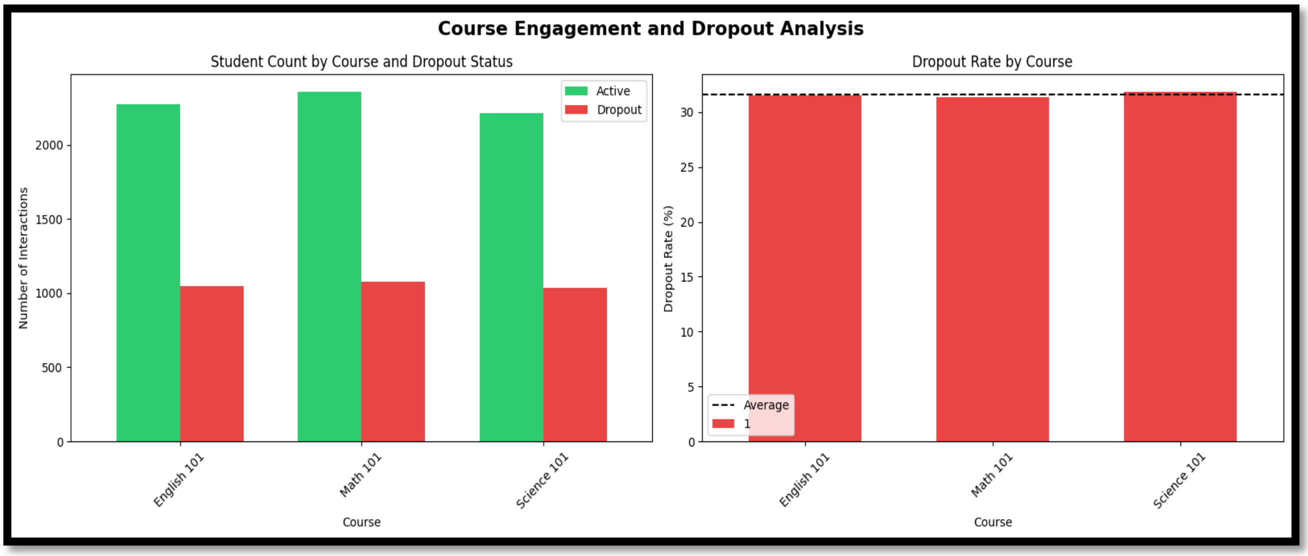

Figure 2. Analysis of Student Engagement and Dropout Analysis

3.1.3. Course-Level Engagement Analysis

While the number of students enrolled in the three courses (English 101, Mathematics 101, Science 101) was different, dropout was not evenly distributed across the courses. There was a significant variation across some courses in the number of students dropping out, even though they had similar numbers of enrolments, indicating that course-specific factors have an impact on students' persistence.

Comparisons of enrolment and dropout rates show that there are specific courses that need focused academic interventions, such as providing greater learner support, improving the design of instruction, providing more opportunities for formative assessment, etc. The results further underline that just using global engagement indicators as predictors for dropouts is not sufficient, but rather that additional context information about the classes should be included.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.014

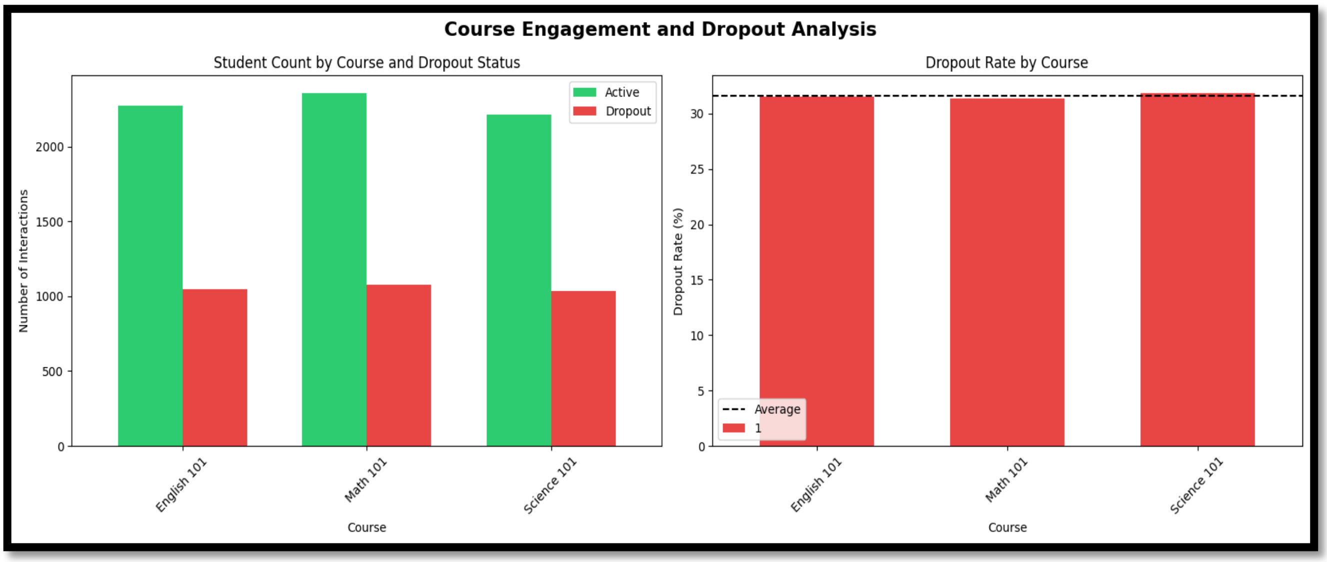

Figure 3. Course Engagement and Dropout Analysis

3.2. Model Performance Evaluation

The predictive performance of the models, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are given in Table 1. The LSTM model achieved slightly better results than the GRU architecture in all of the metrics used. But the differences in the observed results were always small, with similar predictive behaviour seen between the two recurrent architectures. While the overall classification accuracy was above 58%, this alone does not constitute an adequate measure as the classes are not equally balanced (active and dropout classes). More informative measures of performance, such as recall, F1 score and ROC-AUC, showed that both models failed to provide good discrimination between dropout and non-dropout students.

In particular, the two models yielded a low (less than 26%) recall rate, meaning that about 75% of dropping students were misclassified as continuing students. This performance is not sufficient for an education system's early warning system in which the lack of identification of vulnerable learners has a serious academic impact. Likewise, the area under the receiver operating characteristic curve (AUC) is approximately 0.50, suggesting that both recurrent models were at only slightly better than chance.

Table 1. Classification Performance

Metric	LSTM Model	GRU Model	Performance Standard
Accuracy	60.76%	58.76%	Poor (barely above chance)
Precision	33.69%	30.51%	Low (high false positives)
Recall	25.04%	23.93%	Very Low (high false negatives)
F1-Score	28.76%	26.72%	Poor (unbalanced P/R trade-off)
ROC-AUC	0.5032	0.4993	Random (no discrimination)

Source: Authors' Computations (2026)

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.014

3.3. Confusion Matrix Analysis

The results of the models are compared with the observed results in the confusion matrices shown in Table 2, which give additional insight into the behaviour of the models. Both recurrent architectures were able to detect most of the students who were active, while not being able to correctly classify the majority of dropouts. The model was able to correctly classify 1056 students who were active, and correctly identify 158 students who were dropping out. Similarly, the GRU model was able to accurately identify 1,023 active students, while accurately identifying only 151 dropouts.

The largest number of false negatives was the main cause of prediction error. In particular, the LSTM model failed to predict the activity rate of 473 dropout students and the GRU model failed to predict the rate of activity of 480 dropout students. From an educational learning perspective, false negatives are the most serious prediction error, as these students would not be provided with early institutional support when actually at risk of falling out of school. This means that, in general, both models have moderate accuracy, but have not much application in institutional intervention programmes.

Table 2. Confusion Matrix Results

Model	True Negative	False Positive	False Negative	True Positive	Total
LSTM	1,056 (52.9%)	311 (15.6%)	473 (23.7%)	158 (7.9%)	1,998
GRU	1,023 (51.2%)	344 (17.2%)	480 (24.0%)	151 (7.6%)	1,998

Source: Authors' Computations (2026)

3.4. ROC Curve Analysis

Receiver Operating Characteristic (ROC) curves also showed that both the recurrent models had poor discriminative ability. The LSTM obtained an ROC-AUC score of 0.5032, whereas the GRU got 0.4993. The results show that both curves were close to the diagonal reference line, which corresponds to random classification, suggesting that neither of the architectures was able to discriminate between dropout and non-dropout students for any of the different probability levels. These results indicate that the information derived from the series of behaviours in the available interactions is not sufficient to have reliable predictive power for dropout prediction.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.014

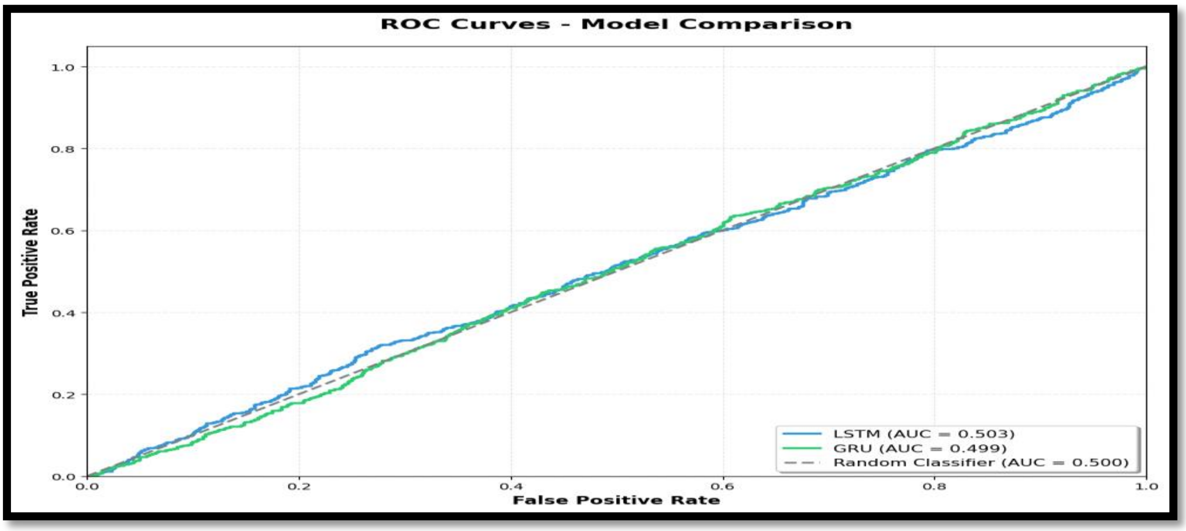

Figure 4. ROC Curves - Model Comparison

3.5. Feature Importance Analysis

Permutation importance analysis (Figure 5) showed no single feature had a significant impact on predictive performance. The contribution of all variables was less than about 1.5% to model accuracy, which suggests that predictive information was rather poorly distributed among the available engagement variables. From this finding, it can be concluded that the performance of the models was limited mainly by the quality of the features and not by the architecture of the network.

Therefore, simply swapping out one recurrent architecture for another for improved prediction accuracy is unlikely to be a major source of improvement. Rather, future research should consider the use of more behavioural, academic, demographic and contextual information, which could offer more robust predictive signals.

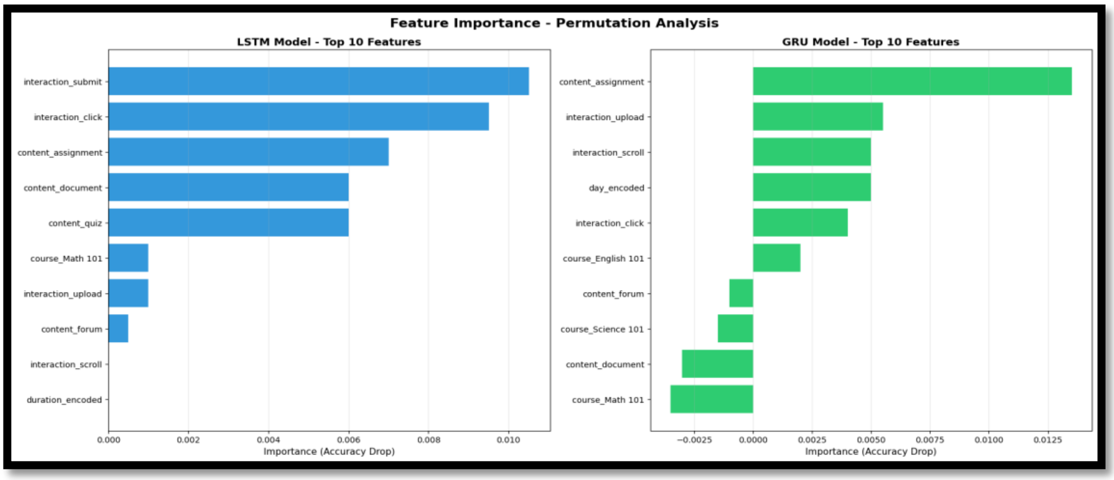

Figure 5. Feature Importance - Permutation Analysis

3.5. Discussion

This study aimed to compare the performance of the recurrent deep learning models in predicting the dropout risk of the students based on sequential interaction data from the LMS. Overall, the results show that the LSTM and the GRU models do not perform well enough to be used in practice. Recurrent neural networks, designed to model temporal dependencies in sequential data, were indeed found to be able to model these interactions, but in the present results, not enough discriminatory information was present in the sequential LMS interactions alone for accurate dropout prediction. The LSTM was marginally superior across all metrics, but was by no means a significant limitation, indicating that the principal limitation was not architectural.

The extremely low recall values are the most significant finding from this study. EWSs for education focus on sensitivity since schools try to detect as many vulnerable students as possible before their withdrawal. The usefulness of the developed models for proactively intervening in students' academic learning is considerably reduced when they are missing from class by about 75% of the actual dropout students.

The ROC-AUC values near 0.50 also suggest that neither of the recurrent networks was able to learn good behavioural representations to distinguish dropout from persistence. These findings aren't due to a problem with the LSTM or GRU models, but rather they reveal that the amount of predictive information in the available LMS engagement sequences is limited.

The results are consistent with previous research (Baker and Inventado, 2014; Ohkawauchi and Tanaka, 2023; Sangeetha and Priya, 2026; Lopez-Muñoz *et al.*, 2026) on dropout rates, which found that predicting dropout rates from behavioural log data alone is often inadequate, especially when there is no academic performance, student background, motivational data, and/or socio-economic data. Recent research (Kim *et al.*, 2023; Kumar *et al.*, 2025; Duan and Chen, 2026; Fan, 2026) in learning analytics have shown that multimodal data integration has a significant impact on predictive performance, as it integrates data from various modalities (behavioural, academic, psychological, institutional) into a single predictive framework.

This is confirmed by the feature importance analysis. The small contribution of all predictor variables shows that there was no dominant behaviour in the present set of data. Rather, predictive information seems to be weak and widely distributed over the interaction variables they have to extract temporal representations from.

In practice, these results highlight the need not to use only the interaction log in LMS to predict dropouts. Early warning systems should be more reliable and be based on sequential behavioural data as well as on other variables, like continuous assessment scores, assignment performance, attendance, previous school achievement, demographic factors, motivation indicators, and course-specific contextual factors.

Last but not least, the slight performance gap between LSTM and GRU indicates that future research should focus on the representation of the data instead of developing other recurrent architectures. Sequential models based on transformer architectures, attention models, graph neural networks, multimodal deep learning, and hybrid models of temporal behavioural sequences and structured academic and contextual information show great promise. These methods would likely be able to achieve significantly better accuracy in prediction than incremental changes to existing recurrent neural networks.

4. Conclusion

While these models can both be used to model the interaction data of an LMS, the results of this study showed that their predictive power was not adequate for successful early identification of the students at risk of dropping out based on their engagement with the LMS when using LMS engagement features as the only features. The classification accuracy of both models was only slightly superior to random chance (ROC-AUC around 0.50 and recall under 25%), which is insufficient for their implementation in institutional EWS. These results indicate that the temporal engagement patterns are insufficient to explain student dropouts, as a number of other factors are involved, and call for more complex and richer features for predicting student dropouts. Ideally, future student dropout prediction systems should incorporate academic performance data, behavioural engagement data, psychological factors, social interaction data, and institutional contextual data, all while respecting ethical standards and data privacy laws. Advanced feature engineering methods, hybrid deep learning architectures, multimodal learning analytics, and cost-effective prediction- intervention frameworks that can provide more accurate, interpretable and actionable predictions to enhance student retention in online learning environments are also worthy of further investigation.

References

Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop an early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, *16*, Article 40. https://doi.org/10.1186/s41239-019-0172-z

Albugami, S., Almagrabi, H., & Wali, A. M. (2024). From data to decision: Machine learning and explainable AI in student dropout prediction. *Journal of E-Learning & Higher Education*, 1–11. https://doi.org/10.5171/2024.246301

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Barber, R., & Sharkey, M. (2012). Course correction: Using analytics to predict course success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 259–262). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330664

Bayer, J., Bydzovská, H., Géryk, J., Obsivac, T., & Popelinsky, L. (2012). Predicting drop-out from social behaviour of students. In *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 103–109). International Educational Data Mining Society.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014a). On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of the 8th Workshop on Syntax, Semantics and Structure in Statistical Translation* (pp. 103–111). Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-4012

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014b). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1179

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modelling. *arXiv*. https://arxiv.org/abs/1412.3555

Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, *135*, 113325. https://doi.org/10.1016/j.dss.2020.113325

Dawson, S., Jovanovic, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 474–478). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027405

Duan, Y., & Chen, X. (2026). An adaptive multi-scale spatio-temporal graph network for robust MOOC dropout prediction. *Scientific Reports*, *16*. https://doi.org/10.1038/s41598-026-40502-w

Early, J., DeBruhl, B., & Koufacos, F. (2017). Identifying student misconceptions in software engineering through machine learning on course data. In *2017 IEEE Frontiers in Education Conference* (pp. 1–5). IEEE. https://doi.org/10.1109/FIE.2017.8190484

El Hajoui, A., Alaoui, O. Y., El Kharki, O., Wahbi, M., Boulaassal, H., & Maatouk, M. (2025). Student performance prediction using RNN models with a variety of optimisation techniques in deep learning. *International Journal of Information and Education Technology*, *15*(6). https://doi.org/10.18178/ijiet

Fan, T. (2026). Enhancing student dropout prediction in educational data mining using sparse feedback-based deep residual network with Xception and optimised feature selection. *European Journal of Education*, *61*(1). https://doi.org/10.1111/ejed.70494

Fakunle, M. A., Ogundoyin, I. K., Ogunleye, T. A., Adeleke, D. K., Ibrahim, M. A., & Popoola, M. A. (2025). Performance comparison study of three ensemble learning models for predicting spatial apparent resistivity distribution around a cemetery site: A case study of Ejigbo Cemetery in Osun State, Nigeria. *Next Research, 2*(4), 1–12. https://doi.org/10.1016/j.nexres.2025.100752

Gokul, B., Dhanush, U., & V. G. S. P. (2025). *Diabetes prediction using machine learning*. Zenodo. https://doi.org/10.5281/zenodo.17144504

Hidayat, N., Afuan, L., & Jannah, H. R. (2026). Prescriptive learning analytics for student dropout: Integrating temporal velocity and counterfactual explanations in longitudinal data. *Journal of Computing Theories and Applications, 3*(4), 627–646. https://doi.org/10.62411/jcta.15920

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu, W., Yu, Y., Yan, J., Zhao, Z., Sun, W., & Shen, X. (2024). Water flow prediction based on improved spatiotemporal attention mechanism of long short-term memory network. *Water, 16*(11), 1600. https://doi.org/10.3390/w16111600

Jamshidi, B., Hakak, S., & Lu, R. (2023). A self-attention mechanism-based model for early detection of fake news. *IEEE Transactions on Computational Social Systems, 11*(4), 5241–5252. https://doi.org/10.1109/TCSS.2023.3322160

Karabacak, E. S., & Yaslan, Y. (2023). Comparison of machine learning methods for early detection of student dropouts. In *Proceedings of the International Conference on Computer Science and Engineering* (pp. 376–381). IEEE. https://doi.org/10.1109/UBMK59864.2023.10286747

Kim, S.-Y., Choi, E.-T., Jun, Y. K., & Lee, S. (2023). Student dropout prediction for university with high precision and recall. *Applied Sciences, 13*(10), 6275. https://doi.org/10.3390/app13106275

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analysing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 170–179). Association for Computing Machinery. https://doi.org/10.1145/2460296.2460330

Krüger, J. G. C., de Souza Britto, A., & Barddal, J. P. (2023). An explainable machine learning approach for student dropout prediction. *Expert Systems with Applications, 233*, 120933. https://doi.org/10.1016/j.eswa.2023.120933

Kumar, N. B. M., Chithrakumar, T., Thangarasan, T., Dhanasekar, J., & Logamurthy, P. (2025). AI-powered early detection and prevention system for student dropout risk. *International Journal of Computational and Experimental Science and Engineering, 11*(1). https://doi.org/10.22399/ijcesen.839

Laili, U. F., Tanzeh, A., Efendi, N., & Gufron, M. (2024). K-means clustering method on academic advising management and early detection of student dropout. *International Journal of Educational Research & Social Sciences, 5*(2), 335–346. https://doi.org/10.51601/ijersc.v5i2.744

Lopez-Muñoz, G. L., Gonzalez Serrano, C., & Sanchez-Ferreira, C. (2026). Student dropout prediction in higher education: A systematic review of machine learning methods and risk factors. *Journal of Information Technology Education: Research*, *25*, Article 20. https://doi.org/10.28945/5776

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, *53*(3), 950–965.

Marcolino, M. R., Porto, T. R., Primo, T. T., Targino, R., Ramos, V., Queiroga, E. M., Muñoz, R., & Cechinel, C. (2025). Student dropout prediction through machine learning optimisation: Insights from Moodle log data. *Scientific Reports, 15*, 9840. https://doi.org/10.1038/s41598-025-93918-1

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems, 33*(1), 107–124.

Mduma, N., Kalegele, K., & Machuve, D. (2019). A survey of machine learning approaches and techniques for student dropout prediction. *Data Science Journal, 18*, Article 14. https://doi.org/10.5334/dsj-2019-014

Nti, I. K., & Ramanayake, S. (2025). *Explainable machine learning for student dropout prediction and tailored interventions in online personalised education*. Research Square. https://doi.org/10.21203/rs.3.rs-6615052/v1

Núñez, A. Á., Del Carmen Santiago Díaz, M., Vázquez, A. C. Z., Marcial, J. P., & Linares, G. T. R. (2024). Early detection of students at high risk of academic failure using artificial intelligence. *International Journal of Combinatorial Optimisation Problems and Informatics*, *15*(5), 155–160. https://doi.org/10.61467/2007.1558.2024.v15i5.573

Ogunleye, T. A. (2025). Predictive modelling of kerogen types using supervised machine learning algorithms: A geochemical study of the Niger Delta Basin, Nigeria. *Journal of Applied Sciences, Information and Computing*, *6*(1), 157–167.

Ohkawauchi, T., & Tanaka, E. (2023). Predicting student dropout risk using LMS logs. *IIAI Letters on Institutional Research, 4*, Article 226. https://doi.org/10.52731/lir.v004.226

Osemwegie, E. E., & Amadin, F. I. (2023). Student dropout prediction using machine learning. *FUDMA Journal of Sciences*, *7*(6), 347–353. https://doi.org/10.33003/fjs-2023-0706-2103

Pan, F., Zhang, H., Li, X.-B., Zhang, M., & Ji, Y. (2024). Achieving optimal trade-off for student dropout prediction with multi-objective reinforcement learning. *PeerJ Computer Science, 10*, e2034. https://doi.org/10.7717/peerj-cs.2034

Pineda-Pertuz, C. M., Martinez, Y., & Diaz, I. (2022). Machine learning algorithms to predict desertion in the faculty of engineering sciences. *IOP Conference Series: Materials Science and Engineering, 1253*(1), 012013. https://doi.org/10.1088/1757-899X/1253/1/012013

Prenkaj, B., Velardi, P., Stilo, G., Distante, D., & Faralli, S. (2020). A survey of machine learning approaches for student dropout prediction in online courses. *ACM Computing Surveys*, *53*(3), Article 70. https://doi.org/10.1145/3388792

Putra, L. G. R., Prasetya, D. D., & Mayadi, M. (2025). Student dropout prediction using random forest and XGBoost methods. *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, *9*(1), 147–157. https://doi.org/10.29407/intensif.v9i1.21191

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355. https://doi.org/10.1002/widm.1355

Sangeetha, K., & Priya, N. S. (2026). Improved GRU-based temporal modelling for robust student dropout classification and risk prediction. In *Proceedings of the International Conference on Smart Electronic Devices and Intelligent Systems* (pp. 453–460). IEEE. https://doi.org/10.1109/ICSEDIS68157.2026.11518148

Segura, M., Mello, J., & Hernández, A. (2022). Machine learning prediction of university student dropout: Does preference play a key role? *Mathematics*, *10*(18), 3359. https://doi.org/10.3390/math10183359

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330661

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226922461.001.0001

Wen, X., & Hu, J. (2023). Early prediction of students' performance using a deep neural network based on online learning activity sequence. *Applied Sciences*, *13*(15), 8933. https://doi.org/10.3390/app13158933

Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout. In *Proceedings of the 8th International Conference on Educational Data Mining*. International Educational Data Mining Society.

Won, H.-S., Kim, M. J., Kim, D., Kim, H. S., & Kim, K. M. (2023). University student dropout prediction using pretrained language models. *Applied Sciences*, *13*(12), 7073. https://doi.org/10.3390/app13127073

Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalised intervention. *Journal of Educational Computing Research*, *57*(3), 547–570. https://doi.org/10.1177/0735633118757015