Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 12, No. 4, December 2024, pp. 758~767
ISSN: 2089-3272, DOI: 10.52549/ijeei.v12i4.5635

Predictive Analysis of Learner's Performance in Online Environments with LSTM and Attention Mechanism

Smruti Nanavaty[1], Ajay Khunteta[2]
[1] Ph.D. Scholar, Department of Computer Science, Poornima University, Jaipur, India
[2] Professor, Faculty of Computer Science and Engineering, Poornima University, Jaipur, India

Article Info

Article history:

Received Jun 2, 2024
Revised Sep 16, 2024
Accepted Oct 8, 2024

Keywords:

At-risk learners
Online learning platforms
Learning outcomes
Periodic activities
Predictive analysis

ABSTRACT

Early identification and supporting at-risk learners is a key problem in digital learning environment. This paper investigates the use of deep learning methods, namely Long Short-Term Memory (LSTM) neurons with cognitive mechanisms to determine those learners that are most likely to be at risk based upon the involvement of the learners in periodic assessment as well as engagement with the learning components in online learning environments. It also accounts for the relevance of dependencies of temporal elements, which adds a degree of precision in forecasting. The findings show how advanced analysis of data can potentially improve student support strategies with online learning systems, thus ensuring the success and retention of learners in consequence. From the test, result yield information concerning the robustness of the LSTM model in predicting the learner's achievement and provides insight into factors that most importantly have an impact on prediction. That suggests the approach of LSTM with attention mechanism is effective to capture periodic behavior of the learner on virtual platforms and early predictions will be useful for administrators to design timely intervention and improve retention rates of learners.

Corresponding Author:

Smruti Nanavaty,
Ph.D Scholar, Department of Computer Science,
Poornima University,
Ramchandrapura, Jaipur – 303905, Rajasthan, India.
Email: smruti.nanavaty@gmail.com

1. INTRODUCTION

In recent years, the growing accessibility and popularity of Virtual Learning Environments (VLEs) have revolutionized education, allowing for flexible, personalized, and data-driven learning experiences. However, predicting student performance remains a significant challenge due to the large, complex, and sequential nature of learning data.The most common challenge faced by online platforms is with regards to identification and support of at-risk learners who may be underperforming or are distracted due to various reasons[1][2]. It is therefore necessary to monitor the learner's activity logs to monitor behaviour as well as performances during periodic assessments continuously during the course. An earlier study has proved that learner engagement levels during live sessions, time spent on the learning content and frequency of visits to archived content are among the useful predictors for identification of at-risk learners as highlighted in a past study [3]. Periodic activities of the learners explain the engagement levels of learners on these online platforms. The data collected for these periodic activities includes periodic assessment and submission, which is analysed using Deep Learning techniques like Artificial Neural Networks (ANN), Recurrent Neural Networks RNN including Long Short-term Memory, and GRU have provided promising results [1][4][5][6][7]. Despite these advances, none of the behavioural aspects when taken alone, correlated to learner's final performance. Therefore, a comprehensive impact of these non-linear predictors should be considered [8][9]. Group Convolution and Dilated Causal Convolution (CGDC) combined with LSTM helps in capturing the impact of various learning periods of the learner's performance [10]. The issues observed in earlier studies where CNN was used for automatic feature recognition is that local correlations structures need to be identified while input tables are created for numerical data and if the data is periodic then use of LSTM needs to be considered [11].

Various machine-learning techniques have been employed in educational settings, including Random Forests, Support Vector Machines, and Neural Networks, with prediction accuracies ranging from 80% to 95% [8]. However, these models often fail to capture the temporal dependencies inherent in student interactions with VLEs. Rescent research [14] has demonstrated the effectiveness of Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, in analyzing such time-series data. LSTM models excel in understanding sequential patterns and have outperformed traditional machine learning methods in predicting complex, temporal behaviors. Furthermore, attention mechanisms have emerged as a powerful addition to LSTM models by helping focus on the most relevant parts of the input data, enhancing the interpretability and accuracy of predictions. For example, studies in natural language processing and image recognition have demonstrated the efficacy of attention mechanisms in highlighting key features. Inspired by these advances, our study proposes the use of LSTM models with attention mechanisms to predict learners' performance in VLEs, addressing gaps in previous models that inadequately captured temporal patterns. Authors [14] presented the eight algorithms included the gradient boosting (GB), bagging, XGBoost, AdaBoost, random forest (RF), logistic regression (LR), support vector machine (SVM), and K-nearest neighbour (KNN) algorithms on the The Open University Learning Analytics Dataset (OULAD) dataset. This study did not explore the specific reasons for this difference dropout in first and second semester, although teachers can be mindful of it in their practice. Authors [15] used Long Short-term Memory (LSTM), in the prediction of students at risk of failing a course offered in a self-paced mode of online education. The study has utilized a freely accessible Open University Learning Analytics Dataset comprising 22,437 students with 69 % pass, and 31 % failed instances. The deep LSTM shows the highest predictive power to classify between pass and fail students, compared to all other alternatives by achieving an accuracy of 84.57 %, precision of 82.24 % and recall of 79.43 %. Researchers [16], presended the studies focus on utilizing LA within BL to predict students who are at risk of failure and to establish early-warning systems. Most of these related studies are relevant to investigation pertaining to BL courses, despite some being based on MOOCs or fully online learning. Authors [17] investigated the early identification of at-risk students in a BL setting by scrutinizing the current learning performance and learning-behavior variables. Their findings indicate that specific factors, such as irregular engagement and subpar performance, can effectively highlight students at risk of academic failure. Authors [18] investigated the use of predictive ML models for small- and medium-sized university courses in a fully online learning environment, where the number of students is limited. Various model-building techniques are presented that can be used to increase the predictive power of models, beginning from less effective techniques to more developed ones. These techniques apply to small or mid-sized university courses and demonstrate monotonically increasing performance metrics over time. Authors [19] predicted student success in BL environments by acquiring diverse data, including student demographics, engagement metrics, and assessment results. They developed predictive models designed to identify students at risk of failure who might benefit from additional support.

The current article addresses integration of timed submission entries and a quiz in to online teaching environments in order to detect academically at-risk learners. Through the employment of machine learning technologies and advanced data analytics, the instructors may be able to discover behaviour of the learners, engagement patterns and evaluation statistics to point at weak areas and help to improve them [2]. Our specialization concerns the implementation of deep learning techniques, for example, Long Short-Term Memory networks with attention mechanisms, to be able to work with students' sequential data flows in order to join connections and make accurate predictions of outcome of a learner.

In addition to the practical implementation of LSTM networks and attention mechanisms, it is crucial to understand the mathematical underpinnings of these techniques. LSTM networks are described by their intricate architecture, involving recurrent connections and gating mechanisms that enable them to capture long-term dependencies in sequential data. Mathematically, the operations within an LSTM cell, including the computation of cell state updates and gate activations, governed by a series of matrix multiplications, element-wise operations, and non-linear activation functions[6][11]. Similarly, attention mechanisms employ mathematical formulations to dynamically weigh the importance of different input features, often through dot-product attention or other attention scoring mechanisms. By delving into the mathematical foundations of these techniques, educators and researchers can gain deeper insights into their functioning and optimize their application for improved prediction accuracy and interpretability.

The study outlines the process of the data preparation stage, the model architecture implementation and experimental results obtained, to prove the effectiveness of LSTM model augmented with attention mechanisms for predictive modelling of online learners' engagement based on their activity records from periodic assignments and the quiz components. Furthermore, we consider the educational research in reference to our conclusions and indicate the possible ways of helping high-risk learners to get earliest counselling concerned their difficulties.

Early warning systems are built to develop administrative policies and intervention strategies to improve educational planing for reducing attrition rates of learners [12]. The model proposed has predictive capabilities and will provide necessary insights to administrators and instructors early during the course so that timely intervention strategies aimed at boosting the support for online learners who may be at risk of dropping out. Necessary startegies may be developed keeping in mind all ethical consideration with an objective of achieving positive academic outcomes and improving the retention rates of learners on virtual platforms.

2. RESEARCH METHOD

LSTM networks are a type of Recurrent Neural Network (RNN) architecture designed to capture long-term dependencies in sequential data[10]. Unlike traditional RNNs, which often struggle to retain information over long sequences due to vanishing gradient problems, LSTM networks overcome this limitation by introducing specialized memory cells and gating mechanisms.

2.1 LSTM Architecture

The most critical components that make up an LSTM cell include:

- Cell State – Maintains the cell's memory, ensuring information preservationd in lengthy sequences.
- Hidden State – As the cell's output, it combines relevant information from the current input and the most distant hidden state.
- Forget Gate – Determines which information should and should not be held in the cell state.
- Input Gate – Controls the cell state's adjustment by monitoring the inflow of new data.
- Output Gate – Determines, which part of the cell state, should be utilized as the output.

2.2 Exploring LSTM Network Training and Attention Mechanisms

While traditional time-series models, such as ARIMA and Exponential Smoothing, perform well in forecasting stationary and linear patterns, LSTMs offer distinct advantages in capturing complex, non-linear, and long-term temporal dependencies. LSTMs can dynamically adapt to varying temporal contexts, handle non-linearities, and maintain memory over long sequences, making them well-suited for analyzing student engagement patterns, which are often irregular and influenced by multiple factors over time.

LSTM networks update their parameters using BackPropagation Through Time (BPTT). It is done by propagating the errors gradients through the entire sequence and adjusting the weights to reduce the value of the loss function. Attention is another component of our model besides the LSTM layers. The attention mechanism enables the ability to focus on different parts of the input sequence dynamically. Instead of treating the feature vectors as uniform, the model may assign more value to some features within a sequence depending on their importance to the model's current decision. This mechanism enables to extract vital data from the input and hence provide a better performance in prediction.

2.3 Hyperparameters

Hyperparameters are the settings of parameters assigned prior to model training, affecting the model's behavior and the performance. They are frequently pre-defined and do not change with data and thus need to be tuned through investigation to achieve the desired model performance. Hyperparameters of an LSTM model network include:

- Number of Units – which is the dimension of the LSTM cell state and the hidden state;
- Learning Rate – which is the step size taken during parameter updates to avoid overshooting the optimal solution;
- Batch Size – the number of samples processed in each iteration;
- Number of Epochs – the number of times the entire training data is passed through the model; and the
- Validation Split – the percentage of training data held out for validation to prevent overfitting the model.

The hyperparameters used for building the model are presented in the Table 1.

Table 1. Hyperparameters for the proposed Model with their respective values

Hyper parameter	Value
LSTM Units	50
Learning Rate (LR)	0.001
Batch Size	32
Number of Epochs	10
Validation Split	0.2

By utilizing, the capabilities of LSTM networks with attention mechanisms, this model can effectively capture time-based dependencies and feature importance in sequential data, thus making it well suited for predicting student performance based on their engagement and performance metrics.

2.4 Dataset

The dataset used for this study was collected from a Virtual Learning Environment (VLE) of an online educational platform. It comprises anonymized data logs from students enrolled in various courses over a period of one academic term (e.g., six months). The dataset contains interactions such as quiz submissions, assignment uploads, discussion forum participation, and overall course activity logs, representing a comprehensive record of learner engagement. The dataset includes the following key features:

- Student ID: A unique identifier for each student, ensuring privacy and enabling longitudinal tracking.
- Quiz Scores: A record of quiz performance, including submission times and scores for each quiz.
- Assignment Submissions: Detailed logs of assignment submission times and grades.
- Forum Participation: Metrics on the number of posts, replies, and interactions in discussion forums.
- Activity Logs: Timestamped logs of login times, video watch durations, and course material views.

2.5 Model Description and Architectural flow

The proposed LSTM model is created using the Keras API, featuring two LSTM layers, each with 50 LSTM units, and an attention mechanism placed between the LSTM layers. This attention mechanism allows the model to selectively focus on key parts of the input sequence, enhancing the predictive performance. A Time Distributed Dense layer with a sigmoid activation function is used for binary prediction at each time step, making the model capable of producing predictions for multiple sequential outputs. Compared to previous models like Random Forest and Support Vector Machines, which have achieved up to 95% accuracy , this LSTM model benefits from its ability to capture temporal dependencies and periodic behaviors, which are crucial for predicting learner performance over time. While other studies in the education domain have utilized feed-forward neural networks and decision trees, they often overlook the importance of the sequence in learning activities. By integrating an attention mechanism, this LSTM model is not only capable of leveraging these sequential dependencies but can also provide insight into which features (such as quiz scores and assignments) have the most significant impact on predictions.

The model architecture thus comprises of two LSTM layers having 50 LSTM units each, an attention mechanism and a Time Distributed Dense layer with sigmoid activation Function and the learning rate (LR) of 0.001.

Figure 1 explains the Architectural Flow of the model illustrating the steps from data selection, loading and preprocessing, model evaluation and outcome of the model.

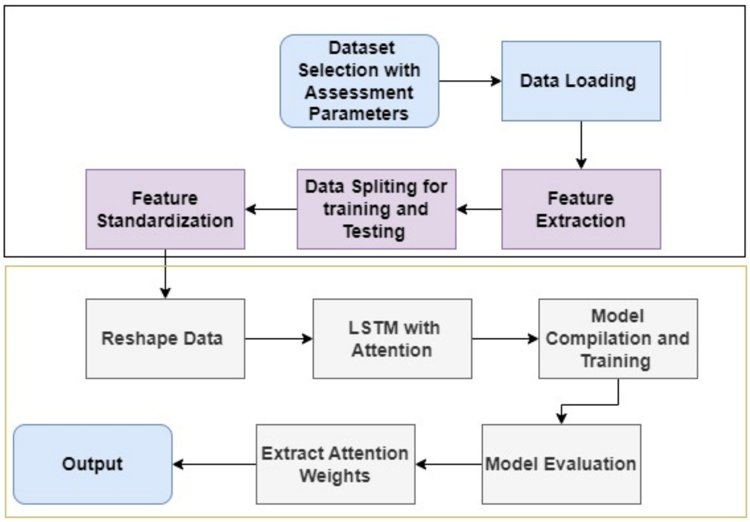

Figure 1. Architectural flow of the Model

3. RESULTS AND DISCUSSION

The results obtained highlights the Model's performance through various metrics, supplying valuable insights for evaluating the effectiveness of LSTM in predicting the learning outcomes of online learners in our study.

3.1 Metrics and Evaluation

The evaluation metrics for the model are described as follows:

1. Loss is a metric used to find the dissimilarity between the predicted and the actual results. Lower values for this metric are indicative of better model performance.
2. Accuracy is a measure for how correctly the model is able to predict the learner's outcomes and how often the model produces correct predictions.
3. Precision reflects the level of accuracy of positive predictions demonstrating the model's capability to identify positive outcomes.
4. Recall is a metrics that focuses on the models ability to capture all the positive outcomes.
5. Area Under the Curve (AUC) signifies the model's ability to distinguish between positive and negative outcomes across various levels of thresholds.
6. Validation results are computed by working with separate dataset assisting the assessment of model's generalised performance
7. The average attention weights are computed for finding the significant contribution of features like quiz component and assignments in model's prediction results.

The performance of the models is remarkable as it showed an accuracy of 98.32%, Precision and recall were computed to be more than 98%, which denotes that the model is able to accurately predict for both the positive as well as negative outcomes. Moreover, the AUC values are 99.68% and impressively explain that the model is effectively able to discriminate between positive and negative outcomes as presented in Table 2.

Table 2. Model Performance

Evaluation	Metrics
Loss	0.050473
Accuracy	0.983194
Precision	0.986244
Recall	0.995818
AUC	0.996750

Table 3. Attention weights for Periodic Assessments

Average	Attention Weights
Quiz1	0.628
Quiz2	- 0.571
Quiz3	- 0.620
Quiz4	0.625
Quiz5	- 0.558
Quiz6	- 0.600
Quiz7	- 0.594
Quiz8	0.620
Quiz9	0.576
Quiz10	0.569
Quiz11	0.588
Quiz12	- 0.430
Quiz13	0.691
Quiz14	0.576
Quiz15	- 0.617
Assignment1	0.596
Assignment 2	- 0.566

The results inclusively suggest that the proposed LSTM model with attention mechanism has been effective in predicting the learner's performance based on periodic submissions and activities like quizzes conducted in phased out manner during the course duration on Virtual Learning Environments (VLEs).

This LSTM model with attention mechanisms achieved an accuracy of 98.32%, a precision of over 98%, and an AUC score of 99.68%. These results are impressive when compared to previous studies that utilized traditional machine learning techniques, which typically achieved accuracy levels ranging from 80% to 95%. The inclusion of the attention mechanism allowed the model to focus on critical features like quiz components and assignment submissions, resulting in more precise and actionable predictions.

Moreover, the average attention weights, as shown in Table 3, disclose that features related to periodic assessments have a more significant influence on prediction outcomes than other features, which could be crucial in identifying at-risk students early on. By integrating attention mechanisms with LSTM networks, this model addresses the shortcomings of previous methods that did not account for temporal patterns and dependencies adequately. This is a significant innovation as it allows for more timely and personalized interventions by educators.

The average attention weights as seen in the Table 3 provides perspectives of the features contributing to the model's predictions. It is observed that the features with higher absolute attention weights have a significant influence in determining the outcomes.

Our study also addresses the shortcomings of earlier methods that did not account for temporal patterns and dependencies adequately. By integrating LSTM networks with attention mechanisms, we can better capture and utilize the sequential nature of student interactions with online platforms, which previous models often overlooked. This is a significant contribution as it allows for more timely and personalized interventions.

The primary limitation of this model lies in its computational complexity. LSTM models, particularly when combined with attention mechanisms, can be resource-intensive, requiring significant computational power for training. Future research should focus on optimizing the model for better scalability, perhaps by incorporating more efficient architectures like GRU (Gated Recurrent Units) or Transformer models, which have been shown to outperform LSTMs in some contexts. Moreover, while our model achieves high predictive accuracy, it is heavily reliant on the quality of input data. Further studies should explore the model's adaptability across different educational settings and datasets to enhance its generalizability.

3.2 Interpretation of Results

The results and findings from the proposed LSTM model with attention mechanism trained using data logs collected from VLEs upholds several key insights. The model evaluation shows that the model performs accurately and yield precise output in terms of AUC and recall to substantiate that the model is reliable for prediction of learner's performance based on periodic submission like assignments and quiz performances. Such a model could be useful to educator for providing early alerts by identifying underperforming learners. Moreover, the model may serve as a useful tool to develop personalized intervention as per the requirement of individual learners.

3.3 Ethical Implications of Predicting Student Performance

The use of predictive models like LSTM with attention mechanisms to forecast student performance and identify at-risk learners can offer tremendous benefits in educational settings. However, such models must be implemented responsibly to avoid potential negative consequences. One of the primary concerns with using predictive models in education is the collection and use of sensitive student data. These models often rely on detailed records of student interactions, assessments, and other personal information, which raises questions about privacy and data security. Models like LSTMs, especially when combined with attention mechanisms, are often complex and lack transparency. This "black-box" nature can make it challenging for educators and students to understand how decisions are being made. Predicting student performance carries the risk of self-fulfilling prophecies, where students who are labeled as "at-risk" might internalize the prediction and exhibit behaviors that align with this label. Such predictions, if not used responsibly, can also influence how teachers perceive and interact with students. To ensure the responsible use of predictive models in educational settings, institutions must balance the potential benefits of early identification and support for at-risk learners with the ethical risks associated with privacy, fairness, and transparency. Establishing robust data governance policies, incorporating bias mitigation strategies, and ensuring human oversight are essential steps in fostering a fair and supportive learning environment that respects the rights and dignity of all students.

3.4 KEY FACTORS

In our study, several key factors have been identified that significantly contribute to the predictive power of the LSTM model with attention mechanisms. Understanding these factors is crucial for educators aiming to enhance student outcomes through timely interventions. Below, we delve into the specific factors derived from the dataset and their respective contributions:

Quiz Scores: Quiz scores emerged as one of the most critical predictors of student performance, as evidenced by the attention weights assigned to various quizzes (e.g., Quiz1 received an average attention weight of 0.628). Higher quiz scores correlate positively with overall performance, suggesting that consistent performance in quizzes indicates strong engagement and understanding of the material. This emphasizes the need for educators to monitor quiz performance closely and provide support for students who struggle with these assessments.

Assignment Submission Timing: The timing of assignment submissions also plays a significant role in predicting student success. Students who submit assignments on time tend to perform better overall. For instance, the average attention weight for Assignment1 was 0.596, indicating its importance in the prediction model. Early submissions can suggest proactive engagement, while late submissions may indicate potential academic challenges or lack of motivation, prompting the need for interventions.

Frequency of Engagement: The data includes various engagement metrics, such as the number of forum posts, replies, and overall interaction with course materials. These factors can serve as indicators of student commitment and interest. High engagement levels often correlate with better performance, suggesting that students actively participating in discussions are likely to perform better academically.

Activity Log Analysis: Timestamped activity logs, including login times and video watch durations, offer insights into student behavior patterns. Consistent login behavior and extended durations spent on course materials can positively influence performance outcomes. For instance, students who frequently access content and participate in synchronous learning opportunities may be more likely to achieve higher grades.

By expanding our analysis of the predictive factors, we establish a clearer understanding of the elements that contribute most to the model's predictive power. The insights gained from this analysis can inform educators about which aspects of student engagement to monitor closely, enabling them to implement timely interventions for at-risk learners. Future research could explore additional dimensions, such as the impact of demographic factors or external commitments on student performance, further enhancing our predictive capabilities in educational settings.

3.5 Practical Implications of The At-Risk Learner Detection Model For Educators:

The developed LSTM model with attention mechanism provides an effective tool for identifying at-risk students early in online learning environments. However, the value of such a predictive model is fully realized only when the insights are translated into actionable strategies for educators.

Early Identification and Proactive Interventions: The primary advantage of using the LSTM model is its ability to identify at-risk learners early in the course based on their engagement patterns and periodic assessment performances. This early detection allows educators to intervene before students experience significant academic difficulties.

Personalized Learning Pathways: The attention mechanism within the model can highlight which specific activities (e.g., quizzes, assignments, or forum participation) have the greatest impact on the prediction of student performance. This information can be used to design personalized learning pathways and resources for each student based on their unique engagement patterns.

Adaptive Course Design: The model's predictions can also inform course design by identifying common points where many students tend to struggle or disengage. By analyzing patterns across multiple cohorts, educators can pinpoint specific content areas or course structures that may need modification.

Real-Time Monitoring and Alerts: The model can be integrated into the learning management system (LMS) to provide real-time monitoring of student engagement and performance. Real-time alerts can be generated when the model predicts a sudden drop in engagement or a shift to a higher risk category.

Facilitating Peer Support and Collaborative Learning: The predictions from the model can be used to foster a collaborative learning environment by pairing or grouping students based on their predicted performance levels and learning needs. For instance, students at risk in certain topics can be paired with those who have a strong understanding of the same topics, encouraging peer learning and support.

Enhanced Communication Strategies: The model's outputs can be used to tailor communication strategies to engage at-risk learners more effectively. For example, students flagged as disengaged can receive personalized messages that emphasize the importance of consistent engagement and offer support resources.

Informing Institutional Policies and Resource Allocation: The aggregate data generated from the model's predictions can provide valuable insights for educational administrators to make informed decisions about resource allocation and policy adjustments.

Ethical and Fair Use of Predictive Models: Finally, while leveraging predictive models in educational settings can have significant benefits, it is crucial to use these tools ethically. Educators should be cautious not to label or stigmatize students based on predictions and should always contextualize the model's outputs with qualitative insights and human judgment.

The practical implications of the LSTM-based at-risk learner detection model extend far beyond predictive accuracy. By providing early identification and insights into student behavior, the model can be leveraged to design personalized learning pathways, inform adaptive course design, and enhance real-time monitoring and support strategies. However, to realize these benefits, the model must be implemented thoughtfully, with a focus on ethical considerations and the inclusion of human judgment. When used responsibly, this predictive tool can significantly improve educational outcomes and foster a more supportive, engaging, and inclusive learning environment.

4. CONCLUSION

To sum up, the application of Long Short-Term Memory (LSTM) networks with attention mechanisms opens new perspectives for educational predictive analytics [13]. The current study demonstrates the effectiveness of these advanced techniques in analyzing temporal patterns and key features, providing essential information about student performance. The LSTM model with attention mechanisms achieved an impressive accuracy of 98.32%, precision and recall of over 98%, and an AUC of 99.68%, outperforming previous studies that utilized traditional machine learning techniques. The novelty of this research lies in its ability to leverage temporal dependencies and periodic behaviors of learners, which previous models often overlooked. By integrating attention mechanisms, the model highlights significant features, such as quiz components and assignment submissions, allowing for more precise and actionable predictions. This innovation enables educators to identify at-risk learners earlier and design personalized interventions, enhancing learner support and retention on virtual learning platforms.

However, the model's performance is heavily reliant on the quality and comprehensiveness of the input data, and the implementation is computationally intensive. Future improvements could focus on enhancing scalability, interpretability, and integration with various virtual learning environments. Open questions remain about adapting the model to different educational contexts and understanding the long-term impacts on learners' outcomes. The prospects for applying LSTM networks with attention mechanisms in educational predictive analytics are vast, with future research needed to refine these models and explore the integration of other deep learning techniques. This study contributes to the theoretical understanding of using deep learning in educational settings, providing a foundation for future research and motivating other researchers to explore similar approaches. The quantitative data and insights highlight the potential of these techniques to revolutionize educational predictive analytics, paving the way for more personalized and effective interventions to support student success.

4.1 Pros and Cons of the Developed At-Risk Learner Detection Method using LSTM

The developed method for at-risk learner detection employs a Long Short-Term Memory (LSTM) model augmented with an attention mechanism. This model is tailored to capture the temporal dependencies and periodic behaviors of learners in online learning environments. While the approach shows promise, it has both strengths and limitations.

Pros:

- Capturing Long-Term Dependencies: The LSTM architecture is well-suited to handle sequential data and long-term dependencies, which are crucial for analyzing student engagement over time.
- Use of Attention Mechanism: The integration of an attention mechanism improves interpretability by enabling the model to dynamically focus on the most critical time periods and features.
- High Predictive Accuracy: According to the paper, the model achieves an impressive predictive accuracy of 98.32%, with high precision and recall. This indicates the method’s robustness in distinguishing between at-risk and non-at-risk learners, making it a reliable tool for early intervention.
- Effective for Periodic Engagement Analysis: The model is particularly effective at capturing the periodic nature of student activities, such as weekly quizzes and assignment submissions.
- Automated Feature Learning: Unlike traditional models, which require manual feature selection and engineering, lstms can automatically learn which temporal aspects are most relevant.
- Scalability Across Different Courses and Platforms: The model’s flexibility makes it adaptable to various online courses and educational platforms.
- Early Detection and Timely Interventions: The LSTM model with attention is capable of providing early alerts, allowing educational institutions to intervene before a learner falls significantly behind.

Cons:

- Lack of Interpretability: Despite the inclusion of attention mechanisms, the overall interpretability of the LSTM model remains limited due to their complex architecture.
- Sensitivity to Data Quality and Quantity: The model’s performance heavily depends on the quality and comprehensiveness of the input data. If the data lacks sufficient granularity or contains missing entries, the LSTM might fail to capture essential patterns, leading to unreliable predictions.
- Potential for Overfitting: With a high number of LSTM units and attention layers, there is a risk of overfitting, especially if the training data is not sufficiently representative of real-world scenarios.
- Computational Complexity: LSTM models, especially when combined with attention mechanisms, are computationally intensive and require significant resources for training and deployment.

- Limited Insight into Causal Factors: While the model is effective at predicting which students are at-risk, it does not provide much insight into the why. This lack of causal understanding limits the ability of educators to design targeted interventions based on specific underlying issues.
- Dependence on Temporal Regularity: The LSTM model performs best when there is a regular temporal pattern in the data. If students exhibit highly irregular or sporadic engagement patterns, the model’s predictive power may degrade.
- Scalability Challenges for Real-Time Analytics: For courses that require real-time monitoring of student engagement, the LSTM model might be too resource-intensive to deploy in a real-time setting.

4.2 Limitations

The LSTM model has various limitations like Interpretability Issues and Potential Biases in Data. Addressing these limitations we can apply a combination of techniques such as explainable AI (XAI) methods to improve interpretability, careful feature engineering, and strategies like re-sampling or weighting to mitigate data biases as future work. The developed LSTM with attention mechanism provides a powerful and accurate method for detecting at-risk learners in online learning environments, with significant strengths in capturing temporal dependencies and periodic engagement patterns. However, challenges related to interpretability, data dependency, computational complexity, and bias must be addressed to ensure that the model is both effective and ethically sound for broader implementation. Future work could focus on improving the model’s interpretability and robustness, as well as developing strategies to handle irregular engagement patterns more effectively.

REFERENCES

[1] H. Waheed, S. U. Hassan, R. Nawaz, N. R. Aljohani, G. Chen, and D. Gasevic, “Early prediction of learners at risk in self-paced education: A neural network approach,” *Expert Syst Appl*, vol. 213, Mar. 2023, doi: 10.1016/j.eswa.2022.118868.

[2] D. Liu, Y. Zhang, J. Zhang, Q. Li, C. Zhang, and Y. Yin, “Multiple features fusion attention mechanism enhanced deep knowledge tracing for student performance prediction,” *IEEE Access*, vol. 8, pp. 194894–194903, 2020, doi: 10.1109/ACCESS.2020.3033200.

[3] Ş. Aydoğdu, “Predicting student final performance using artificial neural networks in online learning environments,” *Educ Inf Technol (Dordr)*, vol. 25, no. 3, pp. 1913–1927, May 2020, doi: 10.1007/s10639-019-10053-x.

[4] Y. Ren, S. Huang, and Y. Zhou, “Deep learning and integrated learning for predicting student’s withdrawal behavior in MOOC,” in *Proceedings - 2021 2nd International Conference on Education, Knowledge and Information Management, ICEKIM 2021*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 81–84. doi: 10.1109/ICEKIM52309.2021.00026.

[5] A. Bouamrane and H. Zarzour, “A Deep Learning-based System for Detection At-Risk Students,” in *4th International Conference on Pattern Analysis and Intelligent Systems, PAIS 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/PAIS56586.2022.9946878.

[6] A. S. Aljaloud *et al.*, “A Deep Learning Model to Predict Student Learning Outcomes in LMS Using CNN and LSTM,” *IEEE Access*, vol. 10, pp. 85255–85265, 2022, doi: 10.1109/ACCESS.2022.3196784.

[7] A. Al-Sulami, M. Al-Masre, and N. Al-Malki, “Deep Learning to Predict At-Risk Students’ Achievement in a Preparatory-year English Courses,” in *1st International Conference in Advanced Innovation on Smart City, ICAISC 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICAISC56366.2023.10085097.

[8] T. Motaung, K. Padayachee, and R. Ajoodha, “Predicting Students At-Risk Using Deep Learning Neural Network: A comparison of performance of different models,” in *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICECCME55909.2022.9988117.

[9] T. Y. Yang, C. G. Brinton, C. Joe-Wong, and M. Chiang, “Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 716–728, Aug. 2017, doi: 10.1109/JSTSP.2017.2700227.

[10] Y. Zhou, K. Niu, H. Lv, G. Lu, and Y. Pan, “CGDC- LSTM: A novel hybrid neural network model for MOOC dropout prediction,” in *Proceedings of the International Joint Conference on Neural Networks*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/IJCNN54540.2023.10191794.

[11] K. Talebi, Z. Torabi, and N. Daneshpour, “Ensemble models based on CNN and LSTM for dropout prediction in MOOC,” *Expert Syst Appl*, vol. 235, Jan. 2024, doi: 10.1016/j.eswa.2023.121187.

[12] M. Skittou, M. Merrouchi, and T. Gadi, “Development of an Early Warning System to Support Educational Planning Process by Identifying At-Risk Students,” *IEEE Access*, pp. 1–1, Dec. 2023, doi: 10.1109/access.2023.3348091.

[13] E. S. Karabacak and Y. Yaslan, “Comparison of Machine Learning Methods for Early Detection of Student Dropouts,” in *UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering*,

Institute of Electrical and Electronics Engineers Inc., 2023, pp. 376–381. doi: 10.1109/UBMK59864.2023.10286747.
[14] Chenglong Zhang, Hyunchul Ahn," E-Learning at-Risk Group Prediction Considering the Semester and Realistic Factors", *Educ. Sci.* 2023, *13*(11), 1130; https://doi.org/10.3390/educsci13111130
[15] Hajra Waheed, Saeed-Ul Hassan, Raheel Nawaz, aif Aljohani, Guanliang Chen, Dragan Gasevic. "Early prediction of learners at risk in self-paced education: A neural network approach", ACM, Volume 213, Issue PA, https://doi.org/10.1016/j.eswa.2022.118868
[16] PanelHajra Waheed, Saeed-Ul Hassan, Raheel Nawaz, Naif Johani, Guanliang Chen, Dragan, " Early prediction of learners at risk in self-paced education: A neural network approach", Expert Systems with Applications, Volume 213, Part A, 1 March 2023
[17] T.A. Kustitskaya, A.A. Kytmanov, M.V. Noskov, "Early student-at-risk detection by current learning performance and learning behavior indicators", Cybernetics and Information Technology, 22(1) 2022, pp. 117-133, 10.2478/cait-2022-0008
[18] L. Bognár, T. Fauszt, B. Nagy, " Machine learning model building techniques for small and medium-sized university courses" International Journal of Artificial Intelligence, 19 (2) (2021), pp. 20-43
[19] S. Van Goidsenhoven, D. Bogdanova, G. Deeva, S.v. Broucke, J. De Weerdt, M. Snoeck "Predicting student success in a blended learning environment", Proceedings of the tenth international conference on learning analytics & knowledge (2020), pp. 17-25, 10.1145/3375462.3375494
[14] N. S. K. M. K. Tirumanadham, T. S, and S. M, 'Evaluating Boosting Algorithms for Academic Performance Prediction in E-Learning Environments', in 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), 2024, pp. 1–8.

BIOGRAPHY OF AUTHORS

Ms. Smruti Nanavaty is a Research Scholar at Poornima Univeristy, Jaipur. Her area of interest include data analytics and its applications in the education sector. Her rescent research articles are based in the area of learning analytics using data in virtual environments.

Dr Ajay Khunteta is an accomplished professional with a rich background in academia and research, spanning over 22 years. Currently serving as a Professor and Dean in the Faculty of Computer Science & Engineering at Poornima University in Jaipur, Rajasthan, India, he has made significant contributions to the field. His expertise lies in the areas of AI, Cloud Computing, Data Security, and Distributed Systems. With a passion for guiding and mentoring students, Dr. Ajay Khunteta has successfully supervised 46 M.Tech dissertations and 6 PhD theses on various topics related to AI and Cloud technology.