BiLSTM-Attention-Delta: A Novel Framework for Predicting Dropout in MOOCs Within Big Data Environments

Thu Nguyen[1,2][a], Hong-Tri Nguyen[3][b] and Tu-Anh Nguyen-Hoang[1,2][c]

[1]*Faculty of Information Science and Engineering, University of Information Technology, Ho Chi Minh City, Vietnam*

[2]*Vietnam National University, Ho Chi Minh City, Vietnam*

[3]*Aalto University, Finland*

fi

Keywords: Predicting Dropout, MOOCs, Big Data Environments, Big Data Architecture, Neural Networks, AI.

Abstract: The high dropout rate on online education platforms like MOOCs is a significant challenge for modern education systems. This wastes resources and diminishes the course's credibility, impacting educational goals and limiting learners' personal development opportunities. Research on predicting dropout rates in MOOCs has achieved significant milestones, with effective predictive models and analysis of influencing factors to reduce dropout rates. However, challenges remain in ensuring data quality, safeguarding personal information, enhancing model interpretability, and addressing implementation difficulties, especially in the context of big data. This study focuses on analyzing big data to develop an AI-powered intelligent education system capable of monitoring and predicting student learning behavior to reduce dropout rates, while also personalizing the learning process and improving the learner's experience. However, the process of extracting big data from MOOCs poses numerous challenges, including ensuring data quality, integrity, and the ability to handle diverse and massive data. Model interpretability and deployment are also complex, requiring rigorous technical solutions and data management to optimize learning quality and experience. To tackle data processing and deployment challenges, the study introduces the BiLSTM-Attention-Delta framework. This model improves dropout prediction by over 10% compared to baselines, optimizes training and prediction times, and leverages the Delta big data architecture (BDA) for effective deployment in MOOCs.

1 INTRODUCTION

Large-scale online learning platforms (MOOCs) have grown rapidly, offering learning opportunities to millions worldwide (Rulinawaty et al., 2023). However, the high dropout rate remains a significant challenge (Wang et al., 2023; Mehrabi et al., 2022), as many learners fail to complete the course or remain inactive for extended periods. To address this, AI and big data analytics offer promising solutions by monitoring activities, predicting dropouts, and personalizing learning (Younus et al., 2022; Cao et al., 2020; Zheng et al., 2023). Challenges in extracting and processing MOOC data persist, including ensuring data quality, security, and managing large, diverse datasets (Ang et al., 2020; Bai et al., 2021). Additionally, model interpretability and deployment require advanced technical solutions to optimize the learning experience.

While AI and big data analytics show promising results, challenges in data processing and sustainable model deployment remain. To address these, this study introduces the BiLSTM-Attention-Delta framework for accurate and efficient dropout prediction in big data environments. This framework offers a comprehensive solution to the dropout problem, enhancing the quality and efficiency of MOOC platforms.

This study is structured into five main sections. Section 2 introduces related research on dropout rates in MOOCs and unresolved challenges. Section 3 describes the BiLSTM-Attention-Delta framework and its implementation in the MOOC context to enhance performance. Section 4 provides experimental data and analyzes model effectiveness, and Section 5 summarizes key contributions and future directions for AI applications in online education.

[a] https://orcid.org/0000-0002-0571-2117
[b] https://orcid.org/0000-0001-6483-0829
[c] https://orcid.org/0000-0001-9283-769X

Nguyen, T., Nguyen, H.-T. and Nguyen-Hoang, T.-A.
BiLSTM-Attention-Delta: A Novel Framework for Predicting Dropout in MOOCs Within Big Data Environments.
DOI: 10.5220/0013202900003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2*, pages 228-235
ISBN: 978-989-758-746-7; ISSN: 2184-5026

2 RELATED WORK

The problem of dropout prediction (binary classification problem) in MOOCs has become a significant research topic in the field of online education. MOOCs, with their large-scale student population and high degree of flexibility, often experience alarming dropout rates, making prediction and intervention a top priority (Jeon et al., 2020). The goal of this problem is to predict whether a learner is likely to drop out before completing the course, thereby enabling timely support and intervention to reduce this rate. MOOCs attract thousands to millions of participants per course, yet most fail to complete them (Talebi et al., 2024; Fu et al., 2021). The flexibility of learning often leads to a loss of motivation, while the diverse cultural and educational backgrounds of students add complexity to predicting individual behavior. MOOC platforms gather extensive data (Sakboonyarat and Tantatsanawong, 2022), including interactions with content (logins, video watch time, assignments submitted, materials downloaded), personal information (age, gender, occupation), community engagement (forum participation), and technical details (device, study time). Leveraging this data is crucial for understanding dropouts and implementing effective interventions.

Predicting student dropout is a critical challenge as educational institutions increasingly use data analytics to improve outcomes and completion rates. Methods range from traditional techniques like decision trees (DT) (Pereira and Zambrano, 2017) and logistic regression (LR) (Cuji Chacha et al., 2020) to advanced models like CNN-LSTM (Talebi et al., 2024) and Multi-layer Perceptron (MLP) (Jeon et al., 2020). With growing datasets, selecting appropriate methods for accurate predictions is crucial. While studies offer diverse approaches, handling large, complex datasets remains a key challenge requiring further research. Our study addresses this issue, with the details summarized in comparison to previous baseline methods in Table 1.

Research on dropout prediction in large data environments has made significant contributions but has also revealed limitations that need to be addressed to enhance practical application effectiveness. Traditional models, such as DT (Pereira and Zambrano, 2017) and LR (Cuji Chacha et al., 2020), offer interpretability and ease of application but are inadequate for scaling with large and complex data, especially unstructured and time-series data. In contrast, more complex models like CLSA (Fu et al., 2021) or CNN-LSTM hybrid model (Talebi et al., 2024) better leverage temporal and spatial information from MOOCs data but face challenges related to computational complexity, requiring substantial resources and lacking interpretability. (Jeon et al., 2020) introduced a MLP learning model to efficiently process clickstream data but still struggled with scalability for big data. Notably, Apache Spark-based model (Sakboonyarat and Tantatsanawong, 2022) demonstrated potential in handling large data efficiently but focused on course recommendation rather than comprehensively addressing dropout prediction.

Recognizing the lack of holistic features in related works, this study focuses on resolving issues related to scalability and computational resource optimization in dropout prediction. To achieve this, the study proposes the BiLSTM-Attention-Delta framework, consisting of two main components: the BiLSTM-Attention model and the Delta BDA. The BiLSTM-Attention model, utilizing a sequential neural network, predicts dropout behavior on MOOC platforms, improving performance by over 10% compared to baseline methods. Training time is reduced by 13 times, and prediction time by 5 times, making it suitable for large-scale data processing systems. The Delta BDA supports large-scale model deployment and optimizes data management, contributing to the improvement of MOOC platforms' quality and efficiency.

3 BiLSTM-ATTENTION-DELTA FRAMEWORK

3.1 The Proposed BiLSTM-Attention Model for Dropout Prediction

The proposed BiLSTM-Attention model used for dropout prediction in Figure 1 consists of three main parts: preprocessing, BiLSTM layers, and the Bahdanau Attention layer combined with a neural network. In the preprocessing phase, the input data is processed similarly to the previous study (Talebi et al., 2024), with the addition of several elements, such as integrating a vector representing the total time for each behavior. Course duration data was collected over a maximum period of 30 days, and dropout prediction was performed at different time points. For each registration with ID Q on day $t \leq T$, there will be a vector representing the learner's behavior, as shown in Equation 1.

$$x = \left[a_t^{(1)}, a_t^{(2)}, a_t^{(3)}, \ldots, a_t^{(7)}\right] \tag{1}$$

In terms of dimensions, $x \in \mathbb{R}^7$, corresponding to the number of behavior types a, representing the fre-

Table 1: Comparison of Baseline Methods and Proposed Approach.

Research studies	**Scope**	**Objectives**	**Technology**	**Data**	**Contribution**
(Pereira and Zambrano, 2017)	The prediction of student dropout at the University of Nariño, Colombia.	Identifying risk factors contributing to student dropout.	DT	Student profile data from 2004 to 2006 includes socioeconomic and academic information.	Providing information to develop appropriate intervention policies.
(Cuji Chacha et al., 2020)	The prediction of student dropout at the Northern Technical University, Ecuador.	Developing a dropout prediction model based on LR.	LR	Data from the academic information system (demographic and academic).	Providing a model to support the identification of students at risk of dropping out.
(Jeon et al., 2020)	Predicting dropout in MOOCs based on clickstream data.	Weekly dropout prediction in MOOCs based on interaction data.	MLP learning: Branch and Bound.	Clickstream data from Coursera courses.	The representation learning method provides interpretable and efficient results.
(Fu et al., 2021)	Predicting MOOCs dropout using the CLSA model.	Developing a deep learning model for predicting dropout in MOOCs.	CNN, LSTM, Self-Attention.	Data from the XuetangX platform (interactions, study time).	The proposed CLSA architecture is more efficient than traditional models.
(Sakboonyarat and Tantat-sanawong, 2022)	Applying big data technology to propose personalized MOOCs.	Developing a personalized MOOC recommendation system based on big data.	Apache Spark, Kappa Architecture, Spark SQL, Spark MLlib.	MOOCs data: courses, user profiles, interaction history.	A personalized course recommendation system with the capability to handle large volumes of data.
(Talebi et al., 2024)	Predicting dropout in MOOCs using a CNN-LSTM model.	Developing an accurate dropout prediction model from MOOCs data.	CNN-LSTM hybrid model.	Data from large MOOCs (interactions, study time, number of clicks).	Combining spatial and temporal information to enhance prediction accuracy.
Our	The research focuses on the issue of predicting student dropout in MOOC platforms.	Developing an intelligent system using AI and big data analytics to predict and reduce dropout rates on MOOC platforms.	The BiLSTM-Attention-Delta framework with the BiLSTM-Attention Model and the Delta BDA.	Big data from MOOC platforms, including student interactions and learning behavior.	Providing a comprehensive solution, and contributing to the field of online education.

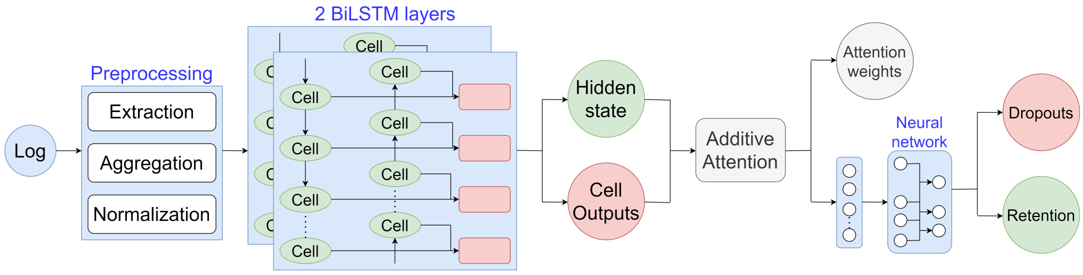

Figure 1: BiLSTM-Attention Model.

quency of each behavior recorded on that day. T is the number of days in the weeks; for example, if the number of weeks used for prediction is 3 weeks, then $T = 21$, so $T = 7, 14, 21, 28, 30$.

The BiLSTM layers (Anand et al., 2023) process data in two stages: the first layer captures daily learning behaviors (timesteps), while the second extracts forward and backward hidden states from each timestep. This information is then passed to the Bahdanau Attention layer, which highlights the most important features in the time sequence. The Attention layer (Itti et al., 1998; Bahdanau, 2014) computes attention scores between BiLSTM input vectors at each timestep and its final hidden state. The softmax function normalizes these scores into attention weights, creating a weighted vector. The final output aggregates weighted features across timesteps, enhancing the model's ability to identify key factors and predict dropouts accurately. The attention calculation steps are detailed in formulas 2, 3, 4, and 5.

$$\text{score}_{ij} = \tanh(W_i h_i + W_g g) \tag{2}$$

$$e_i = \mathbf{v}_a^T \cdot \text{score}_i \tag{3}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_k \exp(e_k)} \tag{4}$$

$$c = \sum_i \alpha_i h_i \tag{5}$$

Where: (1) score_{ij}: The attention score between input feature vector h_i at timestep i and the final hidden state g of the BiLSTM. It assesses the relevance between h_i and g, using weight matrices W_i, W_g, and the tanh function to normalize the output between -1 and 1. (2) e_i: The unnormalized attention score for each timestep i, representing the correlation between the weight vector v_a^T and score_i. It is calculated before applying the softmax function, reflecting the priority of information from timestep i. (3) α_i: The normalized attention weight from the softmax function of e_i, indicating the importance of timestep i. α_i values range from 0 to 1 and sum to 1, representing the model's focus on each timestep. (4) c: The context vector, calculated as the weighted sum of feature vectors h_i using attention weights α_i. It aggregates key information from the input sequence and serves as input for subsequent layers to make predictions.

The context vector is fed into a three-layer neural network, with the final layer using a sigmoid activation function (Equation 6) for binary classification: dropouts (1) or retention (0). The sigmoid maps input x to a probability between 0 and 1, predicting the positive class if the probability exceeds 0.5, and the negative class otherwise.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{6}$$

3.2 Our Framework

The BiLSTM-Attention-Delta framework (Figure 2) addresses high dropout rates on MOOC platforms. By integrating Delta BDA, this framework efficiently manages and processes massive volumes of data derived from users' historical activities and real-time interactions. It is particularly well-suited for educational data, encompassing both historical and real-time datasets. This enables the BiLSTM-Attention model to deliver fast, accurate predictions and real-time solutions, optimizing learning processes and reducing dropout rates. Deploying a BDA on a cloud platform is essential for its scalability, cost-efficiency, and support for distributed processing (Zbakh et al., 2019). This study uses Microsoft Azure Cloud Service, known for its high performance in handling large-scale data processing and management (Ang et al., 2020).

The Ingest component in the architecture (Figure 2) uses Azure Data Factory for batch processing of historical data (e.g., learner profiles, course details) and Azure Event Hubs for real-time streaming data (e.g., activity logs). This setup ensures efficient integration and processing of diverse data types. The Storage component in the architecture (Figure 2) uses a Lakehouse architecture, combining Data Warehouse

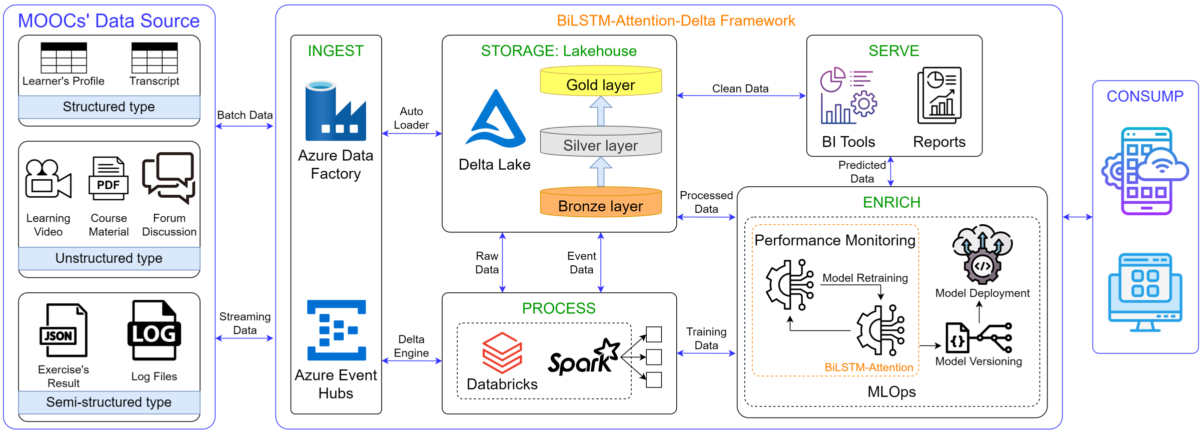

Figure 2: BiLSTM-Attention-Delta Framework.

and Data Lake strengths for centralized, reliable, high-performance data storage. Delta Lake (Armbrust et al., 2020) is recommended, utilizing a Medallion architecture with Bronze, Silver, and Gold layers: (1) Bronze Layer: Stores raw, unprocessed data. (2) Silver Layer: Cleans and transforms data, improving quality. (3) Gold Layer: Contains high-quality, processed data for analysis and decision-making. This structure ensures efficient data management and supports diverse analytical needs. The Process component (Figure 2) handles streaming data, scheduled tasks, and event-based triggers. Pre-configured Azure Databricks (Pala, 2021) jobs retrieve raw data from the Bronze layer, perform integration, transformation, and cleaning, and load refined datasets into the Silver and Gold layers using ACID transactions. Batch data is processed via Azure Data Factory, which extracts, aggregates, and stores raw data in Delta Lake tables. Streaming data from MOOC platforms is collected via Event Hubs, secured with OAuth 2.0, and processed using Azure Databricks with Delta Engine and Spark Streaming (Armbrust et al., 2018). All data is stored in Delta Lake for further processing. The Enrich component (Figure 2) leverages AI and ML for building, versioning, and deploying models, streamlining workflows while ensuring consistency and performance. It deploys the BiLSTM-Attention model using MLOps (Kreuzberger et al., 2023), which automates lifecycle management, including monitoring, retraining, testing, and redeployment when performance drops. The component integrates with MLOps platforms like MLflow to handle experiment tracking, version control, automated deployment, and performance monitoring, ensuring models remain up-to-date, robust, and ready for real-time applications. The Serve and Consumption component (Figure 2) prepares processed data from the Bronze and Silver layers for analysis, storing it in the Gold layer in delta format. This format supports version control, real-time data handling, schema evolution, and audit logging. A Databricks-based dashboard will visualize near real-time data using Spark Structured Streaming. Dropout predictions from the BiLSTM-Attention model will integrate with BI tools and smart education platforms, enabling activity monitoring, behavior prediction, and personalized learning.

The BiLSTM-Attention-Delta framework offers a powerful solution for reducing dropout rates on online education platforms by leveraging the large-scale data processing capabilities of the Delta architecture. By combining historical and real-time data, the proposed BiLSTM-Attention model enables accurate and efficient early prediction of dropout risks. This allows for the timely implementation of tailored strategies to support learners and reduce dropout rates. This framework represents a significant advancement in enhancing the learning experience and addressing the dropout challenge in MOOCs.

4 EXPERIMENTS

4.1 Experimental Setup

In dropout prediction, evaluation metrics are critical for ensuring model accuracy and suitability, especially with complex, imbalanced data. Most dropouts occur in the first two weeks, requiring high sensitivity to early signals. Key metrics include: (1) Recall: Ensures the model identifies most dropouts, critical for timely intervention. (2) F1-score: Balances Precision and Recall, minimizing false alarms while accurately detecting dropouts, especially with imbalanced data. Additionally, training and prediction time are vital for large-scale systems, ensuring the model is both accurate and efficient for real-time applications in dynamic educational environments.

The dataset used for dropout prediction in all experiments of this study, collected from the XuetangX MOOC platform (August 2015–August 2017) (Feng et al., 2019), includes 89 million clickstream records from 254,518 learners across 698 courses, with 467,113 enrollments. It covers seven interaction types: access, discussion, navigation, page_close, problem-solving, video viewing, and wiki search. Among 225,642 labeled samples, 171,133 (dropouts, "1") and 54,509 (retained, "0") highlight a dropout-to-retention ratio of 3.4:1, reflecting significant data imbalance. This dataset is vital for training prediction models, requiring advanced methods to handle its complexity and imbalance. Analyzing interaction trends over time provides insights into learning behaviors, improving course quality, personalizing learning, and identifying at-risk students, as illustrated in Figure 3.

The first chart in Figure 3 shows the highest interaction levels in the first week for both groups. In this week, the dropout group ("1") had more interactions than the completion group ("0") due to a higher number of users. From the second week, dropout group interactions drop sharply, while the completion group declines by only 5–20% weekly, indicating rapid disengagement among dropouts. The second chart reveals that, despite more users in the dropout group, the completion group averaged 3–10 times more weekly actions during the first three weeks, emphasizing the importance of sustained engagement for completion. Over 70% of dropouts occur within the first two weeks, with engagement levels plummeting thereafter. To predict dropouts effectively, models must detect early warning signs, particularly within the first two weeks. Techniques like Attention in BiLSTM-Attention models can focus on critical early behaviors to enhance accuracy. This analysis underscores the importance of capturing early trends, with the BiLSTM-Attention model designed to leverage these features for better predictions.

Details of the implementation of the proposed method along with the baseline methods are presented in Table 2. A summary of the key points and contributions of these baseline methods is presented in Table 1. Our experimental process was conducted in the Google Colab environment, equipped with 2vCPU Intel(R) Xeon(R) @ 2.20GHz, 13GB RAM, and a Tesla T4 GPU. The dataset was divided into three parts: 70% for the training set, 15% for the validation set, and 15% for the test set. We applied the early stopping technique to monitor and save the checkpoint with the highest F1-score on the training set. The training process was carried out over 20 epochs with a batch size of 100, using the Adam optimizer. Notably, the evaluation was performed over the first two weeks of the course, as more than 70% of dropout cases occur during this period. Additionally, the Delta BDA was implemented to facilitate the data collection and processing on various MOOCs platforms. This architecture was deployed on *Azure Cloud Services*, leveraging cloud computing benefits such as flexible scalability, automated workflows, high availability, and reduced infrastructure costs compared to on-premise deployment.

4.2 Results and Discussion

Table 3 compares various dropout prediction methods based on criteria such as Recall, F1-score, training time, prediction time, and deployment on a BDA.

Baseline methods like DT, LR, MLP, CNN, CLSA, LSTM, and Bagging-CNN-LSTM struggle with large-scale time-series data due to limitations in handling temporal relationships and resource efficiency. DT and LR face overfitting and linear assumption challenges, while MLP and CNN fail to capture sequential patterns effectively. Advanced models like CLSA, LSTM, and Bagging-CNN-LSTM are resource-intensive and prone to overfitting or optimization issues. To address these, this study proposes the BiLSTM-Attention-Delta method, combining BiLSTM and Attention to focus on key features, reduce unnecessary data, minimize overfitting, and optimize processing time.

The BiLSTM-Attention-Delta method outperforms others in dropout prediction with high accuracy and stability. It achieves the highest F1-score in the first week (0.7344, over 10% improvement) and a slight drop in the second week (0.7716), demonstrating stability. Its recall is also among the highest, particularly in the second week (0.8281), with consistent performance across both weeks. In terms of time, BiLSTM-Attention-Delta has a much faster training speed compared to complex models like Bagging-CNN-LSTM, taking only 171 seconds compared to 2324 seconds for the complex model, a reduction in training time by more than 13 times. Additionally, the model's prediction time is 2.9 seconds compared to 14.9 seconds, a reduction by more than 5 times. While the training time is longer compared to simpler methods such as DT or LR, the prediction performance is significantly higher. Additionally, when compared to baseline methods such as DT and Bagging-CNN-LSTM, BiLSTM-Attention-Delta demonstrates superiority in both F1 and Recall, proving that the combination of Attention with

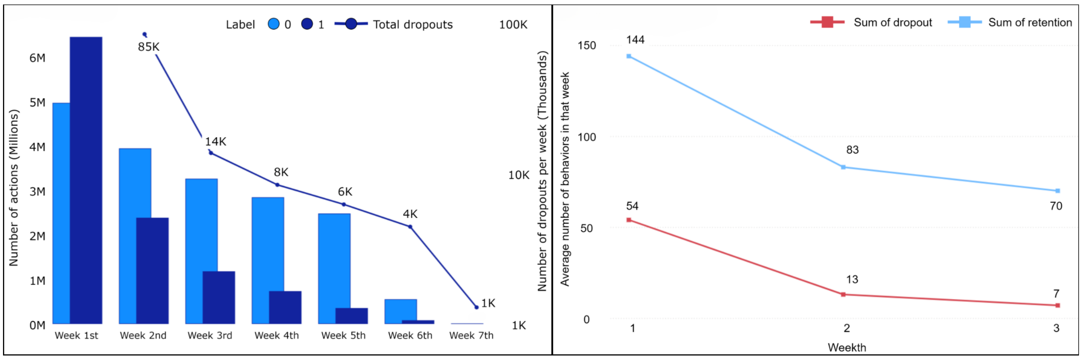

Figure 3: Statistical Results of Experimental Data.

Table 2: Experimental methods for the dropout prediction problem.

Methods	**Neural network**	**CNN-based**	**LSTM-based**	**Attention mechanism**	**BDA**
DT (Pereira and Zambrano, 2017)	No	No	No	No	No
LR (Cuji Chacha et al., 2020)	No	No	No	No	No
MLP (Jeon et al., 2020)	Yes	No	No	No	No
CNN (Talebi et al., 2024)	Yes	Yes	No	No	No
CLSA (Fu et al., 2021)	Yes	Yes	Yes	Yes	No
LSTM (Talebi et al., 2024)	Yes	No	Yes	No	No
Bagging-CNN-LSTM (Talebi et al., 2024)	Yes	Yes	Yes	No	No
BiLSTM-Attention-Delta	Yes	No	Yes	Yes	Yes

Table 3: The experimental results for the dropout prediction problem in the first two weeks.

Methods	**Recall (1st)**	**Recall (2nd)**	**F1 (1st)**	**F1 (2nd)**	**Training time (s)**	**Prediction time (s)**	**BDA**
DT (Pereira and Zambrano, 2017)	0.6605	0.6807	0.6394	0.6614	5	0.1	No
LR (Cuji Chacha et al., 2020)	0.8250	0.8566	0.7189	0.7478	4	0.1	No
MLP (Jeon et al., 2020)	0.6947	0.7281	0.7039	0.7417	31	1.0	No
CNN (Talebi et al., 2024)	0.7942	0.8106	0.7290	0.7612	154	0.9	No
CLSA (Fu et al., 2021)	0.6979	0.8598	0.7112	0.7726	307	2.2	No
LSTM (Talebi et al., 2024)	0.7803	0.8335	0.7320	0.7699	573	4.3	No
Bagging-CNN-LSTM (Talebi et al., 2024)	0.7090	0.7409	0.7183	0.7634	2324	14.9	No
BiLSTM-Attention-Delta	**0.7906**	**0.8281**	**0.7344**	**0.7716**	**171**	**2.9**	**Delta**

BiLSTM yields significant improvements in prediction and accurate classification. Finally, BiLSTM-Attention-Delta has been deployed on a BDA, ensuring stability and the ability to handle large volumes of data, further reinforcing the model's feasibility in practical applications. Thanks to its well-balanced accuracy, stability, and processing speed, BiLSTM-Attention-Delta stands out as the optimal method for this problem.

5 CONCLUSION

The rapid growth of MOOCs and online education platforms offers global learning opportunities but faces the challenge of high dropout rates. This study introduces the BiLSTM-Attention-Delta framework, designed for accurate and efficient dropout prediction in big data environments. The BiLSTM-Attention model improves performance by over 10% compared to baseline methods while significantly reducing training and prediction times, making it ideal for large-scale data. Supported by the Delta BDA, it ensures efficient deployment in MOOC environments. This research enhances online education and big data analytics by addressing dropout issues and improving the quality of MOOCs.

ACKNOWLEDGEMENTS

This research is funded by University of Information Technology-Vietnam National University Ho Chi Minh City under grant number D1-2024-69.

REFERENCES

Anand, G., Kumari, S., and Pulle, R. (2023). Fractional-iterative bilstm classifier: A novel approach to predicting student attrition in digital academia. *SSRG International Journal of Computer Science and Engineering*, 10(5):1–9.

Ang, K. L.-M., Ge, F. L., and Seng, K. P. (2020). Big educational data & analytics: Survey, architecture and challenges. *IEEE access*, 8:116392–116414.

Armbrust, M., Das, T., Sun, L., Yavuz, B., Zhu, S., Murthy, M., Torres, J., van Hovell, H., Ionescu, A., Łuszczak, A., et al. (2020). Delta lake: high-performance acid table storage over cloud object stores. *Proceedings of the VLDB Endowment*, 13(12):3411–3424.

Armbrust, M., Das, T., Torres, J., Yavuz, B., Zhu, S., Xin, R., Ghodsi, A., Stoica, I., and Zaharia, M. (2018). Structured streaming: A declarative api for real-time applications in apache spark. In *Proceedings of the 2018 International Conference on Management of Data*, pages 601–613.

Bahdanau, D. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bai, X., Zhang, F., Li, J., Guo, T., Aziz, A., Jin, A., and Xia, F. (2021). Educational big data: Predictions, applications and challenges. *Big Data Research*, 26:100270.

Cao, W., Wang, Q., Sbeih, A., and Shibly, F. (2020). Artificial intelligence based efficient smart learning framework for education platform. *Inteligencia Artificial*, 23(66):112–123.

Cuji Chacha, B. R., Gavilanes López, W. L., Vicente Guerrero, V. X., and Villacis Villacis, W. G. (2020). Student dropout model based on logistic regression. In *Applied Technologies: First International Conference, ICAT 2019, Quito, Ecuador, December 3–5, 2019, Proceedings, Part II 1*, pages 321–333. Springer.

Feng, W., Tang, J., and Liu, T. X. (2019). Understanding dropouts in moocs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 517–524.

Fu, Q., Gao, Z., Zhou, J., and Zheng, Y. (2021). Clsa: A novel deep learning model for mooc dropout prediction. *Computers & Electrical Engineering*, 94:107315.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259.

Jeon, B., Park, N., and Bang, S. (2020). Dropout prediction over weeks in moocs via interpretable multi-layer representation learning. *arXiv preprint arXiv:2002.01598*.

Kreuzberger, D., Kühl, N., and Hirschl, S. (2023). Machine learning operations (mlops): Overview, definition, and architecture. *IEEE access*, 11:31866–31879.

Mehrabi, M., Safarpour, A. R., and Keshtkar, A. (2022). Massive open online courses (moocs) dropout rate in the world: a protocol for systematic review and meta-analysis. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, 13(2):85–92.

Pala, S. K. (2021). Databricks analytics: Empowering data processing, machine learning and real-time analytics. *Machine Learning*, 10(1).

Pereira, R. T. and Zambrano, J. C. (2017). Application of decision trees for detection of student dropout profiles. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)*, pages 528–531. IEEE.

Rulinawaty, R., Priyanto, A., Kuncoro, S., Rahmawaty, D., and Wijaya, A. (2023). Massive open online courses (moocs) as catalysts of change in education during unprecedented times: A narrative review. *Jurnal Penelitian Pendidikan IPA*, 9(SpecialIssue):53–63.

Sakboonyarat, S. and Tantatsanawong, P. (2022). Applied big data technique and deep learning for massive open online courses (moocs) recommendation system. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, 16(4):436–447.

Talebi, K., Torabi, Z., and Daneshpour, N. (2024). Ensemble models based on cnn and lstm for dropout prediction in mooc. *Expert Systems with Applications*, 235:121187.

Wang, W., Zhao, Y., Wu, Y. J., and Goh, M. (2023). Factors of dropout from moocs: a bibliometric review. *Library Hi Tech*, 41(2):432–453.

Younus, A. M., Abumandil, M. S., Gangwar, V. P., and Gupta, S. K. (2022). Ai-based smart education system for a smart city using an improved self-adaptive leap-frogging algorithm. In *AI-Centric Smart City Ecosystems*, pages 231–245. CRC Press.

Zbakh, M., Essaaidi, M., Manneback, P., and Rong, C. (2019). *Cloud Computing and Big Data: Technologies, Applications and Security*. Springer.

Zheng, L., Wang, C., Chen, X., Song, Y., Meng, Z., and Zhang, R. (2023). Evolutionary machine learning builds smart education big data platform: Data-driven higher education. *Applied Soft Computing*, 136:110114.