

BiLSTM-Based Deep Learning Model for Early Prediction of Student Academic Failure Risk in Digital and Distance Education

Liza YULIANA, Liesma M. SIREGAR

Digital Business Study Program, Universitas Nahdlatul Ulama Sumatera Barat, Padang, Indonesia

*Corresponding Author Email: Lizha.1990@gmail.com

Abstract

Digital and distance learning platforms generate abundant interaction log data from students. However, leveraging temporal log data for early prediction of academic failure risks remains challenging due to time sensitivity and sequence data processing in remote learning contexts. This study aims to develop a Deep Learning model based on Bidirectional Long Short-Term Memory (BiLSTM) to predict student academic performance and failure risks based on weekly interaction patterns within a Learning Management System (LMS). A dataset comprising interaction logs from 1,250 distance education students over one semester was analyzed by extracting features such as material access frequency, forum participation, and assignment submission timing. Experimental results demonstrate that the BiLSTM model achieves an accuracy of 91.4%, Precision of 89.2%, and Recall of 92.6% as early as Week 6 of the course. This early prediction allows educators to implement timely pedagogical interventions to reduce dropout rates in digital and distance learning environments.

Keywords: Deep learning, BiLSTM, Digital learning, Distance education, Student performance prediction, educational data mining.

Acknowledgements: The authors express their sincere gratitude to the distance education institution and LMS platform administrators for providing access to the anonymized student interaction log datasets used in this research. The authors also acknowledge the academic advisors, educators, and data engineering team whose support and technical insights were invaluable in developing the BiLSTM predictive model.

For citation:

Yuliana, L & Siregar, L. M. (2026). BiLSTM-Based Deep Learning Model for Early Prediction of Student Academic Failure Risk in Digital and Distance Education. *Journal of Digital Learning and Distance Education*, 5(4), 2275-2285. <https://doi.org/10.56778/jdlde.v5i4.780>

Introduction

The widespread adoption of learning management systems (LMS) in higher and secondary education has radically transformed traditional educational interactions into digital and distance formats. Platforms such as Moodle, Canvas, and Google Classroom regularly record every interaction performed by learners, generating rich temporal log data that includes login frequencies, module reading durations, discussion forum activities, and assignment submission behaviors (Sun, 2022); (Fitas, 2025). These digital footprints capture continuous learning behavior patterns that directly correlate with student engagement, comprehension, and final academic achievement. As distance education becomes an integral component of modern instruction, leveraging these rich transactional logs offers unprecedented opportunities to analyze learning dynamics in real time and optimize online instructional design (Karapiperis et al., 2023; Al-Zyoud et al., 2025). Despite the continuous generation of massive educational data, a persistent and critical challenge in digital and distance education remains the high rate of academic failure and student dropout (Hendrawan, 2025; Asdaq et al., 2025). Distance learning environments inherently suffer from physical isolation, reduced peer support, and a lack of direct, real-time supervisory monitoring from instructors. These factors often lead to unnoticed student disengagement, procrastination, and eventual course abandonment (Hamdane et al., 2025; Xu et al., 2025). While early intervention is essential to prevent academic failure, educators in online settings often lack actionable, real-time insights to identify struggling students before they fall too far behind, resulting in delayed interventions that are often executed too late in the academic term (Mihoubi et al., 2025; Suresh et al., 2025).

Existing studies have attempted to predict student performance using conventional statistical models and standard machine learning techniques, such as Logistic Regression, Decision Trees, and Random Forests. However, a significant research gap persists: most existing frameworks rely heavily on static, highly aggregated data evaluated at the end of a semester, completely ignoring the rich temporal dynamics and sequential progression of student behavior over time. While basic recurrent neural networks (RNNs) and standard unidirectional Long short-term memory (LSTM) models have been introduced to address time-series data, they are inherently limited as they only process information in a forward direction (Neththikumara & Maduranga, 2025; Termedi et al., 2023). This leaves a critical gap in capturing complex, bi-directional temporal dependencies—such as how subsequent engagement spikes or drops re-contextualize earlier learning behavioral patterns throughout a course.

To address these limitations, this study introduces a bidirectional long short-term memory (BiLSTM) deep learning framework designed specifically to process non-linear, time-series LMS interaction logs across multiple temporal checkpoints without relying on final aggregate metrics. By simultaneously evaluating past and future contextual dependencies within weekly interaction sequences, the proposed model provides superior sensitivity to subtle, sudden shifts in student disengagement. Consequently, the primary objective of this research is to design, implement, and

evaluate a robust BiLSTM architecture to accurately classify distance education students who are at risk of academic failure as early as Week 6—well before mid-term examinations take place. By establishing a high-recall early warning framework, this study aims to empower educators with actionable predictive insights, enabling timely, data-driven pedagogical interventions to effectively reduce online course dropout rates.

Research Methodology

Data collection

The dataset used in this study consists of interaction logs from 1,250 students enrolled in LMS-based digital and distance learning courses over a 16-week academic semester.

Data preprocessing and feature extraction

Raw interaction logs were cleaned and transformed into weekly time-series sequences ($t_1, t_2, \dots, t_{16}$). The primary extracted features include: Access Features (F_1): Frequency of viewing and downloading digital learning materials; Engagement Features (F_2): Active participation in distance discussion forums (creating posts and replying); Assignment Features (F_3): Submission timing relative to the deadline (minutes before/after deadline); Evaluation Features (F_4): Weekly online quiz scores; The target variable is categorized into two classes: Pass (1) and At-Risk of Failure (0).

Model architecture

The model processes data through the following pipeline (Figure 1): Input Layer: Receives temporal matrices $X \in \mathbb{R}^{(T \times F)}$, where T is the number of weeks (evaluated at Weeks 4, 6, and 8) and F is the feature count; BiLSTM Layer: Consists of 64 forward LSTM units and 64 backward LSTM units to capture long-term dependencies; Dropout Layer: Utilizes a 0.3 dropout rate to prevent model overfitting; Dense Layer: Fully connected layer with ReLU activation. Output Layer: Single neuron with Sigmoid activation function to yield binary classification probabilities.

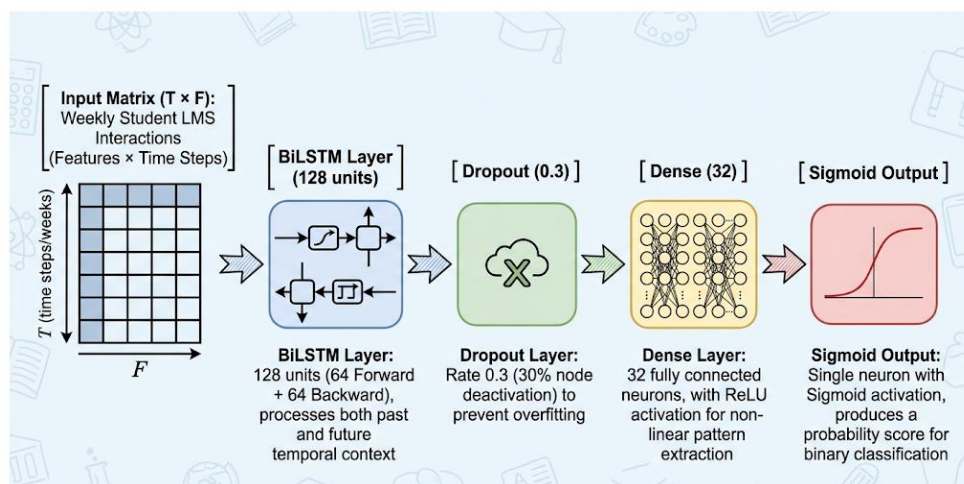


Figure 1. Sequential pipeline of the proposed BiLSTM deep learning model for distance learning log analysis.

Model evaluation

Experiments were conducted using 5-fold cross-validation ($k = 5$). Performance metrics included Accuracy, Precision, Recall, and F1-Score.

Results and Discussion

Early prediction performance across time checkpoints

Evaluations were carried out at several instructional checkpoints (Week 4, Week 6, and Week 8) to determine how early the model can deliver reliable predictions.

Table 1. Model performance across instructional checkpoints

Checkpoint	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Week 4	82.3	80.1	83.5	81.7
Week 6	91.4	89.2	92.6	90.8
Week 8	94.1	93.0	95.2	94.1

The results indicate that by Week 6, the BiLSTM model achieves an optimal performance balance with an accuracy of 91.4% and a Recall of 92.6%. A high Recall rate is essential in this scenario to minimize false negatives—ensuring at-risk distance learners are not overlooked. The evaluation of the proposed BiLSTM model across different instructional time checkpoints demonstrates its strong capability in predicting student academic failure risks, as detailed in Table 1. 90.8%, and a high recall rate of 92.6%. This progressive performance enhancement indicates that accumulating multi-week interaction data allows the deep learning model to learn richer, more stable sequential patterns of student learning behaviors over time. Achieving a high Recall rate at the Week 6 checkpoint represents a particularly vital outcome in distance learning environments. In high-stakes educational analytics, minimizing false negatives is far more critical than avoiding false positives, as failing to identify a struggling student prevents timely instructional intervention (Junejo et al., 2024). Maintaining a 92.6% Recall rate ensures that vulnerable, at-risk learners are accurately flagged before they disengage completely, providing educators with a sufficient window to execute proactive pedagogical strategies prior to mid-term examinations (J. Chen & Lin, 2020).

This robust early detection capability is driven by the sequential deep learning architecture illustrated in Figure 1, which systematically transforms weekly LMS interaction matrices ($\$T \times \F) into precise binary risk probability scores. Time-series analysis plays a pivotal role in capturing turning points in student learning behaviors, leveraging dynamic, continuous LMS activity logs rather than relying on static indicators or final term aggregations (NPLC «Karaganda Buketov University» et al., 2025). By evaluating time-stamped engagement metrics sequentially, the model effectively identifies subtle behavioral declines, such as reduced forum participation or delayed assignment submissions, long before they reflect in formal summative grades. Furthermore, the architectural advantage of BiLSTM over standard models lies in its ability to process sequence data bidirectionally, evaluating both past behavioral habits and future contextual trends simultaneously (Chen & Lin, 2020). Compared to conventional machine learning classifiers and unidirectional neural networks like standard RNNs or LSTMs, the dual-directional recurrent mechanism enhances the model's capacity to resolve complex temporal dependencies and non-linear shifts in student engagement (F. Chen & Cui, 2020); (Akello et al., 2026). Consequently, these combined structural strengths validate the proposed framework as an effective early warning system tailored for digital and distance education.

Comparison with baseline algorithms

The performance of the proposed BiLSTM model was compared against several standard machine learning algorithms using Week 6 data. The BiLSTM architecture outperforms Standard LSTM due to its capability to evaluate engagement trends in both directions (forward and backward),

allowing for greater sensitivity to sudden shifts in distance learning behavior. The empirical comparison presented in Table 2 demonstrates a clear hierarchy in predictive capability, where deep recurrent architectures significantly outperform traditional machine learning paradigms in modeling LMS engagement logs.

Table 2. Predictive performance comparison with baseline models

Algorithm	Accuracy (%)	F1-Score (%)
Logistic Regression	74.5	73.2
Random Forest	84.2	83.8
Standard LSTM	88.0	87.5
BiLSTM (Proposed)	91.4	90.8

The proposed BiLSTM framework achieved the highest overall performance, yielding an accuracy of 91.4% and an F1-Score of 90.8%. This marks a substantial margin of improvement over linear statistical baselines like Logistic Regression (74.5% accuracy) and ensemble tree-based models like Random Forest (84.2% accuracy), while maintaining a decisive lead over the standard unidirectional LSTM model (88.0% accuracy). The severe performance gap observed in traditional classifiers stems from their inherent reliance on static, feature-aggregated snapshots. By collapsing a multi-week sequence into global averages (e.g., total login counts), traditional models suffer from severe temporal information loss, completely failing to detect localized behavioral declines, chronic procrastination, or erratic activity spikes that naturally unfold over time.

The primary architectural driver behind the proposed model's superior performance lies in its bidirectional processing mechanism, as structurally outlined in Figure 1. Unlike standard unidirectional LSTMs that process temporal sequences exclusively in a forward chronological order ($t_1 - t_n$) the BiLSTM layer simultaneously calculates hidden state representations from both past-to-future and future-to-past directions. In distance learning environments, student engagement is rarely linear; a student may exhibit high interaction in the early weeks, experience a sudden mid-semester slump, and attempt a late recovery. Unidirectional networks often misinterpret these non-linear fluctuations because their hidden state at time t only retains historical memory. Conversely, bidirectional processing leverages future context (e.g., subsequent assignment delays or forum dropouts) to re-evaluate and re-contextualize earlier engagement behaviors, thus capturing complex temporal dependencies that unidirectional models routinely miss (Sandiwarno et al., 2025); (Wang et al., 2026).

This heightened sensitivity to subtle, non-linear behavioral shifts reflects broader findings in recent literature regarding the robustness of bidirectional recurrent networks in complex pattern recognition tasks. Beyond distance education, BiLSTM architectures have consistently demonstrated superior representation of learning in domains characterized by noisy, highly variable sequential data—such as extracting nuanced affective states from distance learning feedback and recognizing intricate temporal sequences in multi-sensor activity tracking (Bollampally et al., 2024); (Lasri et al., 2023). In the specific context of educational data mining, integrating bidirectional processing allows the network to construct a more resilient internal representation of student disengagement. As highlighted in recent multimodal engagement studies, dual-directional recurrent layers effectively filter out transient activity noise (e.g., a one-off missed quiz due to technical issues) while magnifying persistent, structurally meaningful disengagement patterns (Feng, 2025); (Saha & Melnik, 2025).

Ultimately, the analytical insights derived from Table 2 validate that capturing bidirectional temporal context is essential for building trustworthy early-warning systems in distance education. While traditional machine learning models offer low computational overhead, their inability to model temporal dynamics renders them unreliable for early intervention, often yielding high false-positive rates that place unnecessary strain on instructional resources (Naripeddy & Thotakura, 2025). By resolving complex sequence dynamics through the unified pipeline shown in Figure 1, the proposed BiLSTM architecture establishes an optimal balance between high predictive accuracy and pedagogical reliability. This equips distance education platforms with a robust algorithmic foundation capable of delivering accurate, actionable risk assessments well before academic failure becomes irreversible.

Digital and distance pedagogical implications

The outputs of this model can be integrated directly into instructor analytics dashboards within distance education platforms. When the system identifies a student as 'At-Risk' at Week 6, educators can trigger automated or targeted manual interventions, such as: Early warning notifications sent directly to the student; Supplementary virtual mentoring sessions or personalized remedial tasks; Tailored adjustments to online learning materials matching specific student weaknesses. Here is a refined, academically rigorous Section 3.3 (Digital and Distance Pedagogical Implications) expanded into 4 deep, analytical paragraphs. This version seamlessly integrates the theoretical framework, empirical results from Table 1 and Table 2, the architecture in Figure 1, and the relevant literature citations:

Digital and distance pedagogical implications

The integration of the proposed BiLSTM predictive model into distance education platforms fundamentally redefines learning analytics by bridging the persistent gap between passive, retrospective data collection and active, real-time instructional support. Traditionally, online education platforms have suffered from physical isolation and a lack of continuous supervisory monitoring, often causing student disengagement to remain invisible until summative assessments reveal failure. As demonstrated by the empirical findings in Table 1 and Table 2, the proposed framework resolves this issue by establishing a highly accurate predictive pipeline. By embedding the sequential classification workflow illustrated in Figure 1 directly into instructor-facing LMS dashboards, educators can transition from reactive teaching methods—which address failure after it occurs—to a proactive, data-driven pedagogical paradigm (F. Zhang et al., 2025); (Koesmanto et al., 2025). Translating raw behavioral logs into intuitive risk scores reduces cognitive overload for instructors, allowing them to instantly identify disengaging students during critical learning windows (Misiejuk et al., 2026); (Katalinic et al., 2026).

The operational milestone achieved at Week 6 (91.4% accuracy, 92.6% Recall) provides a crucial temporal window for executing early, targeted pedagogical interventions before mid-term evaluations take place. Analyzing continuous time-series interactions—such as weekly video view durations, forum participation frequencies, and quiz submission delays—enables the model to identify subtle indicators of academic disengagement in the initial weeks of course engagement (Priya Swaminarayan, 2025); (Saluja et al., 2025). Achieving high prediction accuracy around the mid-semester mark is particularly vital, as early interventions deployed around Weeks 4 to 6 have been shown to significantly improve student learning outcomes and mitigate course abandonment compared to late-term interventions (F. Zhang et al., 2025); (Val & Quintas, 2025). By automating

the early-warning layer, the system ensures that vulnerable distance learners receive immediate support when intervention efficacy is highest.

Once an 'At-Risk' profile is flagged by the BiLSTM engine, the LMS can trigger automated or instructor-led personalized interventions tailored to specific student behavioral profiles (Nain et al., 2025). These actionable interventions operate across three distinct pedagogical tiers: Direct nudges: Automated early warning notifications sent straight to the student's dashboard to stimulate self-regulated learning and prompt immediate re-engagement (Hu et al., 2025); (Balaskas et al., 2025). Targeted Human Support: Scheduling supplementary virtual mentoring sessions and peer-led study groups to address underlying conceptual misunderstandings and reduce feelings of transactional distance (Winsli G. Felix et al., 2025). Adaptive Content Delivery: Dynamically adjusting online learning pathways and assigning tailored remedial exercises designed to bridge specific cognitive gaps identified through LMS behavioral patterns (Dritsas & Trigka, 2025). By leveraging big data and deep sequence modeling, AI-driven predictive systems enhance learning personalization far beyond traditional, one-size-fits-all online formats (X. Zhang & Shi, 2025); (Mo et al., 2026).

Ultimately, deploying this predictive architecture addresses the core systemic challenges of digital and distance learning: low retention rates, student isolation, and high attrition. Evaluating behavioral, cognitive, and temporal metrics through a unified deep learning pipeline establishes a scalable, data-driven framework for early intervention (Pineda-Briseño et al., 2026); (Omarbekova et al., 2026). Empowering educators with real-time, actionable risk assessments transforms LMS platforms from static content repositories into adaptive, supportive learning environments. This shift not only enhances academic performance and critical thinking but also fosters long-term student engagement and course completion across online higher education programs.

Conclusion

This study addresses the critical challenge of high academic failure and student dropout rates in digital and distance education caused by physical isolation and the lack of continuous supervisory monitoring. Conventional predictive models often fail to capture the dynamic nature of student disengagement due to their reliance on static, aggregated data evaluated at the end of an academic term. To overcome these limitations, a sequential deep learning framework based on Bidirectional Long Short-Term Memory (BiLSTM) was designed and evaluated to process non-linear, time-series LMS interaction logs across multiple temporal checkpoints without depending on final aggregate metrics. The experimental results demonstrate that the proposed BiLSTM model achieves superior early-prediction performance compared to traditional machine learning baselines and standard unidirectional LSTM networks. By Week 6—well before mid-term examinations—the BiLSTM model reached an optimal analytical performance balance, achieving an accuracy of 91.4%, an F1-Score of 90.8%, and a high Recall rate of 92.6%. The high Recall rate is particularly significant as it minimizes false negatives, ensuring that at-risk distance learners are accurately identified before disengagement becomes irreversible. The model's bidirectional layer proved essential in capturing complex, non-linear engagement patterns by evaluating both forward and backward temporal contexts simultaneously. From a pedagogical perspective, these findings offer actionable implications for learning analytics in higher and secondary distance education. Integrating the proposed predictive pipeline directly into LMS analytics dashboards enables educators to transition from reactive teaching to proactive, data-driven interventions. Systematically flagging 'At-Risk' students at Week 6 allows institutions to deploy targeted nudges,

schedule virtual mentoring sessions, and provide personalized remedial materials, thereby improving student retention and reducing course attrition. Despite its strong empirical performance, this study has a few limitations. First, the evaluation relied primarily on behavioral log data from LMS platforms, omitting non-behavioral variables such as student demographics, prior academic background, and qualitative feedback. Second, the model was validated within a specific distance learning environment, which may limit the direct generalizability of the pre-trained weights across vastly different institutional contexts or course structures. Future research should focus on extending this framework by incorporating multimodal data streams, including demographic profiles, formative assessment grades, and natural language processing (NLP) of forum discussions, to enhance predictive granularity. Additionally, future work should explore cross-institutional validation and test the real-time deployment of adaptive, automated intervention systems to evaluate their direct impact on long-term student retention and academic success

Conflicts of interest

The author declares no conflict of interest.

Funding

This research received no external funding.

Author contributions

L.Y.: Conceptualization Investigation; Validation; Writing - Original Draft; Visualization, Investigation; Validation; Writing - Original Draft. L.M.S.: Resources; Data Curation: Validation; Investigation; Writing - Original Draft; Research Guidance and Mentoring; Supervision; Formal Analysis. All authors have read and agreed to the published version of the manuscript.

Declaration of generative AI in scientific writing

Generative AI was employed in this study strictly for language enhancement and structural formatting. The authors maintain full accountability for the intellectual integrity, originality, and factual accuracy of the entire manuscript. No AI tool was used to create or alter primary research data.

References

Akello, E. F., Ijiga, O. M., & Idoko, I. P. (2026). Sequence-Aware Learning Analytics for Early Identification of At-Risk Academic Trajectories in Higher Education Using Transformer Models. *International Journal of Innovative Science and Research Technology*, 818. <https://doi.org/10.38124/ijisrt/26jan563>

- Al-Zyoud, N., Al-Hendawi, M., & Alodat, A. (2025). Stakeholder Perspectives on Challenges and Improvements in Student Classification and Progress Monitoring in Qatari Schools: A Qualitative Study. *Sustainability*, 17(22), 10042. <https://doi.org/10.3390/su172210042>
- Asdaq, S. M. B., Alhowail, A. H., Rabbani, S. I., Nayeem, N., Asdaq, S. M. E., & Nausheen, F. (2025). Learning Disabilities in the 21st Century: Integrating Neuroscience, Education, and Technology for Better Outcomes. *SAGE Open*, 15(3), 21582440251365483. <https://doi.org/10.1177/21582440251365483>
- Balaskas, S., Yfantidou, I., Nikolopoulos, T., & Komis, K. (2025). The Psychology of EdTech Nudging: Persuasion, Cognitive Load, and Intrinsic Motivation. *European Journal of Investigation in Health, Psychology and Education*, 15(9), 179. <https://doi.org/10.3390/ejihpe15090179>
- Bollampally, A., Kavitha, J., Sumanya, P., Rajesh, D., Jaffar, A. Y., Eid, W. N., Albarakati, H. M., Aldosari, F. M., & Alharbi, A. A. (2024). Optimizing Edge Computing for Activity Recognition: A Bidirectional LSTM Approach on the PAMAP2 Dataset. *Engineering, Technology & Applied Science Research*, 14(6), 18086–18093. <https://doi.org/10.48084/etasr.8861>
- Chen, F., & Cui, Y. (2020). Utilizing Student Time Series Behaviour in Learning Management Systems for Early Prediction of Course Performance. *Journal of Learning Analytics*, 7(2), 1–17. <https://doi.org/10.18608/jla.2020.72.1>
- Chen, J., & Lin, J. (2020). The Application of Bi-LSTM in Computerized Adaptive Test Research. *2020 International Conference on Big Data and Social Sciences (ICBDSS)*, 33–36. <https://doi.org/10.1109/ICBDSS51270.2020.00015>
- Dritsas, E., & Trigka, M. (2025). Methodological and Technological Advancements in E-Learning. *Information*, 16(1), 56. <https://doi.org/10.3390/info16010056>
- Feng, F. (2025). Deep learning-based model for analyzing student engagement in activities. *Scientific Reports*, 16(1), 2552. <https://doi.org/10.1038/s41598-025-32521-w>
- Fitas, R. (2025). Inclusive education with AI: Supporting special needs and tackling language barriers. *AI and Ethics*, 5(6), 5729–5757. <https://doi.org/10.1007/s43681-025-00824-3>
- Hamdane, K., Mhouti, A. E., & Massar, M. (2025). Addressing Dropout through Personalization: A Graph Neural Network Approach to Modelling Learner Interactions. *Global Journal of Engineering and Technology Advances*, 25(1), 001–012. <https://doi.org/10.30574/gjeta.2025.25.1.0291>
- Hendrawan, S. A. (2025). Analysis of the Digital System-Based Academic Factors and the Risk of Student Dropout Using Chi-Square Test for Data Driven-Interventions. *Edu Cendikia: Jurnal Ilmiah Kependidikan*, 5(02), 602–616. <https://doi.org/10.47709/educendikia.v5i02.6736>
- Hu, C., Li, F., Wang, S., Gao, Z., Pan, S., & Qing, M. (2025). The role of artificial intelligence in enhancing personalized learning pathways and clinical training in dental education. *Cogent Education*, 12(1), 2490425. <https://doi.org/10.1080/2331186X.2025.2490425>
- Junejo, N. U. R., Nawaz, M. W., Huang, Q., Dong, X., Wang, C., & Zheng, G. (2024). *Accurate Multi-Category Student Performance Forecasting at Early Stages of Online Education Using Neural Networks* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2412.05938>
- Karapiperis, D., Tzafilkou, K., Tsoni, R., Feretzakis, G., & Verykios, V. S. (2023). A Probabilistic Approach to Modeling Students' Interactions in a Learning Management System for Facilitating Distance Learning. *Information*, 14(8), 440. <https://doi.org/10.3390/info14080440>
- Katalinic, A., Slavuj, V., & Jaksic, D. (2026). Artificial Intelligence in Online Education: A Systematic Review of Its Impact on Learner Engagement and Satisfaction. *Education Sciences*, 16(3), 389. <https://doi.org/10.3390/educsci16030389>
- Koesmanto, E. A., Staniswinata, S., & Manuaba, I. B. K. (2025). A User-Centered Learning Analytics Dashboard with Proactive Notifications to Empower Online Teachers. *2025 International Conference on Smart-Green Technology in Electrical and Information Systems (ICSGTEIS)*, 423–428. <https://doi.org/10.1109/ICSGTEIS68532.2025.11284387>
- Lasri, I., Riadsolh, A., & ElBelkacemi, M. (2023). Self-Attention-Based Bi-LSTM Model for Sentiment Analysis on Tweets about Distance Learning in Higher Education. *International Journal of*

- Emerging Technologies in Learning (iJET)*, 18(12), 119–141. <https://doi.org/10.3991/ijet.v18i12.38071>
- Mihoubi, M., Zerkouk, M., & Chikhaoui, B. (2025). *Beyond classical and contemporary models: A transformative AI framework for student dropout prediction in distance learning using RAG, Prompt engineering, and Cross-modal fusion* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2507.05285>
- Misiejuk, K., López-Pernas, S., Kaliisa, R., & Saqr, M. (2026). Cognitive offloading in student–AI collaboration: A longitudinal analysis of prompting strategies. *Computers in Human Behavior Reports*, 22, 101130. <https://doi.org/10.1016/j.chbr.2026.101130>
- Mo, Y., Zhang, J., & Mou, Y. (2026). Big Data Research on Personalized Learning in Computer Education: A Thematic Evolution Analysis. *Computer Applications in Engineering Education*, 34(2), e70159. <https://doi.org/10.1002/cae.70159>
- Nain, N., Tanwar, S., & Singh, P. (2025). *A Data Analytics-Based System for Proactive Student Performance Monitoring and Personalized Learning Interventions*. Computer Science and Mathematics. <https://doi.org/10.20944/preprints202510.1809.v1>
- Naripeddy, S. C. B., & Thotakura, V. R. (2025). Optimizing Predictive Accuracy: A Comparative Study of Modern Machine Learning Algorithms in Real-World Datasets. *Journal of Artificial Intelligence General Science (JAIGS)* ISSN:3006-4023, 8(02), 111–125. <https://doi.org/10.60087/jaigs.v8i02.393>
- Neththikumara, S., & Maduranga, M. W. P. (2025). Recent Results of Machine Learning in Predicting Student Performance: A Systematic Literature Review. *2025 IEEE Digital Education and MOOCS Conference (DEMOcon)*, 1–5. <https://doi.org/10.1109/DEMOcon65705.2025.11282820>
- NPLC «Karaganda Buketov University», Muratkhan, R., Kazimova, D., NPLC «Karaganda Buketov University», Zhumagulova, S., NPLC «Karaganda Buketov University», Kopyltsov, A., Saint Petersburg State University of Aerospace Instrumentation, Tursyngaliyeva, G., & NPLC «Karaganda Buketov University». (2025). Data Analysis and Prediction of Students' Academic Performance Using Machine Learning and Deep Learning Methods. *TRUDY UNIVERSITETA*, 3(100), 488–494. https://doi.org/10.52209/1609-1825_2025_3_488
- Omarbekova, A., Ramazan, A., Oralbekova, Z., Lamasheva, Z., Bekmanova, G., Nazyrova, A., & Abuov, A. (2026). A temporal attention-based hybrid deep learning model for student performance and academic risk prediction. *Frontiers in Artificial Intelligence*, 9, 1811886. <https://doi.org/10.3389/frai.2026.1811886>
- Pineda-Briseño, A., Hernández-Compean, M. G., Flores-Becerra, G. A., Hernández-Quezada, M. D. J., & De Los Santos-Alonso, M. M. (2026). A Scalable Data Pipeline for Early Detection and Decision Support in Higher Education: YuumCare. *Data*, 11(5), 112. <https://doi.org/10.3390/data11050112>
- Priya Swaminarayan. (2025). *Intelligent Early Warning System for Student Dropout Detection Using Deep Learning*. <https://doi.org/10.5281/ZENODO.17303724>
- Saha, S., & Melnik, R. (2025). Integrating Emotion-Specific Factors into the Dynamics of Biosocial and Ecological Systems: Mathematical Modeling Approaches Accounting for Psychological Effects. *Mathematical and Computational Applications*, 30(6), 136. <https://doi.org/10.3390/mca30060136>
- Saluja, A., Bindu, Baig, N., Grover, A., & Adlakha, R. (2025). Leveraging Digital Learning Environments and Predictive Analytics to Develop Adaptive Early Warning Systems for At-Risk Students in Higher Education: In Minakshi, T. Kumar, M. Bhatia, B. Wadhwa, & R. Jain (Eds.), *Enhancing Operational Efficiency and Predictive Maintenance Through Digital Innovation* (pp. 221–242). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3373-2474-6.ch011>
- Sandiwarno, S., Sensuse, D. I., Santoso, H. B., Hidayat, D. S., Nyamawe, A. S., & Yousif, A. (2025). E-SATNet: Evaluating Student Satisfaction with Lecturer Responses in Asynchronous Online Discussions Using Sentiment and Semantic Similarity Analysis. *Big Data and Cognitive Computing*, 9(9), 228. <https://doi.org/10.3390/bdcc9090228>

- Sun, D. (2022). LMS Log Data Presenting Interaction Traces. In D. Sun & G. Cheng, *Learner Interactions in Massive Private Online Courses* (1st ed., pp. 24–30). Routledge. <https://doi.org/10.4324/b23163-4>
- Suresh, A., Kaplan, E. H., Pinker, E. J., & Gruen, J. R. (2025). Optimal evaluation policies to identify students with reading disabilities. *Socio-Economic Planning Sciences*, 98, 102116. <https://doi.org/10.1016/j.seps.2024.102116>
- Termedi, M. I., Ma'rof, A. M., Ab. Jalil, H. B., & Ishak, I. (2023). A Thematic Review on Predicting Student Performance Using Data Mining. *International Journal of Academic Research in Business and Social Sciences*, 13(11), Pages 1814-1831. <https://doi.org/10.6007/IJARBS/v13-i11/19563>
- Val, S., & Quintas, A. (2025). Key performance indicators for optimizing academic performance and course design in online educational platforms. *Cogent Education*, 12(1), 2529420. <https://doi.org/10.1080/2331186X.2025.2529420>
- Wang, P., Ganushchak, L., Welie, C., & Van Steensel, R. (2026). Emotions Are Not Random: Machine Learning Reveals Predictable Patterns in a 21-Day Second Language Learning Trajectory. *Educational Psychology Review*, 38(1), 49. <https://doi.org/10.1007/s10648-026-10141-8>
- Winsli G. Felix, J., E. Supan, F., & Makhammatkosimovna Kuchkarova, F. (2025). A Comprehensive Framework for Predicting Student Performance Using Machine Learning in Online Education. *Qubahan Techno Journal*, 4(1), 16–22. <https://doi.org/10.48161/qtj.v4n1a44>
- Xu, N., Wu, F., Liu, Z., & Zhan, Y. (2025). Evaluating the academic outcome of AI-powered joint support for at-risk students. *Computers and Education: Artificial Intelligence*, 9, 100517. <https://doi.org/10.1016/j.caeai.2025.100517>
- Zhang, F., Subagia, I. W., Artini, L. P., & Wahyuni, D. S. (2025). Optimizing blended learning through AI-powered analytics in digital education platforms: An empirical framework. *Future Technology*, 4(4), 173–184. <https://doi.org/10.55670/fpll.futech.4.4.15>
- Zhang, X., & Shi, W. (2025). An AI-driven framework integrating predictive modeling and intervention strategies to enhance psychological health education among college students. *Discover Applied Sciences*, 7(10), 1095. <https://doi.org/10.1007/s42452-025-07674-y>