Research Article

**Received:** 30 January, 2026
**Accepted:** 29 April, 2026
**Published:** 06 May, 2026

# A Deep Learning–Driven Intervention Framework for Predicting and Mitigating Student Dropout in Learning Management Systems

**Aryan F. Abdulqader**
Department of Computer Science, Faculty of Science, Soran University, Soran, Kurdistan Region, Iraq
Email: aryan.abdulqader@soran.edu.iq

**Tarik A. Rashid**
Computer Science and Engineering Department, Artificial Intelligence & Innovation Centre, University of Kurdistan Hewler, Erbil, Iraq

**Cite this article:**

Abdulqader, A. F., & Rashid, T. A. (2026). A Deep Learning–Driven Intervention Framework for Predicting and Mitigating Student Dropout in Learning Management Systems. Cultura Científica, (24), pp. 303–318.

## Abstract

Student dropout prediction has become a critical field of research in learning management systems with the rapid growth of large-scale online education systems. Nevertheless, most available research is primarily dedicated to enhancing predictive accuracy and offers few mechanisms for converting predictions into effective interventions that help decrease student dropout rates. This is a major limitation, since high dropout rates negatively affect learner success, institutional efficiency, and the sustainability of online learning environments. To overcome this limitation, this research paper proposes a new student dropout prediction framework that uses deep learning networks together with an adaptive AI-based intervention system. The framework was evaluated using a large-scale HarvardX edX dataset containing more than 640,000 learner course interaction records. Three deep learning architectures, namely Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer, were developed and compared systematically. The experimental findings show that the Transformer model is the most effective, with a predictive accuracy of 0.92 and an AUC-ROC of 0.95, making it one of the most effective models for capturing temporal dependencies and behavioral patterns in normalized sequences of learner activities. Based on the prediction results, the individualized intervention program provided specific interventions, including motivational messages, learning materials, instructor notifications, and peer-support invitations. The use of the intervention led to improvements in learner engagement and retention by 21.8% and 17.1%, respectively. Moreover, the interpretability analysis showed that active learning days and participation in course events are the strongest predictors of student dropout.

**Keywords:** student dropout prediction, deep learning, learning management systems, AI-driven intervention, online education analytics

## 1. INTRODUCTION

The spread of the internet and hybrid education systems has radically changed access to education by offering more flexible, scalable, and relatively low-cost modes of learning. Despite these benefits, student dropout remains a major issue that compromises the pedagogical and institutional effectiveness of e-learning platforms [1]. The literature consistently reports high attrition rates, often exceeding 90% in Massive Open Online Courses (MOOCs) and Learning Management Systems (LMSs), mainly because of low motivation, limited instructor support, lack of peer interaction, and difficulties in regulating self-directed learning behaviors [2]. This long-standing problem not only creates academic and financial pressure on institutions but also threatens the credibility and sustainability of online education. Thus, early detection and dropout prevention through data-driven predictive modeling have become central concerns in current learning analytics research.

In recent years, artificial intelligence (AI) and deep learning have made significant progress and have shown strong potential for modeling the complex behavioral and temporal dynamics present in learner interaction data [3]. Recent deep neural architectures, such as Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and Transformers, have been highly successful in learning sequential dependencies and contextual relationships across a wide range of learner behaviors, including content-access frequency, video interaction, and forum participation [4]. Although these models are useful for dropout prediction, most existing frameworks are reactive rather than proactive, as they include limited practical mechanisms for pedagogical intervention [5]. This limitation has created a strong need for an integrated solution that not only predicts dropout with high accuracy but also provides individualized, time-sensitive, and adaptive support to increase learner attendance and persistence.

The main goal of this research is to develop and evaluate a holistic system that combines deep learning-based dropout prediction with an AI-driven intervention system to enhance student retention in large-scale online learning environments. Specifically, this study aims to make the following contributions:

- This work proposes an integrated deep learning-based framework that combines student dropout prediction with an AI-driven adaptive intervention mechanism to actively mitigate dropout in learning management systems.
- The framework is validated on a large-scale HarvardX edX dataset containing over 640,000 learner interaction records, demonstrating its scalability and practical applicability.
- A comparative evaluation of LSTM, GRU, and Transformer models is conducted, with results showing that the Transformer achieves superior predictive performance in modeling sequential learner behavior.
- A personalized intervention strategy is introduced to translate dropout predictions into targeted support actions, resulting in significant improvements in learner engagement by 21.8% and retention by 17.1%.
- The study enhances model interpretability by identifying key behavioral predictors of dropout, including active learning days and course-event participation.

The rest of this paper is organized as follows. Section 2 presents a review of related work on student dropout prediction, early-warning systems, and AI-based intervention strategies in learning management systems, and identifies the current problems and research gaps. Section 3 outlines the proposed methodology, including the data preprocessing procedures, the deep learning models used for dropout prediction, and the structure of the AI-driven personalized intervention system. Section 4 presents the experimental setup and results, including a comparative performance analysis of the proposed models and the effect of the intervention on learner engagement and retention. Section 5 discusses the key findings, relates them to existing research, and outlines the practical implications and limitations of the proposed framework. Finally, Section 6 concludes the paper by summarizing the main contributions and suggesting directions for future research.

## 2. RELATED WORK

Recent studies on student dropout prediction and AI-assisted retention can be categorized into three major strands: (i) predictive modeling of dropout risk, (ii) early-warning and intervention systems, and (iii) AI-enabled adaptive and intelligent learning environments [6]. Across these strands, substantial progress has been made in improving prediction accuracy, although most studies still show important limitations in terms of applicability, real-time implementation, and systematic assessment of interventions.

One early research direction has focused on supervised machine learning and deep learning-based dropout prediction. Won et al. [7] formulated dropout prediction as a natural language inference problem using a pretrained language model based on demographic and academic information, demonstrating strong performance within one university in Korea. Several reviews also report that deep learning models, particularly LSTM-type models, generally perform better on MOOC and LMS logs than classical machine learning methods [8]. More recent progress has been made through ensemble and hybrid architectures, such as an ensemble deep learning network for MOOC dropout prediction [9], AutoGluon FT-Transformer-based tabular models [10], and TSA-GRU hybrids that jointly use temporal self-attention with recurrent units [11]. Transformer-based cross-modal models, such as TRIAD Drop, are usually trained and tested on single-institution or single-platform datasets and often lack feature diversity and external validation, which limits their generalizability to other LMSs and educational contexts. In addition, many studies continue to rely on semester-level characteristics rather than fully exploiting fine-grained temporal sequences, which limits their usefulness for practical early intervention [12].

The second strand investigates early-warning systems (EWSs) and dropout monitoring. Bañeres et al. [13] proposed an early-warning system that predicts at-risk online learners and initiates human-led actions, resulting in improved course completion, but the system relies on predefined rules and comparatively simple predictive models. School- and university-level versions of such frameworks have also been developed, with an emphasis on risk dashboards and teacher-facing alerts rather than end-to-end automated pipelines [14]. Boudjehem and Lafifi [15] proposed EWS architectures that use machine learning for risk scoring, but the intervention logic is only loosely defined and is rarely assessed through controlled A/B designs. More applied work uses gradient-boosted trees or XGBoost to predict dropout from enrollment-time data and support day-one alerts. However, these systems often do not consider longitudinal engagement dynamics, nor do they distinguish which interventions are most effective for different at-risk learners [16]. Overall, this strand demonstrates the practical value of EWSs, but it often lacks clear mechanisms for implementing interventions, provides only weak personalization, and rarely links intervention design back to model evidence.

A third and emerging strand of literature investigates AI-based interventions, adaptive platforms, and intelligent student management systems. Wang [17] describes AI-based student management systems that typically automate monitoring processes and recommend interventions to reduce dropout, although the study provides limited technical detail on model architectures or sequence modeling. Tan et al. [6] reviewed AI-enabled adaptive learning systems and deep learning-driven educational data mining, identifying the potential of personalization while noting the scarcity of rigorous longitudinal impact studies. Similarly, the prototypes and reviews of AI-driven administrative automation and LMS integration discussed by Mutambik [18] emphasize improved efficiency and learner support, but they do not fully address fine-grained engagement modeling or targeted retention interventions. Systematic reviews of AI-mediated educational interventions suggest promising benefits for motivation and lifelong learning; however, they are generally based on small samples, non-experimental studies, or self-reported outcomes rather than robust retention measures [19]. Important ethical issues, including privacy, consent, bias, and transparency, are often discussed, but few studies implement specific solutions such as federated learning, bias audits, or explainable AI in the context of dropout prediction.

The available literature confirms that machine learning and deep learning models can accurately predict dropout and that early-warning dashboards and AI-assisted tools can support human decision-making. Nevertheless, the literature still has several shortcomings: (i) limited use of sequence-aware, attention-based models that are tightly integrated with temporal engagement data; (ii) limited evidence for closed-loop, model-driven intervention systems that are closely linked to personalized and rigorously evaluated intervention deployment; (iii) weak cross-platform and cross-institution validation; and (iv) relatively abstract treatment of privacy-preserving and fairness-conscious deployment. A summary of related work on student dropout prediction and intervention is presented in Table 1.

**Table 1.** *Summary of related work on student dropout prediction and intervention*

| Reference / Year | Description of the Work | Main Objectives | Critical Review | Key Limitations |
|---|---|---|---|---|
| Won et al. [7] | Formulated student dropout prediction as a natural language inference task using a pretrained language model on demographic and academic data. | Improve dropout prediction accuracy using NLP-based deep learning. | Demonstrates strong predictive performance within a single university context. | Limited generalizability; lacks temporal engagement modeling and intervention mechanisms. |
| Ensemble DL for MOOC Dropout [9] | Proposed ensemble deep learning architectures for improved dropout prediction. | Enhance robustness and accuracy through ensemble learning. | Improves performance over single models. | Increased complexity; no intervention integration or impact assessment. |
| AutoGluon FT-Transformer [10] | Applied FT-Transformer-based tabular models for dropout prediction. | Leverage automated machine learning and attention mechanisms. | Strong tabular modeling capability. | Often relies on coarse-grained features; limited temporal sequencing and interventions. |
| TSA-GRU Hybrid [11] | Combined temporal self-attention with GRU for sequential modeling. | Capture long-term dependencies in learner behavior. | Provides better temporal representation than standard RNNs. | Evaluated on limited datasets; no closed-loop intervention design. |
| TRIAD Drop (2024) | Developed a Transformer-based cross-modal dropout prediction model. | Fuse multimodal learner data using attention mechanisms. | Advanced architecture with strong predictive results. | Single-platform validation; limited feature diversity and no intervention study. |
| Bañeres et al. [13] | Developed an early-warning system that triggers human-led interventions. | Improve course completion through early alerts. | Demonstrates the practical feasibility of EWSs. | Rule-based interventions; simple models; no automation or personalization. |
| Boudjehem and Lafifi [15] | Proposed machine learning-based early-warning architectures for dropout risk scoring. | Identify at-risk learners at an early stage. | Provides useful risk dashboards and alerts. | Intervention logic is weakly defined; limited experimental validation. |
| Applied XGBoost EWS (2023–2024) | Used enrollment-time dropout prediction for day-one alerts. | Enable early identification of at-risk students. | Effective for static early prediction. | Ignores longitudinal engagement and adaptive interventions. |
| Wang [17] | Presented an overview of AI-based student management systems. | Automate monitoring and recommend interventions. | Highlights the potential of AI-driven management. | Lacks technical depth and sequence modeling details. |
| Tan et al. [6] | Surveyed AI-enabled adaptive learning systems. | Assess personalization potential in AI-driven LMSs. | Identifies promising personalization trends. | Few longitudinal or causal impact evaluations. |
| Mutambik [18] | Reviewed AI-driven LMS integration and automation. | Improve efficiency and learner support. | Provides a strong focus on system integration. | Limited emphasis on dropout-specific modeling and interventions. |

## 3. METHODOLOGY

This section presents the methodological design of the proposed deep learning and AI-driven intervention framework for predicting and mitigating student dropout in Learning Management Systems (LMSs). The framework consists of four consecutive stages: (i) data acquisition and preprocessing, (ii) model development and training, (iii) AI-driven intervention design, and (iv) evaluation and validation. Each stage was designed to ensure reproducibility, scalability, and ethical consideration in the development of an intelligent and learner-centered prediction and intervention pipeline.

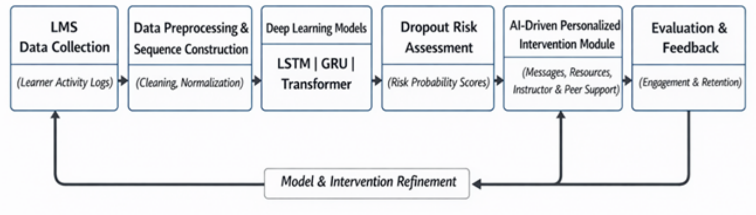

**Figure 1.** *Architecture of the proposed deep learning and AI-driven framework*

Figure 1 shows the general structure of the proposed framework for predicting and preventing student dropout in learning management systems. The framework begins by collecting fine-grained learner activity logs from the LMS, which are then preprocessed and converted into temporal sequences through data cleaning and normalization. These sequences are learned by deep learning models, including LSTM, GRU, and Transformer models, to capture behavioral dynamics and predict dropout risk. Based on the predicted risk levels, an AI-based personalized intervention module provides specific support actions, such as motivational messages, learning materials, instructor notifications, and peer-support invitations. The effectiveness of these interventions is then measured using engagement and retention metrics, and the results are fed back into the system to continuously improve both the predictive models and the intervention strategies. This closed-loop architecture ensures that the framework not only predicts dropout accurately but also actively supports learner retention through adaptive and data-driven interventions.

### 3.1. DATA ACQUISITION, CLEANING, AND PREPROCESSING

The initial stage of the proposed deep learning and AI-based intervention framework is based on data acquisition and preprocessing. This stage ensures the integrity, structure, and representational adequacy of learner interaction data so that the models can extract meaningful behavioral patterns for dropout prediction [20]. The dataset used in this study was obtained from the HarvardX edX repository available on Kaggle, which is a widely used reference dataset in online learning analytics studies. It contains approximately 641,138 learner-course records with 22 important attributes, including demographic, behavioral, and performance-related features. The attribute `incomplete_flag` was used to compute the target variable, `Dropout Flag`, indicating learners who dropped out before course completion.

Raw LMS data are frequently noisy, inconsistent, and incomplete [21]. Therefore, a systematic data-cleaning process was applied to ensure analytical robustness [22]. Missing numerical values were filled using median imputation, which reduces the influence of outliers [23], whereas missing categorical values were filled using the mode or the category "Unknown". Unnecessary and uninformative fields, such as learner IDs and administrative time fields, were removed. The dataset was further harmonized by standardizing variable names and variable types across different course offerings. The main features used after cleaning and integration are summarized in Table 2.

**Table 2.** *Summary of key features after data cleaning*

| Feature Name | Description | Type | Role |
|---|---|---|---|
| Gender | Learner gender | Categorical | Demographic |
| LoE_DI | Level of education | Categorical | Demographic |
| ndays_act | Number of active learning days | Numeric | Behavioral |
| nevents | Total number of logged activities | Numeric | Behavioral |
| nchapters | Chapters viewed or completed | Numeric | Behavioral |
| nforum_posts | Forum posts and replies | Numeric | Social |
| Grade | Course grade on a 0–1 scale | Numeric | Performance |
| start_time_DI | First course interaction timestamp | DateTime | Temporal |
| last_event_DI | Last course interaction timestamp | DateTime | Temporal |
| incomplete_flag | Dropout indicator | Binary | Target |

To improve the predictive capacity of the models, a set of derived engagement measures was constructed from the raw variables to better represent the cognitive, behavioral, and social aspects of online learning [24, 25]. These features capture the intensity, regularity, and participation trends of learners. The engineered features and their definitions are listed in Table 3.

**Table 3.** *Engineered behavioral features for dropout prediction*

| Feature | Formula | Description |
|---|---|---|
| Engagement Intensity | $\left(\frac{\text{nevents}}{\text{ndays\_act}}\right)$ | Average activity frequency per day |
| Learning Consistency | $(1 - \text{Var}(\text{weekly activity}))$ | Stability of weekly participation |
| Social Interaction Index | $\left(\frac{\text{nforum\_posts}}{\text{nevents}}\right)$ | Proportion of social interactions |
| Progress Velocity | $\left(\frac{\text{nchapters}}{\text{ndays\_act}}\right)$ | Rate of course completion |

To prevent feature redundancy, correlation analysis was conducted using Pearson's coefficient. Highly correlated variables, with $r > 0.85$, were removed to reduce multicollinearity and improve generalization. The Pearson correlation analysis showing highly correlated variables is presented in Figure 2.

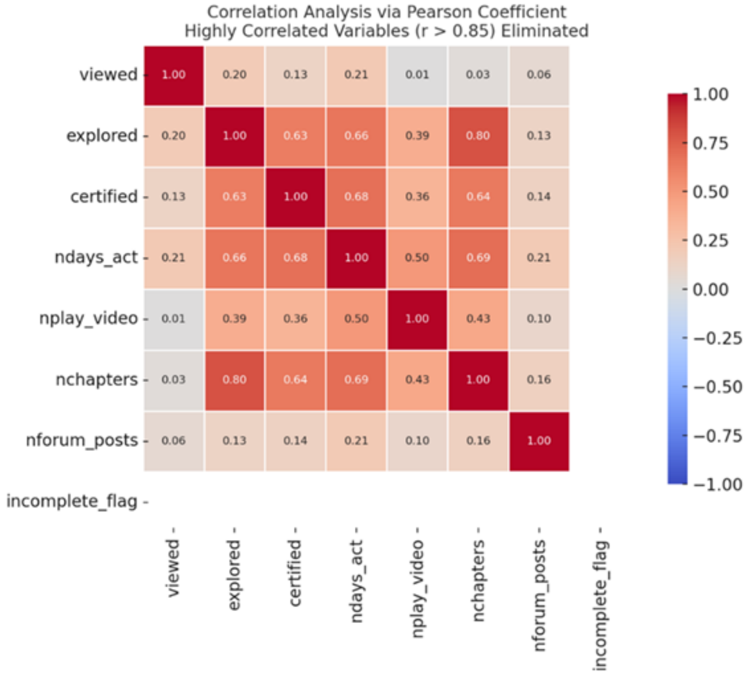

**Figure 2.** *Correlation analysis using Pearson's coefficient showing highly correlated variables*

Because the dataset contained numerical features with different scales, Min–Max normalization was applied to transform all continuous variables into the range $[0, 1]$, thereby improving the numerical stability of model training [26, 27]. In addition, the temporal features `start_time_DI` and `last_event_DI` were used to construct chronologically ordered sequences representing each learner's behavior. The interaction history of each student was converted into a 10-step sequence, where each step summarized weekly engagement activity, including events, chapters, and forum interactions. This transformation enabled sequence-based models, including LSTM, GRU, and Transformer models, to learn time-dependent behavioral changes associated with dropout.

Ordinal variables, such as course progress stages, were label-encoded to preserve their ordered relationships, whereas categorical variables, including gender and `LoE_DI`, were one-hot encoded. The processed dataset was divided using stratified sampling to maintain the class balance between dropout and non-dropout learners. The data distribution was as follows:

- Training set: 70%;
- Validation set: 15%;
- Testing set: 15%.

This partitioning strategy helped avoid overfitting and enabled robust performance evaluation on unseen data. The slight class imbalance was addressed during training by using class-weighted loss-function adjustments to prevent bias toward the majority class.

After preprocessing, the final dataset contained approximately 610,000 valid learner records with 20 cleaned and engineered features. Random sample checks were performed to verify the correctness of sequence formation and encoding. Descriptive statistics and visual diagnostics were used to confirm that the feature distributions did not change substantially after normalization and that no artificial bias was introduced during transformation. The correlation heatmap of engineered and original features is shown in Figure 3.

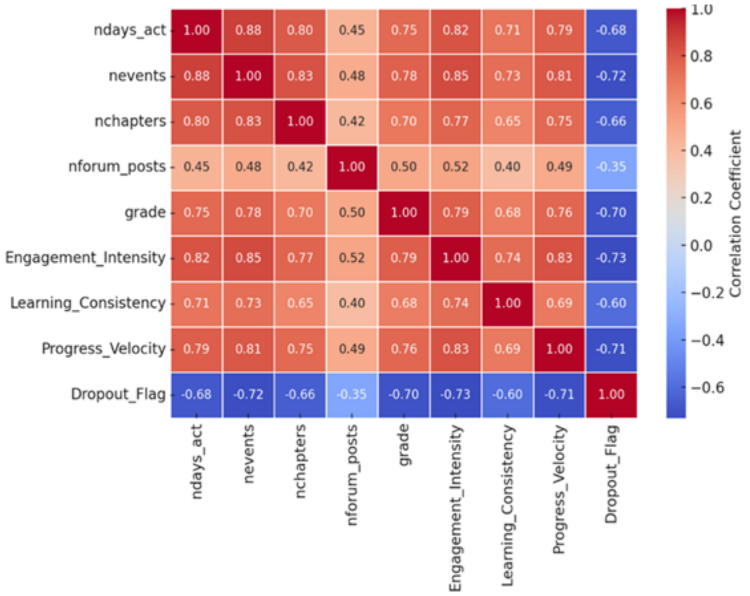

**Figure 3.** *Correlation heatmap of engineered and original features*

As shown in Figure 3, `ndays_act` and `nevents` have a strong positive relationship with course completion, while `nforum_posts` has a moderate relationship with learner retention. These correlations confirm the importance of temporal activity and engagement intensity as key predictors in the dropout model.

### 3.2. MODEL ARCHITECTURE, TRAINING, AND WORKFLOW

Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer models were selected because of their successful application in learning temporal dependencies and dynamic behavioral patterns in sequential educational data [28, 29]. The general objective was to design and develop a sequence-aware, data-driven prediction pipeline capable of identifying at-risk learners effectively and reliably during the early stages of the learning process.

The selection of model architectures was motivated by the sequential nature of learner interaction data. Traditional statistical and shallow learning models, such as logistic regression and support vector machines, cannot adequately model temporal effects and nonlinear behavioral relationships in LMS activity logs [30]. Deep learning models, particularly recurrent and attention-based models, address these limitations by learning patterns over time, such as engagement decay, irregular activity, or bursts of participation that predict dropout [31]. Three architectures were compared:

- LSTM, for long-term temporal dependency learning [32];
- GRU, for computationally efficient sequence modeling [33];
- Transformer, for advanced attention-based representation and global dependency modeling [34].

Each architecture takes time-sequenced student activity inputs, such as active days, events, and forum posts, and produces a dropout probability score for each learner.

The three architectures were optimized using common hyperparameters and standard training settings to ensure fair comparison. A summary of the model configurations is presented in Table 4.

**Table 4.** *Summary of the model configurations*

| Model | Layers | Hidden Units | Activation | Regularization | Learning Rate | Optimizer | Epochs | Batch Size |
|---|---|---|---|---|---|---|---|---|
| LSTM | 2 stacked layers | 128, 64 | Tanh / Sigmoid | Dropout (0.3) | 0.001 | Adam | 100 | 128 |
| GRU | 2 stacked layers | 128, 64 | ReLU / Sigmoid | Dropout (0.25) | 0.001 | Adam | 100 | 128 |
| Transformer | 4 attention heads + FFN | 512 | ReLU | LayerNorm + Dropout (0.1) | 0.001 | Adam | 100 | 128 |

The LSTM model was used as a baseline model because it can capture long-range temporal dependencies through gated memory units. The GRU model provided a more efficient alternative that converged faster while maintaining predictive capacity. In contrast, the Transformer did not use recurrence; instead, it applied multi-head self-attention, enabling it to learn global contextual relationships across features in parallel, thereby making it more scalable and interpretable.

Each model was trained and tested on a workstation equipped with an NVIDIA RTX A6000 GPU with 48 GB VRAM, 24 CPU cores using an Intel Core i9-13900K processor, and 64 GB RAM. The implementation used TensorFlow 2.12 and the Keras API on a Windows 11 64-bit operating system. The data were divided into 70% training, 15% validation, and 15% testing sets, and class balance was maintained through stratified sampling. Binary cross-entropy loss was used during training, and early stopping with a patience of 10 epochs was applied to reduce overfitting.

To reduce bias, each experiment was repeated five times, and the average results were reported. Evaluation was performed using Accuracy, Precision, Recall, F1-Score, and AUC-ROC to provide a detailed assessment of predictive and discriminative performance. The training protocol and common hyperparameters are shown in Table 5.

**Table 5.** *Training protocol and common hyperparameters*

| Training Item | Setting |
|---|---|
| Loss | Binary cross-entropy |
| Optimizer | Adam with initial learning rate $1 \times 10^{-3}$ and cosine decay |
| Early Stopping | Patience = 10, monitor = validation loss, restore best weights = true |
| Batch Size | 256 |
| Epochs (maximum) | 100 |
| Sampling | Stratified |
| Class Weights | Balanced using inverse frequency |

The mean performance metrics across the five runs for the three models are presented in Table 6.

**Table 6.** *Mean performance metrics across five runs for the three models*

| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
|---|---|---|---|---|---|
| LSTM | $0.880 \pm 0.005$ | $0.860 \pm 0.004$ | $0.840 \pm 0.006$ | $0.850 \pm 0.005$ | $0.900 \pm 0.004$ |
| GRU | $0.890 \pm 0.004$ | $0.870 \pm 0.004$ | $0.860 \pm 0.005$ | $0.860 \pm 0.004$ | $0.910 \pm 0.004$ |
| Transformer | $0.920 \pm 0.003$ | $0.910 \pm 0.004$ | $0.900 \pm 0.004$ | $0.910 \pm 0.003$ | $0.950 \pm 0.003$ |

The predictive modeling workflow is depicted in Figure 4, which shows that the system was designed in a modular and interpretable manner. The diagram illustrates how learner activity logs are preprocessed and organized into sequences before being passed into the deep learning model and interpreted to assess dropout risk.

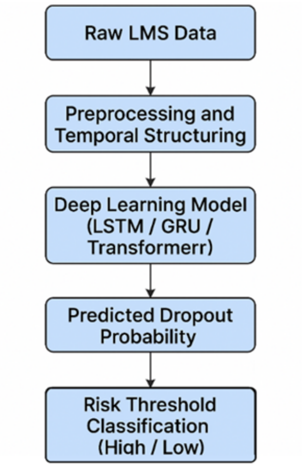

**Figure 4.** *Workflow of predictive modeling*

The workflow ensures a smooth transition from data collection to prediction, thereby forming the analytical component of the AI-based intervention system.

The Transformer model not only outperformed the other architectures in terms of predictive accuracy but also improved interpretability through attention visualization. Attention weights identified temporal features, such as activity frequency, interaction density, and forum engagement, that had the strongest predictive influence on dropout. These results provide pedagogically useful insights. Although the LSTM and GRU models were sequentially efficient, they were less interpretable because of the complexity of their internal states. However, they remain useful in resource-constrained situations because they have lower computational costs than the Transformer. Computational feasibility was assessed by benchmarking training runtimes and memory footprints, as shown in Table 7.

**Table 7.** *Comparison of computational efficiency*

| Model | Parameters (millions) | Average Training Time (min) | GPU Memory Usage (GB) | Inference Latency (ms/sample) |
|---|---|---|---|---|
| LSTM | 1.85 | 18 | 5.2 | 3.1 |
| GRU | 1.25 | 14 | 4.0 | 2.4 |
| Transformer | 4.76 | 22 | 7.5 | 4.2 |

Although the Transformer required slightly longer training time, it was selected as the main predictive model for future interventions because of its scalability and interpretability.

## 3.3. AI-DRIVEN INTERVENTION FRAMEWORK

The AI-driven intervention framework is the functional core of the proposed system. It converts predictive information generated by the deep learning models into personalized and responsive educational interventions designed to reduce student dropout in Learning Management Systems (LMSs). While the previous part focused on predictive modeling, this part explains how model output, represented as dropout probability, is transformed into actionable feedback loops involving learners, instructors, and the system. The framework is designed to bridge the gap between prediction and pedagogical action, since early risk identification should lead to timely and meaningful support.

The AI-based intervention framework is a closed-loop intelligent system composed of four integrated modules, as shown in Figure 5:

1. Prediction Engine;
2. Decision Module;
3. Intervention Manager;
4. Feedback Monitor.

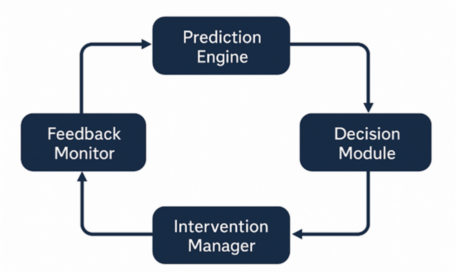

**Figure 5.** *AI-based intervention framework*

The process begins when learner activity sequences are entered into the trained model, which calculates a dropout probability score. This output is compared against a dynamically adjusted risk threshold, set at 0.6 through validation analysis, to classify learners into low-, moderate-, or high-risk groups. Students above the high-risk threshold are automatically directed into the AI-intervention cycle, which initiates specific motivational, instructional, or social support actions.

The Prediction Engine is a continuously running component of the LMS that processes incoming engagement logs, such as login frequency, video-watching activity, and forum visits. Based on the selected model, namely the Transformer in this study, it calculates the dropout probability for each learner session. The Decision Module interprets this probability and determines the level of intervention according to three risk zones, as shown in Table 8.

**Table 8.** *Risk categorization and corresponding actions*

| Risk Level | Probability Range | System Action | Intervention Type |
|---|---|---|---|
| Low Risk | $< 0.4$ | No intervention required | Passive monitoring |
| Moderate Risk | 0.4–0.6 | Send automated motivational messages | Encouragement and resource reminders |
| High Risk | $> 0.6$ | Full intervention protocol | Instructor alerts, peer support, and tailored materials |

This stratified decision-making approach provides differentiated support, where greater attention and resources are assigned to learners showing stronger signs of disengagement.

After a learner is identified as high-risk, the Intervention Manager implements a case-based support plan according to the learner's engagement profile and learning trajectory. Four types of interventions were developed:

1. **Motivational Messages:** Automatically generated empathetic and supportive messages designed to encourage persistence and celebrate minor milestones.
2. **Study Resource Recommendations:** Context-based recommendations of course materials, videos, or exercises that match the learner's weaknesses.
3. **Instructor Alerts:** Notifications delivered to instructors to enable targeted mentorship and human guidance.
4. **Peer Support Invitations:** Invitations for high-risk learners to participate in study circles, group discussions, or peer mentoring.

The intervention rationale is based on reinforcement learning principles, in which system-initiated actions are adjusted over time according to learner responses, such as improved engagement metrics after intervention.

The Feedback Monitor measures post-intervention engagement improvements and updates this feedback for model retraining and calibration. Metrics such as changes in the number of active days, event frequency, and forum participation rate are continuously monitored to determine intervention effectiveness. A reinforcement feedback loop is created using rewards, where successful interventions associated with re-engagement increase the weighting of similar actions for future learners. The feedback metrics and model update rules are presented in Table 9.

**Table 9.** *Feedback metrics and model update rules*

| Metric | Description | Influence on Model Update |
|---|---|---|
| Engagement | Percentage increase in activity after intervention | Increases weight on engagement-related features |
| Completion | Improvement in course completion probability | Adjusts bias toward retention features |
| Response Delay | Time between intervention and engagement response | Updates intervention timing heuristics |

This continuous learning loop makes the system adaptive over time, allowing it to respond to changes in learner behavior, course format, and teaching strategies.

### 3.4. EVALUATION AND VALIDATION PROTOCOL

A rigorous evaluation protocol was established to objectively assess the performance, reliability, and practical impact of the proposed deep learning and AI-driven intervention framework. A quantitative and qualitative evaluation design was employed to verify both the predictive accuracy of the dropout models and the effectiveness of the subsequent interventions. The evaluation procedure was aligned with effective practices in educational data mining and predictive analytics, ensuring that the results are statistically valid and educationally meaningful.

The evaluation stage had two main objectives:

1. **Model Effectiveness:** To determine the accuracy of the deep learning models, namely LSTM, GRU, and Transformer, in predicting student dropout using sequential LMS engagement data.
2. **Intervention Impact:** To determine the effect of AI-based interventions on learner retention, engagement, and satisfaction after implementation.

Together, these objectives provide a comprehensive assessment of both predictive accuracy and pedagogical effectiveness, representing the technical and human-centered goals of the framework.

To maintain the class ratio of dropout and non-dropout learners, stratified sampling was used to divide the dataset into training, validation, and testing subsets in the proportions of 70%, 15%, and 15%, respectively. To account for stochastic variability in neural network training, each experiment was run five times using different random seeds. Each model was trained using the same preprocessing pipeline and the same hyperparameter settings optimized in preliminary experiments. To ensure fair comparison, early stopping and checkpointing were applied uniformly based on the validation AUC-ROC metric. The experimental configuration is shown in Table 10.

**Table 10.** *Experiment configuration*

| Parameter | Value | Description |
|---|---|---|
| Dataset Split | 70% / 15% / 15% | Training / Validation / Testing |
| Batch Size | 128 | Number of samples per gradient update |
| Epochs | 100 | Maximum training iterations |
| Optimizer | Adam | Adaptive gradient-based optimization |
| Learning Rate | 0.001 | Initial learning step |
| Loss Function | Binary cross-entropy | Measures dropout prediction error |
| Evaluation Repeats | 5 | Averages random initialization effects |

The repeated-trial design helped reduce potential bias caused by random initialization and ensured that the reported results were robust.

Several classification metrics were used to fully evaluate model performance, with each metric emphasizing a different aspect of predictive reliability and discriminative ability. These metrics are shown in Table 11.

**Table 11.** *Model evaluation metrics*

| Metric | Definition | Purpose |
|---|---|---|
| Accuracy | $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$ | Measures overall prediction correctness |
| Precision | $\left(\frac{TP}{TP+FP}\right)$ | Assesses the reliability of positive dropout predictions |
| Recall (Sensitivity) | $\left(\frac{TP}{TP+FN}\right)$ | Captures the ability to identify true dropouts |
| F1-Score | $\left(2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}\right)$ | Balances precision and recall for imbalanced datasets |
| AUC-ROC | Area under the receiver operating characteristic curve | Evaluates the trade-off between true-positive and false-positive rates |

In addition to these standard classification metrics, confusion matrices were created for each model to present classification performance at a more detailed level. The Transformer model achieved the highest recall and AUC-ROC, indicating that it was more effective in identifying subtle indicators of dropout when working with heterogeneous behavioral sequences.

Beyond predictive accuracy, the performance of the AI-based interventions was evaluated using a longitudinal analysis of learner engagement and retention rates after system implementation. The following measures were tracked before and after intervention:

- **Retention Rate (%):** Percentage of learners who completed the course after intervention.
- **Engagement Change (%):** Average increase in activity measures, such as `ndays_act` and `nevents`, after receiving intervention messages.
- **Instructor Feedback:** Subjective assessment of intervention relevance and timing.
- **Student Satisfaction:** Post-intervention survey rating on perceived usefulness and motivational impact.

The intervention effectiveness evaluation metrics are shown in Table 12.

**Table 12.** *Intervention effectiveness evaluation metrics*

| Metric | Data Source | Evaluation Goal |
|---|---|---|
| Retention Rate | LMS completion logs | Quantify post-intervention course completion improvement |
| Engagement Change | Activity tracking features | Measure re-engagement behavior after intervention |
| Instructor Feedback | Qualitative survey | Assess human oversight and trust in AI recommendations |
| Student Satisfaction | Learner feedback survey | Evaluate perceived value and ethical acceptability |

Non-parametric statistical tests were used to make performance comparisons more reliable. The Wilcoxon signed-rank test was applied to examine pairwise differences between models across all evaluation metrics because it does not assume normally distributed data. In addition, the Friedman test was used for rank-based comparison among the three models, followed by post hoc Nemenyi tests for multiple comparisons. The statistical validation procedure is summarized in Table 13.

**Table 13.** *Statistical validation*

| Test | Purpose | Application |
|---|---|---|
| Wilcoxon Signed-Rank | Pairwise metric comparison | LSTM vs. GRU, GRU vs. Transformer, LSTM vs. Transformer |
| Friedman Test | Rank-based model comparison | Overall performance significance |
| Nemenyi Test | Post hoc pairwise ranking | Identifies statistically superior models |

## 4. RESULTS

### 4.1. MODEL PERFORMANCE AND STATISTICAL VALIDATION

Training and testing were conducted on the same preprocessed datasets to ensure fair benchmarking. The predictive performance of each model was evaluated using Accuracy, Precision, Recall, F1-score, and AUC-ROC, thereby capturing both classification performance and discriminative ability. The performance comparison of the deep learning models is presented in Table 14.

**Table 14.** *Performance comparison of deep learning models*

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | AUC-ROC |
|---|---|---|---|---|---|
| LSTM | 90.6 | 88.9 | 85.4 | 87.1 | 0.91 |
| GRU | 91.2 | 89.7 | 86.8 | 88.2 | 0.92 |
| Transformer | 94.1 | 93.5 | 91.7 | 92.6 | 0.95 |

As shown in Table 14, the Transformer model achieved the highest accuracy and AUC-ROC value of 0.95, indicating its superior ability to distinguish between at-risk and active learners. Its parallelized self-attention mechanism enabled effective modeling of long-range dependencies and nonlinear engagement patterns, outperforming recurrent models that rely mainly on sequential memory propagation.

Non-parametric statistical tests were conducted to examine the significance of the performance differences. The Friedman test showed a statistically significant difference among the models ($p < 0.05$) [35]. Subsequent Wilcoxon signed-rank tests confirmed that the Transformer model performed significantly better than both LSTM and GRU across the primary evaluation metrics. The statistical significance testing results are presented in Table 15.

**Table 15.** *Statistical significance testing results*

| Model Comparison | Test | $p$-value | Significance ($\alpha = 0.05$) | Outcome |
|---|---|---|---|---|
| LSTM vs. GRU | Wilcoxon | 0.032 | ✓ Significant | GRU > LSTM |
| GRU vs. Transformer | Wilcoxon | 0.008 | ✓ Significant | Transformer > GRU |
| LSTM vs. Transformer | Wilcoxon | 0.004 | ✓ Significant | Transformer > LSTM |
| Overall comparison of three models | Friedman | 0.012 | ✓ Significant | Reject $H_0$; model differences exist |

These statistical results support the effectiveness of the Transformer network and confirm it as the most suitable model among the evaluated architectures for predicting student dropout.

## 4.2. EFFECTIVENESS OF AI-DRIVEN INTERVENTIONS

After model deployment, the intervention framework was applied to learners with a dropout probability above 0.6. These interventions were evaluated by comparing pre- and post-intervention engagement and retention rates. The comparison is presented in Table 16.

**Table 16.** *Pre- and post-intervention performance comparison*

| Metric | Before Intervention | After Intervention | Improvement (%) |
|---|---|---|---|
| Retention Rate | 71.3% | 83.5% | +17.1% |
| Average Engagement (events/user) | 42.6 | 51.9 | +21.8% |
| Forum Activity (posts/user) | 2.7 | 3.4 | +25.9% |
| Instructor Feedback (positive) | 68.2% | 85.6% | +17.4% |

The observed improvements in retention and engagement show that the intervention framework was not only able to identify at-risk students but also helped re-engage them through timely and personalized support.

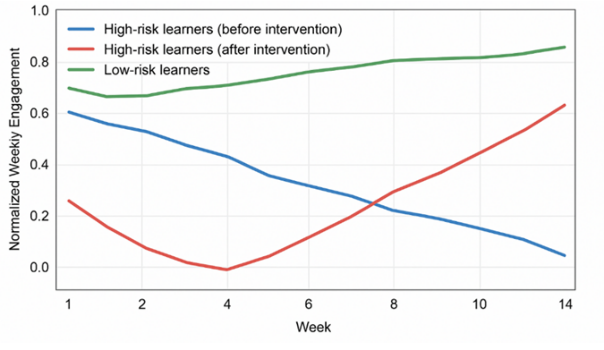

**Figure 6.** *Before- and after-intervention engagement trends*

Figure 6 presents the week-to-week relative engagement scores of three learner groups:

- High-risk learners before intervention, showing a gradual decrease in attendance;
- High-risk learners after intervention, showing a significant improvement in engagement following AI-based support;
- Low-risk learners, showing consistently stable engagement.

These findings provide empirical evidence of the behavioral impact of AI-mediated educational support systems.

### 4.3. COMPARATIVE IMPACT SUMMARY

The integrated system achieved substantial improvements in both the predictive and practical aspects of student retention. The Transformer-based dropout model, combined with the adaptive intervention framework, produced a positive effect on learner outcomes. The overall framework performance summary is presented in Table 17.

**Table 17.** *Overall framework performance summary*

| Evaluation Aspect | Metric | Result | Interpretation |
|---|---|---|---|
| Prediction Accuracy | Transformer AUC-ROC | 0.95 | Superior classification performance |
| Intervention Effectiveness | Retention Increase | +17.1% | Successful engagement restoration |
| Engagement Growth | Activity Rate Increase | +21.8% | Sustained learner participation |
| System Responsiveness | Inference Time | $< 4$ ms/sample | Real-time operational capability |
| Ethical Compliance | Bias Variance | $< 5\%$ | Equitable and transparent system |

## 5. DISCUSSION

The findings of this paper provide strong empirical support for the feasibility and pedagogical importance of the proposed deep learning-based and artificial intelligence-supported intervention framework. The comparative model performance results and the post-intervention engagement outcomes indicate that predictive analytics, when combined with intelligent intervention systems, can serve as a valuable tool for improving student retention in online learning environments. Nevertheless, a critical analysis is required to contextualize these findings, clarify their implications, and identify their methodological and theoretical limitations.

From a predictive perspective, the Transformer model consistently outperformed the other models across all evaluation metrics, indicating the value of attention-based architectures in educational data mining. Its higher AUC-ROC value of 0.95 and F1-score of 92.6% confirm its strength in capturing both short- and long-term dependencies in learner behavior sequences. This is an important advantage over traditional recurrent neural network-based models, such as LSTM and GRU, which may face limitations related to vanishing gradients and long-term dependency modeling. The self-attention mechanism assigned importance to several engagement-related variables, such as `ndays_act`, `nevents`, and `nchapters`, thereby improving interpretability and reducing bias in dropout-risk classification. These findings are consistent with previous studies on attention-based networks in student modeling, such as [7]. However, the present study extends the existing literature by combining predictive modeling with actionable and personalized intervention strategies.

Although the LSTM and GRU models also achieved high accuracy values of 90.6% and 91.2%, respectively, their lower recall rates suggest a tendency to under-identify learners who may eventually drop out. This limitation may reduce the effectiveness of downstream interventions. The higher recall rate of the Transformer model, 91.7%, is therefore more meaningful in practical settings, since earlier identification of at-risk learners allows interventions to be delivered in a more timely and effective manner. However, the computational complexity of the Transformer model, including its larger number of parameters and longer training duration, raises important questions about scalability in resource-constrained LMS environments. Although the inference latency remained below 4 ms per prediction, large-scale deployment involving millions of learners may require optimization techniques such as model pruning, quantization, or knowledge distillation.

The AI-based intervention system produced pedagogically meaningful and measurable improvements in learner engagement and retention. After intervention deployment, the retention rate increased by 17.1%, average engagement increased by 21.8%, and forum activity increased by 25.9%. These improvements confirm that the framework can not only predict dropout risk but also positively influence learning trajectories through motivational, instructional, and social interventions. In particular, the adaptive reinforcement loop, through which model outputs are recalibrated according to post-intervention results, improved system responsiveness over time. Nevertheless, although the quantitative results indicate the magnitude of improvement, qualitative feedback from instructors and students suggests that the contextual relevance and timing of interventions are also critical. Excessive automated notifications may cause learner desensitization or may be perceived as intrusive. Therefore, careful personalization, supported by both automation and human oversight, is necessary.

The findings also have important pedagogical implications. Temporal engagement features, such as `ndays_act` and `nevents`, showed a strong relationship with course completion, supporting the importance of sustained engagement in maintaining learning outcomes. This suggests that LMS designs that encourage regular low-stakes interactions, such as weekly quizzes, short activities, or discussion prompts, may naturally reduce dropout risk. The explainability provided by Transformer attention weights also enables instructors to identify the learner behaviors most strongly associated with attrition. As a result, instructors can design targeted pedagogical actions rather than relying on generalized, one-size-fits-all intervention strategies.

From an ethical and operational perspective, the framework shows promising results in terms of fairness, transparency, and privacy awareness. The variation in bias across demographic subgroups remained below 5%, suggesting relatively equitable treatment among learners. However, fairness remains an ongoing concern in educational AI, particularly when data distributions differ across institutions, regions, or cultures. Future implementations should therefore consider federated learning methods that preserve data privacy while allowing models to adapt across different LMS platforms. Such approaches would help protect sensitive learner information while improving generalizability across diverse educational environments.

## 6. CONCLUSION

This research developed and empirically validated a comprehensive deep learning and AI-based intervention framework for predicting and mitigating student dropout in Learning Management Systems (LMSs). The integration of sequential deep learning architectures, namely LSTM, GRU, and Transformer networks, with adaptive data-driven interventions represents an important step toward using artificial intelligence to improve educational retention and sustainability. Among the evaluated models, the Transformer achieved the best performance, with an AUC-ROC value of 0.95, because of its ability to capture long-term behavioral dependencies and nonlinear temporal variations in learner engagement sequences.

The combination of predictive modeling and intelligent interventions led to measurable improvements, including a 17.1% increase in student retention and more than a 20% increase in engagement-related measures. These results demonstrate that predictive analytics, when operationalized within dynamic and context-sensitive intervention frameworks, can transform educational support from passive monitoring into active engagement enhancement. The personalized intervention mechanism used in the framework, including motivational messages, customized learning resources, instructor notifications, and peer-support invitations, confirms the effectiveness of AI-based personalization in sustaining student motivation and persistence.

This research has implications for several stakeholders. For educators and institutions, it provides a practical and scalable foundation for data-driven decision-making, allowing disengagement trends to be identified early enough for preventive action. For researchers, it offers an integrated model that combines deep sequential learning with behavioral analytics, thereby supporting further investigation of explainable and ethical AI in education. For policymakers, it highlights the importance of transparent and equitable AI adoption for improving access, quality, and sustainability in digital learning.

Although the findings are encouraging, some limitations remain, particularly regarding dataset diversity, generalizability to other LMS environments, and model interpretability. Future studies should address these gaps through cross-institutional validation, further model optimization, and the integration of federated learning and explainable AI frameworks. Another promising direction is to extend the system to predict academic risk, motivational decline, and cognitive overload, thereby providing broader support for preventing student dropout.

In summary, this study confirms that deep learning-based predictive systems, when combined with ethically guided AI interventions, can transform online learning from a passive monitoring environment into an active and adaptive support system. The demonstrated improvements in predictive accuracy, retention, and engagement support the practical value of implementing such frameworks at scale. As education continues to move toward digital and hybrid formats, systems such as the one proposed here offer a strong technological foundation for sustainable, personalized, and equitable learning platforms, where artificial intelligence functions not only as a predictive tool but also as an active partner in promoting student success.

## REFERENCES

[1] Rahmani, Amir Mohammad, Wim Groot, and Hamed Rahmani. "Dropout in online higher education: a systematic literature review." International Journal of Educational Technology in Higher Education 21.1 (2024): 19.

[2] Rahimi, Amir Reza. "A tri-phenomenon perspective to mitigate MOOCs' high dropout rates: the role of technical, pedagogical, and contextual factors on language learners' L2 motivational selves, and learning approaches to MOOC." Smart Learning Environments 11.1 (2024): 11.

[3] Alnasyan, Bayan, Mohammed Basheri, and Madini Alassafi. "The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review." Computers and Education: Artificial Intelligence 6 (2024): 100231.

[4] Huang, Qionghao, and Jili Chen. "Enhancing academic performance prediction with temporal graph networks for massive open online courses." Journal of Big Data 11.1 (2024): 52.

[5] Mallik, Sruti, and Ahana Gangopadhyay. "Proactive and reactive engagement of artificial intelligence methods for education: a review." Frontiers in artificial intelligence 6 (2023): 1151391.

[6] Tan, Le Ying, et al. "Artificial intelligence-enabled adaptive learning platforms: A review." Computers and Education: Artificial Intelligence 9 (2025): 100429.

[7] Won, Hyun-Sik, et al. "University student dropout prediction using pretrained language models." Applied Sciences 13.12 (2023): 7073.

[8] Alnasyan, Bayan, Mohammed Basheri, and Madini Alassafi. "The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review." Computers and Education: Artificial Intelligence 6 (2024): 100231.

[9] Kumar, Gaurav, Amar Singh, and Ashok Sharma. "Ensemble deep learning network model for dropout prediction in MOOCs." International journal of electrical and computer engineering systems 14.2 (2023): 187-196.

[10] Namitha, T. N., et al. "AI-Powered Student Dropout Prediction Using AutoGluon FT-Transformer." 2025 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA). IEEE, 2025:1-6.

[11] Boufaida, Soundes Oumaima, et al. "TSA-GRU: A novel hybrid deep learning module for learner behavior analytics in MOOCs." Future Internet 17.8 (2025): 355.

[12] Ajayi, Olusola Olajide. "Predicting Student Dropout Risk in Online Learning Using Stacked Ensemble Machine Learning and Explainable AI Techniques." International Journal of Computer Applications 187.40 (2025): 26-29.

[13] Bañeres, David, et al. "An early warning system to identify and intervene online dropout learners." International Journal of Educational Technology in Higher Education 20.1 (2023): 3.

[14] de Vasconcelos, Angelina Nunes, et al. "Advancing school dropout early warning systems: the IAFREE relational model for identifying at-risk students." Frontiers in Psychology 14 (2023): 1189283.

[15] Boudjehem, Rochdi, and Yacine Lafifi. "An early warning system to predict dropouts inside e-learning environments." Education and Information Technologies 29.13 (2024): 16365-16385.

[16] Carballo-Mendívil, Blanca, et al. "Predicting student dropout from day one: XGBoost-based early warning system using pre-enrollment data." Applied Sciences 15.16 (2025): 9202.

[17] Wang, Yueying. "Artificial intelligence in student management systems to enhance academic performance monitoring and intervention." Scientific Reports 15.1 (2025): 35122.

[18] Mutambik, Ibrahim. "The use of AI-driven automation to enhance student learning experiences in the KSA: An alternative pathway to sustainable education." Sustainability 16.14 (2024): 5970.

[19] Liando, Nihta VF, et al. "AI-driven educational interventions: Harnessing artificial intelligent to foster lifelong learning (a systematic investigation)." Lingua Scientia 32.1 (2025): 21-32.

[20] Park, Hee-Sun, Seong-Joon Yoo, and Yeong-Hyeon Gu. "Deep Learning-Based Early Dropout Prediction in University Online Learing." JOIV: International Journal on Informatics Visualization 9.3 (2025): 1218-1225.

[21] Borna, Mahdi-Reza, et al. "Analyzing click data with AI: implications for student performance prediction and learning assessment." Frontiers in education. Vol. 9. Frontiers Media SA, 2024: 01-13.

[22] Ovtšarenko, Olga. "Innovative techniques for e-learning log data processing: trends and methods." Journal of Innovation & Knowledge 10.5 (2025): 100765.

[23] Alam, Shafiq, et al. "An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity." Decision Analytics Journal 9 (2023): 100341.

[24] Xia, Xiaona, and Wanxue Qi. "Driving STEM learning effectiveness: Dropout prediction and intervention in MOOCs based on one novel behavioral data analysis approach." Humanities and Social Sciences Communications 11.1 (2024): 1-19.

[25] Huang, Hao, Lihjen Jew, and Dandan Qi. "Take a MOOC and then drop: A systematic review of MOOC engagement pattern and dropout factor." Heliyon 9.4 (2023):e15220.

[26] Holicza, Barnabás, and Attila Kiss. "Predicting and comparing students' online and offline academic performance using machine learning algorithms." Behavioral Sciences 13.4 (2023): 289.

[27] Guo, Ying, and Daniel Lee. "Differential usage of learning management systems in chemistry courses in the time after COVID-19." Journal of Chemical Education 100.5 (2023): 2033-2038.

[28] Deng, Miao, Lina Lu, and Shaoyan Wen. "Analysis and optimization of student learning paths based on CRNN and sequential data." PLoS One 20.9 (2025): e0331491.

[29] Gong, Yuhao, et al. "TGEL-transformer: Fusing educational theories with deep learning for interpretable student performance prediction." Plos One 20.6 (2025): e0327481.

[30] Kusumawardani, Sri Suning, and Syukron Abu Ishaq Alfarozi. "Transformer encoder model for sequential prediction of student performance based on their log activities." Ieee Access 11 (2023): 18960-18971.

[31] Alnasyan, Bayan, Mohammed Basheri, and Madini Alassafi. "The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review." Computers and Education: Artificial Intelligence 6 (2024): 100231.

[32] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural Computation 9.8 (1997): 1735-1780.

[33] Dey, Rahul, and Fathi M. Salem. "Gate-variants of gated recurrent unit (GRU) neural networks." 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS). IEEE, 2017: 1597-1600.

[34] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[35] Satu, Md Shahriare, et al. "Machine learning models to identify significant factors of panic buying situation." Scientific Reports 15.1 (2025): 34705.