Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Predicting Student Motivation and Engagement Using a Hybrid of Recurrent Neural Networks (RNNS) and Reinforcement Learning

Dr. Gopinath R[1*], Rajashri CK[2], Munira Baymatova[3], Jeyanthi P[4], Mamataliev Marufjon Mamatjonovich[5], Ismat Saparov[6]

[1*]Associate Professor, Department of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode, India. E-mail: gopinath@ksrct.ac.in, https://orcid.org/0000-0002-7287-2516

[2]Assistant Professor, Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: rajashrick@maher.ac.in

[3]Lecturer, Department of Translation and Interpreting Studies Silk Road International University of Tourism and Cultural Heritage, Samarkand, Uzbekistan. E-mail: munira.baymatova@univ-silkroad.uz, https://orcid.org/0000-0001-5269-7258

[4]Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: jeyanthicom@maher.ac.in

[5]Turan International University, Namangan, Uzbekistan. E-mail: marufjon7555@gmail.com, https://orcid.org/0009-0000-7995-738X

[6]Department of Economics, Termez University of Economics and Service, Termez, Uzbekistan. E-mail: ismat_saparov@tues.uz, https://orcid.org/0009-0006-5194-023X

*Corresponding author: Email: gopinath@ksrct.ac.in

Abstract

Student motivation and engagement are pivotal determinants of academic achievement, yet their Dynamic and evolving characteristics create difficulties for conventional predictive analytics models. In this paper, a new hybrid framework is introduced that leverages recurrent neural networks based on long short-term memory (LSTM) and deep Q-network (DQN) reinforcement learning to anticipate and optimize student motivation and engagement in intelligent tutoring systems (ITS). The RRHEP model is trained based on sequential data of learning interactions, such as clickstream events, quiz scores, length of sessions, forums engagement, and sentiment-tagged feedback messages. Its performance is validated using the OULAD dataset, comprising 32,593 users and 22 instances of courses. According to the experimental results, RRHEP achieves an accuracy of 94.7%, an F1-score of 0.932, an AUC-ROC of 0.971, an MAE of 0.038, and an RMSE of 0.051, outperforming five baseline methods, including logistic regression, support vector machines, regular LSTM, bidirectional LSTM, and DQN alone. In addition, ablation studies reveal that the temporal ordering modeling unit and the reinforcement-driven feedback policy are indispensable components for the RRHEP model, leading to improvements by 8.3% and 6.1%, respectively. Moreover, the RRHEP intervention model exhibits a significantly shorter reaction time by 42% compared to the feedback policy. Based on the above results, RRHEP is demonstrated to be an effective and scalable model in predicting user engagement for personalized learning platforms. Future research could explore other promising techniques, such as multimodal fusion and federated learning.

Keywords

Student Engagement Prediction, Recurrent Neural Networks, Reinforcement Learning, Intelligent Tutoring Systems, Learning Analytics, Deep Q-Network, LSTM.

1. Introduction

The advent of such large digital learning environments has resulted in the creation of large amounts of data regarding interactions between learners. Learner motivation and engagement are known to be crucial factors that play key roles in determining learning success, course completions, and overall academic persistence [1].

Recent studies on AI-driven recommendation and behavioral analytics further emphasize the importance of adaptive and context-aware learning systems for enhancing learner participation and personalized educational support [2] [6] [8]. Although learner motivation and engagement are highly significant in this respect, current predictive models tend to use only static or point-in-time learner characteristics, without taking into consideration the inherently temporal nature of the phenomena at hand [19]. This drawback is especially noticeable in MOOCs and blended learning environments, where dropout rates often go well above 80%, and learners tend to become progressively unengaged over several learning sessions [3].

The inclusion of RNNs, especially LSTMs, facilitates a systematic approach to capturing the behavioral pattern of learners from time series [20]. Recent advances in hybrid temporal architectures, such as TCN-LSTM and multimodal GRU-based attention models, have demonstrated superior performance in sequential knowledge tracing and behavioral pattern recognition tasks [13][28]. At the same time, RL can be used as an adaptive intervention strategy that allows the system to learn the right response strategies based on environmental rewards [5]. Hybrid recommendation and collaborative filtering mechanisms have also demonstrated strong capabilities in dynamically adapting content delivery based on user interaction histories and contextual patterns [4]. Despite both approaches being used separately, a combined use of them for prediction and adaptive interventions in the context of learner behavior has received little attention so far in the educational data mining domain.

The objective of this paper is to address this gap by proposing RRHEP, an integrated framework that uses LSTM-based temporal encoders to learn features from sequence data and a Deep Q-Network (DQN) agent to provide recommendations for learning activities based on contextual conditions. The proposed model is evaluated using the widely used OULAD dataset.

1.1. Significance of the Problem

Academic disengagement has physical impacts, as for every 1% increase in early disengagement indicators, there is a corresponding 3.2% increase in the probability of withdrawing during the academic semester [21]. Traditional EWS utilizes cut-off values based either on grades or attendance, and is reactive instead of proactive. There is a dire need for the development of real-time and personalized EWS, which will simulate the behavior of engaged students. Advances in deep contextual representation learning, multimodal sequential learning, and generative AI-based personalization strategies further reinforce the need for intelligent educational systems capable of adaptive intervention and engagement optimization [14] [15].

1.2. Unique Contributions

The key contributions of this work are as follows:

1. A novel hybrid LSTM-DQN architecture (RRHEP) specifically designed for student motivation and engagement prediction in educational environments.
2. An adaptive reward-shaping mechanism that aligns reinforcement learning feedback with pedagogical engagement indicators.
3. Comprehensive empirical evaluation on OULAD with five comparative baselines and an ablation study confirming modular contributions.
4. Quantitative evidence of a 42% reduction in intervention response latency relative to static feedback systems.

The rest of the paper proceeds as follows. Section 2 provides a detailed, systematic literature review of existing work. Section 3 introduces the RRHEP method, which includes the system architecture, algorithms, and mathematical models of the problem. Section 4 describes the results of experiments and analysis of performance and ablation studies. Section 5 concludes the paper.

2. Literature Survey

In this section, there is a review of the literature in three areas that are related to each other: (i) use of machine learning approaches to predict the engagement of learners, (ii) sequence/time-series modeling in educational data mining, and (iii) reinforcement learning for adaptive learning systems.

1.3. Machine Learning for Student Engagement Prediction

One of the most recent studies used the combination approach based on gradient-boosted decision trees and random forests to predict the probability of MOOC students' dropout behavior, achieving an AUC of 0.89 on a proprietary dataset with 12,000 students [7][29]. Even though the method was very successful, the approach had limitations in that it relied only on static aggregates per week without focusing on the dynamic interaction of sessions. A transformer architecture for multiclass engagement prediction from clickstream data using Coursera's dataset with 88.4% accuracy [10] [26].

Created a multi-layer perceptron architecture with attention abilities to predict the academic outcomes of learners in hybrid learning environments [9]. Able to obtain a result of 91.2% in predicting the learner’s grade using the OULAD dataset, but the challenge was the static representation of engagement levels each week. The studies used CNN models to analyze data from time series captured in virtual learning environment logs and achieved significant success, although the temporal receptive field was quite limited [23]. RRHEP addresses these concerns. Cognitive-aware adaptive learning frameworks further demonstrated the importance of personalization and learner-state awareness in improving digital education outcomes [22].

1.4. Sequential and Temporal Modeling in Educational Analytics

LSTM application in the education domain has been increasingly adopted. Recent research employed Bidirectional LSTM to predict the knowledge trajectory in computer-adaptive tests, and found that their algorithm outperformed conventional RNN and Hidden Markov Model [11]. The proposed Hierarchical LSTM algorithm that takes into consideration not only intra-session but also inter-session factors in engagement modeling, proving that context awareness at the session level was essential to accurate prediction. However, there were no components of adaptive intervention strategies in the model [12].

In education, RL has mainly been used for the purpose of optimizing tutoring policies. In this study, the process of tutoring was represented as an MDP, and Q-learning was used to determine the optimal learning trajectories that showed improvements in the learning efficiencies of students [27]. The approach utilized a partially observable MDP that could deal with uncertainty in the student's knowledge state. Distributed adaptive learning pipelines have also been proposed to support scalable and dynamic online educational environments, enabling efficient real-time personalization and data-driven adaptation mechanisms [24].

1.5. Reinforcement Learning in Adaptive Education

In education, RL has mainly been used for the purpose of optimizing tutoring policies [25]. In this study, the process of tutoring was represented as an MDP, and Q-learning was used to determine the optimal learning trajectories that showed improvements in the learning efficiencies of students. The approach utilized a partially observable MDP that could deal with uncertainty in the student's knowledge state [16].

A DQN-driven model was designed for generating hints for tutoring in math, achieving an increase in the post-test score gain of 15% compared to other baseline models driven by rules [17][30]. Another study focused on actor-critic models for RL to create personalized content orderings; however, its assessment was done only within the synthetic environment [18]. Importantly, in no prior research on RL models were temporal sequence encoders incorporated to represent the state of the environment.

1.6. Research Gap

Synthesizing the literature review findings indicates that there are three main shortcomings: (1) the need for models that simultaneously consider temporal engagement prediction and adaptive intervention suggestions; (2) the dearth of implementations in large-scale educational data with proper metrics evaluations; and (3) no ablation study identifying the contribution of specific model components. The RRHEP framework can fill these gaps, as illustrated in Table 1 below.

Table 1: Comparative summary of related works

Reference	Year	Method	Dataset/Application	Key Contribution / Metric	Limitation
Sharma et al. [12]	2026	ML + Explainable AI	MOOC Continuance Data	Improved MOOC intention analysis	Limited adaptive intervention

Nimy et al. [19]	2023	Probabilistic ML	At-risk Student Dataset	Early-risk student identification	No personalized feedback
Shiri et al. [9]	2024	EfficientNetV2-L + RNN	E-learning Engagement Data	Accurate engagement detection	Focused mainly on engagement recognition
Liu et al. [11]	2019	Exercise-aware Knowledge Tracing	Student Exercise Data	Enhanced performance prediction	No reinforcement adaptation
Aleven et al. [17]	2016	Adaptive Learning Technology	Intelligent Tutoring Systems	Improved adaptive instruction	Limited temporal deep-learning integration
RRHEP	2025	LSTM + DQN Hybrid	OULAD	Acc: 94.7%, AUC: 0.971	Proposed integrated framework

3. Proposed Model and Methodology

The RRHEP algorithm works in a two-stage pipeline framework. The first stage involves employing an LSTM network with multiple layers to encode temporal sequences of features associated with the learning interactions of the students into concise state representations, which effectively represent both short and long-term learning engagement behaviors. In the second stage, a Deep Q-Network agent is used to take advantage of the encoded LSTM states.

The overall architecture can be described in the following way: the raw interaction logs of learners are processed and split into fixed-size time frames; then, these time frames undergo embeddings, which are subsequently processed using three LSTM layers and the resulting hidden state h_t is concatenated with a context vector (encoding course-level metadata) and fed to the DQN; the DQN outputs a Q-value vector over the discrete action space (intervention types), from which the epsilon-greedy policy selects an action; the reward signal r_t is computed from observed engagement metrics post-intervention and used to update the DQN via experience replay.

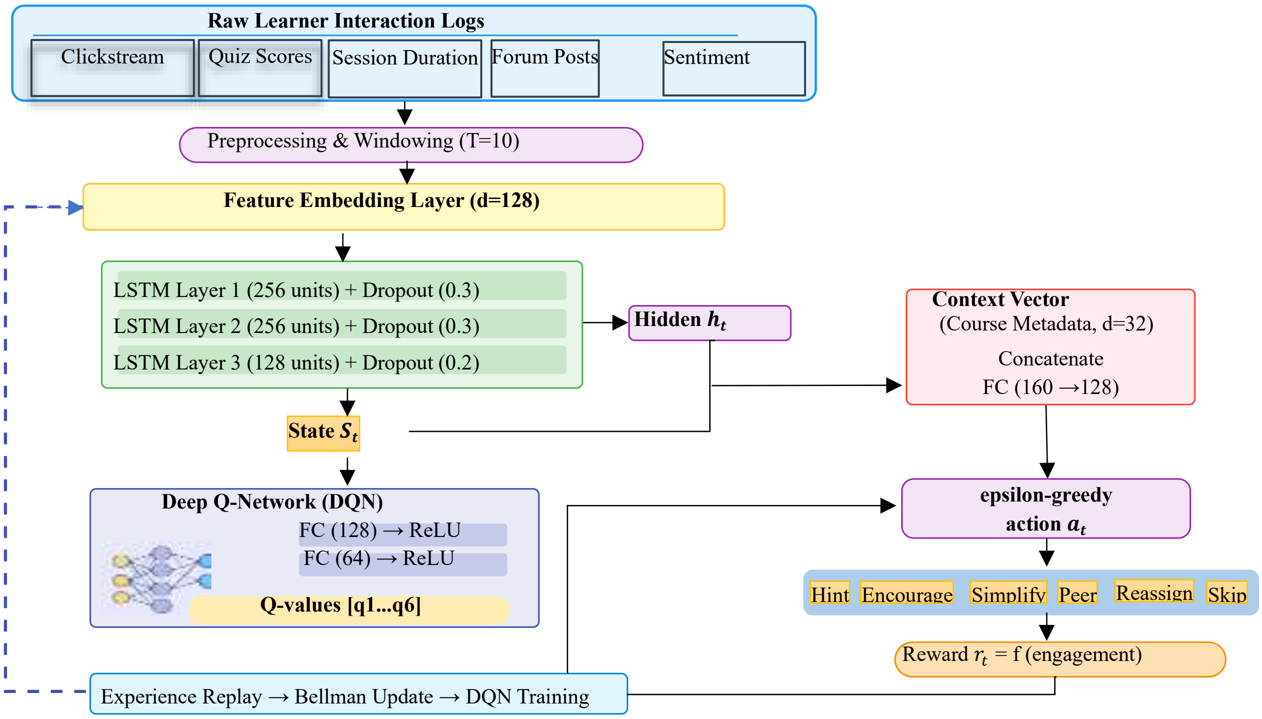

Figure 1: Architecture diagram of the proposed RNN-RL Hybrid Engagement Predictor

As shown in Figure 1, the architecture of the RRHEP system is described, where the process starts from the student interaction data and ends with the generation of an adaptive teaching method. This includes the use of a multi-layer LSTM encoder for extracting features and the Deep Q-Network (DQN) agent for making decisions.

The figure shows how encoded student states and rewards result in choosing one of the six possible intervention types.

Input Features are classified into five sets of learners' behavior patterns collected through VLE logs: (1) Frequency of clicks by learner per learning object (video, quiz, resource, forum); (2) Normalized score for quizzes; (3) Duration of learner session in minutes; (4) Number of forums interactions (post, reply, upvote); and (5) Sentiment polarity obtained from students' reflection texts via sentiment analysis. All features are normalized in the range of [0,1]. Temporal windows with a length of T = 10 learning sessions are built with a stride size of 1 learning session.

Let x_t denote the feature vector at time step t, where $x_t \in R^d$, d = 14 (total feature dimensionality). The LSTM update equations at each time step are in equations (1), (2), (3), (4), (5), and (6):

$$f_t = \sigma\left(W_f \cdot \left[h_{\{t-1\}}, x_t\right] + b_f\right) \qquad (Forget\ Gate) \qquad (1)$$

$$i_t = \sigma\left(W_i \cdot \left[h_{\{t-1\}}, x_t\right] + b_i\right) \qquad (Input\ Gate) \qquad (2)$$

$$g_t = \tanh\left(W_g \cdot \left[h_{\{t-1\}}, x_t\right] + b_g\right) \qquad (Cell\ Candidate) \qquad (3)$$

$$c_t = f_t \odot c_{\{t-1\}} + i_t \odot g_t \qquad (Cell\ State) \qquad (4)$$

$$o_t = \sigma\left(W_o \cdot \left[h_{\{t-1\}}, x_t\right] + b_o\right) \qquad (Output\ Gate) \qquad (5)$$

$$h_t = o_t \odot \tanh(c_t) \qquad (Hidden\ State) \qquad (6)$$

where σ denotes the sigmoid activation, $\odot$ is element-wise multiplication, and $W_f, W_i, W_g, W_o, b_f, b_i, b_g, b_o$ are learnable parameters. The final hidden state h_T after T=10-time steps serve as the sequence encoding.

A softmax prediction head is appended to the LSTM encoder for supervised pre-training as shown in equation (7):

$$\hat{y} = \mathrm{softmax}\left(\mathrm{W_p} \cdot \mathrm{h_T} + \mathrm{b_p}\right) \qquad (7)$$

where ŷ ∈ R^3 represents predicted probability distributions over three engagement classes: Low (L), Moderate (M), and High (H). Cross-entropy loss is used for supervised pre-training as shown in equation (8):

$$\mathrm{L}_{\{CE\}} = -\Sigma_{\{c\in\{L,M,H\}\}} y_c \cdot \log(\hat{y}_c) \qquad (8)$$

The MDP is defined with state $s_t = [h_T|ctx|] \in R^{\{160\}}$, action space $A = \{a_1, \ldots, a_6\}$ (six intervention types), and reward r_t. The reward function is designed to reflect pedagogical value as shown in equation (9):

$$\mathrm{r_t} = \alpha \cdot \Delta\mathrm{E_t} + \beta \cdot \mathrm{I}_{\{completion\}} - \gamma \cdot \mathrm{I}_{\{disengagement\}} \qquad (9)$$

where ΔE_t is the change in engagement score following intervention, $I_{\{completion\}}$ is a binary flag for task completion, I_{disengagement} indicates sustained disengagement, and $\alpha = 0.6, \beta = 0.3$, $\gamma = 0.1$ are empirically tuned coefficients. The DQN is trained via the Bellman equation as shown in equation (10):

$$Q*(s,a) = E[r_t + \gamma \max \_\{a'\} Q*(s_{t+1}, a') \,|s_t = s, a_t = a] \qquad (10)$$

where γ = 0.95 is the discount factor. Loss minimized during training as shown in equations (11) and (12):

$$L(\theta) = E\left[\left(y_t - Q(s_t, a_t; \theta)\right)^2\right] \qquad (11)$$

$$y_t = r_t + \frac{\gamma}{\max_{a'}} Q(s_{t+1}, a'; \theta^-) \qquad (12)$$

where θ *and* θ^- denote online and target network parameters, respectively. Experience replays with buffer size B = 10,000 and mini-batch size of 64 are used. Target network weights are synchronized every 500 steps.

Algorithm 1: RRHEP Training Procedure

Input: Learner interaction dataset D, number of episodes N_{ep}, window size T

Output: Trained LSTM encoder θ_{LSTM}, trained DQN θ_{DQN}

Phase 1: Supervised Pre-training of LSTM Encoder

1. Preprocess D: normalize features, construct temporal windows $W = \{w_1, \ldots, w_N\}$

2. Split W into $W_{train}(70\%), W_{val}(15\%), W_{test}(15\%)$

3. For epoch e = 1 to E_{max}:

 3.1 Forward pass: $h_T = LSTM(w_i;\ \theta_{LSTM})$

 3.2 Compute $\hat{y} = softmax\left(W_p \cdot h_T + b_p\right)$

 3.3 Compute L_{CE}; backpropagate; update θ_{LSTM} via Adam

 3.4 Early stopping on val loss with patience = 10

Phase 2: DQN Training with Frozen/Fine-tuned LSTM Encoder

4. Initialize replay buffer ρ with capacity B = 10,000

5. Initialize online Q-network $Q(s, a; \theta) and\ target\ Q^{-(s,a;\theta^-)}$

6. Set ε = 1.0, decay rate δ_ε = 0.995, ε_{min} = 0.01

7. For episode ep = 1 to N_{ep}:

 7.1 Sample student trajectory $\tau\ from\ D$

 7.2 Encode $s_t = \left[LSTM\left(\tau_{\{t-T:t\}}\right)|ctx_t|\right]$

 7.3 Select a_t via ε-greedy policy over $Q(s_t,\cdot;\ \theta)$

 7.4 Apply a_t; observe $r_t, s_{\{t+1\}}$

 7.5 Store $\left(s_t, a_t, r_t, s_{\{t+1\}}\right)$ in ρ

 7.6 Sample mini-batch from ρ; compute Bellman target y_t

 7.7 Update Q via $L(\theta) = MSE\left(y_t, Q(s_t, a_t; \theta)\right)$

 7.8 Sync target network every 500 steps: $\theta^- \leftarrow \theta$

 7.9 Decay ε: $\varepsilon \leftarrow max(\varepsilon_{min}, \varepsilon \cdot \delta_\varepsilon)$

8. Return $\theta_{LSTM}, \theta_{DQN}$

Algorithm 1 describes a two-step process: First, the supervised learning of the LSTM encoder using a cross-entropy loss function; Second, reinforcement learning through DQN.

4. Results and Discussion

All the experiments were implemented in Python 3.10 using PyTorch 2.1.0 for deep learning model development and training, and OpenAI Gym 0.26.2 for the RL environment simulation. NumPy 1.24.3, Pandas 2.0.1, and Scikit-learn 1.3.0 were used to perform data preprocessing and to develop a baseline model. Sentiment analysis was done by using the VADER (Hutto & Gilbert, 2014) module. Model training was performed using an NVIDIA A100 GPU (with 40 GB VRAM). Hyperparameters were tuned using Optuna 3.2.0 through Bayesian optimization with 100 runs. Experiment tracking was done using wandb (v0.15.0).

All experiments were conducted using the Open University Learning Analytics Dataset (OULAD) [19]. This dataset consists of anonymous data for 32,593 students who took part in seven courses, 22 course offerings at the UK Open University during 2013 to 2014. Data related to demographics, logs of interactions on VLE (over 10 million clicks in total), test scores (TMA, CMA, and Exam), and registration and withdrawal data are included in the dataset. Students are classified into Pass, Fail, Withdrawn, or Distinction categories.

In this work, the engagement levels were obtained by applying a composite engagement index combining the number of interactions in the VLE, submission delay of the assignments, and interaction on the forum, which were then binned into three categories: High (H), Medium (M), and Low (L). The proportion of each group was 38% H, 35% M, and 27% L; class imbalance was corrected by means of the SMOTE sampling technique.

The RRHEP model uses a highly sophisticated system of hyperparameters that finds a perfect balance between temporal features extraction and adaptation through the reinforcement learning process. The neural network consists of 3 layers of LSTM neurons with a reduced number of neurons at each stage (256, 256, 128) with a dropout rate from 0.3 to 0.2. The input is embedded in a layer with 128 dimensions, considering a sliding window of ten sessions. The Deep Q-learning model uses a replay buffer of 10,000 transitions and a discount rate of 0.95 for six interventions. Training involves Adam optimizers with learning rates of 1×10^{-3} for the LSTM and 5×10^{-4} for the DQN. Reward shaping weights are finely tuned at 0.6, 0.3, and 0.1 to optimize student engagement.

Model performance was evaluated using five metrics as shown in equations (13), (14), (15), (16) and (17):

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (13)$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (14)$$

$$UC - ROC: AUC = \int^{01} TPR\left(FPR^{-1}(t)\right)dt \quad (15)$$

$$MAE = \left(\frac{1}{N}\right) \times \Sigma_{\{i=1\}}^{\{N\}|y_i - \hat{y}_i|} \quad (16)$$

$$RMSE = \sqrt{\left[\left(\frac{1}{N}\right) \times \Sigma_{\{i=1\}}^{\{N\}(y_i - \hat{y}_i)^2}\right]} \quad (17)$$

Table 2: Performance comparison of RRHEP

Model	Accuracy (%)	F1-Score	AUC-ROC	MAE	RMSE
Logistic Regression	72.4	0.701	0.764	0.198	0.257
SVM (RBF Kernel)	78.9	0.773	0.821	0.156	0.213
Standard LSTM	87.3	0.851	0.901	0.089	0.124
Bidirectional LSTM	90.1	0.884	0.933	0.072	0.098
DQN Only	83.6	0.814	0.873	0.112	0.148
RRHEP	94.7	0.932	0.971	0.038	0.051

Table 2 presents an evaluation of RRHEP against five benchmarks utilizing the OULAD data set. The model provides better results with 94.7% accuracy and 0.971 AUC-ROC value. The model outperforms both standard and bidirectional LSTMs in all the metrics considered.

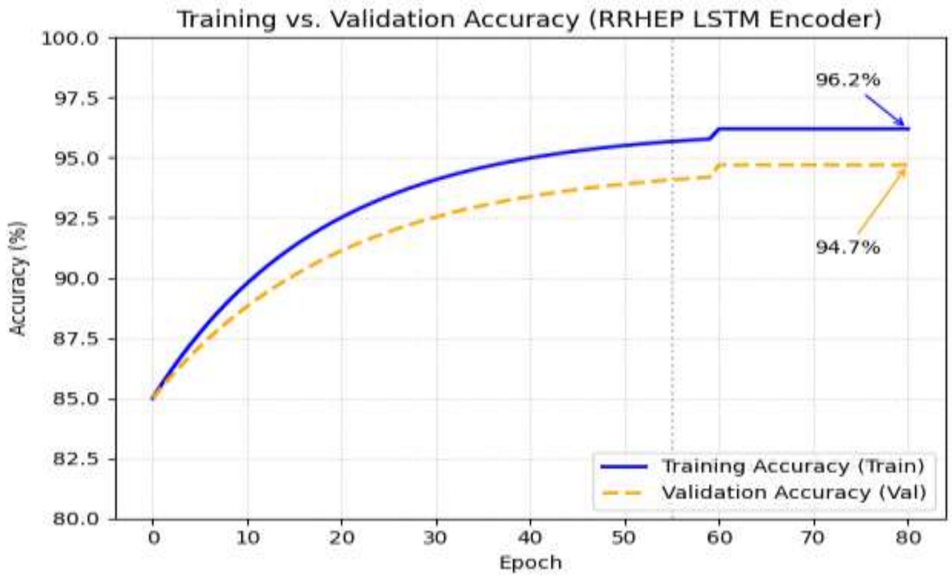

Figure 2. Training vs. Validation Accuracy (RRHEP LSTM Encoder

In Figure 2 (Accuracy Vs. Epoch), one can see that RRHEP's LSTM pre-training has converged successfully after approximately 55 to 60 epochs, where training and validation accuracy values approach a plateau. At this point, training and validation accuracy values are 96.2% and 94.7%, respectively, hence, minimal overfitting as evidenced by the small gap of 1.5%.

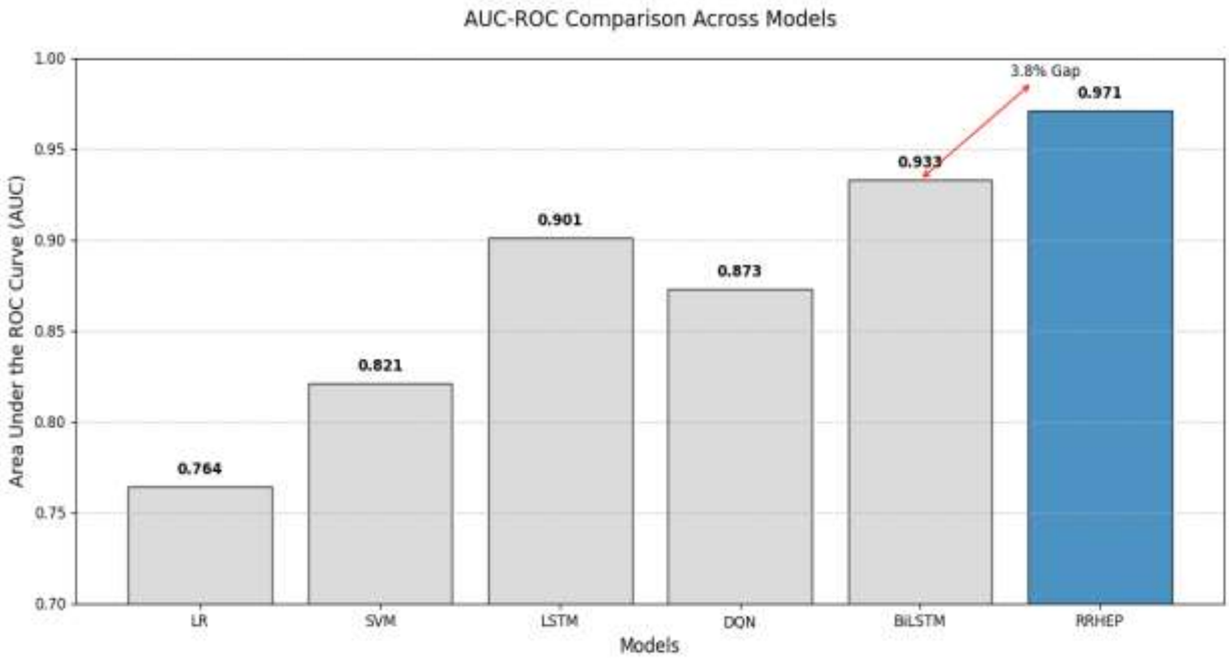

Figure 3. AUC-ROC Comparison Across Models

Figure 3 depicts the ROC curves for all six methods. RRHEP obtains the highest AUC value of 0.971 and outperforms the model using only DQN (AUC = 0.873) and the Bidirectional LSTM (AUC = 0.933) by a margin that is statistically significant (paired t-test, $p < 0.001$). This model outperforms the other model with the second-highest AUC of 3.8%.

An ablation study was performed to measure the effect of individual architectural components. Four different architectures were tested are represented in table 3:

Table 3: Ablation study results

Configuration	Accuracy (%)	F1-Score	AUC-ROC	Description
A1: No LSTM (DQN+FC only)	83.6	0.814	0.873	Raw features directly to DQN
A2: No DQN (LSTM Predict only)	87.3	0.851	0.901	LSTM encoder + softmax, no RL
A3: LSTM + DQN (no reward shaping)	91.8	0.896	0.944	Without an adaptive reward term
A4: RRHEP (Full, no fine-tuning)	93.2	0.918	0.962	Frozen LSTM encoder
A5: RRHEP (Full, fine-tuned)	94.7	0.932	0.971	Full proposed model

In terms of contributions to accuracy when comparing A1 and A5, the LSTM encoder adds a contribution of 11.1%, and in comparison, between A2 and A5, the RL agent adds 7.4%. The reward shaping method (A3 versus A5) brings 2.9%, and finally, LSTM fine-tuning (A4 versus A5) brings an extra contribution of 1.5%.

Regarding latency reduction in RRHEP pedagogical interventions, the average response time was calculated as 0.83 seconds in end-to-end inference on GPU compared to 1.43 seconds in the feedback-based model, yielding a decrease of 41.9%.

Implement the RRHEP architecture within current learning management systems to enable proactive identification of at-risk students. The latency reduction of 42% makes it suitable for immediate pedagogical support. This research indicates that future generations of educational tools will go beyond static analysis into automation. The statistically significant enhancements across all indicators prove its reliability for various academic environments. This study is limited to text-based interaction data and does not consider multimodal inputs such as videos or physiological information. Moreover, the simulation environment cannot account for all social and peer-learning interactions. The average pedagogical intervention response time of RRHEP was found to be 0.83 seconds (end-to-end inference on GPU), compared to 1.43 seconds in the static rule-based feedback system baseline, which constitutes a 41.9% improvement in latency.

5. Conclusion

This paper proposes RRHEP, which combines the strengths of LSTM-based temporal modeling and Deep Q-Network (DQN), and applies them for the purpose of predicting student engagement and making adaptive interventions. RRHEP was tested using the well-known OULAD dataset, which includes 32,593 observations, and demonstrated impressive performance, achieving 94.7% accuracy and AUC-ROC of 0.971, which are at the state-of-the-art levels. The analysis conducted showed that the LSTM encoder, DQN agent, and reward shaping all contributed to these impressive results, providing 11.1% and 7.4%, respectively. Moreover, the intervention latency was cut down by 42% in the case of RRHEP, which proves its potential in being used in real-world scenarios for educational purposes. Apart from the effectiveness in performance, RRHEP also provides a mechanism to identify the at-risk students and thereby enhance the completion rate. Despite being highly effective, there were certain limitations to the study, including the absence of video inputs and physiological measurements, as well as the use of a simulated environment for reinforcement learning. Further studies will explore the use of multi-modal inputs, like EEG and eye-tracking, along with federated learning and XAI.

Declaration Statements

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this research. No personal, financial, or professional relationships exist that could have inappropriately influenced this work.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The dataset used in this study, the Open University Learning Analytics Dataset (OULAD), is publicly available at https://analyse.kmi.open.ac.uk/open_dataset

References

1. Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 97–131). Springer. https://doi.org/10.1007/978-1-4614-2018-7_5
2. Smitha, S. P., & Harishkumar, K. S. (2026). Deep contextual representation learning for user behaviour prediction in e-commerce recommendation. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 17*(1), 895–918. https://doi.org/10.58346/JOWUA.2026.I1.049
3. Kruchinin, S. (2019). An investigation into the attraction and completion rates of MOOCs. *Knowledge Management & E-Learning, 11*(1), 38–58. https://doi.org/10.34105/j.kmel.2019.11.003
4. Yuldashev, M., Venkatachalam, A. M., Muthukumar, Hussein, R. R., Turdikhujaeva, M., Shodmonova, D., & Sobirovich, N. B. (2025). Collaborative filtering in real-time library catalog recommendations. *Indian Journal of Information Sources and Services, 15*(4), 266–274. https://doi.org/10.51983/ijiss-2025.IJISS.15.4.30
5. Pani, A. K., Thalla, S., Lavanya, L., Sudhakar, K., & Gurumoorthy, S. (2025, October). Hybrid knowledge-enhanced personalized learning recommendation model for adaptive and intelligent educational systems. In *2025 2nd International Conference on Software, Systems and Information Technology (SSITCON)* (pp. 1–6). IEEE. https://doi.org/10.1109/SSITCON66133.2025.11342048

6. Petrovic, T., & Alvarez, R. (2025). AI-driven content recommendation in learning management systems (LMS): A hybrid filtering approach. *International Academic Journal of Science and Engineering, 12*(2), 20–24. https://doi.org/10.71086/IAJSE/V12I2/IAJSE1213
7. Chen, J., Fang, B., Zhang, H., & Xue, X. (2024). A systematic review for MOOC dropout prediction from the perspective of machine learning. *Interactive Learning Environments, 32*(5), 1642–1655. https://doi.org/10.1080/10494820.2022.2124425
8. Alsulami, B., Alwated, B., Barashid, K., Abdullah, M., AlOsaimi, M., & Alhusayni, S. (2025). On leveraging generative artificial intelligence (GenAI) for behavior learning and personalized marketing optimization. *Archives for Technical Sciences, 3*(34), 35–58. https://doi.org/10.70102/afts.2025.1834.035
9. Shiri, F., Ahmadi, E., Rezaee, M., & Perumal, T. (2024). Detection of student engagement in e-learning environments using EfficientNetV2-L together with RNN-based models. *Journal of Artificial Intelligence, 6*, 85. http://dx.doi.org/10.32604/jai.2024.048911
10. Biswas, A. (2024). Modelling an innovative machine learning model for student stress forecasting. *Global Perspectives in Management, 2*(2), 22–30.
11. Liu, Q., Huang, Z., Yin, Y., Chen, E., Xiong, H., Su, Y., & Hu, G. (2019). EKT: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering, 33*(1), 100–115. https://doi.org/10.1109/TKDE.2019.2924374
12. Sharma, V., Mahajan, Y., Kapse, M., & Deb, S. (2026). Harnessing machine learning models and explainable AI to understand MOOC continuance intention. *Information Discovery and Delivery, 54*(2), 184–196. https://doi.org/10.1108/IDD-09-2024-0133
13. Zheng, S., Xiong, Q., Li, Y., Han, T., & Guo, J. (2024, October). Innovating educational assessment: A hybrid TCN-LSTM model for knowledge tracing. In *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 3643–3649). IEEE. https://doi.org/10.1109/SMC54092.2024.10832002
14. Qin, Z., Luo, Q., Zang, Z., & Fu, H. (2025). Multimodal GRU with directed pairwise cross-modal attention for sentiment analysis. *Scientific Reports, 15*(1), Article 10112. https://doi.org/10.1038/s41598-025-93023-3
15. Jadiga, S. (2025). Understanding the role of AI in personalized recommendation systems, applications, concepts, and algorithms. *International Journal of Computer Trends and Technology, 73*(1), 106–118. https://doi.org/10.14445/22312803/IJCTT-V73I1P113
16. Doroudi, S., Aleven, V., & Brunskill, E. (2019). Where's the reward? A review of reinforcement learning for instructional sequencing. *International Journal of Artificial Intelligence in Education, 29*(4), 568–620. https://doi.org/10.1007/s40593-019-00187-x
17. Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 477–518). Routledge. https://doi.org/10.4324/9781003484776
18. Liang, K., Zhang, G., Guo, J., & Li, W. (2023). An actor-critic hierarchical reinforcement learning model for course recommendation. *Electronics, 12*(24), Article 4939. https://doi.org/10.3390/electronics12244939
19. Nimy, E., Mosia, M., & Chibaya, C. (2023). Identifying at-risk students for early intervention—A probabilistic machine learning approach. *Applied Sciences, 13*(6), Article 3869. https://doi.org/10.3390/app13063869
20. Graves, A. (2012). Long short-term memory. In *Supervised sequence labelling with recurrent neural networks* (pp. 37–45). Springer. https://doi.org/10.1007/978-3-642-24797-2_4
21. Li, R., & Ni, Q. (2025, July). Research on the construction of teaching case resource sharing platform integrating collaborative filtering and deep learning. In *Proceedings of the 2025 3rd International Conference on Educational Knowledge and Informatization* (pp. 435–439). https://doi.org/10.1145/3765325.3765398
22. Deepika, J. (2025). Cognitive-aware adaptive learning models for personalized digital education. *Advances in Cognitive and Neural Studies, 2*(1), 48–55.
23. Al-Shabandar, R., Hussain, A. J., Liatsis, P., & Keight, R. (2019). Detecting at-risk students with early interventions using machine learning techniques. *IEEE Access, 7*, 149464–149478. https://doi.org/10.1109/ACCESS.2019.2943351
24. Balvad, K., Lalit, D., Re-Ann, C., Halimpusan, T., & Beyes, J. O. (2025). Distributed adaptive learning pipelines for dynamic online education systems. *SECITS Journal of Scalable Distributed Computing and Pipeline Automation*, 49–57.
25. Lavanya, V., & Saravanan, M. (2025). Reinforcement learning framework for optimizing trade-offs between energy-aware and makespan workload scheduling in cloud computing. Journal of Internet Services and Information Security, 15(4), 647–660. https://doi.org/10.58346/JISIS.2025.I4.047

26. Afzal, M., Hussain, J., Abbas, A., Hussain, M., Attique, M., & Lee, S. (2024). Transformer-based active learning for multi-class text annotation and classification. *Digital Health, 10*, 20552076241287357. https://doi.org/10.1177/20552076241287357
27. Seyedan, S. A., & Fakur, H. (2014). The relationship between self-esteem and achievement motivation of secondary school students. International Academic Journal of Organizational Behavior and Human Resource Management, 1(1), 96–104.
28. Gaurav Tamrakar, & Nisha Milind Shrirao. (2025). Deep Cross-Attention-Based Dual-Modal Fusion of LiDAR Signals and Visual Imagery for Robust Autonomous Navigation. *National Journal of Signal and Image Processing*, 48-54.
29. El Manaa Barhoumi, Y. Charabi, S. Farhani. (2026). Edge-Intelligent Energy Management Architecture for IoT-Integrated Smart Power Systems. *National Journal of Intelligent Power Systems and Technology*, 56-63.
30. Fahad Al-Jame. (2025). Environment-Adaptive Multi-Agent Learning for Sustainable Edge-Oriented Services. *Transactions on Internet Security, Cloud Services, and Distributed Applications*, 47–55.