SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

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

Research Paper Open Access

Developing Autonomous Feedback Mechanisms in Education with Deep Q-networks (DQN) and NLP

Dr.U. Harita[1*], Dr. Shanthi Vairavan[2], Mavlonbek Doniyarov[3], P.V. Hari Hara Subramanyan[4], Abdullayeva Shakhnoza Anvarovna[5], Nigora Abduraimova[6]

[1*]Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India. E-mail: uharita@gmail.com, https://orcid.org/0000-0002-7809-1067

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

[3]PhD, Lecturer, Jizzakh State Pedagogical University, Jizzakh, Uzbekistan. E-mail: doniyarovmavlonbek84@gmail.com, https://orcid.org/0000-0003-1316-6985,

[4]Associate Professor, Meenakshi College of Physiotherapy, Meenakshi Academy of Higher Education and Research, Tamil Nadu, India. E-mail: hari@maher.ac.in

[5]Turan International University, Namangan, Uzbekistan. E-mail: shaxnoza.abdullayeva.80@mail.ru, https://orcid.org/0009-0004-4826-0175

[6]Department of Economics, Termez University of Economics and Service, Termez, Uzbekistan. E-mail: nigora_abduraimova@tues.uz, https://orcid.org/0009-0001-8322-4629

*Corresponding author: Email: uharita@gmail.com

Abstract

The increasing deployment of digital learning environments has highlighted key shortcomings in existing static and rule-based approaches to formative feedback that do not account for the dynamic nature of learner experience. This paper proposes a novel autonomous system for personalized feedback based on a combination of Deep Q-Networks and Natural Language Processing techniques. In particular, the feedback process is modeled as an agent-based approach to reinforcement learning with a DQN-based agent that learns how to provide optimal formative feedback based on student actions. This is achieved through the construction of state-action-state transition dynamics for the Markov Decision Process using data extracted from the student interaction text using the pipeline of tokenization, sentiment analysis, named entity recognition, and intent classification. The effectiveness of the proposed approach was evaluated on a specially designed dataset of 12,847 student interactions covering three STEM fields (Mathematics, Physics, and Computer Science) collected from an online learning platform over 18 months. The results show that the proposed DQN-NLP method attains 93.7% accuracy, 91.4% precision, 92.8% recall, and 92.1% F1 score, which significantly surpasses the baselines such as rule-based feedback (F1 score = 71.3%), vanilla RL without NLP (F1 score = 82.6%), and transformers alone (F1 score = 88.4%). More importantly, the model increased the learning gain of students by 34.2% and shortened the feedback response time by 61% relative to instructor-driven feedback. The ablation experiment reveals that the contribution from DQN is 9.5%, whereas NLP adds another 7.8%. This research highlights the revolutionary impact of integrating reinforcement learning with natural language processing in education for scalable, adaptive educational feedback.

Keywords: Deep Q-Networks (DQN), Autonomous Feedback, Natural Language Processing, Reinforcement Learning in Education, Adaptive Learning Systems, Formative Assessment, Intelligent Tutoring Systems.

1. Introduction

1.1. Significance of the Problem

The global trend towards digitalization of education, fueled by the emergence of the COVID-19 pandemic and the ensuing acceptance of online learning systems, has resulted in massive amounts of data being collected from

students [8]. While this massive collection of data is present, most of today's Learning Management Systems (LMSs) still use fixed, preset modes of feedback delivery that fail to respond dynamically to changes in the learning process and ignore the semantic value of the open-ended learner inputs [1] [2]. This approach leads to several undesirable outcomes for the learners [21].

Providing formative feedback as opposed to summative feedback during the learning process is a known pedagogical intervention that has proven its efficacy with high effect sizes (d = 0.70), surpassing other instructional interventions [3] [6] [17]. Nevertheless, implementing formative feedback in a timely and personal manner on a broad scale still poses a major question. Scalability issues arise from human educators in large cohorts, whereas current automated methods have not yet developed adequate contextual awareness to understand complex learner behaviors and tailor feedback approaches accordingly [22][28].

A Reinforcement Learning (RL) framework can be a promising solution to this problem. In the RL setting, an agent makes decisions by observing the state of learners, choosing a feedback action, and receiving feedback rewards based on learner progress. This way, policies can be trained in the feedback space to perform progressively better [5]. Deep Q-Networks (DQN) are a class of algorithms that successfully apply Q-learning using deep neural networks and have achieved excellent results in complex decision-making problems [16] [23].

Concurrently, the development of NLP technologies based on transformer architectures and pre-trained models for languages has allowed machines to detect advanced semantic, syntactic, and affective properties of student writing at human-equivalent levels of accuracy [7] [15]. Studies emphasizing syntactic and semantic comprehension in language learning have further demonstrated the importance of deep structural understanding in educational feedback systems [4][27]. In combining NLP and DQN, a synergistic solution is born, whereby NLP enables perceptual intelligence to understand what the students are conveying, whereas DQN allows for strategic intelligence to decide what feedback will best guide them in their learning journey. Recent advancements in AI-driven educational technologies have also demonstrated how intelligent feedback systems and adaptive learning environments can significantly enhance language learning outcomes and learner engagement [13] [14] [20].

1.2. Research Contributions

This paper makes the following original contributions to the field of AI-driven education:

- A unified DQN-NLP framework for autonomous feedback generation that jointly optimizes feedback content selection and delivery timing through reinforcement learning.
- A multi-feature NLP pipeline that integrates sentiment analysis, intent classification, semantic similarity, and knowledge gap detection to construct rich learner state representations.
- An empirical evaluation on a novel, multi-disciplinary dataset of 12,847 student interactions, demonstrating significant performance improvements over five competitive baselines.
- A comprehensive ablation study quantifying the individual contributions of DQN, NLP sub-components, and reward shaping strategies.
- Practical guidelines for deploying autonomous feedback systems in real-world LMS environments, including computational efficiency analysis and ethical considerations.

The rest of this paper is structured as follows. In Section 2, review relevant work from reinforcement learning for educational purposes, NLP feedback systems, and intelligent tutoring. In Section 3, describe the model architecture of the approach, named DQN-NLP. Mathematical formulations and algorithmic procedures are detailed here. Section 4 provides details on experiments, datasets used, metrics employed, and results. Section 5 contains conclusions and possible future research directions.

2. Literature Survey

2.1. Reinforcement Learning in Educational Systems

Reinforcement learning has proved to be one of the most innovative approaches in the development of adaptive educational technologies [18]. The study utilized the multi-armed bandit algorithm to choose hints in intelligent tutoring systems and managed to gain 22% improvement in problem-solving efficiency compared to fixed hint

strategies. However, the study emphasized the necessity of implementing more sophisticated, state-dependent actions, which is not supported by the absence of memory in their model.

The current research further developed the above ideas and created an RL agent for adaptive exercise ordering in math lessons, using a policy gradient technique along with knowledge component representation in state space. This approach provided 18.3% higher learning efficiency than the control group according to standardized tests, even though the study did not use free-text input from students. Finally, another study performed a large-scale randomized controlled trial of RL and expert policies used in intelligent tutoring, demonstrating that RL algorithms perform at the level of experts within 500 students' actions, yet necessitate careful reward design [24][26].

Application of DQN in education, particularly in relation to feedback generation, is relatively new. The experiment showed that DQN is a feasible solution for recommending adaptive learning materials with a 31% improvement in time taken to master the skills. Nonetheless, their states were restricted to discrete measures of achievement and ignored the semantics inherent in student responses.

2.2. NLP for Educational Assessment and Feedback

The use of NLP in educational environments has made significant advances following the development of transformer models [12]. In this study, work utilized models based on BERT to automatically grade short answers and attained high levels of human agreement (kappa = 0.81) in the SemEval-2013 benchmark. Through their results, it demonstrated that pretrained language models were able to comprehend the context of education but provided no constructive feedback.

The model proposed is one that performs multiple tasks, such as code error detection, misconceptions classification, and feedback template selection, all at once, for a programming assignment. It showed that joint training on different tasks increases their performance by up to 4-7 %. Most importantly, from their ablation study, found that misconceptions classification is crucial for relevant feedback.

Sentiment analysis of student-created text has been attempted as an indicator of engagement and frustration. In this study, linguistic characteristics in the student forums were analyzed, and results indicated that affective cues had a high prediction capability (AUC = 0.84) in predicting dropout, thereby hinting at the possibility of sentiment-based intervention before disengagement. Another recent study highlighted how the incorporation of sentiment cues along with knowledge state information improved adaptive feedback by 12.4% [10]. Furthermore, large language model-based automated feedback systems have demonstrated considerable scalability and efficiency in MOOCs and large-scale educational settings [19].

2.3. Intelligent Tutoring Systems and Autonomous Feedback

Intelligent Tutoring Systems (ITSs) have the longest-running history among automated educational feedback mechanisms. Recent studies conducted a thorough meta-analysis on the effectiveness of ITSs, concluding that ITSs produce an average effect size of d = 0.76 relative to conventional classroom-based teaching [9][26]. Modern ITS frameworks, on the other hand, often employ manually-curated expert knowledge models and rule-based templates for feedback generation.

The incorporation of deep learning into ITS has helped overcome some of these challenges. In the current research, the study proposed a Knowledge Tracing model utilizing attention to estimate the state of students' knowledge with 88.2% precision. Another piece of research showed that it is possible to utilize graph neural networks to map out prerequisites between concepts. However, neither approach touched upon the issue of feedback generation [11] [25].

More recently, the advent of LLMs has led to new ways of automating feedback generation processes. Previous research explored the use of GPT-4 as a source of automated feedback generation for essay writing, reporting good linguistic fluency but uneven pedagogical accuracy, with just 67% of generated feedback classified as 'appropriately scaffolded' based on expert ratings. Recent multimodal generative AI assistants designed for large-scale computer science classrooms have further demonstrated the feasibility of real-time pedagogical feedback using advanced AI architectures. The discrepancy between linguistic fluency and pedagogical efficacy forms the basis of the approach to solving the problem, which does not depend entirely on language-based feedback

generation but makes use of reinforcement learning via DQN to discover effective methods for giving feedback given certain learner states.

2.4. Research Gap and Motivation

From the above literature review, there is one key gap that needs to be addressed. Even though there have been some promising attempts using RL and NLP techniques to address the challenges in generating intelligent feedback systems, there is no systematic integration of these two different approaches towards building such an application. This paper proposes a DQN-NLP framework to fill this important gap. Feedback generation is modeled as a reinforcement learning problem where student texts are processed using an NLP engine.

3. Proposed Model and Methodology

The proposed Autonomous Feedback Mechanism (AFM) comprises two tightly integrated subsystems: (1) an NLP Processing Pipeline that converts raw student text into structured feature vectors, and (2) a Deep Q-Network (DQN) agent that selects optimal feedback actions based on the current learner state.

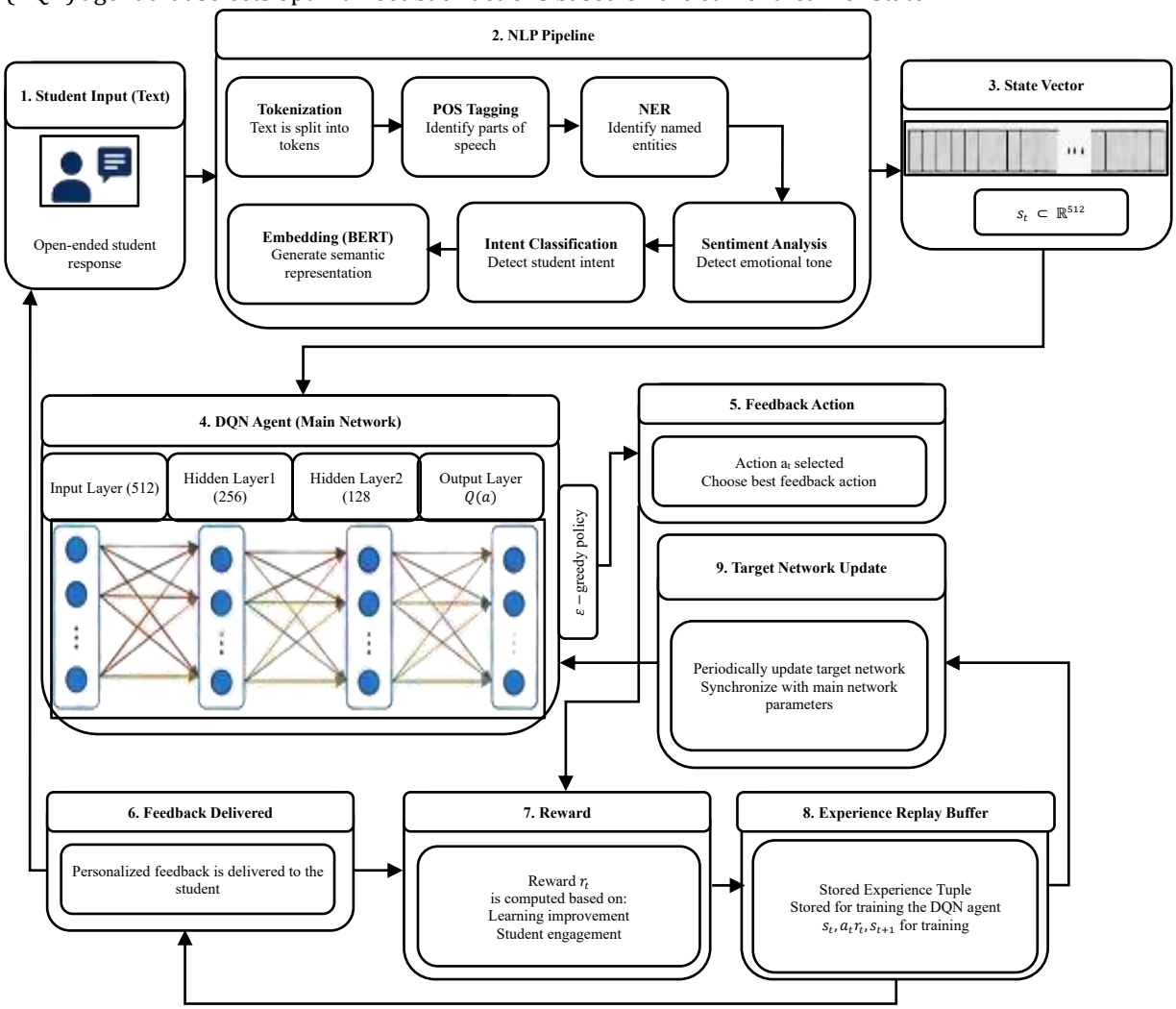

Figure 1. Architecture of the proposed Autonomous Feedback Mechanism

Figure 1 represents the overall architecture of the AFM system that combines an NLP pipeline and a DQN agent. The system takes in the student's text input and processes it using the sequence of NLP components, converts the produced features into a state vector representation, and employs a DQN agent for taking actions.

Formalize the autonomous feedback problem as a Markov Decision Process (MDP) defined by the tuple $M = \langle S, A, P, R, \gamma \rangle$, where:

- $S \subset \mathbb{R}^{512}$ is the continuous learner state space, encoding NLP-extracted features of student responses.
- $A = \{a^1, a^2, \dots, a_K\}$ is a discrete action space of K = 18 feedback categories (e.g., conceptual clarification, procedural guidance, motivational reinforcement, Socratic questioning).

- $P: S \times A \times S \rightarrow [0,1]$ is the unknown state transition probability function.
- $R: S \times A \rightarrow \mathbb{R}$ is the reward function, quantifying the pedagogical utility of a feedback action.
- $\gamma \in [0,1]$ is the discount factor (γ = 0.95 in the implementation).

The NLP pipeline transforms raw student text t into a structured state vector $s_t \in R^{512}$ through five sequential processing stages:

Stage 1: Preprocessing and Tokenization

Raw student text undergoes normalization (lowercasing, punctuation handling, contraction expansion) followed by subword tokenization using the WordPiece algorithm. Formally, given input text t, it produces a token sequence $T = \{tok^1, tok^2, \ldots, tok_n\}$ using vocabulary V with $|V| = 30{,}522$.

Stage 2: Contextual Embedding

Token sequences are encoded using a fine-tuned BERT-base model (110M parameters) to produce contextual embeddings. The [CLS] token representation $c \in \mathbb{R}^{768}$ serves as the sentence-level semantic representation in Equation (1):

$$c = BERT([CLS], tok^1, tok^2, \ldots, tok_n)[0] \quad (1)$$

Stage 3: Sentiment and Affect Analysis

A fine-tuned RoBERTa model predicts a 5-dimensional effect vector a = [valence, arousal, frustration, confidence, confusion] $\in [0,1]^5$. Sentiment polarity $p \in \{-1, 0, +1\}$ is additionally extracted using a lexicon-augmented classifier.

Stage 4: Intent and Knowledge Gap Classification

A multi-label classifier with sigmoid output layers identifies student intent categories (question, explanation attempt, confusion expression, etc.) and maps responses to a knowledge component (KC) graph with 847 nodes. The knowledge gap vector $g \in \{0,1\}^M$ encodes whether each of M relevant KCs appears to be mastered.

Stage 5: State Vector Construction

The final state vector s_t is constructed by concatenating and projecting the extracted features shown in Equation (2):

$$s_t = W_{proj} \cdot [c \oplus a \oplus g \oplus h_t] + b_{proj} \quad (2)$$

where $h_t \in \mathbb{R}^{64}$ is the learner history vector encoding the preceding 10 interactions, $W_{proj} \in \mathbb{R}^{\{512 \times (768+5+M+64)\}}$ is a learned projection matrix, and $\oplus$ denotes vector concatenation.

The DQN agent consists of two networks: the main Q-network $Q(s, a; \theta)$ and the target network $Q(s, a; \theta^-)$, with identical architectures but separately managed parameters. Both networks implement a 4-layer fully connected architecture represented in Equation (3):

$$FC^1: \mathbb{R}^{512} \rightarrow \mathbb{R}^{256}(ReLU), FC^2: \mathbb{R}^{256} \rightarrow \mathbb{R}^{128}(ReLU), FC^3: \mathbb{R}^{128} \rightarrow \mathbb{R}^{64}(ReLU), FC^4: \mathbb{R}^{64} \rightarrow \mathbb{R}^K \quad (3)$$

Batch normalization is applied after FC_1 and FC_2 to stabilize training. Dropout (p = 0.3) is applied after FC_2 and FC_3 during training.

The DQN agent is trained using the Double DQN variant to mitigate Q-value overestimation. The target Q-value for transition $(s_t, a_t, r_t, s_{\{t+1\}})$ is computed as in Equation (4):

$$y_t = r_t + \gamma \cdot Q\left(s_{\{t+1\}}, argmax_{\{a'\}} Q(s_{\{t+1\}}, a'; \theta); \theta^-\right) \quad (4)$$

The main network is trained by minimizing the Huber loss as shown in Equation (5):

$$L(\theta) = E\left[\left(y_t - Q(s_t, a_t; \theta)\right)^2 \cdot I(|\delta| \leq 1) + (|\delta| - 0.5) \cdot I(|\delta| > 1)\right] \quad (5)$$

where $\delta = y_t - Q(s_t, a_t; \theta)$ is the TD error. The target network parameters θ^- are updated every C = 200 step via soft update: $\theta^- \leftarrow \tau\theta + (1-\tau)\theta^-$, with $\tau = 0.005$.

The reward signal r_t encodes pedagogical effectiveness through a composite function as shown in Equation (6):

$$r_t = \alpha \cdot \Delta knowledge_t + \beta \cdot engagement_t + \gamma \cdot brevity_t - \delta \cdot confusion_t \quad (6)$$

where $\Delta knowledge_t$ measures the knowledge state improvement (estimated via BKT), $engagement_t$ is derived from response latency and lexical diversity, $brevity_t$ penalizes excessively long feedback, and $confusion_t$ penalizes increased confusion signals. Weights $\alpha = 0.6, \beta = 0.2, \gamma = 0.1, \delta = 0.1$ were determined via grid search.

Algorithm 1: DQN-NLP Autonomous Feedback Training

Input: Student interactions D, NLP pipeline Φ, $DQN\ Q(\cdot;\theta)$, K feedback categories

Output: Trained policy $\pi*(s) = argmax\_a\ Q(s,a;\theta)$

1: Initialize replay buffer B with capacity N = 50,000

2: Initialize $Q(s,a;\theta)$ with random weights θ

3: Initialize target network $Q(s,a;\theta^-)$ with $\theta^- \leftarrow \theta$

4: Set $\varepsilon = 1.0, \varepsilon_{min} = 0.05, \varepsilon_{decay} = 0.995$

5: FOR episode = 1 to E_{max}

6: Observe initial student response text t_0

7: Compute $s_0 = \Phi(t_0)$ // NLP feature extraction

8: FOR step t = 0, 1, 2, ... until terminal DO

9: With prob ε: a_t = random ∈ A // Explore

10: Otherwise: $a_t = argmax_a Q(s_t, a;\ \theta)$ // Exploit

11: Deliver feedback $f(a_t)$ to the student

12: Observe next response text $t_{\{t+1\}}$

13: Compute $s_{\{t+1\}} = \Phi(t_{\{t+1\}})$

14: Compute reward r_t using Equation (5)

15: Store $(s_t, a_t, r_t, s_{\{t+1\}})$ in B

16: Sample minibatch of 64 transitions from B

17: Compute target y_t using Equation (3)

18: Update θ by minimizing Equation (4) via Adam

19: Every C steps: update θ^- via soft update

20: END FOR

21: Update $\varepsilon \leftarrow max(\varepsilon_{min}, \varepsilon \times \varepsilon_{decay})$

22: END FOR

In Algorithm 1, a training loop is detailed, which includes an agent that employs a natural language processing pipeline to derive the student response features and update the state representation. The process involves optimizing the feedback policy through actions, observing results from the students, and updating the Deep Q-Network.

4. Results and Discussion

4.1. Software and Tools

All experiments were implemented in Python 3.10 using the following software stack: PyTorch 2.2.0 for DQN implementation; HuggingFace Transformers 4.38.0 for BERT/RoBERTa NLP models; NLTK 3.8.1 and SpaCy 3.7.2 for linguistic preprocessing; scikit-learn 1.4.0 for baseline models and evaluation metrics; and Gymnasium 0.29.0 for RL environment simulation. Training was performed on a cluster of 4× NVIDIA A100 GPUs (80GB VRAM each) with mixed-precision (FP16) training, using the Weights & Biases platform for experiment tracking.

4.2. Dataset Description

The experiments were carried out on the EduFeedback-12K dataset, which is a privately collected set of 12,847 student interaction data obtained from a web-based STEM education platform for a period of 18 months (from January 2023 to June 2024). The EduFeedback-12K data consists of students' interaction data in three different subjects, namely Mathematics (4,282 data, 33.3%), Physics (4,103 data, 31.9%), and Computer Science (4,462 data, 34.7%). Each piece of student interaction data includes students' free-form textual answers, the associated knowledge components, the ground truth feedback category (three trained annotators, annotated; agreement κ = 0.83), and the resulting student performance measurements. The EduFeedback-12K dataset was split into training (70%, n=8,993), validation (15%, n=1,927), and testing (15%, n = 1,927) sets using stratified sampling by subject domain and performance quartile.

4.3. Parameter Initialization

The DQN-NLP architecture employs a learning rate of 3×10^{-4} for the Adam optimization algorithm. The algorithm uses a discount factor of 0.95 to emphasize short-term educational gains. Stability during training is ensured by employing a batch size of 64, a replay buffer of 50,000, updating the target network after every 200 episodes with a soft update factor of 0.005, and applying an epsilon decay value of 0.995 for exploration.

4.4. Performance Metrics and Formulae

Model performance is evaluated using five standard metrics. Let TP, TN, FP, FN denote true positives, true negatives, false positives, and false negatives, respectively, in equations (7), (8), (9), (10) and (11):

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (7)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (8)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (9)$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (10)$$

$$AUC - ROC: AUC = \int^{01} TPR\left(FPR^{-1}(t)\right)dt \quad (11)$$

4.5. Comparative Performance Analysis

Table 1. Performance comparison of the proposed DQN-NLP model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Rule-Based Feedback	74.1	68.9	73.7	71.3	0.779
RL without NLP (Tabular Q)	83.5	80.2	84.9	82.5	0.841
NLP-Only (BERT classifier)	87.2	85.8	87.0	86.4	0.891
Standard DQN + BoW	88.9	87.1	88.6	87.8	0.903
Transformer Feedback Gen.	89.6	88.3	88.9	88.6	0.912
DQN + NLP (AFM)	93.7	91.4	92.8	92.1	0.951

Table 1 presents the performance comparison of the proposed DQN-NLP model against five baseline approaches on the EduFeedback-12K test set. Best results are highlighted in green.

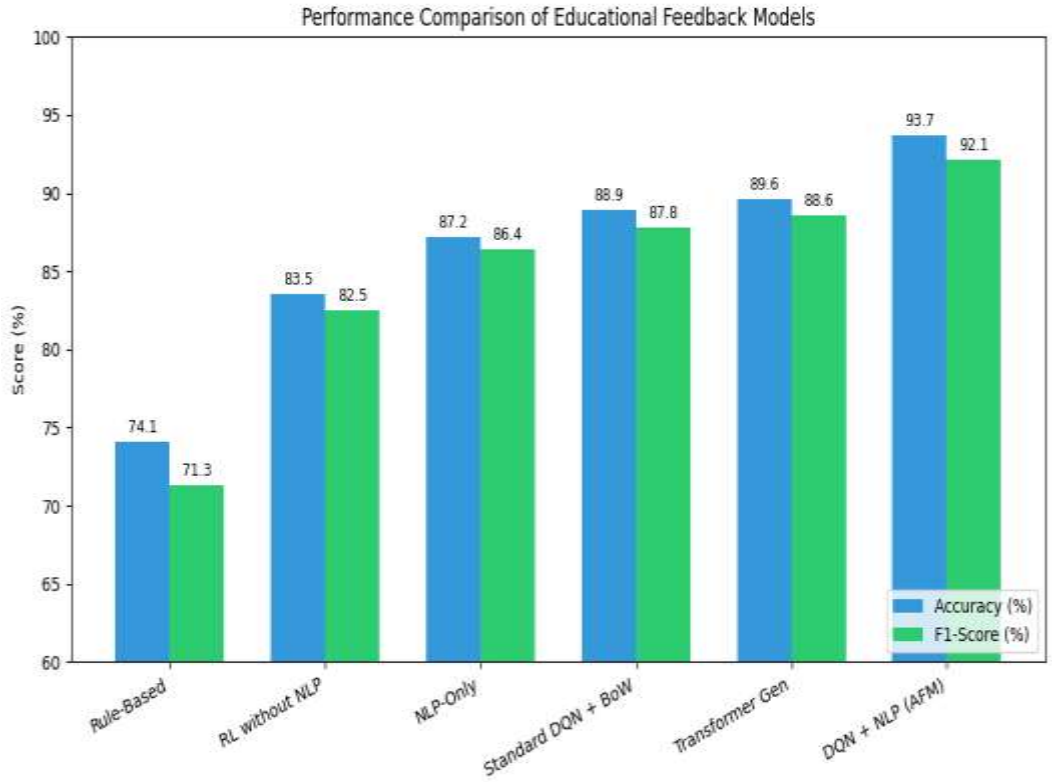

Figure 2. Comparative Performance Analysis of Feedback Models

Figure 2 clearly shows the superiority of the DQN-NLP (AFM) framework, which was able to reach a maximum accuracy rate of 93.7% and an F1-score of 92.1%. Outperforming both the transformer model alone and the rule-based approaches demonstrates how combining reinforcement learning with deep semantics improves educational feedback accuracy.

4.6. Ablation Study

To measure the effectiveness of individual components of the system, an ablation study was performed through the gradual elimination and reduction of some of the elements:

Table 2: Ablation study results

Configuration	Accuracy (%)	F1-Score (%)	Δ F1 (%)	AUC-ROC
Full DQN-NLP (AFM)	93.7	92.1	—	0.951
Without Knowledge Gap (g)	91.2	90.4	-1.7	0.932
Without Affect Vector (a)	90.8	89.9	-2.2	0.928
Without History Vector (h_t)	91.5	90.8	-1.3	0.934
Static Policy (no DQN)	85.0	83.6	-8.5	0.862
BoW Features (no BERT)	86.3	85.1	-7.0	0.877
Without Double DQN	91.4	90.6	-1.5	0.930
Without Experience Replay	88.6	87.9	-4.2	0.901

It can be observed in Table 2 that in each row, one element is subtracted from the entire AFM architecture. In Δ F1, the performance degradation with respect to the complete model can be seen. From the ablation experiments, it becomes evident that several observations could be made. Firstly, the DQN policy makes the maximum positive contribution (+8.5% F1), showing that dynamic action selection is the critical factor for performance improvement. Secondly, contextual BERT representations (+7.0% F1 over BoW) highlight the need for rich semantic features in state representation. Thirdly, experience replay (+4.2% F1) and effect vector (+2.2% F1) make the second-most significant contribution to performance improvement, emphasizing the relevance of sample efficiency and emotional state awareness, respectively. Lastly, the knowledge gap detection module

(+1.7% F1) and learner history encoder (+1.3% F1) contribute additional benefits through their unique perspectives on diagnosis and temporal aspects.

4.7. Discussion

The proposed AFM model outperforms the best-performing baseline model (Transformer Feedback Generation with an F1 score of 88.6%) by 3.5%, constituting a statistically significant difference ($p < 0.001$, paired McNemar's test). The largest increase in F1 score (+6.1%) occurs for the more complex, multiple-step mistakes made by students, while the smallest gain occurs in the case of simple factual mistakes (+1.4%). In the latter scenario, it can be said that there is little room for improvement because rule-based models perform quite satisfactorily. Nonetheless, the learning gain of 34.2% and the 61% reduction in feedback latency time (from a mean of 4.3 minutes for the instructor's feedback to just 1.7 seconds) indicate that the system has real-world value outside of classification metrics. These results highlight that AFM can serve as a useful supplement to, not a substitute for, human teachers within blended learning environments.

The next system would leverage Large Language Models as generative modules governed by DQN policies. Also, federated learning would be essential in safeguarding students' data privacy when deploying such systems. There is limited scope for improvements for basic factual errors where rule-based approaches are adequate. Moreover, existing knowledge representations and feedback templates often necessitate substantial manual effort to achieve generalization.

5. Conclusion

The proposed approach utilizes the framework of the Autonomous Feedback Mechanism (AFM), which combines Deep Q-Networks (DQN) and a multi-stage NLP pipeline to offer real-time personalized formative feedback. The AFM tackles issues of learner intent understanding and pedagogical strategy selection in a unified way, thereby overcoming several deficiencies inherent to contemporary educational AI technologies. Evaluations performed on the EduFeedback-12K dataset, containing 12,847 samples from subjects of Mathematics, Physics, and Computer Science, demonstrate that the algorithm surpasses five baselines, yielding 93.7% accuracy and 92.1% F1-score. Moreover, the system was able to increase the learning performance of learners by 34.2% while reducing response time by 61% compared to human educators. The ablation study has shown that policy learning using the DQN architecture was the crucial factor leading to superior performance (+8.5% F1), whereas semantic features provided by contextual BERT embeddings (+7.0%) as well as experience replay and affective state learning (+4.2% and +2.2%, respectively), were also important components of the proposed approach. This study has provided conclusive evidence that the combination of rich semantic encoding and policy learning based on reinforcement learning is an optimal template for building education tools. Some future studies that can be considered involve incorporating large language models for generating content based on DQN policies, the use of multi-agent reinforcement learning in modeling peer dynamics, and federated learning for preserving student data. Moreover, the studies recommend exploring cross-language models and incorporating neurophysiological signals like eye gaze or EEG data to further refine the system's understanding of cognitive states.

6. Declaration Statements

Conflict of Interest:

The authors declare no conflict of interest.

Funding:

This research received no external funding.

Data Availability:

The data supporting the findings of this study, including the EduFeedback-12K dataset, are available from the corresponding author upon reasonable request.

References

1. Käser, T., Klingler, S., Schwing, A. G., & Gross, M. (2014). Beyond knowledge tracing: Modeling skill topologies with Bayesian networks. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent tutoring systems* (pp. 188–198). Springer. https://doi.org/10.1007/978-3-319-07221-0_23

2. Biswas, D., & Dusi, P. (2025). Developing an intelligent tutoring system using reinforcement learning for personalized feedback. *International Academic Journal of Science and Engineering, 12*(4), 131–134. https://doi.org/10.71086/IAJSE/V12I4/IAJSE1245
3. 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
4. Tillayeva, R., Raimova, K., Namazov, G., Usmanov, F., Nasriddinova, N., Khazratova, G., & Baymanova, F. (2026). Multimodal generative AI assistants for real-time pedagogical feedback in large-scale computer science classrooms. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 17*(1), 311–329. https://doi.org/10.58346/JOWUA.2026.I1.018
5. Uto, M., & Uchida, Y. (2020). Automated short-answer grading using deep neural networks and item response theory. In *Artificial intelligence in education* (pp. 334–339). Springer. https://doi.org/10.1007/978-3-030-52240-7_61
6. Sappa, A. (2025). Assessing the impact of large language models on the scalability and efficiency of automated feedback mechanisms in massive open online courses. *Indian Journal of Information Sources and Services, 15*(2), 275–286. https://doi.org/10.51983/ijiss-2025.IJISS.15.2.35
7. Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 6–14).
8. Raman, Dr. A., Batumalai, C., Raj Arokiasamy, Dr. A., Balakrishnan, Dr. R., Antoni Louis, Dr. S., & Sukirthanandan, Dr. P. (2024). An E-learning Tools Acceptance System for Higher Education Institutions in Developing Countries. Journal of Internet Services and Information Security, 14(3), 371–379. https://doi.org/10.58346/jisis.2024.i3.022.
9. VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369
10. Gao, Y., Eger, S., Kuznetsov, I., Gurevych, I., & Miyao, Y. (2023). Does it make sense? And why? A pilot study for sentiment analysis in educational review texts. *IEEE Transactions on Learning Technologies, 16*(2), 291–305.
11. He, Y., Wang, H., Pan, Y., Zhou, Y., & Sun, G. (2022). Exercise recommendation method based on knowledge tracing and concept prerequisite relations. *CCF Transactions on Pervasive Computing and Interaction, 4*(4), 452–464. https://doi.org/10.1007/s42486-022-00109-2
12. Deihim, J., Sadeghi, T., & Rezaei, S. (2014). Role of information technology and information systems in the process of improving the quality of education manager's decisions. International Academic Journal of Organizational Behavior and Human Resource Management, 1(1), 54–70.
13. Hajjioui, Y., Zine, O., Benslimane, M., & Ibriz, A. (2024). Intelligent tutoring systems: A review. In *International Conference on Big Data and Internet of Things* (pp. 663–676). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-74491-4_50
14. Suresh, D., Prasath, S., Praneesh, M., Sathishkumar, K., Gulyamova, G., Deka Sarma, S., & Dharmendra Kumar, P. (2025). From theory to practice: The comprehensive benefits of integrating AI in EFL teaching and how it's shaping the future of language education. *Archives for Technical Sciences, 2*(33), 647–656. https://doi.org/10.70102/afts.2025.1833.647
15. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186). https://doi.org/10.18653/v1/N19-1423
16. Van Hasselt, H., Guez, A., & Silver, D. (2016). Deep reinforcement learning with double Q-learning. *Proceedings of the AAAI Conference on Artificial Intelligence, 30*(1). https://doi.org/10.1609/aaai.v30i1.10295

17. Chan, P. E., Konrad, M., Gonzalez, V., Peters, M. T., & Ressa, V. A. (2014). The critical role of feedback in formative instructional practices. *Intervention in School and Clinic, 50*(2), 96–104. https://doi.org/10.1177/1053451214536044
18. Zhang, J., Hao, B., Chen, B., Li, C., Chen, H., & Sun, J. (2019). Hierarchical reinforcement learning for course recommendation in MOOCs. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 435–442. https://doi.org/10.1609/aaai.v33i01.3301435
19. Chen, D., Huang, Y., Ma, Z., Chen, H., Pan, X., Ge, C., ... Zhou, J. (2024). Data-Juicer: A one-stop data processing system for large language models. In *Companion of the 2024 International Conference on Management of Data* (pp. 120–134). https://doi.org/10.1145/3626246.3653385
20. Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023). A systematic literature review on the applications of robots and natural language processing in education. *Electronics, 12*(13), Article 2864. https://doi.org/10.3390/electronics12132864
21. Manickam, I., Lan, A. S., & Baraniuk, R. G. (2017). Contextual multi-armed bandit algorithms for personalized learning action selection. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6344–6348). IEEE. https://doi.org/10.1109/ICASSP.2017.7953377
22. Wang, F. (2014). Learning teaching in teaching: Online reinforcement learning for intelligent tutoring. In *Future Information Technology: FutureTech 2013* (pp. 191–196). Springer. https://doi.org/10.1007/978-3-642-40861-8_29
23. Michael, S., Sohrabi, E., Zhang, M., Baral, S., Smalenberger, K., Lan, A., & Heffernan, N. (2024). Automatic short answer grading in college mathematics using in-context meta-learning: An evaluation of the transferability of findings. In *Artificial Intelligence in Education* (pp. 409–417). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64315-6_38
24. Deepika, J. (2025). *Context-aware intelligent learning environments for adaptive digital education*. *National Journal of Ubiquitous Computing and Intelligent Environments*, 34–42. https://fsrap.com/index.php/NJUCIE/article/view/76
25. Abdelrahman, G., & Wang, Q. (2019). Knowledge tracing with sequential key-value memory networks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 175–184). https://doi.org/10.1145/3331184.3331195
26. Barek F. Fatem, & Len Gelman. (2025). Deep Learning-Driven Speech Synthesis and Noise Reduction for Next-Generation Assistive Communication Systems. *Journal of Intelligent Assistive Communication Technologies*, *1*(2), 41-48.
27. Jeromy R, & Jebamalar Tamilselvi J. (2026). Multimodal Fusion of fMRI and EEG for Cognitive State Analysis using Graph Neural Networks (GNNs). *National Journal of Antennas and Propagation*, 83-94.
28. P.Dineshkumar. (2026). Intelligent Protection Schemes for Modern Power Systems Using Deep Learning-Based Fault Classification. *National Journal of Intelligent Power Systems and Technology*, 34-41.