STEM Education

STEM Education, 6(4): 606–631
DOI: 10.3934/steme.2026025
Received: 02 March 2026
Revised: 08 May 2026
Accepted: 13 May 2026
Published: 10 June 2026

https://www.aimspress.com/journal/steme

Research article

AI-based Q-learning model for personalized learning strategies for students with disabilities

Theyazn H.H Aldhyani[1,*], Samina Amin[2], Mossab Saud Alholiby[3] and M. Irfan Uddin[4]

[1] Applied College in Abqaiq, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia; taldhyani@kfu.edu.sa
[2] Institute of Computing, Kohat University of Science and Technology, Pakistan; samina@kust.edu.pk
[3] Educational Leadership Department, Education College, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia; Malholiby@kfu.edu.sa
[4] Department of Computer Science, University of Swabi, Swabi 23540, KP, Pakistan; irfanuddin@uoswabi.edu.pk

*** Correspondence:** Email: taldhyani@kfu.edu.sa; Tel: 00966504937970.

Academic Editor: Jun Shen

Abstract: Students with disabilities find learning challenging in traditional learning environments due to the lack of instructional strategies that are not personalized or adaptive. Various studies have focused on identifying learning difficulties such as dyslexia, dysgraphia, and dyscalculia, which require a multi-step screening process under the supervision of psychologists. Identifying these difficulties is challenging but essential, as it impacts a student's learning and academic success. Everyone's comprehension ability depends on several factors, including the experience and knowledge they bring to the learning environment. With the evolution of technology and the advancement of e-learning platforms, adaptive e-learning has bridged the gap between students' needs and educational institutions' extra classes, enabling students to select targeted courses aligned with their interests. Since the onset of COVID-19, universities have recognized the necessity of online learning and have continued to use these platforms for student assessment. Educational institutions seek innovative strategies to enhance personalized learning (PL) for students with disabilities. The use of technology in schools has created new opportunities for PL, especially for students with disabilities. Traditional learning systems frequently neglect the diverse focuses and distinct requirements of students with visual, cognitive, motor, or auditory impairments. Because of this lack of flexibility,

students with disabilities are less likely to be engaged and to complete tasks. Identifying and using effective learning strategies that work for each student remains a significant challenge in education. To address this issue, this paper proposes a reinforcement learning–based PL system for students with disabilities (PLS-SD) using Q-learning. We suggest PL actions, including audio instructions, augmented reality, and text-based resources. A reward system based on real-world outcomes, i.e., how well students complete their work and how engaged they are, helps them learn. Experimental results demonstrate that the proposed approach effectively identifies optimal actions for different learner states, with immersive and adaptive strategies, such as augmented reality and interactive content, consistently achieving higher rewards. The model shows stable learning behavior across training episodes and successfully adapts its policy to maximize learner engagement and task completion. As indicated by comparing the results with currently advanced models, the proposed method outperforms approaches that are considered traditional by providing context-aware and adaptable recommendations. These findings highlight the potential of reinforcement learning to support scalable and personalized educational solutions for diverse learners.

Keywords: reinforcement learning, Q learning, artificial intelligence, neural network, modeling

1. Introduction

Inclusive education aims to provide equitable learning opportunities for all students, regardless of their physical, cognitive, or sensory abilities. In recent years, technological advancements have emerged as powerful tools for enhancing accessibility in education [1,2]. PL has been emphasized as a key goal and reform initiative in modern education [3]. PL methods are especially useful for students with disabilities, who often struggle to find jobs because they do not have equal learning chances. This problem has grown worse with online learning. While PL systems can help, many e-learning platforms still do not support the needs of learners with cognitive disabilities [4,5].

PL path recommendation creates a learning path based on a learner's goals, abilities, and personal characteristics. While an adaptive learning environment offers personalized content to learners for self-directed study [6,7], current methods, such as global optimal and local iterative recommendations, suggest a fixed order of learning materials. This lack of flexibility makes learning harder for students. In addition, these methods cannot fully adjust the learning path as a student's knowledge changes over time [8–10].

A learning style is the way an individual learns, shaped by their preferences, strengths, and weaknesses [11]. Universities often offer support for students with disabilities, but few programs specifically address the needs of those with intellectual or developmental disabilities [12–16].

In recent years, there have been notable developments in utilizing RL (Reinforcement Learning) techniques to enhance the personalization capabilities of RSs (Recommender Systems). RL is one of the subcategories of machine learning (ML). RL agents learn optimal actions through trial-and-error interactions with the environment [17]. RL determines the optimal action or path in a given scenario. Unlike supervised learning, RL does not have a training set with a defined answer and instead trains the agent through experience in unpredictable environments [18]. The agent takes actions, iteratively learning how to complete the task. The reward serves as a gauge of how successful the preceding action was in achieving the objective. RL algorithms dynamically adjust learning content based on the

student's progress, strengths, and challenges. For instance, if a student with dyslexia struggles with reading comprehension, the system prioritizes visual or auditory learning materials.

However, despite these advances, limited work has focused on applying RL specifically for students with disabilities by explicitly modeling learner states (e.g., disability type and engagement level), defining targeted learning actions, and designing reward structures based on engagement and task completion.

This research aims to enhance learning engagement for students with disabilities by identifying and recommending the most effective learning strategies and designing a reward system that reflects real-world learning outcomes, such as task completion rates and engagement improvements, for various states.

The subsequent sections are organized as follows: A brief overview of background studies on related work is presented in Section 2. Section 3 shows a system design of the proposed work, while Section 4 presents the results and evaluation. Section 5 concludes the paper and outlines directions for future research work.

2. Literature review

Various studies have focused on identifying learning difficulties such as dyslexia, dysgraphia, and dyscalculia, which require a multi-step screening process under the supervision of psychologists. This remains a challenging but important problem in educational systems. Dutt et al. [19] proposed an intelligent tutoring system framework to identify learning disabilities, and 24 participants (both with and without learning disabilities) were assessed. The intelligent tutoring system design includes a pre-test analysis, followed by system-based screening of the child's responses to detect learning difficulties. Neural network classifiers, specifically the fuzzy min-max neural network, were applied to create learner profiles and identify disabilities. Fuzzy sets were used in the supervised learning process to classify and profile learners with disabilities in the intelligent tutoring system [19]. While this approach is effective for identifying learning difficulties, it mainly focuses on diagnosis rather than recommending adaptive learning strategies.

Minoofam et al. [20] presented RALF, an adaptive RL framework using Cellular Learning Automata to automatically create content for dyslexic students. RALF begins by generating simplified alphabet models, then generates Persian words through algorithms that account for character states. This work highlights the potential of RL in supporting specific disabilities such as dyslexia; however, it is limited to content generation for a single disability type and does not generalize across multiple learner profiles.

Learning disabilities can be classified into different categories, and analyzing them helps students improve their weak areas. Existing models for such analysis are often complex and not scalable. To address this, Modak and Gharpure [21] proposed a model that assessed students with and without LD using multimodal analysis. It collects real-time responses to various question types, processes them through a correlative engine, and evaluates performance based on accuracy, response time, and other factors. Khabbaz et al. [22] introduced a customized serious game designed to assess social skills in children with autism spectrum disorder (ASD). The game adapts to each child's abilities using RL and evaluates social skills through fuzzy logic. An intelligent agent adjusts the game's difficulty by modifying elements, shapes, movements, and speed based on the child's progress. If communication skills improve during gameplay, challenges increase dynamically to ensure continuous evaluation. The

system estimates the player's level by analyzing factors like gaze points and correct responses. Three children with ASD participated in five game sessions each, and results showed that the approach is effective in the long run. These approaches demonstrate the use of adaptive systems for assessment and skill development, but they primarily focus on evaluation rather than continuous personalized learning strategy recommendation.

An online learning platform has emerged as a vital resource for learners, offering convenient access and abundant educational content. To address the dynamic short-term needs and long-term learning objectives of online learners, Ma et al. [23] proposed user modeling and learning path recommendation. They introduced a novel approach, cDQN-PathRec, that integrates multi-behavior user modeling with cascading deep Q-networks to learn path recommendations. The proposed model uses a knowledge graph-based multi-behavior transformer architecture for user state modeling, incorporating factors such as a learner's knowledge background, learning styles, settings, and preferences. A cascading DQN with a two-tier reward function is employed to guide the agent toward achieving both balanced global and local optima. Unlike disability-focused approaches, this work targets general learners and does not explicitly consider disability-specific needs or engagement variations.

F. Zhang et al. [8] introduced a process-type learning path model and a recommendation approach that presents learning paths as flowcharts. It dynamically suggests branching paths based on learners' evolving knowledge states throughout their learning journey. Specifically, deep knowledge tracing is used to annotate learners' knowledge states from historical logs, while process mining is employed to generate a personalized process-type learning path that captures sequences, parallel relationships, and selection relationships among learning objects. [24] proposed a deep Q-learning (DQN) method informed by marketing psychology. While these studies improve adaptive learning and recommendation systems, they are not specifically designed for students with disabilities or disability-aware personalization.

Shawky and Badawi [25] proposed an adaptive learning approach tailored to the dynamic needs of individual learners and diverse educational contexts, including both solo and collaborative environments. The system uses RL to build an intelligent framework that not only suggests PL materials but also adapts to the learner's changing states (such as engagement and knowledge) and acceptance of educational technologies. We demonstrate the feasibility of the approach through simulation-based evaluations, and the results indicate strong potential for real-world use. We present a data-driven PL platform enhanced by reinforcement learning and big data tools, developed by [26], which combines knowledge tracking and adaptive AI, leading to a 62% progress increase over traditional methods. Imamah et al. [27] used the ant colony algorithm and item response theory (IRT) for learning improvements. Authors in [28] introduced an AI-enabled intelligent assistant for higher education. Similarly, the work presented in [29] applied an AI-based dashboard for educational leaders, while [25] emphasized the necessity of teaching AI in schools to prepare students for a technology-driven future. Overall, these studies demonstrate increasing adoption of AI in education, but their range tends to overlook disability-aware adaptive decision-making.

Recent research also acknowledges the influential, positive presence of AI to help engage learners and enhance creativity and digital literacy in education. Research conducted in [30] outlines that learners' perceptions toward AI and their attitude toward trust impact the effectiveness and adoption of AI-based education tools in digital learning. Research surrounding the presence of generative AI in education, conducted by [31], outlines the positive presence of generative AI in enhancing creativity,

learner engagement, and language proficiency through the support of an interactive and adaptive digital learning environment. Further work [32] and [33] has also explored how AI dependency and learner perception affect motivation and vocabulary acquisition, emphasizing the psychological and behavioral aspects of AI-driven learning. These findings collectively suggest that beyond algorithmic personalization, successful educational systems must also consider learner engagement, trust, and interaction when designing adaptive learning environments.

To our knowledge, current research highlights the importance of machine learning techniques for adapting educational content to learners' levels. Despite progress in machine learning–based educational systems, there is still limited work on reinforcement learning frameworks that jointly model disability type, engagement level, and adaptive learning intervention selection based on real-world feedback. This study addresses this gap by proposing a Q-learning-based personalized learning system for students with disabilities.

3. Materials and methods

This section presents a proposed RL-based framework for personalized learning strategies for students with disabilities. Figure 1 shows the flowchart of the proposed model. The explanation of the model is described in the following subsections.

3.1. System components (states and actions)

The number of states(s) and actions (a) forms the dimensions of the Q-table, which will store learned values for state-action pairs.

3.1.1. States

States represent different conditions of students' disabilities, such as engagement level or task performance. These states account for a) type of disability (e.g., cognitive, visual, hearing, motor, or general disabilities), b) engagement level (e.g., low or high), and c) task completion rate (e.g., slow or fast), and can be seen in Table 1.

Table 1. Types of states.

S. No	State types
1.	Cognitive impairment: low engagement
2.	Visual impairment: low engagement
3.	Hearing impairment: low engagement
4.	Motor impairment: low engagement
5.	Cognitive impairment: high engagement
6.	Visual impairment: high engagement
7.	Hearing impairment: high engagement
8.	Motor impairment: high engagement
9.	General disabilities: task completion rate: slow
10.	General disabilities: task completion rate: fast

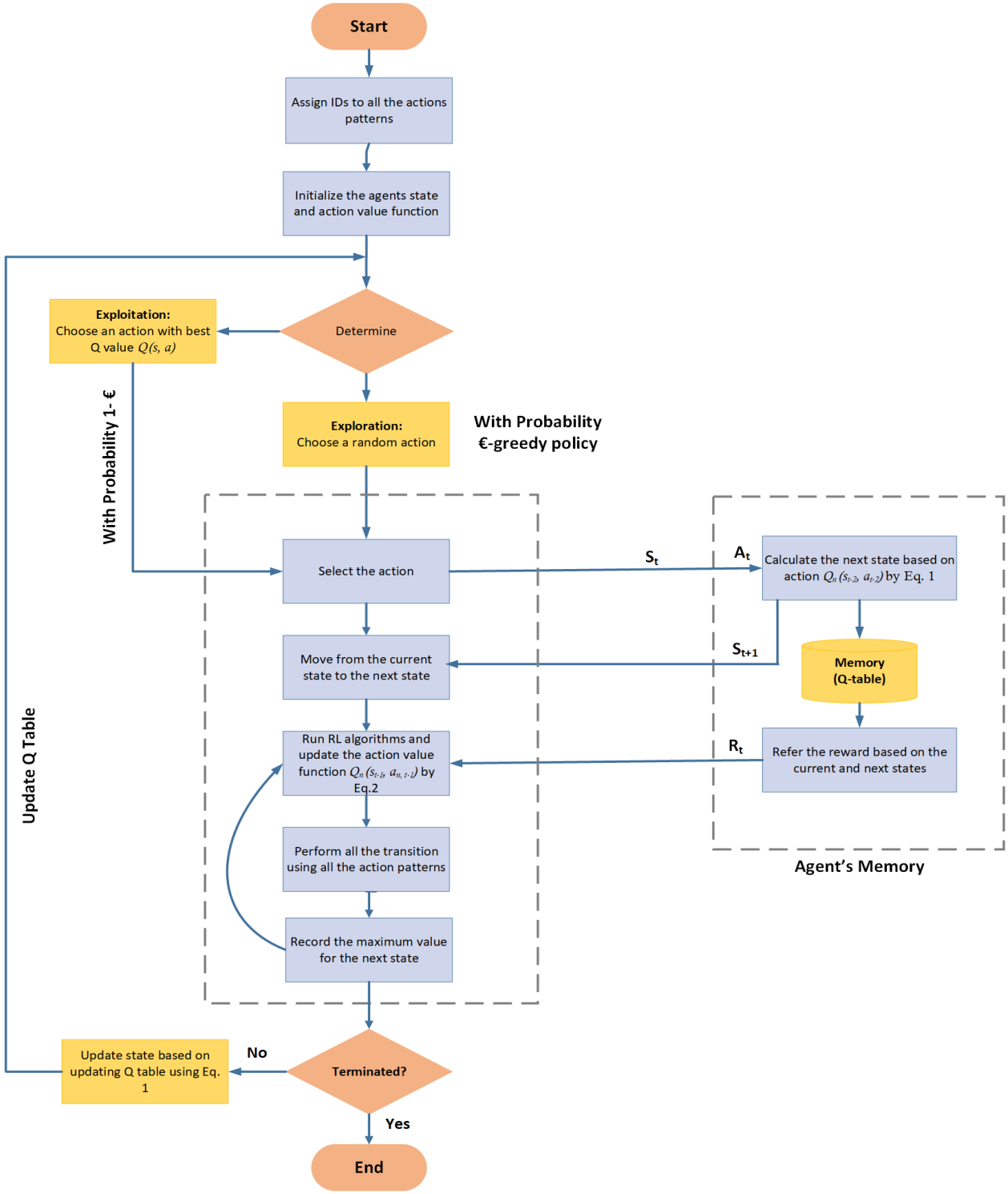

Figure 1. Proposed flowchart for PLS-SD using RL.

3.1.2. Actions

Actions correspond to PL strategies that can be applied to improve engagement and learning outcomes. The selected actions are presented in Table 2.

Table 2. Defined actions.

S. No	Actions
1.	Text-to-speech
2.	Voice-activated navigation
3.	Real-time captioning
4.	Adaptive pacing
5.	Gamified learning
6.	PL pathways
7.	Augmented reality
8.	Speech-to-text

3.2. Q-table initialization

A Q-table is created with dimensions equal to the number of states × number of actions. It is initialized with random values in the range [0, 1]. Each value in the Q-table represents the "quality" (or expected cumulative reward) of performing a specific action in a specific state.

Reward matrix (r)

A reward matrix is created, where each state is associated with a list of rewards for all possible actions. We have assigned the rewards to state–action pairs as outlined in Table 3. For example, the reward for using "text-to-speech" for "cognitive impairment: low engagement" might be higher than for "augmented reality". Rewards quantify the impact of applying a specific action in each state. The design of the reward function ensures that the system encourages strategies that lead to improved engagement and learning performance.

Table 3. Reward assignments for states and actions.

State	Action	Reward
Cognitive impairment: low engagement	Text-to-speech	+5
Cognitive impairment: low engagement	Adaptive pacing	+7
Cognitive impairment: high engagement	Gamified learning	+8
Cognitive impairment: high engagement	PL pathways	+9
Visual impairment: low engagement	Voice-activated navigation	+6
Visual impairment: low engagement	Augmented reality	+8
Visual impairment: high engagement	Text-to-speech	+9
Visual impairment: high engagement	Adaptive pacing	+7
Hearing impairment: low engagement	Real-time captioning	+7
Hearing impairment: low engagement	Speech-to-text	+6
Hearing impairment: high engagement	Gamified learning	+9
Hearing impairment: high engagement	PL pathways	+8
Motor impairment: low engagement	Voice-activated navigation	+8
Motor impairment: low engagement	Adaptive pacing	+7
Motor impairment: high engagement	Augmented reality	+9
Motor impairment: high engagement	PL pathways	+8
General disabilities: task completion rate: slow	Adaptive pacing	+6
General disabilities: task completion rate: slow	Gamified learning	+7
General disabilities: task completion rate: fast	PL pathways	+9
General disabilities: task completion rate: fast	Augmented reality	+8

3.2.1. Q-learning parameters

Learning rate (α): Controls how much new information overrides old knowledge. A value of 0.1 allows gradual updates. We have selected this value to ensure that the values are gradually updated in the Q-table, as it prevents the model from overreacting to experiences recently encountered or if there is noise in rewards, while maintaining stability in training. Particularly in situations where the rewards fluctuate, this learning rate helps in updating the values slowly and enabling the agent to converge reliably to an optimal policy over time.

Discount factor (γ): Balances immediate rewards and long-term benefits. A value of 0.9 gives more weight to future rewards. The values also place a strong emphasis on long-term rewards while still considering immediate gains. The value encourages the agent to plan and make a decision that helps in maximizing cumulative rewards rather than focusing on short-term benefits. It is beneficial in situations with delayed consequences and to prioritize a long-term strategy.

Epsilon (ε): Determines exploration (choosing random actions) vs. exploitation (choosing the best-known action). With ε = 0.2, 20% of the time, random actions are taken to explore.

3.2.2. Q-learning algorithm

Q-learning is an off-policy, model-free learning approach [34]. Off-policy suggests that it is not necessary to adhere to a certain policy; the agent's behaviors could instead be arbitrary, and it will still be capable of discovering the best course of action [35,36].

RL creates a Q-table for an environment, with dimensions (s x a), where s and a are the number of states and actions, respectively. The state-action values in the Q-table determine the likelihood of selecting a particular action for each state. The learned value is the key aspect of Q-learning. The sum of all the actions in state s' is used to determine the estimation of the ideal future value. If there are two feasible actions a_1 and a_2 for a state s, the action with the higher Q-value will be picked. If both have identical Q-values, a random action is taken.

The structural design of Q-learning is depicted in Figure 2. The data structure developed to help determine the highest projected future rewards for action at each state is termed the Q-table. In essence, this table will show us what to do in each state. Q-learning is used to learn the values in the Q-table. Figure 3 is a flow of the Q-learning inspired by a similar flow diagram published in [37]. The stages involved in building a Q-table for an agent that must learn to execute, fetch, and sit are graphically visualized in Figure 3.

Environment: The environment represents the learning system that interacts with the agent. It is formally defined using a transition function T(s′ | s,a), which determines the probability distribution over the next states based on the current state and selected action. It also provides reward feedback R(s,a) and updates the state accordingly.

Agent: Observes the present situation, acts following its policy, and is rewarded for upgrading (learning) its policy.

State: s stands for the state of an agent in an environment at any given time.

Action: The agent's step when it is in a certain state is called Action *a*.

Rewards: The agent will obtain good or poor rewards for each action.

Episodes: When an agent arrives at a terminating phase and is incompetent to execute any additional actions.

Q-values: Used to evaluate the effectiveness of an Action, a, taken at a specific state, s, $Q(s,a)$.

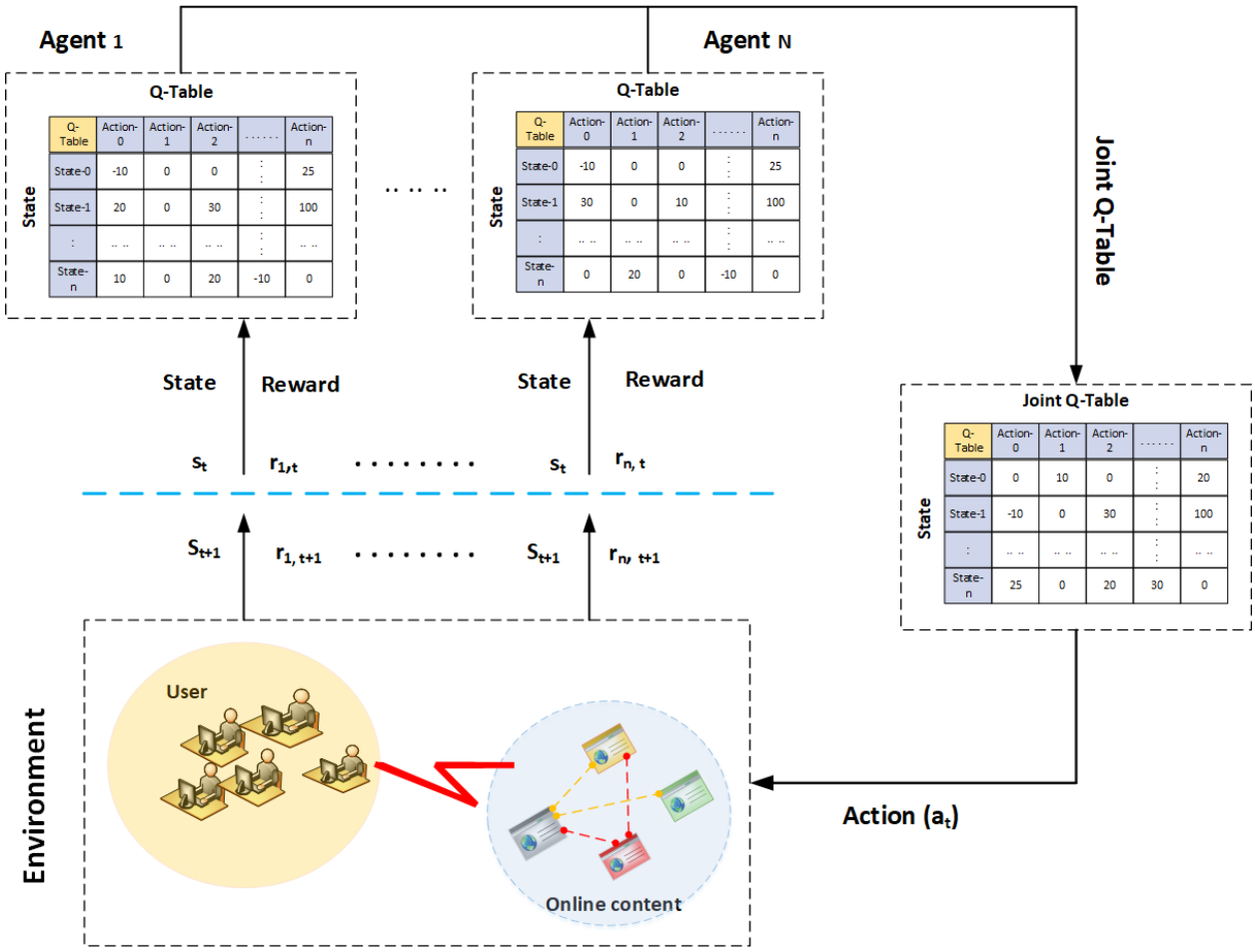

Figure 2. Framework of multi-agent Q-learning structure.

The Bellman equation is employed to evaluate a state and estimate how advantageous it is to remain in or adopt that state. The maximum reward will be received by the optimal value in the ideal state. The value function is updated by value-based methods using an equation (particularly Bellman equation 4). Equation 4 finds the future state of an agent by using the present state and the reward connected to that state, the maximum anticipated reward, and a discount rate that evaluates its significance to the current state. The learning rate controls how rapidly or slowly the model will pick up new information.

$$Q(s,a) = Q(s,a) + \alpha\,[R\,(s,a)] + \gamma\, max\, Q^{'}\,(s^{'},\ a^{'}) - Q(s,a) \quad (1)$$

Where $Q(s,a)$ on the left side shows the new *Q-value*, and $Q(s,a)$ on the right side represents the current *Q-value*, α shows the learning rate, $R\,(s,a)$ shows the reward value, γ is the discount rate, and $max\, Q^{'}\,(s^{'},\ a^{'})$ is the maximum expected future reward. Algorithm 1 demonstrates the detailed implementation of the proposed work. The algorithm adaptively recommends learning strategies based on a student's cognitive ability, engagement level, and skill proficiency. It initializes a Q-table, updates action-value estimates through exploration and exploitation, and determines the best intervention based on learned experience. The reward matrix given in Table 3 is synthetically defined for simulation purposes and represents heuristic estimates of relative pedagogical effectiveness rather than real-world measured values. It does not reflect empirically validated educational outcomes. In

practical settings, reward values should be derived from expert evaluation and/or empirical learner interaction data to ensure pedagogical validity.

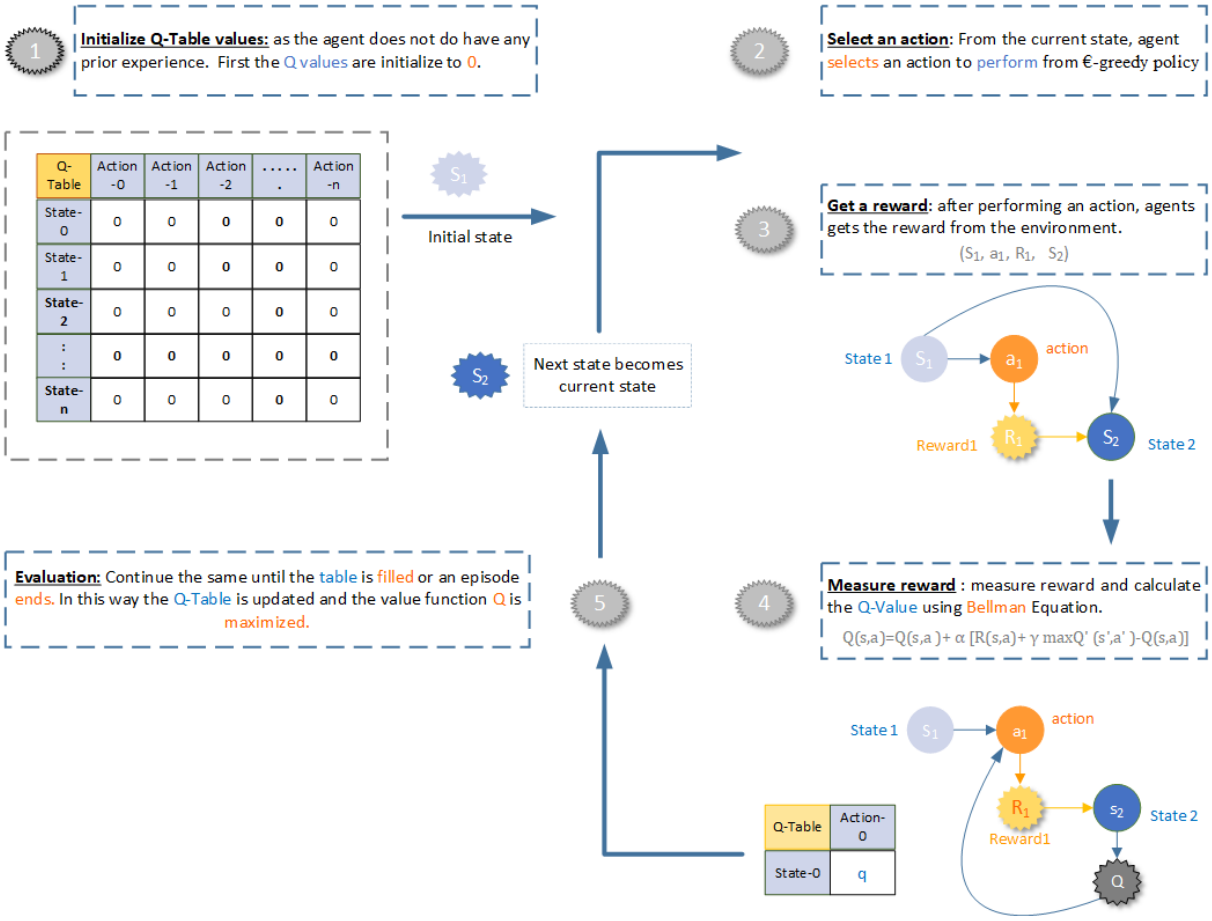

Figure 3. Processing of Q-learning algorithm.

Algorithm 1: Q-learning-based adaptive learning strategy for students with disabilities

BEGIN
DEFINE states as a list of cognitive abilities, engagement levels, skill levels, etc.
DEFINE actions as a list of possible interventions
INITIALIZE Q-table with zeros of size (*number of states* × *number of actions*)
SET learning_rate = 0.1, discount_factor = 0.9, epsilon = 0.2, episodes = 1000
INITIALIZE reward matrix based on predefined heuristic estimates (Table 3)

FOR each episode **DO**
SET state **TO** random state
WHILE true **DO**
IF random < epsilon **THEN**
SELECT action randomly (exploration)
ELSE
SELECT action with max Q-value (exploitation)
END IF
SET next_state **TO** a T(state, action)
SET reward **TO** rewards[state][action]

```
            UPDATE Q-value using formula:
                Q(s,a) = Q(s,a) + α [R (s,a)] + γ max Q′ (s′, a′) − Q(s,a)
          SET state TO next_state
          IF random < 0.1 THEN
            BREAK LOOP
          END IF
        END WHILE
      END FOR
PRINT "Learned Q-table"
PRINT Q-table

DEFINE decide_action(state_index)
  RETURN action with max Q-value
END FUNCTION

DEFINE T(state, action)
  DEFINE probability distribution P(s′ | state, action)
  IF action improves engagement THEN
    ASSIGN higher probability to higher engagement state
  ELSE
    ASSIGN higher probability to same or nearby states
END IF
SAMPLE next_state from P(s′ | state, action)
RETURN next_state
END Function

SET recommended_action = decide_action (0)
PRINT recommended_action
END
```

4. Results and discussion

The learned Q-table represents the outcome of applying the Q-learning algorithm to optimize PL strategies for students with disabilities. Values in the Q-table denote the future value of rewards associated with each state–action, with greater values implying actions that are more successful for that particular state. Moreover, each state in the Q-table refers to a particular state, which denotes categories of students with regard to their disability and involvement in studies. Every column in each row refers to a particular action in that particular state, whose value is denoted by the Q-value.

Table 2 analysis reveals the most effective interventions for various student profiles. For visual impairment with low engagement, augmented reality (Action 4, Q-value: 483.82) and text-to-speech (Action 1, Q-value: 481.84) are highly effective. For cognitive impairment with high engagement, Action 7 (Q-value: 488.08) and Actions 2 and 8 show strong results. In cases of learning disability with moderate engagement, Action 3 (Q-value: 485.90) is the top performer. For physical disability with moderate engagement, Action 8 (Q-value: 487.24) is most beneficial. Students with hearing impairment and high engagement respond best to Action 3 (Q-value: 485.16). For autism spectrum disorder with high engagement, Action 3 (Q-value: 487.97) is highly effective. Visual impairment with high engagement benefits most from Action 3 (Q-value: 486.74). For multiple disabilities with

low engagement, Action 1 (Q-value: 487.28) is the best choice. Students with dyslexia and moderate engagement respond well to Action 4 (Q-value: 483.18). Lastly, for speech impairment with low engagement, Action 4 (Q-value: 485.85) and Actions 6 and 8 are highly effective. These findings highlight the importance of personalized interventions, such as augmented reality, text-to-speech, and interactive tools, to improve engagement for students with diverse disabilities.

Table 4 shows that different actions are effective in different situations. The action with the highest Q-value is recommended. In the state (visual impairment: low engagement), the action (augmented reality) had the highest Q-value of 483.56. For example, augmented reality, along with other immersive technologies, is useful for addressing the requirement of visually impaired learners with low engagement. Similar is the case with (hearing impairment: low engagement), where the maximum Q value was recorded as 487.52, again by (augmented reality). Thus, it is evident that augmented reality is very useful in terms of engaging learners with hearing impairment. In the case of the state (motor impairment: low engagement), the action (personalized exercises) had the maximum Q value of 485.40. Thus, it clearly emphasizes the need for personalization of the interventions offered to the learners suffering from motor impairment. The state (general disability: low engagement) had a Q value of 487.11 for the action (interactive videos).

Table 4. Generated rewards value across 1000 episodes.

Action state	**Action 1**	**Action 2**	**Action 3**	**Action 4**	**Action 5**	**Action 6**	**Action 7**	**Action 8**
Visual impairment: low engagement	481.84	477.68	478.77	483.82	476.84	477.16	474.08	480.25
Cognitive impairment: high engagement	484.45	481.96	477.87	479.97	478.01	482.07	488.08	485.12
Learning disability: moderate engagement	475.95	480.09	485.90	481.30	478.65	479.14	476.31	477.61
Physical disability: moderate engagement	483.20	486.07	478.71	482.08	479.82	481.39	483.68	487.24
Hearing impairment: high engagement	480.92	479.31	478.69	485.16	484.70	481.55	479.63	481.69
Autism spectrum disorder: high engagement	486.01	484.43	487.97	482.44	485.33	486.11	486.67	486.28
Visual impairment: high engagement	479.38	482.65	486.74	483.09	481.79	484.66	479.12	480.23
Multiple disabilities: low engagement	487.28	483.41	479.94	485.68	484.76	482.64	485.94	486.14
Dyslexia: moderate engagement	476.29	479.74	477.56	483.18	478.68	480.99	476.60	477.51
Speech impairment: low engagement	481.17	480.42	478.47	485.85	484.11	487.66	481.22	485.61

The efficiency of actions improved along with increased engagement. "Interactive videos" obtained a Q-value of 485.64 on "visual impairment: medium engagement" state. This means that

interactive videos may sustain user interest. "Augmented reality" continued being the most efficient action for "hearing impairment: medium engagement" with the Q-value equaling 487.82. In addition, interactive videos were considered the best option for "motor impairment: medium engagement" because they provided a Q-value of 486.73. Finally, when "general disability" and "medium engagement" conditions were observed, "augmented reality" turned out to be the best action to perform; its Q-value amounted to 486.89.

It should be noted that "interactive exercises" proved efficient for the "visual impairment" state by providing a Q-value of 482.52, while "interactive videos" was the optimal action on "hearing impairment" (Q-value of 487.46). The learned Q-table (refer to Table 5) revealed that "augmented reality" could serve as an optimal choice in many cases, because the highest Q-values were associated with this type of action in a number of states. For instance, in the "visual impairment: low engagement" state, the Q-value was 483.5612. For "hearing impairment: low engagement", it also got the highest Q-value of 487.5216, which shows that it can provide hearing-impaired learners with immersive and rewarding experiences.

Table 5. Q-value obtained in the process of using different learning methods, considering the type of disability, to analyze the effectiveness of the proposed method.

Visual impairment: low engagement	**Augmented reality**	**483.5612**	**The Q-value indicates that using augmented reality leads to the highest reward for engaging learners with visual impairments and low engagement.**
Hearing impairment: low engagement	Augmented reality	487.5216	Augmented reality provides the best outcome for engaging learners with hearing impairments.
Motor impairment: low engagement	Personalized exercises	485.4032	Personalized exercises maximize engagement for learners with motor impairments.
General disability: low engagement	Interactive videos	487.1054	Interactive videos achieve the highest Q-value, making them the most effective action for general disabilities and low engagement.
Visual impairment: medium engagement	Interactive videos	485.6444	For medium engagement, interactive videos work best to improve engagement for learners with visual impairments.
Hearing impairment: medium engagement	Augmented reality	487.8207	Augmented reality remains the optimal action for hearing-impaired learners with medium engagement.
Motor impairment: medium engagement	Interactive videos	486.7315	Interactive videos provide the best reinforcement for motor-impaired learners with medium engagement.
General disability: medium engagement	Augmented reality	486.8939	Augmented reality proves most effective for general disabilities with medium engagement.

However, one should note that the results presented in Tables 4 and 5 are not only about ranking the actions in terms of their performance; they represent an illustration of how the Q-learning model chooses different behaviors according to specific conditions. Contrary to what might be expected, these findings prove that each approach can be highly beneficial in certain situations, depending on the kind of disability as well as on the level of engagement. When there are significant differences between Q-values, it means that the preference is obvious; in other situations, when differences are slight, several approaches can be used simultaneously. This analysis reveals that some strategies are more

efficient regardless of certain conditions, whereas others can depend on them. Therefore, it proves that the proposed approach provides opportunities for flexible policy creation.

The results in Figures 5–8 prove that the learning process of the Q-learning model is characterized by stable policy convergence throughout the training sessions. The data demonstrate how the cumulative reward and Q-value variance increase steadily during training iterations. Moreover, the effect of various learning rates shows that the model achieves stability at a moderately fast rate.

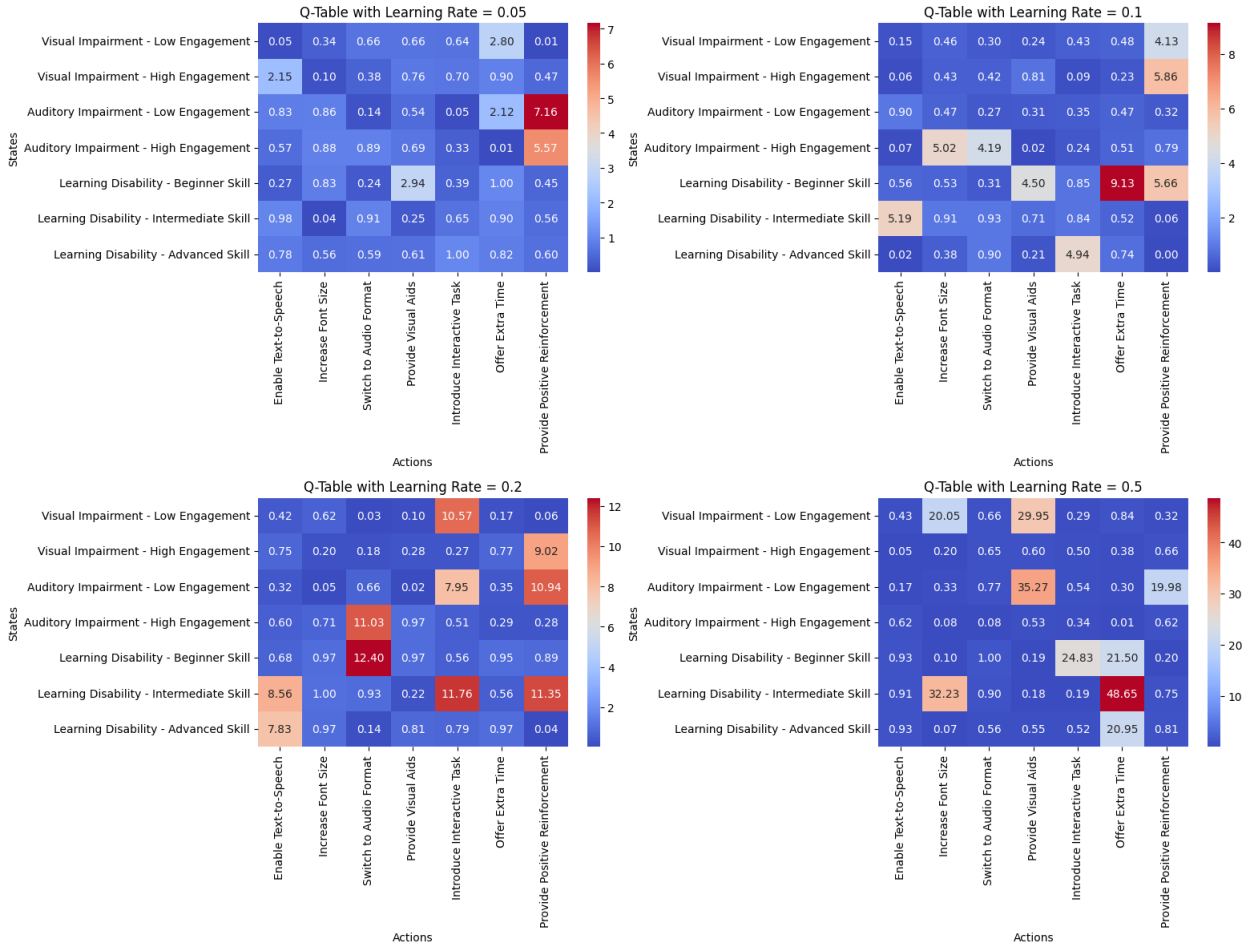

Figure 4. Heatmap visualization of the Q-values.

Figure 4 depicts the expected rewards (Q-values) for different actions in various states, personalized to students with disabilities. Each row represents a state based on disability type, engagement level, and task completion rate, while columns represent possible actions like adjusting difficulty or changing content format. Higher Q-values show more efficient actions for specific states, such as highly engaged students with visual impairments responding well to certain strategies. Lower Q-values suggest less efficient actions, such as for students with low engagement.

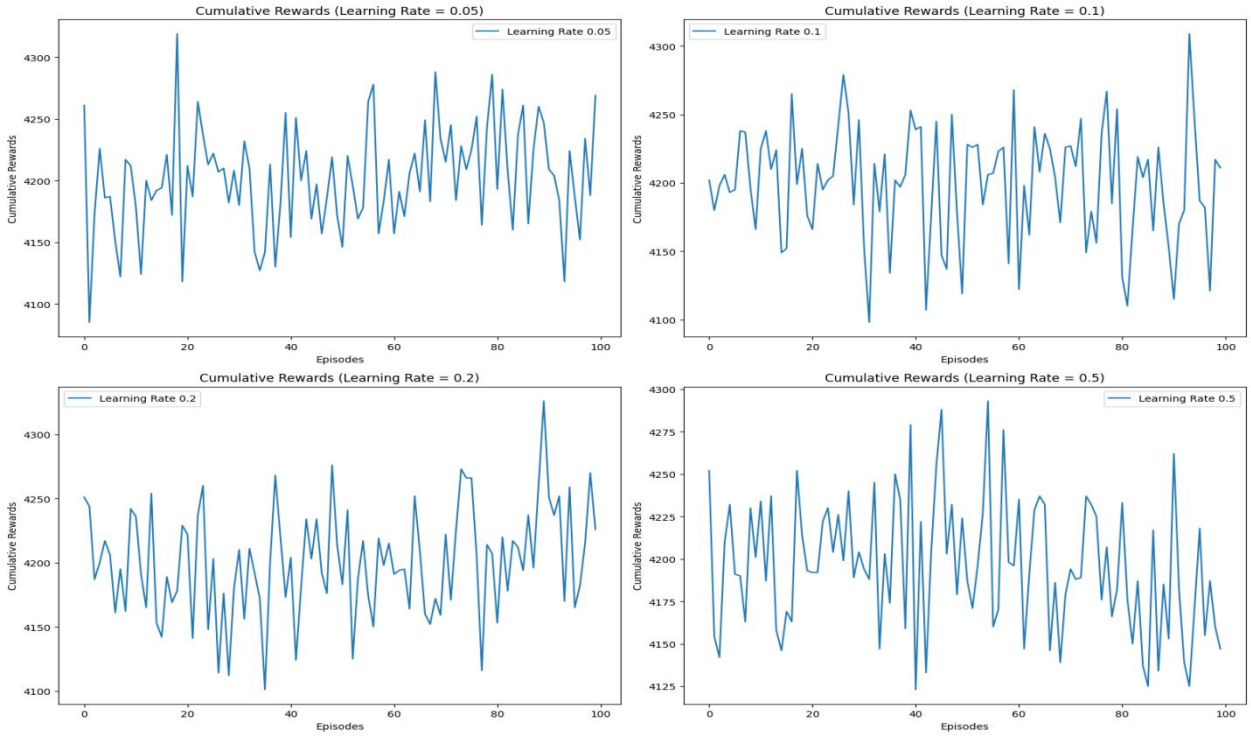

Figure 5. Cumulative rewards for each state–action over 100 episodes using various LR = 0.05, 0.1, 0.2, 0.5.

Figure 5 represents the cumulative rewards for each state–action pair over 100 episodes using various learning rate (LR) values, such as LR = 0.05, 0.1, 0.2, and 0.5. The results show that the Q-learning model successfully learned the optimal actions for various states, with a high success rate and a generally stable policy after several iterations. However, the number of changes in policy for different states reflects the varying levels of complexity in learning the optimal action for each state.

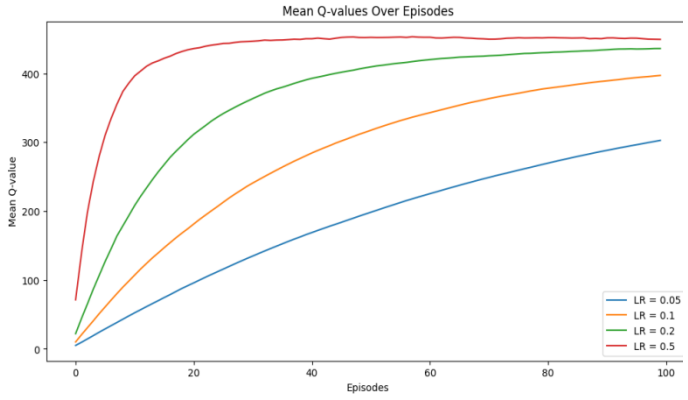

Figure 6. Mean Q-values over episodes.

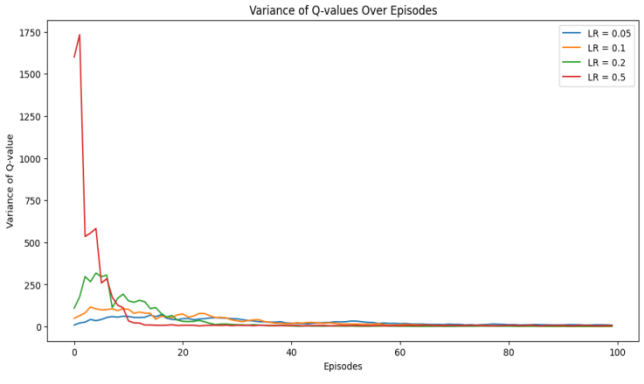

Figure 7. Variance of Q-values over episodes.

Figure 6 shows the average expected cumulative rewards (Q-values) of an RL agent over 100 training episodes for different learning rates (LR) (LR = 0.05, 0.1, 0.2, 0.5). The x-axis (episodes) shows how far along the training has come, and the y-axis (mean Q-value) shows how well the agent is doing. Higher Q-values mean that strategies have been learned better. If LR = 0.5, then high learning is expected; however, it can be quite unstable. In contrast to the high LR, low LR = 0.05 represents a

gradual growth of Q-values. In addition, moderate LR values such as 0.1 or 0.2 provide good results when it comes to the learning process. Thus, moderate LR rates should be considered optimal since they balance speed and stability and make the learning process possible. In this case, it may be assumed that moderate LR helps to retain the interest of the disabled students since exploration and exploitation become balanced.

Figure 7 shows different variances of Q-values (the reward gained) depending on the rate used (0.05, 0.1, 0.2, 0.5) throughout the training process. Episodes represent the training process, while variance illustrates changes in Q-values. High rates (LR = 0.5) provide high variance of Q-values, which means an instability of the system due to the very fast growth of Q-values. Low rates (LR = 0.05), in turn, provide less variance in Q-values. Moderate learning rates (0.1, 0.2) illustrate that balance can be obtained without much variability.

Figure 8 shows the level of Q-value stability with respect to time. The x-axis in both graphs indicates the number of episodes or training iterations. High values for cumulative rewards indicate that the agent is performing well, while Q-value stability represents how consistently the agent learns and applies its policies. Learning is effective, and the policies improve consistently when cumulative rewards increase continuously while Q-value stability increases. Stability variations, including spikes and dips, may indicate the exploration phase or instability due to the high LR value. Consistent, high levels of reward received by disabled students due to their adaptive strategy help ensure reliability in the learning process.

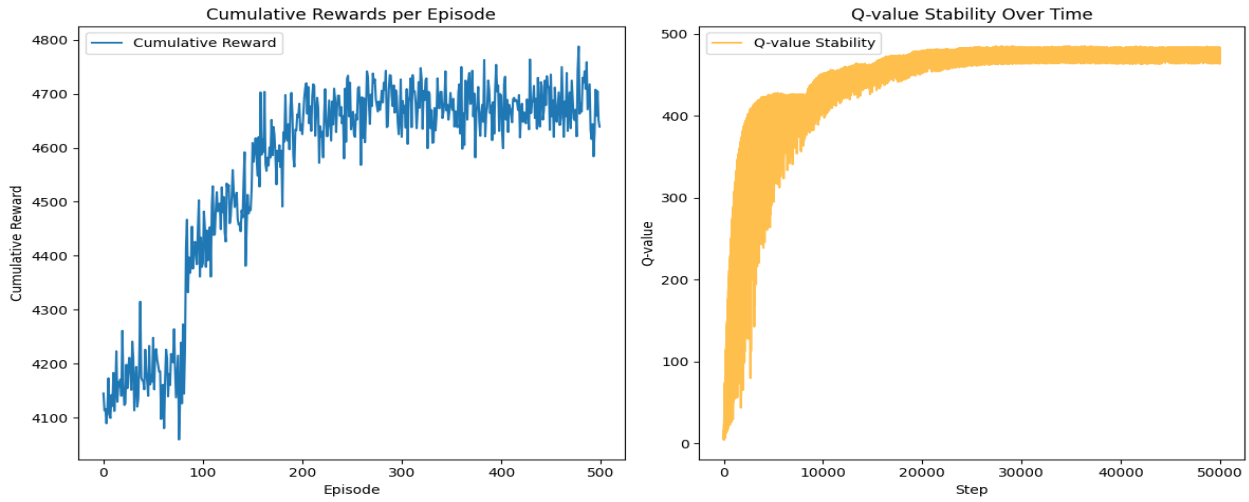

Figure 8. Cumulative rewards per episode and Q-value stability over time.

Figure 9 below depicts the expected reward values (Q-values) for each state and action obtained through 1000 iterations for disabled learners. A Q-table is used to illustrate the rewards obtained using a heatmap. The rows correspond to states determined by the type of disability, engagement, and completion of tasks. Actions refer to changes in tasks, like making them more difficult or changing the content form. High Q-values such as 486.49 suggest that actions are effective in certain scenarios. For instance, highly engaged visually impaired students may be motivated using particular approaches. On the other hand, low Q-values such as 474.92 imply that particular actions are unlikely to work since the learners are not enthusiastic about learning.

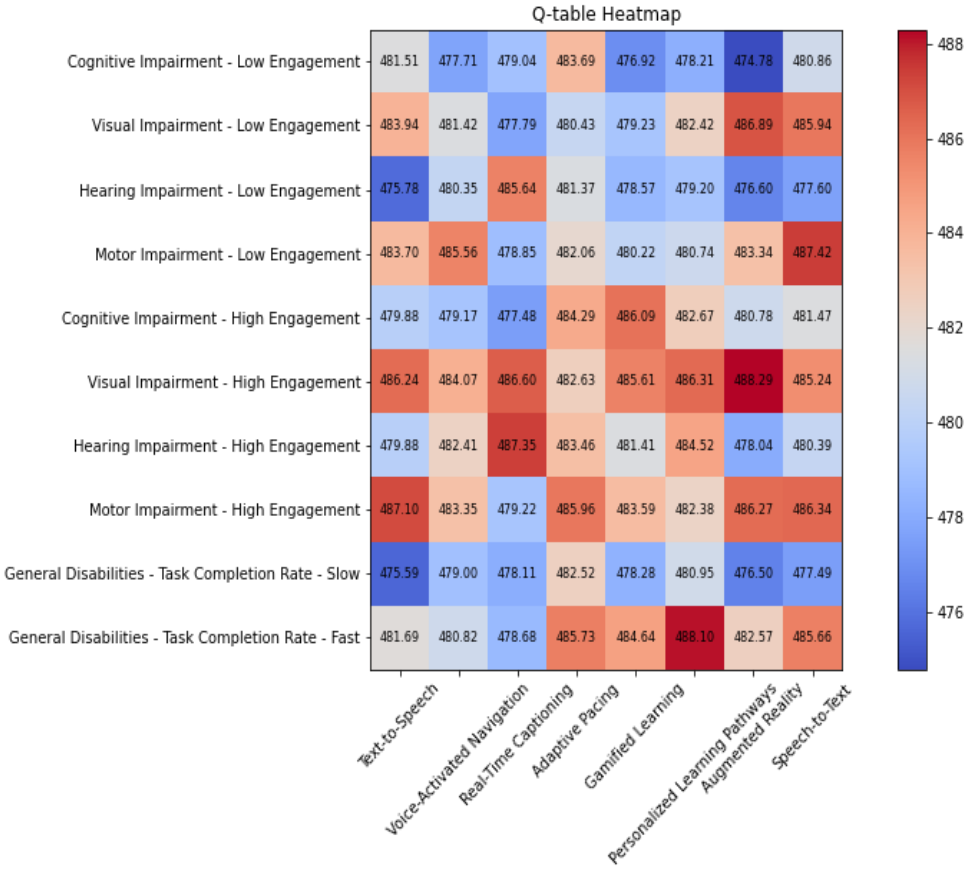

Figure 9. Heatmap visualizing the expected rewards based on Q-values.

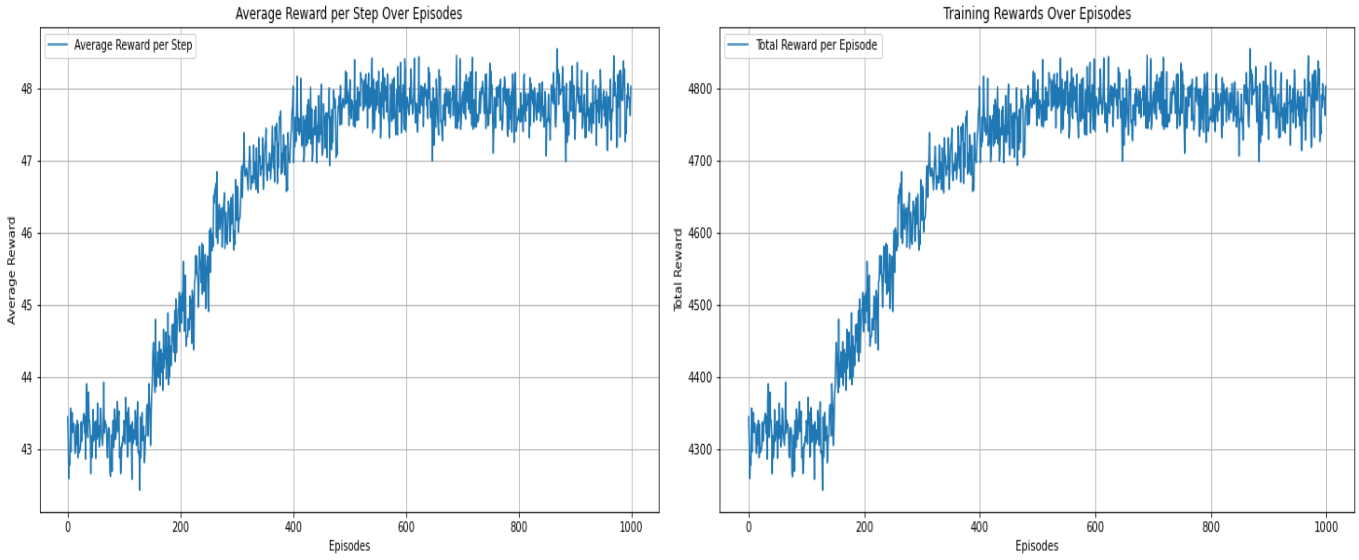

Figure 10. Average and training total reward per episode.

Figure 10 depicts that an RL agent receives a certain amount of reward on average in each of its training sessions. The x-axis indicates the number of training sessions (episodes), and the y-axis shows the amount of total reward received in each session. A positive slope indicates the improvement in the

agent's strategy, and variations depict the tradeoff between exploring and exploiting actions based on their rewards. The graph shows how things like augmented reality or text-to-speech (which can be found by looking at Q-values) can make learning more interesting for students with disabilities. Higher rewards are linked to better actions. Rewards that stay the same over time show that the proposed RL model's adaptive learning strategies are stable, which is important for inclusive education.

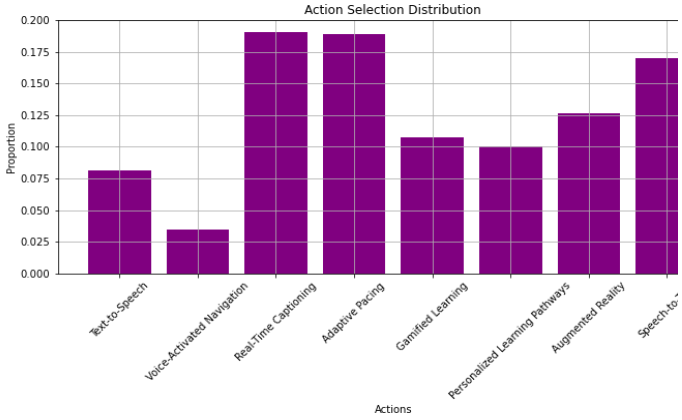

Figure 11. Action selection distribution.

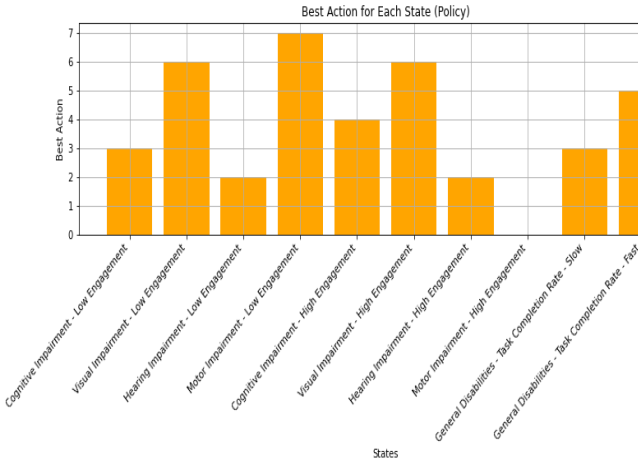

Figure 12. Best action for each state.

Figure 11 shows the probability distribution of action selection across different states. Optimal action selection for each state is depicted in Figure 12, while reward distribution across all state-action pairs can be seen in Figure 13. In RL, the reward distribution over all state-action pairs is computed to characterize how rewards are assigned to actions taken in different states.

Figure 13. Reward distribution for all state–actions pairs.

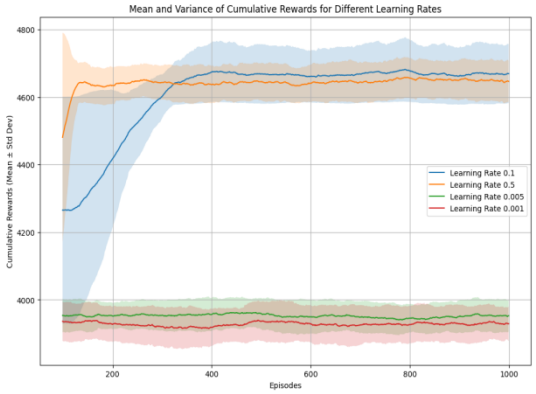

Figure 14. Mean and variance of cumulative rewards for different LRs.

Figure 14 shows the average performance (mean) and variability (standard deviation) of cumulative rewards over 1000 episodes for various LRs. The x-axis (episodes) shows training progression, while the y-axis (cumulative rewards, 4000–4800) reflects the agent's success. Higher LR (e.g., 0.5) likely shows greater variability (wider shaded areas) due to aggressive updates, while lower rates (e.g., 0.001) exhibit slower reward growth but steadier performance. A moderate rate (e.g., 0.1) balances speed and stability when optimizing LR for adaptive strategies. This figure validates how PL rates improve reliability in personalized education for students with disabilities, to validate consistent outcomes without overfitting.

The performance of the RL model throughout 1000 episodes is indicated in Figure 15 in terms of the success rate. In this graph, the number of episodes is placed on the x-axis, while the rolling success rate (4000–4800) is on the y-axis. Each line corresponds to the learning rates of 0.1, 0.5, 0.005, and 0.001. Meanwhile, the shading represents the standard deviation, which reflects how successful the model becomes. The smaller the standard deviation, the better the performance is.

As seen from the figure, the learning rate of 0.1 performs consistently at a relatively high success rate. Meanwhile, 0.5 produces fluctuating success rates. On the other hand, learning rates of 0.005 and 0.001 require a considerably longer time to improve the success rate, suggesting their inefficient use. Therefore, a learning rate of 0.1 achieves the most optimal results.

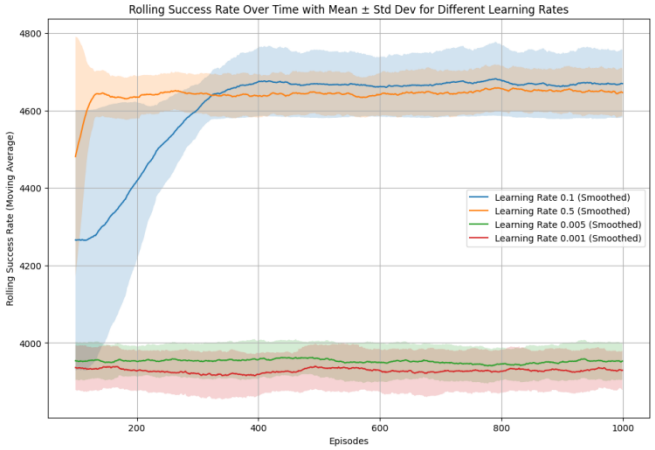

Figure 15. Rolling success rate over time with mean ± std dev for different learning rates (0.1, 0.5, 0.005, 0.001).

Figure 16. Cumulative rewards per episode with mean ± std dev.

Figure 16 provides the results of reinforcement learning by running different episodes, where the reward is summed up under different learning rates (0.1, 0.5, 0.005, and 0.001). Actual data represent the real value of rewards, while smoothed data provide the trend of performance. The shaded areas represent the degree of variance in performance; thus, the smaller the shaded area, the higher the consistency. A learning rate of 0.1 provides consistent performance with high rewards, making it suitable as an equilibrium point for the learning rate. A learning rate of 0.5 performs better than other learning rates with regard to fluctuations in performance. Learning rates below 0.01 perform poorly compared to others, meaning that their contribution toward reinforcement learning is insignificant.

Figure 17 displays the findings for the following LR values: 0.001, 0.005, 0.1, and 0.5. When LR is set to 0.001, learning will be stable but very slow. When LR is set to 0.005, learning becomes slightly faster but is still controlled. An ideal value for LR would be 0.1, as it allows learning to progress swiftly and smoothly. On the other hand, an LR that is too high, such as 0.5, might expedite learning but may result in overly divergent policy updates. It may also cause instability during learning.

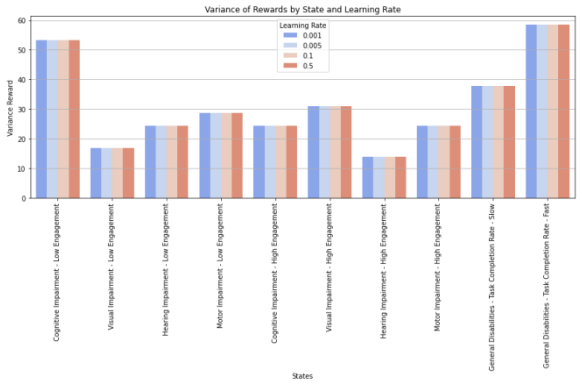

Figure 17. Variance of rewards by states and learning rates.

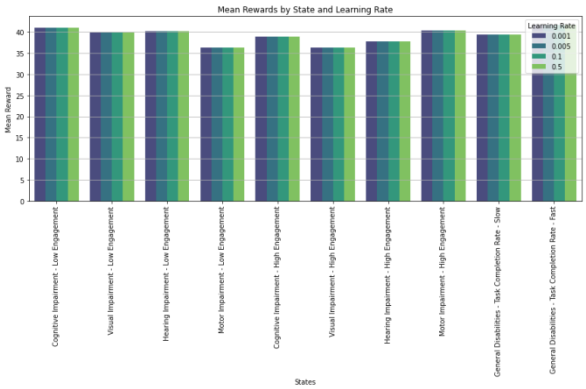

Figure 18. Mean rewards by state and LR.

LRs for RL are presented in Figure 18 (0.001, 0.005, 0.1, 0.5). Learning rate plays an essential role in improving the performance of the RL model. Low values of LRs (0.001, 0.005), although stable, may lead to slower learning. On the other hand, high LRs (for example, 0.5) allow learning much faster, although they may be unstable. Although the high learning rate can speed up learning, it can also lead to instability or overshooting, since many of the things known by the agent are being changed. Medium values of LRs (for instance, 0.1) seem to be optimal, as they facilitate quick learning.

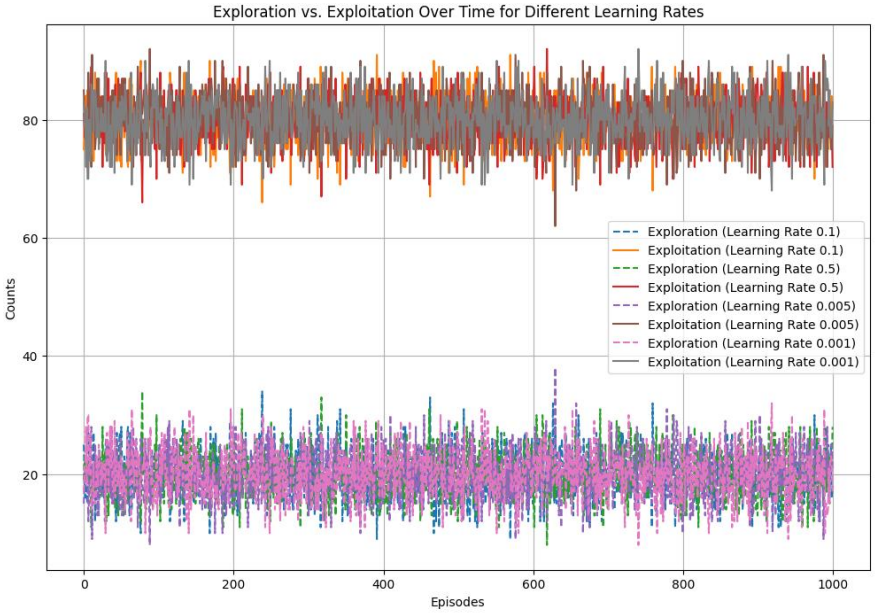

Figure 19. Exploration vs. exploitation over time for different learning rates.

As illustrated in Figure 19, the RL model maintains an appropriate balance between exploration and exploitation in 1000 episodes. The vertical axis depicts either the amount of exploration or exploitation, while the horizontal axis represents the number of episodes. There are several trends regarding exploration and exploitation, and there are various LRs (0.1, 0.5, 0.005, and 0.001). If the learning rate is equal to 0.1, the shift from exploration to exploitation is smooth, indicating proper learning. If the learning rate is equal to 0.5, it may quickly jump to exploitation, implying that it misses out on optimal actions since there was not enough exploration. For LRs of 0.005 and 0.001, exploration will be more favored, meaning that it would take longer to achieve exploitation. Both

phases work well with a moderate learning rate (like 0.1), but extreme rates (like 0.5 or 0.001) can lead to bad behavior, like exploiting too soon or exploring for too long. Overall, the experimental results demonstrate that the proposed reinforcement learning framework successfully learns adaptive intervention policies that stabilize over time and effectively personalize learning strategies based on disability type and engagement level.

Table 6. Comparative analysis of the proposed reinforcement learning approach with baseline methods in terms of adaptivity, personalization, performance, and stability.

Method	Adaptivity	Personalization	Performance	Stability	Key limitation
Rule-based approach	Low	Low	Low	High	Fixed rules
Static recommendation	None	Low	Low	Very high	No adaptation
Supervised learning	Medium	Medium	Medium	Medium	Limited flexibility
Proposed RL method	High	High	High	High	Training required

To evaluate the efficiency of the proposed technique more effectively, a comparison is made between it and a few other existing techniques in Table 6. The results show that the rule-based technique, along with other static techniques, faces restrictions because of being constrained within its methodology and lacking the ability to adapt according to the different requirements and engagement levels of the learners. Although supervised learning techniques are relatively better compared to these techniques due to partial personalization features, they lack the capability of adapting themselves once they have been trained. This allows it to provide more flexible and relevant interventions for different situations. Overall, the comparison highlights that the main advantage of the proposed approach is its ability to adapt over time, making it more suitable for supporting diverse learners in inclusive educational settings.

5. Conclusions

Each learner has a unique set of skills, knowledge, and learning capacities that they bring to the learning process. Adaptive e-learning platforms seek to match learners with the best courses possible depending on their knowledge and abilities. E-learning platforms for online learning have become necessary to the educational process. Institutions and academics hunt for fresh, ideal techniques and strategies that might enhance PL strategies for students with disabilities. The integration of technology in education has opened new avenues for PL, especially for students with disabilities. In this study, we propose PLS-SD using Q-learning. We address various disability types, such as visual, cognitive, motor, and hearing impairments, to recommend PL actions like augmented reality, audio instructions, and text-based resources. In essence, a reward scheme that focuses on practical achievements in terms of engagement and task-completion ratios facilitates learning. Technically, the learning process in the model shows stability during training sessions, and better convergence is seen at a moderate learning rate, such as LR = 0.1. It was also noted that the choice of a learning rate affects performance: a high learning rate creates variability, while a low learning rate slows down convergence. In addition, it was observed that the policy reaches its peak after enough number of episodes, demonstrating the efficacy of Q-learning in recognizing a consistent preference for action across various disability-engagement states. In conclusion, "augmented reality" proved to be a consistent winner among actions for a majority of states, especially when it came to hearing disabilities and general disabilities. However,

"interactive videos" also worked very well, thus proving the versatility of this type of strategy in terms of ensuring medium-to-high levels of engagement. "Personalized exercises" were quite successful in the case of motor impairments. These results highlight that the use of targeted approaches is needed for addressing particular types of disabilities. The efficiency of the proposed model is evaluated through the analysis of success rate, cumulative rewards, and stability of the policy in order to estimate whether this model is capable of supporting not only particular but also general types of disabilities. Despite the positive results, there are some limitations. The results of our experiment were achieved within a simulation setting, where reward values and state transitions are assumed and not based on actual students' behavior. As such, the model can be regarded as a prototype and not an educational strategy to be put into practice. In future studies, we hope to confirm the viability of this model based on actual student interaction data. Moreover, we hope to improve the reward function by receiving expert feedback and conducting pilot experiments.

Author contributions

Theyazn H.H Aldhyani: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing; Samina Amin: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing; Mossab Saud Alholiby: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing; M. Irfan Uddin: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgment

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No **KFU262138].**

Conflict of interest

The authors declare no conflicts of interest.

Ethics declaration

The research was conducted in accordance with accepted ethical standards and guidelines for scientific research and publication.

References

1. Alsolami, A.S., The Effectiveness of Using Artificial Intelligence in Improving Academic Skills of School-Aged Students with Mild Intellectual Disabilities in Saudi Arabia. *Research in Developmental Disabilities*, 2025, 156: 104884. https://doi.org/10.1016/j.ridd.2024.104884

2. Fu, L., The Role of STEM Teachers' Emotional Intelligence and Psychological Well-Being in Predicting Their Artificial Intelligence Literacy. *Acta Psychologica*, 2025, 253: 104708. https://doi.org/10.1016/j.actpsy.2025.104708
3. Zhang, L., Basham, J.D. and Yang, S., Understanding the Implementation of Personalized Learning: A Research Synthesis. *Educational Research Review*, 2020, 31: 100339. https://doi.org/10.1016/j.edurev.2020.100339
4. Hocine, N. and Sehaba, K., A Systematic Review of Online Personalized Systems for the Autonomous Learning of People with Cognitive Disabilities. *Human–Computer Interaction*, 2024, 39 (3–4): 174–205.
5. Nganji, J.T. and Brayshaw, M., Disability-Aware Adaptive and Personalised Learning for Students with Multiple Disabilities. *The International Journal of Information and Learning Technology*, 2017, 34(4): 307–21. https://doi.org/10.1108/IJILT-08-2016-0027
6. Normadhi, N.B.A., Shuib, L., Nasir, H.N.M., Bimba, A., Idris, N. and Balakrishnan, V., Identification of Personal Traits in Adaptive Learning Environment: Systematic Literature Review. *Computers & Education*, 2019, 130: 168–90. https://doi.org/10.1016/j.compedu.2018.11.005
7. Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G. and Vasileva, T., AI-driven Adaptive Learning for Sustainable Educational Transformation. *Sustainable Development*, 2025, 33(2): 1921–1947.
8. Zhang, F., Feng, X. and Wang, Y., Personalized Process–Type Learning Path Recommendation Based on Process Mining and Deep Knowledge Tracing. *Knowledge-Based Systems*, 2024, 303: 112431. https://doi.org/10.1016/j.knosys.2024.112431
9. Pliakos, K., Joo, S.H., Park, J.Y., Cornillie, F., Vens, C. and Van den Noortgate, W., Integrating Machine Learning into Item Response Theory for Addressing the Cold Start Problem in Adaptive Learning Systems. *Computers & Education*, 2019, 137: 91–103. https://doi.org/10.1016/j.compedu.2019.04.009
10. Gligorea, I., Cioca, M., Oancea, R., Gorski, A.T., Gorski, H. and Tudorache, P., Adaptive Learning Using Artificial Intelligence in E-Learning: A Literature Review. *Education Sciences*, 2023, 13(12): 1216.
11. Hung, Y.H., Chang, R.I. and Lin, C.F., Hybrid Learning Style Identification and Developing Adaptive Problem-Solving Learning Activities. *Computers in Human Behavior*, 2016, 55: 552–61. https://doi.org/10.1016/j.chb.2015.07.004
12. Hills, K., Andersen, K. and Davidson, S., Personalized Learning and Teaching Approaches to Meet Diverse Needs: A Prototype Tertiary Education Program. *Reimagining Christian Education: Cultivating Transformative Approaches*, 2018, 233–57. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-0851-2_16
13. Amin, S., Uddin, M.I., Alarood, A.A., Mashwani, W.K., Alzahrani, A. and Alzahrani, A.O., Smart E-Learning Framework For Personalized Adaptive Learning and Sequential Path Recommendations Using Reinforcement Learning. *IEEE Access*, 2013, 11: 89769–90. https://doi.org/10.1109/ACCESS.2023.3305584
14. Amin, S., Uddin, M.I., Alarood, A.A., Mashwani, W.K., Alzahrani, A.O. and Alzahrani, H.A., An Adaptable and Personalized Framework for Top-N Course Recommendations in Online Learning. *Scientific Reports*, 2024, 14(1): 10382. https://doi.org/10.1038/s41598-024-56497-1

15. Essa, S.G., Celik, T. and Human-Hendricks, N.E., Personalized Adaptive Learning Technologies Based on Machine Learning Techniques to Identify Learning Styles: A Systematic Literature Review. *IEEE Access,* 2023, 11: 48392–409.
16. Isabona, J., Imoize, A.L. and Kim, Y., Machine Learning-Based Boosted Regression Ensemble Combined with Hyperparameter Tuning for Optimal Adaptive Learning. *Sensors,* 2022, 22(10): 3776.
17. Sutton, R.S. and Barto, A.G., *Reinforcement Learning : An Introduction*, 2nd ed. The MIT Press, 2018.
18. Müller, H., Berg, L. and Kudenko, D., Using Incomplete and Incorrect Plans to Shape Reinforcement Learning in Long-Sequence Sparse-Reward Tasks. *Neural Computing and Applications,* 2025, 37(23): 18851–66. https://doi.org/10.1007/s00521-024-10615-2
19. Dutt, S., Ahuja, N.J. and Kumar, M., An Intelligent Tutoring System Architecture Based on Fuzzy Neural Network (FNN) for Special Education of Learning Disabled Learners. *Education and Information Technologies,* 2022, 27(2): 2613–33.
20. Minoofam, S.A.H., Bastanfard, A. and Keyvanpour, M.R., RALF: An Adaptive Reinforcement Learning Framework for Teaching Dyslexic Students. *Multimedia Tools and Applications*, 2022, 81(5): 6389–6412.
21. Modak, M.M., Gharpure, P. and M, S., Adaptive Learning and Correlative Assessment of Differential Usage Patterns for Students With-or-without Learning Disabilities via Learning Analytics. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2023, 22(12): 1–25.
22. Khabbaz, A.H., Pouyan, A., Fateh, M. and Abolghasemi, V., An Adaptive Learning Game for Autistic Children Using Reinforcement Learning and Fuzzy Logic. *Journal of AI and Data Mining,* 2019, 7(2): 321–29.
23. Ma, D., Zhu, H., Liao, S., Chen, Y., Liu, J., Tian, F., et al., Learning Path Recommendation with Multi-Behavior User Modeling and Cascading Deep Q Networks. *Knowledge-Based Systems*, 2024, 294: 111743. https://doi.org/10.1016/j.knosys.2024.111743
24. Liu, Z., Hou, J., Ning, D., Zhou, C., Liang, G. and Zhang, F., Improving Deep Q Network Based on Marketing Psychology for AUV Path Planning in Unknown Marine Environments. *IEEE Internet of Things Journal*, 2024, 12(5): 5476–5487. https://doi.org/10.1109/JIOT.2024.3487129
25. Shawky, D. and Badawi, A., Towards a Personalized Learning Experience Using Reinforcement Learning. *Machine Learning Paradigms: Theory and Application*, 2018, 169–87. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-02357-7_8
26. Islam, M.Z., Ali, R., Haider, A., Islam, M.Z. and Kim, H.S., PAKES: A Reinforcement Learning-Based Personalized Adaptability Knowledge Extraction Strategy for Adaptive Learning Systems. *IEEE Access,* 2021, 9: 155123–37. https://doi.org/10.1109/ACCESS.2021.3128578
27. Yuhana, U.L., Djunaidy, A. and Purnomo, M.H., Enhancing Students Performance through Dynamic Personalized Learning Path Using Ant Colony and Item Response Theory (ACOIRT). *Computers and Education: Artificial Intelligence,* 2024, 7: 100280. https://doi.org/10.1016/j.caeai.2024.100280
28. Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D. and Demir, I., Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education. *Information,* 2024, 15(10): 596.

29. Demartini, C.G., Sciascia, L., Bosso, A. and Manuri, F., Artificial Intelligence Bringing Improvements to Adaptive Learning in Education: A Case Study. *Sustainability*, 2024, 16(3): 1347.
30. Khoso, A.K., Honggang, W. and Darazi, M.A., Trust and attitude towards AI as pathways to creativity: a TAM Model study of EFL students' digital literacy and AI acceptance. *Humanit Soc Sci Commun*, 2026, 13: 69. https://doi.org/10.1057/s41599-025-06362-x
31. Khoso, A.K., Honggang, W., Tahir, S.R., Nusrat, A. and Younas, M., Integrating Sustainable Development Goals (SDGs) and Generative AI to Enhance Language Digital Literacy and Creativity in EFL Learning Environments. *J Vis Exp*, 2026, (228): e69445. https://doi.org/10.3791/69445
32. Khoso, A.K., Honggang, W., Younas, M., Salam E.D., DA The Dual Forces of AI: How Generative AI and Perceived AI Dependency Influence Fear of Missing Out (FoMO) and EFL Students' Vocabulary Acquisition. *J Vis Exp*, 2025, 226: e69637. https://doi.org/10.3791/69637
33. Khoso, A.K., Honggang, W. and Darazi, M.A., Empowering creativity and engagement: The impact of generative artificial intelligence usage on Chines EFL students' language learning experience. *Computers in Human Behavior Reports*, 2025, 18: 100627. https://doi.org/10.1016/j.chbr.2025.100627
34. Jang, B., Kim, M., Harerimana, G. and Kim, J.W., Q-Learning Algorithms: A Comprehensive Classification and Applications. *IEEE Access*, 2019, 7: 133653–67.
35. Fan, J., Wang, Z., Xie, Y. and Yang, Z., A Theoretical Analysis of Deep Q-Learning. In *Learning for Dynamics and Control*, 2020, 486–89. PMLR.
36. Spano, S., Cardarilli, G.C., Di Nunzio, L., Fazzolari, R., Giardino, D., Matta, M., et al., An Efficient Hardware Implementation of Reinforcement Learning: The q-Learning Algorithm. *Ieee Access*, 2019, 7: 186340–51.
37. Doshi, K., *Reinforcement Learning Explained Visually*, 2020. Available from: https://towardsdatascience.com/reinforcement-learning-explained-visually-part-4-q-learning-step-by-step-b65efb731d3e

Author's biography

Theyazn H.H Aldhyani: In 2017, he was awarded the Ph.D. degree in Computer Science and Information Technology from NMU University. His areas of research interest are Artificial Intelligence, Machine Learning, Soft Computing, Big Data, Healthcare information, deep learning, cybersecurity, and IoT. He is currently an associate professor in the Faculty of Computer Science and Information Technology at King Faisal University. He has published over 35 research papers in highly reputable journals published by MDPI, Springer, and IEEE. He is a Reviewer in MDPI, Springer, IEEE, and Elsevier.

Samina Amin: A passionate researcher in Computer Science, specializing in Artificial Intelligence (AI) and Machine Learning (ML). She earned his Ph.D. in Computer Science in December 2024 from the Institute of Computing, Kohat University of Science & Technology, Pakistan, where she also completed my Master's degree in 2021. Her research primarily focused on leveraging reinforcement learning (RL) to enhance online learning through intelligent algorithms that

recommend personalized course content.

Mossab Saud Alholiby: Dr. Mossab Saud Alholiby: Associate Professor of Educational Leadership and Executive President of the Applied College at King Faisal University. He received his Ph.D. in Higher Education Management from the University of Glasgow and his Master's degree in Educational Administration from King Faisal University. He has held several academic and administrative leadership positions, including Assistant Vice President for Academic Affairs and Advisor to the Vice President for Academic Affairs at King Faisal University. His research interests include higher education leadership, quality assurance and academic accreditation, strategic planning, digital transformation, organizational effectiveness, and the use of artificial intelligence in improving administrative and educational processes. He has published a number of research papers in the fields of educational leadership, governance, quality in higher education, and organizational development, and has actively contributed to strategic and academic initiatives at both university and national levels.

M. Irfan Uddin: With a solid educational foundation and over two decades of teaching and research experience at renowned academic institutions, he possesses a robust academic and research background across diverse domains of computer science. He is an active member of several esteemed scientific societies, including IEEE, ACM, HiPEAC, CSTA, IAENG, KSS, and Science-i. He has played a leading role in organizing numerous national and international seminars, workshops, and conferences. His research contributions include over 130 research articles published in JCR and Scopus/ISI-indexed journals, as well as national/international conferences, in addition to three authored books published by reputed publishers. He has also published two patents with the United States Patent and Trademark Office. He has served as (PI/Co-PI/Collaborator) in various nationally/internationally funded research projects. Additionally, he actively contributes as a reviewer, editorial board member, and technical program committee member for several prestigious journals and conferences.

AIMS Press