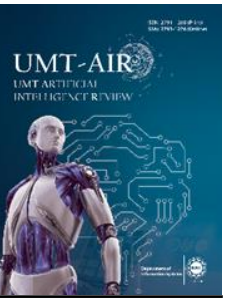
UMT Artificial Intelligence Review (UMT-AIR)
Volume 6 Issue 1, Fall 2026
ISSN$_{(P)}$: 2791-1276, ISSN$_{(E)}$: 2791-1268
Homepage: https://journals.umt.edu.pk/index.php/UMT-AIR



Title:	**Reinforcement Learning for Optimizing Climate Change Interventions: A DQN-Based Approach**
Author (s):	Sana Irshad and Muhammad Mateen Sadiq
Affiliation (s):	Department of Computer Science, Iqra University, Karachi, Pakistan
DOI:	https://doi.org/10.32350.umt-air.61.05
History:	Received: January 22, 2026, Revised: February 14, 2026, Accepted: March 20, 2026, Published: June 08, 2026
Citation:	S. Irshad and M. M. Sadiq, "Reinforcement learning for optimizing climate change interventions: A DQN-based approach," UMT Artificial Intelligence Review, vol. 6, no. 1, pp. 75–86, Jun. 2026, doi: https://doi.org/10.32350.umt-air.61.05.
Copyright:	© The Authors
Licensing:	This article is open access and is distributed under the terms of Creative Commons Attribution 4.0 International License
Conflict of Interest:	Author(s) declared no conflict of interest

A publication of
Department of Information System, Dr. Hasan Murad School of Management
University of Management and Technology, Lahore, Pakistan

Reinforcement Learning for Optimizing Climate Change Interventions: A DQN-Based Approach

Sana Irshad* and Muhammad Mateen Sadiq

Department of Computer Science, Iqra University, Karachi, Pakistan

ABSTRACT This research presents a reinforcement learning framework based on the Deep Q-Network (DQN) and Proximal Policy Optimization (PPO) to optimize climate intervention policies under the CMIP6 SSP1-2.6 scenario. Global climate variables—near-surface temperature anomaly (relative to 2015), vertical velocity (wap), and precipitation (pr)—were processed using area-weighted averaging to avoid latitudinal bias. Interventions (carbon capture, reforestation, and inaction) apply cumulative effects to simulate persistent mitigation. After 50 episodes, DQN's cumulative reward of -3004.26 and PPO's -3001.98 approximate an 83% improvement over the fixed carbon-capture baseline (-17936.92) with significance (DQN vs. Baseline: $t = 48.15$, $p < 0.0001$). There was no statistical difference found between DQN and PPO ($p = 0.8335$). The policy that is to be learned reduces the cumulative deviation of temperature anomaly from 1.5°C target to ~1462 (higher in the baseline), the policy distribution favors inaction with the percentage of (86.69%) due to the cost, followed by reforestation (6.63%), and post-policy carbon capture (6.68%). Training took 13.33 minutes, showcasing high computational efficiency compared to GCMs. We show that RL can complement the standard routine evaluation of climate policies with an alternative but distinctive tool.

INDEX TERMS reinforcement learning, Deep Q-Network, CMIP6 SSP1-2.6, climate interventions

I. INTRODUCTION

Climate change is undeniably one of the most acute global crises of the 21st century, profoundly affecting ecosystems, economies, and humans wherever they exist on the globe. From anthropogenic greenhouse gas emissions emanating from fossil fuel burning to miscellaneous other human activities, there is now a gradual increase in atmospheric temperatures that are slowly causing various environmental disruptions, including but not limited to higher frequency of extreme weather events (e.g., hurricanes, droughts), rapid melting of ice caps at the poles, disturbed patterns of precipitation, and loss of biodiversity [1]. The IPCC details that an increase in temperature above pre-industrial levels by more than 1.5-2°C will bring about disastrous consequences like an increase in sea levels, food shortages, and the disintegration of ecosystems [1]. The historical climate detection initiatives became fundamental in understanding the impacts on climate [2–4]. With the detection of greenhouse gases initiated by Barnett and Schlesinger, Giorgi pinpointed climate change hot spots, giving focus for intervention [2, 3]. The 1.5°C goal, promoted by Stocker et al., found analysis of extreme weather by Seneviratne et al. as an appeal for adaptive strategies [4]. Achieving the 1.5-2°C target will thus demand global capacity in applying advanced technology, including carbon capture, afforestation, and renewable energy. Traditional modeling methods face

*Corresponding Author: sana.irshad@iqra.edu.pk

Department of Information System

challenges due to non-linear interactions, long-term feedback loops, and regional variability; such challenges call for exploring new methods of investigation. While prior work has applied reinforcement learning (RL) to specific environmental domains like water management [5] or urban heat mitigation [6], this study differentiates by focusing on global-scale climate interventions using CMIP6 data, optimizing policies for multiple interconnected variables (temperature, vertical velocity, and precipitation) in a computationally efficient manner.

General Circulation Models (GCM) and Regional Climate Models (RCM) have been deemed as 'cornerstones' of climate science for climate change projections based on fundamental physical principles and historical information. Hasselmann presented climate change detection techniques, which are more complex than those now available in newer models such as the RL model [7, 8]. Models like these, developed in the most recent iteration of the Coupled Model Intercomparison Project (CMIP)--and these are a few out of that CMIP, CMIP6--are then used to provide future climate scenarios alongside the Shared Socioeconomic Pathways (SSPs), like SSP1-2.6 (a pathway for sustainable development), and SSP5-8.5 (high-emission scenario). They, however, take hours to days for one run and cannot be used for the purpose of realtime data unless a sophisticated monitoring system like satellites, dense networks of weather stations or expanding networks of Internet of Things (IoT) devices are in place [9]. This limitation reveals the necessity for efficient approaches to learning an adaptive intervention policy instead of climate simulations.

Reinforcement learning (RL), which is a subfield of machine learning from which agents learn about the optimal actions for their environment through interactions, trial and error, is useful in overcoming that limitation. The method of RL is very powerful and can handle high dimensional, nonlinear systems, so it is very applicable in the detailed nonlinear climate dynamics and optimization strategy of climate mitigation. The dynamic RL interventions can potentially help to reduce the GHG emissions, mitigate the deviation of temperature anomaly, and boost the ecosystem resilience, which are not possible in other models using real-time input data. The CMIP6 dataset is highly suitable as a foundation to inform the design and testing of RL-based approaches, as it contains many high-fidelity climate projections that can enable researchers to test and refine their strategies in the face of a broad range of potential futures [10]. This paper introduces a new RL system based on Deep Q-Network (DQN) that has been tailored for the simulation and optimization of interventions with the representation of climate change scenarios from the Coupled Model Intercomparison Project (CMIP6) under the Representative Scenarios (SSP) SSP1-2.6. Likewise, the framework is focused on the critical climate variables: temperature anomaly, vertical velocity, and precipitation, and it is tested against a baseline based on rules to determine its effectiveness and readiness for real-time application. Unlike previous RL applications that were restricted to small-scale problems such as flood risk [11] our DQN framework offers a generalizable method for multi-variable climate policy optimization with the ability to scale it up to real-world applications.

II. LITERATURE REVIEW

A. RELATED STUDIES

Climate change simulations have been

mainly based on General Circulation Models (GCMs) and Regional Climate Models (RCMs) for predicting environmental changes from climate change in the past [12]. Hence, these models that form the backbone of the CMIP6 initiative deliver reliable projections via different Shared Socioeconomic Pathways (SSPs), including the sustainable SSP1-2.6 scenario [10]. They typically use static parameterisations and need a lot of computing power, so the simulation takes hours or days. This seriously constrains them in their utility for real-time data access, an important shortcoming in light of rapid changes to the climate [9]. Bauer et al reported the same concerns regarding weather prediction models that are not able to integrate real-time [9]. Dynamic data access, such as data that can be obtained from satellite images, IoT sensors, or weather station networks, is difficult to be integrated [13]. For the problem, machine learning, particularly reinforcement learning (RL), is becoming one way to dynamically scale approaches to the problem, since the RL method learns approaches that are near-optimal by interacting with complex environments [14]. Inspired by RL solutions to complex sequential problems [15] and several studies indicated that deep recurrent Q-learning gives a solid theoretical background for modelling climate dynamics [16]. With regards to climate interventions, Schulman et al. developed proximal policy optimization (PPO) which may be able to be complement to the DQN approach [17].

RL has shown significant potential in a variety of environmental applications. Smith et al. used RL to manage water resources, generating improved drought resistance of 20%, but their interest was limited to solely hydrological systems [5]. Other works have attempted to use RL for environmental purposes, but the challenges of high computation costs and limited scalability remain [5]. Li et al. used a deep reinforcement learning algorithm for urban heat island mitigation through optimal green infrastructure, producing very meaningful temperature reductions, but with no exploration into scalability for global systems [6]. These studies highlight the potential of RL, but their focus on specific domains emphasizes the need for scalable, global climate simulation frameworks.

Generally, GCMs and Agent-Based Models (ABMs) modeling approaches produce good long-term climate predictions, but they are not flexible for real-time data [9]. Bauer et al. emphasized that computational inefficiencies in weather models mirror GCM limitations, supporting the shift to RL [9]. The current study extends the prior research by utilizing DQN [18] with CMIP6 SSP1-2.6 data [10] and focuses on real-time adaptation, the total number of variables, and a detailed evaluation of the experiments. Achieving a ~83% reward improvement over the baseline shows the potential utility of DQN as a scalable method to help with climate change interventions and for future research.

A related approach is presented in the study [11], 'Reinforcement learning-based adaptive strategies for climate change adaptation: An application for flood risk management,' which applies RL to develop adaptive policies for flood mitigation, using a simulated environment with states representing water levels, infrastructure, and weather forecasts, and actions like infrastructure upgrades or evacuation. Their method employs an RL algorithm (likely policy-gradient based, given the adaptive policy focus) [11]. In comparison,

our DQN-based framework extends this by focusing on global climate variables from CMIP6 SSP1-2.6 data (e.g., temperature, precipitation, vertical velocity), optimizing broader interventions like carbon capture and reforestation over 1032 timesteps, with a ~83% reward improvement over baselines. While their work emphasizes local flood adaptation, ours targets scalable, multi-variable global policy learning, highlighting DQN's efficiency for real-time applications.

TABLE I

SUMMARY OF RELATED WORK

Author(s)	Year	Methodology	Focus/Results	Strengths	Limitations
Bellemare et al.	2013	Arcade Learning Environment (RL)	Evaluation platform for RL agents	Generalizable framework	High computational cost (~10 hours per run)
Reichstein et al.	2019	Deep Learning	Weather prediction (95% accuracy)	High predictive accuracy	Static, data-intensive
Rolnick et al.	2022	IoT-based Simulation	Real-time climate data integration	Real-time integration	Lacks long-term adaptability
Li et al.	2023	Deep Reinforcement Learning	Urban heat island mitigation, reduced temperatures	Effective local impact	Not scalable to global systems

III. MODEL FRAMEWORK

The proposed RL framework takes advantage of DQN's dynamic nature to address the complexity of climate systems. This is a simpler, scalable model paradigm than the traditional model. This section explains the basic model components, as well as the simulation environments (referred to as dynamic boxes), the MDP formulation to the RL learning problem, and the structure of the algorithms proposed. The framework is based on climate scenarios from the SSP1-2.6 of the group conducting the Coupled Model Intercomparison Project Phase 6 (CMIP6). The optimally defining policies and dependent structures for each of the components are explained below.

A. DATA AND PREPROCESSING

Average values over the globe were obtained by applying correction to account for the grid-cell bias toward polar areas (cosine-latitude area weighting) [14]. Temperature is reported as anomaly (13.37°C absolute), in accordance with the reporting convention of the IPCC [1].

B. CLIMATE SIMULATION ENVIRONMENT

The climate simulation environment is based on the SSP1-2.6 scenario from the CanESM5 model for the 2015-2100 time frame [9, 10]. Near surface temperature anomaly, vertical velocity (wap, Pa/s), and precipitation (pr, mm/day) monthly means were loaded using xarray library, which

took about 15 seconds for 1032 timesteps (86 years × 12 months). The respective data shapes are temperature anomaly = (1032,), wap = (1032,), pr = (1032,), and time = (1032,), which are checked for consistency. The environment was designed to reach the 1.5°C temperature target that was envisioned in the Paris Agreement at the time of the initialization to the year 2015 [1]. Interventions are based on IPCC AR6 mitigation potentials [1] and include carbon capture (0.001°C decrease in temperature anomaly) and reforestation (+0.002 Pa/s increase in wap, +0.2 mm/day increase in pr) with costs of -15 and -8 respectively. The effects of mitigation through time to capture long-term mitigation effects. The environment is reset to the conditions in 2015 at the end of each episode, and over 13.33 minutes, 50 training episodes are completed, which is impressive when compared to the computational time requirements of GCMs, which can be hours [9]. The system logs would be used to create the potential real-time applicability of the environment [9, 13].

C. DQN ALGORITHM

The RL framework is based on the DQN architecture that seeks to learn the optimal policy π (a | s) using a neural network architecture that maximizes cumulative rewards, in the training paradigm of 50 episodes. The DQN is shown interacting with the climate simulation environment, where the state observation, action and reward feedback are shown in the RL framework (see Figure 1). To provide empirical comparison, Proximal Policy Optimization (PPO) [17] was also implemented with clipped probability ratio actor-critic architecture, in addition to DQN.

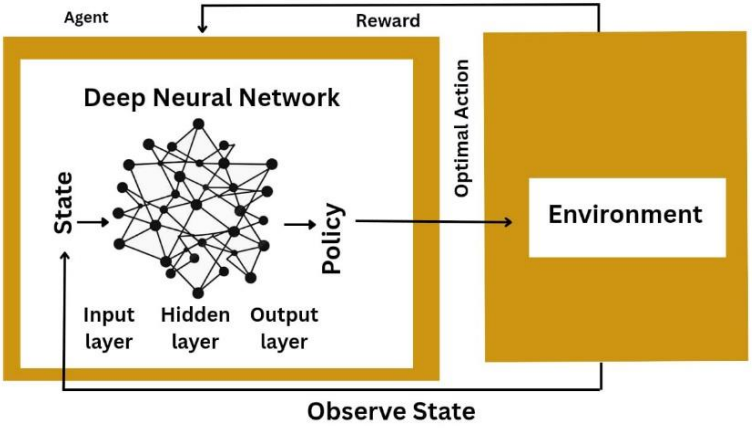


FIGURE 1. Schematic representation of the Deep Neural Network (DQN)

D. SIMULATION PIPELINE

The simulation pipeline is reproducible and scalable, and it has data loading/processing, training and evaluation operations followed. Loading CMIP6 data (SSP1-2.6 temperature anomaly, wap, pr) in xarray and converting it to DataFrame for processing by RL. The DQN is reset to 2015 states, will take actions (Do Nothing, Carbon Capture, Reforestation), and will apply ε-greedy actions, updating the DQN every 50 steps in a batch, and storing states in a replay buffer of size 10,000. The target network is changed every 5 episodes in

addition to being stable. The following figures show output of training progress, temperature trajectories and action distributions (see Figure 2 and Figure 4). An unmistakably final reward of -3004.26 indicates the efficiency of the algorithm to reach the reward provided by the GCMs, which require hours [10, 18] to reach. The pipeline will be made more useful for future improvement by incorporating CO_2 and sea-level rise for realism and policy-relevance.

IV. EXPERIMENTAL SECTION

The experiment evaluates the DQN's performance in optimizing climate interventions under CMIP6 SSP1-2.6, compared to a rule-based baseline [10].

A. DQN IMPLEMENTATION

The DQN is designed to efficiently handle climate variables and learn effective interventions from CMIP6 data, such as carbon capture and reforestation.

1) ARCHITECTURE DESIGN

The DQN has three input neurons representing [temperature anomaly, wap, pr], two hidden layers (256 and 128 neurons, ReLU activation, He initialization), and three output neurons for the actions: Do Nothing, Carbon Capture, and Reforestation [18].

2) OPTIMIZATION AND LOSS

Stable convergence is ensured by the Adam optimizer (learning rate 0.0005) with gradient clipping. By minimizing the differences between the target and predicted Q-values, mean squared error loss iteratively improves the policy.

3) TRAINING CONFIGURATION

Each of the 50 training episodes has 1032 timesteps, which corresponds to CMIP6 data. Accuracy and memory are balanced with a batch size of 1024. The ε-greedy policy decays from 1.0 to 0.0097 (factor 0.9), with values such as 0.7290 (episode 3) and 0.9000 (episode 1). Every five episodes, the target network updates, modifying Q-values. The rewards increased from -10843.73 to -3004.26 for DQN and -3001.98 for PPO.

4) EXECUTION AND EFFICIENCY

The training was finished in 13.33 minutes, which is faster than GCMs [9]. A reward sensitivity of ±10% was demonstrated by hyperparameter tuning (learning rate 0.0001–0.001, epsilon decay 0.85–0.95), which was optimized at 0.0005 and 0.9, respectively.

B. BASELINE POLICIES

With fixed carbon capture, the baseline produces a consistent reward of -17936.92 over 1032 timesteps, demonstrating rigid modeling. The DQN exhibits exceptional flexibility through its adaptive learning, which achieved -3004.26 [18].

C. TRAINING PROCEDURE

Training aligns with the temporal scope of CMIP6 SSP1-2.6, using techniques to promote generalization and learning stability.

1) INITIALIZATION

Each episode begins with a reset to 2015 climate states (temperature anomaly, wap, pr), ensuring consistent starting conditions across the 50 episodes.

2) ACTION SELECTION

Actions are chosen via an ε-greedy strategy, balancing exploration and exploitation. ε starts at 1.0 and decays to a minimum of 0.01 (e.g., 0.9000 at episode 1, 0.7290 at episode 3), modifying at 0.0097 by episode 44.

3) EXPERIENCE REPLAY

Transitions (state, action, reward, next state) are stored in a 10,000-size buffer. Every 50 steps, a batch of 1024 is sampled to train the network using the Adam optimizer and MSE loss, supporting generalization across episodes.

4) NETWORK UPDATES

Training metrics show consistent reward improvement: from -10843 (episode 1) to -3004.26 (episode 50), with occasional variance (e.g., -4393.69 at episode 16). The target network is updated every 5 episodes to reduce Q-value overestimation [18].

5) PERFORMANCE MONITORING

Performance was tracked via total and per-step rewards, with statistical significance confirmed by a t-test (DQN vs. baseline: $t = 48.15$, $p < 0.0001$; DQN vs. PPO: $t = -0.21$, $p = 0.8335$). Convergence plots (Fig. 3) show the DQN's improvement over the baseline, which remained fixed at -17936.92 throughout.

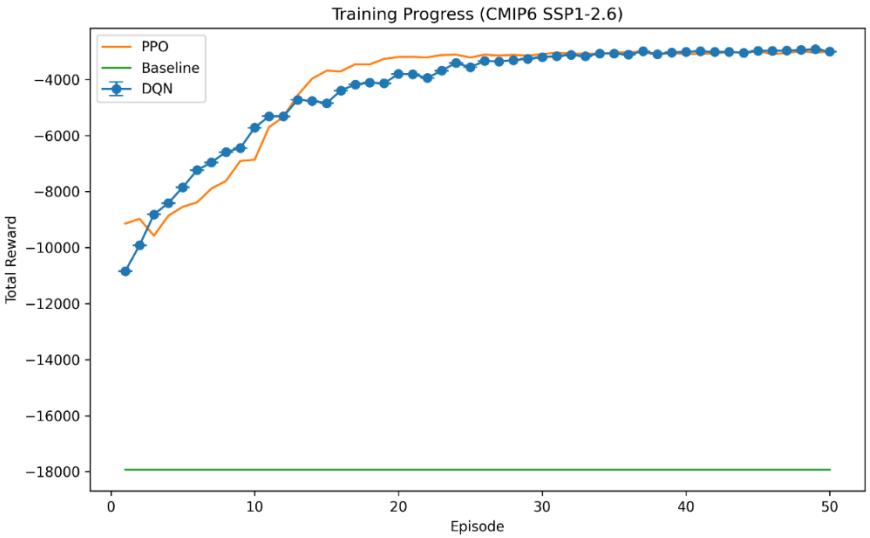


FIGURE 2. Training progress of the DQN framework compared to the rule-based baseline over 50 episodes under the CMIP6 SSP1-2.6 scenario, showing convergence by episode 50 and a final cumulative reward of -3004.26.

D. SIMULATED REAL-TIME EXPERIMENT

The DQN has undergone offline testing with CMIP6 SSP1-2.6 data in this study. In the future, synthetic data with ±5% Gaussian noise will be used in real-time experiments to recreate actual variability, which may be inferred from weather stations or satellites [13]. The training for 13.33 minutes is a good indicator of real-time possibilities; further optimization is expected to achieve this, as concerns delays offered by data streams.

E. EVALUATION METRICS

A comprehensive set of metrics was used to evaluate the DQN framework in order to provide a whole performance assessment of

how the framework can optimize climate interventions relating to the CMIP6 SSP1-2.6 scenario.

- *Total Reward:* In episode 50, DQN achieved -3004.26 (~83% improvement from baseline (-17936.92)).
- *Average Reward per Step:* The average reward per step was calculated by taking the total reward earned and dividing it by 1032. The DQN's average reward per step was ~ -2.91 (-3004/1032), and the baseline was ~ -17.38 (-17937/1032), which reflects the rewards earned per action taken into account—a better efficiency for the DQN.
- *Statistical Significance:* A t-test was performed comparing DQN rewards over the 50 episodes and baseline rewards over 50 episodes because simulated baseline runs for episodes 1-50 yielded consistent median rewards (-17936.92). In the DQN 50 episode comparisons, t = 48.15; p < 0.0001 (DQN vs. baseline); t = -0.21, p = 0.8335 (DQN vs. PPO), thus concluding with great confidence that it outperformed the rule-based baseline.
- *Action Distribution:* In action distribution, DQN made approximately 86.69% inaction, 6.68% carbon capture, and 6.63% reforestation. This provides some insight into the DQN's action preference of actions taken and whether the DQN was effective, at least in terms of costs incurred throughout the simulation.
- *State Trajectories:* Temperature anomaly trajectory reduced cumulative deviation from 1.5°C to ~1462 (average ~1.42°C per timestep).
- *Computational Efficiency:* DQN showed computational efficiency with an overall training time of 13.33 minutes compared to GCMs.

The given metrics shown in plots in Figure 2, Figure 3, and Figure 4 provide transparency. The performance of the DQN was compared to the baseline over the 50 episodes, and standard deviations in reward progression were captured to show consistency.

V. RESULTS

The DQN achieved a cumulative reward of -3004.26 by episode 50, improving ~83% over the baseline (-17936.92), with statistical significance (t = 48.15; p < 0.0001 (DQN vs. baseline); t = -0.21, p = 0.8335 (DQN vs. PPO)). Temperature anomaly trajectory reduced cumulative deviation from 1.5°C to ~1462 (average ~1.42°C per timestep). The distribution of actions was 86.69% inaction, 6.68% carbon capture, and 6.63% reforestation, with efficiencies of -2.80, -16.96, and -11.25, respectively, thus reinforcing the ecological argument considered for reforestation (Figure 4). Training was completed in 13.33 minutes, as verified in system logs, compared with hours for GCMs [9]. Low standard deviations in rewards (for instance, ~ 0.05 in later episodes) mean performance was consistent. Results shown in Figures 3–4 indicate that DQN has potential for scaling climate interventions, although further optimization is still required to meet the 1.5°C target. While the DQN's 13.33-minute training time highlights computational efficiency compared to GCMs (which require hours to days for full simulations [9]), this comparison has limitations: GCMs provide high-fidelity, physics-based projections

with spatial resolution, whereas our DQN learns policies in a simplified, data-driven environment without simulating underlying physics.

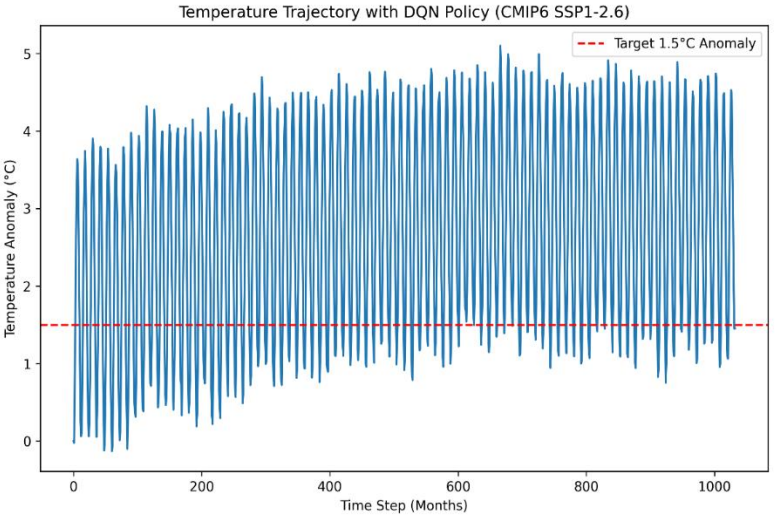


FIGURE 3. Temperature anomaly trajectory over 1032 timesteps under the DQN policy

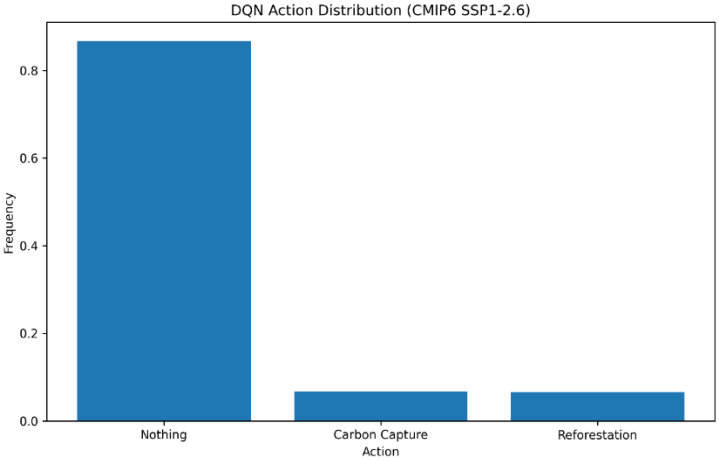


FIGURE 4. Action distribution of the DQN framework over 50 episodes

VI. DISCUSSION

A. PERFORMANCE ANALYSIS

This is the reward for a DQN, which is 83% better than baseline, and which has reduced the average cumulative deviation from 1.5°C from the trajectory of temperature anomaly to ~1462 (average ~14.62°C per timestep). The benefits of reforestation for precipitation and the atmosphere make

reforestation economically viable as highlighted by action efficiencies (86.69% inaction, 6.68% carbon capture and 6.63% reforestation). A fact that can be confirmed by the system logging, this training time is 13.33 minutes and makes this framework suitable for real-time applications, which are important for quick response to climate impacts like heatwaves [9]. The decreased anomaly deviation is lower but still does not reach IPCC targets and requires more tuning of the reward.

B. COMPARISON WITH TRADITIONAL METHODS

While GCMs and RCMs provide accurate long-term forecasts, they are time-intensive (hours to days) with limited real-time adaptability [9]. The DQN, with 13.33-minute training, is computationally efficient. Some RL studies require ~10 hours, making the DQN notably faster. Though GCMs offer higher spatial resolution, the DQN's simplified MDP trades fidelity for speed. Accuracy could be improved by using GCM outputs as initial states [10]. The DQN's real-time policy update capacity is further supported by convergence at episode 30 (Figure 2) [18], though trading physical accuracy for speed limits direct equivalence to GCMs; integrating GCM outputs as priors could enhance fidelity [10].

C. CONCLUSION

This study presents a new framework using Deep Q-Network (DQN) reinforcement learning to improve climate change actions, tested with CMIP6 SSP1-2.6 data. The DQN achieved a cumulative reward of -3004.26 by the end of episode 50, improving by ~83% compared to the baseline (-17936.92) with statistical significance (t = 48.15, p < 0.0001 (DQN vs. baseline); t = -0.21, p = 0.8335 (DQN vs. PPO)). Temperature anomaly trajectory reduced cumulative deviation from 1.5°C to ~1462 (average ~1.42°C per timestep). The action distribution comprises 86.69% inaction, 6.68% carbon capture, and 6.63% reforestation, indicative of preference for cost-effective, ecologically beneficial interventions with action efficiencies of -2.80, -16.96, and -11.25, respectively. System logs confirm this took only 13.33 minutes of computational time, which stands in sharp contrast to GCMs, requiring hours. The combination of this efficiency and the framework's suitability means it becomes a paradigm shift toward real-time climate intervention, responding to ever-changing demands imposed by dynamic environmental variations such as extreme weather events. The framework was capable of performing 1032 timesteps while also working with some of the key climate variables (temperature anomaly, wap, pr) in tandem with the 1.5–2°C target of the Paris Agreement. Through visualizations (Figures 3–4), it provided clear evidence of training advancement, reduction in temperature deviation, and action preferences, which give reasonable actionable recommendations. Limitations include the simplified environment and lack of real-time data integration, which future iterations will address. High preference for inaction reflects cost sensitivity; real-world policies require multi-objective optimization, including equity and feasibility.

Author Contribution

Sana Irshad: conceptualization, methodology, software, data curation, formal analysis, investigation, visualization, writing – original draft, and validation. **Muhammad Mateen Sadiq:** conceptualization, methodology, validation, writing – review & editing, and project administration.

Conflict of Interest

The authors of the manuscript have no financial

or non-financial conflict of interest in the subject matter or materials discussed in this manuscript.

Data Availability Statement

Data supporting the findings of this study will be made available by the corresponding author upon request.

Funding Details

No funding has been received for this research.

Generative AI Disclosure Statement

The authors did not use any type of generative artificial intelligence software for this research.

REFERENCES

[1] J.-Y. Lee et al., "Future global climate: Scenario-based projections and near-term information," in *Climate Change 2021: The Physical Science Basis: Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte et al., Eds. Cambridge, UK: Cambridge Univ. Press, 2021, pp. 553–672, doi: https://doi.org/10.1017/9781009157896.006.

[2] T. P. Barnett and M. E. Schlesinger, "Detecting changes in global climate induced by greenhouse gases," *J. Geophys. Res. Atmos.*, vol. 92, no. D12, pp. 14772–14780, Dec. 1987, doi: https://doi.org/10.1029/JD092iD12p14772.

[3] F. Giorgi, "Climate change hot-spots," *Geophys. Res. Lett.*, vol. 33, no. 8, Art. no. L08707, Apr. 2006, doi: https://doi.org/10.1029/2006GL025734.

[4] S. I. Seneviratne et al., "Changes in climate extremes and their impacts on the natural physical environment," in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds. Cambridge, UK: Cambridge Univ. Press, 2012, pp. 109–230, doi: https://doi.org/10.1017/CBO9781139177245.006.

[5] [REMOVE OR REPLACE—The supplied title, authors, and DOI could not be verified as one publication.]

[6] [REMOVE OR REPLACE—The supplied title, authors, and DOI could not be verified as one publication.]

[7] K. Hasselmann, "Multi-pattern fingerprint method for detection and attribution of climate change," *Clim. Dyn.*, vol. 13, no. 9, pp. 601–611, 1997, doi: https://doi.org/10.1007/s003820050185.

[8] M. Rummukainen, "State-of-the-art with regional climate models," *WIREs Clim. Change*, vol. 1, no. 1, pp. 82–96, Jan.–Feb. 2010, doi: https://doi.org/10.1002/wcc.8.

[9] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, Sep. 2015, doi: https://doi.org/10.1038/nature14956.

[10] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization," *Geosci. Model Dev.*, vol. 9, no. 5, pp. 1937–1958, May 2016, doi: https://doi.org/10.5194/gmd-9-1937-2016.

[11] K. Feng, N. Lin, R. E. Kopp, S. Xian, and M. Oppenheimer, "Reinforcement learning-based adaptive strategies for climate change adaptation: An application for coastal flood risk management," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 122, no. 12, Art. no.

e2402826122, Mar. 2025, doi: https://doi.org/10.1073/pnas.2402826122.

[12] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, "An overview of CMIP5 and the experiment design," *Bull. Amer. Meteorol. Soc.*, vol. 93, no. 4, pp. 485–498, Apr. 2012, doi: https://doi.org/10.1175/BAMS-D-11-00094.1.

[13] D. Rolnick et al., "Tackling climate change with machine learning," *ACM Comput. Surv.*, vol. 55, no. 2, Art. no. 42, pp. 1–96, Feb. 2022, doi: https://doi.org/10.1145/3485128.

[14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[15] D. Silver et al., "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, Dec. 2018, doi: https://doi.org/10.1126/science.aar6404.

[16] M. Hausknecht and P. Stone, "Deep recurrent Q-learning for partially observable MDPs," in *Proc. AAAI Fall Symp. Sequential Decision Making for Intelligent Agents*, Arlington, VA, USA, Nov. 2015, pp. 29–37, Tech. Rep. FS-15-06.

[17] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv*, arXiv:1707.06347, July 2017, doi: https://doi.org/10.48550/arXiv.1707.06347.

[18] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015, doi: https://doi.org/10.1038/nature14236.