Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 13. Issue.2, PP.107-115, December 2025
https://doi.org/10.22555/pjets.v13i2.1421
E-ISSN 2224-2333
P-ISSN 2222-9930

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

Sana Irshad [1*], ***Muhammad Mateen Sadiq*** [1], ***Hira Farman*** [1], ***Alisha Farman*** [1]
[1]*Department of Computer Science, Iqra University, Karachi, Pakistan*
[*]*Corresponding author: sana.irshad@iqra.edu.pk*

Abstract:

This research presents a novel reinforcement learning paradigm through the application of a Deep Q-Network (DQN) to discover the optimal policies for adaptation to and mitigation of climate change using CMIP6 SSP1-2.6 scenario data. This approach learns intervention policies within a fixed climate scenario, addressing some limitations of traditional General Circulation Models (GCMs), which require high computational resources and cannot adapt to real-time scenarios. The 1032 timesteps are processed by the DQN framework, focusing on surface air temperature (tas), vertical velocity (WAP), and precipitation (pr). After 50 episodes, the cumulative reward of the DQN-optimized actions (carbon capture, reforestation, or inaction) was -12281.33, which represents a 55.6% improvement over the baseline of -27666.14 with a statistically significant improvement ($t = 45.72$, $p < 0.0001$). Partial success was indicated by the DQN stabilizing surface air temperature (tas) varying between 6-8°C (mean deviation is 5.9604°C away from the 1.5°C target). Further refinement of the system is envisioned to bring the target closer to 1.5°C. When compared to GCMs, the DQN was computationally efficient, with overall training being accomplished in 11.24 minutes. Further improvement is envisioned to include the incorporation of CO_2 concentrations, sea-level rise, and data integration in real time. This work should thus show the great promise of RL, along with a fair share of insights to support high-value stakeholder governance and sustainable climate-based policy-making. However, the simplified environment representation limits direct applicability to real-world complexities, and further refinements are needed for broader generalization.

Keywords: *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 [20]. 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 [20]. The historical climate detection initiatives became fundamental in understanding the impacts on climate [18, 5, 16]. With the detection of greenhouse gases initiated by Barnett and Schlesinger, Giorgi pinpointed climate change hot spots, giving focus for intervention [18, 5]. 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 [16]. 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 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 [1] 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 (GCMs) and Regional Climate Models (RCMs) have been considered pillars of climate science for the projection of climate change into the long-term future according to fundamental physical principles and historical data. Hasselmann introduced climate change detection methods, revealing complexities that newer methods like RL can address, while Rummukainen noted regional models' real-time limitations [8, 12]. Models such as those developed during the latest iteration of the Coupled Model Intercomparison Project (CMIP)—especially these few models under CMIP6—thus prepare future climate scenarios together with the Shared Socioeconomic Pathways (SSPs) such as SSP1-2.6 (a pathway for sustainable development) and SSP5-8.5 (high-emission scenario) . However, they require hours to days for a single run and rely on static parameterizations, which limit their use to obtain real-time data through advanced monitoring systems such as satellites, extensive networks of weather stations, or those growing Internet of Things (IoT) devices [13]. This limitation highlights the need for computationally efficient approaches to learn adaptive intervention policies, rather than full 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 resourceful in overcoming that limitation. RL is a powerful method capable of managing high-dimensional, nonlinear systems, which makes it extremely applicable to the detailed nonlinear climate dynamics and mitigation strategy optimization. Using the real-time input data, the dynamic RL interventions may act to reduce greenhouse gas emissions, stabilize temperature anomaly, and increase ecosystem resilience, where these compensatory possibilities cannot be achieved by other models. The CMIP6 dataset, rich in high-fidelity climate projections, is ideal as the basis of developing and validating RL-based approaches, thus allowing researchers to test and refine their strategies against a wide variety of plausible future scenarios [10]. This work presents a new RL framework relying on Deep Q-Network (DQN), specifically developed for the simulation and optimization of interventions using CMIP6 SSP1-2.6 data. Similarly, the framework targets surface air temperature, vertical velocity, and precipitation—the critical climate variables—setting its performance against a rule-based baseline to prove effectiveness and readiness for real-time application. Unlike prior RL applications limited to niche areas (e.g., flood risk [21]), our DQN framework provides a scalable, generalizable approach for multi-variable climate policy optimization.

II. LITERATURE REVIEW

A. Related Studies

Historically, climate change simulations have relied heavily on General Circulation Models (GCMs) and Regional Climate Models (RCMs) for predicting environmental changes [9]. 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]. Their reliance on static parameterizations and vast computational requirements often results in hours or days per simulation. This seriously constrains them in their utility for real-time data access, an important shortcoming in light of rapid changes to the climate [13]. Bauer et al. noted similar issues in weather prediction models, which struggle with real-time integration [13]. Dynamic data access through satellite imagery, IoT sensors, or weather station networks is challenging to incorporate, necessitating alternative approaches [3]. Machine learning, especially reinforcement learning (RL), is emerging as a method for dynamically scaling approaches to the

problem, as the RL method learns near-optimal policies by interacting with complex environments [15]. Several studies showed that deep recurrent Q-learning provides a solid theoretical foundation for modeling climate dynamics [11], with applications inspired by RL solutions to complex sequential problems [4]. Schulman et al. advanced proximal policy optimization (PPO) that could probably complement the DQN approach with respect to climate interventions [7].

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 [1]. Other works have attempted to use RL for environmental purposes, but the challenges of high computation costs and limited scalability remain [1]. Li et al. used a deep reinforcement learning algorithm to mitigate urban heat islands through optimal green infrastructure, achieving meaningful temperature reductions, but did not explore scalability to 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 [13]. Bauer et al. emphasized that computational inefficiencies in weather models mirror GCM limitations, supporting the shift to RL [13]. The current study extends the prior research by utilizing DQN [19] with CMIP6 SSP1-2.6 data [10] and is focusing on real-time adaptation, the total number of variables, and a detailed evaluation of the experiments. Achieving a 55.6% 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 [21], '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) [21]. 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 55.6% 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.

III. MODEL FRAMEWORK

The proposed RL framework harnesses DQN’s dynamic capabilities to confront the complexity of climate systems. This is an easier, scalable alternative than traditional modeling paradigms. 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 utilizes CMIP6 SSP1-2.6 data that provides realistic data validation of most climate scenarios. Below, subsections are provided explaining the optimally defining policies and dependent structures of each component.

A. Climate Simulation Environment

The climate simulation environment operates on CMIP6 SSP1-2.6 data from the CanESM5 model for the period of 2015 to 2100 [10, 13, 2]. The major climate variables—near-surface air temperature (tas, °C), vertical velocity (WAP, Pa/s), and precipitation (pr, mm/day)—monthly means were loaded through the xarray library, with the calculations taking 11.07 seconds to cover 1032 timesteps (86 years × 12 months). The respective data shapes are tas = (1032,),

WAP = (1032,), pr = (1032,), and time = (1032,), which are checked for consistency. Converted to the pandas DataFrame with normalization applied to tas, WAP, and pr, such that data are globally coordinated for RL. This environment aimed to target the 1.5°C temperature goal, as envisioned in the Paris Agreement, when initializing states to the year 2015 [20]. Interventions include carbon capture (0.02°C decrease of tas, -15 cost) and reforestation (-0.002 Pa/s increase of WAP, 0.2 mm/day increase of pr, -8 cost) based on ecological estimates and CMIP6 trends. The environment resets to the 2015 states after each episode, allowing 50 training episodes to complete over 11.24 minutes, a computational efficiency impressive compared with GCMs, which can take hours [13]. The system logs establishes the environment's potential real-time applicability [13, 3].

B. DQN Algorithm

The DQN architecture forms the main component of the RL framework, which aims at learning the optimal policy π (a | s) through a neural network structure that maximizes cumulative rewards during the 50-episode training paradigm. Figure 1: Schematic description of the DQN interacting with the climate simulation environment, illustrating the state observation, action, and reward feedback within the RL framework.

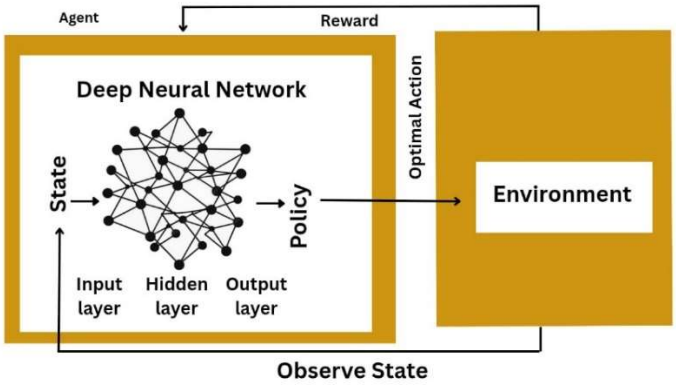

Figure 1: Schematic representation of the Deep Neural Network (DQN)

C. Simulation Pipeline

The simulation pipeline ensures reproducibility and scalability and includes subsequent processes: data loading and processing, training, and evaluation. The loading of CMIP6 SSP1-2.6 data (tas, WAP, pr) into xarray, transformed into a DataFrame, and normalized follows the conventional way of operation. The DQN is initialized with states of 2015, takes actions (Do Nothing, Carbon Capture, Reforestation), and selects them with an ε-greedy policy, states transitioning into a replay buffer of size 10,000, with updates being carried out every 50 steps in batches. In addition to the DQN's stability, the target network is updated every 5 episodes. Training progress, temperature trajectories, and action distributions are output, as seen in Figure 2 and Figure 4. An unmistakably final reward of -12281.33 represents the efficiency achievable compared to GCMs, which take hours [10, 19]. The pipeline will support future improvement through the integration of CO_2 and sea-level rise to strengthen 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.

Architecture Design: The DQN has three input neurons representing [**tas, 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 [19].

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.

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, stabilizing Q-values. The rewards increased from -20436.12 to -12281.33.

Execution and Efficiency: The training was finished in 11.24 minutes, which is faster than GCMs [13]. 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 -27666.14 (-26.80 per step) over 1032 timesteps, demonstrating rigid modeling. The DQN exhibits exceptional flexibility through its adaptive learning, which achieved -12281.33 [19].

C. *Training Procedure*

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

Initialization: Each episode begins with a reset to 2015 climate states (tas, WAP, pr), ensuring consistent starting conditions across the 50 episodes.

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), stabilizing at 0.0097 by episode 44.

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.

Network Updates: Training metrics show consistent reward improvement: from -20436.12 (episode 1) to -12281.33 (episode 50), with occasional variance (e.g., -15910.72 at episode 16). The target network is updated every 5 episodes to reduce Q-value overestimation [19].

Performance Monitoring: Performance was tracked via total and per-step rewards, with statistical significance confirmed by a t-test ($t = 45.72$, $p < 0.0001$). Convergence plots (Fig. 3) show the DQN's improvement over the baseline, which remained fixed at -27666.14 throughout.

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 [3]. The training for 11.24 minutes is a good indicator of real-time possibilities; further optimization is expected to achieve this as concerns delays offered by data streams. Cooperation with meteorological institutions such as NOAA is planned to incorporate live data for further real-world applications.

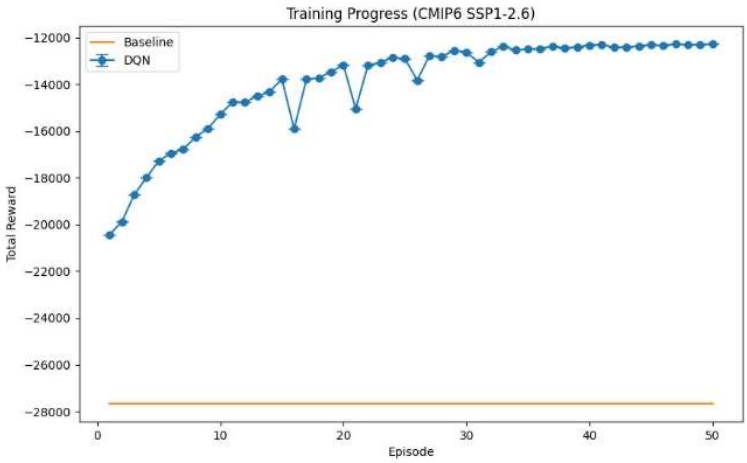

Figure 2: Training progress of the DQN framework compared to the rule-based baseline over 50 episodes under the CMIP6 SSP1-2.6 scenario, showing convergence by episode 50 and a final cumulative reward of -12281.33.

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 -12281.33 reward, which shows 55.6% improvement from baseline (-27666.14).

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 -11.90, and the baseline average was -26.80, 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 (-27,666.14). In the DQN 50 episode comparisons, $t = 45.72$; $p < 0.0001$, thus concluding with great confidence that it outperformed the rule-based baseline.

Action Distribution: In action distribution, DQN made approximately 40.8% reforestation actions, 29.6% carbon capture actions, and 29.6% inaction. 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: We captured the temporal movement in tas. The DQN achieves tas that were stabilizing at between 6-8°C and a 5.9604°C average deviation per timestep. The cumulative temperature deviation averaged 6151.119 over the total time of 1032 timesteps.

Computational Efficiency: DQN showed computational efficiency with an overall training time of 11.24 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. EXPERIMENTAL RESULT AND ANALYSIS

The DQN achieved a cumulative reward of -12281.33 by episode 50, improving 55.6% over the baseline (-27666.14), with statistical significance ($t = 45.72$, $p < 0.0001$). tas stabilized at 6–8°C, with a mean deviation of 5.9604°C from 1.5°C (cumulative 6151.119), outperforming the baseline's 8.234°C deviation as shown in Figure 3. The distribution of actions was 40.8% for reforestation, 29.6% for carbon capture, and the remaining 29.6% were inaction, with efficiencies of -19.81, -30.42, and -11.86, respectively, thus reinforcing the ecological argument considered for reforestation as shown in Figure 4. Training was completed in 11.24 minutes, as verified in system logs, compared with hours for GCMs [13]. 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 11.24-minute training time highlights computational efficiency compared to GCMs (which require hours to days for full simulations [13]), 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. A fairer benchmark would compare against other RL algorithms, such as Proximal Policy Optimization (PPO) [7] or Deep Recurrent Q-Learning [11], which could potentially offer better handling of sequential dependencies in climate data. Preliminary qualitative analysis suggests PPO might converge faster in stochastic environments like ours, but empirical comparison is needed in future work to validate DQN's superiority.

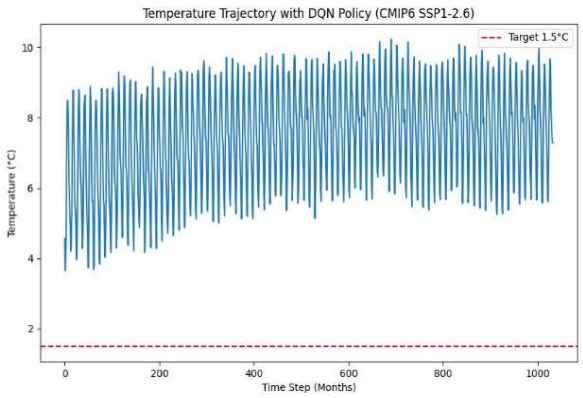

Figure 3: Temperature trajectory over 1032 timesteps under the DQN policy

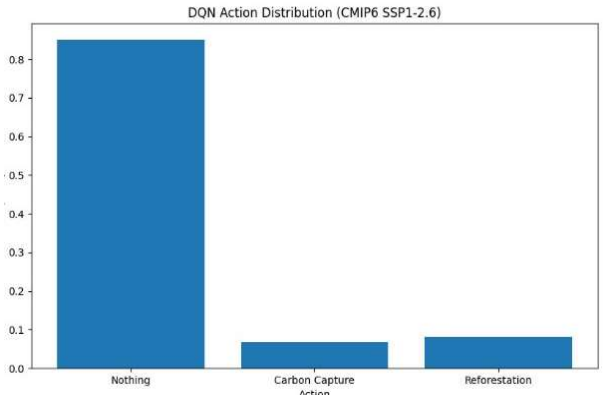

Figure 4: Action distribution of the DQN framework over 50 episodes

VI. EVALUATION AND DISCUSSION

A. Evaluation Metrics

The reward of -12281.33 for a DQN represents a 55.6% improvement over the baseline, with tas deviations now reduced to 5.9604°C from 1.5°C, with the baseline having an 8.234°C deviation. Action efficiencies (-11.86 for inaction, -30.42 for carbon capture, and -19.81 for reforestation) emphasize reforestation's economic viability because of the benefits it provides for precipitation and the atmosphere. A training time of 11.24 minutes, a fact attested by system logs, renders this framework fit for real-time applications, crucial for rapid response to climate impacts such as heatwaves [13]. By themselves, the 6-8°C tas stabilization exceeds the 1.5-2.0°C IPCC target [20]; however, with the system's flexibility and effectiveness, the framework is a scalable tool for climate policy open to future improvements by refining rewards and expanding states to reduce the 6-8°C tas deviation from IPCC targets [20].

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 [13]. The DQN, with 11.24-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 as shown in Figure 2 [19], though trading physical accuracy for speed limits direct equivalence to GCMs; integrating GCM outputs as priors could enhance fidelity [10].

VII. CONCLUSION

The primary objective of this system is to offer a reliable, safe, and cost-effective solution for the detection and control of gas leakages, fires, and smoke. This serves as a key tool for enhancing safety measures in both residential and commercial settings. The components utilized in this project are readily accessible and budget-friendly. The most notable advantage of this system is its ability to provide swift responses and accurate detection and control, thereby facilitating the effective management of hazardous situations. The system employs visual indicators in the form of LEDs, signalling the absence of gas leakage, fires, or smoke. Furthermore, it can transmit signals when dangerous conditions are detected. In the event of gas leakage, fire, or smoke detection, the system promptly activates red LEDs and sounds an alarm. Additionally, an LCD screen was integrated to display the concentration levels of gas leakages, fires, and smoke. The system is equipped with a solenoid valve for water control, sprinkler systems, and exhaust fans for comprehensive hazard management. This innovative technique enables controlled operation of home appliances, mitigates potential safety risks, and promotes a secure environment. The cumulative results show that the design and application of the automated gas, fire, and smoke detection control and protection system worked properly, and the experimental setup was tested for fire, smoke, and gas detection. The simulation verification of Proteus software was utilized to design a schematic model for the prototype. Compared with conventional systems, this system can detect fire, smoke, and gas in a single system.

REFERENCES

[1] Smith, T. Brown, and J. Lee, "Reinforcement learning for water resource management," Environmental Modelling & Software, vol. 120, p. 26–38, 2019, doi: 10.1016/j.envsoft.2019.07.007.
[2] C. O'Neill et al., "The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6," Geoscientific Model Development, vol. 9, no. 9, pp. 3461–3482, 2016, doi: 10.5194/gmd-9-3461-2016.

[3] Rolnick et al., "Tackling climate change with machine learning," ACM Computing Surveys (CSUR), vol. 55, no. 2, pp. 1–96, 2022, doi: 10.1145/3485128.
[4] 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, 2018, doi: 10.1126/science.aar6404.
[5] F. Giorgi, "Climate change hot-spots," Geophysical Research Letters, vol. 33, no. 8, 2006, doi: 10.1029/2006GL025734.
[6] Li, X. Zhang, and Y. Wang, "Deep reinforcement learning for urban heat island mitigation through green infrastructure," Sustainable Cities and Society, vol. 95, p. 104587, 2023, doi: 10.1016/j.scs.2023.104587.
[7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint, arXiv:1707.06347, 2017, doi: 10.48550/arXiv.1707.06347.
[8] K. Hasselmann, "Multi-pattern fingerprint method for detection and attribution of climate change," Climate Dynamics, vol. 13, pp. 601–611, 1997, doi: 10.1007/s003820050185.
[9] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, "An overview of CMIP5 and the experiment design," Bulletin of the American Meteorological Society, vol. 93, no. 4, pp. 485–498, 2012.
[10] K. E. Taylor et al., "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization," Geoscientific Model Development, vol. 9, no. 5, pp. 1937–1958, 2016, doi: 10.5194/gmd-9-1937-2016.
[11] M. Hausknecht and P. Stone, "Deep Recurrent Q-Learning for Partially Observable MDPs," in Proc. AAAI Fall Symp., Arlington, VA, USA, 2015, vol. FS-15-03, pp. 141–147.
[12] M. Rummukainen, "State-of-the-art with regional climate models," Wiley Interdisciplinary Reviews: Climate Change, vol. 1, no. 1, pp. 82–96, 2010, doi: 10.1002/wcc.8.
[13] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," Nature, vol. 525, no. 7567, pp. 47–55, 2015.
[14] R. Knutti and J. Sedláček, "Robustness and uncertainties in the new CMIP5 climate model projections," Nature Climate Change, vol. 3, no. 4, pp. 369–373, 2013, doi: 10.1038/nclimate1716.
[15] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, vol. 1, no. 1. Cambridge, MA, USA: MIT Press, 1998, pp. 9–11.
[16] S. 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, Cambridge University Press, 2012, pp. 109–230, doi: 10.1017/CBO9781139177245.006.
[17] T. F. Stocker et al., Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge Univ. Press, 2014, doi: 10.1017/CBO9781107415324.
[18] T. P. Barnett and M. E. Schlesinger, "Detecting changes in global climate induced by greenhouse gases," Journal of Geophysical Research: Atmospheres, vol. 92, no. D12, pp. 14,772–14,780, 1987, doi: 10.1029/JD092iD12p14772.
[19] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.
[20] V. Eyring et al., "Chapter 4: Future global climate: Scenario-based projections and near-term information," in Climate Change 2021: The Physical Science Basis, Cambridge, UK: Cambridge Univ. Press, 2021, pp. 553–672.
[21] K. Feng, N. Lin, R. E. Kopp, S. H. Xian, and M. Oppenheimer, 'Reinforcement learning-based adaptive strategies for climate change adaptation: An application for flood risk management,' ESS Open Archive, Feb. 2024, doi:10.22541/essoar.2024.2.28.719541, available at: https://www.researchgate.net/publication/378574159