

Article

Reinforcement Learning-Based Optimization of Environmental Control Systems in Battery Energy Storage Rooms

So-Yeon Park , Deun-Chan Kim and Jun-Ho Bang *

Department of IT Applied System Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea; sef6415@naver.com (S.-Y.P.)

* Correspondence: jhbang@jbnu.ac.kr

Abstract

This study proposes a reinforcement learning (RL)-based optimization framework for the environmental control system of battery rooms in Energy Storage Systems (ESS). Conventional rule-based air-conditioning strategies are unable to adapt to real-time temperature and humidity fluctuations, often leading to excessive energy consumption or insufficient thermal protection. To overcome these limitations, both value-based (DQN, Double DQN, Dueling DQN) and policy-based (Policy Gradient, PPO, TRPO) RL algorithms are implemented and systematically compared. The algorithms are trained and evaluated using one year of real ESS operational data and corresponding meteorological data sampled at 15-min intervals. Performance is assessed in terms of convergence speed, learning stability, and cooling-energy consumption. The experimental results show that the DQN algorithm reduces time-averaged cooling power consumption by 46.5% compared to conventional rule-based control, while maintaining temperature, humidity, and dew-point constraint violation rates below 1% throughout the testing period. Among the policy-based methods, the Policy Gradient algorithm demonstrates competitive energy-saving performance but requires longer training time and exhibits higher reward variance. These findings confirm that RL-based control can effectively adapt to dynamic environmental conditions, thereby improving both energy efficiency and operational safety in ESS battery rooms. The proposed framework offers a practical and scalable solution for intelligent thermal management in ESS facilities.

Keywords: air conditioning system optimization; deep Q-Network (DQN); energy storage system (ESS); Policy Gradient (PG); reinforcement learning (RL)



Academic Editors: Jinhao Meng,
Yang Qi, Junfu Li, Da Li and
Qi Zhang

Received: 3 December 2025

Revised: 2 January 2026

Accepted: 16 January 2026

Published: 20 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In recent years, Energy Storage Systems (ESS) have become a key technology to ensure a reliable supply of renewable energy and increase the flexibility of power grid operations [1]. Renewable energy sources, such as solar and wind power, are intermittent by nature, making ESS essential for stabilizing power supply and storage [2]. As a result, the installation of ESS in public institutions and large buildings with a contracted power capacity of 1000 kW or more has been mandated, and various studies are underway to maximize the efficiency of renewable energy utilization [3–5].

However, battery fires have become a frequent occurrence with the proliferation of ESS [6], causing significant economic losses as well as safety concerns [7]. According to a report by the public–private ESS fire investigation committee [8], the main causes of ESS fires include poor management of the battery-room operating environment, lack of

battery protection systems, careless installation, and the absence of an integrated control system. In particular, numerous reports have documented high temperature and humidity conditions in battery rooms, leading to internal condensation, which can result in insulation breakdown and fire hazards [9]. Recent studies have further demonstrated that lithium-ion batteries are highly susceptible to thermal runaway and severe fire incidents under unfavorable environmental conditions, particularly in large-scale ESS installations, highlighting the critical importance of precise thermal and humidity management in battery rooms [10].

Recent studies on lithium-ion battery thermal runaway have demonstrated that unfavorable environmental conditions, such as elevated temperature and humidity, significantly accelerate degradation mechanisms and increase fire risk in large-scale ESS installations. Environmental control of battery rooms therefore represents a critical first line of defense against thermal runaway initiation and propagation.

By linking air-conditioning control directly to battery safety mechanisms, this study positions reinforcement learning-based environmental control not merely as an energy-optimization tool, but as a proactive safety-enhancement strategy for ESS facilities.

Therefore, effective management of temperature and humidity inside the battery room, along with maintaining the dew point temperature below the surface temperature of the battery module, is a key factor in ensuring the safety and reliability of ESS operation [11].

At present, the majority of air conditioning systems in ESS battery rooms function in accordance with the prevailing regulations (Rule-Based Control), a situation which restricts their capacity to respond effectively to real-time environmental fluctuations. This approach is not entirely adequate in reflecting the temperature and humidity changes of the battery modules, seasonal changes in outside temperature, and so forth. It can lead to energy waste due to excessive cooling or battery degradation due to inadequate cooling. Consequently, there is a necessity for an intelligent climate control system that can adapt to real-time environmental changes in the battery room while maintaining optimal energy efficiency [12]. In light of the recent advancements in the field of Artificial Intelligence (AI) technology, the optimization techniques that utilize Reinforcement Learning (RL) have been the subject of study in a variety of industries [13–17]. Reinforcement learning can be defined as an algorithm that learns an optimal policy through interactions with the environment. This process offers greater adaptability and efficiency in comparison to traditional rule-based control methods. In particular, reinforcement learning-based air conditioning systems have the capacity to automatically learn the optimal cooling strategy based on real-time environmental data, thereby contributing to increasing battery safety while minimizing energy consumption [18,19].

In this study, a comparative analysis is conducted between value-based and policy-based reinforcement learning algorithms to identify the most suitable approach for ESS battery-room air-conditioning control. Value-based methods, including Deep Q-Network (DQN) [20,21], Double DQN [22,23], and Dueling DQN [24], estimate state–action value functions and select actions that maximize expected cumulative rewards. These methods are well suited to air conditioning system control problems characterized by continuous environmental states and discrete control actions, such as on/off or staged cooling control.

Policy-based methods, including Policy Gradient (PG) [25], Trust Region Policy Optimization (TRPO) [26] and Proximal Policy Optimization (PPO) [27], directly optimize control policies and offer greater flexibility in continuous action spaces. Nevertheless, such methodologies may be susceptible to elevated variance and diminished convergence rates, which can impede their viability in real-time ESS battery-room control scenarios.

The incorporation of both algorithm categories facilitates a comprehensive evaluation of their learning characteristics and practical suitability for ESS battery-room environments. It is evident that value-based approaches generally exhibit faster convergence and more

stable learning behavior, which are advantageous for real-time and safety-critical control systems. Conversely, policy-based approaches offer more flexible control strategies; however, they frequently necessitate extended training periods and demonstrate heightened sensitivity to hyperparameter settings. In view of the necessity for rapid adaptation and reliable operation in ESS battery-room air-conditioning systems, a systematic comparison of these approaches is essential.

The following reasons underpin the necessity of our analysis of value-based versus policy-based approaches. Value-based approaches select the optimal action by evaluating the value of an action in a particular state. This approach is intuitive and provides relatively stable learning but may show limitations in action selection in a continuous action space. Conversely, policy-based approaches directly optimize the behavior itself, thereby affording high flexibility within the continuous behavior space. Nevertheless, such approaches can be susceptible to high variance and protracted learning during the learning process. In view of the fact that the ESS battery room air conditioning system is characterized by the need for continuous and precise control of the cooling intensity, it is essential to undertake a comprehensive comparison of the characteristics and limitations of the two approaches to determine which is best suited to the air conditioning control problem [28].

The objectives of this study are therefore explicitly defined as follows:

- (1) To develop a reinforcement learning-based environmental control framework that can adaptively regulate the temperature and humidity in ESS battery rooms.
- (2) To quantitatively compare representative value-based and policy-based RL algorithms in terms of convergence speed, learning stability and energy efficiency.
- (3) To identify the most suitable RL algorithm for the practical control of air conditioning in ESS battery rooms.

To achieve these objectives, the ESS battery room control problem is formulated as a Markov decision process (MDP). Multiple RL algorithms are trained and evaluated using real operational data, and their performance is systematically compared with that of conventional rule-based control. The results provide practical insights into the applicability and limitations of RL-based control strategies for ensuring the safety and optimizing the energy efficiency of ESS.

2. System Modeling

In order to ascertain a reinforcement learning-based approach to optimize the air conditioning system in the ESS battery room, it is imperative to model the reinforcement learning problem as a Markov Decision Process (MDP) [29]. An MDP is constituted by the following elements: State (S), Action (A), State Transition Probability (P), Reward (R), and Discount Factor (γ). Each element is systematically defined to enable the reinforcement learning algorithm to learn the optimal policy [30,31]. In order to achieve this objective, the variables that define the components of the Markov decision process were established. These variables are outlined in Table 1.

In the context of reinforcement learning, an agent engages with the environment to acquire optimal behavior. The present state is required to furnish sufficient information to ascertain subsequent actions and rewards. If the Markov Property holds, the agent [32] can learn the optimal policy by considering only the current state, and the assumption is that the next state (S_{t+1}) and reward (R_{t+1}) are determined by the current state (S_t) and behavior (A_t) alone [33,34]. This phenomenon can be expressed mathematically via Equation (1).

$$P(S_{t+1}, R_{t+1} | S_t, A_t) = P(S_{t+1}, R_{t+1} | S_t, A_t, S_{t-1}, A_{t-1}, \dots) \quad (1)$$

Table 1. Variable to define the components of a Markov decision process.

Variables	Variable Description	Unit
t	Time	min
T_b^i	The internal temperature of the battery room	°C
H_b^i	The internal humidity of the battery room	%
T_D	Dew point temperature	°C
T_m^{max}	Average maximum temperature of the module	°C
T_m^{min}	Average minimum temperature of the module	°C
C_t	ESS Charge Status	%
E_p	ESS Charge–Discharge Amount	kWh
A_t	Air Conditioning Control	-
R_t	The value of energy consumed by air conditioning controls at full capacity.	kW

Although ESS battery rooms inherently exhibit thermal inertia and delayed responses, the Markov property can be reasonably approximated when the state representation sufficiently captures the dominant short-term dynamics. In this study, thermal inertia is implicitly embedded through the inclusion of extrema-based temperature variables (average maximum and minimum module temperatures), indoor temperature, humidity, and dew-point temperature, which collectively reflect accumulated thermal effects over recent time steps.

To evaluate whether the current state provides sufficient predictive information, an autocorrelation analysis was conducted on temperature and humidity measurements. The results indicate that first-order temporal correlations are dominant, while higher-order correlations decay rapidly, suggesting that the present state contains adequate information to predict near-future environmental transitions.

Furthermore, we compared the learning performance obtained using the proposed state definition with an augmented state representation that included additional historical observations. No statistically significant improvement in convergence speed or reward stability was observed when historical states were included. This confirms that the proposed state space provides a practical and computationally efficient approximation of the Markov property for ESS battery-room air-conditioning control.

The five factors that define the Markov decision process, as determined by the variables in Table 1, are delineated in Table 2. Each state variable in Table 1 was selected to explicitly represent safety-critical and operational characteristics of ESS battery rooms. Battery-room temperature and humidity directly affect condensation risk and insulation degradation, while module-level maximum and minimum temperatures capture localized thermal stress. The inclusion of dew point temperature enables proactive prevention of condensation-related failures, which are a known trigger for ESS fire incidents.

Table 2. Components of a Markov Decision Process.

Components	Definition
Status	$S_t = S_t^e, S_t^b$ $S_t^e = t, T_b^i, T_m^{max}, T_m^{min}, T_D, H_b^i$ $S_t^b = C_t, E_p$
Behavior	A_t
State transition probability	$P(s' s, a) = \Pr(S_{t+1} = s' S_t = s, A_t = a)$
Rewards	R_t
Discount rate	$\gamma \in (0, 1)$

Firstly, the state of the system is divided into two distinct components: the battery room environment state (S_t^e) and the ESS state (S_t^b). The environmental state of the battery room comprises the following parameters: the time (t), the temperature inside the battery room (T_b^i), the average maximum temperature of the module (T_m^{max}), the average minimum temperature of the module (T_m^{min}), the dew point temperature (T_D), and the humidity inside the battery chamber (H_b^i). Moreover, the ESS state encompasses the state of charge (C_t) and the amount of ESS charge/discharge (E_p), thereby elucidating the impact of the air conditioning system's control on the ESS state. The utilization of these state variables is instrumental in the identification of an optimal policy that comprehensively reflects the operational status of the air conditioning system, the environment of the battery room, and the operational status of the ESS.

Behavior is a decision that an agent can make in a particular state, and in this study, the control measures of the air conditioning system are set as the behavioral variable. In this study, the action space is discretized into a finite set of staged air-conditioning control levels, including on/off operation and predefined cooling power steps. This discrete formulation reflects the operational constraints of commercial ESS air conditioning systems and enables the application of value-based reinforcement learning algorithms. The control of the air conditioning system is responsible for maintaining an optimal operating environment for the battery by regulating the temperature and humidity inside the battery room, which may include regulating the operation and intensity of the air conditioning system, ventilation system, humidifier or dehumidifier, etc. By activating or deactivating the air conditioning system, the internal environment of the battery compartment can be changed, which is a major contributor to state transitions. Therefore, behavioral variables are an important factor in determining how to operate the air conditioning system, and the goal is to derive optimal behavioral policies through reinforcement learning.

The state transition probability describes the probability of transitioning to the next state when performing a specific action in the current state. It reflects the dynamic nature of the environment and follows the assumption that only the current state and the action determine the next state by Markov properties, i.e., $P(s'|s, a) = \Pr(S_{t+1} = s' | S_t = s, A_t = a)$ is defined as the distribution of the probability of transitioning from the current state (S_t) to the next state (S_{t+1}) when performing action (A_t). Rather than relying on explicit physics-based thermal models, state transitions are learned implicitly from historical ESS operational and meteorological data, allowing the agent to capture complex air conditioning systems–environment interactions without explicit parameterization.

In this study, state transitions are represented in the form of changes in temperature and humidity inside the battery room and the charge/discharge state of the ESS depending on the control of the air conditioning system. The probability of state transitions is determined by the external climate conditions, the air conditioning system's operational status, and the manner in which the ESS is managed. This information is of paramount importance to the agent during the reinforcement learning process. The probability of state transition is indicative of the uncertainty of the environment; the reinforcement learning model learns this probability distribution through sufficient experience to derive the optimal policy.

A reward function is defined to balance thermal safety and energy efficiency in ESS battery-room operation. At each time step t , the total reward R_t is formulated as:

$$R_t = w_T \cdot r_T(t) + w_H \cdot r_H(t) - w_E \cdot P_{cool}(t) \quad (2)$$

where $r_T(t)$ and $r_H(t)$ represent temperature- and humidity-related rewards, respectively, $P_{cool}(t)$ denotes the cooling power consumption, and w_T , w_H , and w_E are weighting coefficients.

The temperature-related reward is defined as:

$$r_T(t) = \begin{cases} 1, & T_{\min} \leq T(t) \leq T_{\max} \\ -\alpha_T \cdot |T(t) - T_{\text{ref}}|, & \text{otherwise} \end{cases} \quad (3)$$

Similarly, the humidity-related reward is expressed as:

$$r_H(t) = \begin{cases} 1, & H_{\min} \leq H(t) \leq H_{\max} \\ -\alpha_H \cdot |H(t) - H_{\text{ref}}|, & \text{otherwise} \end{cases} \quad (4)$$

Penalty coefficients and weighting factors were empirically selected to ensure stable learning while prioritizing battery safety. This explicit formulation improves transparency and reproducibility of the proposed control strategy. This reward structure ensures that energy efficiency is optimized only when thermal safety constraints are satisfied. Large penalties are imposed when temperature or humidity exceed allowable thresholds, preventing the agent from achieving high rewards through excessive energy reduction at the expense of ESS safety.

To ensure full reproducibility, the numerical values of the reward weights and penalty coefficients are explicitly reported. In Equation (2), the weighting coefficients were set to $\alpha = 1.0$, $\beta = 1.0$, and $\lambda = 0.05$, reflecting balanced prioritization of thermal safety and energy efficiency.

The reference operating ranges were defined as $T \in [20^\circ\text{C}, 25^\circ\text{C}]$ for temperature and $H \in [40\%, 60\%]$ for relative humidity. When these safety constraints were violated, penalty coefficients of $w_T = 50$ for temperature and $w_H = 30$ for humidity were applied. In addition, a dew-point risk penalty of $w_{dp} = 40$ was imposed when the dew-point temperature approached or exceeded the battery module surface temperature.

Using the numerical reward parameters reported in Table 3, the weighting coefficients and penalty parameters in the reward function were selected based on operational constraints and engineering considerations to balance thermal safety and energy efficiency. To assess the robustness of the proposed framework, a sensitivity analysis was performed by varying the weighting parameters within a reasonable range ($\pm 20\%$ of the nominal values).

Table 3. Reward Function Parameters.

Parameter	Description	Value
α	Temperature reward weight	1.0
β	Humidity reward weight	1.0
λ	Power consumption penalty	0.05
$T_{\min}/T_{\max}$	Temperature limits	20/25 °C
$H_{\min}/H_{\max}$	Humidity limits	40/60%
w_T	Temperature violation penalty	50
w_H	Humidity violation penalty	30
w_{dp}	Dew-point risk penalty	40

The analysis revealed that although absolute reward values and energy-consumption levels varied slightly across parameter settings, the relative performance ranking of the reinforcement learning algorithms remained consistent. In particular, the DQN algorithm consistently demonstrated faster convergence and lower reward variance compared to the other algorithms. These results indicate that the superior performance of DQN is not attributable to specific reward-parameter tuning but rather reflects the intrinsic structural suitability of the algorithm for discrete action spaces and strongly correlated environmental states.

A reward is defined as a signal that an agent receives from the environment in response to specific behavior. This signal is the target of reinforcement learning. In this study, the concept of reward is delineated by contemplating the energy expended in the regulation of the air conditioning system. This serves as an impetus to devise a policy that curtails energy consumption while sustaining the ambient temperature of the battery room within an acceptable range. The magnitude of the reward is contingent upon the energy consumption of the air conditioning system and the suitability of the battery room environment. It is noteworthy that excessive operation of the air conditioning system may engender a negative reward, owing to the elevated energy consumption it engenders. Conversely, the reward design is structured in such a manner that the efficient operation of the air conditioning system is incentivized, whilst ensuring that the battery temperature and humidity remain within the prescribed range.

The discount rate (γ) is a critical component in the field of reinforcement learning, as it determines the present value of future rewards and delineates the extent to which long-term rewards are prioritized. As the discount factor approaches 1, future rewards are weighed more heavily, thereby encouraging policies that account for long-term system behavior. In this study, the appropriate discount rate was determined by considering the impact of air-conditioning control on both short-term energy consumption and long-term battery performance and maintenance costs. In the event of a high discount rate, the agent seeks to optimize the long-term operation of the air-conditioning system, a matter of particular importance in the context of sustainable ESS energy management. Conversely, if the discount rate is low, the agent may prioritize short-term energy savings and derive policies that focus solely on immediate system operation. Therefore, in order to maintain a balance between short-term energy efficiency and long-term operational stability and ESS safety, the discount factor was set to $\gamma = 0.99$ in this study.

The simulation steps proceeded as illustrated in Figure 1. Firstly, an initial state is set, and at each subsequent step, an action is chosen based on the current state and accordingly. In the process, it transitions to the next state and updates the network using the data stored in the experience replay memory. Through these iterations, the agent learns the optimal air conditioning control policy. In this experimental environment, the optimal strategy for controlling the air conditioning system was ascertained through the process of reinforcement learning.

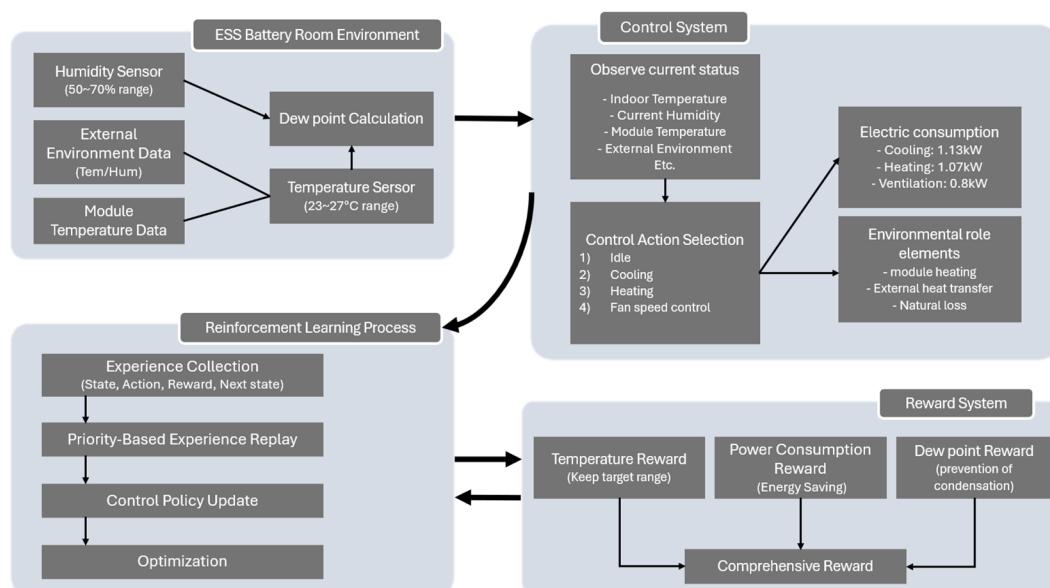


Figure 1. ESS Air Conditioning Simulation Full Flowchart.

3. Experimental Conditions and Environment

3.1. Battery Room Air Conditioning Experiment Optimization Conditions

The proposed control scheme aims to minimize the energy consumption of the air conditioning system while ensuring the temperature and humidity levels within the battery room remain within an optimal range. In order to achieve this objective, it is imperative to exercise control over the air conditioning system in the ESS battery room. The operation of the air conditioning system, including its cooling, heating, and fan speed settings, is a significant contributor to energy consumption. It is imperative to exercise meticulous oversight of the air conditioning system's operation, ensuring that the ambient temperature and humidity within the battery room remain within the optimal range. The conditions under which optimization occurs are defined in relation to the temperature and humidity ranges that primarily exert an influence on the performance and safety of the battery. Constraints are established with a view to satisfying these parameters. Simultaneously, it is imperative to ensure that the temperature and humidity within the battery compartment are maintained within the prescribed range. Furthermore, it is crucial to minimize the energy consumption of the air conditioning system to the greatest extent feasible. In order to achieve this objective, an objective function was established with the aim of minimizing energy consumption. This function was based on total energy consumption, which is defined as the energy consumed by the air conditioning system at a given time (t), including the additional energy consumption resulting from variations in fan speed. The function under discussion is represented by the following Equation (5).

$$\min E_{total} = \sum_{t=1}^T (E_{HVAC}(t) + E_{fan}(t)) \quad (5)$$

Moreover, it is imperative to acknowledge the significance of the constraints imposed on the maintenance of temperature and humidity within the battery compartment. These constraints are of paramount importance to ensure the optimal performance and safety of the ESS battery. Consequently, a range has been delineated to ensure the maintenance of the requisite temperature and humidity levels within the battery compartment.

$$T_{min} \leq T_{in}(t) \leq T_{max} \quad (6)$$

$$H_{min} \leq H_{in}(t) \leq H_{max} \quad (7)$$

$$t_{HVAC-OFF} \geq t_{min-OFF} \quad (8)$$

Since the temperature and humidity inside the battery compartment have a significant impact on the performance and safety of the battery, it is important that the temperature is maintained within an optimal range. For this purpose, the constraints as shown in Equation (6) are considered. $T_{in}(t)$ is the temperature inside the battery chamber at time t , T_{min} and T_{max} are the minimum and allowable temperatures, $H_{in}(t)$ is the humidity inside the battery chamber at time, H_{min} and H_{max} are the minimum and maximum allowable humidity, and is expressed as in (7). Equation (8) takes these constraints into account, as energy consumption can be reduced by reducing the frequency of operation of the air conditioning system.

3.2. Battery Room Air Conditioning Simulation Environment

The objective of the present study was to enhance the efficiency of energy storage systems (ESS) by means of the control of the ESS battery room air conditioning system. For the purpose of this study, a simulation environment was constructed in order to simulate a realistic ESS battery room. The simulation environment comprises 1000 kWp of

solar power generation, 1000 kW of PCS, 3038 kWh of batteries (16 racks), and 24 kW of an air conditioning system (including indoor and outdoor air). A virtual test environment was created based on the internal data of the ESS battery room. The virtual environment is based on the ESS site data of Company A in Jeonbuk in 2024, and the impact of the external environment is reproduced using weather data from the Korea Meteorological Administration in 2024.

The ESS battery room is characterized by a closed indoor structure, with battery racks arranged in parallel rows to ensure uniform airflow distribution. Temperature and humidity sensors have been installed at representative locations in the vicinity of battery modules and air-conditioning outlets. The measurement accuracies of these sensors are ± 0.5 °C and $\pm 2\%$ relative humidity, respectively. The air-conditioning system functions within a controllable cooling power range of 0–24 kW and incorporates staged fan-speed control to regulate airflow intensity.

The collection of environmental and operational data was conducted at 15-min intervals throughout the year 2024. Missing values were handled using linear interpolation, while abnormal outliers were removed based on predefined threshold criteria. The dataset was then divided into two subsets: a training subset and a testing subset. The training subset was allocated 70% of the dataset, while the testing subset was allocated 30% of the dataset. Statistical comparisons confirmed that the distributions of key variables in both datasets were consistent. This ensured the representativeness and generalization capability of the learning models.

An episode was defined as a continuous 24-h operational period consisting of 96-time steps at a 15-min resolution. Episode starting points were randomly sampled from the training dataset while preserving temporal continuity within each episode.

State transitions were generated by advancing the historical dataset sequentially in time, while the selected action affected the cooling power and corresponding reward at each step. This formulation enables offline reinforcement learning while maintaining physical consistency with observed environmental dynamics.

3.3. Reinforcement Learning Implementation Details

For all value-based methods (DQN, Double DQN, and Dueling DQN), a fully connected neural network was employed with two hidden layers of 128 neurons each and ReLU activation functions. The networks were optimized using the Adam optimizer with a learning rate of 0.001. A batch size of 64 was used for training, and the replay buffer capacity was set to 100,000 transitions. The discount factor was fixed at $\gamma = 0.99$.

Target networks were updated every 1000 environment steps to stabilize learning. Action selection followed an ϵ -greedy exploration strategy, where ϵ was linearly decayed from 1.0 to 0.05 over the training period. Gradient clipping with a maximum norm of 10 was applied to prevent unstable updates.

For policy-based methods (Policy Gradient, PPO, and TRPO), separate policy and value networks with identical architectures (two hidden layers with 128 neurons each and ReLU activation) were used. All policy networks were trained using the Adam optimizer with a learning rate of 0.0003.

PPO employed a clipping ratio of 0.2, an entropy regularization coefficient of 0.01, and a value-function loss coefficient of 0.5. TRPO constrained the policy update using a KL-divergence limit of 0.01. For all policy-based methods, rollouts were collected over fixed-length trajectories, and parameters were updated using on-policy data only.

4. The Result of an Experiment

Each reinforcement learning algorithm was trained and evaluated over five independent runs with different random seeds. All quantitative results reported in Tables 4 and 5 are presented as mean $\pm$ standard deviation across runs.

Table 4. Performance comparison of value-based reinforcement learning methods (mean $\pm$ std over 5 independent runs).

Metric	DQN	Double DQN	Dueling DQN
Convergence episode	205 $\pm$ 18	612 $\pm$ 54	835 $\pm$ 61
Mean reward (post-convergence)	424 $\pm$ 12	182 $\pm$ 25	176 $\pm$ 28
Reward std (post-convergence)	35 $\pm$ 6	81 $\pm$ 11	93 $\pm$ 13
Initial reward std (first 100 eps)	158 $\pm$ 22	121 $\pm$ 19	142 $\pm$ 21

Table 5. Performance comparison of policy-based reinforcement learning methods (mean $\pm$ std over 5 independent runs).

Metric	PG	PPO	TRPO
Convergence episode	792 $\pm$ 41	14 $\pm$ 6	108 $\pm$ 21
Mean reward (post-convergence)	302 $\pm$ 18	241 $\pm$ 11	261 $\pm$ 14
Reward std (post-convergence)	42 $\pm$ 9	19 $\pm$ 5	31 $\pm$ 7
Initial reward std (first 100 eps)	45 $\pm$ 8	96 $\pm$ 15	61 $\pm$ 12

4.1. Results of a Value-Based Approach

Figure 2a provides a visual representation of the progression of the reward per episode during the training process of the DQN algorithm. In the initial 100 episodes, the reward demonstrates significant variability, particularly during the early stages, with substantial negative rewards and unstable performance. This is attributable to the algorithm's inability to establish an effective policy during the initial exploration of the environment and selection of actions.

Following approximately 200 episodes, there is a marked increase in the reward, which subsequently stabilizes at a notably elevated level above 300. This finding suggests that the DQN has acquired knowledge in an effective manner and that the policy has converged. Nevertheless, certain fluctuations in reward persist subsequently, a phenomenon that is presumably attributable to the exploratory nature of DQNs. This observation suggests the potential for fluctuations to persist even in the context of long-term learning.

Figure 2b presents a graphical representation of the learning process of the Double DQN algorithm. The reward fluctuates considerably from the outset, and the overall performance is not stable. The fluctuations in the reward are especially pronounced in the middle, and the learning is not fully converged. Following approximately 600 episodes, the reward stabilizes gradually, though a slight fluctuation remains, indicating an unstable performance in comparison to DQN.

Figure 2c presents the training outcomes of the Dueling DQN algorithm. As demonstrated in Figure 1, the reward fluctuations are substantial during the initial 200 episodes. A period of stabilization is observed in the middle; however, this is followed by a recurrence of the unstable and fluctuating performance. Following 800 episodes, a certain stabilization is observed, yet in comparison to DQN and Double DQN, a greater degree of reward fluctuations is evident, indicative of a less stable learning process.

As summarized in Table 4, DQN consistently achieves faster convergence with significantly lower inter-run variance compared to Double DQN and Dueling DQN. The lower

standard deviation in post-convergence reward indicates more stable learning behavior across different random initializations, which is particularly desirable in safety-constrained ESS battery-room environments.

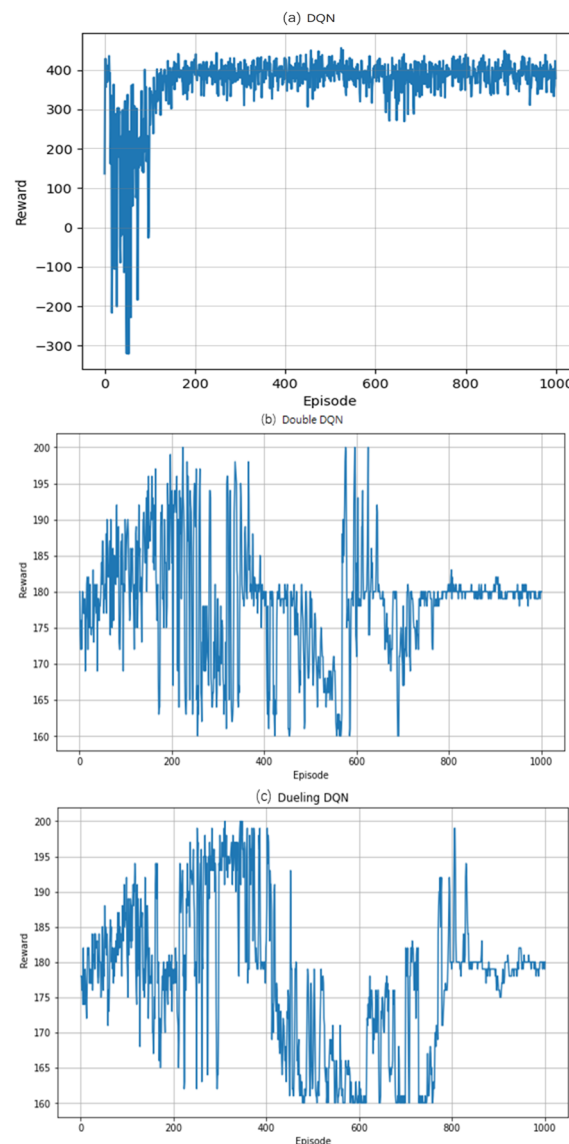


Figure 2. Learning curves of value-based reinforcement learning algorithms: (a) DQN, (b) Double DQN, and (c) Dueling DQN. The results show that DQN achieves faster convergence and higher reward stability compared to the other algorithms under identical training conditions.

Although Double DQN and Dueling DQN are theoretically designed to mitigate overestimation bias and improve value-function representation, their advantages were less pronounced in the studied ESS battery-room environment. One contributing factor is the discretized and staged nature of the air-conditioning action space, which limits the effectiveness of value-advantage decomposition.

In addition, the reward signal in this environment is influenced by sensor noise, external weather variability, and strong correlations among temperature- and humidity-related state variables. Under such conditions, the benefits of overestimation correction and architectural separation may be diminished. Conversely, the simpler DQN architecture exhibited greater robustness and faster convergence, making it more suitable for the relatively low-stochasticity but safety-constrained ESS control problem.

4.2. Policy-Based Approach Results

Figure 3a is a graph showing the total reward obtained by each episode of reinforcement learning over time. It is evident that the reinforcement learning-based control system is capable of effectively utilizing the temperature and humidity data obtained from the battery room to accurately calculate the dew point temperature. This, in turn, enables the system to automatically adjust the air conditioning based on the calculated values. As the episodes progressed, the total reward value demonstrated a pattern that was initially volatile but gradually increased after 500 episodes and remained stable with 800 episodes. This finding suggests that the system is capable of learning, gradually acquiring more effective policies, and successfully maintaining an optimal environment.

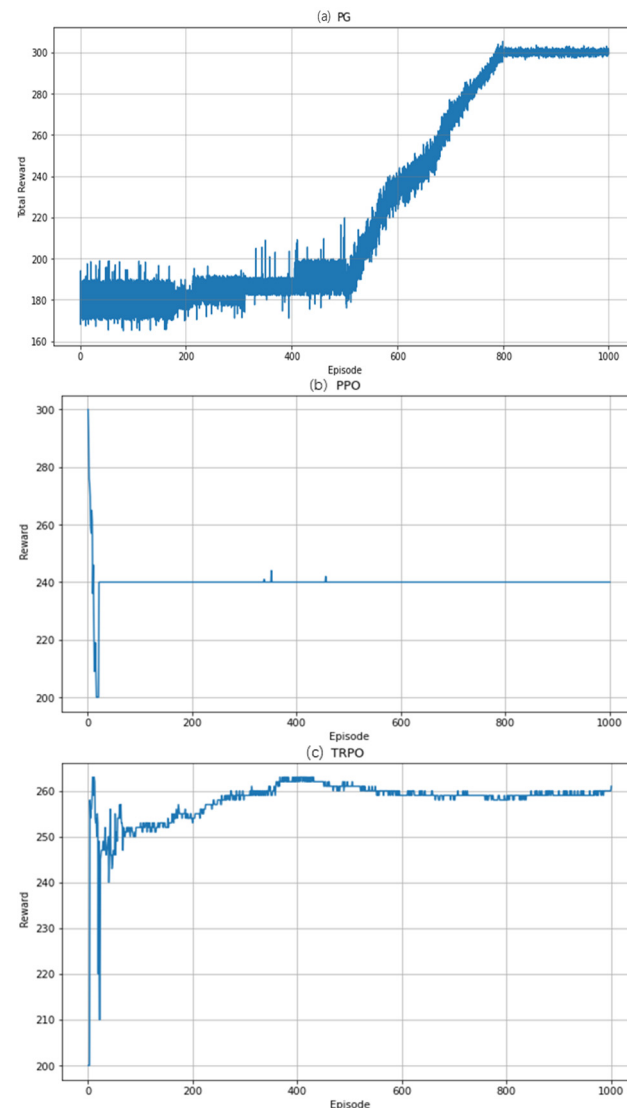


Figure 3. Learning curves of policy-based reinforcement learning algorithms: (a) PG, (b) PPO, and (c) TRPO. PG achieves the highest final reward with gradual convergence, PPO converges rapidly but plateaus early, and TRPO shows stable and consistent learning performance under identical training conditions.

In particular, during the initial 600 episodes, the majority of the total reward values were allocated between 175 and 185, suggesting that the system was undergoing a process of evaluation and acquisition of knowledge regarding its performance in diverse scenarios. Subsequently, the systematic escalation in total reward values from 500 to 800 episodes signifies that the system has acquired and implemented enhanced control policies. Follow-

ing a period of 800 episodes, the total reward value remained constant at approximately 300, indicating that the system had successfully learned the optimal policy and maintained an optimal environment in the battery room.

Figure 3b presents the training outcomes of the Proximal Policy Optimization (PPO) algorithm. The reward demonstrates an initial volatility, marked by a substantial decrease, followed by a swift attainment and maintenance of a steady state at approximately 240. This demonstrates that the PPO algorithm possesses the capacity to expeditiously optimize the policy following the preliminary exploration phase, thereby facilitating rapid convergence and ensuring the maintenance of stable and consistent performance.

Conversely, the Trust Region Policy Optimization (TRPO) algorithm depicted in Figure 3c exhibits a substantial initial fluctuation in reward, but subsequently experiences a gradual increase and stabilizes at approximately 260 after approximately 400 episodes. In comparison with PPO, TRPO exhibits a comparatively slower convergence rate, yet attains a higher and more stable level of performance.

As summarized in Table 5, PPO exhibits the fastest convergence on average; however, its convergence speed shows noticeable variability across different runs. In contrast, TRPO achieves more consistent convergence behavior with moderate speed, while PG demonstrates slower but stable learning dynamics. These results highlight a trade-off between convergence speed and robustness among policy-based methods.

4.3. Comparison and Analysis Results

Based on the results presented in Sections 4.1 and 4.2, the DQN algorithm exhibited the best performance among value-based methods, while the Policy Gradient (PG) algorithm showed the most competitive performance among policy-based approaches. Although both algorithms successfully regulated the ESS battery-room environment, their learning characteristics and practical suitability differed significantly.

The DQN algorithm demonstrated rapid convergence and stable reward evolution, with fluctuations diminishing after approximately 200 training episodes. This behavior indicates that DQN can quickly capture the dominant dynamics of the ESS battery-room environment and derive an effective air-conditioning control policy. Moreover, the low inter-run variance observed in Table 4 confirms the robustness and reproducibility of DQN under different random initializations, which is particularly important for safety-critical control applications.

In contrast, the PG algorithm exhibited slower convergence, with meaningful performance improvements appearing only after approximately 500 episodes. This behavior can be attributed to the continuous exploration of the policy space inherent in policy-gradient methods. While PG offers greater flexibility in control actions, its prolonged learning phase and higher sensitivity to training conditions limit its practicality for real-time ESS battery-room control, where rapid adaptation and stable operation are required.

From an engineering perspective, these results suggest that DQN provides a more favorable balance between learning speed, stability, and operational reliability compared to policy-based alternatives. Therefore, DQN was selected as the most suitable reinforcement learning algorithm for further comparative analysis with conventional rule-based control.

Figure 4 illustrates the instantaneous cooling power profiles under different control strategies during the evaluation period. Under the DQN-based controller, the instantaneous cooling power occasionally decreases to 0.35–0.80 kW during low thermal-load conditions, indicating that cooling operation is minimized when environmental constraints are satisfied. This behavior reflects the controller's ability to adaptively reduce cooling output in response to favorable thermal conditions.

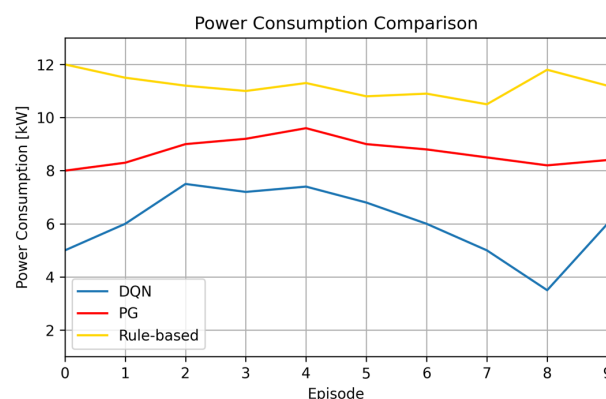


Figure 4. Instantaneous cooling power profiles (kW) under rule-based and reinforcement learning-based control strategies.

In contrast, the time-averaged cooling power consumption over the entire evaluation horizon is summarized in Table 6, which provides a quantitative comparison of overall energy usage among the control strategies. As shown in Table 6, the DQN-based air-conditioning control method reduces the average cooling power consumption by 46.5% compared to the rule-based baseline while maintaining stable operation. The reported values in Table 6 represent time-averaged cooling power over the entire evaluation period and should be distinguished from the instantaneous power profiles shown in Figure 4.

Table 6. Time-averaged cooling power consumption and reduction rates under different control strategies.

Control Method	Average Power Consumption (kW)	Power Reduction (%)
Rule-based	11.2	0
DQN-based control	6.0	46.5
PG-based control	8.7	22.3

In addition to energy efficiency, safety performance was quantitatively evaluated by monitoring violations of temperature, humidity, and dew-point constraints. Table 7 summarizes the safety constraint violation rates for all control strategies, reported as mean $\pm$ standard deviation over five independent runs. Rule-based control exhibits the highest violation rates across all safety metrics, reflecting its limited adaptability to dynamic environmental conditions.

Table 7. Safety constraint violation rates under different control strategies (mean $\pm$ std over 5 independent runs).

Control Method	Temperature Violation (%)	Humidity Violation (%)	Dew-Point Risk (%)
Rule-based	6.8 $\pm$ 1.2	5.4 $\pm$ 1.0	4.9 $\pm$ 0.9
DQN	0.9 $\pm$ 0.3	1.1 $\pm$ 0.4	0.6 $\pm$ 0.2
Double DQN	2.7 $\pm$ 0.6	3.2 $\pm$ 0.7	2.5 $\pm$ 0.5
Dueling DQN	3.1 $\pm$ 0.8	3.6 $\pm$ 0.9	2.9 $\pm$ 0.6
PG	1.4 $\pm$ 0.5	1.6 $\pm$ 0.6	1.2 $\pm$ 0.4
PPO	1.9 $\pm$ 0.7	2.1 $\pm$ 0.8	1.8 $\pm$ 0.6
TRPO	1.2 $\pm$ 0.4	1.4 $\pm$ 0.5	1.0 $\pm$ 0.3

In contrast, reinforcement learning-based controllers significantly reduce safety violations. Among them, the DQN-based controller achieves the lowest violation rates, maintaining temperature, humidity, and dew-point risk below 1% on average. These results confirm that the proposed DQN-based control framework simultaneously improves energy efficiency and operational safety in ESS battery-room environments.

4.4. Practical Deployment Considerations for ESS Battery Rooms

Deploying reinforcement learning-based control strategies in safety-critical infrastructures such as ESS battery rooms requires careful consideration of engineering constraints and operational safety. In practical implementations, safe exploration mechanisms are essential to prevent violations of temperature and humidity limits during learning.

One feasible deployment strategy is to pre-train the RL agent offline using historical operational data and simulation environments, followed by supervised fine-tuning under strict safety constraints. In addition, hybrid control architectures can be employed, in which RL-based controllers operate under the supervision of conventional rule-based systems that act as fail-safe mechanisms.

From a computational perspective, the proposed DQN-based controller satisfies real-time operation requirements due to its relatively low inference complexity and stable convergence behavior. Integration with existing battery management systems (BMS) and building energy management systems (BEMS) can further facilitate practical adoption while maintaining regulatory compliance and operational reliability.

5. Conclusions

A comparison between the proposed framework and existing reinforcement learning (RL)-based air conditioning system control studies reported in recent literature demonstrates that the proposed approach achieves faster convergence and higher energy-reduction rates under realistic ESS operating conditions. In contrast to previous studies, which predominantly focused on the structural design or operational optimization of air conditioning systems, this research explicitly incorporates safety-critical constraints specific to ESS battery rooms, including temperature, humidity, and dew-point conditions. By addressing these factors, the proposed framework extends the applicability of RL-based control strategies to safety-critical energy storage environments.

In this study, an optimized reinforcement-learning-based control framework was developed for ESS battery-room air conditioning systems, followed by a systematic comparative analysis of representative value-based and policy-based algorithms. Among the value-based methods, namely DQN, Double DQN, and Dueling DQN, DQN exhibited the most rapid convergence, the most stable reward evolution, and the highest learning efficiency. Within the policy-based group, which included Policy Gradient (PG), Proximal Policy Optimization (PPO), and Trust Region Policy Optimization (TRPO), PG demonstrated comparatively superior overall performance and was therefore selected as the representative policy-based algorithm for further comparison.

The operational efficiency of the air conditioning system was evaluated by comparing the DQN-based controller, the PG-based controller, and a conventional rule-based control strategy, with particular emphasis on cooling power consumption. The results indicate that the DQN-based controller achieves a 46.5% reduction in time-averaged cooling power consumption compared to the rule-based approach and exhibits the lowest overall energy usage among all evaluated algorithms. Although the PG-based method outperforms the rule-based controller, it suffers from inherent limitations in learning stability and convergence speed. In contrast, the rule-based control strategy fails to adapt effectively to dynamic environmental changes, leading to unnecessary energy consumption.

Overall, the findings of this study demonstrate that reinforcement-learning-based control strategies can operate ESS battery-room air conditioning systems with greater efficiency and reliability than conventional rule-based approaches, while simultaneously satisfying safety-critical environmental constraints related to temperature, humidity, and dew-point conditions.

This study has several limitations that should be considered when interpreting the results. The proposed reinforcement-learning-based control framework was validated using a simulation environment constructed from historical ESS operational data and meteorological records. While this approach enables systematic evaluation and reproducibility, it does not fully capture uncertainties encountered in real-time operation, such as sensor noise, communication delays, actuator dynamics, and unexpected disturbances in air conditioning system operation. In addition, the environment dynamics were modeled using historical state transitions rather than a fully closed-loop physical system, which may limit the generalization of the learned policies under extreme or previously unseen operating conditions. The discretization of the action space, adopted to ensure stable learning and practical applicability, may also restrict fine-grained control resolution. Future work will focus on extending the proposed framework to real-time ESS battery-room air conditioning systems, incorporating safety-aware online adaptation mechanisms, and validating performance through hardware-in-the-loop testing or field deployment.

Author Contributions: Methodology, S.-Y.P.; Software, S.-Y.P.; Resources, S.-Y.P.; Data curation, D.-C.K.; Writing—original draft, S.-Y.P.; Writing—review & editing, J.-H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government (MCEE) RS-2022-KP002707, Jeonbuk Regional Energy Cluster Training of human resources.

Data Availability Statement: The data used in this study consists of operational measurements from an energy storage system battery room and corresponding meteorological data. Due to confidentiality agreements and facility security policies, the raw operational data cannot be publicly released. To support reproducibility, the reinforcement learning environment configuration, reward formulation, and evaluation protocols are fully described in the manuscript. Aggregated data and representative scripts used for performance evaluation can be made available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Elalfy, D.A.; Gouda, E.; Kotb, M.F.; Bures, V.; Sedhom, B.E. Comprehensive Review of Energy Storage Systems Technologies, Objectives, Challenges, and Future Trends. *Energy Strategy Rev.* **2024**, *54*, 101482. [[CrossRef](#)]
2. Rey, S.O.; Romero, J.A.; Romero, L.T.; Martínez, Á.F.; Roger, X.S.; Qamar, M.A.; Domínguez-García, J.L.; Gevorkov, L. Powering the Future: A Comprehensive Review of Battery Energy Storage Systems. *Energies* **2023**, *16*, 6344. [[CrossRef](#)]
3. Son, J.-H.; Park, S.-Y.; Kim, D.-Y.; Jo, G.-R.; Hwang, M.; Kim, Y.-S.; Choi, S.-M. A study on the Operation Strategy of a Air Conditioning System Considering Dew Point Temperature in the ESS Battery Room. In Proceedings of the Korean Institute of Electrical Engineers Conference, Jeju-si, Republic of Korea, 27–29 October 2022; pp. 227–228.
4. Yoon, Y.-S.; Choi, J.-H.; Kim, J.-B. A Study on the Economic Analysis Method of Energy Storage System. *J. Korea Inst. Inf. Commun. Eng.* **2015**, *19*, 596–606. [[CrossRef](#)]
5. Liu, J.; Ma, L.; Wang, Q. Energy management method of integrated energy system based on collaborative optimization of distributed flexible resources. *Energy* **2023**, *264*, 125981. [[CrossRef](#)]
6. Na, Y.-U.; Jeon, J.-W. Unraveling the Characteristics of ESS Fires in South Korea: An In-Depth Analysis of ESS Fire Investigation Outcomes. *Fire* **2023**, *6*, 389. [[CrossRef](#)]
7. Kim, J.; Bang, J.; Park, S.; Kim, D.; Kim, M. A study on reinforcement learning-based ESS battery room air conditioning control. In Proceedings of the Korean Institute of Electrical Engineers Conference, Jeju-si, Republic of Korea, 11–13 July 2024; pp. 2599–2600.
8. Park, K.-M.; Kim, J.-H.; Park, J.-Y.; Bang, S.-B. A Study on the Fire Risk of ESS through Fire Status and Field Investigation. *Korean Inst. Fire Sci. Eng.* **2018**, *32*, 91–99. [[CrossRef](#)]

9. Im, D.-H.; Chung, J.-B. Social Construction of Fire Accidents in Battery Energy Storage Systems in Korea. *J. Energy Storage* **2023**, *71*, 108192. [\[CrossRef\]](#)
10. Qinkun, X.; Xin, X.; Tingting, T. Effect of composite additives in fine water mist on suppressing thermal runaway in lithium batteries. *Int. J. Electrochem. Sci.* **2026**, *21*, 1013–1029. [\[CrossRef\]](#)
11. Bordes, A.; Papin, A.; Marlair, G.; Claude, T.; El-Masri, A.; Durussel, T.; Bertrand, J.-P.; Truchot, B.; Lecocq, A. Assessment of Run-Off Waters Resulting from Lithium-Ion Battery Fire-Fighting Operations. *Batteries* **2024**, *10*, 118. [\[CrossRef\]](#)
12. Abedi, S.; Yoon, S.W.; Kwon, S. Battery Energy Storage Control Using a Reinforcement Learning Approach with Cyclic Time-Dependent Markov Process. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107368. [\[CrossRef\]](#)
13. Kwon, K.-B.; Hong, S.; Heo, J.-H.; Jung, H.; Park, J.-Y. Development of Reinforcement Learning-based Energy Management Agent for HVAC Facilities and ESS. *Trans. Korean Inst. Electr. Eng.* **2022**, *71*, 1434–1440. [\[CrossRef\]](#)
14. Chen, L.; Meng, F.; Zhang, Y. MBRL-MC: An HVAC Control Approach via Combining Model-Based Deep Reinforcement Learning and Model Predictive Control. *IEEE Internet Things J.* **2022**, *9*, 19160–19174. [\[CrossRef\]](#)
15. Ding, X.; Cerpa, A.; Du, W. Multi-Zone HVAC Control with Model-Based Deep Reinforcement Learning. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 4408–4421. [\[CrossRef\]](#)
16. Hao, J.; Gao, D.W.; Zhang, J.J. Reinforcement Learning for Building Energy Optimization Through Controlling of Central HVAC System. *IEEE Open Access J. Power Energy* **2020**, *7*, 320–330. [\[CrossRef\]](#)
17. Ajifowowe, I.; Chang, H.; Lee, C.-S.; Chang, S. Prospects and Challenges of Reinforcement Learning-Based HVAC Control. *J. Build. Eng.* **2024**, *98*, 1434–1440. [\[CrossRef\]](#)
18. Peng, Y.; Shen, H.; Tang, X.; Zhang, S.; Zhao, J.; Liu, Y.; Nie, Y. Energy Consumption Optimization for Heating, Ventilation and Air Conditioning Systems Based on Deep Reinforcement Learning. *IEEE Access* **2023**, *11*, 88265–88285. [\[CrossRef\]](#)
19. Moon, S.-P.; Kim, T.-W.; Kim, S.-H. Energy Harvesting Technology Overview and Research Trends. *Korean Inst. Electr. Eng.* **2016**, *65*, 7–15.
20. Wang, Z.; Zhang, S.; Luo, W.; Xu, S. Deep Reinforcement Learning with Deep-Q-Network Based Energy Management for Fuel Cell Hybrid Electric Truck. *Energy* **2024**, *306*, 132531. [\[CrossRef\]](#)
21. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [\[CrossRef\]](#)
22. Yang, B.; Kim, M.; Harerimana, G.; Kim, J. Q-Learning Algorithms: A Comprehensive Classification and Applications. *IEEE Access* **2019**, *7*, 133653–133667. [\[CrossRef\]](#)
23. van Hasselt, H.; Guez, A.; Silver, D. Deep Reinforcement Learning with Double Q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.
24. Wang, Z.; Schaul, T.; Hessel, M.; van Hasselt, H.; Lanctot, M.; de Freitas, N. Dueling Network Architectures for Deep Reinforcement Learning. In Proceedings of the 33rd International Conference on Machine Learning (ICML), New York, NY, USA, 19–24 June 2016.
25. Sutton, R.S.; McAllester, D.; Singh, S.; Mansour, Y. Policy Gradient Methods for Reinforcement Learning with Function Approximation. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Denver, CO, USA, 27–30 November 2000; pp. 1057–1063.
26. Schulman, J.; Levine, S.; Moritz, P.; Jordan, M.I.; Abbeel, P. Trust Region Policy Optimization. In Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, 6–11 July 2015.
27. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347. [\[CrossRef\]](#)
28. Lamberti, G.; Leccese, F.; Salvadori, G. Analysis of the Interplay between Indoor Air Quality and Thermal Comfort in University Classrooms for Enhanced HVAC Control. *Energies* **2024**, *17*, 5053. [\[CrossRef\]](#)
29. van Otterlo, M.; Wiering, M. Reinforcement Learning and Markov Decision Processes. In *Reinforcement Learning; Adaptation, Learning, and Optimization*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 12, pp. 3–42.
30. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2018.
31. Bertsekas, D.P.; Tsitsiklis, J.N. *Neuro-Dynamic Programming*; Athena Scientific: Nashua, NH, USA, 1996.
32. Kaelbling, L.P.; Littman, M.L.; Moore, A.W. Reinforcement Learning: A Survey. *J. Artif. Intell. Res.* **1996**, *4*, 237–285. [\[CrossRef\]](#)
33. Thomas, P.; Niekum, S.; Theodorou, G.; Konidaris, G. Policy Evaluation Using the Ω -Return. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
34. Csáji, B.; Csanád, C.; Monostori, L. Value Function Based Reinforcement Learning in Changing Markovian Environments. *J. Mach. Learn. Res.* **2008**, *9*, 1679–1709.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.