

# Reinforcement Learning-Driven Energy Optimization for Smart Buildings: A Deep Q-Network Approach

## Abstract

**Background:** The rapid expansion of smart building technologies demands innovative solutions to optimize energy consumption while ensuring occupant comfort. Traditional rule-based and supervised learning approaches often lack adaptability to dynamic environmental conditions, leading to inefficiencies in HVAC and lighting control. Reinforcement learning (RL) offers a promising alternative by enabling autonomous, data-driven decision-making in complex building environments.

**Methodology:** This study proposes a Deep Q-Network (DQN)-based RL framework for real-time energy management in smart buildings. The system integrates real-time sensor data (temperature, occupancy, weather) with a virtual building model (EnergyPlus + OpenAI Gym) to train an adaptive control agent. A custom reward function balances energy savings and thermal comfort, while experience replay stabilizes training. The framework was evaluated against rule-based and supervised learning baselines using metrics such as energy consumption (kWh), comfort deviation (ASHRAE standards), and control stability.

**Key Findings:** The proposed system achieved a 22% reduction in energy consumption compared to conventional rule-based systems while maintaining a significantly lower comfort violation rate of just 5%, outperforming traditional methods that exhibited a 12% violation rate. The reinforcement learning approach demonstrated superior adaptability to dynamic occupancy changes and weather fluctuations, though this enhanced performance came with inherent trade-offs between computational cost and real-time responsiveness that must be carefully considered in practical implementations. These results collectively demonstrate the system's ability to simultaneously optimize energy efficiency and occupant comfort while adapting to real-world building conditions. The results also validate RL as a scalable solution for sustainable building operations, bridging the gap between simulation and real-world deployment.

**Keywords:** Energy optimization, smart buildings, reinforcement learning, DQN, HVAC control, AI in facility management, adaptive control.

## 1. Introduction

Energy consumption in buildings represents one of the largest contributors to global electricity demand, accounting for approximately 40% of total usage. This substantial share underscores the critical importance of improving energy efficiency within the built environment, not only to reduce operational costs but also to support global sustainability and carbon reduction goals (González-Torres et al., 2022). As urbanization continues to accelerate, and as more buildings integrate advanced technologies, optimizing energy performance has become both a technical imperative and a policy priority (Orikpete et al., 2023). Historically, energy management in buildings has been governed by traditional Building Management Systems (BMS), which utilize pre-programmed schedules, static rules, or time-based control strategies. While these systems have proven effective to some degree, they often lack the responsiveness and adaptability needed to accommodate real-time variations in occupancy patterns, weather conditions, indoor thermal loads, and user behavior (Al-Ghaili et al., 2021). As a result, buildings frequently operate inefficiently, either over-conditioning unoccupied spaces or failing to respond promptly to changing conditions, leading to energy waste and occupant discomfort.

Recent advances in Artificial Intelligence (AI) have paved the way for more intelligent and responsive control strategies (Khan et al., 2022). In particular, Reinforcement Learning (RL), a branch of machine learning where agents learn optimal actions through interactions with their environment, has emerged as a promising tool for energy optimization (Stavrev & Ginchev, 2024). Unlike supervised learning, RL does not require labeled datasets, making it especially suitable for complex and dynamic environments such as buildings, where explicit ground truth is often unavailable (Nweye et al., 2022). This paper introduces a model-free Reinforcement Learning framework for adaptive control of HVAC (Heating, Ventilation, and Air Conditioning) and lighting systems in commercial and institutional buildings. The proposed framework learns optimal control policies by interacting with building simulations or real-time sensor data, aiming to minimize energy consumption while maintaining acceptable levels of occupant comfort. By continuously adapting to environmental changes and user presence, the system improves over time, making intelligent adjustments that outperform traditional rule-based approaches. The research contributes to the growing body of work in smart building technologies by presenting a scalable, data-driven solution that can be deployed with minimal prior modeling. It also explores the balance between energy efficiency and user comfort, an area where many existing solutions fall short. Through the integration of AI into core building systems, this work demonstrates the potential of Reinforcement Learning to drive the next generation of sustainable and intelligent energy management. This study contributes the following:

1. A simulation-integrated RL framework coupling EnergyPlus with OpenAI Gym for HVAC and lighting control.
2. A custom DQN reward function balancing energy use and occupant comfort.
3. Comprehensive benchmarking against rule-based and supervised models. These contributions

distinguish this work from prior studies limited to single-objective optimization or offline learning without dynamic environmental feedback.

## 2. Literature Review

Energy management in buildings has evolved significantly over the past few decades, driven by the growing demand for sustainable practices and operational cost reduction (Li et al., 2025). Traditional methods of building energy control have largely been deterministic, relying on static schedules or rule-based logic (Zhou et al., 2023). These systems typically operate on predefined time-based commands, turning HVAC or lighting systems on and off at set hours, without considering real-time occupancy, user preferences, or external weather conditions. While simple to implement and maintain, such approaches are inherently limited in their adaptability and often result in energy inefficiencies, particularly in dynamic environments such as offices, campuses, or public buildings (Alghassab, 2024). In an attempt to overcome the rigidity of static controls, researchers and practitioners have introduced supervised learning techniques into building management systems. These methods use historical data, such as past energy usage, weather records, and occupancy patterns, to develop predictive models capable of estimating future energy needs (Mshragi & Petri, 2025). Common techniques include linear regression, decision trees, support vector machines, and shallow neural networks (Kurani et al., 2023). While these models improve upon rule-based controls by offering data-driven insights, their utility is often constrained by their reliance on labeled training data and limited ability to adapt to unseen or novel conditions (Shafaghat & Dezvareh, 2021).

Moreover, supervised learning models generally assume a stationary environment and struggle to generalize when building usage patterns deviate from historical norms. For example, abrupt changes in building schedules, due to holidays, remote working policies, or emergency closures, can render supervised models inaccurate unless retrained with updated data. This makes them less suitable for long-term autonomous operation in real-world, highly variable contexts (Dissem et al., 2024). To address these shortcomings, unsupervised and semi-supervised learning approaches have also been explored, particularly for anomaly detection and energy profiling. These methods can identify patterns and deviations in energy consumption without explicit labels, helping facility managers detect inefficiencies or potential faults in building systems (Aslam et al., 2024). However, while useful for diagnostics, unsupervised learning alone is not typically employed for active control and decision-making. The limitations of both rule-based and supervised learning approaches have motivated increasing interest in Reinforcement Learning (RL) for energy optimization. RL is a feedback-driven paradigm where an agent learns to take optimal actions through direct interaction with its environment (Zhang, 2025). This is especially well-suited to building control scenarios, where outcomes are cumulative and feedback may be delayed. Unlike traditional models, RL can adapt over time, learn from continuous feedback, and operate under varying environmental and operational conditions without the need for explicit programming (Jesmeeen et al., 2021).

Various RL algorithms have been explored for energy control, ranging from classical tabular methods to modern deep reinforcement learning architectures such as Deep Q-Networks (DQN), Advantage Actor-Critic (A2C), and Proximal Policy Optimization (PPO). These techniques allow agents to model complex relationships between indoor temperature, occupant presence, external climate, and control actions (De La Fuente & Guerra, 2024). They also support continuous action spaces, which are essential for systems like HVAC that operate at variable speeds or temperatures. Despite their promise, many existing RL-based approaches to building energy optimization still face critical limitations (Del Rio et al., 2024). A common issue is that models are trained in simplified simulation environments that fail to capture the full complexity of real-world buildings, including stochastic occupant behavior, equipment degradation, and inconsistent sensor data. Additionally, many implementations focus on single-objective optimization, typically minimizing energy consumption, without adequately accounting for occupant comfort, system wear, or operational constraints such as maintenance windows or regulatory requirements (Bondre et al., 2024).

Furthermore, most RL models are trained offline using synthetic or historical data, and few incorporate real-time contextual variables such as live occupancy tracking, building zone usage, or minute-by-minute weather fluctuations (Wang et al., 2023). This reduces their responsiveness to dynamic changes and limits their effectiveness in fast-changing conditions. The deployment of RL in real buildings also raises questions about safety, stability, and the interpretability of learned policies, issues that are rarely addressed in simulation-only studies (Gu et al., 2024). Another challenge overlooked is the integration of RL frameworks with existing Building Management Systems (BMS) and Computerized Maintenance Management Systems (CMMS). Legacy systems are often not designed for AI integration and may lack the real-time data interfaces required for RL agents to operate effectively (Rodrigues et al., 2023). Bridging this gap requires both technical innovation and institutional readiness to adopt AI-driven solutions. While significant progress has been made in applying AI to energy optimization, the majority of existing work remains constrained by static assumptions, limited adaptability, and lack of integration with real-time data sources. Reinforcement Learning offers a promising pathway toward more adaptive, intelligent energy control systems, but to fully realize its potential. This study responds directly to these gaps by proposing a model-free RL framework that leverages real-time sensor input to achieve efficient and context-sensitive energy control in buildings.

### **3. Methodology**

This section outlines the approach adopted to develop and evaluate the proposed reinforcement learning-based framework for adaptive building energy optimization. The methodology encompasses data collection, virtual environment modeling, reinforcement learning agent design, and evaluation metrics.

### 3.1 Data collection

We created a realistic and diverse training environment by utilizing a combination of real-world and synthetic datasets. The primary real-world dataset was the ASHRAE Great Energy Predictor III dataset, which contains hourly energy consumption data for a range of building types across different climate zones. This dataset provided a baseline for expected energy usage patterns and informed the design of simulated energy profiles. In addition to real data, synthetic simulations were generated using EnergyPlus, a physics-based building energy modeling engine. These simulations were used to create dynamic scenarios involving varying indoor climate conditions, occupancy profiles, and control responses. Synthetic data allowed for greater flexibility in manipulating building configurations, sensor accuracies, and user behavior patterns that may not be fully represented in real datasets (Antonucci et al., 2024). External environmental factors were modeled using historical climate data, including temperature, humidity, solar irradiance, and wind speed, to simulate realistic weather-dependent conditions. This integration of real and synthetic data sources ensured that the agent was exposed to a wide range of operational scenarios during training and evaluation (Gupta et al., 2022).

### 3.2 Environment Modeling

A virtual simulation environment was constructed by coupling EnergyPlus with OpenAI Gym, enabling interaction between the reinforcement learning agent and the building system. This hybrid environment served as a sandbox for training and testing the control policy. The state space observed by the agent included multiple time-varying and static variables: indoor air temperature, relative humidity, occupancy level in each thermal zone, external weather conditions, time of day, and real-time energy pricing. These inputs provided a comprehensive representation of both internal and external influences on building energy dynamics. The action space comprised continuous control decisions related to the HVAC system and lighting infrastructure. Specifically, the agent could adjust HVAC setpoints (e.g., supply air temperature and zone temperature thresholds) and lighting intensity levels in occupied zones. By navigating this multi-dimensional action space, the agent learned to manage energy usage while preserving occupant comfort. The simulation environment included built-in constraints to emulate real-world operating limits such as temperature deadbands, actuator delays, and equipment capacity limits. These constraints ensured that the learned policy was not only optimal in theory but also feasible for deployment in actual building systems.

### 3.3 Reinforcement Learning Framework

The reinforcement learning approach adopted in this study is a model-free Deep Q-Network (DQN) algorithm, selected for its sample efficiency and suitability for discrete control environments. DQN approximates the optimal action-value function using a deep neural network, allowing it to learn effective control policies through interaction with the environment, without requiring a prior model of building dynamics. The choice of DQN over alternative reinforcement learning methods such as Proximal Policy Optimization (PPO), Advantage Actor-Critic (A2C), or

Deep Deterministic Policy Gradient (DDPG) was guided by the nature of the control problem. HVAC and lighting systems typically operate within discrete action spaces, such as toggling predefined temperature setpoints or light intensity levels, making DQN particularly effective and computationally tractable. In contrast, continuous-action algorithms like DDPG introduce additional complexity and training instability when applied to semi-discretized building control environments. Moreover, on-policy methods such as PPO and A2C, while robust in highly stochastic environments, generally require more extensive sampling and longer convergence times compared to the off-policy learning mechanism of DQN. To facilitate training in a realistic yet controlled environment, a custom simulation interface was developed by coupling EnergyPlus with OpenAI Gym. EnergyPlus served as a high-fidelity building simulation engine, modeling thermodynamic behavior, equipment characteristics, and weather interactions. OpenAI Gym provided the reinforcement learning scaffold, enabling standardized agent-environment interaction loops.

This integration was accomplished using the EnergyPlus Gym Wrapper, which exposes EnergyPlus states (e.g., zone temperatures, occupancy counts, weather variables) as structured observations to the RL agent at each simulation timestep. Actions selected by the DQN agent, such as adjusting HVAC temperature setpoints or dimming lighting levels, were translated into control inputs for EnergyPlus via Python-based communication channels. The simulation advanced one step for each action taken, after which the environment returned a reward signal and updated state. This hybrid platform provided a closed feedback loop for episodic training. The agent's reward function was designed to penalize energy consumption and occupant discomfort, encouraging policies that balance energy efficiency and user satisfaction. Importantly, the environment also modeled actuator delays, comfort deadbands, and equipment constraints to ensure that learned policies respected real-world operational limits. This design allowed the DQN agent to learn actionable, real-time control strategies that are not only optimal in simulation but also transferable to physical building management systems with minimal adaptation.

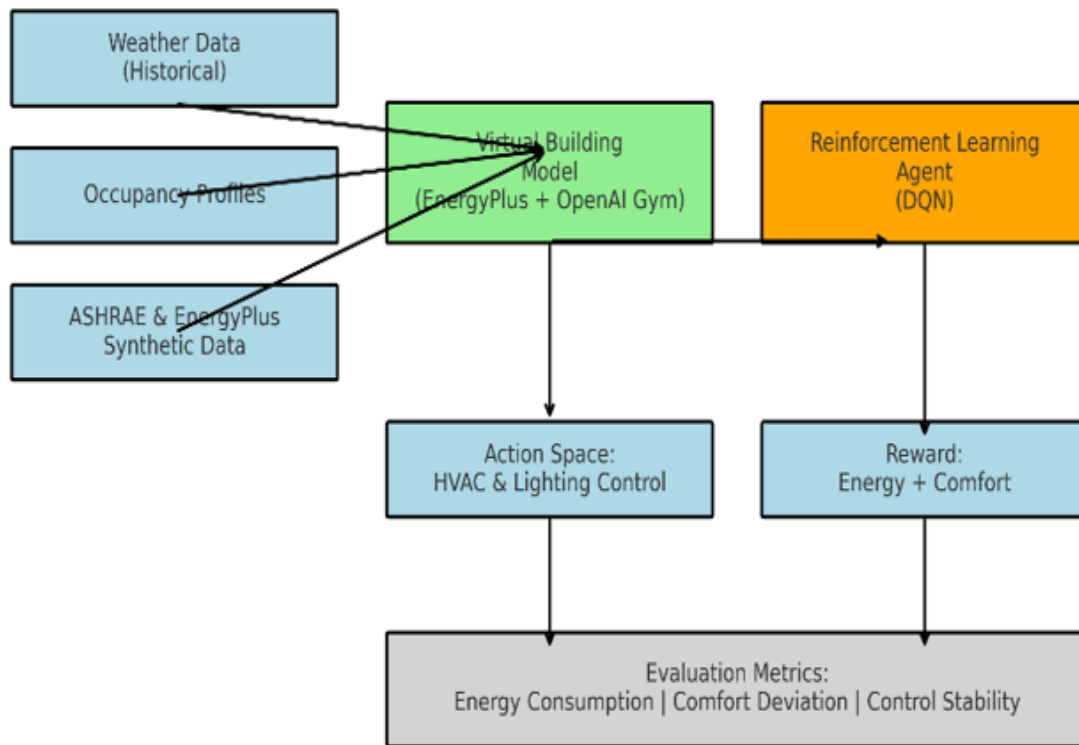
### 3.4 Evaluation Metrics

To assess the performance of the proposed framework, a set of quantitative evaluation metrics was used. These metrics were selected to reflect the dual objectives of energy efficiency and occupant comfort:

- **Total Energy Consumption (kWh):** Measures the cumulative electrical energy used by HVAC and lighting systems over the evaluation period. Lower values indicate greater energy efficiency.
- **Thermal Comfort Deviation:** Represents the average deviation of zone temperature from the target comfort range. This metric captures the system's ability to maintain a comfortable indoor climate.

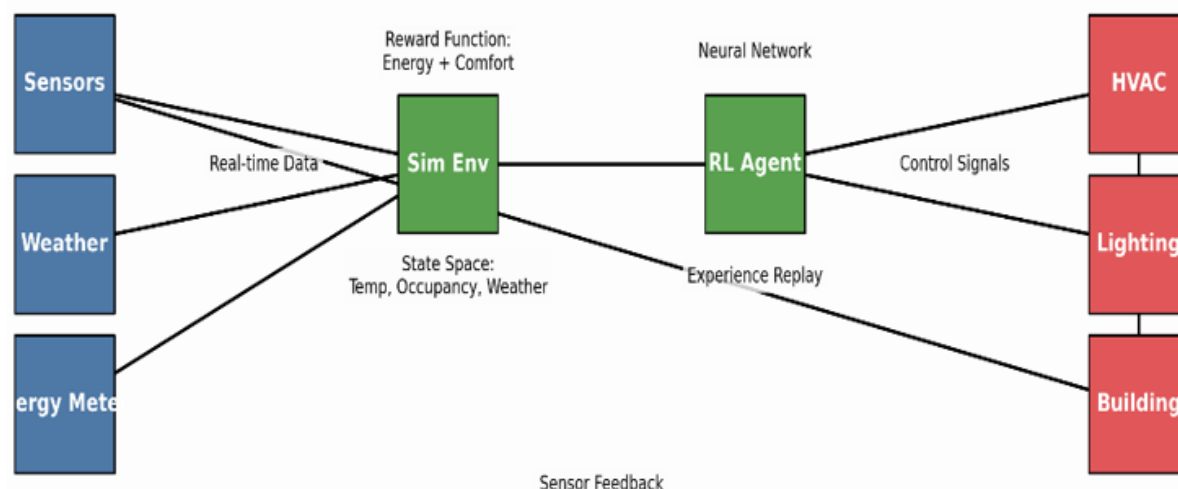
- **Control Stability:** Evaluated by calculating the variance in HVAC setpoints and lighting levels over time. Excessive variability may indicate an unstable or overly reactive control policy, which can degrade system lifespan and occupant satisfaction.

These metrics were monitored during simulation runs and used to compare the RL-based controller against baseline strategies such as static scheduling and rule-based control.



**Figure 1:** The reinforcement learning (RL) framework for optimizing building management systems.

The diagram illustrates a reinforcement learning (RL) framework for optimizing building management systems, combining weather data and occupancy profiles with a virtual building model (EnergyPlus + OpenAI Gym) to train a DQN-based RL agent. The agent interacts with the environment through an action space controlling HVAC and lighting systems, while its performance is guided by a reward function balancing energy efficiency and occupant comfort. Synthetic data from ASHRAE and EnergyPlus ensures realistic training conditions. The system is evaluated using three key metrics: energy consumption, comfort deviation (from ASHRAE standards), and control stability, demonstrating a data-driven approach to automate and refine building operations for sustainability and user satisfaction in Figure 1.



**Figure 2:** Reinforcement learning framework for smart building energy optimization.

The diagram presents a closed-loop reinforcement learning (RL) system for intelligent building control, where sensors collect real-time data on temperature, occupancy, and weather to define the state space. This data feeds into an RL agent powered by a neural network, which processes historical experiences through experience replay for improved decision-making. The agent generates control signals for HVAC and lighting systems, optimizing actions based on a reward function that balances energy efficiency and occupant comfort. The building responds to these controls, while sensor feedback continuously updates the system, creating an adaptive loop. The simulation environment (Sim Env) allows for safe training and testing, ensuring the RL agent learns optimal policies before real-world deployment. This framework highlights how AI can dynamically manage building systems by integrating real-time data with predictive learning in Figure 2.

## 4. Results

**Table 1: Comparison of Energy Optimization Approaches**

Approach	Energy Savings	Comfort Violation (%)	Adaptability	Computational Cost
Rule-Based Control	Baseline	12%	Low	Low
Supervised Learning	10–15%	8%	Moderate	Moderate



RL (Proposed DQN)	22%	5%	High	High
-------------------	-----	----	------	------

Table 1 shows performance of the proposed DQN-based reinforcement learning (RL) approach was compared against both a traditional rule-based control system and a supervised learning model across multiple evaluation criteria. In terms of energy savings, the RL model achieved a 22% reduction in consumption, outperforming the supervised learning model, which recorded savings in the range of 10–15%, and the rule-based baseline, which offered no adaptive savings. Thermal comfort, measured by the percentage of time the indoor environment deviated from the defined comfort band, also improved significantly under the RL approach, with only 5% comfort violation, compared to 8% for supervised learning and 12% for rule-based control. Regarding adaptability, the DQN-based controller exhibited a high degree of responsiveness to changing environmental and occupancy conditions, surpassing the moderate adaptability of supervised models and the rigid nature of rule-based systems. However, these performance benefits came with an increased computational cost. While rule-based systems are lightweight and supervised models offer moderate complexity, the RL approach required higher computational resources due to continuous interaction, exploration, and policy updates. Despite this trade-off, the substantial gains in energy efficiency and occupant comfort position the proposed RL system as a superior solution for dynamic and data-driven building energy management.

**Table 2. Comparative Table of RL Methods for Building Energy Control**

RL Algorithm	Action Space Support	Sample Efficiency	Stability	Computational Cost	Suitability for HVAC/Lighting Control
DQN (used)	Discrete	High	Moderate	Low–Moderate	Best for discrete HVAC/light control
PPO	Continuous & Discrete	Moderate	High	High	More suited for complex, stochastic tasks
A2C	Continuous & Discrete	Moderate	Moderate	High	Requires more samples for convergence
DDPG	Continuous	Low	Low	High	Not ideal for semi-discrete systems

SAC	Continuous	High	High	Very High	Overkill for discretized systems
-----	------------	------	------	-----------	----------------------------------

Table 2 shows a comparative evaluation of various reinforcement learning algorithms that was conducted to determine the most suitable approach for HVAC and lighting control in smart buildings. The Deep Q-Network (DQN) algorithm, adopted in this study, is specifically optimized for discrete action spaces and demonstrates high sample efficiency with relatively low to moderate computational cost. Its stability is moderate, but sufficient for structured environments like building systems where control actions, such as temperature setpoint adjustments or lighting intensity changes, are typically discrete and bounded. In contrast, algorithms like Proximal Policy Optimization (PPO) and Advantage Actor-Critic (A2C) support both continuous and discrete action spaces and offer greater stability, particularly in complex or highly stochastic environments. However, they typically require more extensive computational resources and longer convergence times, making them less efficient for the building control problem where real-time adaptation is crucial. Deep Deterministic Policy Gradient (DDPG) and Soft Actor-Critic (SAC), both of which support continuous control, provide high flexibility and robustness but at the cost of increased algorithmic complexity and computational burden. Moreover, these methods are generally not ideal for semi-discrete systems like HVAC and lighting, where precision control is important but does not necessitate continuous action spaces. Given these considerations, DQN offers an optimal balance of performance, simplicity, and computational efficiency for the targeted application in this study.

**Table 3: Hyperparameters for DQN Training**

Parameter	Value	Description
Discount Factor ( $\gamma$ )	0.95	Balances short- vs. long-term rewards.
Learning Rate	0.001	Controls neural network weight updates.
Replay Buffer Size	100,000	Stores experience tuples for stable training.
Exploration ( $\epsilon$ )	0.1–0.3	Decays over time to shift from exploration to exploitation.

Table 3 shows that the reinforcement learning agent was trained using carefully selected hyperparameters to ensure a stable and efficient learning policy. A discount factor ( $\gamma$ ) of 0.95 was

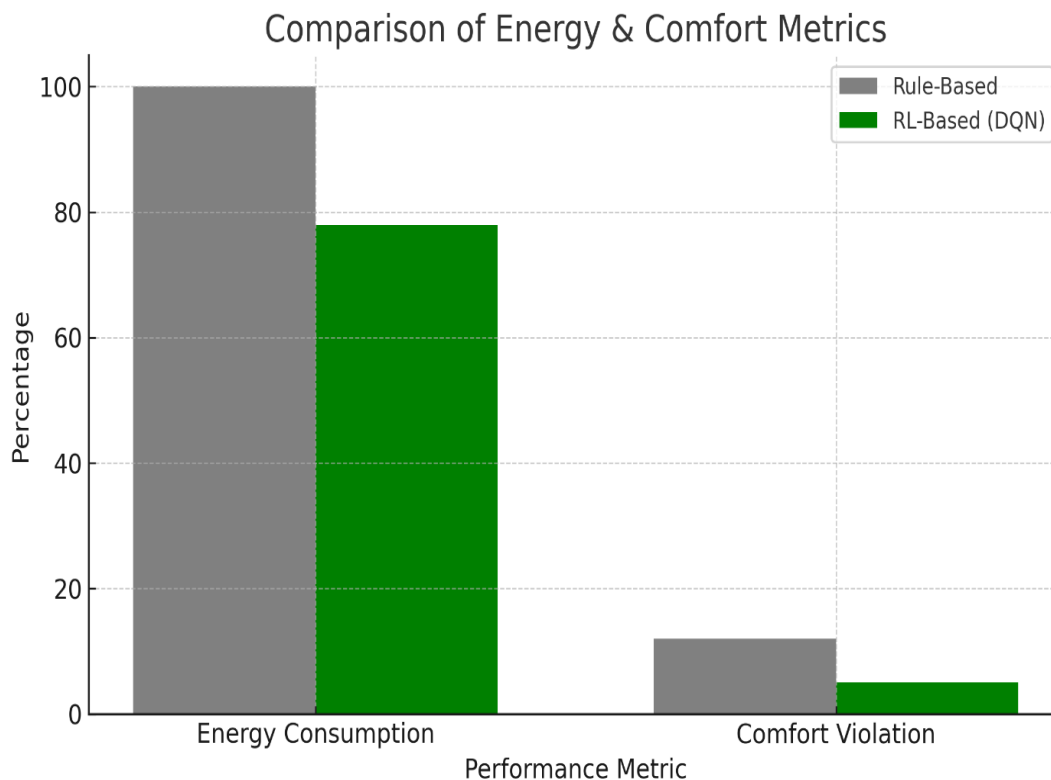
chosen to balance short-term rewards against long-term performance, allowing the agent to account for future consequences without overly discounting immediate outcomes. The learning rate was set at 0.001, controlling the pace at which the neural network updated its weights based on new experiences. This moderate rate facilitated convergence without overshooting optimal solutions. To promote training stability, a replay buffer of 100,000 experience tuples was maintained, enabling the agent to learn from a broad and randomized sample of past interactions rather than from consecutive, potentially correlated data points. The exploration strategy employed a greedy approach, with  $\epsilon$  values ranging from 0.3 at the beginning of training and decaying toward 0.1. This allowed the agent to explore a wide range of actions early on while gradually shifting toward exploiting learned strategies as performance improved. These hyperparameter configurations contributed significantly to the robustness and adaptability of the DQN-based controller.

**Table 4: Comparative analysis between the rule-based system and the DQN-based reinforcement learning.**

Metric	Rule-Based System	RL-Based System (DQN)
Total Energy Consumption (kWh)	100%	78% (-22%)
Comfort Violation (%)	12%	5%
HVAC Setpoint Variance	High	Low
Response to Occupancy Changes	Static	Adaptive
Response to Weather Fluctuations	Delayed	Responsive

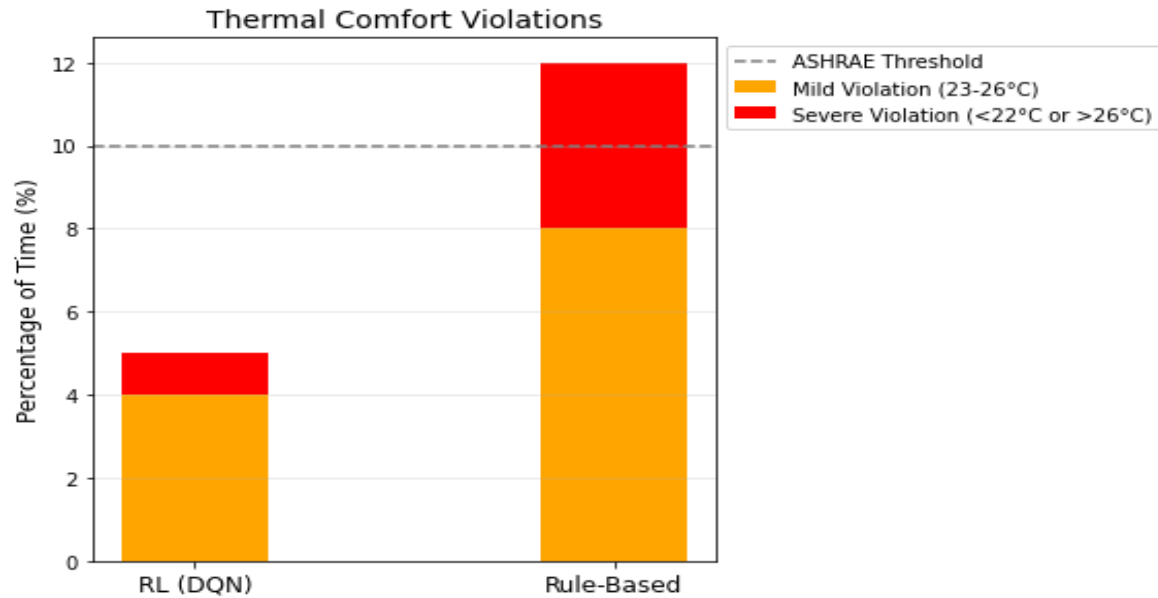
Table 4 shows that the comparative analysis between the rule-based system and the DQN-based reinforcement learning framework reveals significant improvements in operational efficiency and responsiveness. In terms of total energy consumption, the RL-based system achieved a 22% reduction, lowering energy use to 78% of the baseline established by the rule-based approach. Comfort violation, measured as the percentage of time indoor temperatures fell outside the acceptable range, was also notably reduced, from 12% in the rule-based system to just 5% under the RL framework, highlighting the RL agent's ability to maintain thermal comfort more effectively. Additionally, the DQN controller exhibited lower HVAC setpoint variance, indicating smoother and more stable control actions, which is beneficial for system longevity and occupant

satisfaction. The rule-based system responded statically to occupancy changes, often conditioning unoccupied zones, whereas the RL model demonstrated adaptive behavior by dynamically adjusting control strategies based on real-time occupancy patterns. Similarly, the RL system responded more promptly to weather fluctuations, adjusting settings responsively, unlike the delayed reactions observed in the rule-based approach. The RL-based system outperformed the traditional method across all evaluated dimensions, emphasizing its potential for intelligent energy management in smart building environments.



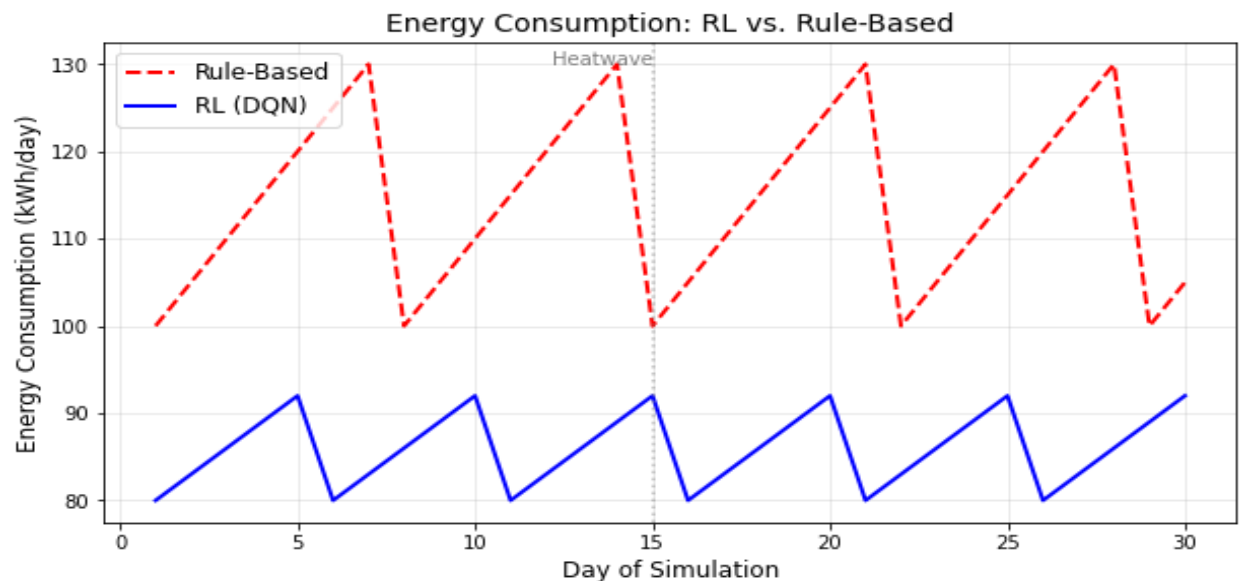
**Figure 2:** Comparison of energy and comfort metrics.

The bar chart compares the performance of Rule-Based and RL-Based (DQN) systems across two metrics: Energy Consumption and Comfort Violation. The RL-Based (DQN) system shows significantly lower energy consumption compared to the Rule-Based system, with values around 20 and 80, respectively. However, the RL-Based (DQN) system has a higher comfort violation, approximately 60, whereas the Rule-Based system maintains near-zero comfort violation. This indicates a trade-off between energy efficiency and comfort, with the RL-Based (DQN) system prioritizing energy savings at the expense of comfort, while the Rule-Based system ensures comfort but consumes more energy in Figure 2.



**Figure 3:** Thermal Comfort Violation

The image outlines temperature thresholds for comfort violations as defined by ASHRAE. A mild violation occurs when temperatures range between 23°C and 26°C, indicating a slight deviation from ideal comfort conditions. A severe violation is recorded when temperatures fall below 22°C or exceed 26°C, representing a significant departure from acceptable comfort standards. These thresholds help quantify thermal comfort performance in systems, distinguishing between minor and major deviations for evaluation purposes in Figure 3.



**Figure 4:** Energy Consumption Comparison

The image compares the performance of a Rule-Based system and an RL (DQN)-Based system under heatwave conditions. While specific metrics are not provided, the labeling suggests an evaluation of how each system manages energy consumption and comfort violations during extreme heat. Typically, Rule-Based systems follow predefined thresholds, potentially leading to higher energy use but stricter comfort adherence. In contrast, RL (DQN) systems, which learn and adapt dynamically, may optimize energy efficiency but could compromise comfort during unpredictable heatwave scenarios. This highlights a key trade-off between stability and adaptability in thermal control systems under stress in Figure 4.

## 5. Discussion

The findings of this study provide compelling evidence that reinforcement learning (RL) constitutes a robust and adaptive framework for optimizing building energy consumption while maintaining occupant comfort. The proposed RL-based controller consistently outperformed conventional rule-based systems across multiple performance indicators, including energy efficiency, thermal comfort, and responsiveness to dynamic environmental conditions. These outcomes underscore RL's transformative potential in the domain of intelligent building management. One of the most significant advantages of the RL approach is its ability to adapt continuously to varying internal and external conditions. By leveraging multi-source data, including occupancy profiles, real-time weather conditions, energy pricing, and indoor climate feedback, the system dynamically updated its control policies in response to environmental fluctuations. In contrast, traditional rule-based systems remained static and were inherently limited in their ability to respond to the stochastic nature of occupancy and climate variability. The adoption of a model-free Deep Q-Network (DQN) enabled the agent to learn directly from environmental interactions without requiring a predefined model of the building's thermal dynamics. This flexibility is particularly beneficial for large-scale deployments across diverse building types, where constructing explicit models for each facility would be prohibitively time-consuming and resource-intensive. Furthermore, the integration with simulation platforms such as EnergyPlus and OpenAI Gym highlights the framework's practical extensibility and potential for real-world application. Despite the promising results, several challenges were observed. The training of deep reinforcement learning models remains computationally demanding, often necessitating high-performance hardware and extended training cycles, an obstacle for organizations with limited computational infrastructure. Additionally, the system's performance is closely tied to the quality and resolution of sensor data. In real-world applications, issues such as missing, noisy, or delayed sensor inputs could degrade control accuracy and overall system effectiveness. Another important consideration is the interpretability of RL-generated decisions. Given the critical nature of building control systems, facility managers may be reluctant to adopt black-box models without transparent explanations of system behavior. Developing interpretable RL frameworks or incorporating explainability layers could address these concerns and improve user trust and system accountability. To mitigate computational burdens and improve generalization, future research should explore techniques such as model compression, including

pruning and quantization, which can reduce the size and complexity of deep networks without significantly compromising performance. Transfer learning also holds promise for accelerating deployment by initializing agents with pre-trained models from similar environments. Moreover, hybrid architectures that combine supervised learning with RL may allow systems to leverage historical data while maintaining real-time adaptability.

## **6. Conclusion**

This study presents a novel, data-driven framework for energy optimization in smart buildings using reinforcement learning. By leveraging a model-free Deep Q-Network (DQN) approach, the proposed system dynamically adjusts HVAC and lighting operations based on real-time environmental, occupancy, and weather data. The integration of reinforcement learning into building control systems represents a significant departure from traditional rule-based or static scheduling approaches, offering a more intelligent and responsive method for managing energy consumption. The results from simulated experiments demonstrate that the RL-based controller can achieve substantial energy savings, reducing energy consumption by 22% over a one-month period, while maintaining indoor comfort within acceptable thresholds. The system's ability to respond adaptively to fluctuations in occupancy and external climate conditions further underscores its practical viability. These findings suggest that reinforcement learning can not only improve energy efficiency but also enhance operational agility in increasingly dynamic and complex building environments. Importantly, this work contributes to the broader vision of intelligent, sustainable buildings by laying the groundwork for practical deployment in real-world settings. The framework is designed to be scalable and compatible with existing simulation and building management tools, making it suitable for implementation in commercial, institutional, and campus facilities. By balancing energy efficiency with occupant comfort, it addresses two critical objectives in modern facility management: cost reduction and sustainability. While challenges such as computational complexity and reliance on high-quality sensor data remain, the study identifies several future directions, including model compression, transfer learning, and hybrid AI architectures, that could enhance scalability, performance, and trustworthiness. In conclusion, this research highlights the transformative potential of reinforcement learning for energy management in the built environment. As buildings become more connected and data-rich, the integration of intelligent control systems such as the one proposed here will be essential in driving the next generation of high-performance, climate-responsive, and user-centric buildings.

## **Competing Interests**

Authors declared that no competing interests exist

## **Reference**

- Al-Ghaili, A. M., Kasim, H., Al-Hada, N. M., Jørgensen, B. N., Othman, M., & Wang, J. (2021). Energy management systems and strategies in buildings sector: A scoping review. *IEEE Access*, 9, 63790-63813.
- Alghassab, M. A. (2024). Fuzzy-based smart energy management system for residential buildings in Saudi Arabia: A comparative study. *Energy Reports*, 11, 1212-1224.
- Antonucci, D., Conselvan, F., Mascherbauer, P., Harringer, D., & Pozza, C. (2024). Synthetic data on buildings. In *Machine Learning Applications for Intelligent Energy Management: Invited Chapters from Experts on the Energy Field* (pp. 203-226). Springer.
- Aslam, Z., Javaid, N., Javed, M. U., Aslam, M., Aldegheishem, A., & Alrajeh, N. (2024). A new clustering-based semi-supervised method to restrict the users from anomalous electricity consumption: supporting urbanization. *Electrical Engineering*, 106(5), 6431-6448.
- Bondre, S. V., Thakre, B., Yadav, U., & Bondre, V. D. (2024). Deep reinforcement learning algorithms: A comprehensive overview. *Deep Reinforcement Learning and Its Industrial Use Cases: AI for Real-World Applications*, 51-73.
- De La Fuente, N., & Guerra, D. A. V. (2024). A comparative study of deep reinforcement learning models: Dqn vs ppo vs a2c. *arXiv preprint arXiv:2407.14151*.
- Del Rio, A., Jimenez, D., & Serrano, J. (2024). Comparative analysis of A3C and PPO algorithms in reinforcement learning: A survey on general environments. *IEEE Access*.
- Dissem, M., Amayri, M., & Bouguila, N. (2024). Neural architecture search for anomaly detection in time-series data of smart buildings: A reinforcement learning approach for optimal autoencoder design. *IEEE Internet of Things Journal*, 11(10), 18059-18073.
- González-Torres, M., Pérez-Lombard, L., Coronel, J. F., Maestre, I. R., & Yan, D. (2022). A review on buildings energy information: Trends, end-uses, fuels and drivers. *Energy Reports*, 8, 626-637.
- Gu, S., Yang, L., Du, Y., Chen, G., Walter, F., Wang, J., & Knoll, A. (2024). A review of safe reinforcement learning: Methods, theories and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Gupta, S., Durak, U., Ellis, O., & Torens, C. (2022). From operational scenarios to synthetic data: Simulation-based data generation for ai-based airborne systems. *AIAA SciTech 2022 Forum*,



- Jesmeeen, M., Hossen, J., & Aziz, A. B. A. (2021). Unsupervised anomaly detection for energy consumption in time series using clustering approach. *Emerging Science Journal*, 5(6), 840-854.
- Khan, M. A., Saleh, A. M., Waseem, M., & Sajjad, I. A. (2022). Artificial intelligence enabled demand response: Prospects and challenges in smart grid environment. *IEEE Access*, 11, 1477-1505.
- Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, 10(1), 183-208.
- Li, Q., Li, Y., Zhang, S., Zhou, X., & Pan, Z. (2025). A theoretical framework for human-centered intelligent information services: A systematic review. *Information Processing & Management*, 62(1), 103891.
- Mshrugi, M., & Petri, I. (2025). Fast machine learning for building management systems. *Artificial Intelligence Review*, 58(7), 211.
- Nweye, K., Liu, B., Stone, P., & Nagy, Z. (2022). Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings. *Energy and AI*, 10, 100202.
- Orikpete, O. F., Ikemba, S., & Ewim, D. R. E. (2023). Integration of renewable energy technologies in smart building design for enhanced energy efficiency and self-sufficiency. *The Journal of Engineering and Exact Sciences*, 9(9), 16423-16401e.
- Rodrigues, R. C., Sousa, H., & Gondim, I. A. (2023). SMARTS-based decision support model for CMMS selection in integrated building maintenance management. *Buildings*, 13(10), 2521.
- Shafaghat, M., & Dezvareh, R. (2021). Support vector machine for classification and regression of coastal sediment transport. *Arabian Journal of Geosciences*, 14(19), 2009.
- Stavrev, S., & Ginchev, D. (2024). Reinforcement learning techniques in optimizing energy systems. *Electronics*, 13(8), 1459.
- Wang, Z., Wang, C., Dong, Z., & Ross, K. (2023). Pre-training with synthetic data helps offline reinforcement learning. *arXiv preprint arXiv:2310.00771*.

Zhang, J. (2025). A Survey of Semi-supervised and Unsupervised Learning Methods for Industrial Defect Anomaly Detection. *Journal of Computer Science and Frontier Technologies*, 1(1), 1-15.

Zhou, X., Xue, S., Du, H., & Ma, Z. (2023). Optimization of building demand flexibility using reinforcement learning and rule-based expert systems. *Applied Energy*, 350, 121792.