

Optimal Control of a Data Center Chilled Water Plant with Deep Q-Network[#]

Wanpeng Zhang¹, Guanjing Lin^{1*}, Huaiyu Yang¹

¹ The Institute for Future Human Habitat, Shenzhen International Graduate School, Tsinghua University

(Corresponding Author: linguanjing@sz.tsinghua.edu.cn)

ABSTRACT

The optimal control of chilled water (CHW) plants is critical for enhancing the energy efficiency of data centers. The effectiveness of traditional rule-based control (RBC) largely relies on the expertise of control engineers, while the application of model-based predictive control (MPC) is constrained by its dependence on extensive historical data and highly accurate models. Recently, reinforcement learning (RL)-based control has attracted increasing research attention due to its characteristics of adaptive control. However, RL-based control studies in HVAC domain are mainly for air-side systems, with relatively limited exploration of CHW plants. To address this gap, this paper evaluated the application of Deep Q-Network (DQN) for the cooling water system optimal control in a CHW plant. A detailed simulation environment model was developed using the real data from a data center CHW plant, and the DQN algorithm was subsequently assessed within this environment. The results showed that, compared with RBC, DQN-based control achieved 15.4% monthly energy savings, closely approaching the 15.9% energy savings attained by MPC. Moreover, the control actions generated by DQN-based control converged toward patterns similar to those of MPC. These findings suggested that DQN-based optimal control holds strong potential to improve energy performance for CHW plants without sufficient historical data.

Keywords: deep Q-network, optimal control, chilled water plant, energy saving

NONMENCLATURE

Abbreviations

RL	Reinforcement Learning
RBC	Rule Based Control
MPC	Model-based Predictive Control
AI	Artificial Intelligence
DQN	Deep Q Network

COP	Coefficient of Performance
ANN	Artificial Neural Network
AHU	Air Handling Unit
CHW	Chilled Water
<i>Symbols</i>	
state_dim	State Dimension
action_dim	Action Dimension
Hidden_size	Number of Hidden Layers
Relu	Activation Function
Adam	Optimizer
CL	Cooling Load
CH _{chwst}	Temperature of the Chilled Water Supplied by the Chiller to (AHU)
CH _{chwrt}	Water returned by the Chiller from AHU
CT _{cwrt}	Temperature of the Cooling Water Supplied by the Cooling Tower to the Chiller
OA _{wbt}	Outdoor Air Wet Bulb Temperature
plr	Partial Load Ratio
f _{cw}	Frequency of the Cooling Water Pump
f _{ct}	Frequency of the Cooling Tower
M _{chw}	Mass Flow Rate of the Chilled Water Pump
M _{cw}	Mass Flow Rate of the Cooling Water Pump
P _{cwp}	Power of the Cooling Water Pump
P _{chwp}	Power of the Chilled Water Pump
P _{ct}	Power of the Cooling Tower

1. INTRODUCTION

With the rapid advancement of artificial intelligence (AI) in recent years, data centers have become a critical component of modern infrastructure. A substantial portion of their energy consumption arises from cooling systems, which are essential for dissipating heat generated by IT equipment. In 2022, global data centers consumed an estimated 240–340 terawatt-hours (TWh) of electricity, representing approximately 1–1.3% of

[#] This is a paper for the 17th International Conference on Applied Energy (ICAE2025), December 8-12, 2025, Bangkok, Thailand.

worldwide electricity use [1]. In the United States, data centers accounted for about 4.4% of national electricity consumption in 2023, with forecasts suggesting this share may increase to 6.7–12% by 2028 [2]. Within data centers, HVAC systems typically consume 30–40% of total energy [3]. Chilled water (CHW) plants account for the highest share of energy use of HVAC systems at traditional data centers. Consequently, enhancing the energy efficiency of CHW plants is of critical importance to achieve overall energy conservation while ensuring reliable thermal requirement for data centers.

Currently, approaches for optimizing data center CHW plant operations can be broadly classified into three categories: rule-based control (RBC), model-based predictive control (MPC), and model-free control. RBC is characterized by its simplicity and low implementation cost. However, it lacks flexibility under dynamic loads and complex environmental conditions, and its performance strongly depends on control engineer’s experience and the quality of predefined rules [4]. MPC predicts future system states using physical or mechanistic models to optimize control strategies, thus

[12] used Q-learning to optimize the cooling load distribution among chillers, resulting in 4.6% energy savings. Considering that Q-learning method discretizes both state and action spaces, its performance in the CHW plant optimization may be degraded if the state and action space dimensions are too high, leading to excessive computational complexity and slow convergence. Recent studies started to investigate the potential of DQN in CHW plant optimization. Wang et al. [13] applied DQN to control a CHW plant’s operation modes, achieving energy savings of 8.7–8.9% on selected operation days and 11.8% across a cooling season.

This paper applied DQN reinforcement learning approach for optimization of the cooling water system in a CHW plant and compared its performance with RBC and MPC.

2. METHODOLOGY

The overall methodology of this study is illustrated in Fig. 1. The four parts are described in Sec 2.1-2.4, respectively.

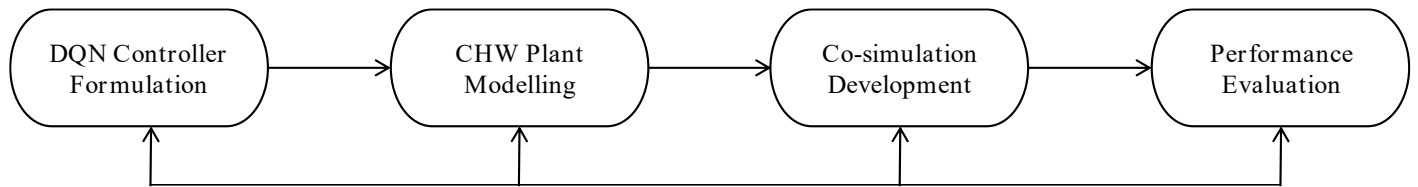


Fig. 1 Flowchart of methodology

can improve cooling efficiency and reduce energy consumption. Yet, its effectiveness relies on highly accurate models, large volumes of historical data, and detailed prior knowledge for model calibration [5-6]. Model-free control based on reinforcement learning (RL) eliminates the need for precise models, instead learns optimal strategies directly from system interactions. This approach offers adaptability to dynamic conditions, effectiveness in handling nonlinear systems, and promise for multi-objective optimization (e.g., energy efficiency, thermal comfort).

Model free reinforcement learning is well suited for addressing control problems as sequential decision-making tasks. Previous RL-based optimal control studies targeted mainly on air-side systems, for example, optimizing AHU valve openings [8], control strategies for variable air volume systems [9], and heating/cooling setpoints [10] for improving building energy efficiency. Several simulation studies have used RL for the optimal control of CHW plants. Zhang et al. [11] applied Q-learning to optimize the condenser water system in a CHW plant, achieving energy savings of 14.7%. Qiu et al.

2.1 Construction of DQN Controller

Reinforcement learning (RL) is a dynamic process that promotes the stable development and optimization of control strategies through trial-and-error interactions between an agent controller and its environment. The DQN algorithm is a RL method that combines Q-learning with deep neural networks to approximate the action-value function, to address the challenges of high-dimensional state and action spaces [14]. Furthermore, by employing a target Q-network and experience replay, DQN alleviates the instability issues in neural network training that arise when dealing with such high-dimensional problems. The DQN-based optimal control in this study follows the framework illustrated in Fig. 2 and consists of the following components.

State Space: It consists of 2 elements: 1) Outdoor wet-bulb temperature (OA_{wbt}), 2) Partial load ratio (plr), defined as the ratio of the system's cooling load to rated cooling capacity.

Action Space: The control parameters are the operating frequencies of the cooling water pump and

cooling tower. Based on on-site recommendations, their

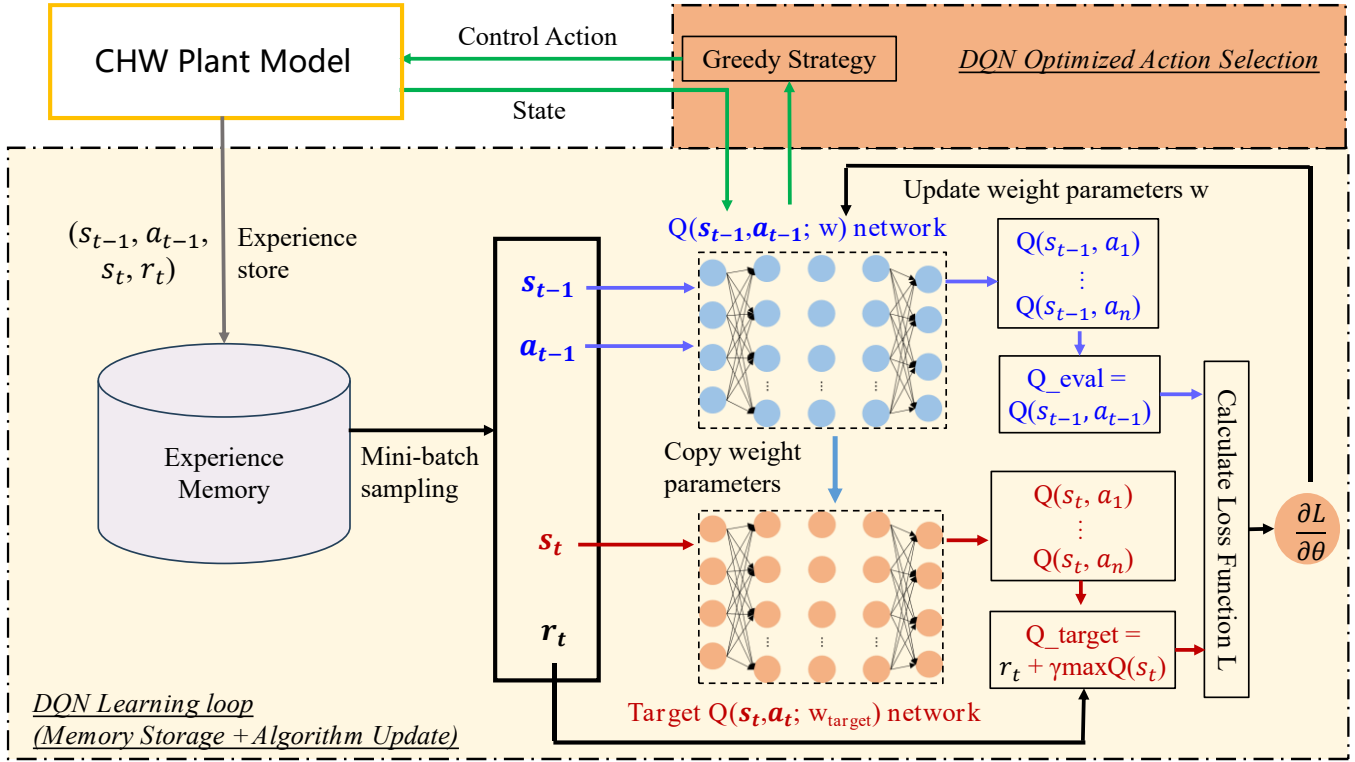


Fig. 2 Co-simulation of CHW plant model and DQN controller

ranges are 25–50Hz. The action interval is 3 Hz for both cooling water pumps and cooling towers. A summary of the action space is provided in Table 1.

Table 1. Action space definition

Control actions	Action space
Cooling water pump	25,28, ... ,49
Cooling tower	25,28, ... ,49

Reward Function: The coefficient of performance (COP) of the cooling water system is used as the optimization objective. Accordingly, the reward function is defined to maximize COP, as expressed in Equation (1).

$$COP_{cws} = \frac{CL}{P_{chillers} + P_{cwps} + P_{towers}} \quad (1)$$

Q Network: In the DQN framework, a multi-layer feedforward neural network is used to approximate the state–action value function. The input layer corresponds to the dimension of the state space (state_dim). Two fully connected hidden layers, each with hidden_size neurons, are employed, with ReLU activation functions enhancing nonlinear representation. The output layer matches the dimension of the action space (action_dim) and estimates Q-values for all candidate actions, which guide the agent's action selection.

The hyperparameters for DQN are shown in Table 2.

Table 2. Hyperparameter settings for DQN

Hyperparameter	Value
State_dim	2
Hidden_size	16
Action_dim	81
Number of neurons in each layer	16
Batch_size	256
Optimizer	Adam
Loss function	MSE
Memory_size	5000
α	0.01
ϵ	0.1
γ	0.01

2.2 CHW plant modelling

This study applies the proposed control method to the cooling water system of a data center CHW plant. The case system consists of six variable-frequency cooling towers, three variable-frequency cooling water pumps, three water-cooled chillers, and three fixed-frequency chilled water pumps. All chillers, cooling towers, and pumps in the system are of identical models.

Operational data from a data center cooling plant was employed to construct a data-driven system model. The model provides the basis for subsequent simulations and validation of optimization control strategies.

(1) The COP model of the chiller and the cooling tower outlet water temperature model were constructed with artificial neural network (ANN). Input variables included cooling load, chilled water supply and return temperatures, chilled water flow rate, and outdoor wet-bulb temperature, as expressed in Equations (2)-(3).

(2) The energy consumption of chilled water pumps, cooling water pumps, and cooling tower fans was characterized by a cubic polynomial function of frequency–power, with parameters fitted to measured data. Chilled water and cooling water flow rates were described by linear regression models, with coefficients obtained from historical data. The corresponding mathematical expressions are given in Equations (4)–(8):

$$\text{COP}_{\text{chiller}} = \text{ANN}(\text{CL}, \text{CH}_{\text{chwst}}, \text{CH}_{\text{chwrt}}, \text{M}_{\text{chw}}, \text{M}_{\text{cw}}) \quad (2)$$

$$\text{CT}_{\text{cwst}} = \text{ANN}(\text{CT}_{\text{chwrt}}, f_{\text{cw}}, \text{OA}_{\text{wbt}}, \text{M}_{\text{cw}}) \quad (3)$$

$$P_{\text{cw}} = a_0 + a_1 \times f_{\text{cw}} + a_2 \times f_{\text{cw}}^2 + a_3 \times f_{\text{cw}}^3 \quad (4)$$

$$P_{\text{chwp}} = 46.2 \text{ kW} \quad (5)$$

$$P_{\text{ct}} = b_0 + b_1 \times f_{\text{ct}} + b_2 \times f_{\text{ct}}^2 + b_3 \times f_{\text{ct}}^3 \quad (6)$$

$$\text{M}_{\text{cw}} = c_0 + c_1 \times f_{\text{cw}} \quad (7)$$

$$\text{M}_{\text{chw}} = 653.0 \text{ kg/s} \quad (8)$$

CL represents the cooling load of the chiller, CH_{chwst} denotes the temperature of the chilled water supplied by the chiller to the AHU, CH_{chwrt} indicates the temperature of the chilled water returned from the AHU to the chiller, CT_{chwrt} signifies the temperature of cooling water returning from the chiller to the cooling tower, OA_{wbt} is the outdoor wet-bulb temperature, f_{cw} is the cooling water pump frequency, f_{ct} is the cooling tower frequency, and other parameters are model fitting coefficients.

The training results are presented in Table 3. All models achieved R^2 values above 0.9 and CVRMSE values below 10%, indicating acceptable performance.

In RBC, both the cooling water pumps and cooling towers operate at a fixed frequency of 50 Hz. In MPC, their frequencies are optimized at each time step by solving an energy minimization problem.

Table 3. Model performance evaluation results

Model	R^2	CVRMSE
$\text{COP}_{\text{chiller}}$	0.979	2.41%
CT_{cwst}	0.984	7.13%
P_{cwp}	0.993	2.14%
P_{ct}	0.983	5.24%

2.3 Co-Simulation Interaction

The co-simulation model first reads wet-bulb temperature and cooling load data from an excel file, then publishes this data to the DQN controller and the CHW plant model. After the controller determines the

actions, the results are transmitted to CHW plant model, which then executes the actions and generates the results of reward. The reward is then passed back to the controller to update the Q-network. This process is repeated until all state data in the Excel file is processed, constituting one episode. During the whole episode, all three chillers operated continuously to satisfy the data center's cooling load.

2.4 Convergence Criteria

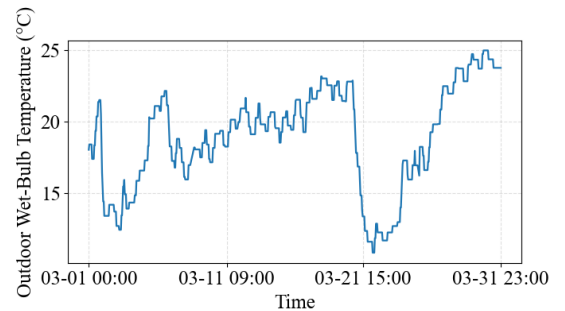
The RL convergence condition is defined as a change rate of less than 0.02% in the moving average of energy consumption across five consecutive episodes. Specifically, at time t , the relative change rate between each pair of adjacent moving averages within the most recent five episodes must remain below the 0.02% threshold, as shown in Equations (9)–(11).

$$\bar{E}_t = \frac{1}{5} \sum_{i=0}^4 E_{t-i} \quad t \geq 5 \quad (9)$$

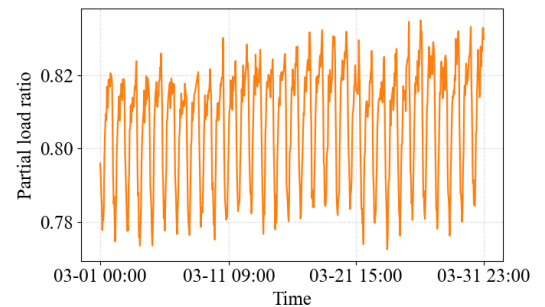
$$\Delta_t = \frac{|\bar{E}_t - \bar{E}_{t-1}|}{\bar{E}_{t-1}} \times 100\% \quad (10)$$

$$\Delta_{t-k} \leq 0.02\% \quad \forall k \in \{0,1,2,3,4\} \quad (11)$$

where \bar{E}_t denotes the average energy consumption over the latest five episodes ending at time t . Δ_t represents the relative change rate between adjacent moving average values.



(a) Outdoor wet-bulb temperature



(b) Partial load ratio

Fig. 3 Outdoor wet-bulb temperature and partial load ratio data

3. CASE STUDY

3.1 Experiment Description

The operation of the CHW plant from one month in 2021 was simulated on an hourly basis under RBC, MPC and RL control. During this period, the outdoor wet-bulb temperature ranged from 11°C to 25°C, and the partial load ratio varied between 0.76 and 0.82, as shown in Fig. 3. The total energy consumption of the CHW plant under RBC, MPC, and DQN-based control was compared to assess the effectiveness of the DQN controller.

3.2 Results Analysis

The simulation results indicate that the monthly energy consumption of the RBC controller is 814.7 kWh, whereas the MPC controller consumed 685.2 kWh, corresponding to an energy savings rate of 15.9%. The subsequent section presents the performance analysis of the DQN controller. Table 4 shows the energy performance of CHW plant under DQN-based control during the first five episodes.

Table 4. Energy consumption and energy savings

Training time	Energy consumption (kWh)	Energy savings (%)
Episode 1	708.2	13.1
Episode 2	698.3	14.3
Episode 3	699.1	14.2
Episode 4	698.8	14.2
Episode 5	698.8	14.2

Table 4 indicates that energy savings were realized in the first episode with DQN-based control, achieving a savings rate of 13.1%. From the first to the second episode, it increased to 14.3%, while energy consumption decreased to 698.3 kWh. Thereafter, the DQN controller's performance was more stabilized, with energy savings varied around 14.2 and energy consumption fluctuating between 698.8 kWh - 699.1 kWh during this period.

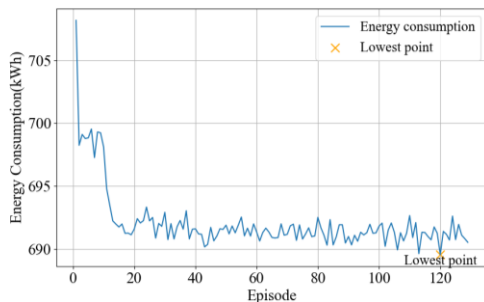


Fig. 4 Long-term energy performance of the DQN controller

The long-term energy performance of the DQN controller is shown in Fig. 4. The energy consumption exhibits a rapid decline during the first 20 episodes, followed by minor oscillations below 692.5 kWh.

According to the convergence criteria defined in equations (9)–(11) in this paper, the algorithm converged at the 17th episode with an energy consumption of 691.27 kWh and an energy savings rate of 15.2%. After convergence, the energy savings rate gradually approached 15.4% at episode 119, with the energy consumption reduced to the lowest level at 689.6 kWh.

Fig. 5 illustrates that during episode 119 the control actions generated by DQN controller converged toward patterns similar to those of MPC.

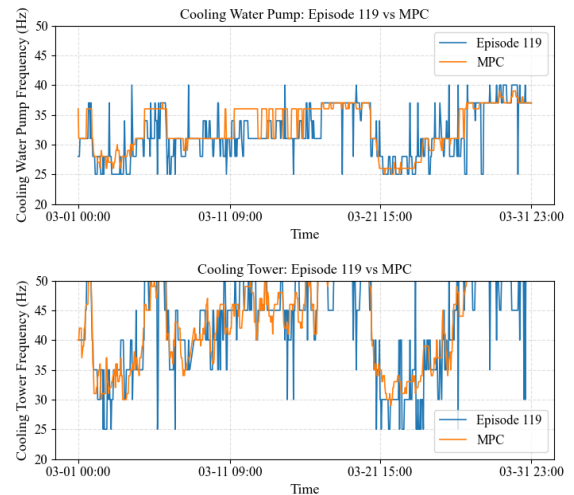


Fig. 5 Comparison of control actions between the DQN-based control and MPC

4. CONCLUSION AND FUTURE WORK

Effective control of data center CHW plant is very important for improving overall energy efficiency. This study investigated the implementation of DQN, a RL method that integrates Q-learning with deep neural networks, for optimizing the cooling water system control of a data center CHW plant. Using RBC as a reference, the DQN controller achieved 13.1% energy savings in the first episode. Over long-term operation, its performance improved, ultimately approaching energy savings of 15.4%, slightly below the 15.9% achieved by MPC. These findings demonstrate that it is promising to apply DQN for energy-efficient CHW plant control when there is lack of historical data.

Future research will focus on extending the approach to additional control variables and CHW plants with other configurations, as well as improving the convergence speed of the RL algorithm. These efforts aim to further enhance the applicability of RL-based optimal control in real-world CHW plant systems.

ACKNOWLEDGEMENT

This work was supported by scientific research start-up funds grant QD2024005C from Tsinghua Shenzhen International Graduate School, Tsinghua University, China.

REFERENCE

- [1] International Energy Agency. (2023). *Data centres and data transmission networks*. Retrieved from <https://www.iea.org/reports/datacentres-and-data-transmission-networks>.
- [2] Shehabi, A., Marston, L., and Greenberg, S. (2023). *U.S. data center energy usage report* (CRS Report No. R48646). Congressional Research Service. Retrieved from <https://crsreports.congress.gov/product/pdf/R/R48646>.
- [3] Capozzoli, A., and Primiceri, G. Cooling systems in data centers: state of art and emerging technologies. *Energy Procedia* 2015;83:484-493.
- [4] Adegbenro, A., Short, M., and Angione, C. An integrated approach to adaptive control and supervisory optimisation of HVAC control systems for demand response applications. *Energies* 2021;14(8):2078.
- [5] Gao, C., and Wang, D. Comparative study of model-based and model-free reinforcement learning control performance in HVAC systems. *Journal of Building Engineering* 2023;74:106852.
- [6] Grahovac, M., Ehrlich, P., Hu, J., and Wetter, M. Model-based data center cooling controls comparative co-design. *Science and Technology for the Built Environment* 2024;30(4):394-414.
- [7] Moriyama, T., De Magistris, G., Tsubori, M., Pham, T. H., Munawar, A., and Tachibana, R. (2018, October). Reinforcement learning testbed for power-consumption optimization. 2018 AsiaSim (Asian simulation conference), 2018 30 October – 1 November, Singapore, Singapore.
- [8] Fan, Y., Fu, Q., Chen, J., Wang, Y., Lu, Y., and Liu, K. A deep reinforcement learning control method for multi-zone precooling in commercial buildings. *Applied Thermal Engineering* 2025;260:124987.
- [9] Wang, H., Chen, X., Vital, N., Duffy, E., and Razi, A. Energy optimization for HVAC systems in multi-VAV open offices: A deep reinforcement learning approach. *Applied Energy* 2024;356:122354.
- [10] Deng, X., Zhang, Y., and Qi, H. Towards optimal HVAC control in non-stationary building environments combining active change detection and deep reinforcement learning. *Building and environment* 2022;211:108680.
- [11] Zhang W, Lin G, and Yang H. Reinforcement learning-based optimal control of condenser water system in chiller plant[C]. 2025 ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) Annual Conference, 2025 21-25, June, Phoenix, Arizona, USA.
- [12] Qiu, S., Li, Z., Li, Z., and Zhang, X. Model-free optimal chiller loading method based on Q-learning. *Science and Technology for the Built Environment* 2020;26(8):1100-1116.
- [13] Wang, X., Yang, H., Sun, Y., Tan, J., Chen, J., Kang, X., and Yan, D. Cooling water system control of a commercial complex based on reinforcement learning. International IBPSA (International Building Performance Simulation Association) Building Simulation Conference, 2023 4-6, September, Shanghai, China.
- [14] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... and Hassabis, D. Human-level control through deep reinforcement learning. *Nature* 2015;518(7540):529-533.