

AI-empowered online control optimization for enhanced efficiency and robustness of building central cooling systems

Lingyun Xie^a, Kui Shan^{a,b}, Hong Tang^{a,*}, Shengwei Wang^{a,b,*}

^a Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

^b Research Institute of Smart Energy, The Hong Kong Polytechnic University, Kowloon, Hong Kong

ARTICLE INFO

Keywords:

Optimal control
Artificial intelligence
Air-conditioning
Energy efficiency
Buildings

ABSTRACT

Adopting Artificial Intelligence for optimizing building system controls has gained significant attention due to the growing emphasis on building energy efficiency. However, substantial gaps remain between academic research and the practical implementation of AI-based algorithms. Key factors hindering implementation include computational efficiency requirements and concerns about reliability in online applications. This paper addresses these challenges by presenting AI-empowered online control optimization technologies designed for practical implementation. A simplified deep learning-enabled Genetic Algorithm is developed to accelerate optimization processes, ensuring optimization intervals are short enough for online applications. This algorithm also significantly reduces CPU and memory usage, enabling deployment on miniaturized control station for field implementation. To enhance stability and reliability, a robust assurance scheme is introduced, which switches to expert knowledge-based control under abnormal conditions. Hardware-in-the-loop tests validate the proposed strategy's computation efficiency, control performance and operational robustness using a physical smart station controlling a simulated real-time dynamic cooling system. Test results show that the optimal control strategy achieves 7.66 % energy savings and exhibits strong operational robustness.

1. Introduction

Global energy consumption has emerged as a pressing concern in recent years due to its environmental and economic implications. To achieve a sustainable future, innovative solutions must be explored and implemented to mitigate the adverse effects of energy consumption. Buildings serve residential, commercial and institutional purposes, demanding substantial energy usage to facilitate the operation of heating, cooling, lighting and other facilities. According to statistics from the International Energy Agency, buildings account for approximately 40 % of global energy consumption [1]. In Hong Kong, this percentage exceeds 80 %, with nearly 40 % attributed to air conditioning systems [2]. Hence, implementing efficient control optimization in air conditioning systems is essential for conserving energy and reducing the environmental impacts of excessive energy consumption.

Traditionally, building control methods have relied predominantly on simple on/off control and PID control techniques. These methods regulate systems by activating/deactivating actuators or minimizing tracking errors based on fixed or manually set points [3]. However, these methods face limitations due to the continuously changing building

loads and operating conditions [4]. The lack of online tuning and monitoring presents significant challenges in adapting to changing operating conditions. Additionally, their simplistic logic is insufficient for achieving optimal performance in nonlinear systems [5].

Researchers have developed a range of mathematical optimization methods to address control problems, including dynamic programming [6], quadratic programming [7], linear programming [8], among others. Additionally, specialized optimization methods have been developed to address specific challenges, adapt to environmental changes, and meet real-time control requirements. Escobar et al. [9] utilized advanced fuzzy-logic-based controllers to dynamically adjust HVAC settings in buildings, optimizing comfort and efficiency by processing real-time contextual data such as occupancy and weather conditions. Baptiste et al. [10] proposed a temperature adaptive control strategy (TACS) that optimizes the use of natural cooling by adjusting the air temperature throughout the year, allowing the system to maximize environmental conditions for cooling. Zhang et al. [11] introduced a hybrid algorithm to optimize indoor air supply parameters. They employed multivariate regression analysis to halve the variable space and used a fuzzy logic controller (FLC) to achieve a 35.7 % reduction in computational costs. Zhang et al. [12] implemented Model Predictive Control (MPC) in a

* Corresponding authors at: Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

E-mail addresses: honghong.tang@polyu.edu.hk (H. Tang), beswwang@polyu.edu.hk (S. Wang).

Nomenclature		<i>SD</i>	Standard deviation
<i>ANN</i>	Artificial Neural Network	<i>SMA</i>	Simple moving average
<i>BAS</i>	Building Automation System	<i>T</i>	Temperature, °C
<i>COP</i>	Coefficient of Performance	<i>W</i>	Power consumption, kWh
<i>DDC</i>	Direct Digital Control	<i>Subscripts</i>	
<i>EER</i>	Energy efficiency ratio	<i>chws</i>	Supply chilled water
<i>FLOPs</i>	Floating-point operations per second	<i>comp</i>	Compressor
<i>GA</i>	Genetic Algorithm	<i>cw</i>	Cooling water
<i>HVAC</i>	Heating Ventilation Air Conditioning	<i>cwr</i>	Return cooling water
<i>k</i>	Penalty factor	<i>cws</i>	Supply cooling water
<i>M</i>	Volume flow rate, m^3h^{-1}	<i>db</i>	Dry-bulb
<i>N</i>	Number of operational units	<i>sw</i>	Seawater
<i>PLR</i>	Partial load ratio	<i>sws</i>	Supply seawater
<i>Q</i>	Heat exchange capacity, kJ	<i>wb</i>	Wet-bulb

small commercial building to manage space conditioning, commercial refrigeration, and the battery system, resulting in a 12 % annual electricity cost savings.

The advancement of computer science has fostered the integration of artificial intelligence (AI) with previously mentioned methods, enhancing their performance and adaptability to dynamic operational conditions [13]. These AI-enhanced methods utilize machine learning approaches [14] and heuristic algorithms [15] to effectively address system complexity and dynamics, thereby enhancing overall energy efficiency and ensuring operational stability. For example, Ye et al. [16] proposed a heuristic algorithm for the evolutionary optimization of tribal intelligence, demonstrating its overall superiority in terms of optimization accuracy, standard deviation, and calculation time through comparative experiments. The algorithm ultimately achieved energy savings of up to 19.43 % in a central HVAC system. Miao et al. [17] evaluated and compared the performance of ten swarm intelligence optimization algorithms by developing a real HVAC system model. Their study found that the artificial bee colony algorithm outperformed others, achieving an energy saving rate of 24.07 % in the simulation of a typical day. Dai et al. [18] proposed an iterative learning control strategy to optimize cooling distribution during the morning start-up period of an air conditioning system. A reinforcement learning (RL) method was employed to adjust the controller parameters, resulting in a reduction in daily precooling energy consumption by 5.1 % to 17.8 %. Huang et al. [19] optimized the load distribution and operating speeds of fans and pumps in a heat pump system, enhancing computational efficiency through the use of an artificial neural network (ANN) model and achieving an energy savings of 14.8 %. Ma et al. [20] proposed an energy efficiency optimization control strategy for a central chiller system, utilizing simplified linear self-tuning models and a genetic algorithm (GA) to optimize temperature setpoints, achieving a daily energy savings of 2.55 %.

Despite many advanced methods being proposed and demonstrating their energy saving potential, the transition from academic research to practical applications remains a formidable challenge. The primary concern lies in the adaptability and reliability of the optimization method's implementation. Specifically, the performance and efficiency of the optimization algorithm must be effectively balanced to ensure that the control can be executed in real time [21]. Additionally, platforms running AI algorithms must address challenges related to computing power and operational robustness [22]. Previous research has often relied on cloud or central computers to provide adequate computing power [23]. However, these approaches face challenges such as bandwidth limitations, network connectivity issues, and latency [24]. Unstable connections or insufficient bandwidth can result in delays and data loss, ultimately compromising the precise control of building systems. In this scenario, it is crucial to develop a feasible and reliable

approach for implementing AI-empowered optimal control at the field level. This would allow for more stable online monitoring, optimization calculations, and decision support [25].

This research, therefore, proposes a comprehensive set of innovative AI-empowered online control optimization technologies for central cooling systems. The main original contributions of this study include:

- A deep learning-enabled genetic algorithm is proposed to reduce computational complexity, enabling rapid online responses to dynamic conditions, and facilitating field implementation.
- A robust assurance scheme is designed and adopted to ensure the stability and reliability of the AI-empowered control optimization in online application.

The developed optimization algorithm is deployed on a physical smart station built on an embedded single-board computer, demonstrating its practical applicability and effectiveness in real-world scenarios. The proposed control optimization technologies are tested and validated on a simulated dynamic cooling system through hardware-in-the-loop testing.

2. Description of AI-empowered optimization strategy

2.1. Overview of control strategy

Fig. 1 illustrates the generation and control process of the proposed AI-empowered optimization strategy, designed to reduce computational complexity and enhance application robustness. The core concept involves simplifying the AI-based optimization algorithm to enhance its suitability for field applications on miniaturized stations, and integrating an anomaly detection mechanism to maintain the reliability of the optimization process.

The optimization algorithm is a deep learning-enabled genetic algorithm, in which the fitness function's optimization model is trained as a hybrid model to achieve a balance between efficiency and reliability. A physical model developed in Dymola is utilized for comprehensive simulation, generating a complete dataset of system operation. Following data processing, sensitivity analysis, and feature selection, the hybrid model is established and integrated into the genetic algorithm (GA) for optimization calculations. To further enhance system robustness, an anomaly detection mechanism based on a dynamic threshold detection method is incorporated, forming a robust assurance scheme. This mechanism enables automatic switching to an expert knowledge-based scheme whenever anomalies are detected, ensuring continuous and stable operation.

Under conventional control, the Direct Digital Control (DDC) controller follows a fixed setpoint strategy, continuously acquiring

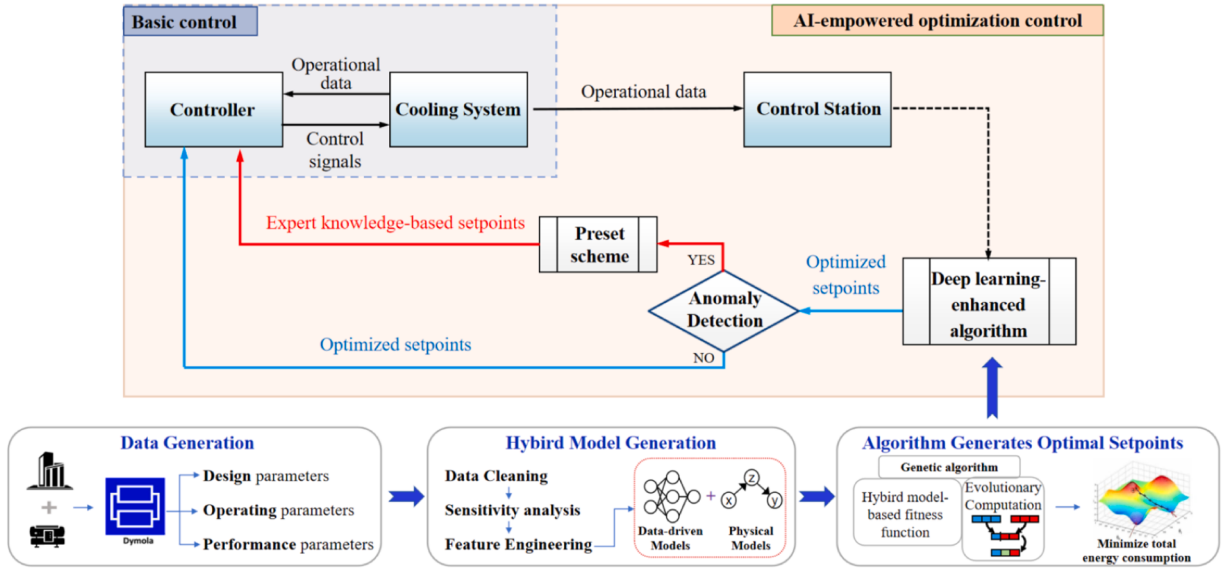


Fig. 1. Generation and control process of the AI-empowered optimization strategy.

cooling system parameters and adjusting pump and chiller operations via PID feedback control. In the optimal control mode, the AI-empowered optimization algorithm is deployed on a control station, where it processes real-time operational data to generate optimized setpoints and transmits them to the DDC controller. The anomaly detection mechanism is embedded within the controller to validate the received setpoints. If no anomalies are detected, the controller updates the setpoints accordingly; otherwise, it switches to a predefined robust assurances scheme to maintain system stability.

The detailed working mechanism of the deep learning-enabled genetic algorithm and the robust assurance scheme is elaborated in Section

2.2.

2.2. Formulation of the AI-empowered optimization strategy

2.2.1. Deep learning-enabled genetic algorithm

One of the main challenges in applying advanced optimization algorithms in the field is meeting computational demands. Traditional optimization algorithms, such as genetic algorithms, typically find optimal results by evaluating the effectiveness of various schemes through repeated assessments and selections. This process consumes significant computing resources and time, making it less practical for

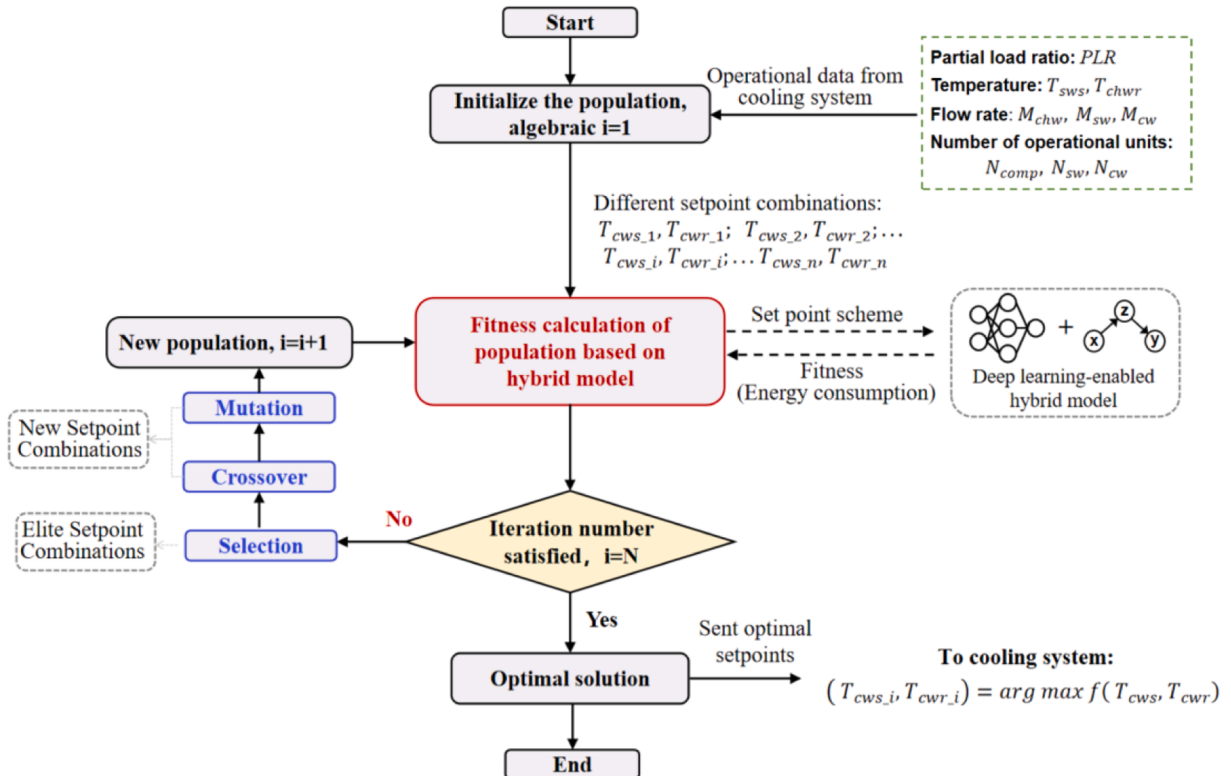


Fig. 2. Flowchart of the deep learning-enabled genetic algorithm.

real-time applications.

The deep learning-enabled genetic algorithm proposed in this paper is a hybrid model-based approach where key nonlinear models of the cooling system are trained as Artificial Neural Network (ANN) models. While pure black-box models can significantly enhance computational efficiency, their lack of interpretability can limit the reliability of applications. The hybrid model, which retains most physical laws, effectively improves adaptability during implementation. The flowchart of the deep learning-enabled genetic algorithm for optimal control is depicted in Fig. 2. Initially, the control station receives real-time operational parameters from the cooling system and generates an initial population, where each individual represents a different setpoint combination. These initial solutions are then fed into the hybrid model, which evaluates the fitness score for each offspring based on energy consumption. The best candidates are selected for evolutionary operations—selection, crossover, and mutation. Finally, the algorithm determines the optimal setpoint combination, enabling the cooling system to operate efficiently. Additionally, the genetic algorithm used in this study combines roulette wheel selection and elitism retention strategies to balance exploration and exploitation, ensuring faster convergence and stable optimization results. Through extensive testing, the following key parameters were selected to ensure efficient convergence of the algorithm: (1) The initial population size is 20; (2) The number of generations is 16; (3) The crossover rate is 0.75; (4) The mutation rate is 0.05.

Through sensitivity analysis, it is found that the power consumption of the chiller is highly correlated with the flow rates of seawater and cooling water, while the power consumption of the pump is significantly influenced by the corresponding control temperature. Based on these findings, this study selects the cooling water supply temperature (T_{cws}) and the cooling water return temperature (T_{cwr}) as the optimization target parameters for the genetic algorithm. The fitness calculation formula is defined in Eq. (1), where ϵ is a small constant to prevent division by zero. The total system energy consumption, denoted as W_{sum} , primarily consisting of the energy consumption of the seawater pump (W_{sw}), cooling water pump (W_{cw}), and chiller unit (W_{comp}). To drive the optimization process toward energy efficiency, the reciprocal of W_{sw} is incorporated into the fitness function, ensuring that higher fitness values correspond to lower energy consumption. During the evolution process of the genetic algorithm, selection, crossover, and mutation operations are iteratively performed, evaluating the fitness of each individual (setpoint combination).

Additionally, a penalty function for the chilled water supply temperature is introduced into Eq. (2) to ensure that the optimization results align with the building's operational requirements. In this function, k indicates the penalty factor, T is the temperature, and the subscripts chw and set represent the chilled water and set value. In this study, the identification of the penalty factor k is based on repeated tuning through simulation tests. The initial range of k was discretized into a grid, and tests were conducted under various load and external environmental conditions. The effect of different k values on the chilled water supply temperature (T_{chws}) was carefully observed. When the setpoint $T_{chws, set}$ was set to 280.15 K, a value of $k = 5 \times 10^{-8}$ was chosen to ensure the chilled water supply temperature does not exceed 281.15 K.

$$W_{sum} = W_{sw} + W_{cw} + W_{comp} \quad (1)$$

$$Fitness = \frac{1}{W_{sum} + \epsilon} - k \cdot |T_{chws} - T_{chws, set}| \quad (2)$$

2.2.2. The robust assurance scheme

The robust assurance scheme is designed between the control station and the Direct Digital Control (DDC) controller to enhance system reliability. As depicted in Fig. 1, the optimal setpoints generated by the AI-empowered algorithm are sent to a detection node at the controller, where their validity is assessed. A dynamic threshold detection method,

based on a simple moving average (SMA), is introduced to evaluate the reliability of the control station's outputs.

As highlighted in formulas (3–6), the detection method involves the calculation of the moving average and standard deviation (SD) of a set temperature point over a specified historical data window. The system identifies abnormal deviations in the optimization setpoints by establishing dynamic thresholds at two SDs above and below the moving average. Such deviations may indicate potential algorithmic failures or inefficiencies, triggering an automatic switch to a preset scheme based on expert knowledge. In addition, if the DDC controller does not receive an updated setpoint from the control station within 20 min, a system alert is triggered, indicating a possible failure in communication. In such cases, the system automatically switches to robust assurance scheme to maintain stable operation. Once the controller resumes receiving valid setpoints, the system switches back to the optimal control scheme. This robust assurance mechanism ensures proper system operation and prevents potential operational failures.

$$SMA_t = \frac{1}{N} \sum_{i=t-N+1}^t x_i \quad (3)$$

$$Upper\ Threshold_t = SMA_t + k \times SD_t \quad (4)$$

$$Lower\ Threshold_t = SMA_t - k \times SD_t \quad (5)$$

$$SD_t = \sqrt{\frac{\sum_{i=t-N+1}^t (x_i - \bar{x})^2}{N - 1}} \quad (6)$$

where, SMA_t represents the simple moving average at time t , N is the size of the moving average window, and x_i is the data point at position i . The SD_t indicates the standard deviation within the window, k is the threshold width coefficient, and $\bar{x}$ is the mean of the data points within the window.

In addition, an empirical fitting formula based on expert knowledge is introduced to provide set points under abnormal conditions. According to theoretical research and experiments conducted by a chiller manufacturer, the cooling water temperature optimization set point is influenced by several factors, including the outdoor wet-bulb temperature (T_{wb}) and its design value, the partial load ratio of the chiller system (PLR), and the flow rate of the cooling water (m_{cw}):

$$T_{cwr, set} = A \times T_{wb} + B \times PLR - C \times T_{wb, set} - D \times m_{cw} / Load + 37 \quad (7)$$

3. Setup of hardware-in-the-loop tests

Hardware-in-the-loop test integrates real-world hardware components into a virtual simulation environment. It is adopted in this study to test the strategy deployed on a physical control station for controlling complex systems under comprehensive dynamic conditions. This section provides a detailed description of the test platform and the test arrangement.

3.1. Description of smart station and test platform

A smart station was constructed using a Raspberry Pi in this study, which features a quad-core ARM Cortex-A72 processor, supporting up to 8GB of LPDDR4 RAM, and is compatible with Linux-based operating systems. Fig. 3 presents both the schematic diagram and the implementation diagram of the hardware-in-the-loop test platform. The physical smart station, equipped with the deep learning-enabled genetic algorithm, controls a simulated, real-time virtual dynamic cooling system. The cooling system model was developed in Dymola using the Modelica language and packaged as a Functional Mock-up Interface (FMI) file, serving as a "digital twin in a box" hosted on the computer. Within the communication framework, the TCP/IP protocol is employed over the local area network, simulating the real-time data transmission

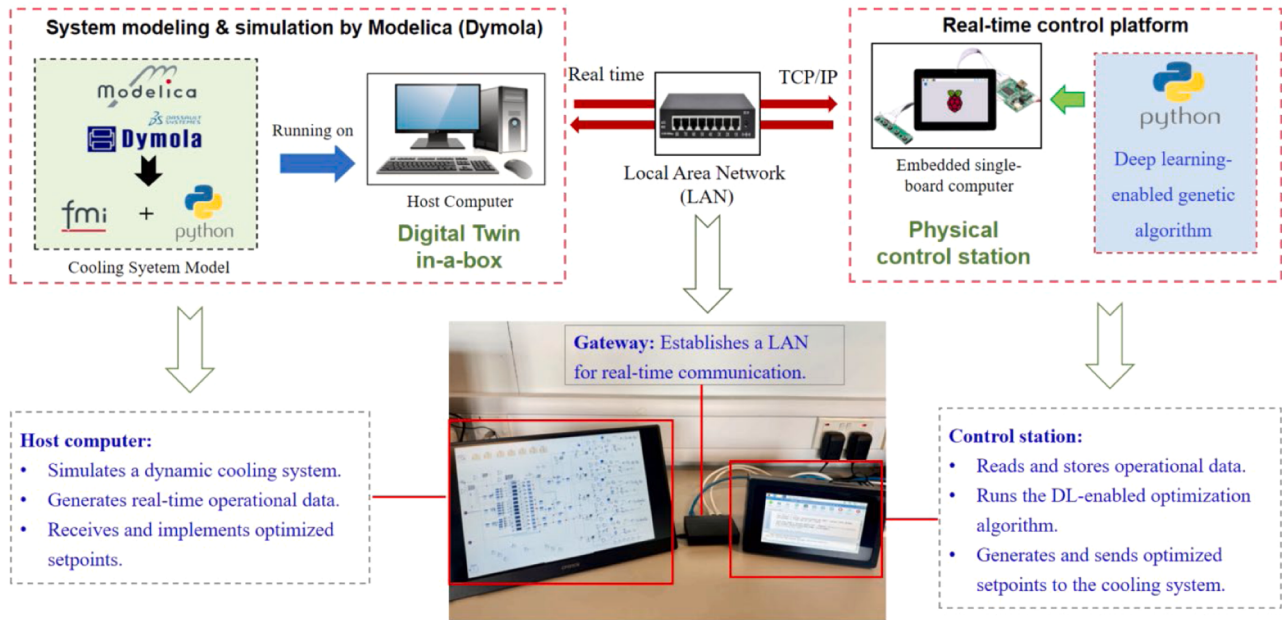


Fig. 3. Hardware-in-the-loop test platform: schematic and implementation.

process during optimization.

In this study, a Raspberry Pi-based control station is developed to implement the optimization strategy in a hardware-in-the-loop test, evaluating the hardware requirement and its feasibility in a low-cost and easily deployable setup. Considering the Raspberry Pi is primarily designed for education and personal projects, it may face processing limitations when applied to larger and more complex systems. For practical implementation, an NVIDIA Jetson edge AI device is chosen to replace the Raspberry Pi, ensuring higher computational efficiency, reliability, and real-time processing capabilities.

3.2. The test arrangement

The proposed AI-empowered optimization strategy is deployed on the smart station, and evaluated using the hardware-in-the-loop test platform. Real-time validation tests are conducted over a day using weather data representative of a typical summer day in Hong Kong. The analysis begins with an evaluation of computation time per iteration, comparing the results before and after integrating the hybrid model to quantify improvements in computational efficiency.

Subsequently, comparative studies are conducted to evaluate the performance of the proposed AI-empowered optimization strategy against the traditional control strategy.

- **Baseline control strategy:** This is a commonly used method for chiller systems that does not involve optimization. Fixed setpoints for the cooling water supply temperature and return water temperature are maintained throughout the operation.
- **AI-empowered optimization control strategy:** A deep learning-enabled genetic algorithm is used to optimize the cooling water supply and return temperatures at fixed intervals, improving the system's efficiency.

During the tests, an optimization process within each control interval (5 min in this study) is divided into four distinct steps:

- **Sampling:** collecting real-time system data,
- **Calculation:** performing the optimization computation,
- **Sending:** sending optimized setpoints to the system,
- **Waiting:** waiting until the subsequent optimization interval.

To validate the robustness assurance scheme, three hours during a typical day are selected for the validation test. The controller is programmed to send a set of optimal set points every optimization interval, allowing for the evaluation of the Direct Digital Control (DDC) in responding to normal reception, data anomalies and interruptions.

4. Model development and description

4.1. Reference building and cooling system

A commercial building located in West Kowloon, Hong Kong, is selected as the reference building for optimization tests. As shown in Fig. 4, the building comprises four tower buildings (1A, 1B, 2A, 2B) and a podium, with a total area of 288,010 m². The total design cooling load is 55,331 kW. To satisfy this cooling demand, the building is equipped with 21 seawater-cooled chillers, including 15 chillers rated at 3517 kW and 6 chillers rated at 1758 kW, as summarized in Table 1.

The central cooling systems described above are simulated in Dymola, a software platform based on the Modelica language [26], which is capable of simulating and analyzing complex dynamic systems with high accuracy and flexibility. As shown in Fig. 5, the system structure is organized from left to right, representing the seawater side, cooling water circulation, chiller units, and load side, respectively. Additionally, from top to bottom, the structure follows the hierarchy of Podium, Building 1, and Building 2, as illustrated in Fig. 4. The models for components, such as pumps, heat exchangers and chillers, are developed using the "Buildings" library. Basic control logic, including PID feedback control and device start-stop control, is incorporated into the model to simulate fundamental operational behaviors. During the simulation, external environmental parameters, including seawater temperature and building load variations, are input into the seawater side and load side of the Modelica model for dynamic execution. The load profiles are collected from a large existing building of the same owner nearby. The seawater temperature is based on the Hong Kong Environmental Protection Department (EPD) [27].

4.2. Hybrid energy models of cooling systems for online optimization

A hybrid energy model, serving as the fitness function for the genetic algorithm (GA) in online optimization, enables efficient computation of

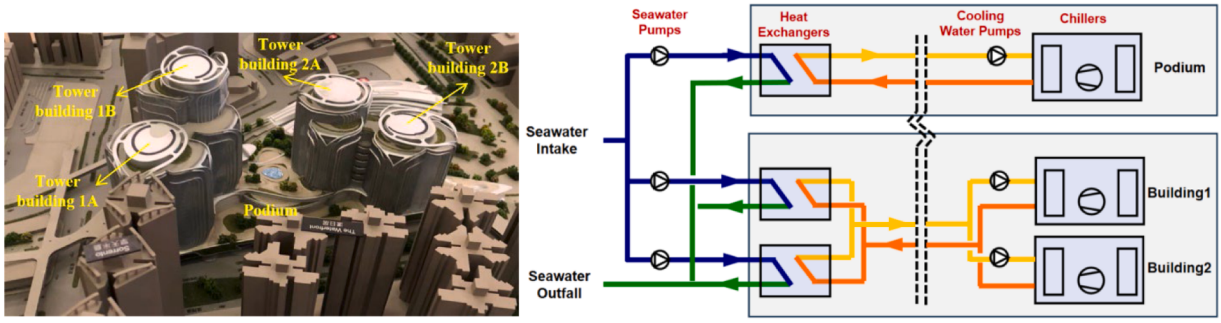


Fig. 4. Overview of the buildings and the central cooling system layout.

Table 1

Key building parameters and cooling system specifications.

Building	Building 1A	Building 1B	Building 2A	Building 2B	Podium
Gross Floor Area (m ²)	58,770	77,000	52,490	43,750	56,000
Air Conditioning Area (m ²)	48,528	60,351	43,751	36,694	13,156 (Retail) 11,744 (F&B) 13,687 (Arcade)
Block load (kW)	10,544	12,960	9415	7906	4526, 4846, 5135
Load ratio	0.19	0.23	0.17	0.14	0.26
Chiller plants	3 × 3517kW +1 × 1758kW	3 × 3517kW +1 × 1758kW	2 × 3517kW +2 × 1758kW	2 × 3517kW +1 × 1758kW	5 × 3517kW +1 × 1758kW

the total power consumption of a central cooling system. Fig. 6 illustrates the structure of this hybrid model and its function in the optimization process. This model is primarily responsible for fitness evaluation in the genetic algorithm optimization, generating fitness scores correlated with total system energy consumption to determine the optimal

offspring. In this hybrid model, Artificial Neural Network (ANN) models are used for chillers and heat exchangers due to their significant nonlinearity and complex parameter settings. The cooling water pump and seawater pump models adopt polynomial fitting derived from their design parameters, as described in formulas (8) and (9). Each component in the models is interconnected according to its physical connection and physical laws, ensuring reliability and interpretability during application.

$$W_{sw} = \pi_1 + \pi_2 \left(\frac{M_{sw}}{M_{sw,nom}} \right) + \pi_3 \left(\frac{M_{sw}}{M_{sw,nom}} \right)^2 + \pi_4 \left(\frac{M_{sw}}{M_{sw,nom}} \right)^3 \quad (8)$$

$$W_{cw} = \mu_1 + \mu_2 \left(\frac{M_{cw}}{M_{cw,nom}} \right) + \mu_3 \left(\frac{M_{cw}}{M_{cw,nom}} \right)^2 + \mu_4 \left(\frac{M_{cw}}{M_{cw,nom}} \right)^3 \quad (9)$$

where, W and M are the energy consumption and flow rate respectively. The subscript nom represents the nominal condition. The parameters $\pi_1 \sim \pi_4$ and $\mu_1 \sim \mu_4$ can be obtained from manufacturers or by curve-fitting using operational data.

The details of the ANN models for the heat exchangers and chillers are illustrated in Fig. 7. Both models utilize a two-hidden-layer structure, with each hidden layer consisting of ten nodes. For the heat exchanger model, there are four input parameters: seawater flowrate (M_{sw}), seawater supply temperature (T_{sws}), cooling water flowrate (M_{cw}), and cooling water return temperature (T_{cwr}). The output of this model is

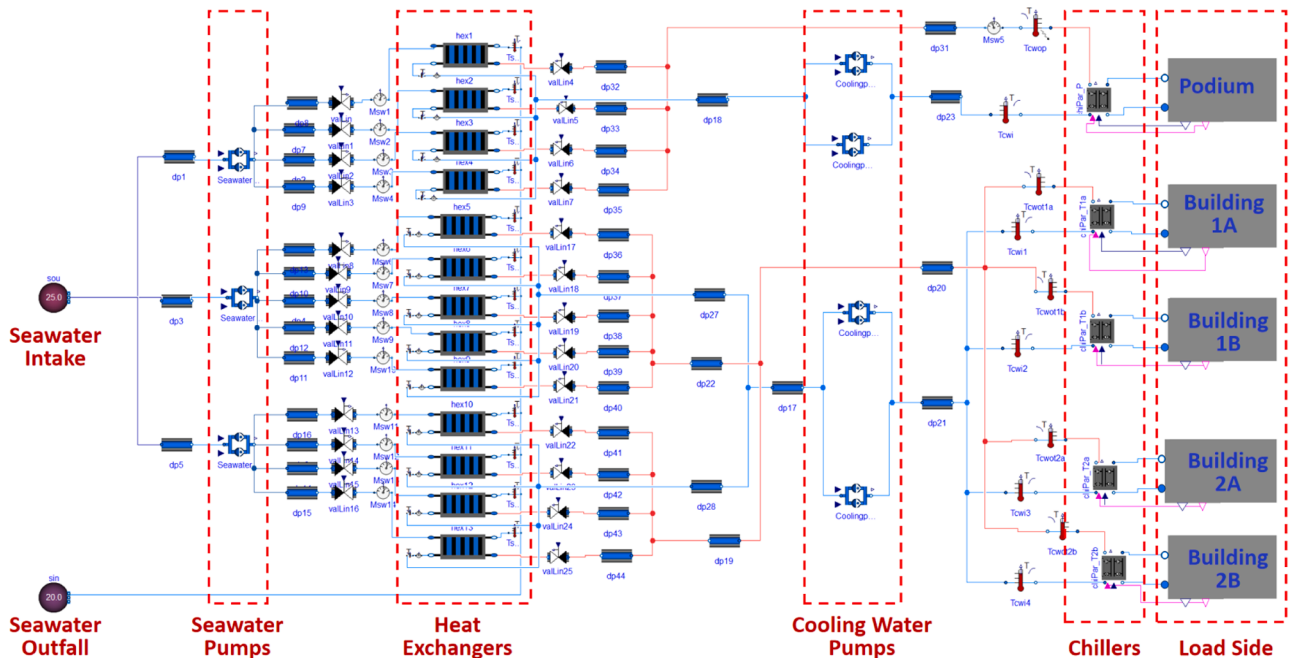


Fig. 5. Seawater-cooled central cooling system in Dymola.

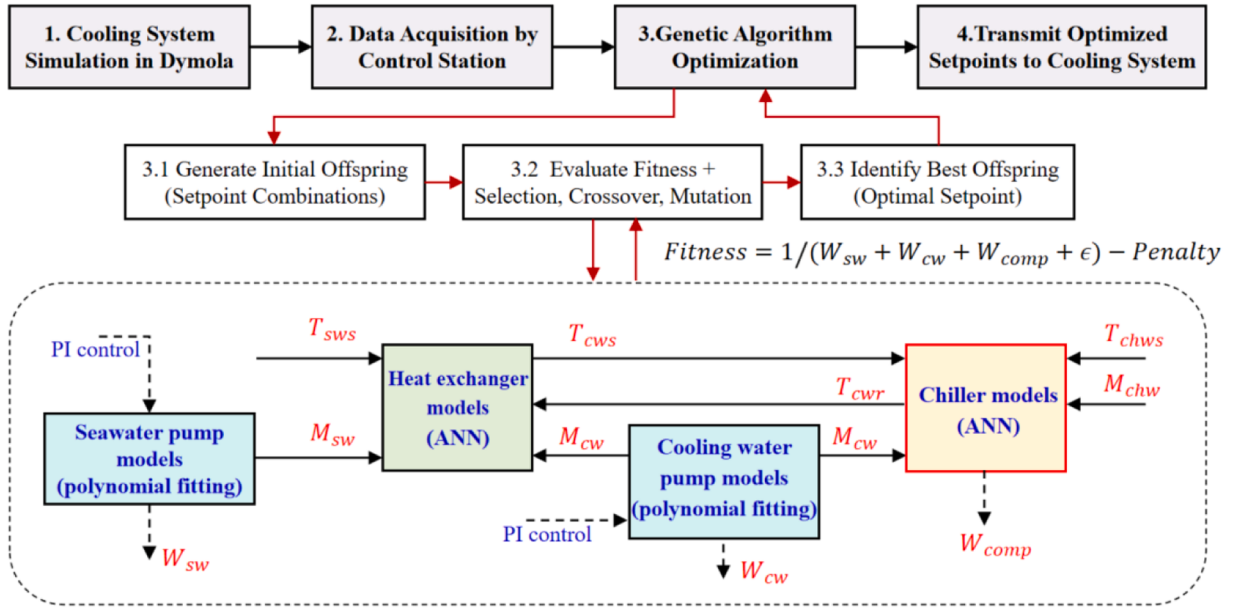


Fig. 6. Hybrid model structure and its function in optimization process.

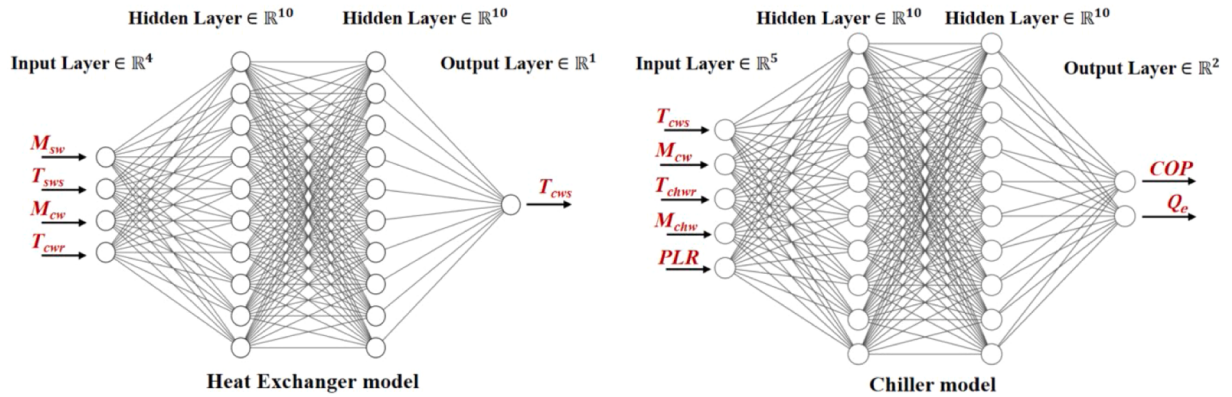


Fig. 7. Overview of ANN model configurations, inputs, and outputs.

the cooling water supply temperature (T_{cws}), which is crucial for coupling with the chiller unit. The chiller model uses five input parameters: cooling water supply temperature (T_{cws}), cooling water flow (M_{cw}), chilled water supply temperature (T_{chws}), chilled water flow (M_{chw}), and the partial load ratio (PLR). The outputs of the chiller model include the coefficient of performance (COP) of the unit and the cooling capacity (Q_e), which can be further used to calculate the chiller's power consumption. The data for building the ANN models comes from comprehensive full-operation simulations of the Modelica model, covering all possible setpoint combinations and environmental parameters under various system operating conditions. A total of 60,000 valid datapoints are generated for machine learning. This simulation approach aims to provide the ANN model with a diverse and representative dataset, enabling it to accurately simulate the behavior of the heat exchanger and chiller under various conditions.

For model validation, 4000 sets of data, which are different from the data used for model training, were randomly generated by the Modelica models (shown in Fig. 7). These validation datasets were used to assess the performance of the trained ANN models. The validation results are shown in Fig. 8. The output of the heat exchanger model is a temperature value, which exhibits a relatively small fluctuation range and a strong logical correlation with the input parameters. As a result, the average relative error of the T_{cws} is 0.01 %. For the chiller model, the average

relative errors of the output values for the COP and Q_e are 0.29 % and 0.18 %, respectively. These results demonstrate that the ANN models trained in this study meet the accuracy requirements.

However, in real-world applications, challenges such as sensor errors, noise, and other system biases are likely to arise. In such cases, the ANN, as a data-driven model, can easily adapt and correct itself through online learning and self-updating, thereby enhancing its robustness to noise and system deviations. Furthermore, in this study, the combination of physical models and ANN models significantly improves the system's interpretability and reliability. By integrating the physical principles of the system with the data-driven learning capabilities of the ANN, the hybrid model-based approach offers both efficient computation and stable operation.

5. Results and discussion

5.1. Optimization efficiency

The number of floating-point operations per second (FLOPs), serving as a standard measure to evaluate the processing power of a computing device, is used to quantify the performance of the physical control station. The theoretical value of FLOPs can be calculated using: $\text{FLOPs} \approx \text{Clock Frequency (GHz)} \times \text{Floating-point Operations per Cycle}$.

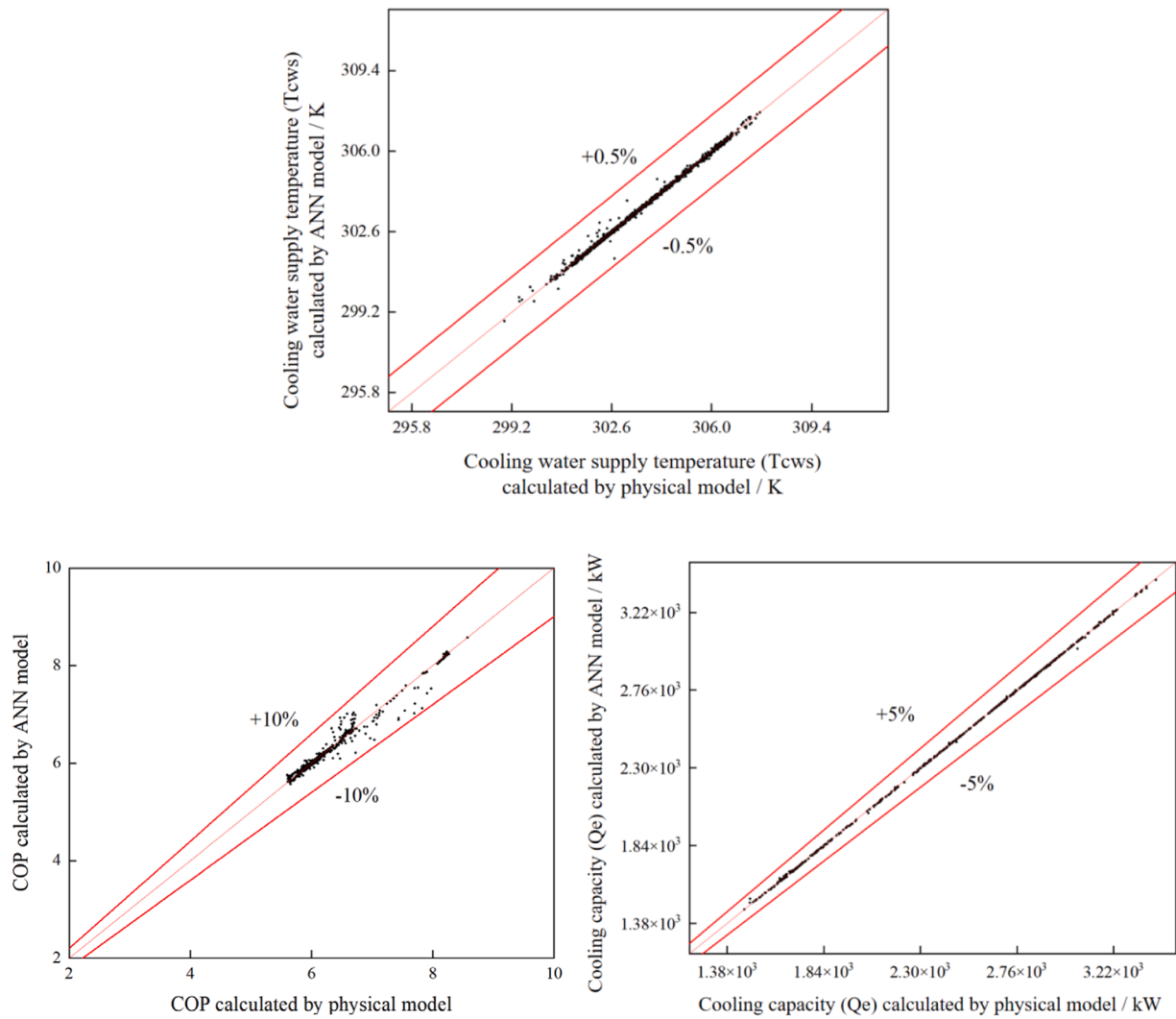


Fig. 8. Comparison between the data and the model prediction.

The smart station used in the hardware-in-the-loop test is equipped with a quad-core ARM Cortex-A72 processor, with each core capable of a maximum clock frequency of 1.5 GHz and supporting up to three floating-point operations per cycle. Thus, the theoretical value of FLOPs can reach 18 GFLOPs ($1.5 \text{ GHz} \times 4 \text{ cores} \times 3$). An Intel Core i7 920

processor typically achieves around 63 GFLOPs, suggesting that the station is only suitable for lightweight computing tasks.

Fig. 9 shows a comparison between the computation times for 200 iterations without and with the hybrid model. In the former, the genetic algorithm uses a fitness function based on a purely physical model.

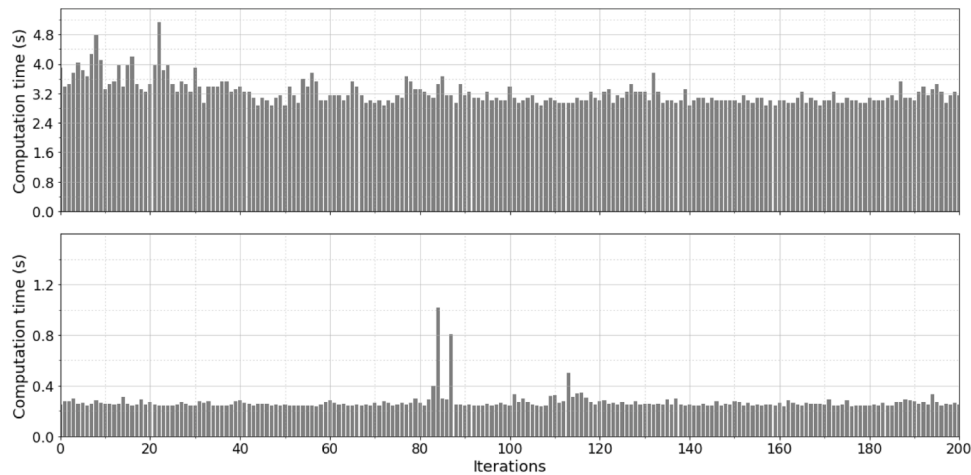


Fig. 9. Iterative computation time: comparison before (top) and after (bottom) introducing the hybrid model.

Specifically, the heat exchanger model is established using the logarithmic mean temperature difference (LMTD) method, while the chiller model consists of the compressor, condenser, expansion valve, and evaporator, also modeled using the heat and mass transfer principle. The detailed modeling methods have been thoroughly discussed in previous research [28]. The results show that the average computation time per iteration using the physical model is about 3.22 s, while the hybrid model reduces this to 0.27 s. When applied to a genetic algorithm with an initial population of 16 and 8 generations, the physical model requires approximately 6.9 min to complete a single optimization computation. By replacing key models with ANN models, the optimization time is reduced to 0.6 min under the same conditions, enabling practical real-time applications of AI-empowered optimization control.

5.2. Control performance

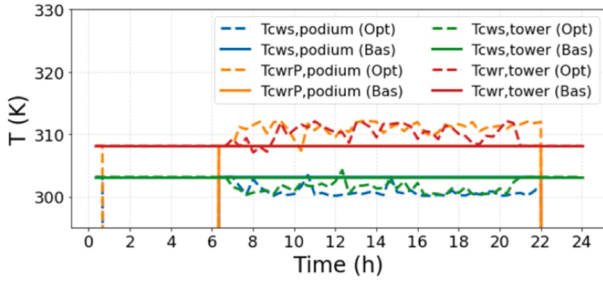
Fig. 10 presents the profiles of major system operation variables controlled by the proposed AI-empowered optimization strategy compared to those under the control of the baseline strategy with fixed setpoints. Fig. 10(a) shows that, compared to the baseline, the optimized

cooling water return temperature increases by 2.30 K on average, while the supply temperature decreases by 1.93 K. These adjustments reduce cooling water flow rates but increase seawater flow rates, as illustrated in Fig. 10(b).

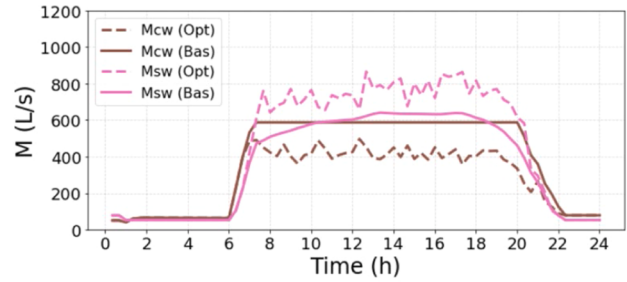
The operational numbers of chillers and pumps follow a rule-based mechanism, ensuring that each unit operates within an acceptable efficiency range. As a result, the number of seawater pumps increases to handle the higher flow rate, while the total numbers of chillers and cooling water pumps remain unchanged, as shown in Fig. 10(c).

Fig. 10(d) demonstrates the impact of this strategy on power consumption. While seawater pump power consumption increases, cooling water pump power consumption decreases significantly. Furthermore, although the higher condenser temperature difference increases compressor energy consumption, this effect is mitigated by a rise in chilled water temperature, as shown in Fig. 10(e). Optimization constraints and penalty functions effectively limit the chilled water temperature increase, ensuring that comfort requirements are not compromised.

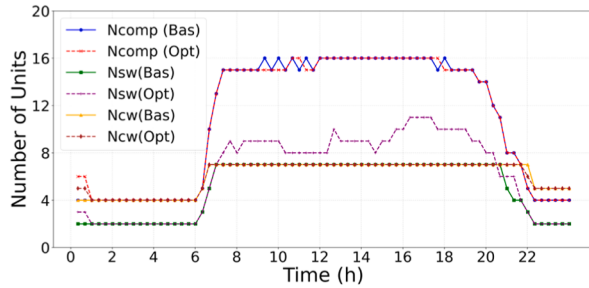
Finally, Fig. 10(f) shows the overall reduction in total energy consumption. By properly balancing the energy use of seawater pumps,



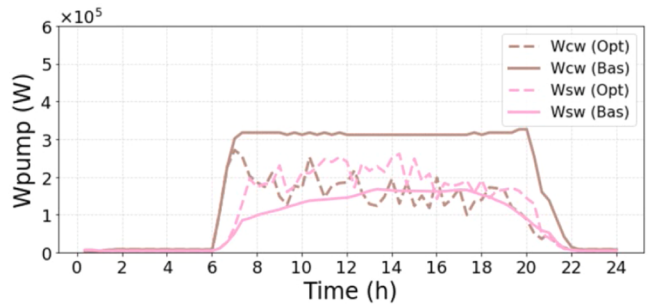
(a) Cooling water temperature set points



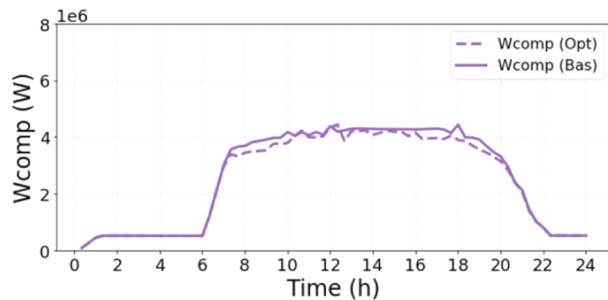
(b) Flow rate of pumps



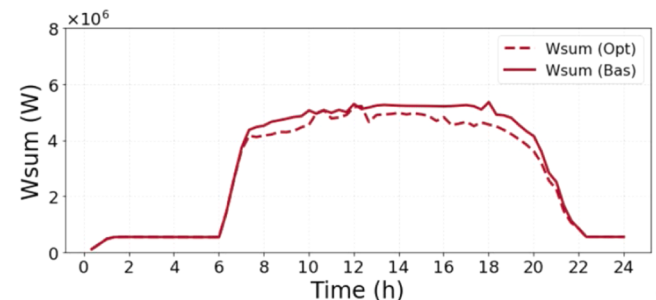
(c) Number of units in operation



(d) Power consumption of pumps



(e) Power consumption of chillers



(f) Total power consumption

Fig. 10. Main operation variables under two control strategies in a typical day.

cooling water pumps, and chillers, the AI-empowered optimization algorithm achieves an energy saving of 7.66 % compared to the baseline case.

Fig. 11 illustrates the detailed breakdown of hourly energy consumption and the profile in the Energy Efficiency Ratio (EER). The hourly energy consumption trends provide a clear representation of the instantaneous energy usage fluctuations under different control strategies. The EER is defined as the ratio of building load to total system energy consumption, serving as an indicator of overall system efficiency. In the basic control strategy, both the water pumps and the chillers operate based on fixed setpoints, resulting in relatively constant energy consumption ratios for each component. In contrast, the optimized control scheme utilizes global optimization through a genetic algorithm, which leads to a more efficient distribution of energy.

The results show that the energy savings of the cooling pumps and chillers significantly outweighed the increased energy consumption of the seawater pumps, leading to a higher EER. As a result, the red line representing the optimized EER in the figure shows a higher value compared to the black line of the basic control strategy.

Detailed energy savings are presented in Table 2. With the global optimization control, chiller energy consumption was reduced by 4.45 %, saving approximately 2605.7 kW. The cooling water pumps achieved substantial energy savings of 47.9 % (4340.1 kW), due to the reduced cooling water flow rate. Since seawater does not require long-distance transport as cooling waters, the seawater pump power consumption is about half that of the cooling water pump. As a result, under the global optimization, seawater pump energy consumption increased by 38.1 % (1469.3 kW). Eventually, the total energy consumption for the basic control scheme was 71,496.6 kW, while the optimized control strategy achieved a total energy consumption of 66,020.1 kW, and an overall energy savings of 7.66 % (5476.5 kW).

Given the increasing importance of the sustainability of HVAC systems, it is essential to analyze the potential reduction in carbon footprint associated with energy conservation. In this paper, the average carbon emission factor of electricity in Hong Kong (0.529 kg CO₂/kWh) is used to calculate the energy savings. The results show that, under typical daily summer conditions, the proposed strategy can reduce CO₂ emissions by approximately 2.9 tons per day.

5.3. Robust assurance in practical operation

Various abnormal disturbances are also introduced in the optimization tests. The operation of DDC controller is depicted in Fig. 12. For

Table 2

Comparison of energy consumption throughout the day.

Energy consumption (kW)	Basic operation	Optimized operation	Comparison
Compressors	58,572.1	55,966.4	4.45 %
Cooling water pumps	9066.6	4726.5	47.9 %
Seawater pumps	3857.9	5327.2	−38.1 %
Total	71,496.6	66,020.1	7.66 %

regular optimization control processes, the controller precisely achieves the optimal set-point received from the smart station. When the data from the optimization algorithm overshoots due to external noise or other transmission problems, the controller automatically switches to the expert knowledge-based scheme, as illustrated in Fig. 12(b). Fig. 12 (c) shows the control reaction when it is unable to receive information from the smart station for an extended period, whether due to a physical layer failure (such as disconnection, poor connection, or port damage) or a network issue (such as packet loss). The test results demonstrate that the robust assurance scheme can mitigate the impact of these abnormal situations by generating set-points based on the current operational state and previous optimization experience. Although it exhibits some fluctuations and cannot achieve global optimization accurately, this scheme effectively ensures the stability and reliability of AI-empowered optimization in real-time operation.

6. Conclusion

This paper presents an AI-empowered optimization strategy for practical implementation at the field level. A hardware-in-the-loop test is conducted to validate the computational efficiency, control performance, and operational robustness of the proposed strategy. The results show that the proposed AI-empowered optimization strategy accelerates the optimization process by a factor of 6.9 and achieves an energy saving of 7.66 %. The robust assurance scheme effectively mitigates the impacts of abnormal disturbances through an expert-based switching mechanism, ensuring stable and reliable system operation. This research advances the boundaries of AI applications in the building and energy system fields, however, there is still much to explore regarding efficiency and adaptability under dynamic working conditions in practical applications. Future research should further address practical application issues, such as developing general AI models and transitioning toward adaptive control.

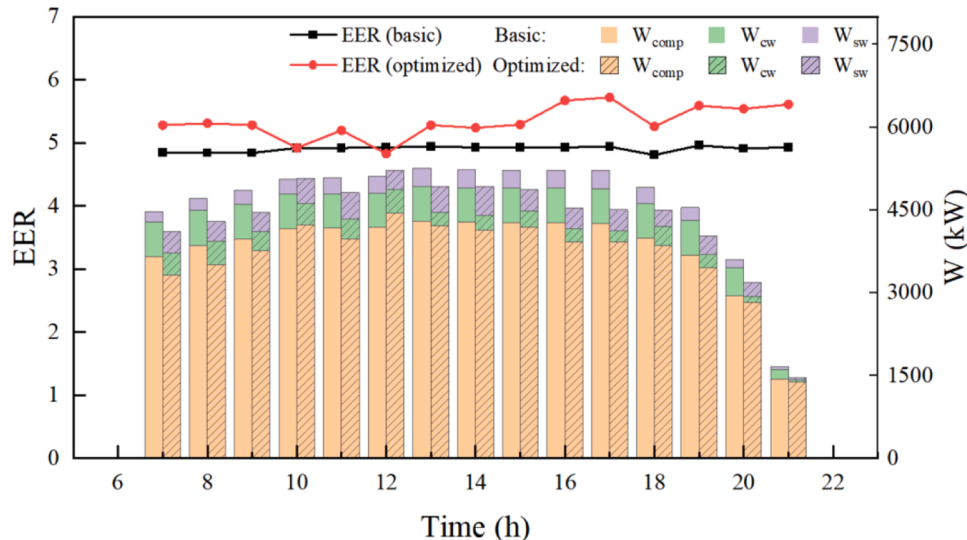


Fig. 11. Comparison between hourly energy consumption and Energy Efficiency Ratio under baseline and AI-empowered optimization strategies.

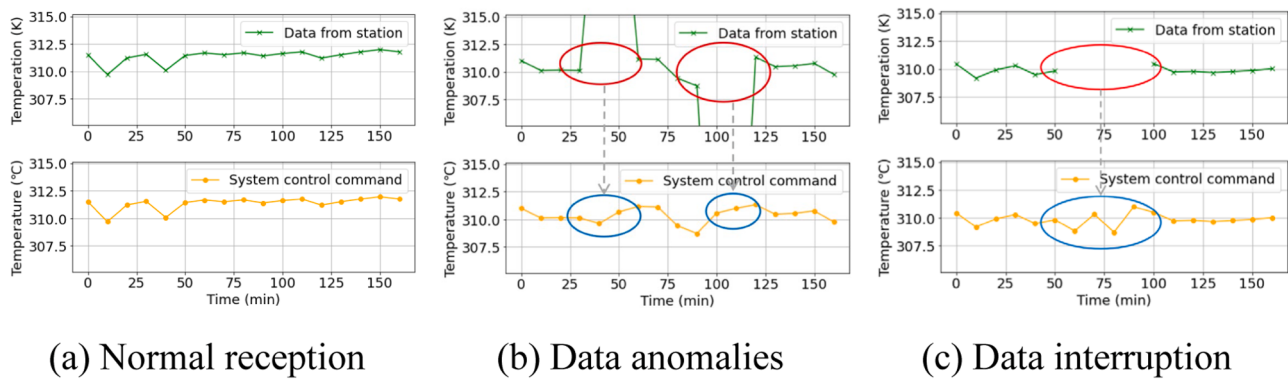


Fig. 12. Controller response in different situations.

CRedit authorship contribution statement

Lingyun Xie: Writing – original draft, Software, Methodology, Investigation, Data curation. **Kui Shan:** Writing – review & editing, Supervision. **Hong Tang:** Writing – review & editing, Conceptualization. **Shengwei Wang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The research presented in this paper is financially supported by a General Research Fund (No.152053/21) of the Research Grant Council (RGC) of the Hong Kong SAR and a grant for collaborative research from Sun Hung Kai Properties.

Data availability

Data will be made available on request.

References

- [1] Net zero by 2050 - a roadmap for the global energy sector. International Energy Agency; 2021.
- [2] Chen Z, Xiao F, Guo F, Yan J. Interpretable machine learning for building energy management: a state-of-the-art review. *Adv Appl Energy* 2023;9:100123. <https://doi.org/10.1016/j.adapen.2023.100123>.
- [3] Aste N, Manfren M, Marenzi G. Building Automation and Control Systems and performance optimization: a framework for analysis. *Renew Sustain Energy Rev* 2017;75:313–30. <https://doi.org/10.1016/j.rser.2016.10.072>.
- [4] Ulpiani G, Borgognoni M, Romagnoli A, Di Perna C. Comparing the performance of on/off, PID and fuzzy controllers applied to the heating system of an energy-efficient building. *Energy Build* 2016;116:1–17. <https://doi.org/10.1016/j.enbuild.2015.12.027>.
- [5] Parvin K, Lipu MSH, Hannan MA, Hannan MA, Ker PJ, Abdullah MA, et al. Intelligent controllers and optimization algorithms for building energy management towards achieving sustainable development: challenges and prospects. *IEEE Access* 2021;9:41577–602. <https://doi.org/10.1109/access.2021.3065087>.
- [6] Vignali RM, Borghesan F, Piroddi L, Strelec M, Prandini M. Energy management of a building cooling system with thermal storage: an approximate dynamic programming solution. *IEEE Transact Autom Sci Eng* 2017;14:619–33. <https://doi.org/10.1109/TASE.2016.2635109>.
- [7] OLSON RT, LIEBMAN JS. Optimization of a chilled water plant using sequential quadratic programming. *Eng Optimiz* 1990;15:171–91. <https://doi.org/10.1080/03052159008941151>.
- [8] Deng K, Sun Y, Chakraborty A, Lu Y, Brouwer J, Mehta PG. Optimal scheduling of chiller plant with thermal energy storage using mixed integer linear programming. In: 2013 American Control Conference; 2013. p. 2958–63. <https://doi.org/10.1109/ACC.2013.6580284>.
- [9] Escobar LM, Aguilar J, Garces-Jimenez A, Gutierrez De Mesa JA, Gomez-Pulido JM. Advanced fuzzy-logic-based context-driven control for HVAC management systems in buildings. *IEEE Access* 2020;8:16111–26. <https://doi.org/10.1109/ACCESS.2020.2966545>.
- [10] Durand-Estebe B, Bot CLe, Mancos JN, Arquis E. Simulation of a temperature adaptive control strategy for an IWSE economizer in a data center. *Appl Energy* 2014;134:45–56. <https://doi.org/10.1016/j.apenergy.2014.07.072>.
- [11] Zhang T, Liu Y, Rao Y, Li X, Zhao Q. Optimal design of building environment with hybrid genetic algorithm, artificial neural network, multivariate regression analysis and fuzzy logic controller. *Build Environ* 2020;175:106810. <https://doi.org/10.1016/j.buildenv.2020.106810>.
- [12] Zhang K, Prakash A, Paul L, Blum D, Alstone P, Zoellick J, et al. Model predictive control for demand flexibility: real-world operation of a commercial building with photovoltaic and battery systems. *Adv Appl Energy* 2022;7:100099. <https://doi.org/10.1016/j.adapen.2022.100099>.
- [13] Zhao N, Zhang H, Yang X, Yan J, You F. Emerging information and communication technologies for smart energy systems and renewable transition. *Adv Appl Energy* 2023;9:100125. <https://doi.org/10.1016/j.adapen.2023.100125>.
- [14] Zhang L, Wen J, Li Y, Chen J, Ye Y, Fu Y, Livingood W. A review of machine learning in building load prediction. *Appl Energy* 2021;285:116452. <https://doi.org/10.1016/j.apenergy.2021.116452>.
- [15] Ifaei P, Esfhankalateh AT, Ghobadi F, Mohammadi-Ivatloo B, Yoo C. Systematic review and cutting-edge applications of prominent heuristic optimizers in sustainable energies. *J Clean Prod* 2023;414:137632. <https://doi.org/10.1016/j.jclepro.2023.137632>.
- [16] Yao Y, Hong X, Xiong L. Study on a new metaheuristic algorithm – Tribal intelligent evolution optimization and its application in optimal control of cooling plants. *Appl Energy* 2025;383:125339. <https://doi.org/10.1016/j.apenergy.2025.125339>.
- [17] Miao Y, Yao Y, Hong X, Xiong L, Zhang F, Chen W. Research on optimal control of HVAC system using swarm intelligence algorithms. *Build Environ* 2023;241:110467. <https://doi.org/10.1016/j.buildenv.2023.110467>.
- [18] Dai M, Li H, Wang S. A reinforcement learning-enabled iterative learning control strategy of air-conditioning systems for building energy saving by shortening the morning start period. *Appl Energy* 2023;334:120650. <https://doi.org/10.1016/j.apenergy.2023.120650>.
- [19] Huang S, Lu X, Zuo W, Zhang X, Liang C. Model-based optimal operation of heating tower heat pump systems. *Build Environ* 2019;160:106199. <https://doi.org/10.1016/j.buildenv.2019.106199>.
- [20] Ma Z, Wang S. Supervisory and optimal control of central chiller plants using simplified adaptive models and genetic algorithm. *Appl Energy* 2011;88:198–211. <https://doi.org/10.1016/j.apenergy.2010.07.036>.
- [21] Su B, Wang S, Li W. Impacts of uncertain information delays on distributed real-time optimal controls for building HVAC systems deployed on IoT-enabled field control networks. *Appl Energy* 2021;300:117383. <https://doi.org/10.1016/j.apenergy.2021.117383>.
- [22] Hannan MA, Faisal M, Ker PJ, Mun LH, Parvin K, Mahlia TMI, Blaabjerg F. A review of Internet of energy based building Energy Management systems: issues and recommendations. *IEEE Access* 2018;6:38997–9014. <https://doi.org/10.1109/ACCESS.2018.2852811>.
- [23] Bird M, Daveau C, O'Dwyer E, Acha S, Shah N. Real-world implementation and cost of a cloud-based MPC retrofit for HVAC control systems in commercial buildings. *Energy Build* 2022;270:112269. <https://doi.org/10.1016/j.enbuild.2022.112269>.
- [24] Wollschlaeger M, Sauter T, Jasperneite J. The future of industrial communication: automation networks in the era of the Internet of Things and Industry 4.0. *IEEE Ind Electron Mag* 2017;11:17–27. <https://doi.org/10.1109/MIE.2017.2649104>.
- [25] Pan Y, Zhu M, Lv Y, Yang Y, Liang Y, Yin R, et al. Building energy simulation and its application for building performance optimization: a review of methods, tools, and case studies. *Adv Appl Energy* 2023;10:100135. <https://doi.org/10.1016/j.adapen.2023.100135>.
- [26] Fritzson P, Engelson V. Modelica — A unified object-oriented language for system modeling and simulation. In: Jul E, editor. *ECOOP'98 — object-oriented*

- programming. Berlin, Heidelberg: Springer; 1998. p. 67–90. <https://doi.org/10.1007/BFb0054087>.
- [27] Environmental Protection Department. Marine water quality in Hong Kong in 2023. The Government of the Hong Kong Special Administrative Region; 2023. p. 2023. https://www.epd.gov.hk/epd/sc_chi/environmentinhk/water/hkwqrc/waterquality/marine-2.html.
- [28] Xie L, Huang S, Ye Y, Zhang X. Regional suitability of heating tower heat pumps based on technical and economic performance maps. Energy Build 2021;245: 111048. <https://doi.org/10.1016/j.enbuild.2021.111048>.