Novel energy consumption forecasting method employing weighted occupant behavior probabilities and physics-informed network with thermodynamic constrain in office buildings

Chengyu Zhang[1], Jiaming Wang[1], Yuan Su[2], Yacine Rezgui[3], Zhiwen Luo[4], Yifan Wu[5], Chang Sun[1,6], Ben Jiang[1], Yiting Wang[1], Tianyi Zhao[1,*]

1 School of Infrastructure Engineering, Dalian University of Technology, Dalian, China

2 School of Architecture &s Fine Art, Dalian University of Technology, Dalian, China

3 School of Engineering, Cardiff University, Cardiff, UK

4 Welsh School of Architecture, Cardiff University, Cardiff, UK

5 Department of Building Science, Tsinghua University, Beijing, China

6 China National Engineering Research Center for Human Settlements, Beijing, China

* Corresponding author: zhaotianyi@dlut.edu.cn

Abstract: Addressing the global energy crisis and excessive emissions has heightened the critical importance of reducing energy consumption and carbon emissions in the building sector, making accurate building energy forecasting a fundamental research focus. While existing methods predominantly prioritize forecasting accuracy by advanced algorithms, considerations of computational efficiency and model interpretability remain scarce. To bridge this gap, this study proposes a novel forecasting method that simultaneously optimizes for accuracy, efficiency, and interpretability. The method integrates three strategies: (a) incorporating weighted occupant behavior probabilities as novel inputs; (b) incorporating physics-informed loss function calculated by thermal resistance-capacitance (R-C) models; and (c) developing a hybrid CNN-LSTM-Attention algorithm that integrates convolutional neural networks and an attention mechanism with a long short-term memory network. Validation of 48 cases from four office buildings shows the proposed method significantly enhances performance. These three strategies reduce the mean absolute percentage error (MAPE) by 25.78% and the coefficient of variation of the root mean square error (CV-RMSE) by 21.31%, and average contributions are 40%, 15% and 45% for Strategies (a)–(c), respectively. Strategy (c) is the primary contributor to efficiency gains, which can reduce time consumption by 7343.69s and 146.81s compared to Transformer-LSTM-Adaboost and LSTM-SSA, respectively. Strategies (a) and (b) improve interpretability by embedding occupant behavior patterns and thermal constraints. Moreover, the priority of these strategies for buildings with varying behavioral and functional complexities is analyzed. In summary, based on theoretical considerations and practical validation, the proposed method can improve the accuracy, efficiency, and interpretability simultaneously.

Keyword: Building energy consumption forecasting, Physics-informed algorithms, Occupant behavior probabilities, Building thermal R-C network models, Attention mechanism

Highlights:

1. This paper proposes a forecasting method with novel inputs, algorithm and loss function;
2. This paper uses thermal R-C models for physic-informed loss to enhance interpretability;
3. This paper uses weighted behavior probabilities as novel inputs to enhance accuracy;
4. This paper uses novel algorithms (CNN-LSTM-Attention) to enhance accuracy and efficiency;
5. This paper discusses the contribution and priority for different optimization strategies.

1 Introduction

Global energy consumption has consistently increased over the past three decades [1]. According to the National Human Activity Pattern Survey (NHAPS), respondents spend an average of 87% of their time inside buildings [2]. Recent studies indicate that buildings account for 41% of total primary energy use in the U.S. and even up to 65% in Europe [3]. In China, the building sector contributes 44.8% of the total national energy consumption; even when considering only the operation phase, this value reaches 22.0% [4].

Consequently, low-carbon and energy-efficient building operations have become a global research priority, with energy consumption forecasting serving as a foundational element [5]. Accurate hourly energy consumption forecasting may be widely beneficial to groups and individuals. For example, by accurate forecasting, energy companies can adjust energy pricing in real time, support demand response programs, and react more promptly and precisely to extreme weather conditions [6]. Energy managers can optimize the schedules of energy storage and photovoltaic systems to enhance supply and demand matching [7]. Additionally, energy users can shift flexible loads, such as laundry operations, from high-price to low-price hours, thereby reducing energy costs [9]. Building energy consumption forecasting methods are generally categorized into three types: physical modeling, data-driven, and hybrid methods.

Regarding physical modeling methods, these methods involve simulation software, such as EnergyPlus, which requires detailed inputs including climate data, building information, envelope and device performance, and operation schedules. For example, Neto et al., developed an EnergyPlus model for a university office building, which shows a forecasting error range of ±13% for 80% of the tested database [10]. However, these methods possess limited online applicability and struggle to represent the complex and real-time interactions among occupants, indoor environment, and energy-consuming devices accurately [11]. Furthermore, some researchers raised concerns regarding the common accuracy of this method [12]. Consequently, for applications such as energy efficiency assessment and management, data-driven and hybrid methods are often preferred over those relying solely on physical modeling.

Regarding Data-driven methods, these methods focus on two improvements including identifying optimal input features and developing advanced training algorithms. For example, Wang et al. developed a novel federated learning framework that integrates a sparse Mixture-of-Experts model with a lightweight MetaFormer model, reporting a 10–40% improvement in forecasting accuracy [13]. Feng et al. implemented adaptive LSTM networks optimized with a beluga whale optimization algorithm, reducing the RMSE by 11% and the CV-RMSE by 6.3% [14]. Cao et al. proposed a PSO-stacking ensemble model with priority feature selection, which achieved significant reductions in RMSE [15]. Zhang et al. proposed a suitable input selection method, which can reduce the MAPE by 13.94% on average in seven buildings [16]. Zhang et al. proposed a short-term power load forecasting method for industrial buildings with temporal convolutional network, informer, and bi-directional gated recurrent unit (GRU), which can reduce the mean absolute error (MAE) by 48.54% on average [17]. Overall, common input features encompass factors influencing building energy performance, such as physical properties, indoor and outdoor environmental conditions, occupant behaviors,

envelope and device performance, economic, and social factors [18]. And training algorithms frequently employ some advanced machine learning algorithms such as LSTM and GRU [19], with further optimizations achieved through swarm intelligence for hyperparameter optimization [20], Transformer frameworks for weight optimization [21], CNN for data processing [22], and transfer learning to address data scarcity [23]. However, despite these advancements, challenges in terms of technology remain. Accurately integrating complex, dynamic inputs such as detailed occupant behavior and environmental fluctuations into models is still difficult. Moreover, while current research predominantly focuses on improving predictive accuracy, other critical aspects such as computational efficiency and model interpretability are often neglected.

Regarding hybrid methods, these methods are generally categorized into two approaches: the first extends datasets using physical modeling for subsequent data-driven forecasting, while the second incorporates physical models as the constraints or inputs into data-driven methods. For example, Ali et al. generated large-scale synthetic building data through parametric simulation and integrated end-use demand segregation with ensemble learning, achieving a forecasting accuracy of 91% compared to 76% for traditional methods [24]. Similarly, Song et al. used building energy simulation software to generate data and employed the whale optimization algorithm to optimize hyperparameters of the Bi-LSTM, resulting in the lowest MAPE among benchmark techniques [25]. Zhang et al. introduced occupant energy-use behavior probability models as algorithm inputs, reducing the MAPE by 19.54% on average [26]. Lee et al. developed a spatial-temporal graph neural network with an encoder-decoder framework that embedded physical constraints from mass and energy conservation laws, which enhanced interpretability and improved forecasting accuracy by 44.7% [27]. Michalakopoulos et al. applied a physics-informed neural network that incorporated heat loss constraints, yielding the RMSE lower than purely data-driven equivalents [28]. Despite their promising performance, hybrid methods often require complex validation and expertise across disciplines. A significant technical challenge also remains the limited ability of existing methods to fully capture the complex and real-time interactions among occupants, environment, and devices. Consequently, further simultaneous improvements in the accuracy, efficiency, and interpretability of hybrid forecasting methods are necessary.

Among the hybrid methods, physics-informed neural network (PINN) has emerged as a promising paradigm that seamlessly integrates physical principles with data-driven learning [29–31, 34]. In the context of building energy analyses, common PINN practices may embed governing equations or thermodynamic constraints into the loss function or input features. The PINN can improve generalizability beyond the training data distribution, improve interpretability by grounding predictions in established physical laws, and reduce reliance on massive, high-quality datasets. However, these benefits come also with inherent challenges. PINN increases model complexity and requires domain knowledge to build physical constraints. Consequently, while the potential of PINN is widely recognized, verifying its adaptability in buildings with different characteristics and achieving an optimal balance among accuracy, efficiency, and interpretability are important.

To address these challenges and fill existing knowledge gaps, this study proposes a novel energy consumption

forecasting method for office buildings, designed to enhance efficiency, accuracy, and interpretability simultaneously. The framework of the proposed method is shown in Figure 1, with the key components, objectives, and case studies. The potential contributions and primary works are summarized as follows:

(1) Introducing the weighted occupant AC, light and plug usage behavior probabilities as the novel inputs, to improve the interpretability and accuracy;

(2) Using the combined loss function with physical-informed (calculated by the thermal R-C network models) and data-informed loss replacing the only data-informed loss function (MAE), to improve the interpretability;

(3) Proposing the algorithmic improved modules into the basic training algorithm LSTM, including CNN, Attention, SSA, Transformer, etc., to improve the efficiency and accuracy;

(4) Comparing the performance with different optimization strategies in different buildings, and discussing the contributions, suitability and prioritization.

The remaining sections of this paper are structured as follows: Section 2 details the methodology, encompassing data collection, occupant behavior modeling, thermodynamic resistance-capacitance network modeling, and all the algorithmic modules. Section 3 presents case studies and results, analyzing the performance improvements achieved by different optimization strategies. Section 4 discusses the contributions and priority, compares the findings with those of other recent studies, and notes the limitations of this study. Finally, Section 5 provides concluding remarks.

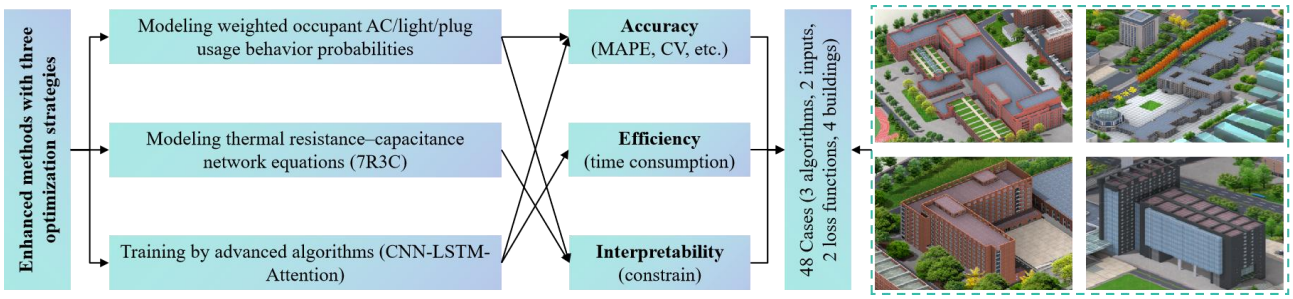


Figure 1 The framework of the proposed enhanced building energy consumption forecasting method

2 Methodology

2.1 Information of example buildings

The case studies comprise four office buildings on a campus in cold region in Dalian, China, characterized by a coastal warm-temperate continental monsoon climate with a mean annual temperature of 10.4°C [32]. Building A houses the school of environment and biology. Building B houses the school of economics and management, the skill training center, and the international exchange office. Building C houses four schools including electrical engineering, information and communication engineering, control science and engineering, and computer science and technology. Building D houses the school of energy and power solely. All these office buildings contain professor's offices, student workstations, conference rooms, lecture halls, and laboratories, etc. Indoor environment and electricity monitoring systems are installed in these buildings, providing sufficient data for the case studies. The layout of the monitoring systems is shown in Figure 2, and detailed building information is summarized in Table 1.

Table 1 Detailed information of typical office building

Codes	Building function and energy-use branches	Sizes	Appearances
A	ACs: Common split ACs in almost all rooms Lights: Fluorescent lamps and LED lamps Plugs: computers, experiment devices, heaters, etc. EUI: 65.79 kWh/m^2	Length: 84.06 m (a), 75.06 m (b) Width: 81.86 m (a), 67.86 m (b) Height: 29.10 m (a), 23.60 m (b) Area: 40324 m^2 South W2W ratio: 0.20	
B	ACs: VRF AC system Lights: Fluorescent lamps and LED lamps Plugs: computers, servers, printers, heaters, etc. EUI: 29.88 kWh/m^2	Length: 233.98 m Width: 81.32 m Height: 23.55 m Area: 35998 m^2 South W2W ratio: 0.18	
C	ACs: Central AC system with screw-type chillers Lights: Fluorescent lamps and LED lamps Plugs: computers, servers, printers, heaters, etc. EUI: 39.44 kWh/m^2	Length: 15.30 m (a), 51.90 m (b) Width: 40.20 m (a), 20.40 m (b) Height: 85.80 m (a), 55.80 m (b) Area: 36500 m^2 South W2W ratio: 0.25	
D	ACs: Common split ACs in almost all rooms Lights: Fluorescent lamps and LED lamps Plugs: computers, experiment devices, heaters, etc. EUI: 90.20 kWh/m^2	Length: 57.70 m Width: 64.90 m Height: 43.90 m Area: 27500 m^2 South W2W ratio: 0.22	

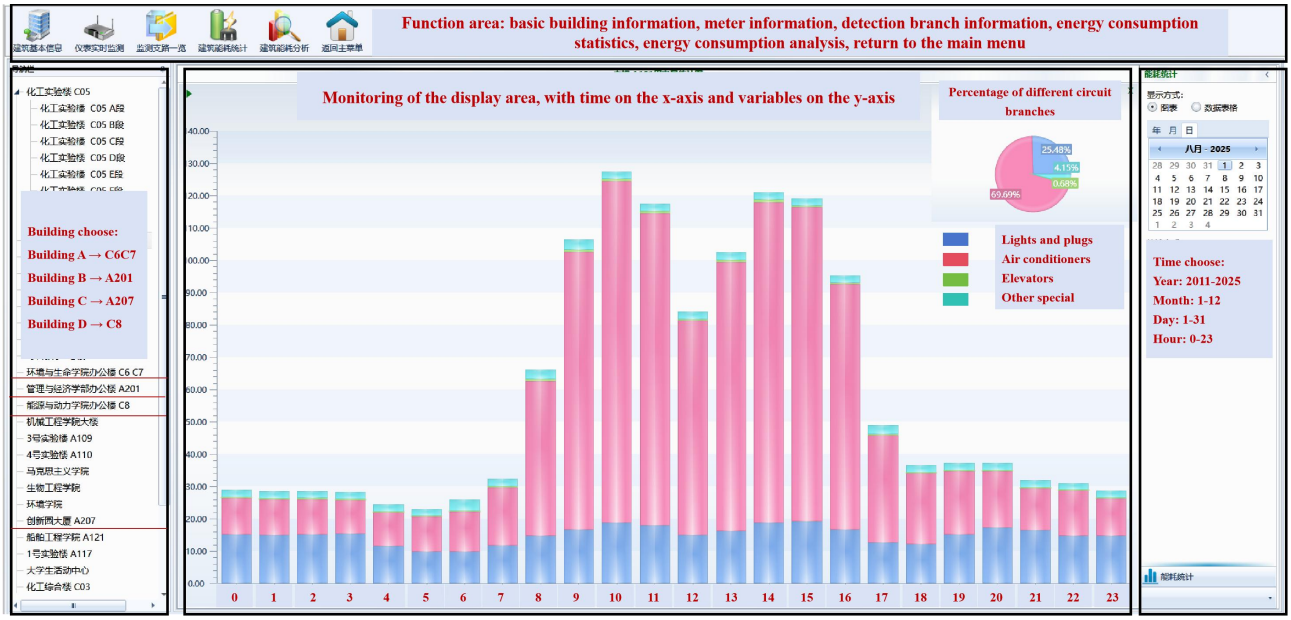
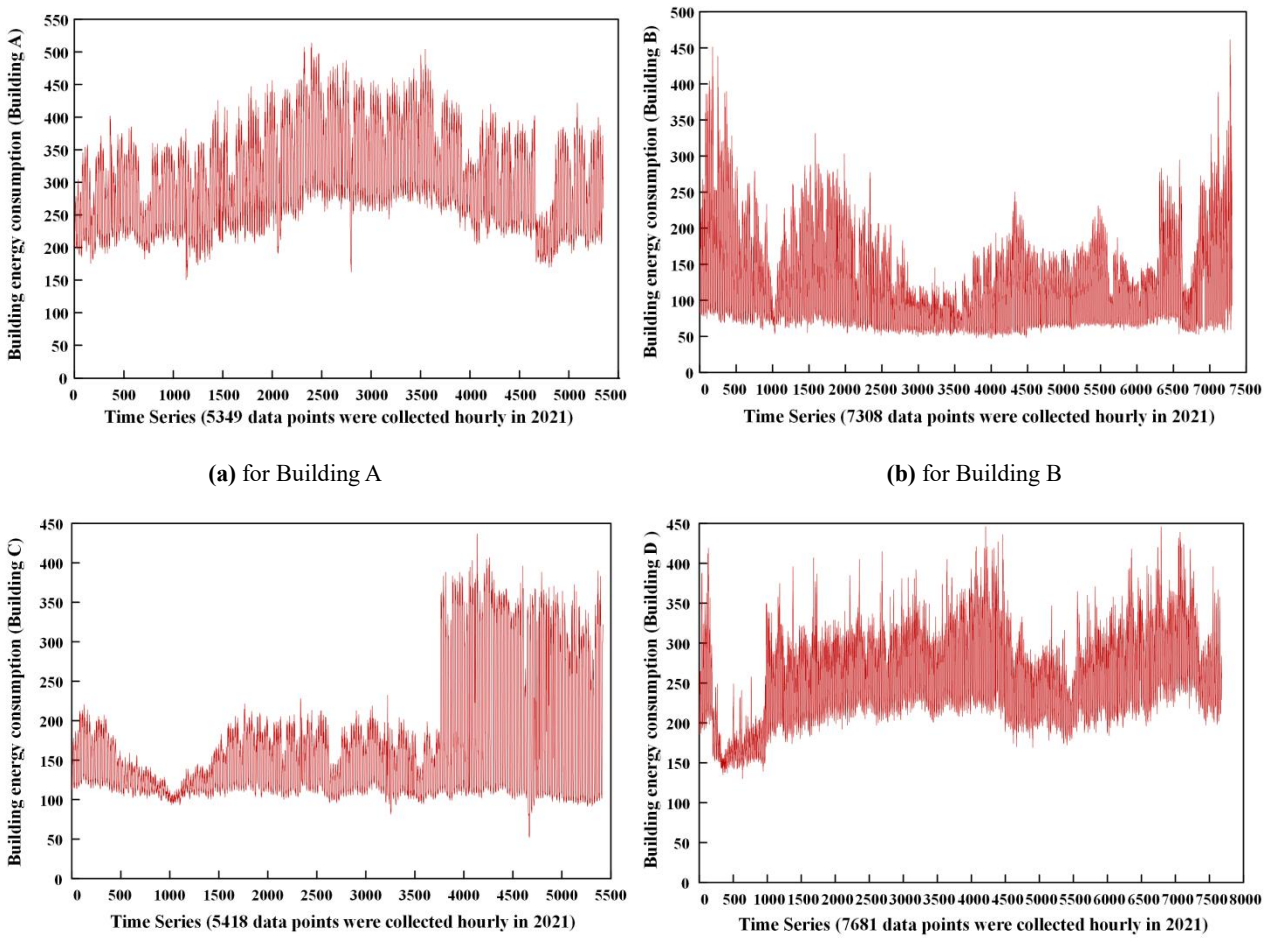


Figure 2 The monitoring system for building energy consumption

Based on the monitored data, Figures 3(a)–3(d) shows the energy consumption fluctuation characteristics for different buildings. Although these buildings are all offices, the energy consumption characteristics exhibit significant variations. Given that environmental air temperatures and meteorological conditions are similar, it demonstrates that

existing prediction frameworks may struggle to achieve sufficient adaptability clearly. There is an urgent need for improvements across various aspects, including algorithms and input data.



(a) for Building A **(b)** for Building B

(c) for Building C **(d)** for Building D

Figure 3 Annual energy consumption fluctuation characteristics for the four example buildings in 2021

2.2 Enhanced energy consumption forecasting

The proposed framework for building energy consumption forecasting incorporates three optimization strategies, with the simplified complete process shown in Figure 1.

First, weighted occupant behavior (for AC, light, plug usage) probabilities are introduced as novel inputs. This enhancement improves both the interpretability and accuracy of the forecast model [33].

Second, the physics-informed loss term, calculated using a building thermal R-C model, is integrated into the training process alongside the traditional data-informed loss, which occurs a combined loss function shown in equation (1). This enhancement improves the interpretability by grounding it in physical principles [34].

$$LOSS_{total} = \lambda_1 \times MSE(E_f, E_a) + \lambda_2 \times \frac{1}{n} \times \sum_{i=1}^{n} (LOSS_{physics})^2 \quad (1)$$

Where, $LOSS_{total}$ is the total loss in algorithmic training process; MSE is the function to calculate mean-square error which is the data-informed loss; $LOSS_{physics}$ is the physical-informed loss; n is the number of the training point; λ_1,

and λ_2 are the weights of data-informed loss and physical-informed loss, which are all 0.5 in this paper; E_f, and E_a are the forecast and actual values of energy consumption, kWh.

Third, a CNN module and an attention module are incorporated into the LSTM for data processing and dynamic weight optimization. This enhancement improves accuracy and efficiency [35]. Furthermore, some additional algorithmic modules, including SSA, transformer, etc., were also implemented for comparative performance analyses.

2.2.1 Occupant behavior probabilities and weights

Occupant behavior can be categorized into three types: environment-driven, time-driven, and random behaviors. This study focuses on energy-related behaviors, including AC, light, and plug usage within buildings [36].

Environment-driven behavior is primarily initiated by environmental stimulus. The probability can be calculated using equations (2–3). Equation (2) models behaviors triggered when an environmental stimulus exceeds a threshold, such as AC activated due to indoor air temperature surpassing a threshold. Conversely, Equation (3) models behaviors triggered when a stimulus falls below a threshold, such as turning on lights due to low indoor illuminance.

$$P = \begin{cases} 1 + P_{su} - e^{-\left(\frac{X - X_{su}}{l}\right)^k \Delta\tau}, X \geq X_{su} \\ P_{su}, \ X < X_{su} \end{cases} \tag{2}$$

$$P = \begin{cases} 1 + P_{su} - e^{-\left(\frac{X_{su} - X}{l}\right)^k \Delta\tau}, X \leq X_{su} \\ P_{su}, \ X > X_{su} \end{cases} \tag{3}$$

Where, P is the behavior probabilities; X is the environment stimulus, such as indoor air temperature for AC-usage behaviors [37] and outdoor solar radiation for light-usage behaviors [33]; X_{su} is the threshold of stimulus; P_{su} is the original probabilities, accounting for devices that remain normally operational, such as safety indicators; k, l, $\Delta\tau$ are all the fitting parameters, representing the occupant thermal sensitivity, the dimensionless constant, and the time-scale parameter respectively.

Time-driven behavior is primarily initiated by specific times or events, such as turning off lights when leaving. The probability can be calculated by equation (4).

$$P = \begin{cases} P_{su}, \ t \in [t_a, t_b] \\ 0, \ t \notin [t_a, t_b] \end{cases} \tag{4}$$

Where, t is the present moment; t_a, and t_b are the start and end moment for some time-driven behaviors.

Random behavior lacks explicit driving factors. Based on human dynamics theory [38], human behavior often exhibits a pattern of a short-term burst followed by a long period of silence, showing a non-uniform temporal distribution. Consequently, current behavior may be influenced by previous behaviors, and since energy consumption reflects behavioral patterns, the probability of random behavior can be modeled using equation (5–6). Other recent studies also suggest that random behavior can also be calculate by the number of occupants, or some similar parameters such as occupancy and indoor CO_2 concentration, which can indirectly model the random behaviors.

$$P = Bi\text{-}LSTM(E_{t-1}, E_{t-2}, E_{t-3}, t) \tag{5}$$

$$P = f(\Delta t) \tag{6}$$

Where, *Bi-LSTM* means a black-box algorithm; $E_{t\text{-}1}$, $E_{t\text{-}2}$, and $E_{t\text{-}3}$ are the energy consumption at one, two, and three previous time steps, kWh; f is a mapping relationship; Δt is the time interval between two behaviors.

Based on the survey, in the office buildings examined in this study, AC and light usage behaviors are determined by both environmental and temporal drivers. Plug usage is predominantly random, and some instances are also time-influenced. As the contribution of AC, light, and plug usage to total building energy consumption may vary, the weights of these behaviors are determined by multiple regression analysis, as shown in equation (7).

$$E_t^{total} = \varepsilon_0 + \varepsilon_1 P_t^{AC} + \varepsilon_2 P_t^{light} + \varepsilon_3 P_t^{plug} \tag{7}$$

Where, P_t^{AC}, P_t^{light}, and P_t^{plug} are the probabilities of AC, light and plug usage at moment t; E_t^{total} is the building energy consumption in moment t, kWh; the intercept ε_0 represents the base energy load independent of occupant behavior, kWh; while ε_1, ε_2, and ε_3 quantify the sensitivity of the total energy consumption to the corresponding behavior probabilities, effectively converting the dimensionless probabilities into energy contributions, kWh.

2.2.2 Building thermal network model (7R-3C)

The building thermal R-C model is employed to calculate the physics-informed loss term in equation (1). R-C models are widely adopted for room-level thermodynamic simulation and exist in various configurations [34]. Common variants include 3R-2C model for facade analyses and 5R-1C model for indoor air modeling [39–42]. As this study does not focus on precise thermal control, the complex convective and radiation between internal and external surfaces will be simplified. With assuming homogeneous thermal process, a simplified 7R-3C model will be used for building-level thermal modeling, as shown in Figure 4 and equation (8), which can be transformed into the following equation (9).

$$C_{build}\frac{dT_{in}}{dt} = \frac{T_{out}-T_{in}}{R_{wall}} + \frac{T_{out}-T_{in}}{R_{window}} + \frac{T_{out}-T_{in}}{R_{soil}} + \frac{T_{out}-T_{in}}{R_{roof}} + \frac{T_{device,s}-T_{in}}{R_{device}} + \frac{T_{human,s}-T_{in}}{R_{human}} \tag{8}$$

$$C_{build}\frac{dT_{in}}{dt} = \frac{T_{out}-T_{in}}{R_{wall}} + \frac{T_{out}-T_{in}}{R_{soil}} + \frac{T_{out}-T_{in}}{R_{roof}} + Q_{device} + Q_{solar} + Q_{human} \tag{9}$$

Where, T_{out}, and T_{in} are the outdoor and indoor air temperature on average, ℃; $T_{device,s}$, and $T_{human,s}$ are the average surface temperature of the devices and human, ℃; C_{build} is the overall heat capacitance of indoor air, kJ/K; R_{wall}, R_{window}, R_{soil}, R_{roof}, R_{device}, and R_{human} are the thermal resistance from facade walls, windows (including all transparent envelope), floors, roofs, devices, and occupants to the indoor air, K/kW; Q_{device} is the heat generation from devices and AC heating provided, kW; Q_{solar} is solar radiation heat generation through transparent enclosures, kW; Q_{human} is the heat generation from human metabolism, kW.

Some terms in equation (9) can be further solved as follows.

$$Q_{device} = \varepsilon_1 \times P_{plug} + \varepsilon_2 \times P_{light} + a \times \varepsilon_3 \times P_{ac} \tag{10}$$

$$Q_{solar} = \beta \times SHGC \times A_w \times I_{solar} \tag{11}$$

$$Q_{human} = N \times \gamma \tag{12}$$

$$\frac{1}{R_{o\text{-}envelope}} = \frac{R_{soil}R_{roof} + R_{wall}R_{roof} + R_{wall}R_{soil}}{R_{wall}R_{soil}R_{roof}} \tag{13}$$

Where, P_{plug}, P_{light}, and P_{ac} are the power of plug, light, and AC usage, kW; ε_1, and ε_2 are the heat generation coefficient for electrical plug loads (such as computers) and lights; ε_3 is the cooling or heating efficiency of AC

systems; a is the cooling-heating switching coefficient (+1 for heating and -1 for cooling); β is the ratio of solar radiation heat gain absorbed by indoor thermal mass to the total solar radiation heat gain transmitted through transparent windows; *SHGC* is the solar heat gain coefficient; $A_w$is the window area, m²; I_{solar} is the solar radiation intensity, kJ/m²; N is the number of occupants in the building; γ is the average metabolic heat per person, kW/person; $R_{o\text{-}envelope}$ is the total thermal resistance of the opaque envelope (including exterior walls, roof, and floor), K/kW.

If the occupant count is not directly measured, it can be estimated using design occupancy density or inferred from dynamics indoor CO_2 concentration, as shown in equations (14–15). Consequently, equation (8) is reformulated as equation (16). When there are only building energy data instead of all device energy data, an experiment can be conducted to comprehensively test the heat generated by lighting fixtures, electronic devices such as computers, and the heating or cooling from ACs, converting these into the total thermal generation of all devices in buildings, calculated by total energy consumption, shown in equation (17). Based on these equations, the physics-informed loss in equation (1) can be calculated by equation (18).

$$N = \rho \times A_b \tag{14}$$

$$V_b \frac{dC_{CO_2}^{in}}{dt} = N \times G_{CO_2} + Q_v \times (C_{CO_2}^{out} - C_{CO_2}^{in}) \tag{15}$$

$$C_{build} \frac{dT_{in}}{dt} = \frac{T_{out} - T_{in}}{R_{o\text{-}envelope}} + \varepsilon_1 \times P_{plug} + \varepsilon_2 \times P_{light} + a \times \varepsilon_3 \times P_{ac} + \beta \times SHGC \times A \times I_{solar} + N \times \gamma \tag{16}$$

$$C_{build} \frac{dT_{in}}{dt} = \frac{T_{out} - T_{in}}{R_{o\text{-}envelope}} + \delta_1 \times P_{build} + \delta_2 \times I_{solar} + \delta_3 \times N \tag{17}$$

$$LOSS_{physics} = \frac{T_{out} - T_{in}}{R_{o\text{-}envelope}} + \varepsilon_1 \times P_{plug} + \varepsilon_2 \times P_{light} + a \times \varepsilon_3 \times P_{ac} + \beta \times SHGC \times A \times I_{solar} + \rho \times A_b \times \gamma - C_{build} \frac{dT_{in}}{dt} \tag{18}$$

Where, ρ is the design occupancy density, person/m²; A_b is the building area, m²; V_b is the building volume, m³; Q_v is the ventilation volume, m³/h; $C_{CO_2}^{out}$, and $C_{CO_2}^{in}$ are the outdoor and indoor CO_2 concentration at moment *t,* ppm; G_{CO_2} is the metabolism CO_2 production per person per hour, m³/(person·h); δ_1, δ_2, and δ_3 are fitting parameters.

Referring to recent studies and the ASHRAE standard [43–45], ε_1 is taken as 0.88, and ε_2 is 0.60. Based on the device performance and weather information, ε_3 is taken as 4.0, and *SHGC* is 0.48. The area of buildings, walls and windows are shown in Table 1. And the T_{in}, T_{out}, P_{plug}, P_{light}, P_{ac}, and I_{solar} are monitored in real time. The remaining unknown parameters, C_{build} and $R_{o\text{-}envelope}$, can be determined by nonlinear least-squares fitting to minimize the $LOSS_{physics}$ using *lsqnonlin* function in MATLAB 2024a [46], as shown in equation (18) and Figure 4.

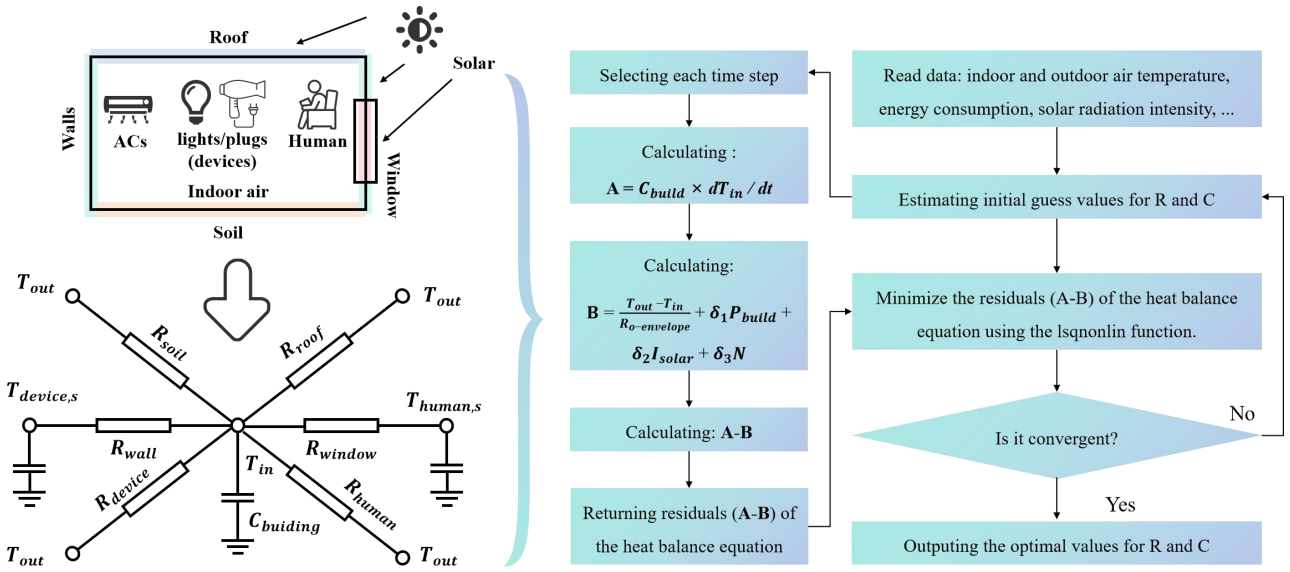


Figure 4 The calculation process of R and C of the thermal R-C models

2.2.3 Training algorithm modules

The LSTM network serves as the foundational forecasting algorithm due to its proven efficacy in modeling time-series data. To enhance its performance, several advanced algorithmic modules are integrated including CNN, SSA, Transformer, and attention mechanism. This study proposes a CNN-LSTM-Attention as the forecast model. For comparison, other advanced hybrid algorithms, including Transformer-LSTM-Adaboost and LSTM-SSA, are also implemented [47–51], which incorporate SSA, Transformer, and Adaboost modules into the LSTM framework incorporating optimization for hyperparameter tuning, weight adjustment, and data processing.

For CNN-LSTM-Attention, the CNN module performs feature extraction and dimensionality reduction to improve generalization, and the attention mechanism dynamically assigns weights to different input feature channels, enabling the model to focus on the most relevant information. This integration enhances prediction accuracy while maintaining computational efficiency [33, 47–48]. The overall process is shown in Figure 5.

For LSTM-SSA, it uses SSA to optimize the key LSTM hyperparameters such as learning rate, the number of iterations and the number of hidden units [49]. For Transformer-LSTM-Adaboost, the Transformer module employs self-attention mechanism within an encoder-decoder structure, which uses query, key, and value matrices to capture dependencies among different positions in the input sequence dynamically. The Adaboost module serves as a meta-learner, adaptively adjusting weights by iterative training of the Transformer-LSTM (basic learner), to improve the model generalization ability [50–51].

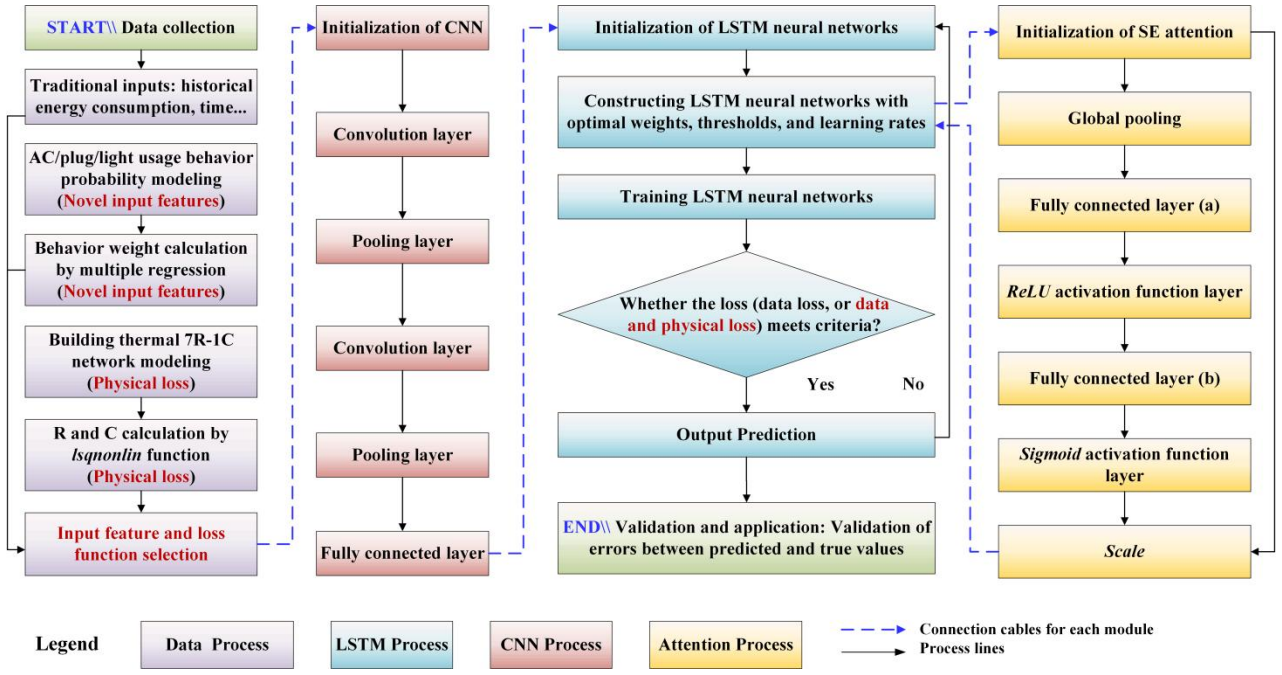


Figure 5 The process and connection of the CNN-LSTM-Attention with novel inputs and loss function

As these algorithmic are not novel contributions of this study, their detailed mathematical formulations are omitted from the main text for brevity, provided in Appendix A. The key hyperparameters are shown in Table 2.

Table 2 The key hyperparameters of different algorithmic modules

Modules	Hyperparameters	Values	Modules	Hyperparameters	Values
CNN	Maximum number of iterations	100	LSTM	Number of hidden layers	64
CNN	Initial learning rate	0.005	WOA	Population size	10
CNN	Learning rate decay period	80	WOA	Maximum number of iterations	20
CNN	Learning rate decay coefficient	0.8	SSA	Population size	10
CNN	Minimum batch size	24	SSA	Maximum number of iterations	20
CNN	Activation function	*ReLU*	SSA	Threshold of danger	70%
CNN	Convolutional kernel size	64	SSA	Proportion of producers	40%
LSTM	Maximum number of iterations	1000	SSA	Proportion of realizing danger	20%
LSTM	Initial learning rate	0.01	Attention	Activation function	*Sigmoid, ReLU*
LSTM	Learning rate decay period	800	Attention	Number of neurons	128
LSTM	Learning rate decay coefficient	0.8	Transformer	Position encoding dimension	256
LSTM	Minimum batch size	240	Transformer	Number of self attention heads	4
LSTM	Activation function (Output gate)	*Tanh*	Transformer	Key/Value Vector Dimension	128
LSTM	Activation function (Input/Forget gate)	*Sigmoid*	Adaboost	Number of basic learners	10
LSTM	Number of neurons	32	No	Ratio of training to test	7:3

2.3 Evaluation parameters

The performance of the proposed method is evaluated using five indices: R^2, MAE, MAPE, RMSE, and CV-RMSE, computed using equations (19–23).

$$R^2\ (X,Y) = 1 - \frac{\sum_1^n (X-Y)^2}{\sum_1^n (Y-\overline{Y})^2} \tag{19}$$

$$MAE\ (X,Y) = \frac{1}{n}\ \sum_1^n |X - Y| \tag{20}$$

$$MAPE\ (X,Y) = \frac{1}{n}\ \sum_1^n |\frac{X-Y}{Y} \times 100\%| \tag{21}$$

$$RMSE(X,Y) = \sqrt{\frac{1}{n}\sum_1^n (X-Y)^2} \tag{22}$$

$$CV - RMSE(X,Y) = \frac{RMSE(X,Y)}{\overline{Y}} \tag{23}$$

Where, X and Y is the forecast and actual energy consumption, kWh. $\overline{Y}$ is the average of the actual value, kWh. In addition, it is important to evaluate the time required for implementation. In this study, the time consumption is recorded using the timing program provided by MATLAB. The program runs on a fixed computer (Model: Lenovo, TianYi 510 pro, RAM: 16GB; CPU: i7-13700; GPU: NVIDIA GeForce GT 730), with a consistent operator and other influencing factors such as parameter selection kept constant throughout the analyses.

3 Cases and results

Comprehensive case studies were designed to evaluate forecasting performance, encompassing a total of 48 unique cases. These cases were generated by combining four variables: three forecasting algorithms, two kinds of loss functions, two kinds of input features, and four different buildings. And the output is the building energy consumption.

The two kinds of inputs are traditional and novel. Traditional inputs include time, indoor air temperature and CO_2 concentration, outdoor air temperature, solar radiation intensity, cloudiness, and historical energy consumption. Novel inputs include all traditional inputs and weighted occupant AC, light, and plug usage behavior probabilities.

The three kinds of algorithms are Transformer-LSTM-Adaboost, CNN-LSTM-Attention, and LSTM-SSA.

The two kinds of loss functions are data-informed loss, and combined data-informed and physics-informed loss.

While the input features include historical energy consumption, the model is structured as a multi-variate regression forecast model rather than a pure time-series forecast model. The lagged energy consumption is treated as a supplementary static feature that provides valuable information about the building operation state, similar to indoor air temperature and occupant behavior probabilities. The model learns a direct mapping from the feature vector at a given time to the energy consumption at the same time. Consequently, since the prediction for each time step is made independently based on its corresponding features, the entire dataset is partitioned into a training set (70%) and a testing set (30%) randomly, and the results are the average results based on five-time simulation for each case.

3.1 Results on behavior probabilities

The probabilities for AC, light, and plug usage behaviors, calculated by equations (2–6), were categorized into environment-driven, time-driven, and random types. Based on energy monitoring and field investigations, five common behavioral modes were identified and summarized as follows:

Table 3 summarizes the predominant behavior modes identified for each building. For example, while lights in Building C could be operated freely, field investigation revealed that occupants typically turned all lights on when entering and off when leaving. Therefore, Mode b is selected for light usage behavior in Building B instead of Mode c.

It is important to note that energy-related behavior is defined broadly in this context. As the studied office buildings include scientific laboratories, AC systems were sometimes used to maintain precise temperature and humidity levels in unoccupied areas and time (for example, data centers, experimental animal facilities, and precision instrument rooms). The resulting calculated behavior probabilities are shown in Figures 6–9.

Table 3 The energy usage modes and schedules in different buildings by field investigation

Building	AC usage modes	Light usage modes	Plug usage modes
Building A	Heating: Mode a Cooling: Mode a	Mode c	Mode e
Building B	Heating: Mode a Cooling: Mode a	Mode c	Mode e
Building C	Heating: No usage Cooling: Mode c	Mode b	Mode e
Building D	Heating: No usage Cooling: Mode a	Mode c	Mode e

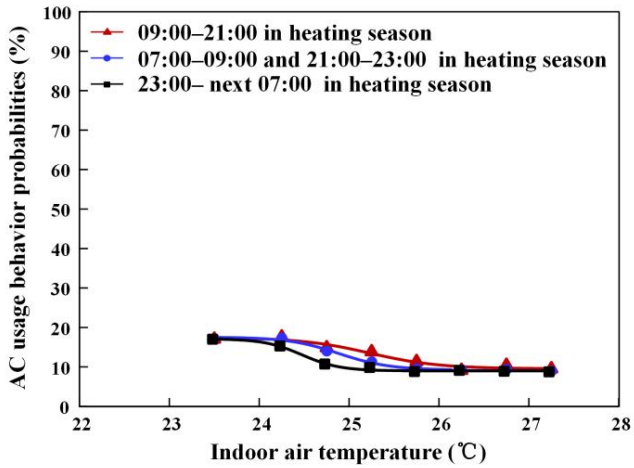
Mode a: Human tend to turn on, turn off, and adjust the devices only due to their own feelings influenced by environment parameters;

Mode b: Human tend to turn on the devices when entering and off when leaving without adjustments;

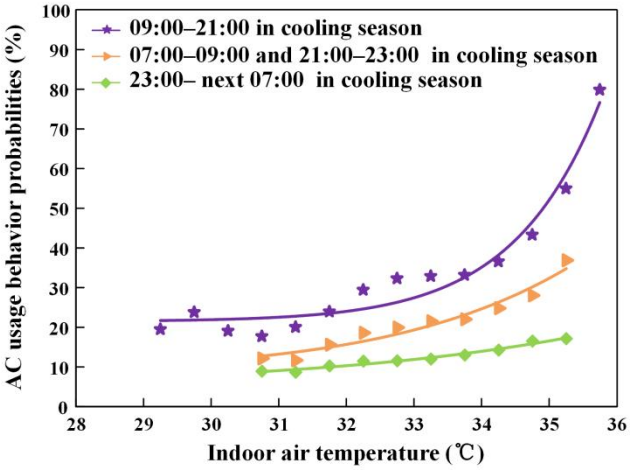
Mode c: Human tend to turn on and adjust the devices due to their own feelings influenced by environment parameters, and turn off the devices when leaving;

Mode d: Human tend to turn on and adjust the devices when entering, and turn off the devices only due to their own feelings influenced by environment parameters, and turn off the devices when leaving.

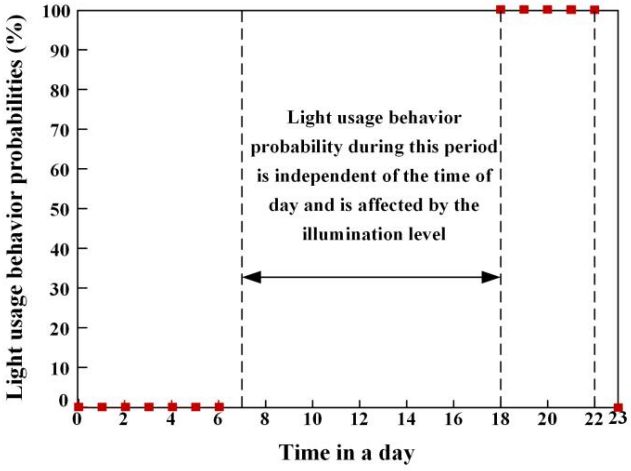
Mode e: Human use the devices completely at random, unaffected by environment or time.



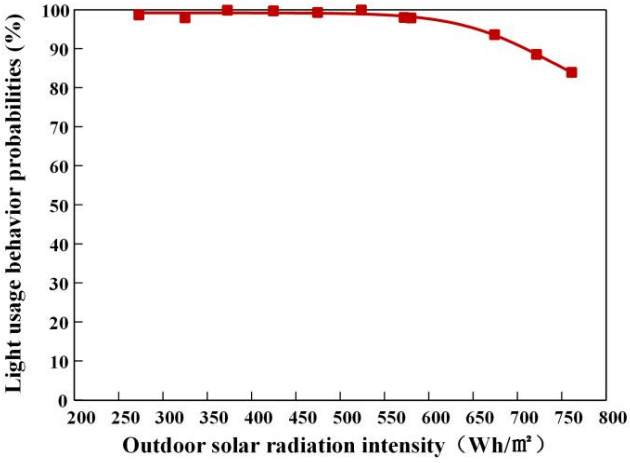
(a) AC usage in the heating season



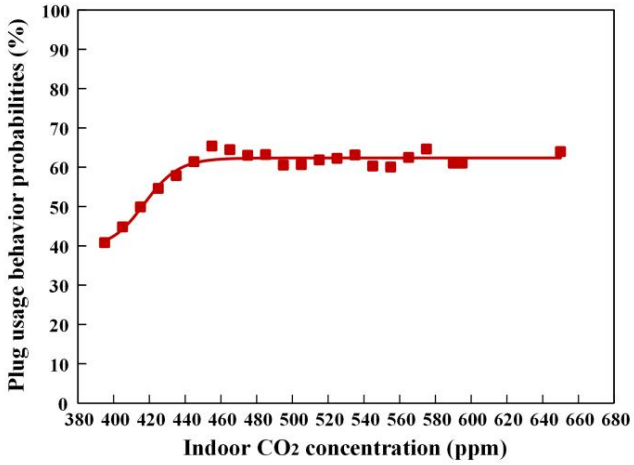
(b) AC usage in the cooling season



(c) Light usage during 18:00–next 7:00

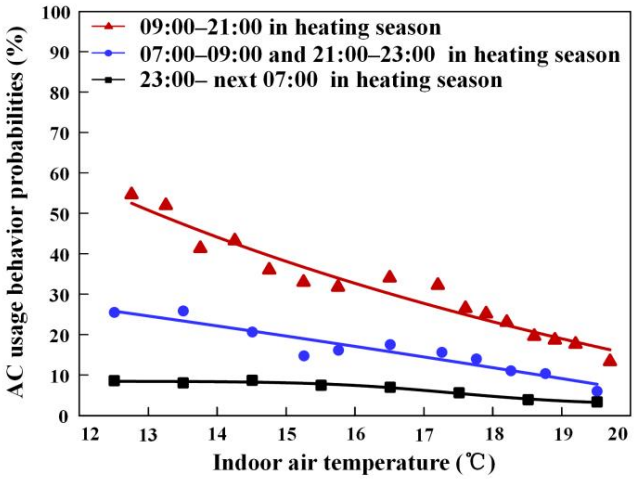


(d) Light usage during 7:00–18:00

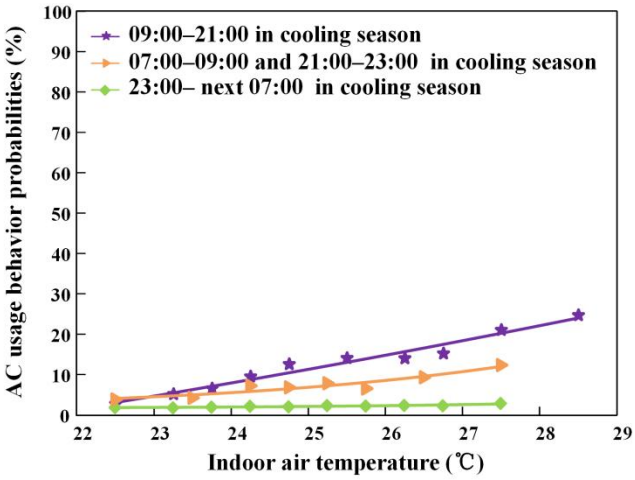


(e) Plug usage

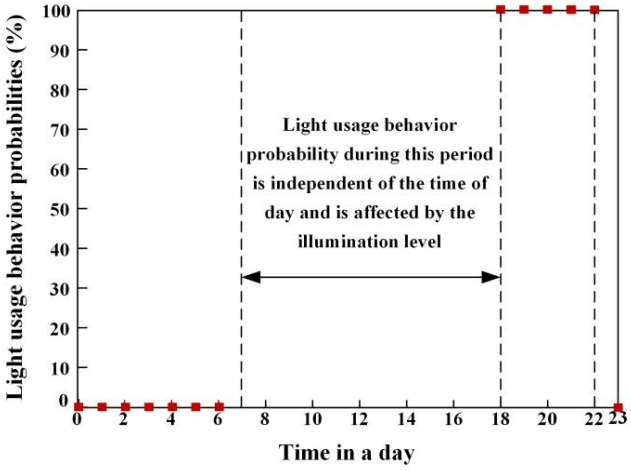
Figure 6 The occupant behavior probabilities on average in Building A



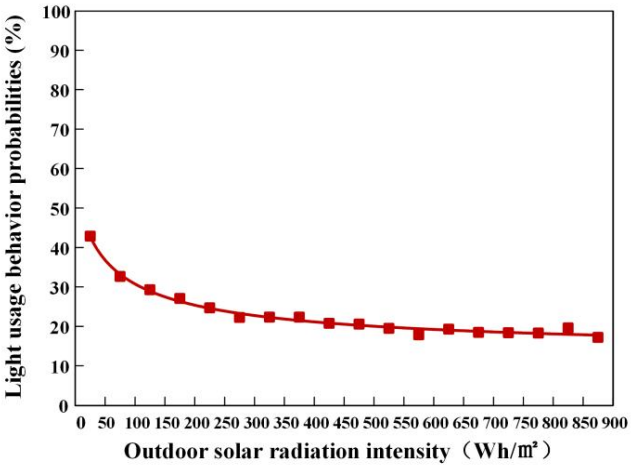
(a) AC usage in the heating season



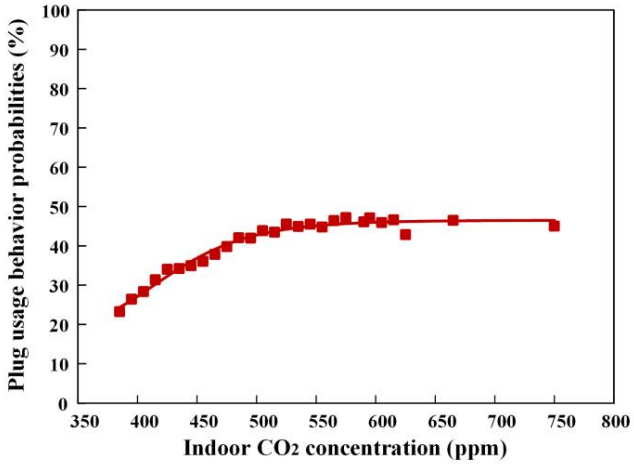
(b) AC usage in the cooling season



(c) Light usage during 18:00–next 7:00

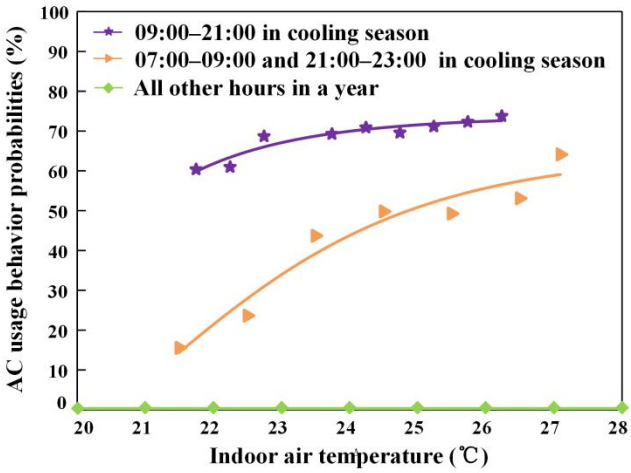


(d) Light usage during 7:00–18:00

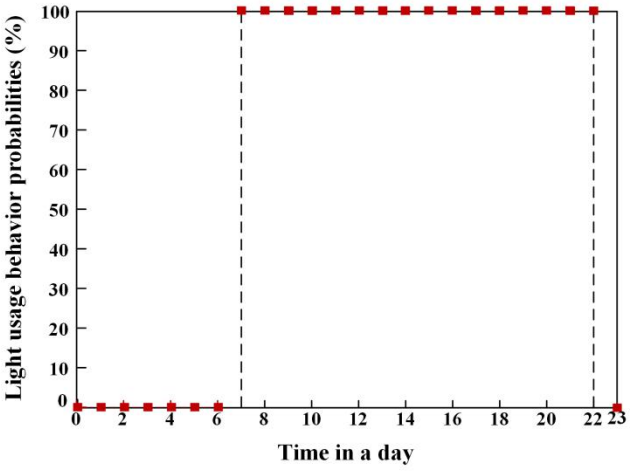


(e) Plug usage

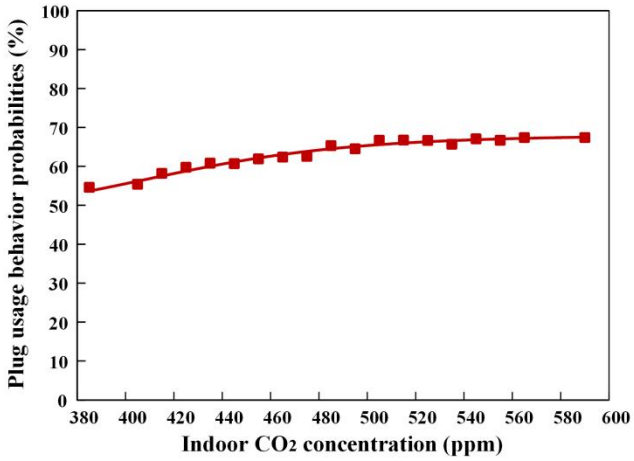
Figure 7 The occupant behavior probabilities on average in Building B



(a) AC usage

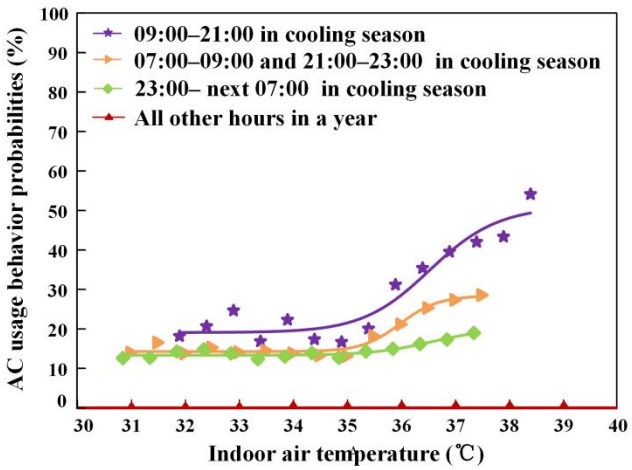


(b) Light usage

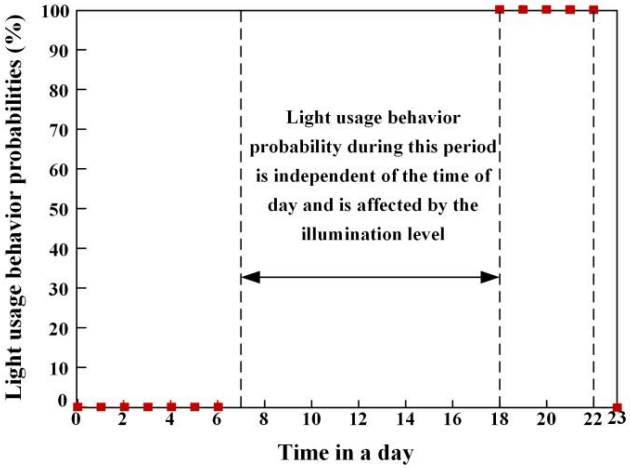


(c) Plug usage

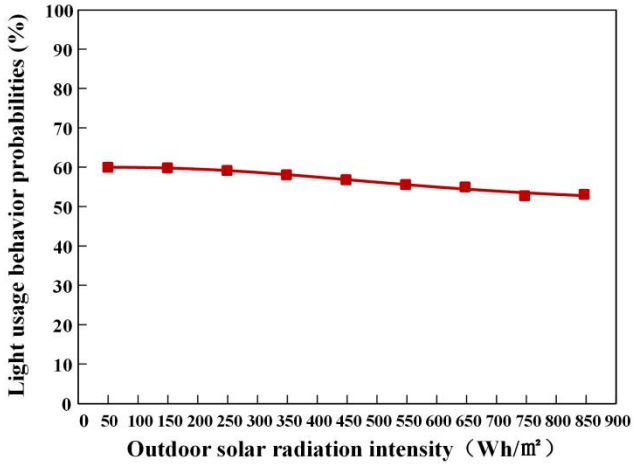
Figure 8 The occupant behavior probabilities on average in Building C



(a) AC usage



(b) Light usage during 18:00–next 7:00

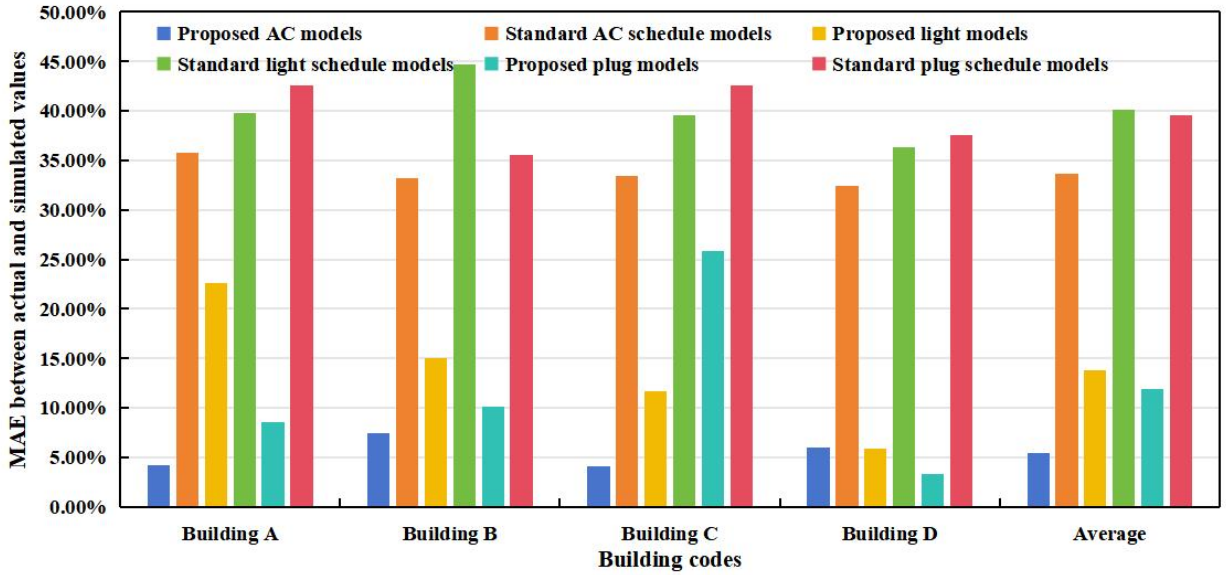


(c) Light usage during 7:00–18:00

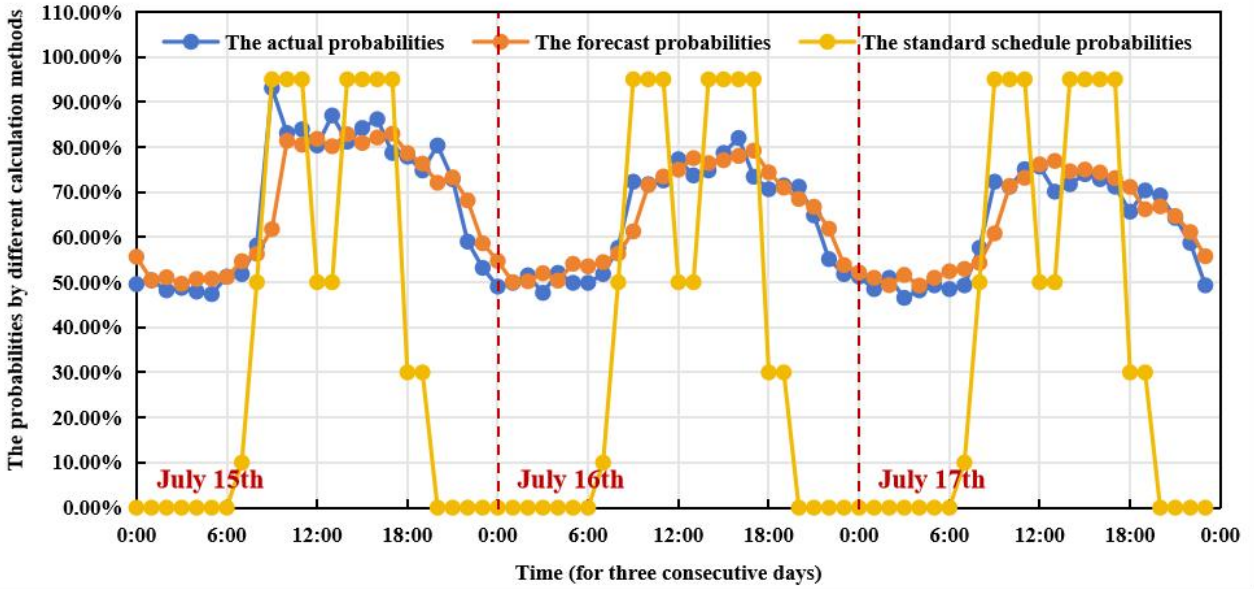
Figure 9 The occupant behavior probabilities on average in Building D (plug usage calculated by Bi-LSTM rather than fitting)

The modeling accuracy for these behavior probabilities was quantitatively assessed, of which MAE ranges were

4.06%–7.47% (mean: 5.44%) for AC use, 5.84%–22.58% (mean: 13.78%) for light use, and 3.34%–25.90% (mean: 11.96%) for plug use. For comparison, based on the *Design Standard for Energy Efficiency of Public Buildings* (*GB 50189-2015*) in China [52], the MAE values between standardized schedules and actual probabilities are significantly higher, which are 33.69% for AC use, 40.09% for light use, and 39.53% for plug use on average. A comparison of MAE values between the proposed models and the standard schedules is shown in Figures 10(a)–10(b). It shows that standard schedules fail to accurately reflect actual energy use patterns, particularly for specialized end-uses such as safety lighting and HVAC for data centers or laboratories. These results substantiate the necessity of employing the proposed occupant behavior modeling approach.



(a) The average MAE for the whole year



(b) For three consecutive days

Figure 10 Comparison of behavior probabilities calculated by standard schedules and the proposed models

3.2 Results on building energy consumption forecasting

Table 4 shows the forecast results of different cases with different inputs, algorithms, and loss functions.

Table 4 The forecast results of different cases with different inputs, algorithms, and loss functions

NO.	Building	Inputs	Algorithms	Loss	MAE	MAPE	RMSE	CV-RMSE	R^2	TIME(s)
1	A	Traditional	CNN-LSTM-Attention	P-D loss	16.9023	5.73%	20.7442	0.0732	0.9002	534
2	A	Novel	CNN-LSTM-Attention	P-D loss	16.525	5.59%	20.3085	0.0725	0.9119	525
3	A	Traditional	CNN-LSTM-Attention	D loss	17.1762	5.92%	21.9923	0.0753	0.8995	514
4	A	Novel	CNN-LSTM-Attention	D loss	16.8122	5.69%	21.9364	0.0737	0.9058	534
5	A	Traditional	LSTM-SSA	P-D loss	17.8947	6.08%	22.7135	0.0758	0.8472	7230
6	A	Novel	LSTM-SSA	P-D loss	16.915	5.74%	21.9661	0.0742	0.8514	7090
7	A	Traditional	LSTM-SSA	D loss	18.2412	6.28%	24.2284	0.0783	0.8357	7450
8	A	Novel	LSTM-SSA	D loss	17.8355	6.01%	23.7699	0.0782	0.8438	8840
9	A	Traditional	Transformer-LSTM-Adaboost	P-D loss	17.9904	6.28%	23.7848	0.0774	0.8908	735
10	A	Novel	Transformer-LSTM-Adaboost	P-D loss	17.9702	6.23%	23.584	0.0771	0.8926	703
11	A	Traditional	Transformer-LSTM-Adaboost	D loss	19.0016	6.50%	24.6803	0.0815	0.8824	684
12	A	Novel	Transformer-LSTM-Adaboost	D loss	18.0005	6.25%	23.9298	0.078	0.8895	666
13	B	Traditional	CNN-LSTM-Attention	P-D loss	13.59	11.01%	21.3246	0.172	0.8872	407
14	B	Novel	CNN-LSTM-Attention	P-D loss	13.1539	10.31%	20.3636	0.1466	0.8932	388
15	B	Traditional	CNN-LSTM-Attention	D loss	14.8471	11.95%	23.3239	0.1922	0.8566	420
16	B	Novel	CNN-LSTM-Attention	D loss	13.2646	10.60%	21.5031	0.1681	0.8926	412
17	B	Traditional	LSTM-SSA	P-D loss	15.6603	12.77%	25.2528	0.2157	0.8197	7570
18	B	Novel	LSTM-SSA	P-D loss	15.15	12.28%	23.3859	0.1848	0.8349	7490
19	B	Traditional	LSTM-SSA	D loss	17.5397	13.35%	27.6699	0.2419	0.8127	8090
20	B	Novel	LSTM-SSA	D loss	15.7739	13.10%	23.4475	0.2035	0.8346	6870
21	B	Traditional	Transformer-LSTM-Adaboost	P-D loss	19.9723	18.42%	26.7767	0.2055	0.8154	566
22	B	Novel	Transformer-LSTM-Adaboost	P-D loss	15.0608	12.75%	20.7797	0.1644	0.8888	554
23	B	Traditional	Transformer-LSTM-Adaboost	D loss	22.4751	22.04%	28.8096	0.2142	0.7863	526
24	B	Novel	Transformer-LSTM-Adaboost	D loss	15.4593	13.20%	21.5116	0.1681	0.8808	548
25	C	Traditional	CNN-LSTM-Attention	P-D loss	13.0685	7.47%	22.841	0.146	0.8989	541
26	C	Novel	CNN-LSTM-Attention	P-D loss	11.2351	6.46%	19.9104	0.1228	0.9254	584
27	C	Traditional	CNN-LSTM-Attention	D loss	13.4631	7.92%	23.667	0.1479	0.8878	514
28	C	Novel	CNN-LSTM-Attention	D loss	12.3376	7.43%	21.6855	0.1291	0.9129	552
29	C	Traditional	LSTM-SSA	P-D loss	14.4486	8.88%	23.9846	0.1598	0.8693	6980
30	C	Novel	LSTM-SSA	P-D loss	13.6927	7.40%	23.1098	0.1382	0.9046	7410
31	C	Traditional	LSTM-SSA	D loss	14.4814	9.06%	27.2999	0.171	0.8306	8050
32	C	Novel	LSTM-SSA	D loss	13.9669	7.69%	24.3809	0.1454	0.8908	9280

33	C	Traditional	Transformer-LSTM-Adaboost	P-D loss	25.9803	17.24%	33.4838	0.1959	0.7938	674
34	C	Novel	Transformer-LSTM-Adaboost	P-D loss	16.4865	9.73%	24.2819	0.1502	0.8916	728
35	C	Traditional	Transformer-LSTM-Adaboost	D loss	26.6521	17.44%	35.0189	0.2044	0.7729	602
36	C	Novel	Transformer-LSTM-Adaboost	D loss	16.5604	9.84%	25.665	0.1569	0.878	721
37	D	Traditional	CNN-LSTM-Attention	P-D loss	13.2241	5.10%	18.1837	0.0706	0.8947	417
38	D	Novel	CNN-LSTM-Attention	P-D loss	11.7376	4.50%	15.9993	0.0617	0.9203	416
39	D	Traditional	CNN-LSTM-Attention	D loss	14.0007	5.47%	18.4779	0.071	0.8938	435
40	D	Novel	CNN-LSTM-Attention	D loss	11.9814	4.59%	16.387	0.0628	0.9132	438
41	D	Traditional	LSTM-SSA	P-D loss	14.0237	5.32%	19.1384	0.0751	0.8905	9880
42	D	Novel	LSTM-SSA	P-D loss	11.839	4.66%	16.4688	0.0645	0.9161	5310
43	D	Traditional	LSTM-SSA	D loss	15.72	6.31%	20.1459	0.0762	0.8689	8740
44	D	Novel	LSTM-SSA	D loss	12.4199	4.76%	16.922	0.0655	0.9111	8850
45	D	Traditional	Transformer-LSTM-Adaboost	P-D loss	16.1136	6.45%	20.5644	0.0776	0.8694	560
46	D	Novel	Transformer-LSTM-Adaboost	P-D loss	13.5889	5.28%	17.975	0.0669	0.9015	604
47	D	Traditional	Transformer-LSTM-Adaboost	D loss	16.367	6.64%	20.7385	0.0788	0.8635	504
48	D	Novel	Transformer-LSTM-Adaboost	D loss	13.6328	5.37%	18.0387	0.0686	0.8949	605

Note: P-D loss mean the combined data-informed and physics-informed loss, while D loss is the data-informed loss.

To evaluate performance improvement achieved by incorporating behavior-related inputs, physics-informed loss, and CNN-LSTM-Attention algorithm, the MAPE, CV-RMSE, and R^2 were employed as evaluation metrics, since these dimensionless metrics facilitate direct comparison with other studies.

(1) Regarding the proposed methods with all three optimization strategies (such as Case 10), it will be compared to traditional methods (such as Cases 3, 7, and 11). The results show the enhancements across all these strategies.

For accuracy: The proposed method significantly improves forecasting accuracy, achieving an average reduction in MAPE of 25.78%, a reduction in CV-RMSE of 21.31%, and an improvement in R^2 of 7.71%.

For efficiency: A substantial reduction in computational time was observed, with an average saving of 2566 seconds. Even after excluding the LSTM-SSA algorithm due to its high computational time consumption and the CNN-LSTM-Attention own, the proposed method still yielded a net time saving of 101s on average, effectively offsetting the additional overhead from introducing novel inputs and loss function.

For interpretability: The incorporation of occupant behavior rules and building thermodynamics can enhance the model interpretability by providing a physical basis for the predictions.

In summary, the integrated method shows a balance among forecasting accuracy, computational efficiency, and model interpretability. The average performance improvement in different buildings is shown in Figure 11.

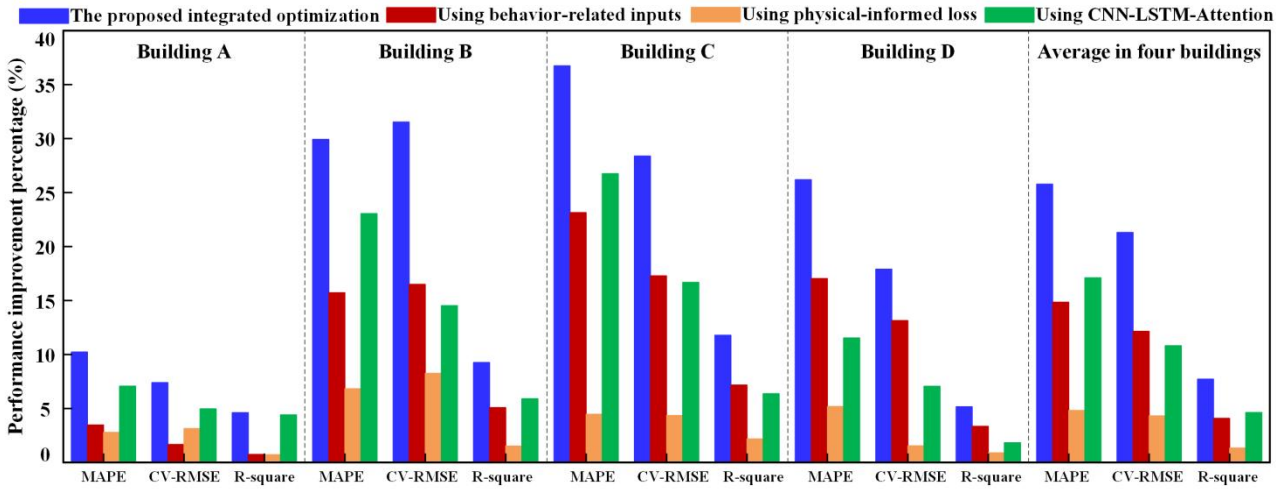


Figure 11 The average performance improvement percentage in different buildings by using different optimization modules

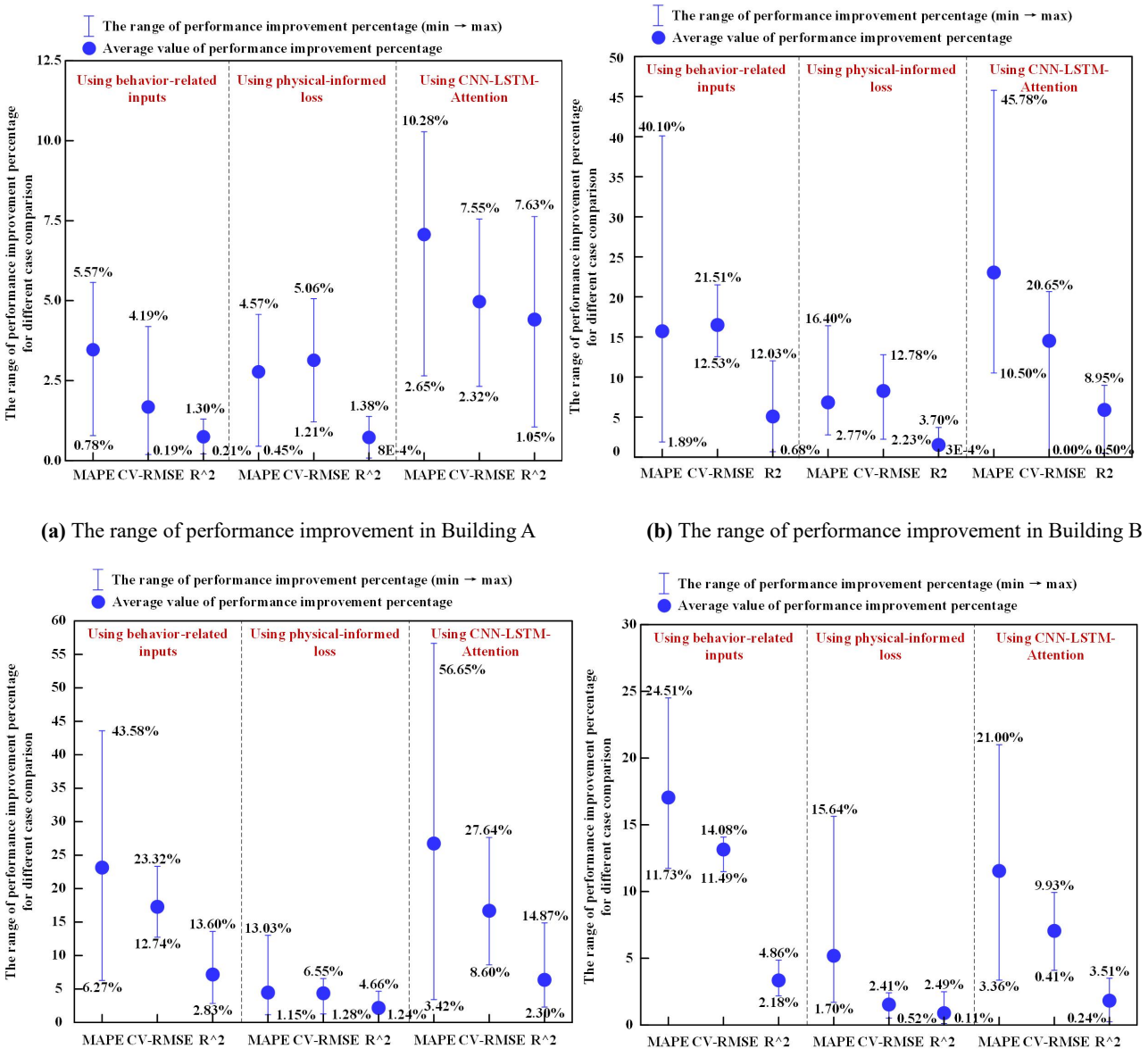
(2) Regarding the proposed methods with individual optimization strategies, the contribution of each strategy was isolated and analyzed as follows:

For accuracy: Comparing Case 1 to Case 2 and similar case combination, incorporating behavior-related inputs can reduce MAPE by 0.78%–43.58% (mean: 14.84%), CV-RMSE by 0.19%–23.32% (mean: 12.15%), and R^2 by 0.21%–13.60% (mean: 4.08%) on average. Comparing Case 1 to Case 3 and similar case combination, incorporating physical-informed loss can reduce MAPE by 0.45%–16.40% (mean: 4.82%), CV-RMSE by 0.52%–12.78% (mean: 4.32%), and R^2 by 0.03%–4.66% (mean: 1.33%) on average. Comparing Case 1 to Cases 5 and similar case combination can find that incorporating CNN-LSTM-Attention can reduce MAPE by 2.65%–56.65% (mean: 17.10%), CV-RMSE by 0%–27.64% (mean:10.81%), and R^2 by 0.24%–14.87% (mean: 4.63%) on average. Figure 12 shows the average performance improvement in different buildings by comparing the proposed methods and others.

For efficiency: Incorporating CNN-LSTM-Attention can reduce the time consumption by 7343.6875s compared to Transformer-LSTM-Adaboost, and by 146.8125s compared to LSTM-SSA on average. Although the inclusion of behavior-related inputs and the physics-informed loss increases the number of simulation steps, the associated rise in computational time was not significant.

For interpretability: Similar to the integrated method, the individual introduction of behavior-related inputs and the physics-informed loss function can improve model interpretability.

In conclusion, all three optimization strategies contribute to improved forecasting accuracy, with different weights as shown in Figure 12. Furthermore, CNN-LSTM-Attention algorithm offers the additional advantage of significantly enhancing computational efficiency. Finally, the behavior-related inputs and physics-informed loss can all improve model interpretability without incurring substantial additional computational time consumption.



(a) The range of performance improvement in Building A

(b) The range of performance improvement in Building B

(c) The range of performance improvement in Building C

(d) The range of performance improvement in Building D

Figure 12 The range of performance improvement in different buildings through different improvement modules

4 Discussions

4.1 The contribution and priority of different enhanced strategies

As shown in Figures 11–12, the three optimization strategies collectively enhance both the accuracy and efficiency of the model. Nevertheless, their relative contributions vary across different building types, as shown in Figure 13.

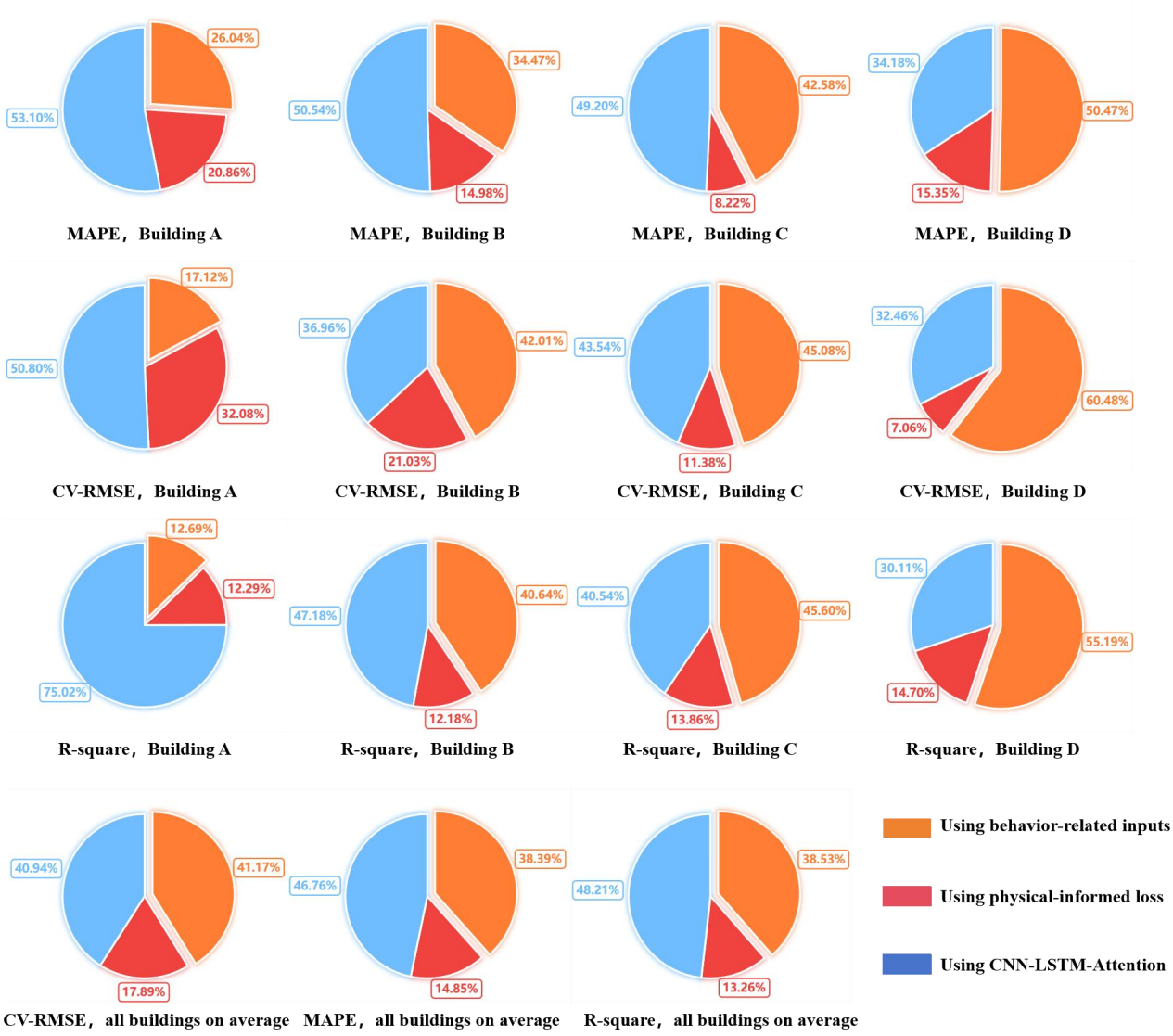


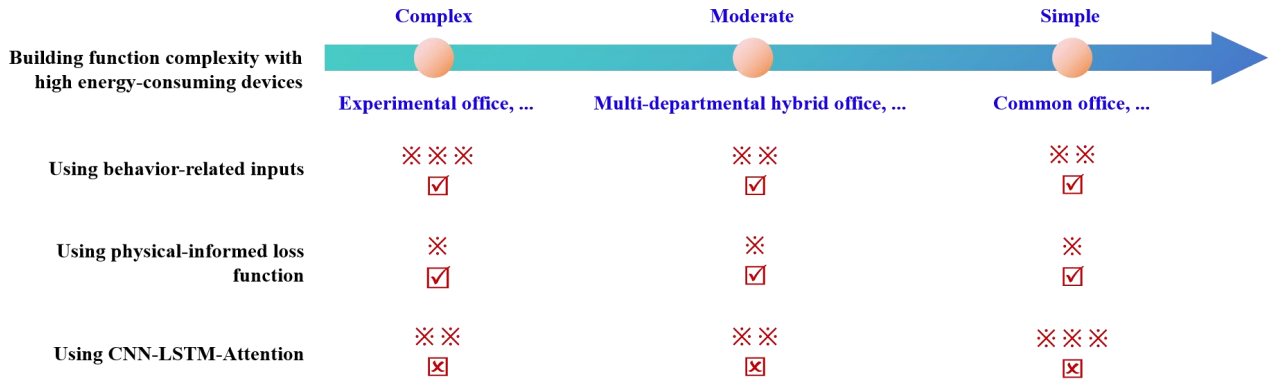
Figure 13 The average contribution ration from different optimization modules in different buildings

In summary, behavior-related inputs contributed to 26.04–50.47% (mean: 38.39%) of the MAPE reduction, 17.12–60.48% (mean: 41.17%) of the CV-RMSE reduction, and 12.69–55.19% (mean: 38.53%) of the R^2 improvement. Physics-informed loss function contributed to 8.22–20.86% (mean: 14.85%) of the MAPE reduction, 7.06–32.08% (mean: 17.89%) of the CV-RMSE reduction, and 12.18–14.70% (mean: 13.26%) of the R^2 improvement. Finally, CNN-LSTM-Attention contributed to 34.18–53.10% (mean: 46.76%) of the MAPE reduction, 32.46–50.80% (mean: 40.94%) of the CV-RMSE reduction, and 30.11–75.02% (mean: 48.21%) of the R^2 improvement. On average, the contributions of behavior-related inputs, the physics-informed loss function, and the CNN-LSTM-Attention may be approximately 40%, 15%, and 45%, respectively.

Among these strategies, the physics-informed loss function exhibited the most consistent contribution across buildings. In contrast, although average contributions of behavior-related inputs and the CNN-LSTM-Attention algorithm are around 40% and 45%, respectively, their impacts varied considerably. Specifically, behavior-related inputs had a weaker effect in Building A but a stronger one in Building D, while the CNN-LSTM-Attention algorithm

showed the opposite trend. The contributions for Buildings B and C were closer to the average values.

The observed discrepancies can be attributed to the following factors. In Building A, occupant behavior is highly irregular, and energy consumption is dominated by specialized equipment such as biological cultivation devices and clean AC systems, of which operation is stochastic and less influenced by conventional environmental cues. Consequently, behavior-related inputs offered limited forecasting accuracy improvement. Conversely, the CNN-LSTM-Attention algorithm proves particularly effective in this context, as its attention mechanism can dynamically weight temporal dependencies, capture long-range periodic patterns, and reduce information dilution. By ensuring the forecasting for specific operational states (such as the peak hours for experiments) are primarily informed by historically similar states, it can improve the accuracy significantly. While in Building D, which serves as a conventional office space, occupant behavior is the primary determinant of energy usage, causing behavior-related inputs highly important. Buildings B and C represent intermediate cases, with mixed usage patterns that neither fully resemble pure office spaces nor contain high-energy specialized equipment. Thus, the effectiveness of each strategy lies between those observed in Buildings A and D. Based on these findings, the priority of application for the three strategies are shown in Figure 14.



Legend: these symbols (※ for accuracy, ☒ for efficiency, ☑ for interpretability) represent the priority of different strategies. In different kinds of buildings, the higher number of the symbols indicates the higher priority, while the same number indicates similar priorities, and the absence of a symbol indicates that this strategy is not relevant to this performance metric.

Figure 14 The priority of these optimization strategies for accuracy, efficiency, and interpretability improvement

4.2 The performance level of the proposed energy consumption forecasting

To further evaluate the practical performance of the proposed forecasting method, a comparison with several recently developed forecasting methods is shown in Table 5. Although direct identification of a superior method is infeasible due to differences in building characteristics, weather conditions, and geographical locations, the proposed method may achieve comparable accuracy to recent advanced methods. Furthermore, it is observed that most existing studies focus primarily on algorithmic improvements to enhance accuracy, while few address input feature optimization, and virtually none consider interpretability. This observation aligns with the literature review conducted in this study and indirectly underscores the value of the proposed method, which balances accuracy, efficiency, and interpretability.

Table 5 Advanced hourly energy consumption or load forecasting methods in recent studies

Ref.	The proposed forecasting method	Forecasting performance
This study	Using a hybrid method with introducing occupant behavior probabilities as novel input features, combined physics-informed and data-informed loss function, and CNN-LSTM-Attention algorithm	MAPE: 4.50%–10.31% CV-RMSE: 0.0617–0.1466 R^2: 0.8932–0.9254
Ma et al. [53]	Using a hybrid algorithm with whale optimization, extreme gradient boosting, and double Bi-LSTM attention Q-network (WXGB-DBAQN)	MAPE: 7.331%
Dong et al. [54]	Using a hybrid algorithm with introducing graph attention network and K-Medoids algorithm with DDTW distance into LSTM (GAN-LSTM, K-Medoids-LSTM) for different nodes in buildings	R^2: 0.83, and 0.89 (different conditions)
Jiang et al. [55]	Using a hybrid algorithm with CNN, and attention, and temporal distribution characterization (TDC–CNN–AttLSTM)	CV-RMSE: 0.034–0.145 R^2: 0.931–0.995
Ma et al. [56]	Using a hybrid algorithm with grey wolf optimizer, adaptive Neuro-Fuzzy inference system, and recurrent deep deterministic policy gradient with dynamic action adjustment (GWO-ANFIS-RD3PG).	MAPE: 11.29% R^2: 0.980
Yesilyurt et al. [57]	Using a hybrid algorithm with novel input feature named air conditioning demand (0 or 1) and the deep neural networks	MAPE: 5.095%–6.839% R^2: 0.939–0.971
Lian et al. [58]	Using a hybrid method with the physical model in EnergyPlus to generate database (includes more than 25.14 million data cases) and the LightGBM algorithm to extract feature variables and build the prediction model	MAPE: 12.42% (cooling) MAPE: 7.97% (heating)
Wang et al. [59]	Using a hybrid method with random forest algorithm and considering the time (holiday, etc.) and chilled plant start-up state optimization.	MAPE: 4.05%–8.40% CV-RMSE: 0.0648–0.1719

4.3 Limitation and future directions

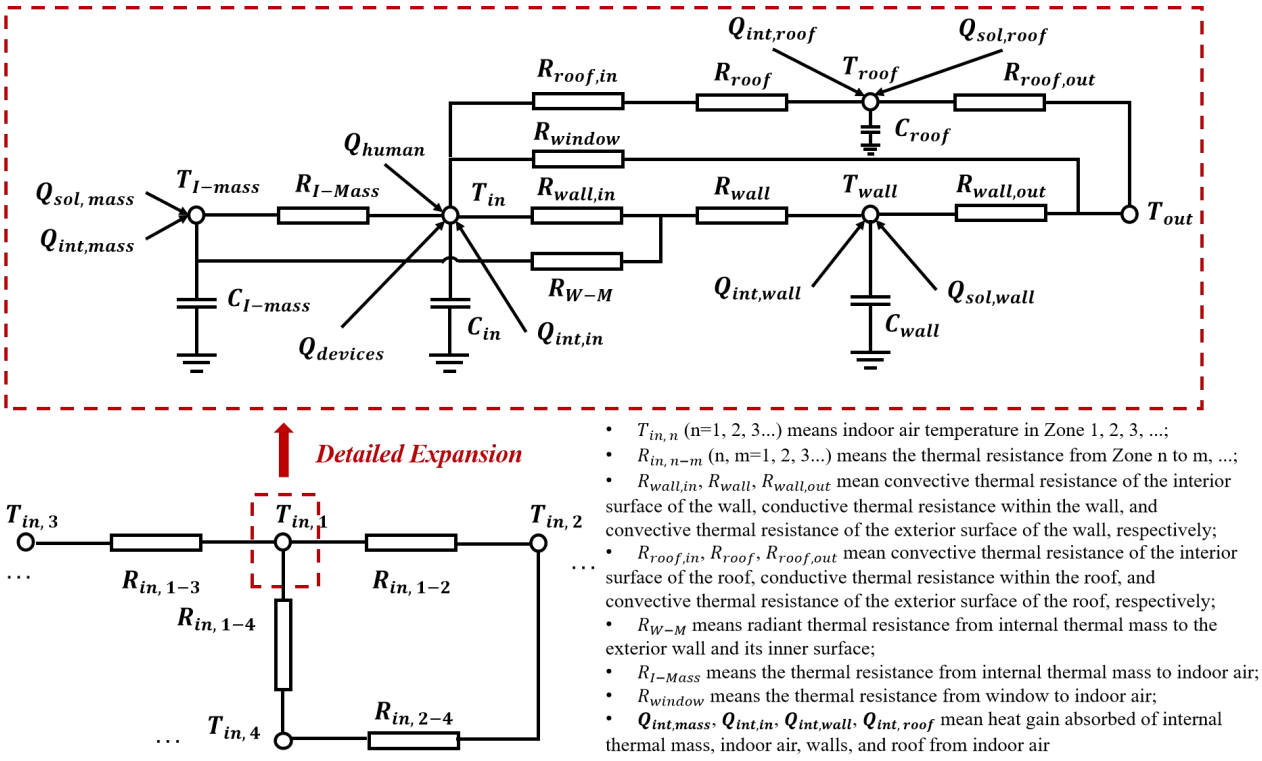
Despite the results shows the effectiveness of the proposed method, some limitations must be acknowledged. First, treating the building as a single entity may oversimplify its dynamics and complex heat exchange processes in the building, including combined convective and radiative heat transfer among the building envelope, internal thermal mass, and indoor air. Second, although the three proposed optimization strategies can improve accuracy, efficiency, and interpretability both theoretically and practically, the generalization of the priority framework requires further validation. In addition, quantitative metrics for precisely evaluating functional and behavioral complexity remain underdeveloped and should be established in future work. Finally, the proposed method relies building intelligence system, including comprehensive energy and environmental monitoring, which may restrict its applicability in buildings with limited sensor infrastructure or in non-public building types.

Future research will primarily focus on the following directions:

(1) Developing a suitable zoning method to partition the building into multiple thermal zones, constructing detailed R-C models (such 8R-4C) for each typical thermal zone, and aggregating these into a full-building R-C

network using appropriate interconnection rules, as shown in Figure 15.

(2) Incorporating a wider range of buildings to validate the proposed framework, and establishing a classification system based on quantitative parameters to confirm and refine the priority rules (Figure 15) to statistically confirm and refine the prioritization spectrum, such as the proportion of some specific equipment energy consumption to total energy consumption, the fluctuation of occupant behaviors.



- $T_{in,n}$ (n=1, 2, 3...) means indoor air temperature in Zone 1, 2, 3, ...;
- $R_{in,n-m}$ (n, m=1, 2, 3...) means the thermal resistance from Zone n to m, ...;
- $R_{wall,in}$, R_{wall}, $R_{wall,out}$ mean convective thermal resistance of the interior surface of the wall, conductive thermal resistance within the wall, and convective thermal resistance of the exterior surface of the wall, respectively;
- $R_{roof,in}$, R_{roof}, $R_{roof,out}$ mean convective thermal resistance of the interior surface of the roof, conductive thermal resistance within the roof, and convective thermal resistance of the exterior surface of the roof, respectively;
- R_{W-M} means radiant thermal resistance from internal thermal mass to the exterior wall and its inner surface;
- R_{I-Mass} means the thermal resistance from internal thermal mass to indoor air;
- R_{window} means the thermal resistance from window to indoor air;
- $Q_{int,mass}$, $Q_{int,in}$, $Q_{int,wall}$, $Q_{int,roof}$ mean heat gain absorbed of internal thermal mass, indoor air, walls, and roof from indoor air

Figure 15 The further detailed thermal R-C models for buildings with more than one zone

5 Conclusions

This study proposes a novel energy consumption forecasting method for office buildings that integrates three key optimization strategies: weighted occupant behavior probabilities as novel input features, physics-informed loss function derived from thermal R-C models, and CNN-LSTM-Attention algorithm. The integrated method effectively balances forecasting accuracy, computational efficiency, and model interpretability. Validation across 48 test cases shows that the proposed method can reduce MAPE by 25.78%, CV-RMSE by 21.31%, and improve R^2 by 7.71% on average, with minimum time consumption also.

The weighted occupant behavior probabilities and the physics-informed loss function enhances both accuracy and interpretability, whereas the CNN-LSTM-Attention algorithm primarily improves efficiency and accuracy. Although all three strategies contribute to predictive accuracy, their relative impacts differ. On average, the use of behavior-related inputs, physics-informed loss, and CNN-LSTM-Attention algorithm account for 40%, 15%, and 45% of the total accuracy improvement, respectively. In terms of efficiency, the CNN-LSTM-Attention is the dominant contributor, reducing computational time by a7343.69s compared to the Transformer-LSTM-Adaboost and by 146.81s

compared to LSTM-SSA on average, which can fully offset the slight increase in time consumption caused by the changes in the inputs and loss function. Regarding interpretability, behavior probabilities and thermodynamic constraints can offer substantial improvements by incorporating domain knowledge and physical realism.

The highest overall performance is achieved when all the three strategies are combined. In cases where simultaneous implementation is not feasible, priority should be given to the CNN-LSTM-Attention for buildings with complex functions and high-energy equipment (such as experimental offices), due to its ability to capture dynamic operational patterns. While for common office buildings, introducing occupant behavior probabilities may be prioritized. Finally, the use of LSTM-SSA is generally not recommended due to its excessive computational requirements during training.

Acknowledgement

The authors are grateful for the Postdoctoral Fellowship Program and China Postdoctoral Science Foundation (Grant No. BX20250055, and No. GZB20240379), the National Natural Science Foundation of China (Grant No. 52478080, No. 52408101, and No. 52508123), 2025 International Exchange Foundation Project of "Co-Creation of Excellence Program" from Dalian University of Technology (Grant: No. DUTIO-ZG-202503), and the Natural Science Foundation Joint Fund of Liaoning Province (Grant: No. 2023-MSBA-023).

Appendix A

(1) CNN-LSTM-Attention

CNN-LSTM-Attention introduces the CNN and Attention modules into LSTM. The CNN module can reduce the dimensionality and improve the generalization ability, while the Attention module can assign weights to each channel of the input features, so the proposed algorithm can improve the prediction accuracy while ensuring the computational efficiency [33, 47–48]. The mathematical process is as follows.

The CNN module has four steps:

Step 1 (Convolution Layer): CNN module begins with a convolutional layer, where filters are applied to perform convolution operations on input data, extracting features at different spatial locations, as shown in equation (A.1).

Step 2 (Activation Layer): A non-linear transformation is applied to the output of the convolutional layer, using the *ReLU* activation function.

Step 3 (Pooling Layer): The pooling layer reduces the spatial dimensions of the feature maps generated by the convolutional layer, thereby decreasing computation complexity and mitigating over-fitting by max-pooling methods, as shown in equation (A.2).

Step 4 (Flattening Layer): The outputs from the convolutional layers and pooling layers are flattened into a one-dimensional vector, which serves as the input to the fully connected layers and is subsequently fed into the following LSTM network.

$$Y_{i,j} = \sum_{m=1}^{M} \sum_{n=1}^{N} W_{m,n} \cdot X_{i+m-1,\, j+n-1} \tag{A.1}$$

$$Y_{i,j} = MaxPooling\left(X_{i\times s+m,\, j\times s+n}\right) = MAX_{m,\, n}\left(X_{i\times s+m,\, j\times s+n}\right) \tag{A.2}$$

Where, Y represents the output of the convolutional layer, obtained by convolving the input features X with the sliding filter W; s denotes the stride; m and n iterate over the pooling window.

The LSTM module has four steps:

Step 1 (Input Gate): The LSTM receives the input sequence and employs the *sigmoid* activation function to determine which information to retain or discard. The input gate (i_t) regulates the incorporation of new input (x_t) into the memory cell (MC) state (C_t), as shown in equation (A.3).

Step 2 (Forget Gate): The forget gate decides which information from the cell state should be forgotten by using the *sigmoid* activation function, which evaluates the extent to the previous cell outputs influences the current MC, as shown in equation (A.5). The forget gate (f_t) effectively preserves relevant portions of the previous cell state ($C_{t\text{-}1}$).

Step 3 (Update Gate): The *sigmoid* activation function is used to identify which values need to be updated. The *tanh* activation function generates new candidate values, and using new information updating new cell state. The update gate computes candidate values ($\widetilde{C}_t$) for the new cell state (C_t), as shown in equations (A.6–A.7).

Step 4 (Output Gate): Based on the updated cell state, the output gate uses the *sigmoid* and *tanh* activation functions to determine the final output information, as shown in equations (A.8–A.9). The output gate (o_t) controls the extent to the cell state (C_t) contributes to the next hidden state (h_t).

$$i_t = sigmoid\,(W_{ix} \cdot x_t + W_{ih} \cdot h_{t-1} + b_i) \tag{A.3}$$

$$sigmoid = (1 + e^{-x})^{-1} \tag{A.4}$$

$$f_t = sigmoid\,(W_{fx} \cdot x_t + W_{fh} \cdot h_{t-1} + b_f) \tag{A.5}$$

$$\widetilde{C}_t = tanh\,(W_{cx} \cdot x_t + W_{ch} \cdot h_{t-1} + b_c) \tag{A.6}$$

$$C_t = \widetilde{C}_t \odot i_t + C_{t-1} \odot f_t \tag{A.7}$$

$$o_t = sigmoid\,(W_{ox} \cdot x_t + W_{oh} \cdot h_{t-1} + b_o) \tag{A.8}$$

Where, t is the current moment while (t-1) is the previous moment; i_t is the input gate; x_t is the input data; C_t is the cell state; f_t is the forget gate; h_t is the input of the hidden state; o_t is the output gate; $\odot$ is the Hadamard product; *tanh* is the tangent activation function; *sigmoid* is the sigmoid activation function.

The Attention mechanism in this study is the squeeze-and-excitation attention mechanism, which has three steps:

Step 1 (Squeeze): While preserving the number of feature channels, features are compressed along the spatial dimensions. Specifically, a global pooling operation is applied to the input features to obtain channel-wise statistics, thereby capturing the importance of each channel. Each two-dimensional feature channel is compressed into a real number U_k by using equation (A.9).

Step 2 (Excitation): A compact neural network is used to capture inter-channel dependencies. This network takes the channel-wise statistics obtained from the squeeze step as input, and produces a set of channel-wise weights, representing the significance of each channel, as shown in equation (A.10). this step involves the feature dimension increase or decrease.

Step 3 (Re-weighting): The original feature maps are re-weighted using the attention scores (that are the weights) from the excitation step, emphasizing informative features and suppressing less useful ones.

$$U_k = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} u_{m,n}}{M \times N},\ m, n = 1, 2, \ldots \tag{A.9}$$

$$S = Sigmoid\,[W_2 \times ReLU\,(W_1 \times U)] \tag{A.10}$$

Where, *Sigmoid* and *ReLU* are used as the two activation functions; W_1 and W_2 denote two fully connected layers; M and N denote the spatial dimensions.

(2) Transformer-LSTM-Adaboost

Transformer-LSTM-Adaboost introduces the Transformer module and Adaboost module into LSTM. The core of Transformer is the self-attention mechanism, with an encoder-decoder structure. Transformer relies on the query, key, and value matrix. By simulating the interaction among these three matrices, it can dynamically capture dependency relationship of different positions in the input sequence. And the Adaboost module can adaptively adjust weights by multiple training for the basic learner (Transformer-LSTM), thereby enhancing overall generalization capability [50–51].

The core of the Transformer is the self-attention mechanism, which relies on the query matrix (***Q***), key matrix (***K***), and value matrix (***V***). By simulating query-key-value interaction pattern, it can dynamically capture dependencies among different positions in the input sequence. The overall framework of Transformer follows an encoder-decoder structure. The multi-head self-attention mechanism can be viewed as the concatenation of outputs from multiple

self-attention mechanisms. The detailed computational procedure is given by equations (A.11–A.16).

$$\boldsymbol{Q} = \boldsymbol{X}_f \ \boldsymbol{W}^Q \tag{A.11}$$

$$\boldsymbol{K} = \boldsymbol{X}_f \ \boldsymbol{W}^K \tag{A.12}$$

$$\boldsymbol{V} = \boldsymbol{X}_f \ \boldsymbol{W}^V \tag{A.13}$$

$$Attention\ (\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = softmax\ (\frac{\boldsymbol{Q}\,\boldsymbol{K}^T}{\sqrt{d_k}})\ \boldsymbol{V} \tag{A.14}$$

$$h_i = Attention\ (\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})_i \tag{A.15}$$

$$MultiHead\ (\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = Concat\ (\sum_{i=1}^{m}\ h_i\ ) \tag{A.16}$$

Where, $\boldsymbol{Q}$ is the query matrix, used to compute similarity with the key matrix ($\boldsymbol{K}$) at other positions, determining which parts of the input to attend to; $\boldsymbol{K}$ represents the key matrix, which matches with the query matrix ($\boldsymbol{Q}$) to generate attention weights; $\boldsymbol{V}$ stands for the value matrix, storing the actual information at each position. After weighting by the attention weights, these values are aggregated to produce the final output. $\boldsymbol{X}_f$ is the input sequence matrix; $\boldsymbol{W}^Q$, $\boldsymbol{W}^K$, and $\boldsymbol{W}^V$ refer to the corresponding weight matrices, which are trainable parameters; d_k denotes the dimension of the matrices $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$; m is the number of attention heads; d_k is the computation results of the ith attention head; the *softmax* function is a normalized exponential function converting a real-valued vector into a probability distribution; the *Concat* function is used to concatenate the outputs of the individual attention heads.

The LSTM process has been listed in equations (A.3–A.8).

In the AdaBoost process, the training dataset is first assigned initial weights. After being processed by a weak classifier, the weight of this basic classifier is computed based on the classification error, and the algorithm proceeds to the next iteration. The learning rate is adjusted in each round of training. Finally, a strong classifier is obtained by summing all basic classifiers. This procedure adaptively adjusts sample weights, allowing the model to progressively focus on samples that are difficult to classify correctly, thereby enhancing overall generalization capability.

Assuming the training set is the $\boldsymbol{D}$=$\{(x_1,y_1), (x_2,y_2), \ldots, (x_N,y_N)\}$, X is the training data, y is the forecast data, $\boldsymbol{W}^D$ is the weight matrix. The initial weight distribution of the training data is set as shown in equation (A.17). The basic learner is trained to obtain a weak classifier $\boldsymbol{G}_t(x_j)$, and the error of this classifier is given by equation (A.18). The weight of this current weak classifier in the final composite classifier is calculated via equation (A.19). The updated sample weights are provided in equations (A.20–A.211). The above steps are repeated until all classifiers are assigned weights. The final strong classifier is obtained by combining the weak classifiers through weighted majority voting, as shown in equation (A.22).

$$w_{t=1}(j) = \frac{1}{N} \tag{A.17}$$

$$\varepsilon_t = P\ (\boldsymbol{G}_t(x_j) \neq y_j) = \sum_{j=1}^{N}\ w_t\,(j)\ II(\boldsymbol{G}_t(x_j) \neq y_j) \tag{A.18}$$

$$\alpha_t = \frac{1}{2}\ ln\ (\frac{1 - \varepsilon_t}{\varepsilon_t}) \tag{A.19}$$

$$w_t\,(j) = \frac{w_t(j)}{z_t(j)}\ exp\ (-\alpha_t\, y_j\ \boldsymbol{G}_t(x_j)) \tag{A.20}$$

$$z_t(j) = \sum_{j=1}^{j=N} w_t(j)\ exp\ (-\alpha_t\ y_j\ \boldsymbol{G}_t(x_j)) \tag{A.21}$$

$$\boldsymbol{G}(x_j) = sign\ (\sum_{t=1}^{t=T} \alpha_t\ \ \boldsymbol{G}_t(x_j)) \tag{A.22}$$

Where, t is the iteration count, t = 1, 2, ..., T; $w_t(j)$ is the sample weight at iteration t; N is the total number of samples; P is the probability; ε_t is the error of the weak classifier ($\boldsymbol{G}_t$) at iteration t; II is the indicator function, taking the value 1 for classification errors and 0 otherwise; α_t is the classifier weight at iteration t; $w_t(j)$ is the weight corresponding to the jth training sample at iteration t; $z_t(j)$ is the normalization factor, ensuring the sum of weights in equation (A.17) equals 1; *sign* is the sign function.

(3) LSTM-SSA

LSTM-SSA introduces the SSA into LSTM, which can solve the optimal solutions of the learning rate, the number of iterations and the number of hidden units in the LSTM process [49].

The LSTM process has been listed in equations (A.3–A.8).

The sparrow search algorithm (SSA) simulates the foraging behavior of sparrows, which includes food searching, group cooperation, and information sharing. The position vectors of all sparrows are shown by equation (A.23), and their corresponding fitness values are shown in equation (A.24). The sparrow population is divided into producers and scroungers. Producers are responsible for seeking food and directing the movement of the population; their positions are updated by using equation (A.25). Scroungers consistently monitor the producers and will move to contest if they discover food. If successful, they acquire the food found by the producers; otherwise, they continue monitoring. The position update rule of scroungers is shown in equation (A.26). Additionally, the algorithm considers sparrows that are aware of danger, accounting for 20% of the population, which act as scouts. The initial positions of these scouts are generated randomly, and their positions are updated using equation (A.28).

$$X_i = \begin{pmatrix} x_{1,1} & \cdots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,d} \end{pmatrix} \tag{A.23}$$

$$F_X = \begin{pmatrix} f([x_{1,1}, x_{1,2}, \ldots, x_{1,d}]) \\ f([x_{2,1}, x_{2,2}, \ldots, x_{2,d}]) \\ \cdots \\ f([x_{n,1}, x_{n,2}, \ldots, x_{n,d}]) \end{pmatrix} \tag{A.24}$$

$$X_{i,j}(t+1) = \begin{cases} X_{i,j}(t) \times\ exp\left(\frac{-i}{\alpha \times iter_{max}}\right), \ if\ R_2 < ST \\ X_{i,j}(t) + Q \times L, \ if\ R_2 \geq ST \end{cases} \tag{A.25}$$

$$X_{i,j}(t+1) = \begin{cases} Q \times\ exp\left(\frac{X_{worst}(t) - X_{i,j}(t)}{i^2}\right), \ if\ i > \frac{n}{2} \\ X_p(t+1) + |X_{i,j}(t) - X_p(t+1)| \times A^+ \times L, \ otherwise \end{cases} \tag{A.26}$$

$$A^+ = A^T\ (A\ A^T\ )^{-1} \tag{A.27}$$

$$X_{i,j}(t+1) = \begin{cases} X_{best}(t) + \beta \times |X_{i,j}(t) - X_{best}(t)|, \ if f_i > f_g \\ X_{i,j}(t) + K \times \left(\frac{|X_{i,j}(t) - X_{worst}(t)|}{(f_i - f_w) + \epsilon} \right), \ if f_i = f_g \end{cases} \quad \text{(A.28)}$$

Where, n is the total number of sparrows; d is the dimension of the variables; each value ($f([x_{n,1}, x_{n,2}, \ldots, x_{n,d}])$ in the matrix F_X is the fitness value of an individual; X_i is the position vector of the ith sparrow; t is the current iteration; j is the dimension (j = 1, 2, ..., d); i is the sparrow index (i = 1, 2, ..., n). $X_{i,j}^t$ is the position of the ith sparrow in the jth dimension at iteration t. $iter_{max}$ is the maximum number of iterations; α is a random number between 0 and 1; R_2 is the alarm value (range: [0, 1]); ST is the safety threshold (range: [0.5, 1.0]); Q is a random number following the normal distribution. When $R_2 <$ ST, it indicates that the producer enters a wide-area search mode with no danger around; while when $R_2 \geq$ ST, it means some sparrows find the danger, and all sparrows need to move to a safe region. X_p is the best position currently occupied by the producer; X_{worst} is the current global worst position; A is a ($1 \times d$) matrix where each element is randomly assigned as either 1 or -1; A^+ is the resulting matrix after applying the operation rule in equation (A.27). β is a step size control parameter (a random value from a normal distribution with mean 0 and variance 1); K is a random number within the range [-1, 1]; f_g and f_w are the current global best and worst fitness values, respectively; $f_i > f_g$ means that the ith sparrow is at the edge of the population; $X_{best}(t)$ is the central position of the population, which is considered safe; whereas $f_i = f_g$ means that a sparrow located near the center has detected danger and needs to move closer to other sparrows.

References

[1] H. Ritchie, P. Rosado, M. Roser. Energy Production and Consumption: Explore data on how energy production and use varies across the world. https://ourworldindata.org/energy-production-consumption, 2024 (accessed 1 September 2025).

[2] N. E. KLEPEIS, W. C. NELSON, W. R. OTT, et al. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. Journal of Exposure Science & Environmental Epidemiology, 2001, 3, 11, 231-252. https://doi.org/10.1038/sj.jea.7500165

[3] E. Belloni, T. Ferrucci, D. Fioriti, et al. Global diffusion and key features of Energy Communities with a main focus on building loads modelling and management: A review. Applied Energy, 2025, 399, 126461. https://doi.org/10.1016/j.apenergy.2025.126461

[4] Professional Committee of Building Energy and Emissions at China Association of Building Energy Efficiency. Research Report on Carbon Emissions in the Field of Urban and Rural Development in China. https://www.cabee.org/site/content/25288.html, 2025 (accessed 1 September 2025).

[5] C. Wang, K. Liu, J. Peng, et al. High-precision energy consumption forecasting for large office building using a signal decomposition-based deep learning approach. Energy, 2025, 314, 133964. https://doi.org/10.1016/j.energy.2024.133964

[6] A. Rahman, V. Srikumar, A. D. Smith. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied Energy, 2018, 212, 372-385. https://doi.org/10.1016/j.apenergy.2017.12.051

[7] W. Dong, H. Sun, Z. Li, et al. Design and optimal scheduling of forecasting-based campus multi-energy complementary energy system. Energy, 2024, 309, 133088. https://doi.org/10.1016/j.energy.2024.133088

[8] F. Rodríguez, E. Maqueda, M. Fernández, et al. A novel methodology for day-ahead buildings energy demand forecasting to provide flexibility services in energy markets. International Journal of Electrical Power & Energy Systems, 2024, 161, 110207. https://doi.org/10.1016/j.ijepes.2024.110207

[9] P. Fan, D. Wang, W. Wang, et al. A novel multi-energy load forecasting method based on building flexibility feature recognition technology and multi-task learning model integrating LSTM. Energy, 2024, 308, 132976. https://doi.org/10.1016/j.energy.2024.132976

[10] A. H. Neto, F. A. S. Fiorelli. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy and Buildings, 2008, 12, 40, 2169-2176. https://doi.org/10.1016/j.enbuild.2008.06.013

[11] D. Yan, T. Hong, B. Dong, et al. IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. Energy and Buildings, 2017, 156, 258-270. https://doi.org/10.1016/j.enbuild.2017.09.084

[12] H. Sha, P. Xu, C. Yan, et al. Development of a key-variable-based parallel HVAC energy predictive model. Building Simulation, 2022, 7, 15, 1193-1208. https://doi.org/10.1007/s12273-021-0885-0

[13] R. Wang, L. Bai, R. Rayhana, et al. Personalized federated learning for buildings energy consumption forecasting. Energy and Buildings, 2024, 323, 114762. https://doi.org/10.1016/j.enbuild.2024.114762

[14] Z. Feng, J. An, M. Han, et al. Office building energy consumption forecast: Adaptive long short term memory networks driven by improved beluga whale optimization algorithm. Journal of Building Engineering, 2024, 91, 109612. https://doi.org/10.1016/j.jobe.2024.109612

[15] Y. Cao, G. Liu, J. Sun, et al. PSO-Stacking improved ensemble model for campus building energy consumption forecasting based on priority feature selection. Journal of Building Engineering, 2023, 72, 106589. https://doi.org/10.1016/j.jobe.2023.106589

[16] C. Zhang, Z. Luo, Y. Rezgui, et al. Enhancing multi-scenario data-driven energy consumption prediction in campus buildings by selecting appropriate inputs and improving algorithms with attention mechanisms. Energy and Buildings, 2024, 311, 114133. https://doi.org/10.1016/j.enbuild.2024.114133

[17] H. Zhang, M. Zhou, Y. Chen, et al. Short-term power load forecasting for industrial buildings based on decomposition reconstruction and TCN-Informer-BiGRU. Energy and Buildings, 2025, 347, 116317. https://doi.org/10.1016/j.enbuild.2025.116317

[18] Y. Yang, Q. Duan, F. Samadi. A systematic review of building energy performance forecasting approaches. Renewable and Sustainable Energy Reviews, 2025, 223, 116061. https://doi.org/10.1016/j.rser.2025.116061

[19] V. H. Rondón-Cordero, L. Montuori, M. Alcázar-Ortega, et al. Advancements in hybrid and ensemble ML models for energy

consumption forecasting: results and challenges of their applications. Renewable and Sustainable Energy Reviews, 2025, 224, 116095. https://doi.org/10.1016/j.rser.2025.116095

[20] M. Zhou, L. Wang, F. Hu, et al. ISSA-LSTM: A new data-driven method of heat load forecasting for building air conditioning. Energy and Buildings, 2024, 321, 114698. https://doi.org/10.1016/j.enbuild.2024.114698

[21] R. Spencer, S. Ranathunga, M. Boulic, et al. Transfer learning on transformers for building energy consumption forecasting—A comparative study. Energy and Buildings, 2025, 336, 115632. https://doi.org/10.1016/j.enbuild.2025.115632

[22] T. Kim, S. Cho. Predicting residential energy consumption using CNN-LSTM neural networks. Energy, 2019, 182, 72-81. https://doi.org/10.1016/j.energy.2019.05.230

[23] F. Qian, W. Gao, Y. Yang, et al. Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption. Energy, 2020, 193, 116724. https://doi.org/10.1016/j.energy.2019.116724

[24] U. Ali, S. Bano, M. H. Shamsi, et al. Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. Energy and Buildings, 2024, 303, 113768. https://doi.org/10.1016/j.enbuild.2023.113768

[25] Y. Song, H. Xie, Z. Zhu, et al. Predicting energy consumption of chiller plant using WOA-BiLSTM hybrid prediction model: A case study for a hospital building. Energy and Buildings, 2023, 300, 113642. https://doi.org/10.1016/j.enbuild.2023.113642

[26] C. Zhang, Z. Luo, Y. Rezgui, et al. Enhancing building energy consumption prediction introducing novel occupant behavior models with sparrow search optimization and attention mechanisms: A case study for forty-five buildings in a university community. Energy, 2024, 294, 130896. https://doi.org/10.1016/j.energy.2024.130896

[27] J. Lee, S. Cho. Forecasting building operation dynamics using a Physics-Informed Spatio-Temporal Graph Neural Network (PISTGNN) ensemble. Energy and Buildings, 2025, 328, 115085. https://doi.org/10.1016/j.enbuild.2024.115085

[28] V. Michalakopoulos, S. Pelekis, G. Kormpakis, et al. Data-driven building energy efficiency prediction using physics-informed neural networks. 2024 IEEE Conference on Technologies for Sustainability (SusTech), Portland, OR, USA, 2024, 84-91. https://doi.org/10.1109/SusTech60925.2024.10553513

[29] B. Li, I. Severinsen, T. Walmsley, et al. Physics-informed neural networks for extrapolating press washer unit operations with heuristic physical knowledge and scarce data. Computers & Chemical Engineering, 2026, 204, 109440. https://doi.org/10.1016/j.compchemeng.2025.109440

[30] J. Lee, S. Cho. Forecasting building operation dynamics using a Physics-Informed Spatio-Temporal Graph Neural Network (PISTGNN) ensemble. Energy and Buildings, 2025, 328, 115085. https://doi.org/10.1016/j.enbuild.2024.115085

[31] S. Semeraro, F. Vecchi, R. Stasi, et al. Physics-Informed Neural Networks for predicting indoor temperature and cooling demand in historic buildings. Journal of Building Engineering, 2025, 115, 114392. https://doi.org/10.1016/j.jobe.2025.114392

[32] Y. Su, H. Cheng, Z. Wang, et al. Impacts of the COVID-19 lockdown on building energy consumption and indoor environment: A case study in Dalian, China. Energy and Buildings, 2022, 263, 112055. https://doi.org/10.1016/j.enbuild.2022.112055

[33] C. Zhang, Z. Luo, Y. Rezgui, et al. Enhancing building energy consumption prediction introducing novel occupant behavior models with sparrow search optimization and attention mechanisms: A case study for forty-five buildings in a university community. Energy, 2024, 294, 130896. https://doi.org/10.1016/j.energy.2024.130896

[34] Y. Chen, H. Wang, Z. Chen. Sensitivity analysis of physical regularization in physics-informed neural networks (PINNs) of building thermal modeling. Building and Environment, 2025, 273, 112693. https://doi.org/10.1016/j.buildenv.2025.112693

[35] C. Chang, G. Ma, J. Zhang, et al. Investigation on the CNN-LSTM-MHA-based model for the heating energy consumption prediction of residential buildings considering active and passive factors. Energy, 2025, 333, 137508. https://doi.org/10.1016/j.energy.2025.137508

[36] H. Li, R. Appel-Meulenbroek, T. A. Arentze, et al. Conceptualizing sustainable occupant behaviour in offices from an interdisciplinary point of view: A systematic review. Energy and Built Environment, 2024, https://doi.org/10.1016/j.enbenv.2024.08.004

[37] Y. Sun, X. Luo, X. Liu, et al. Impact of interactive evolution of heterogeneous occupants' air-conditioning usage behavior on air-conditioning energy consumption in university office buildings. Energy and Buildings, 2025, 331, 115378.

https://doi.org/10.1016/j.enbuild.2025.115378
[38] X. Zhou, Y. Mei, L. Liang, et al. Modeling of occupant energy consumption behavior based on human dynamics theory: A case study of a government office building. Journal of Building Engineering, 2022, 58, 104983. https://doi.org/10.1016/j.jobe.2022.104983
[39] J. Vivian, A. Zarrella, G. Emmi, et al. An evaluation of the suitability of lumped-capacitance models in calculating energy needs and thermal behaviour of buildings. Energy and Buildings, 2017, 150, 447-465. https://doi.org/10.1016/j.enbuild.2017.06.021
[40] J. Yang, H. Wang, L. Cheng, et al. A review of resistance–capacitance thermal network model in urban building energy simulations. Energy and Buildings, 2024, 323, 114765. https://doi.org/10.1016/j.enbuild.2024.114765
[41] Z. Wang, Y. Chen, Y. Li. Development of RC model for thermal dynamic analysis of buildings through model structure simplification. Energy and Buildings, 2019, 195, 51-67. https://doi.org/10.1016/j.enbuild.2019.04.042
[42] G. Gokhale, B. Claessens, C. Develder. Physics informed neural networks for control oriented thermal modeling of buildings. Applied Energy, 2022, 314, 118852. https://doi.org/10.1016/j.apenergy.2022.118852
[43] R. Li, J. Zhang. Heat dissipation models by convection and radiation during the real-time operation of office equipment: A case study of computers. Energy and Buildings, 2023, 300, 113592. https://doi.org/10.1016/j.enbuild.2023.113592
[44] R. Li, J. Zhang. Real-time heat dissipation model of electronic equipment for determining the dynamic cooling demand of office buildings. Journal of Building Engineering, 2022, 45, 103465. https://doi.org/10.1016/j.jobe.2021.103465
[45] R. Li, J. Zhang. A transfer function model of lighting heat dissipation for detecting time-delay effect and dynamic cooling demand during the real-time operation of buildings. Energy and Buildings, 2022, 254, 111538. https://doi.org/10.1016/j.enbuild.2021.111538
[46] U. G. Abdulla, R. Poteau. Identification of parameters for large-scale kinetic models. Journal of Computational Physics, 2021, 429, 110026. https://doi.org/10.1016/j.jcp.2020.110026
[47] J. Zhu, J. Yang, M. Peng, et al. A CNN-LSTM-attention (CLA) hybrid model with PID-based search algorithm (PSA) for heat load forecasting in district heating system. Journal of Building Engineering, 2025, 111, 113325. https://doi.org/10.1016/j.jobe.2025.113325
[48] C. Li, J. Shi. A novel CNN-LSTM-based forecasting model for household electricity load by merging mode decomposition, self-attention and autoencoder. Energy, 2025, 330, 136883. https://doi.org/10.1016/j.energy.2025.136883
[49] M. Zhou, L. Wang, F. Hu, et al. ISSA-LSTM: A new data-driven method of heat load forecasting for building air conditioning. Energy and Buildings, 2024, 321, 114698. https://doi.org/10.1016/j.enbuild.2024.114698
[50] R. Zhang, S. Bu, Y. Zheng, et al. A novel multi-task learning model based on Transformer-LSTM for wind power forecasting. International Journal of Electrical Power & Energy Systems, 2025, 169, 110732. https://doi.org/10.1016/j.ijepes.2025.110732
[51] C. Zhai, X. He, Z. Cao, et al. Photovoltaic power forecasting based on VMD-SSA-Transformer: Multidimensional analysis of dataset length, weather mutation and forecast accuracy. Energy, 2025, 324, 135971. https://doi.org/10.1016/j.energy.2025.135971
[52] Ministry of Construction of the People's Republic of China. Design Standard for Energy Efficiency of Public Buildings (GB 50189-2015). https://www.chinesestandard.net/PDF/English.aspx/GB50189-2015?English_GB%2050189-2015, 2015 (accessed 1 September 2025).
[53] X. Ma, J. Wang, J. Huang, et al. Energy consumption and carbon emission modeling and forecasting study with novel deep learning methods. Expert Systems with Applications, 2025, 290, 128314. https://doi.org/10.1016/j.eswa.2025.128314
[54] X. Dong, Y. Luo, S. Yuan, et al. Building electricity load forecasting based on spatiotemporal correlation and electricity consumption behavior information. Applied Energy, 2025, 377, 124580. https://doi.org/10.1016/j.apenergy.2024.124580
[55] B. Jiang, Y. Li, Y. Rezgui, et al. Multi-source domain generalization deep neural network model for predicting energy consumption in multiple office buildings. Energy, 2024, 299, 131467. https://doi.org/10.1016/j.energy.2024.131467
[56] X. Ma, J. Fan, J. Wang, et al. GWO-ANFIS-RD3PG: A reinforcement learning approach with dynamic adjustment and dual replay mechanism for building energy forecasting. Journal of Building Engineering, 2024, 97, 110726. https://doi.org/10.1016/j.jobe.2024.110726
[57] H. Yesilyurt, Y. Dokuz, A. S. Dokuz. Data-driven energy consumption prediction of a university office building using machine learning algorithms. Energy, 2024, 310, 133242. https://doi.org/10.1016/j.energy.2024.133242

[58] H. Lian, Y. Ji, M. Niu, et al. A hybrid load prediction method of office buildings based on physical simulation database and LightGBM algorithm. Applied Energy, 2025, 377, 124620. https://doi.org/10.1016/j.apenergy.2024.124620
[59] J. Wang, X. Li. Optimization of machine learning load forecasting models based on building operation characteristics: a case study for office buildings. Energy and Buildings, 2025, 346, 116233. https://doi.org/10.1016/j.enbuild.2025.116233