buildings

MDPI

Article

The Cost-Optimal Control of Building Air Conditioner Loads Based on Machine Learning: A Case Study of an Office Building in Nanjing

Zhenwei Guo [1,2], Xinyu Wang [2,*], Yao Wang [3,4], Fenglei Zhu [4], Haizhu Zhou [5], Miao Zhang [5] and Yuxiang Wang [2]

1 State Key Laboratory of Building Safety and Built Environment, Beijing 100013, China; sunshinetone@126.com
2 Chinese Society for Urban Studies, Beijing 100835, China; 18649350410@163.com
3 School of Architecture, Tsinghua University, Beijing 100190, China; wangyao22@mails.tsinghua.edu.cn
4 Beijing Glory PKPM Technology Co., Ltd., Beijing 100013, China; zhufenglei@pkpm.com.cn
5 China Academy of Building Research, Beijing 100013, China; zhznhm@163.com (H.Z.); zhangmiao117@126.com (M.Z.)
* Correspondence: wangxinyu356533529@126.com

Abstract: Building envelopes and indoor environments exhibit thermal inertia, forming a virtual energy storage system in conjunction with the building air conditioner (AC) system. This system represents a current demand response resource for building electricity use. Thus, this study centers on the CatBoost algorithm within machine learning (ML) technology, utilizing the LASSO regression model for feature selection and applying the Optuna framework for hyperparameter optimization (HPO) to develop a cost-optimal control method for minimizing building AC loads. This method addresses the challenges associated with traditional load forecasting and control methods, which are often impacted by environmental temperature, building parameters, and user behavior uncertainties. These methods struggle to accurately capture the complex dynamics and nonlinear relationships of AC operations, making it difficult to devise AC operation and virtual energy storage scheduling strategies effectively. The proposed method was applied and validated using a case study of an office building in Nanjing, China. The prediction results showed coefficient of variation in root mean square error (CV-RMSE) values of 6.4% and 2.2%. Compared with the original operating conditions, the indoor temperature remained within a comfortable range, the AC load was reduced by 5.25%, and the operating energy costs were reduced by 24.94%. These results demonstrate that the proposed method offers improved computational efficiency, enhanced model performance, and economic benefits.

Keywords: building air conditioner load; machine learning; load control; energy costs; indoor temperature

check for updates

Citation: Guo, Z.; Wang, X.; Wang, Y.; Zhu, F.; Zhou, H.; Zhang, M.; Wang, Y. The Cost-Optimal Control of Building Air Conditioner Loads Based on Machine Learning: A Case Study of an Office Building in Nanjing. *Buildings* **2024**, *14*, 3040. https://doi.org/10.3390/buildings14103040

Academic Editors: Elisa Belloni and Gabriele Maria Lozito

Received: 15 August 2024
Revised: 10 September 2024
Accepted: 18 September 2024
Published: 24 September 2024

1. Introduction

1.1. Background and Motivation

With the increasing electrification of buildings, the demand for electricity in building operations has progressively increased [1], exhibiting significant fluctuations. These peak and off-peak electricity usage characteristics impose substantial pressure on the grid and power supply [2]. Demand response (DR) is recognized as an effective demand-side energy management and optimization technique capable of guiding users in actively adjusting their energy usage patterns through monetary incentives [3]. This strategy can be used to reduce electricity demand during peak periods and emergencies [4,5], thereby decreasing the variability in electricity demand [6].

When the grid requires buildings to adjust their load, buildings can achieve load reductions or shifts through temperature regulation, by shutting down non-essential equipment, by using energy storage systems, and by implementing building energy management system (BEMS). Among these methods, energy storage devices, particularly electrochemical storage, are the most direct measures. They not only compensate for the volatility of the power system but also enhance the safety of building electricity use [7–9]. However, due to issues such as initial investment costs, maintenance, and lifecycle concerns, this approach is challenging to implement on a large scale. In recent years, virtual energy storage (VES) in buildings has gained considerable attention. A VES avoids the aforementioned initial investment costs and space limitations, making it a viable alternative to battery storage [10], and is considered one of the most valuable DR resources [3,11]. Building a VES originates from the thermal buffering capacity of building envelopes and indoor environments. When heating, ventilation, and air conditioning (HVAC) systems cool or heat indoor spaces, the stored cold or heat is retained within the building envelope, indoor equipment, and air. Thus, reducing the cooling or heating power of an HVAC system decreases the electricity load without immediately changing the indoor temperature, but rather by releasing the stored cold or heat, resulting in a gradual temperature increase or decrease. This process is analogous to charging and discharging energy storage devices [12–14].

A building air conditioning (AC) system serves both as an energy provider for a VES and as a key component in the DR system of buildings that maintains the requirements of comfortable temperatures. Kaliyamoorthy et al. [15] evaluated the charging and discharging patterns of loads from the perspectives of power and energy, validating the feasibility of using building AC combined with a VES for DR. Braun et al. [16] demonstrated that pre-cooling buildings overnight reduced the cumulative load by 23% compared with buildings that were not pre-cooled, highlighting the load adjustment potential of such DR systems. Lu et al. [13] explored methods for achieving hourly load-balancing services using integrated AC systems, indicating that further optimization is possible by combining building AC systems with VES.

Building AC systems exhibit diverse dynamic characteristics [17], gradual and continuous response traits [18], and are susceptible to uncertainties such as environmental temperature, building parameters, and user behavior. Capturing these complex dynamics and nonlinear relationships to develop AC operation strategies for VES scheduling is therefore complex and challenging [19]. In this study, we propose using machine learning (ML) algorithms to minimize AC load costs while meeting the requirement of comfortable temperatures as dual objectives, aiming to enhance predictive analysis accuracy and to develop automated adjustment strategies.

1.2. Literature Review

Three main prediction methods exist for AC load regulation: simulation-based methods, mechanism-based methods, and data-driven methods. AC load and indoor environmental temperature are influenced by factors such as outdoor weather, building envelope insulation, and indoor occupancy and behavior, which introduce significant randomness into real-world operations. Thus, building VES adjustment models using simulation software is highly challenging [20], and their prediction results can be significantly inaccurate when precise data are lacking [21]. Consequently, simulation-based methods are often used for steady-state and hypothetical load adjustment analyses. Mechanism-based methods involve constructing mathematical models to describe physical processes. For example, Hu et al. [22] developed a thermal model for a simplified room to support a model predictive controller for load adjustment and validated the method's cost-effectiveness through simulations. Li et al. [23] built a linear state-space model to capture indoor temperature dynamics and used a quasi-steady-state approach to represent the AC system for control purposes. Similarly, Hughes et al. [24] employed a simple first-order linear dynamic model to analyze the flexibility of residential AC systems. The authors in [25] conceptualized rooms equipped with variable frequency air conditioners as "thermal battery models" and incorporated them into load adjustment scheduling. Overall, both methods rely on certain assumptions and simplifications, which can limit their ability to capture the true dynamics of building energy systems [26], thereby affecting the accuracy of the models and strategies, and the effectiveness of any load adjustments.

Research indicates that data-driven modeling approaches can enhance the accuracy of results [27,28]. Data-driven methods utilize ML algorithms from the field of artificial intelligence to perform black-box modeling using extensive historical data. These methods not only correctly interpret but also learn from external data and adapt accordingly to achieve specific goals and tasks [29,30]. ML has broad applications in the building sector, including energy system forecasting [31,32], operations [33,34], control [35,36], and diagnostics [37–39]. Although ML methods have been applied in building energy systems to some extent, research interest in their use for load adjustment has only significantly increased in recent years [20]. ML technologies enable the real-time analysis of grid states, demand prediction, and energy optimization, thereby achieving peak shaving, mismatch mitigation, enhanced energy efficiency, and reduced operational costs [40,41]. For instance, Zhou et al. [42,43] used ML algorithms for short-term load forecasting in residential DR scenarios, showing good predictive accuracy. Behl et al. [44] provided a model-based control method using regression tree algorithms (mbCRTs) for the comprehensive closed-loop control of load adjustment strategies in large commercial buildings. Ben et al. [45] employed a General Regression Neural Network (GRNN) to predict cooling loads, achieving a good fit between the predicted and actual building load curves.

1.3. Contributions

This study represents the first application of the CatBoost algorithm in the prediction of building AC loads considering the role of building VESs. Historical data generated by real-time building energy monitoring systems were used as input parameters, rather than assumed preset simulation parameters, enhancing the practical applicability of this research. During the algorithm model optimization process, we employed the LASSO regression model for feature selection and utilized the Optuna framework for automated hyperparameter optimization (HPO). A comparison of the model's performance before and after optimization demonstrated a significant improvement in accuracy with the adjusted algorithm model.

Building on these results, we also developed a multi-objective prediction and decision-making method for AC load optimization, constrained by minimizing costs and ensuring indoor comfort temperatures. This approach makes the automatic load adjustment of an AC system under DR requirements feasible, offering significant value by reducing energy consumption and cost expenditures, as well as lowering the carbon emissions associated with building operations. The research framework is illustrated in Figure 1.

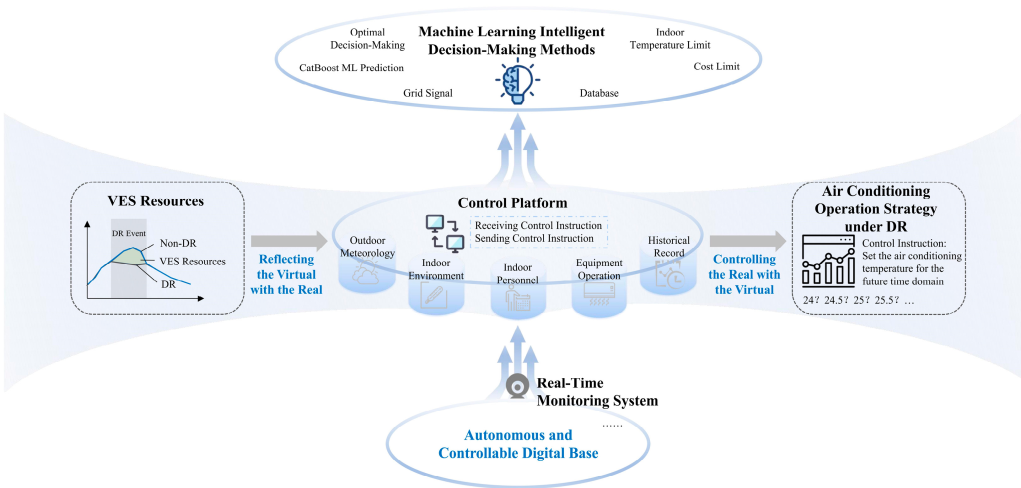


Figure 1. Research framework.

2. Materials and Methods

Boosting is an ML technique used to reduce bias in supervised learning and is known for its outstanding performance [3]. The Boosting family includes algorithms such as CatBoost, XGBoost, and AdaBoost. Bian et al. [46] utilized the CatBoost and AdaBoost algorithms to predict an HVAC system's power consumption, demonstrating their excellent predictive performance. Kaligambe et al. [47] designed and trained several XGBoost models to estimate the indoor temperature, relative humidity, and carbon dioxide concentration, confirming the accuracy of these models and their suitability for optimal HVAC system control in smart buildings. Wang et al. [3] introduced an HVAC system predictive control strategy based on the XGBoost algorithm to enhance daytime DR. CatBoost is one of the principal algorithms in the Boosting family, proficient in handling both numerical and categorical data. It is noted for its robustness, versatility, platform applicability, and prediction speed [48]. In this study, we leveraged the CatBoost algorithm for model development and training, followed by a demand response decision analysis based on the prediction results. The basic process involved first developing and training a machine learning-based model for predicting the AC load and indoor temperature. Feature selection was conducted using a LASSO regression model, and HPO was performed following the Optuna framework to optimize computational efficiency and model performance, further enhancing the model's performance and generalization capabilities. Subsequently, simulations were conducted based on real-time measurements from building equipment and environmental data. Using the prediction results, decisions were made to set the temperature for cost optimization within a comfortable indoor temperature range. Finally, temperature setting commands were issued to the AC's temperature controllers via the building equipment management system, adjusting the building environment to schedule building a VES and to achieve user-side DR control. The principle behind this process is illustrated in Figure 2.

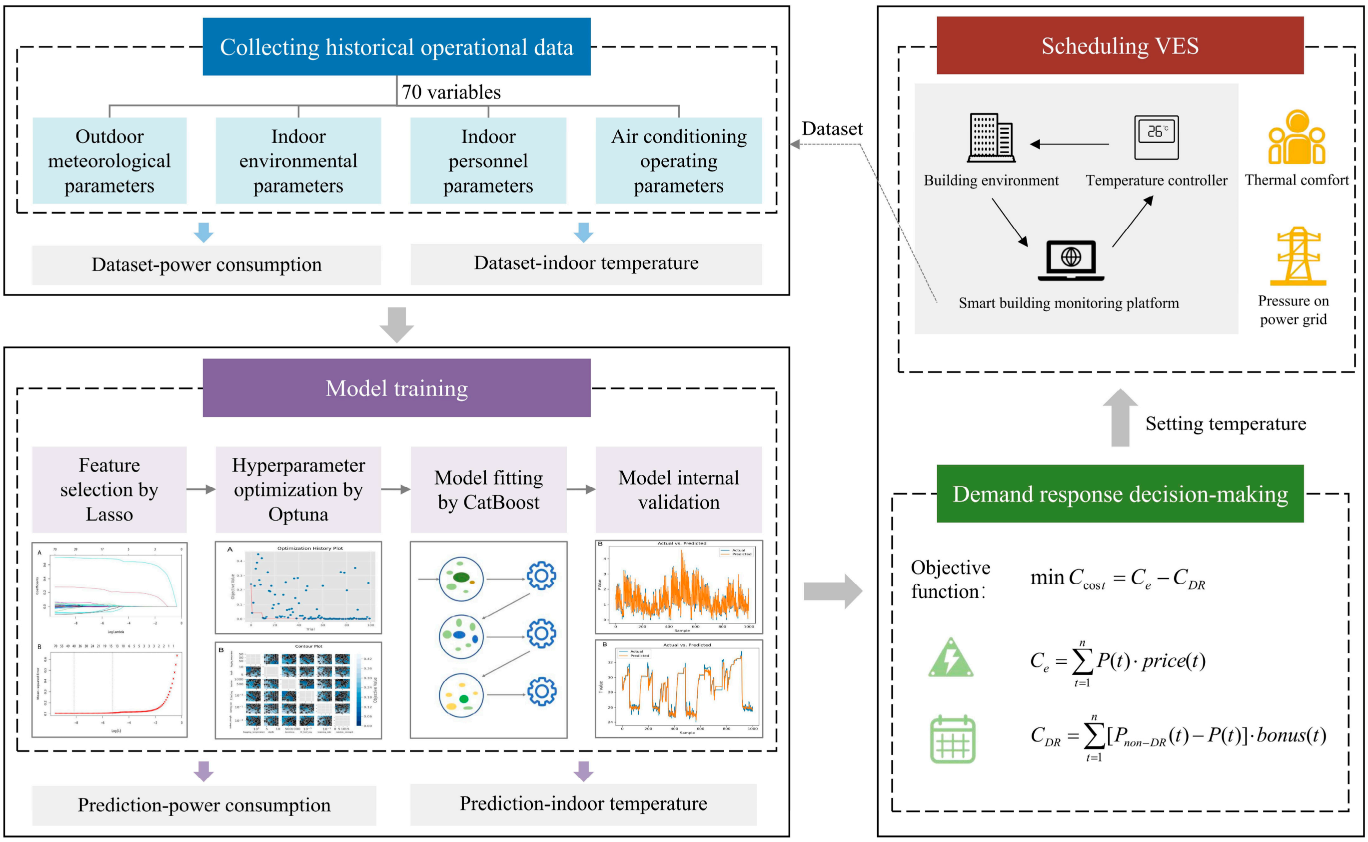


Figure 2. Principle of cost minimization for building AC load regulation based on ML.

2.1. Data Acquisition and Selection of Input Variables

Then, the indoor environmental parameters, equipment operating parameters, and occupancy information were monitored. These parameters were transmitted in real time to an intelligent monitoring platform via a wireless network. Sensors transmitted the data every 10 min, which were then stored in a responsive database. Data extracted from 3 June 2023 to 27 August 2023 were utilized for the analysis.

The relationship between the inputs and outputs in this study was established by predicting the current AC power consumption and indoor temperature using historical monitoring data. The impact of the outdoor environmental parameters on AC power consumption and the indoor temperature was significant and included factors such as the outdoor temperature and solar radiation. These factors were thus incorporated as input variables in the predictive model. Additionally, the indoor environment conditions, number of occupants, and time information also correlated with AC power consumption and the indoor temperature and were considered. The historical information reflected certain thermal inertia trends and was therefore relevant. Consequently, historical data from the past 10, 20, 30, 40, and 50 min, including outdoor temperature, outdoor humidity, eastward wind speed, northward wind speed, net solar radiation intensity, total solar radiation intensity, indoor humidity, number of occupants, set temperature, HVAC power consumption, indoor temperature, and time information were used as input variables for the model.

2.2. Feature Selection

Feature selection is a crucial step in ML, essential for building an efficient and accurate prediction model. The primary goal of feature selection is to choose the most informative features from the raw data to simplify the model, enhance its performance, and reduce the risk of overfitting. In this study, we addressed these issues by employing the LASSO regression model for optimal feature selection. LASSO regression, which stands for "Least Absolute Shrinkage and Selection Operator", is a widely used method for feature selection and regularization in regression analyses. The fundamental principle behind LASSO is the introduction of an L1 regularization term into the ordinary least squares method, thereby minimizing the objective function to achieve feature selection and coefficient sparsity. The advantage of LASSO regression lies in its ability to automatically perform feature selection, resulting in a more streamlined and interpretable model. Currently, LASSO regression is considered one of the most commonly used feature selection methods [49]. In this study, 70 potential input features were screened, and LASSO regression was used to select those with significant correlations to AC power consumption and indoor temperatures.

2.3. Hyperparameter Optimization

Like most ML algorithms, the CatBoost model has several hyperparameters, and different combinations of these hyperparameters can significantly impact the model's performance. By fine-tuning hyperparameters such as learning rate, tree depth, and number of trees, the predictive accuracy of the model can be improved. Since each dataset is unique, specific hyperparameter settings are required to achieve an optimal performance for a given problem. Additionally, appropriate hyperparameter settings help balance model complexity, preventing overfitting and underfitting. In this study, we utilized a widely adopted HPO framework: Optuna. Optuna provides an efficient method for automatically searching for optimal hyperparameters. It employs an algorithm known as the "Tree-structured Parzen Estimator" (TPE) to select the hyperparameters most likely to enhance the model's performance [50]. Compared with traditional grid search and random search methods, this approach is typically more efficient and precise. Furthermore, Optuna automates the hyperparameter search process, including defining the search space, running optimization trials, and recording the results. This significantly reduces the manual effort required for HPO [51].

2.4. CatBoost Model Training

Here, we developed two key predictive models based on the CatBoost algorithm—the AC power consumption prediction model and the indoor temperature prediction model—which provide crucial references for building AC load regulation.

Firstly, actual AC operation data with a temporal sequence were collected, with the first 80% used as the training set and the remaining 20% reserved for internal validation. This deliberate partitioning preserved the inherent temporal characteristics of the time series data. Next, LASSO regression was employed for feature selection, and the Optuna framework was used for HPO. After HPO, the model was fitted using the optimal hyperparameter combination and subsequently validated. To assess the performance of the predictive models, four key metrics were used: two scale-dependent metrics—mean absolute error (MAE) and root mean square error (RMSE)—and two scale-independent metrics—R-squared (R^2) and the coefficient of variation of RMSE (CV-RMSE). The scale-dependent metrics provide an intuitive understanding of prediction errors in specific scenarios, while the scale-independent metrics facilitate comparisons with similar studies.

2.5. Load Adjustment Decision

Building VES resources allows for the adjustment of or shift in the required loads by modifying the AC set temperature without violating the constraints of comfortable indoor temperatures. In this study, we follow the same principle for load adjustment decisions. Additionally, under the DR requirements of the grid, cost-oriented load adjustment has significant potential for widespread adoption. Therefore, the primary decision objective for setting the temperature in this study is to minimize AC operating costs while maintaining the constraints of comfortable indoor temperatures. The specific process for deciding on the set temperature is illustrated in Figure 3. First, the method involves manually dividing the AC set temperature range and then defining the adjustment range from 23.5 °C to 26.5 °C with a granularity of 0.5 °C, which serves as the candidate set temperatures [3]. Simultaneously, the comfortable indoor temperatures were constrained to between 23 °C and 27 °C. This range was chosen based on historical operating temperatures to ensure that the indoor environment remains within an acceptable range for occupants. Given the constraints of economic costs, the cost formula is as follows:

$$\min C = C_e - C_{DR} \tag{1}$$

where C is the total cost; C_e is the cost of electricity, as shown in Equation (2); and C_{DR} is the cost return during the DR period, as shown in Equation (3).

$$C_e = \sum_{t=1}^{n} P(t) \cdot price(t) \tag{2}$$

where *price* is the energy consumption electricity price, which is determined by local electricity management regulations.

$$C_{DR} = \sum_{t=1}^{n} \left[P_{non-DR}(t) - P(t)\right] \cdot bonus(t) \tag{3}$$

where $P_{non\text{-}DR}$ represents power consumption under non-DR conditions, which can be determined based on historical operational patterns, and *bonus* is the price offered for participating in DR, as determined by local electricity regulations.

The values for air conditioning power consumption and indoor temperature during the decision-making process were derived from the optimized, high-performance CatBoost models, as given in the following equations:

$$P_t = CatBoost\,1(\mathbf{X}_{P,t}) \tag{4}$$

$$T_{in,t} = CatBoost\ 2(\mathbf{X}_{Tin,t}) \quad (5)$$

where P_t represents the predicted AC power consumption at time t, $\boldsymbol{X}_{P,t}$ denotes the input variables for predicting AC power consumption at time t, $T_{in,t}$ is the predicted indoor temperature at time t, and $\boldsymbol{X}_{Tin,t}$ represents the input variables for predicting the indoor temperature at time t.

The final set temperature was determined based on cost minimization. Among the candidate set temperatures, the one with the lowest cost that maintained the indoor temperature within the specified comfort range was selected as the final set temperature.

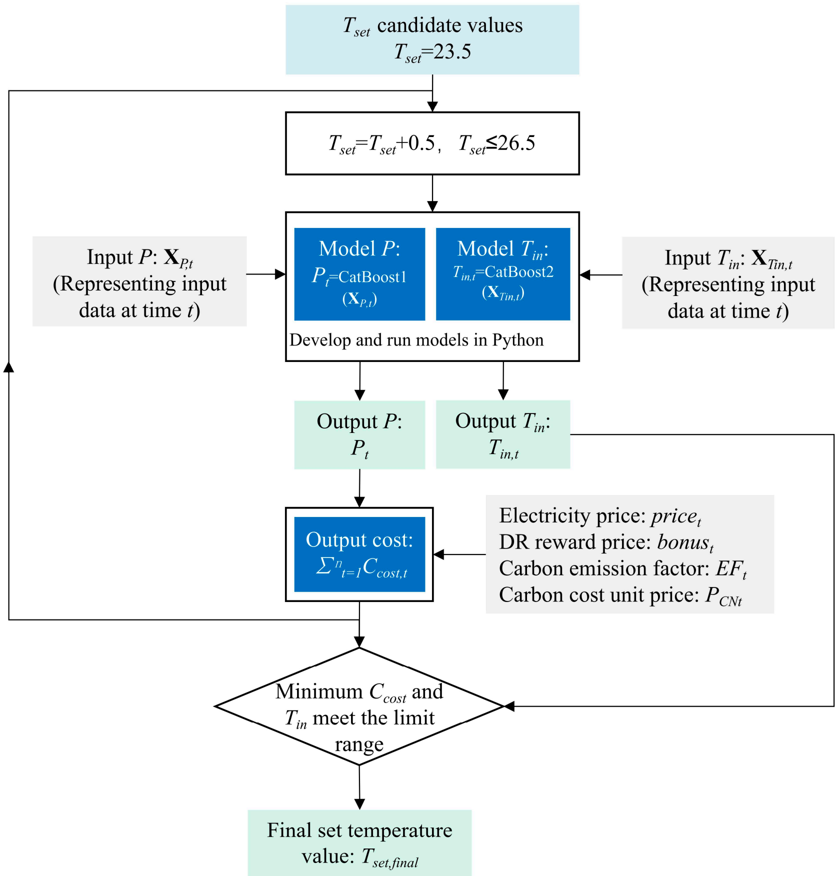


Figure 3. Framework for deciding on the set temperatures.

3. Results and Analysis

3.1. Prediction Model

3.1.1. Data Source

The data used in this study were collected from an office building in Nanjing, China. The building's equipment management system is illustrated in Figure 4. Approximately 12,500 data entries were collected. After excluding data points related to the non-operational periods of the AC system and removing extreme outliers, the dataset was refined to about 5000 continuous entries. The initial feature set included 12 variable categories (X1–X12), comprising current time (H), indoor humidity (D_{in}), indoor occupancy (M_{in}), outdoor temperature (T_{out}), outdoor humidity (D_{out}), radial wind speed (V_{radial}), zonal wind speed (V_{zonal}), net solar radiation intensity ($R_{net\text{-}solar}$), total solar radiation intensity ($R_{total\text{-}solar}$), set temperature (T_{set}), AC power consumption (P), and indoor temperature (T_{in}). Table 1 presents statistical descriptions of these variables. The data were systematically updated every 10 min, resulting in a comprehensive dataset with 70 variables (X1–X12 for the past 10 to 50 min and X1–X10 for the current time). The distributions of the 12 variables are shown in Figure 5.

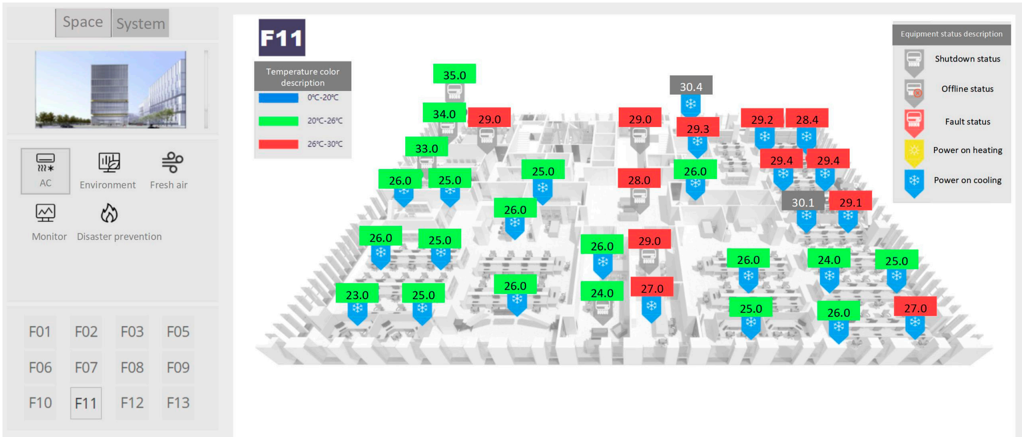


Figure 4. Building equipment management system.

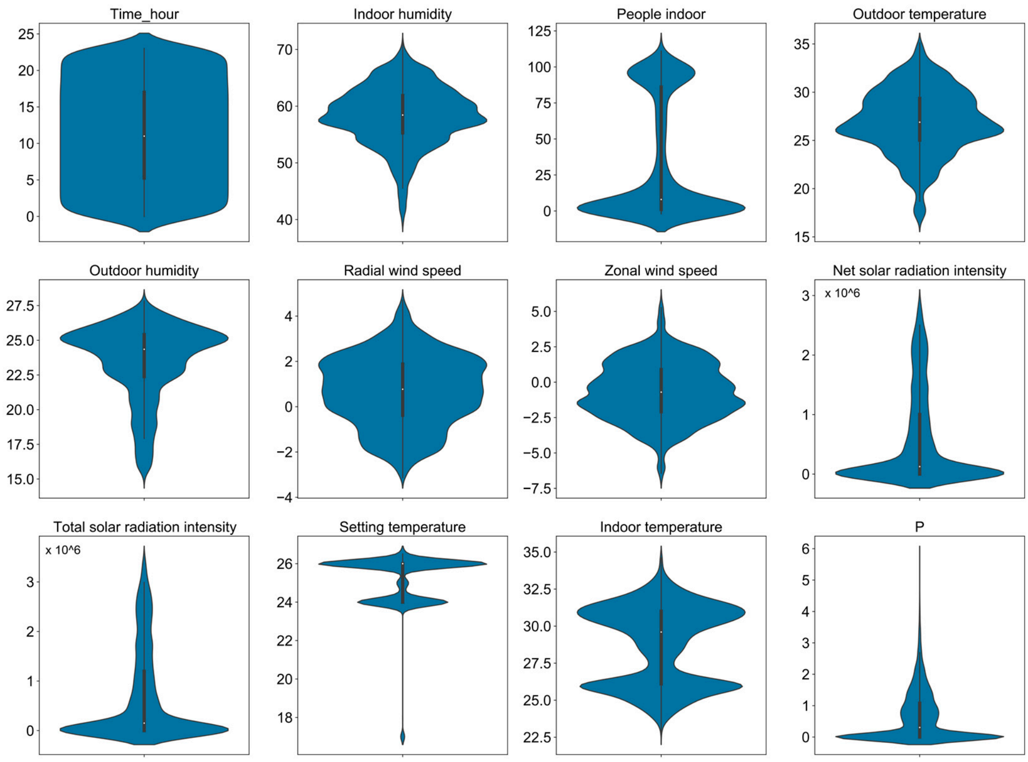


Figure 5. Distribution of 12 key variables in the dataset through violin plots.

Table 1. Statistical descriptions of the data.

Variable	Unit	Description	Range	Average	Median
X1: H	-	current hours	[7, 21]	14.49 ± 3.85	15
X2: D_{in}	%	indoor relative humidity	[41.29, 71.29]	57.28 ± 4.39	57.57
X3: M_{in}	people	indoor occupancy	[0, 111]	66.23 ± 36.40	88
X4: T_{out}	°C	outdoor temperature	[18.13, 35.1]	28.98 ± 2.59	29.16
X5: D_{out}	%	outdoor relative humidity	[28.18, 99.99]	75.57	79.63
X6: V_{radial}	m/s	radial wind speed	[−3.1, 99.99]	0.71 ± 1.67	0.85
X7: V_{zonal}	m/s	zonal wind speed	[−6.03, 5.92]	−0.45 ± 2.19	−0.52
X8: $R_{net\text{-}solar}$	J/m^2	net solar radiation intensity	[0, 2,862,042]	1,023,792 ± 808,399	904,292
X9: $R_{total\text{-}solar}$	J/m^2	total solar radiation intensity	[−4, 3,439,454]	1,223,246 ± 965,701	1,077,806
X10: T_{set}	°C	set temperature	[17, 26.5]	25.38 ± 1.02	26
X11: P	kWh	power consumption	[0.00, 3.49]	1.19 ± 0.70	1.07
X12: T_{in}	°C	indoor temperature	[22.90, 34.50]	27.33 ± 2.39	26.10

3.1.2. The Result of Feature Selection

LASSO regression was employed to handle any redundant variables and to enhance the model's practicality. The variable selection process in LASSO regression is depicted in Figure 6. As the penalty coefficient (Lambda) increased, the coefficients of certain variables approached zero, indicating their limited impact on the model. After applying LASSO regression, 14 variables were retained for the final model (Model 1: AC power consumption prediction model). These variables were M_{in}, D_{out}, T_{set}, $T_{out\_20\ min}$, $D_{out\_20\ min}$, $P_{\_20\ min}$, $P_{\_30\ min}$, $T_{in\_30\ min}$, $D_{in\_40\ min}$, $P_{\_40\ min}$, $M_{in\_50\ min}$, $T_{out\_50\ min}$, $T_{set\_50\ min}$, and $P_{\_50\ min}$. To explore the importance of the model input features, Shapley additive explanation (SHAP) values were used to rank the significance of the variables used in the model's construction (Figure 7). For Model 2 (the indoor temperature prediction model), nine variables were included: H, M_{in}, $D_{in\_10\ min}$, $P_{\_10\ min}$, $D_{in\_20\ min}$, $P_{\_20\ min}$, $T_{in\_20\ min}$, $T_{set\_40\ min}$, and $T_{in\_50\ min}$.

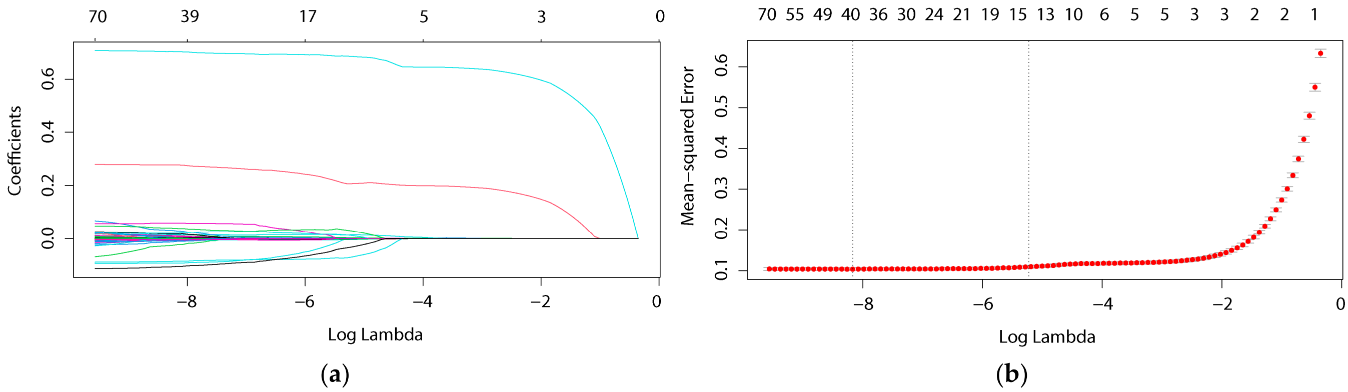


Figure 6. Feature selection via LASSO analysis. (**a**) Plots for LASSO analysis coefficients; (**b**) cross validation plot for the penalty term.

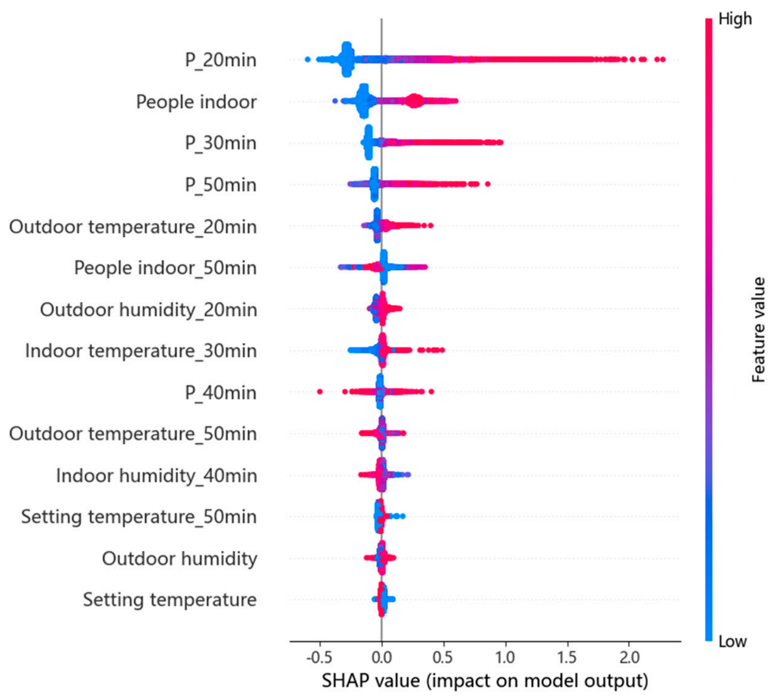


Figure 7. Feature importance via SHAP. Dots are colored based on feature values for each sample, accumulating vertically to show density. The *y*-axis color ranges from blue to red, indicating feature values (blue for low, red for high). The *x*-axis represents the impact on the model output, with positive impacts on the right and negative impacts on the left.

3.1.3. The Result of Hyperparameter Optimization

HPO was conducted for Model 1 to enhance its prediction performance, as shown in Figure 8. After 100 trials, the CatBoost model with the lowest RMSE was identified. The final hyperparameter settings obtained from the search are detailed in Table 2. To validate the effectiveness of the optimization process (Figure 9), a comparison was made between the model's performance before and after HPO, using the best combination of the obtained parameters. The results indicate that the CatBoost model showed improved prediction capabilities after HPO. The optimization process for Model 2 was similar, demonstrating a significant performance enhancement post-optimization, and the results are shown in Figure 10.

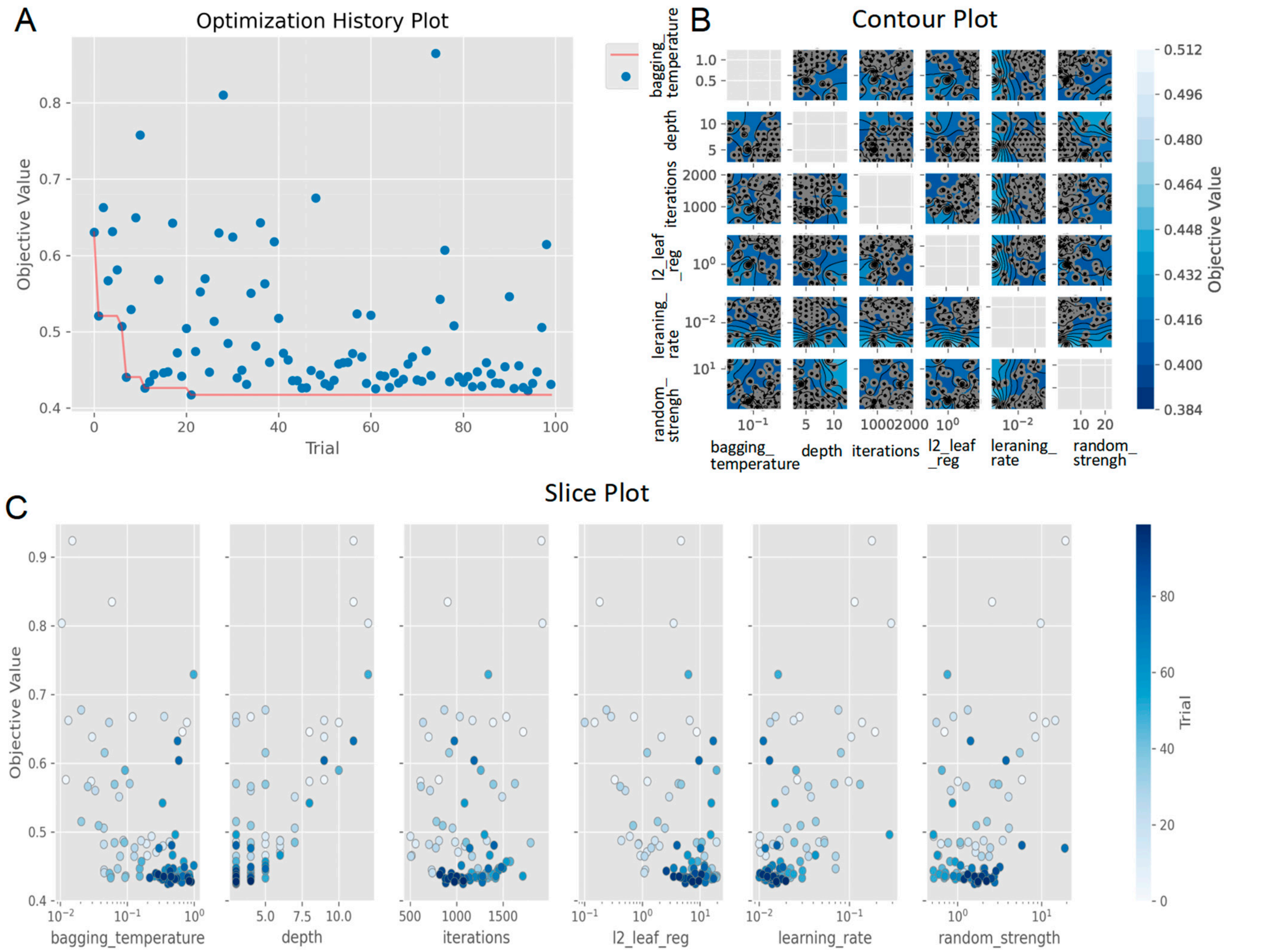


Figure 8. HPO process of Model 1. (**A**) Optimization history plot: each blue point represents the result of a trial, and the dark orange line represents the best objective function value; (**B**) Contour plot: the horizontal and vertical axes represent two different hyperparameter values, while the contour lines indicate the objective function value. Different colored regions represent various ranges of the objective function value, with darker colors indicating lower values. (**C**) Slice plot for hyperparameters: the shade of color represents the influence of every parameter on the model's performance.

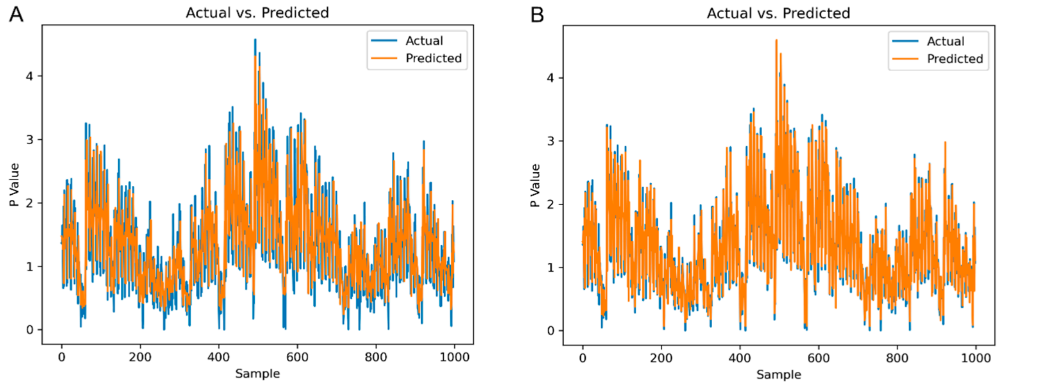


Figure 9. Comparisons of prediction performance for Model 1 before (**A**) and after (**B**) hyperparameter optimization. The blue line indicates the actual value; the yellow line indicates the predictive value.

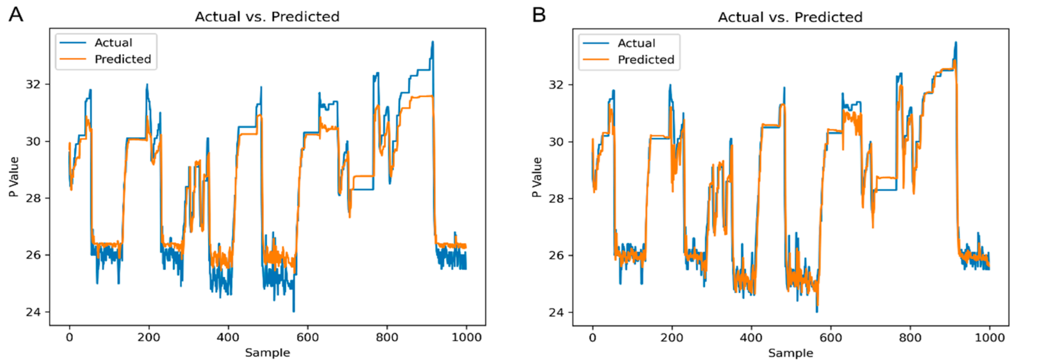


Figure 10. Comparisons of prediction performance for Model 2 before (**A**) and after (**B**) hyperparameter optimization. The blue line indicates the actual value; the yellow line indicates the predictive value.

Table 2. Search domain and values set for the hyperparameters.

Hyperparameter	Search Domain	Set Value
iterations	[100, 1000]	880
depth	[4, 10]	10
learning_rate	[0.01, 0.3]	0.2949
random_strength	[1×10^{-9}, 10]	0.0092
bagging_temperature	[0.01, 100]	0.0340
l2_leaf_reg	[1×10^{-8}, 10]	1.0519×10^{-6}

3.1.4. Model Validation

Based on the above methods, the optimal CatBoost models were obtained. The final 20% of the dataset was used to fit the models, designated as the internal validation set. Detailed results of the model validation are presented in Tables 3 and 4. The results show that Model 1 achieved an MAE of 0.06, an RMSE of 0.08, an R-squared value of 0.98, and a CV-RMSE of 6.4%. Model 2 exhibited an MAE of 0.36, an RMSE of 0.62, an R-squared value of 0.94, and a CV-RMSE of 2.2%. Both models demonstrated low error rates and superior prediction performances. Additionally, K-fold cross-validation was conducted to further assess model robustness and performance, providing additional validation of the models' effectiveness (Figure 11).

Table 3. Model performance for predicting power consumption.

Metrics	Scale-Dependent Metrics		Scale-Independent Metrics	
	MAE	RMSE	R^2	CV-RMSE
Pre-HPO	0.22	0.30	0.84	22.5%
After-HPO	0.06	0.08	0.98	6.4%

Table 4. Model performance for predicting indoor temperature.

Metrics	Scale-Dependent Metrics		Scale-Independent Metrics	
	MAE	RMSE	R^2	CV-RMSE
Pre-HPO	0.56	0.74	0.90	2.6%
After-HPO	0.36	0.62	0.94	2.2%

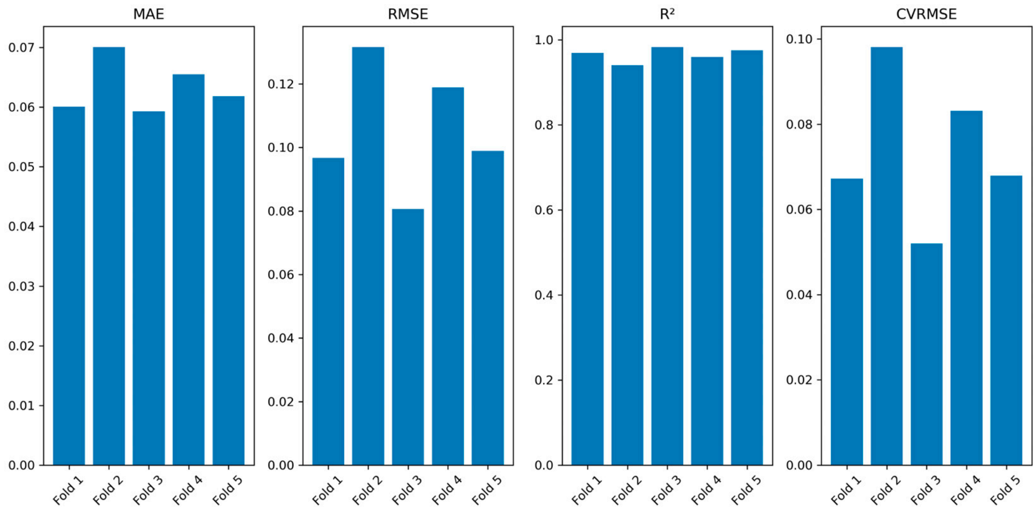
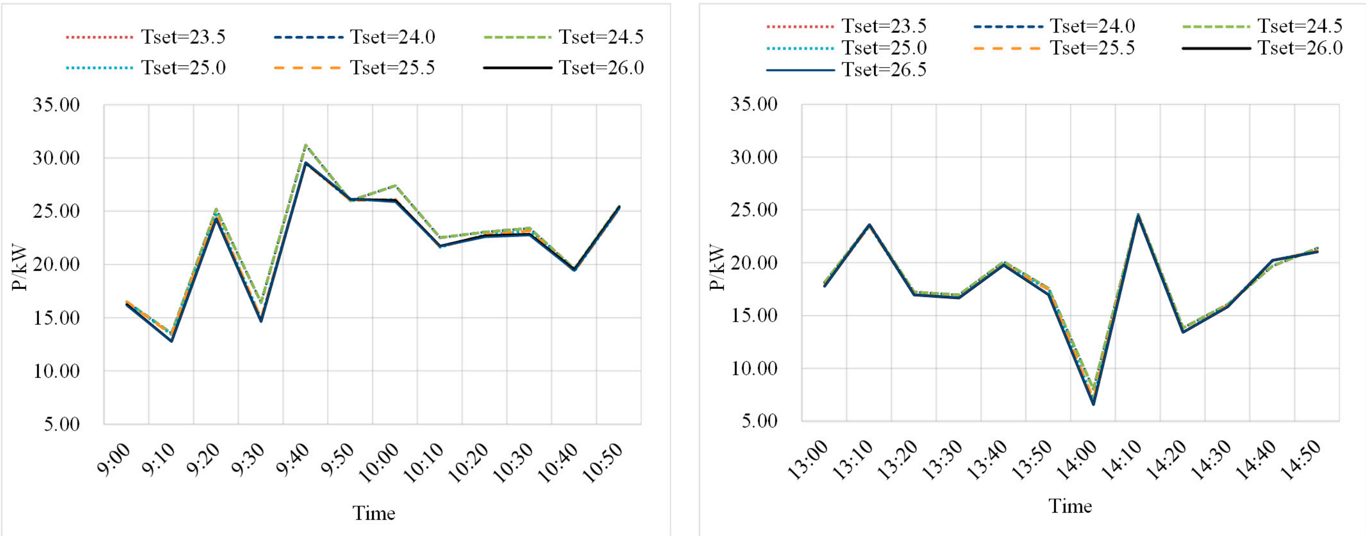


Figure 11. K-fold cross-validation performance metrics.

3.2. Set Temperature Adjustment Analysis

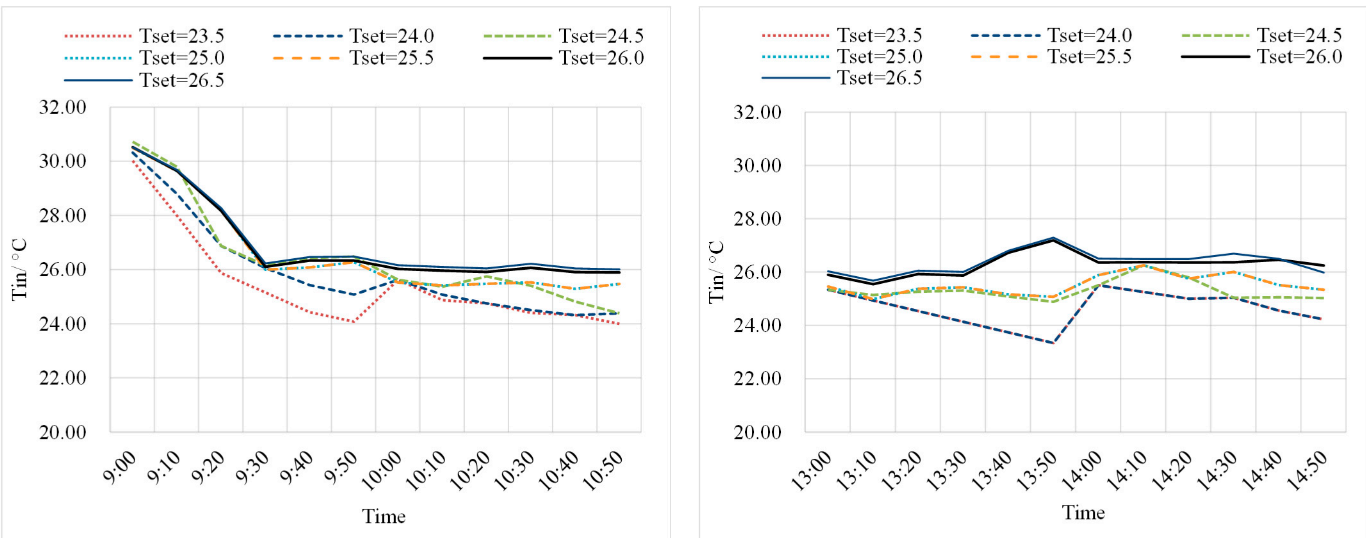
Using the designed control strategy, a specific day during a demand response event was selected to demonstrate the response process and economic benefits of the strategy. Of note, the data from the adjustment day were excluded from the modeling phase to ensure the integrity of the predictive model. In this study, the peak and off-peak electricity prices were 0.86 CNY/kWh and 0.66 CNY/kWh, respectively. The peak load periods were from 10:00 to 11:00 and 14:00 to 15:00, with the DR incentive assumed to be 4.8 CNY/kWh. The AC system performed adjustments, either by pre-cooling or increasing the temperature, at different time points (09:00, 10:00, 13:00, and 14:00). For each time point, seven candidate set temperatures were used. These potential set temperatures were substituted into Equations (4) and (5), and the CatBoost model was employed to obtain the AC power consumption and indoor temperature at different times. The results are illustrated in Figures 12 and 13. Figure 12 shows the impact of adjusting the set AC temperature on power consumption at different time points, while Figure 13 depicts the effect on the indoor temperature.

Through a decision-making process constrained by cost minimization, the optimal set temperature combination was determined. In detail, during peak demand periods, set temperatures were adjusted to 25.0 °C at 09:00, 26.5 °C at 10:00, 25.0 °C at 13:00, and 26.0 °C at 14:00. Starting from 08:00, the total electricity consumption and cost for the corresponding day were calculated, as depicted in Figures 14 and 15. In Figure 14, the power consumption during the DR event corresponds to the AC power consumption at the optimal set temperature combinations. Furthermore, in Figure 15, the cost under the DR event represents the comprehensive cost associated with the AC power consumption at these optimal set temperatures, including both electricity costs and any cost returns during the DR period, as calculated according to Equation (1). Compared with the scenario without demand response, the AC power consumption was reduced by 5.25% and the cost savings amounted to 24.94%, effectively shifting peak loads and demonstrating the economic benefits of this strategy. Specifically, the model effectively balanced the flexibility of temperature adjustment with economic efficiency by using finely tuned set temperature candidates (ranging from 23.5 °C to 26.5 °C in 0.5 °C increments) and comprehensive cost calculations (including both electricity costs and DR incentives), ensuring its effectiveness and cost-efficiency in practical applications.



(**a**) (**b**)

Figure 12. Predicted P under different potential set temperatures during two time intervals: (**a**) 09:00~10:50; and (**b**) 13:00~14:50.



(**a**) (**b**)

Figure 13. Predicted T_{in} under different potential set temperatures during two time intervals: (**a**) 09:00~10:50; and (**b**) 13:00~14:50.

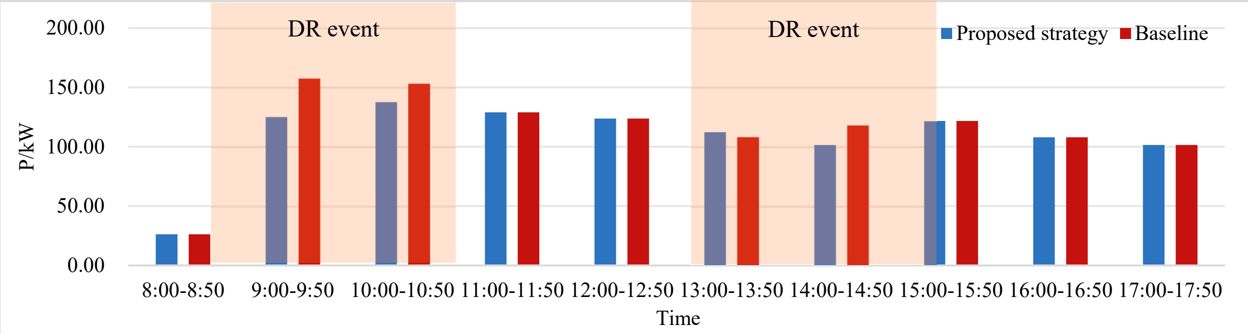


Figure 14. Comparison of daily AC power consumption.

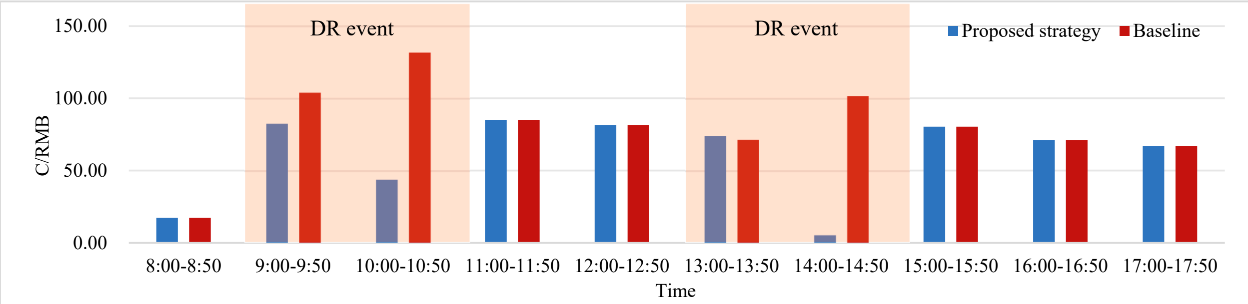


Figure 15. Comparison of daily economic costs.

4. Discussion

Building energy efficiency is closely linked to maintaining a comfortable indoor temperature. Given the insulation and thermal conditions of building envelopes, the operational regulation of AC systems is key to energy savings. Therefore, investigating and identifying the most cost-effective adjustment strategies within a comfortable temperature range utilizing building VESs and by considering the operating states and characteristics of AC systems is essential.

The proposed AC load forecasting and adjustment method, validated through a real-world case study, achieved a 5.25% reduction in electricity consumption and a 24.94% reduction in costs. These reductions highlight the effectiveness of coupling building VESs with AC load regulation. The cost savings are particularly notable because they demonstrate the potential for real-time DR programs to reduce operational expenses, especially during peak periods when electricity prices are higher. Furthermore, the electricity savings indicate a direct reduction in energy demand, which alleviates the burden on the grid during peak hours.

Additionally, unlike previous studies that employed long-term data [52], this study developed and validated the model using only two months of historical data from the case building at 10 min intervals. However, the prediction accuracy for AC power consumption and indoor temperatures, as shown in Tables 3 and 4, indicates that the CatBoost algorithm, calibrated with actual multiparameter operational data rather than simplistic assumed data, achieves high accuracy in its prediction. The model's fit between actual and predicted values is satisfactory, with CV-RMSEs of 6.4% (for AC power consumption) and 2.2% (for the indoor temperature), which are significantly lower than the ASHRAE guideline of 30% CV-RMSE for hourly data [53]. This also translates to savings in time and effort for the simulation analysis.

The physical essence of these results lies in the system's ability to optimize AC operations without compromising indoor comfort. By leveraging the building's thermal inertia, the AC system can be adjusted to reduce energy consumption during peak demand while maintaining a stable indoor environment. This showcases the potential of building VESs to act as a buffer, shifting energy demand and alleviating the stress on power grids during high-demand periods. Additionally, this study demonstrates that real-time data from building operations enhances the accuracy of predictive models, making them more reflective of actual energy dynamics compared to models based on static assumptions or long-term averages.

Buildings' operational energy consumption accounts for approximately 30% of the total energy consumption in a city. Thus, reducing energy use in buildings is a primary focus for countries to address climate change and reduce carbon emissions from building operations [54]. When compared with further enhancing the thermal performance of building envelopes or improving the efficiency of building equipment, fully leveraging the existing building VESs and operational adjustments of AC systems to reduce loads is clearly more economically advantageous. This study demonstrates that, by simply applying the proposed AC load forecasting and adjustment method within an existing building environment and its equipment, the building energy–savings rate can be increased by 5.25%, indicating a strong potential for wider applications. Once scaled, this approach could alleviate grid supply pressures during peak load periods at the regional power supply level, thereby addressing the issue of high operational costs faced by power generation during peak demand [55].

Although this study focuses on a commercial office building using the CatBoost algorithm, the methodology is not limited to office buildings. In demand-side energy management, DR strategies can be applied at various levels, including home, building, and building cluster energy management. In the future, this approach can be expanded to entire buildings and building clusters, incorporating more VES resources to enhance DR and achieve larger energy savings. Moreover, this AI/ML-based DR scheduling method can be applied to other energy management scenarios, such as residential buildings, schools, hospitals, and hotels. The method also extends beyond HVAC scheduling. In DR events, resources like energy storage devices, EVs, and appliances can be integrated with HVAC systems to optimize building energy efficiency. Combining data from these sources enables a comprehensive DR strategy.

This study introduces a novel DR strategy model using the CatBoost algorithm to optimize electricity use through the use of VESs when building HVAC systems. Key innovations include: (1) The innovative application of the CatBoost algorithm: unlike traditional linear models for HVAC systems, this research applies the CatBoost algorithm for its superior nonlinear modeling, high-dimensional data handling, and robust performance, marking its debut in DR control for building HVAC systems. (2) Feature selection and HPO: this study utilizes LASSO regression for efficient feature selection and the OPTUNA framework for automatic hyperparameter tuning, enhancing model performance and interpretability. (3) The use of real-time operational data: while related research typically relies on energy simulation platform data, this study utilizes real-time operational data from office buildings for model development and validation. This approach demonstrates the economic potential of DR through VESs and provides a more accurate representation of building dynamics, reflecting the impact of various real-world factors on the model's performance.

However, this study has certain limitations. Firstly, the predictive model built using ML was validated only with internal data, lacking external validation, which may affect the model's generalization performance. Secondly, the DR incentive mechanism explored in this study is relatively narrow, as it does not fully consider diverse models and incentive strategies, potentially limiting the ability to fully harness VESs potential under DR scenarios. Finally, although the prediction model performs well overall, there are still some instances where the predictions are less accurate. Future work will thus involve expanding the number of validation cases and optimizing the model and data processing methods to enhance its broader applicability.

5. Conclusions

In this study, we investigated a cost-minimizing adjustment method for building AC loads within a comfortable indoor temperature range, considering the VESs in buildings. The proposed approach was validated on a typical office building in Nanjing, China, demonstrating the following key contributions: (1) Accurate Load Forecasting: the model showed high accuracy in forecasting AC power consumption and indoor temperatures, contributing to more reliable demand response strategies. (2) Effective Operational Adjustment: the proposed adjustment method effectively optimized set temperatures, leading to a 5.25% reduction in energy consumption and a 24.94% reduction in operational costs, showcasing the potential for significant energy savings. (3) Rapid Response to Demand: the method allows for quick adjustments in response to real-time energy demand, enhancing the system's ability to participate in demand response events and reduce peak load stress.

As smart buildings continue to develop rapidly, the operational adjustment of building equipment, particularly AC systems, is a key area of research. Supported by sensors, controllers, and building energy management systems, the method developed in this study endows AC systems with self-learning and self-regulation capabilities. This allows for reduced building energy consumption without compromising occupant comfort, effectively supporting the development of smart buildings and unlocking their potential for sustainability.

Author Contributions: Conceptualization, Z.G. and X.W.; methodology, Y.W. (Yao Wang); software, F.Z.; validation, H.Z., M.Z. and Y.W. (Yuxiang Wang); resources, Z.G.; writing—original draft preparation, X.W.; writing—review and editing, Z.G.; visualization, Y.W. (Yao Wang); supervision, H.Z., M.Z.; project administration, F.Z.; funding acquisition, Y.W. (Yuxiang Wang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Opening Funds of State Key Laboratory of Building Safety and Built Environment & National Engineering Research Center of Building Technology, grant number BSBE2022-12. The APC was funded by the Young Elite Scientists Sponsorship Program by CAST, grant number 2022QNRC001.

Data Availability Statement: The data will be made available upon reasonable request to the corresponding author.

Conflicts of Interest: Authors Yao Wang and Fenglei Zhu are employed by the Beijing Glory PKPM Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. IEA. *Buildings*; IEA: Paris, France, 2022. Available online: https://www.iea.org/reports/buildings (accessed on 1 July 2024).
2. D'Ambrosio, D.; Schoenfisch, M. *Electricity Sector*; IEA: Paris, France, 2022. Available online: https://www.iea.org/reports/electricity (accessed on 1 July 2024).
3. Wang, H.; Chen, Y.; Kang, J.; Ding, Z.; Zhu, H. An XGBoost-Based predictive control strategy for HVAC systems in providing day-ahead demand response. *Build. Environ.* **2023**, *238*, 110350. [CrossRef]
4. Li, W.; Xu, P.; Lu, X.; Wang, H.; Pang, Z. Electricity demand response in China: Status, feasible market schemes and pilots. *Energy* **2016**, *114*, 981–994. [CrossRef]
5. Vahid-Ghavidel, M.; Javadi, M.S.; Gough, M.; Santos, S.F.; Shafie-Khah, M.; Catalao, J.P. Demand Response Programs in Multi-Energy Systems: A Review. *Energies* **2020**, *13*, 4332. [CrossRef]
6. Parrish, B.; Heptonstall, P.; Gross, R.; Sovacool, B.K. A systematic review of motivations, enablers and barriers for consumer engagement with residential demand response. *Energy Policy* **2020**, *138*, 111221. [CrossRef]
7. Tyagi, V.V.; Chopra, K.; Kalidasan, B.; Chauhan, A.; Stritih, U.; Anand, S.; Pandey, A.K.; Sarı, A.; Kothari, R. Phase change material based advance solar thermal energy storage systems for building heating and cooling applications: A prospective research approach. *Sustain. Energy Technol. Assess.* **2021**, *47*, 101318. [CrossRef]
8. Ghahary, K.; Abdollahi, A.; Rashidinejad, M.; Alizadeh, M.I. Optimal reserve market clearing considering uncertain demand response using information gap decision theory. *Int. J. Electr. Power Energy Syst.* **2018**, *101*, 213–222. [CrossRef]
9. Guo, Z.; Pinson, P.; Chen, S.; Yang, Q.; Yang, Z. Chance-Constrained Peer-to-Peer Joint Energy and Reserve Market Considering Renewable Generation Uncertainty. *IEEE Trans. Smart Grid* **2020**, *12*, 798–809. [CrossRef]
10. Mu, Y.; Zhang, Y.; Jia, H.; Yu, X.; Zhang, J.; Jin, X.; Deng, Y. Day-ahead optimal scheduling of building energy microgrids based on time-varying virtual energy storage. *IET Renew. Power Gener.* **2022**, *17*, 376–388. [CrossRef]
11. Bu, Y.; Yu, H.; Ji, H.; Song, G.; Xu, J.; Li, J.; Zhao, J.; Li, P. Hybrid data-driven operation method for demand response of community integrated energy systems utilizing virtual and physical energy storage. *Appl. Energy* **2024**, *366*, 123295. [CrossRef]

12. Ji, Y.; Xu, Q.; Luan, K.; Yang, B. Virtual energy storage model of air conditioning loads for providing regulation service. *Energy Rep.* **2020**, *6* (Suppl. 2), 627–632. [CrossRef]
13. Lu, N.; Zhang, Y. Design Considerations of a Centralized Load Controller Using Thermostatically Controlled Appliances for Continuous Regulation Reserves. *IEEE Trans. Smart Grid* **2013**, *4*, 914–921. [CrossRef]
14. Cheng, M.; Sami, S.S.; Wu, J. Benefits of using virtual energy storage system for power system frequency response. *Appl. Energy* **2016**, *194*, 376–385. [CrossRef]
15. Kaliyamoorthy, V.; Krishnasamy, V.; Kandasamy, N. Prediction of virtual energy storage capacity of the air-conditioner using a stochastic gradient descent based artificial neural network. *Electr. Power Syst. Res.* **2022**, *208*, 107879.
16. Braun, J.E.; Lawrence, T.M.; Herrick, R.W.; Klaassen, C.J.; House, J.M. Demonstration of Load Shifting and Peak Load Reduction with Control of Building Thermal Mass. 1970. Available online: www.energytaxincentives.org (accessed on 8 July 2024).
17. Bampoulas, A.; Saffari, M.; Pallonetto, F.; Mangina, E.; Finn, D.P. A fundamental unified framework to quantify and characterise energy flexibility of residential buildings with multiple electrical and thermal energy systems. *Appl. Energy* **2021**, *282*, 116096. [CrossRef]
18. Beil, I.; Hiskens, I.; Backhaus, S. Frequency Regulation From Commercial Building HVAC Demand Response. *Proc. IEEE* **2016**, *104*, 1. [CrossRef]
19. Vijayalakshmi, K.; Vijayakumar, K.; Nandhakumar, K. An ensemble learning model for estimating the virtual energy storage capacity of aggregated air-conditioners. *J. Energy Storage* **2023**, *59*, 106512. [CrossRef]
20. Cao, J.; Zhao, W.; Song, J.; Peng, J.; Yin, R. Development of a dynamic demand response quantification and control framework for fan-coil air-conditioning systems based on prediction models. *Appl. Therm. Eng.* **2024**, *239*, 122098. [CrossRef]
21. Cox, S.J.; Kim, D.; Cho, H.; Mago, P. Real time optimal control of district cooling system with thermal energy storage using neural networks. *Appl. Energy* **2019**, *238*, 466–480. [CrossRef]
22. Hu, M.; Xiao, F.; Jørgensen, J.B.; Wang, S. Frequency control of air conditioners in response to real-time dynamic electricity prices in smart grids. *Appl. Energy* **2019**, *242*, 92–106. [CrossRef]
23. Klein, K.; Herkel, S.; Henning, H.M.; Felsmann, C. Load shifting using the heating and cooling system of an office building: Quantitative potential evaluation for different flexibility and storage options. *Appl. Energy* **2017**, *203*, 917–937. [CrossRef]
24. Hughes, J.T.; Domínguez-García, A.D.; Poolla, K. Identification of Virtual Battery Models for Flexible Loads. *IEEE Trans. Power Syst.* **2016**, *31*, 4660–4669. [CrossRef]
25. Wang, C.; Wang, B.; Cui, M.; Wei, F. Cooling seasonal performance of inverter air conditioner using model prediction control for demand response. *Energy Build.* **2022**, *256*, 111708. [CrossRef]
26. Krstić, H.; Teni, M. Review of Methods for Buildings Energy Performance Modelling. *J. IOP Mater. Sci. Eng.* **2017**, *245*, 042049. [CrossRef]
27. Zhou, X.; Du, H.; Xue, S.; Ma, Z. Recent advances in data mining and machine learning for enhanced building energy management. *Energy* **2024**, *307*, 132636. [CrossRef]
28. Antonopoulos, I.; Robu, V.; Couraud, B.; Kirli, D.; Norbu, S.; Kiprakis, A.; Flynn, D.; Elizondo-Gonzalez, S.; Wattam, S. Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. *Renew. Sustain. Energy Rev.* **2020**, *130*, 109899. [CrossRef]
29. Joiner, I.A. Artificial intelligence: AI is nearby. In *Emerging Library Technologies*; Joiner, I.A., Ed.; Chandos Publishing: Oxford, UK, 2018; pp. 1–22.
30. Kaplan, A.; Haenlein, M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Bus. Horiz.* **2019**, *62*, 15–25. [CrossRef]
31. Gonzalez, P.A.; Zamarreno, J.M. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy Build.* **2005**, *37*, 595–601. [CrossRef]
32. Kalogirou, S.A.; Bojic, M. Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy* **2000**, *25*, 479–491. [CrossRef]
33. Cabrera DF, M.; Zareipour, H. Data association mining for identifying lighting energy waste patterns in educational institutes. *Energy Build.* **2013**, *62*, 210–216. [CrossRef]
34. Santamouris, M.; Mihalakakou, G.; Patargias, P.; Gaitani, N.; Sfakianaki, K.; Papaglastra, M.; Pavlou, C.; Doukas, P.; Primikiri, E.; Geros, V.; et al. Using intelligent clustering techniques to classify the energy performance of school buildings. *Energy Build.* **2007**, *39*, 45–51.
35. Zeng, Y.; Zhang, Z.; Kusiak, A. Predictive modeling and optimization of a multi-zone HVAC system with data mining and firefly algorithms. *Energy* **2015**, *86*, 393–402. [CrossRef]
36. Perera AT, D.; Kamalaruban, P. Applications of reinforcement learning in energy systems. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110618. [CrossRef]
37. Fan, B.; Du, Z.; Jin, X.; Yang, X.; Guo, Y. A hybrid FDD strategy for local system of AHU based on artificial neural network and wavelet analysis. *Build. Environ.* **2010**, *45*, 2698–2708. [CrossRef]
38. Mulumba, T.; Afshari, A.; Yan, K.; Shen, W.; Norford, L.K. Robust model-based fault diagnosis for air handling units. *Energy Build.* **2015**, *86*, 698–707. [CrossRef]
39. Namburu, S.M.; Azam, M.S.; Luo, J.; Choi, K.; Pattipati, K.R. Data-Driven Modeling, Fault Diagnosis and Optimal Sensor Selection for HVAC Chillers. *IEEE Trans. Autom. Sci. Eng.* **2007**, *4*, 469–473. [CrossRef]

40. Yuan, J.; Xiao, Z.; Chen, X.; Lu, Z.; Li, J.; Gang, W. A Temperature & Humidity Setback Demand Response Strategy for HVAC Systems. *Sustain. Cities Soc.* **2021**, *75*, 103393.
41. Yoon, A.Y.; Kim, Y.J.; Zakula, T.; Moon, S.I. Retail electricity pricing via online-learning of data-driven demand response of HVAC systems. *Appl. Energy* **2020**, *265*, 114771. [CrossRef]
42. Zhou, D.; Balandat, M.; Tomlin, C. Residential demand response targeting using machine learning with observational data. In Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016.
43. Zhou, D.; Balandat, M.; Tomlin, C. A Bayesian Perspective on Residential Demand Response Using Smart Meter Data. In Proceedings of the 2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 27–30 September 2016.
44. Behl, M.; Smarra, F.; Mangharam, R. DR-Advisor: A data-driven demand response recommender system. *Appl. Energy* **2016**, *170*, 30–46. [CrossRef]
45. Ben-Nakhi, A.E.; Mahmoud, M.A. Cooling load prediction for buildings using general regression neural networks. *Energy Convers. Manag.* **2004**, *45*, 2127–2141. [CrossRef]
46. Bian, J.; Wang, J.; Yece, Q. A novel study on power consumption of an HVAC system using CatBoost and AdaBoost algorithms combined with the metaheuristic algorithms. *Energy* **2024**, *302*, 131841. [CrossRef]
47. Kaligambe, A.; Fujita, G.; Keisuke, T. Estimation of Unmeasured Room Temperature, Relative Humidity, and CO_2 Concentrations for a Smart Building Using Machine Learning and Exploratory Data Analysis. *Energies* **2022**, *15*, 4213. [CrossRef]
48. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363.
49. Kok, B.C.; Choi, J.S.; Oh, H.; Choi, J.Y. Sparse Extended Redundancy Analysis: Variable Selection via the Exclusive LASSO. *Multivar. Behav. Res.* **2021**, *56*, 426–446. [CrossRef] [PubMed]
50. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.
51. Li, L.; Jamieson, K.; Desalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *J. Mach. Learn. Res.* **2018**, *18*, 1–52.
52. Langtry, M.; Wichitwechkarn, V.; Ward, R.; Zhuang, C.; Kreitmair, M.J.; Makasis, N.; Conti, Z.X.; Choudhary, R. Impact of data for forecasting on performance of model predictive control in buildings with smart energy storage. *Energy Build.* **2024**, *320*, 114605. [CrossRef]
53. George, R. *(ASHRAE Guideline Project Committe14P). ASHRAE GUIDELINE 14-2002 Measurement of Energy and Demand Savings*; American Society of Heating, Refrigerating and Air-Conditioning Engineers: Atlanta, GA, USA, 2002; pp. 40–41.
54. Nozarian, M.; Hajizadeh, A.; Fereidunian, A. A methodological review on management of building cluster meso energy hubs in accordance with the agile framework: Exploring flexibility upward the building-cluster-city hierarchy. *Sustain. Energy Technol. Assess.* **2024**, *67*, 103834. [CrossRef]
55. Saleem, M.U.; Shakir, M.; Usman, M.R.; Bajwa, M.H.T.; Shabbir, N.; Shams Ghahfarokhi, P.; Daniel, K. Integrating smart energy management system with internet of things and cloud computing for efficient demand side management in smart grids. *Energies* **2023**, *16*, 4835. [CrossRef]