Build Simul (2024) 17: 1439–1460
https://doi.org/10.1007/s12273-024-1152-3

Energy consumption dynamic prediction for HVAC systems based on feature clustering deconstruction and model training adaptation

Huiheng Liu[1], Yanchen Liu[1] (✉), Huakun Huang[2], Huijun Wu[1], Yu Huang[1]

1. School of Civil Engineering and Transportation, Guangzhou University, Guangzhou 510006, China
2. School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China

Abstract

The prediction of building energy consumption offers essential technical support for intelligent operation and maintenance of buildings, promoting energy conservation and low-carbon control. This paper focused on the energy consumption of heating, ventilation and air conditioning (HVAC) systems operating under various modes across different seasons. We constructed multi-attribute and high-dimensional clustering vectors that encompass indoor and outdoor environmental parameters, along with historical energy consumption data. To enhance the *K*-means algorithm, we employed statistical feature extraction and dimensional normalization (SFEDN) to facilitate data clustering and deconstruction. This method, combined with the gated recurrent unit (GRU) prediction model employing adaptive training based on the Particle Swarm Optimization algorithm, was evaluated for robustness and stability through *k*-fold cross-validation. Within the clustering-based modeling framework, optimal submodels were configured based on the statistical features of historical 24-hour data to achieve dynamic prediction using multiple models. The dynamic prediction models with SFEDN cluster showed a 11.9% reduction in root mean square error (RMSE) compared to static prediction, achieving a coefficient of determination (R^2) of 0.890 and a mean absolute percentage error (MAPE) reduction of 19.9%. When compared to dynamic prediction based on single-attribute of HVAC systems energy consumption clustering modeling, RMSE decreased by 12.6%, R^2 increased by 4.0%, and MAPE decreased by 26.3%. The dynamic prediction performance demonstrated that the SFEDN clustering method surpasses conventional clustering method, and multi-attribute clustering modeling outperforms single-attribute modeling.

Keywords

HVAC system energy consumption
clustering analysis
deep learning
model adaptation
dynamic prediction

Article History

Received: 28 December 2023
Revised: 21 May 2024
Accepted: 02 June 2024

1 Introduction

1.1 Background

The building sector contributes significantly to global energy consumption and carbon emissions, accounting for approximately 36% and 38% respectively (Global Alliance for Buildings and Construction 2020). In China, the energy consumption and carbon emissions of buildings account for 21% and 22% of the national total, respectively. Among them, the energy consumption of buildings in 2019 was 1.02 billion tons of coal equivalent (Chinese Society for Urban Studies 2022). Given the current situation and the goal of achieving carbon peaking by 2030 and carbon neutrality before 2060, we face the formidable challenge of saving building energy consumption and decreasing building carbon emissions.

The prediction of building energy consumption, a key strategy for enhancing building energy efficiency, has numerous applications in various fields such as building energy control (Cholewa et al. 2022; Elnour et al. 2022; Erfani et al. 2024; Li et al. 2024), design optimization (Kim and Suh 2021; Shen et al. 2024), retrofit evaluation (Li et al. 2019; Kim and Kim 2020; Seo et al. 2022), energy pricing guidance (Jota et al. 2011; Atalay et al. 2019), fault diagnosis (Fan et al. 2017), demand-side management (Sala-Cardoso et al. 2018;

E-mail: liuych@gzhu.edu.cn

List of symbols

a	whole data of one parameter
$\boldsymbol{B}_{\text{conventional}}$	cluster matrix for conventional cluster method
$\boldsymbol{B}_{\text{SFEDN}}$	cluster matrix for SFEDN cluster method
CO_2	indoor CO_2 concentration
CV	coefficient of variation
d	Euclidean distance
E	HVAC systems energy consumption
J	solar radiation intensity
k	fold of across validation
K	number of clusters
m	attribute of cluster vector
max	the maximum value
min	the minimum value
R^2	coefficient of determination
RH_{in}	indoor relative humidity
RH_{out}	outdoor relative humidity
T_{in}	indoor dry bulb temperature
T_{out}	outdoor dry bulb temperature
X	judging condition to model matching
$\boldsymbol{x}$	vector of one cluster parameter
Y	cluster center
$y_{1,j}$	measured value
$y_{2,j}$	predicted value
$\overline{y_1}$	average of measured value

Abbreviations

ANN	artificial neural network
DNN	deep neural network
DTW	dynamic time warping
GA	genetic algorithm
GRU	gate recurrent unit
HVAC	heating, ventilation, and air conditioning
LSTM	long short-term memory neural network
MAPE	mean absolute percentage error
PSO	particle swarm optimization
RMSE	root mean square error
RNN	recurrent neural network
SCOA	sine cosine optimization algorithm
SVR	support vector regression

Mohammed et al. 2021), indoor discomfort evaluation (Boithias et al. 2012) and COVID-19 prevention and control (Risbeck et al. 2021; Li et al. 2022). Prediction methods can be broadly categorized into white-box, black-box, and gray-box approaches (Afram and Janabi-Sharifi 2014; Wei et al. 2018). White-box and gray-box models rely on physical principles, detailed building energy characteristics and assumptions in the modeling process (Afram and Janabi-Sharifi 2014; Liu et al. 2023). In contrast, black-box models analyze large datasets using algorithms to identify patterns in the data, enabling automated decision-making without relying on thermodynamic principles or specific building energy characteristics. However, the maximum achievable performance of black-box models is constrained by the characteristics of the data and its quality (Chen et al. 2022). In particular, the intricate patterns of energy systems and the seasonal influence of environmental changes have heightened the challenges in prediction. Consequently, numerous researchers have turned to clustering methods as a strategy to alleviate these prediction challenges. This involves subdividing the overall dataset into subsets with similarities, aiming to enhance the accuracy and effectiveness of predictions.

The clustering algorithms mainly include fuzzy clustering, hierarchical clustering, density-based clustering, and *K*-means clustering. The current clustering metrics encompass squared cosine distance, Euclidean distance, Manhattan distance, Hamming distance, correlation distance, and dynamic time warping (DTW). They are used to measure the correlation between data. Subsequently, the elbow method (based on the sum of squared errors), silhouette method, and Dunn index are utilized to determine the optimal number of clusters. Czétány et al. (2021) used the hierarchical, fuzzy *K*-means, and *K*-means clustering methods to perform cluster analysis on the daily and annual electricity load time series of nearly a thousand single-family households in Hungary to determine energy consumption patterns. The elbow method, silhouette method, and Dunn index were employed to evaluate the clustering results, all indicating that clustering performance of *K*-means algorithm was the best. Moreover, a review of 127 studies on the application of clustering methods in building energy system design and energy policy formulation found that *K*-means is the most widely used clustering method, especially demonstrating good robustness in problems involving time series clustering (Kang et al. 2023). Ruiz et al. (2020) researched on time series feature extraction in energy, compared the clustering effects of squared cosine distance, Euclidean distance, Manhattan distance, Hamming distance, and correlation distance in *K*-means, Hierarchical clustering, *K*-medoids, and Gaussian mixture algorithms. They found that the clustering results of using squared Euclidean distance is better than other cluster methods.

Commonly, clustering is one of the effective methods to recognize energy patterns and help prediction models to extract features in complex energy dataset. Studying on the

prediction of day-ahead hourly electric load of commercial buildings with non-stationary operation, Chen et al. (2017) employed a fuzzy clustering algorithm to cluster daily and hourly electric loads by rainy condition, day type, average hourly electric consumption, standard deviation of the hourly electric consumption, minimum, maximum and average ambient temperature for a whole day. And then, to reflect on the seasonal changes in outdoor climate, Luo et al. (2020) utilized *K*-means clustering to extract historical meteorological features throughout the year for predictive modeling of building energy usage. Furthermore, to analyze the seasonal changes and complex trends in energy consumption data, Somu et al. (2021) utilized the *K*-means algorithm to cluster the energy consumption data of complex and dynamic from an academic building. Li et al. (2021a) established a multi-attribute high-dimensional (the dimensions of every attribute are 24) clustering model that integrates the electricity, heating, cooling, and gas loads for complex load analysis of regional comprehensive energy systems. Bourdeau et al. (2021) employed cluster methods to study the impact of data collection time step and time frame on the data characteristics of daily electric load profiles in 14 educational buildings on the same campus. The results indicated that when the time frame of the dataset exceeded three months, all methods were influenced by seasonality, resulting in lower classification accuracy. Additionally, relying solely on the distribution of daily power loads for classification was limited, and additional metadata was required for explanatory variable investigation. Moreover, when additional variables were included in clustering, numerical differences and the correlation among the various data had potential interference with the clustering effect. To illustrate the correlation between clustering attributes and enhance the clustering effectiveness of the *K*-means algorithm, Chen et al. (2022) introduced a weighted Euclidean distance based on the Pearson correlation coefficient between the features (input variables) and cooling load (output variable). The prediction results demonstrated that the improved method outperformed the scenario without clustering, achieving a reduction in mean absolute percentage error (MAPE) of over 14%. Overall, concerning the dimension and attribute of clustering vectors remains an issue. Although the previous study mentioned the dimensions or attributes of clustering vectors, there was no comparison of clustering using different attributes and dimensions, let alone an analysis of prediction results based on this issue.

On the other hand, determining how to partition the overall dataset into training set and testing set for cluster method and prediction models poses a pivotal challenge. In a departure from the conventional approach of seasonal division, Zhang et al. (2019) opted for the use of the *K*-means algorithm to cluster historical cooling loads, creating training sets for distinct load patterns. Ensuring each dataset comprised time series groups with similar features, trends, and seasonal patterns, Kohli et al. (2022) applied *K*-means based on DTW and silhouette coefficient to categorize the dataset into K clusters. Subsequently, 60% of the data from each cluster was extracted for the formation of a pre-training set, 20% for model parameter fine-tuning, and the remaining 20% for testing. Moreover, the meticulous matching of clustering training sets with testing sets holds equal importance in enhancing the predictive capability of energy models. In essence, the selection and effective utilization of well-trained models during the testing phase are crucial. Zhang et al. (2019) used *K*-means clustering to partition the training dataset and employed the *K*-nearest neighbor algorithm to match the testing set with the best training set. Zhang et al. (2020) used the weighted Manhattan distance to quantify the dissimilarity between the test set and the training dataset and found that the predictive accuracy of the test set decreased as the weighted Manhattan distance increased. This is because in testing set, the prediction performance of models primarily depended on the experiences learned from the training progress. Therefore, the essence of the model selection problem lies in how to achieve a match between testing set and training set, which can be further interpreted as the similarity of them. Chen et al. (2022) used the weighted Euclidean distance to select the training data with the closest distance to the predicted sample as the current model's training sample, thereby achieving a match between the forecasting and training dataset. However, the uncertainty of future conditions cannot be acquired in advance. It is unreasonable that prioritize the improvement of predictive accuracy by making prior assumptions about future conditions and ignoring the existence of uncertainty.

After clustering, another critical consideration arises: how to effectively train models across different clusters. The persistent use of a single fixed architecture for model training (Hsu 2015; Ko et al. 2017), grid search (Chitalia et al. 2020), or manual tuning, is employed to enhance adaptability to each dataset. However, relying solely on these methods fails to meet practical requirements for simplicity. Consequently, recognizing the need to enhance the adaptive capability of models, numerous scholars have advocated for the integration of optimization algorithms such as particle swarm optimization (PSO), genetic algorithm (GA), sine cosine optimization algorithm (SCOA), with predictive models like recurrent neural network (RNN), artificial neural network (ANN), support vector regression (SVR). Luo et al. (2020) utilized GA to optimize deep neural network (DNN) models for adaptive training in clustering results, resulting

in corresponding sub-models. The prediction results on the test set showed a reduction of 11.9% in MAPE. Luo and Oyedele (2021) used GA to select the optimal structure of the long short-term memory neural networks (LSTM) in order to improve its prediction accuracy and robustness. However, the GA algorithm was prone to local optima when optimizing hyperparameters through crossover and mutation methods. Kim and Cho (2019) found that the random initialization mechanism of PSO can effectively avoid the limitations of GA. By using PSO for global search of hyperparameters in the CNN-LSTM model, and after conducting 10-fold cross-validation, the prediction accuracy of the CNN-LSTM model was improved. Pawar et al. (2020) constructed a PSO-SVR model to predict building energy consumption on an hourly basis and one day in advance. The predictive accuracy of this model was superior to ANN, SVR, and PSO-ANN models. Zhou et al. (2020) utilized the artificial bee colony (ABC) and PSO to optimize a multilayer perceptron (MLP) neural network for estimating the cooling and heating loads of energy-efficient residential buildings. The results indicated that through the application of ABC and PSO, the mean absolute error for heating loads decreased by 22.32% and 24.28% respectively, while for cooling loads, it decreased by 10.36% and 12.00%. Somu et al. (2021) improved SCOA to optimize the hyperparameters of the CNN-LSTM model. The improved SCOA utilizes mutation operators based on Haar wavelets to update positions, thereby avoiding local optima. Li et al. (2023) used PSO to optimize the hyperparameters of the Informer model and achieved a 56.11% improvement in accuracy compared to the original model. Fan et al. (2023) proposed a PSO-BOA-LSTM-SVR model to address the volatile and variable nature of electricity load forecasting. By utilizing PSO, the model enhances the global optimization capability of the Bayesian optimization algorithm, thereby preventing it from getting stuck in local optima. This approach allows for better utilization of LSTM in extracting feature variables related to electricity consumption. Finally, the model integrates with an SVR model that exhibits strong mapping capabilities for nonlinear data. Overall, while much research has proposed various prediction models based on training adaptation, there has been less focus on validating the model's generalization in the training sets. It is essential to implement cross-validation for effective model training. Furthermore, model adaptive training techniques have rarely been applied in conjunction with clustering methods.

1.2 Research gap and objectives

Cluster analysis was initially employed to recognize building energy characteristics and determine energy consumption trends over time and seasons. However, when applied to energy prediction, the clustering effect based on energy was limited. Despite introducing additional energy-related data to expand clustering vector attributes and dimensions, the performance of distance-based clustering algorithms was constrained by high-dimensional data. Therefore, there is an urgent need for a method that can reduce the dimensions of cluster vectors and weaken the correlation influence between data to strengthen cluster algorithms for model's prediction. Meanwhile, in existing research, the testing set were clustered, and its role is often confined to static predictions using a single model, failing to fully exploit the essential functionality of cluster modeling. Lastly, various adaptive models can support the clustering data to train flexibly, lacking the validation of the generalization ability and robustness of adaptive models in different subsets from clustering.

This study integrated meteorological data, indoor environmental data, and historical energy consumption data to construct multi-attribute and high-dimensional vectors. A method of statistical feature extraction was employed to reduce dimensions while preserving the original multi-attribute information. Subsequently, based on the principles of Euclidean distance clustering, the normalization was performed according to vector dimension instead of the conventional method according to attribute. This approach aimed to ensure the retention of original information while mitigating the impact of data attribute differences and correlations on clustering effects. Secondly, in order to enhance the training adaptability of the deep learning model, the particle swarm algorithm was employed to optimize the key parameters of the gated recurrent unit (GRU) model. Additionally, *k*-fold cross-validation was introduced to determine the sub-model architecture with the best performance. Thirdly, to achieve dynamic predictions, the study utilized historical 24-hour feature data to assess the uncertainty of future conditions and match the best sub-model for dynamic prediction using multiple models.

2 Methods

The research framework of this paper is shown in Figure 1. In the data preprocessing stage, the overall training-validation set is clustered by the *K*-means algorithm of statistical feature extraction and dimensional normalization (SFEDN). During the model training and validation stage, PSO is used to the parameter optimization for the GRU model, enabling adaptive training within each cluster. The *k*-fold cross- validation is introduced to assess the robustness and stability of the submodels. In the model matching stage, the best sub-model is selected by calculating the similarity between the statistical features based on the historical 24 hours and the cluster centers. In the dynamic prediction stage, feedback information

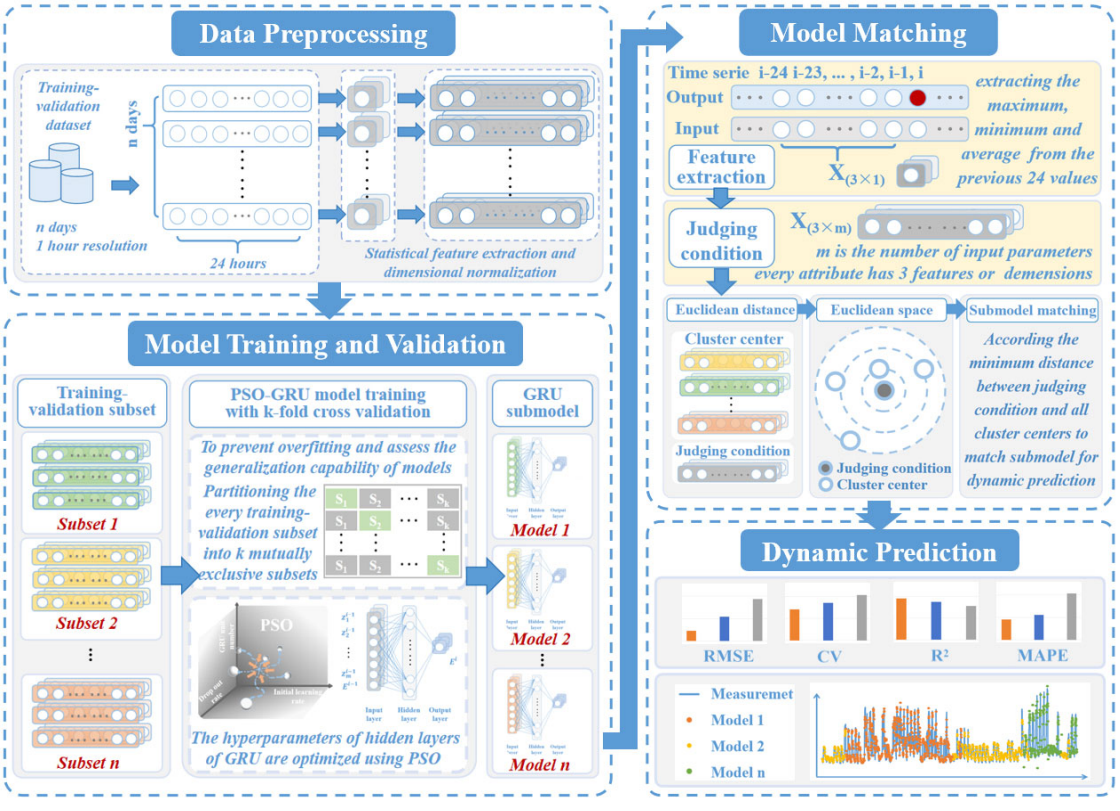


Fig. 1 Research framework

from the model matching mechanism is used to achieve dynamic prediction using multiple models throughout the process.

2.1 Data preprocessing

The raw data consists of outdoor meteorological conditions, indoor environmental parameters, and data related to building energy consumption systems. To ensure the accuracy of the analysis, the raw data underwent necessary preprocessing. This involved supplementing missing values through linear interpolation. Outliers were also identified and removed using quartiles (Xiao and Fan 2014), with rejected outliers further supplemented through linear interpolation. Before inputting the data into the prediction models, normalization of inputs was performed.

The dataset was divided into a training-validation set and a testing set, with the training-validation set comprising 80% of the data and the testing set comprising 20%. The process of constructing the multi-attribute high-dimensional clustering vectors, as shown in Figure 2, begins by reshaping the *n*-day hourly training-validation dataset into

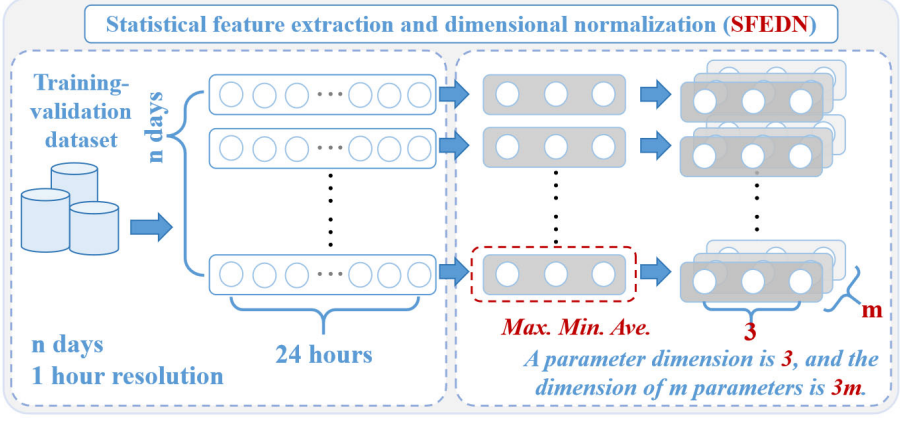


Fig. 2 Data preprocessing for clustering of the *K*-means algorithm of statistical feature extraction and dimensional normalization (SFEDN)

24-dimensional (hourly) vectors based on daily units. The maximum, minimum, and average values within each day are statistic as the features for each attribute. If there are a total of m attributes, the dimension of the vector becomes $3m$. This results in an n-row, $3m$-column clustering matrix. Subsequently, considering the calculation principle of Euclidean distance point to point, this method differs from the conventional normalization method, which is typically carried out attribute-wise. Instead, it normalizes according to the dimensions of the clustering vectors. This means that the normalization reflected on the dimensions of the clustering vectors for each attribute's feature values, rather than normalizing each attribute independently (Figure 3). Normalization can reduce the sensitivity of clustering algorithms to variations in the magnitude of different data, thereby enhancing the robustness of the clustering algorithm (Fan et al. 2021). The equations of the min-max normalization can refer to Equation (1) and Equation (2).

The raw cluster matrix ($\boldsymbol{A}$) comprises m attributes ($\boldsymbol{x}$), each derived from m parameters.

$$\boldsymbol{A} = [x_1, ..., x_m]$$

There are $24n$ data points (representing n whole days) in each attribute ($\boldsymbol{x}_{\text{conventional}}$), and its dimension is $24n$.

$$\boldsymbol{x}_{\text{conventional}} = [a_{ij}]_{n\times 24}, \quad (1 \le i \le n, 1 \le j \le 24)$$

With statistical feature extraction, the dimension of each attribute ($\boldsymbol{x}_{\text{SFEDN}}$) decreases to $3n$.

$$b_{i1} = \max\{a_i\}, b_{i2} = \min\{a_i\}, b_{i3} = \text{ave}\{a_i\}$$

$$\boldsymbol{x}_{\text{SFEDN}} = [b_{i1}, b_{i2}, b_{i3}]_{n\times 3}, \quad (1 \le i \le n)$$

The conventional normalization method is performed attribute-wise. In contrast, the SFEDN method normalizes based on the dimensions of the clustering vectors. Both methods utilize data min-max normalization, and the equations for min-max normalization are as follows:

$$a'_{ij,\text{conventional}} = \frac{a_{ij} - \min\{a\}}{\max\{a\} - \min\{a\}}, \quad (1 \le i \le n, 1 \le j \le 24) \tag{1}$$

$$b'_{ij,\text{SFEDN}} = \frac{b_{ij} - \min\{b_j\}}{\max\{b_j\} - \min\{b_j\}}, \quad (1 \le i \le n, 1 \le j \le 3) \tag{2}$$

Then, the cluster matrix reconstructed as

$$\boldsymbol{B}_{\text{conventional}} = [a'_{ij,\text{conventional}}]_{n\times 24m}, \quad (1 \le i \le n, 1 \le j \le 24m)$$

$$\boldsymbol{B}_{\text{SFEDN}} = [b'_{ij,\text{SFEDN}}]_{n\times 3m}, \quad (1 \le i \le n, 1 \le j \le 3m)$$

where, m is the attribute of cluster vector or the number of cluster parameters, $\boldsymbol{x}$ is the vector of one cluster parameter, a is the whole data of one parameter, $\max\{a_i\}$ is the maximum value for a whole day, $\min\{a_i\}$ is the minimum value for a whole day, $\text{ave}\{a_i\}$ is the average value for a whole day, $\min\{a\}$ is the minimum of one parameter, $\max\{a\}$ is the maximum of one parameter, $\min\{b_j\}$ is the minimum of all of b_j, $\max\{b_j\}$ is the maximum of all of b_j, $\boldsymbol{B}_{\text{conventional}}$ is the cluster matrix for conventional cluster method, $\boldsymbol{B}_{\text{SFEDN}}$ is the cluster matrix for SFEDN cluster method.

2.2 Cluster processing and feature representation

This paper referenced Kang et al. (2023) and Ruiz et al. (2020) to utilize the K-means algorithm for clustering, Euclidean distance as the clustering metric, and the optimal number of clusters was determined by silhouette coefficient. K-means is one of the robust clustering algorithms used to group similar data points into a predefined n number of clusters. By employing randomly initialized cluster centers,

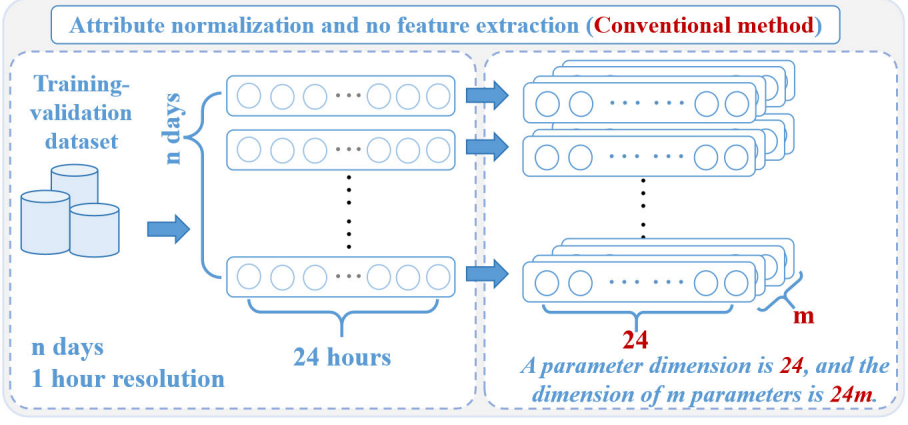


Fig. 3 Data preprocessing for clustering of the K-means algorithm of attribute normalization and no feature extraction (conventional method)

points are partitioned into n clusters based on their proximity to these centers. New cluster centers are generated through the average of points within each cluster. This process iterates until the centroids no longer change, completing the clustering. The silhouette coefficient method (Arbelaitz et al. 2013; Li et al. 2020) is used to evaluating the clustering effectiveness based on the proximity and separation of the clusters. The proximity refers to the average distance between a vector and other vectors within the same cluster, while the separation refers to the average distance between a vector and all vectors in other clusters. The silhouette coefficient value closer to 1 indicating better clustering performance. The Euclidean distance has been proven to be more stable in analyzing energy-related features compared to other distances, as discussed in Bourdeau et al. (2021). Moreover, in Euclidean space, the Euclidean distance represents the length of the line connecting two points and is commonly used as a measure of similarity. Therefore, this paper adopts the Euclidean distance as the metric for K-means algorithm clustering.

As shown in Figure 4, different subsets are obtained through clustering analysis of the training-validation set using the K-means algorithm based on Euclidean distance and silhouette coefficient method. Each subset will be utilized for adaptive model training and verification of clustering effectiveness, while the optimized sub-models will offer model selection for dynamic prediction applications. Based on the Pearson correlation coefficient (Sala et al. 2021), it accurately represents the feature differences among various subsets by examining the relationship between the model's input and output.

2.3 PSO enabled adaptive GRU model

The GRU model, which is a variant of LSTM, aims to tackle the problem of gradient disappearance in long-term memory and back propagation (Chung et al. 2014). It boasts fast response speed and high prediction accuracy. The PSO is a classic intelligent algorithm. It analogizes the optimization problem as a group of raptors, the solution space as the flying space of a bird flock, and each bird's position in the space simulates the complementary feeding process within the bird flock. The solution space represents the solution to be optimized. The core idea is the sharing of information among the particles in the group. In the problem-solving space, the movement of particles transitions from disorder to order, thereby obtaining the optimal solution to the problem (Gad 2022).

As shown in Figure 5, the GRU model achieves adaptive optimization in different training sets through PSO, thereby generating sub-models for energy consumption prediction that adapt to different features. Subsequently, each sub-model needs to undergo k-fold cross-validation to select models with good robustness and stability.

The k-fold cross validation is commonly used for establishing baseline models and evaluating model accuracy. It helps to prevent overfitting and assess the generalization capability of the model (Peng et al. 2020; Li et al. 2021b; Abdallah et al. 2022). As shown in Figure 6(a), the training set consists of the union of $k - 1$ subsets, while the remaining subset is designated as the testing set. And then, the predicted results are averaged over k cross-validation experiments. This article adopted 5-fold cross validation,

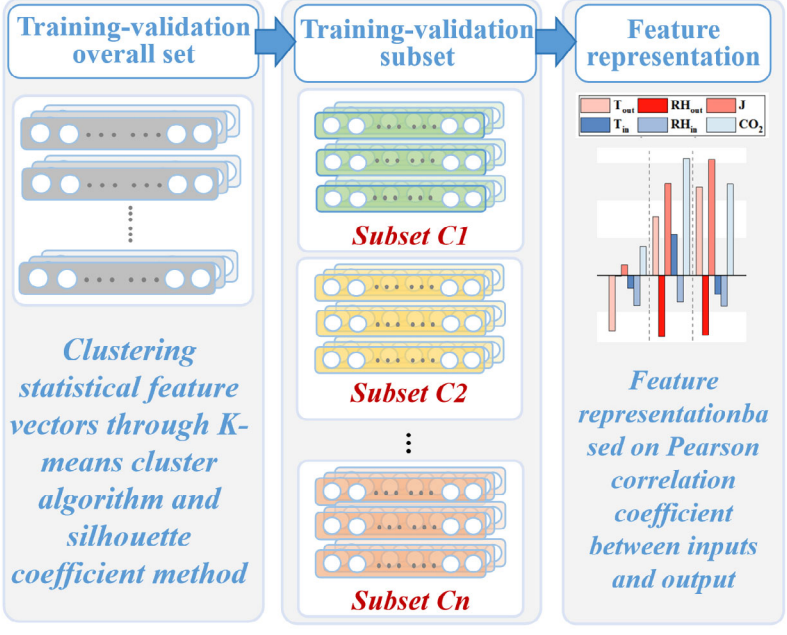


Fig. 4 The cluster processing and feature analysis of training-validation dataset

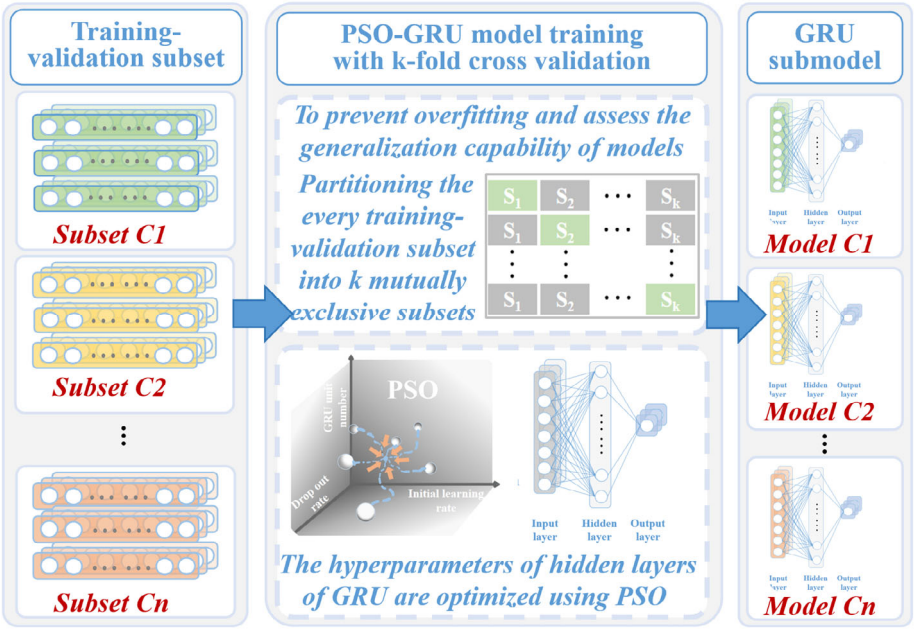
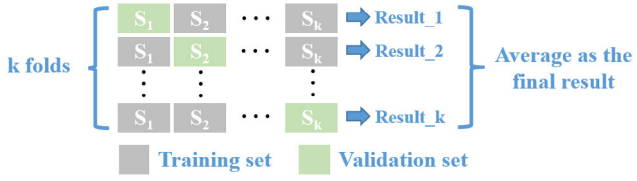
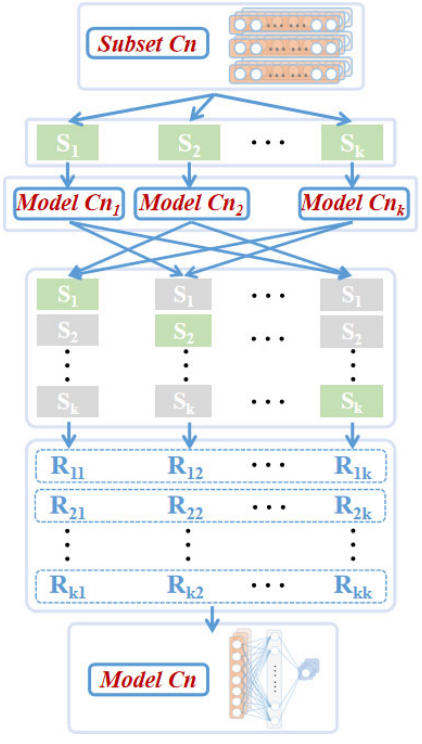


Fig. 5 The modeling process of training adaptation based on PSO and the assessment of generalization capability based on *k*-fold cross validation



(a) The calculation method of prediction results in validation set



(b) The training process of prediction models using *k*-fold cross validation

Fig. 6 The process of *k*-fold cross validation

where the training-validation set is divided into a 4:1 ratio (Vijayalakshmi et al. 2023). As Figure 6(b) shows, each subset will undergo *k* validation to generate k candidate sub-models. Due to the changes in the validation set, the hyperparameters of the trained models will differ, indicating that the performance of the models will also vary. These sub-models with different parameters will then be cross-validated in the other $k - 1$ training sets. Finally, based on the minimum of MAPE, a sub-model with good robustness and stability in the *k* sub-models will be selected for dynamic prediction.

2.4 Model matching for dynamic prediction

As shown in the Figure 7, while predicting the HVAC system energy consumption of *i* time sample, the key step is to choose the optimal prediction model. According to the judging condition $X_{(3\times m)}$, the first step is to calculate the distance between of $X_{(3\times m)}$ and cluster center *Y*. And then, the second step is model matching. For instance, if the judging condition indicates that the best submodel for predicting energy consumption at time *i* is Model C2. After that, the third step is to determine the input parameters of the model. Lastly, the fourth step is predicting the HVAC system energy consumption of *i* time sample. This process entails selecting a model based on the judging condition for each prediction moment and utilizing the chosen model to make the prediction. The final presentation of results reflects the alternating prediction of different models.

As shown in Figure 8, when predicting the energy

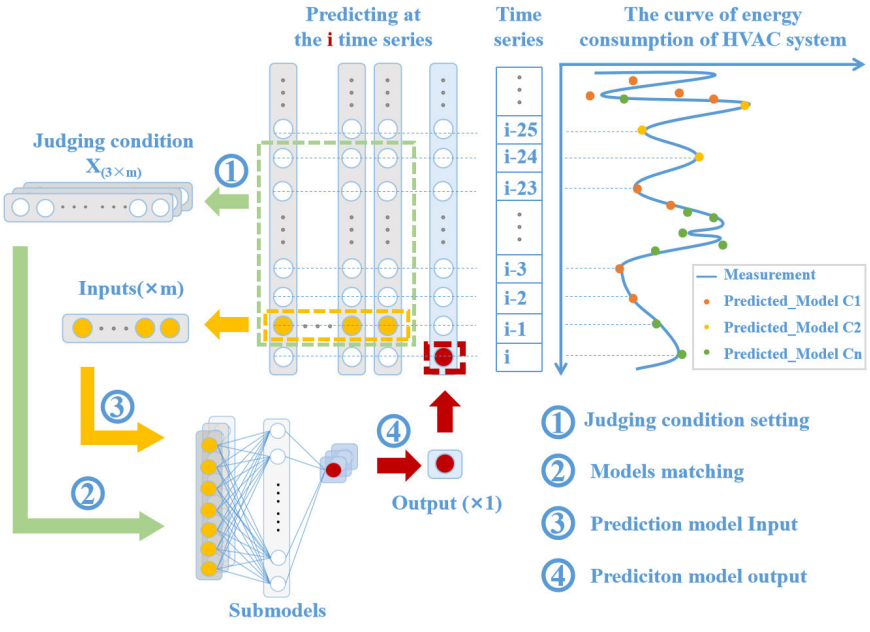


Fig. 7 The process of dynamic prediction

consumption at time i, the maximum, minimum, and average values are extracted from the historical data of time steps $i - 24$ to $i - 1$ to form elements of the judgment condition vector X. If there are m input parameters, the number of elements in X is $3m$. As illustrated in Figure 9, by calculating the Euclidean distance (d) between X and the cluster centers Y of all training subsets, the best model prediction is selected. For example, when the distance between X and Y_2 is the smallest, Model C2 is the optimal model.

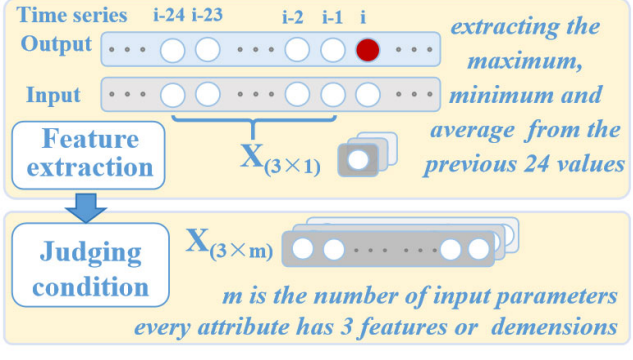


Fig. 8 The process of judging condition generation in dynamic prediction

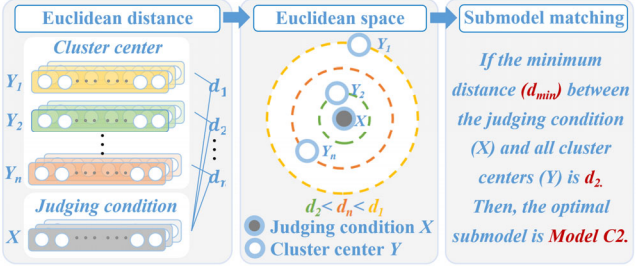


Fig. 9 The method of model matching in dynamic prediction

2.5 Parameters setting of PSO-GRU model

This paper referenced the range for the hyperparameter optimization of deep learning models from (Luo and Oyedele 2021), and combined with the process of manual tuning to determine the key hyperparameters affecting the predictive performance of the GRU model: GRU units, dropout rate, and initial learning rate. The optimization range for the model parameters and the structural settings are shown in Table 1 and Table 2, respectively.

Table 1 Decision variables of PSO for GRU models

Parameter	Range
Number of GRU units	20–120
Dropout rate	0.01–0.5
Initial learning rate	0.001–0.1
Learning rate drop factor	0.1-0.5

Table 2 PSO parameters

Parameter	Value
Initial population size	20
Maximum iteration	5
Particle dimension	4
Inertia weight	0.4–0.9
Learning rate C1	2
Learning rate C2	2

2.6 Evaluation metric of prediction performance

The evaluation metric of prediction accuracy of the models included the root mean square error (RMSE), coefficient of variation (CV), R^2, and mean MAPE. Higher prediction accuracy was indicated by smaller values of RMSE, CV, MAPE, and larger values of R^2. The formulas of metrics are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{t}\left(y_{1,j} - y_{2,j}\right)^2}{t}} \tag{3}$$

$$\mathrm{CV} = \frac{1}{y_1}\mathrm{RMSE} \times 100\% \tag{4}$$

$$R^2 = 1 - \frac{\sum_{j=1}^{t}\left(y_{1,j} - y_{2,j}\right)^2}{\sum_{j=1}^{t}\left(y_{1,j} - \overline{y_1}\right)^2} \tag{5}$$

$$\mathrm{MAPE} = \frac{1}{t}\sum_{j=1}^{t}\left|\frac{y_{1,j} - y_{2,j}}{y_{1,j}}\right| \times 100\% \tag{6}$$

where, $y_{1,j}$ is the measured value, $\overline{y_1}$ is the average of measured value, $y_{2,j}$ is the predicted value.

3 Case study

3.1 Building information

The case study was a commercial office building with heating, ventilation, and air conditioning (HVAC) systems located in Tianjin, China (Figure 10). This building was divided into six floors with a total area almost 5,232 m^2. The annual energy consumption of this building was 46.1 kWh/(m^2·a). Building energy consumption consists mainly of HVAC systems, plugs, and lighting. The energy consumption of HVAC systems accounted for 40% of a building's annual energy consumption, whereas plug and lighting energy consumption accounted for 28% and 14%, respectively.

(a) Outdoor

(b) Indoor

Fig. 10 The view of the office building

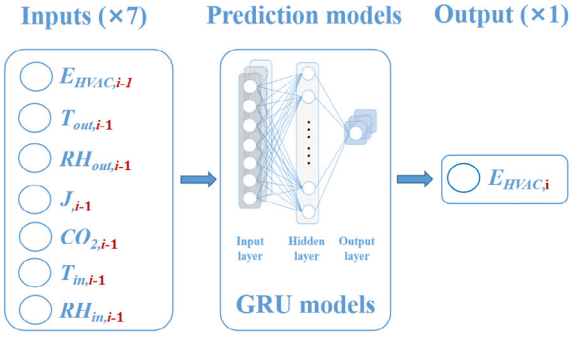


Fig. 11 The inputs and output of prediction models

3.2 Data set and the input variables of model

As show in Figure 11, the output of prediction model is HVAC systems energy consumption ($E_{\mathrm{HVAC},i}$), and the inputs, 7 variables, are the values of historical HVAC systems energy consumption ($E_{\mathrm{HVAC},i-1}$), outdoor dry bulb temperature ($T_{\mathrm{out},i-1}$), outdoor relative humidity ($\mathrm{RH}_{\mathrm{out},i-1}$), solar radiation intensity (J_{i-1}), indoor CO_2 concentration ($CO_{2,i-1}$), indoor dry bulb temperature ($T_{\mathrm{in},i-1}$), and indoor relative humidity ($\mathrm{RH}_{\mathrm{in},i-1}$) before the i time series.

The data collection period was from September 1 to August 30 of next year. The overall data size is 365 d, and the temporal resolution is 1 h. It is ensuring that testing set and training-validation set comprised time series groups with similar features, trends, and seasonal patterns (Kohli et al. 2022). The testing set consists of data from the last six days of each month (approximately accounted 20% by the overall data). The statistics features of main parameters are presented in Table 3.

3.3 The attribute and dimension of clustering vector

As shown in Table 4, two cases are set for studying the performance difference of the SFEDN method to cluster multiple attributes and single multiple attributes. In Case 1,

Table 3 The statistics features of main parameters

		Training-validation set			Testing set		
Parameter	Unit	Max	Min	Average	Max	Min	Average
T_{out}	°C	35.6	−16.2	13.8	36.1	−9.8	13.2
RH_{out}	%	99%	7%	58%	98%	7%	56%
J	W/m^2	865	0	154	866	0	164
T_{in}	°C	27.6	18.4	23.7	27.7	16.7	23.7
RH_{in}	%	84%	7%	39%	83%	12%	36%
CO_2	ppm	1706	814	1009	1599	832	1017
E	kW·h	161.5	7.2	41.2	157.6	11.0	42.8

Table 4 The attribute and dimension of cluster vector

		Dimension of cluster vector	
Case	Attribute of cluster vector	Conventional	SFEDN
1	T_{out}, RH_{out}, J, T_{in}, RH_{in}, CO_2	24 × 6	3 × 6
2	E	24	3

there are six attributes involving clustering of T_{out}, RH_{out}, J, T_{in}, RH_{in}, and CO_2. In Case 2, it is one attribute and involves clustering of E.

4 Results

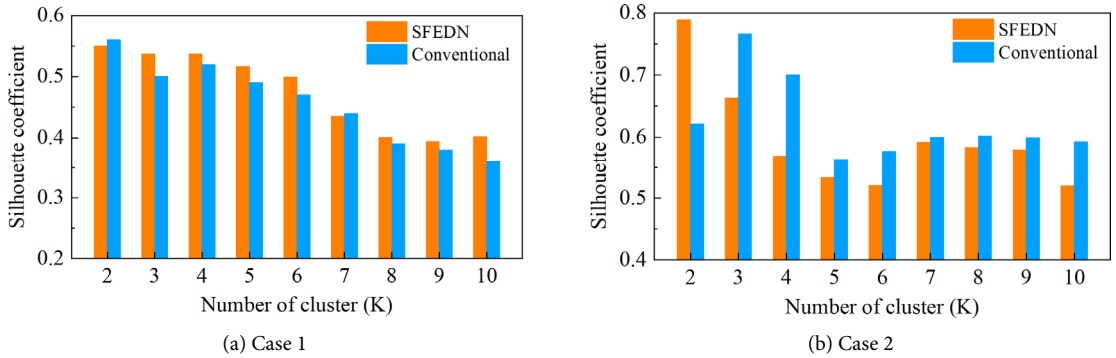
In Section 4.1, for multi-attribute clustering, both the SFEDN method proposed in this paper and the conventional method determined the optimal number of clusters to be 2 based on the silhouette coefficient. In Section 4.2, an analysis was conducted on the differences in the characteristics of Pearson correlation coefficients among the subsets after clustering. In Section 4.3, the predicted results (MAPE) of each subset after k-fold cross-validation were presented to assess the robustness and stability of the model. In Section 4.4, a comparison was made between the performance indicators of the multi-modal dynamic prediction method using SFEDN clustering modeling and the single-model static prediction method on the test set. The results showed that dynamic prediction using SFEDN outperformed static prediction in Case 1, both RMSE and CV are reduced by 11.9%, R^2 reaches 0.89, and MAPE is reduced by 19.9%. The dynamic prediction using conventional clustering reduced RMSE and CV by 7.8%, reaching an R^2 of 0.879 and reducing MAPE by 10.6% compared to static prediction. Meanwhile, the dynamic prediction performance demonstrated that the SFEDN clustering method surpassed conventional clustering method, and multi-attribute clustering modeling outperformed single-attribute modeling.

4.1 The number of clusters

As shown in Figure 12, the optimal number of clusters between the SFEDN method and the conventional method. According to the principle that the larger the value of the silhouette coefficient, the better the clustering is (Li et al. 2020; Sala et al. 2021). In Case 1, when K is 2, the clustering effect of SFEDN method and conventional method are best. In Case 2, the SFEDN clustering method has an optimal K of 2, while the conventional method has an optimal K of 3. In order to expand the comparison, we explored the case where the number of clusters in Case 1-Case 2 is 2 and 3. The size of data subsets and the mapping cluster distribution on calendar were shown in the Appendix.

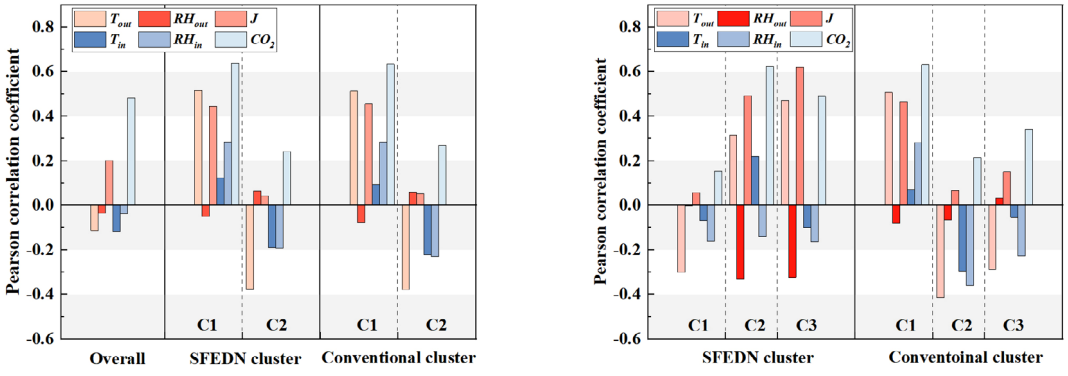
4.2 The Pearson correlation coefficient in subsets

As shown in Figure 13 and Figure 14, the Pearson correlation coefficient before and after clustering in the training-validation set has undergone significant changes, indicating that the dataset has been decomposed into multiple feature-significant subsets through clustering. As shown in Figure 13(a), compared to the overall set, the variables with correlation coefficients exceeding 0.4 with the HVAC system energy consumption in subset C1 are 3 (CO_2, T_{out}, and J), while the correlation coefficients with the HVAC system energy consumption in subset C2 do not exceed 0.4, but the correlation coefficients of T_{out}, T_{in}, and



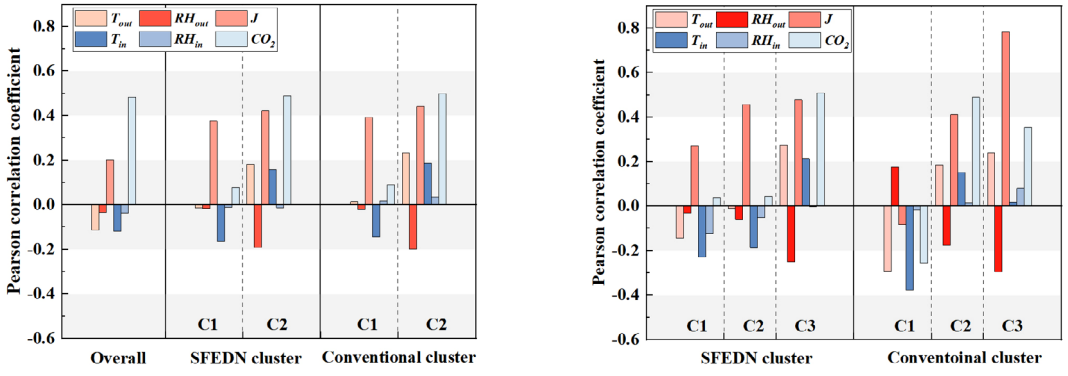
(a) Case 1 (b) Case 2

Fig. 12 The determination of cluster number with silhouette coefficient method



(a) $K = 2$ (b) $K = 3$

Fig. 13 The Pearson correlation coefficient after clustering (Case 1)



(a) $K = 2$ (b) $K = 3$

Fig. 14 The Pearson correlation coefficient after clustering (Case 2)

RH_{in} were improved compared to the overall set. According to Figure 13(b), when clustering into 3 subsets, the subset C1 advocated by SFEDN method has significantly weaker correlation compared to the other two subsets, while the subset C3 of conventional clustering also has weaker significance compared to the other two subsets. The comparison of subset features based on single attribute (HVAC system energy consumption) clustering shown in Figure 14(a). The subset C2 of both clustering methods has two variables with a correlation coefficient of more than 0.4 with HVAC system energy consumption (CO_2 and J), while the parameters in both subsets C1 do not reach 0.4 or fall below −0.4. The features of subset C2 are significantly better than those of subset C1. As shown in Figure 14(b), the features of subset C3 of both clustering methods are clearly superior to other subsets. The quality of energy consumption feature parameters directly affects the prediction performance of the model.

4.3 The prediction results in training set with *k*-fold cross validation

According to the model's training process with *k*-fold cross-validation as shown in Figure 6, we used the minimum of MAPE obtained from *k*-fold cross-validation to assess the robustness and stability of the model. Furthermore, a sub-model with good robustness and stability in the k subset will be selected for dynamic prediction. The prediction results in the training set with *k*-fold cross-validation are presented from Figure 15 to Figure 18. For example, in the multi-attribute clustering using the SFEDN method, the best sub-model for subset C1 generated through 5-fold cross-validation is Model 2 (Figure 15(a1)).

4.4 The performance of dynamic prediction

As shown in Figure 19, the comparison between the dynamic prediction using clustering-based modeling and the static prediction using a single model was conducted. Additionally, the dynamic predictions based on the SFEDN clustering method were compared with those using conventional clustering. The dynamic prediction performance demonstrated that the SFEDN clustering method surpassed conventional clustering method, and multi-attribute clustering modeling outperformed single-attribute modeling.

When cluster number is 2, the performance metrics showed that the dynamic prediction of multiple-attribute

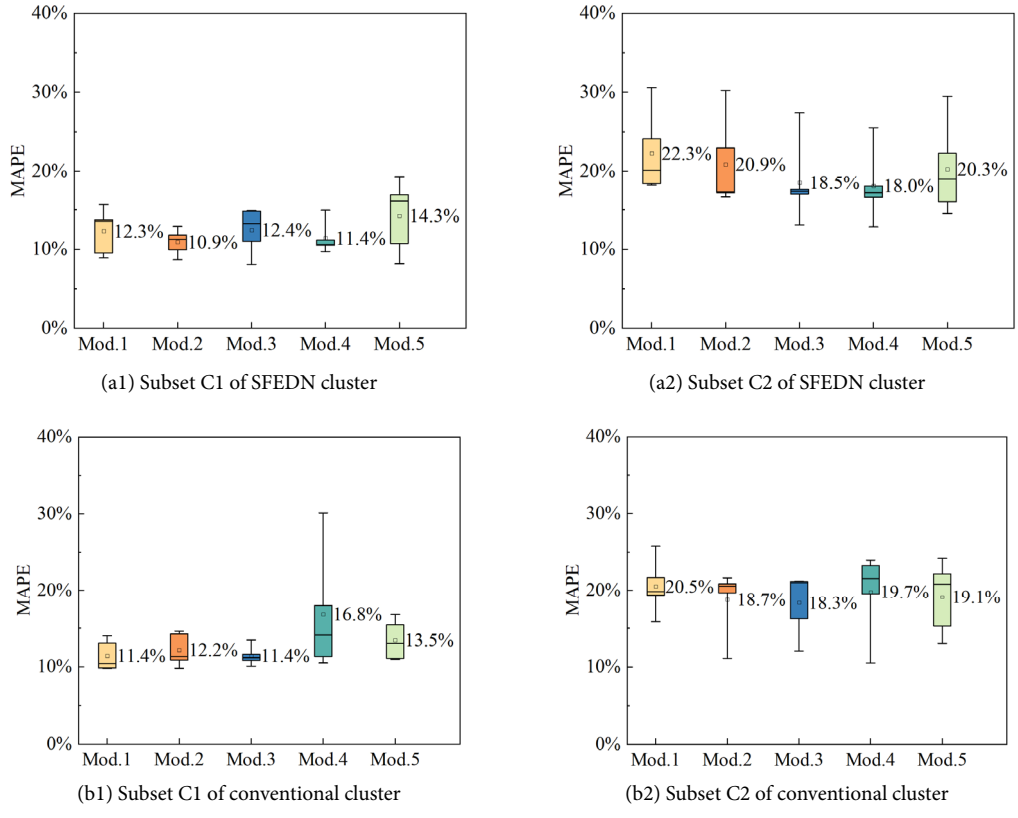


Fig. 15 The MAPE of submodels with *k*-fold cross validation in training set (Case 1, K =2)

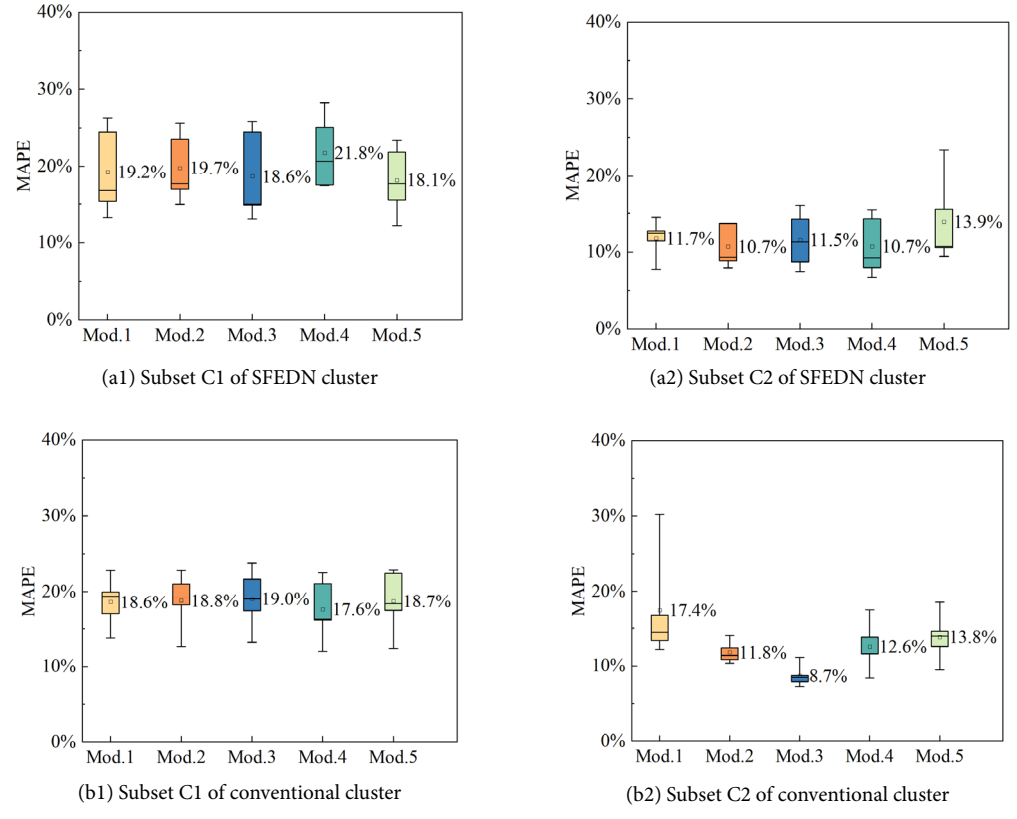
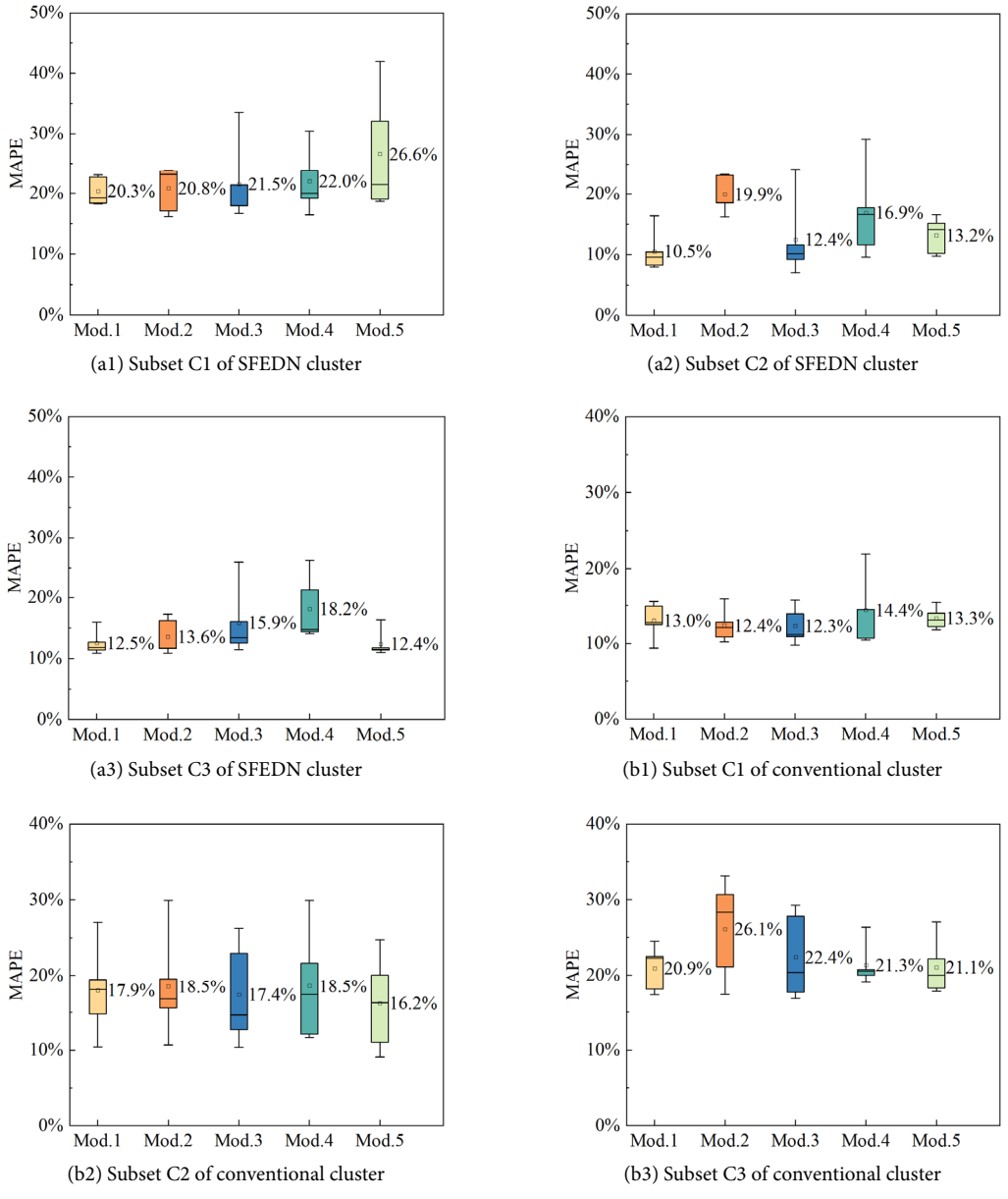


Fig. 16 The MAPE of submodels with *k*-fold cross validation in training set (Case 2, K = 2)

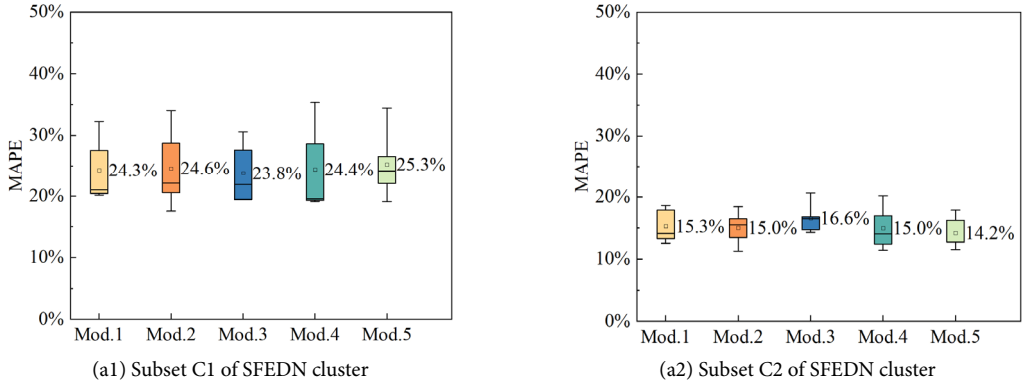


(a1) Subset C1 of SFEDN cluster (a2) Subset C2 of SFEDN cluster

(a3) Subset C3 of SFEDN cluster (b1) Subset C1 of conventional cluster

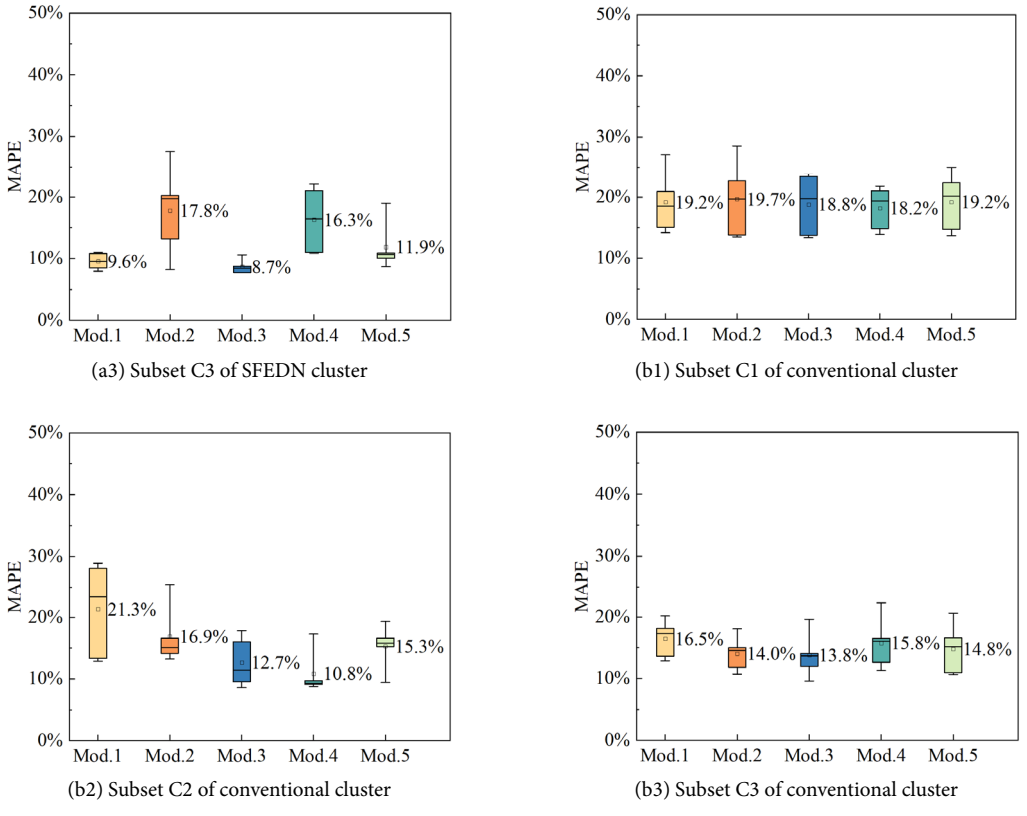
(b2) Subset C2 of conventional cluster (b3) Subset C3 of conventional cluster

Fig. 17 The MAPE of submodels with *k*-fold cross validation in training set (Case 1, $K = 3$)



(a1) Subset C1 of SFEDN cluster (a2) Subset C2 of SFEDN cluster

Fig. 18 The MAPE of submodels with *k*-fold cross validation in training set (Case 2, $K = 3$)



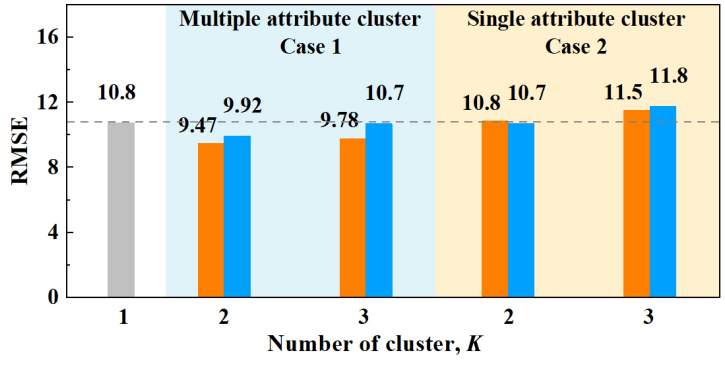
(a3) Subset C3 of SFEDN cluster

(b1) Subset C1 of conventional cluster

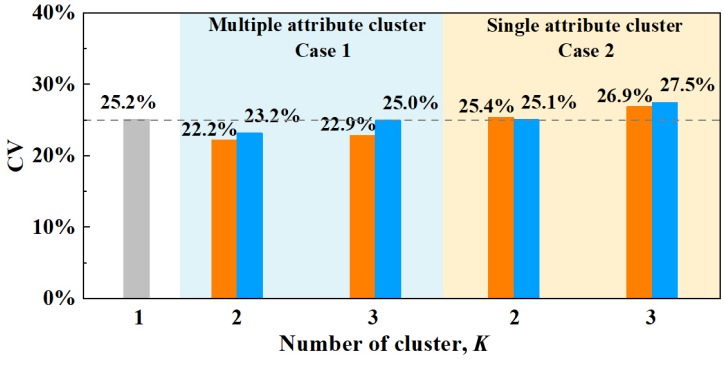
(b2) Subset C2 of conventional cluster

(b3) Subset C3 of conventional cluster

Fig. 18 The MAPE of submodels with *k*-fold cross validation in training set (Case 2, $K = 3$) (Continued)

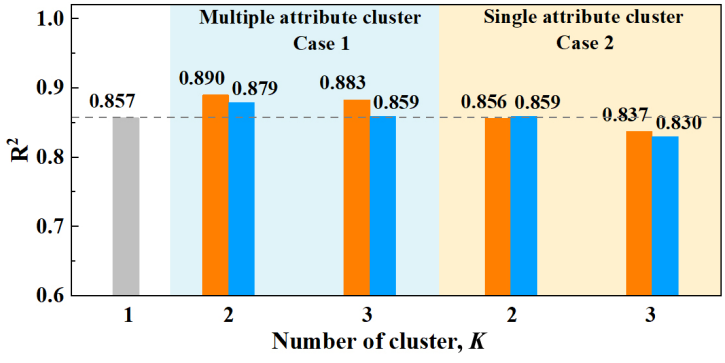


(a) RMSE

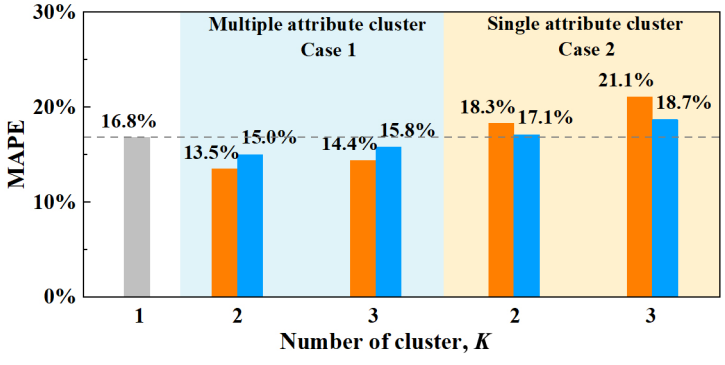


(b) CV

Fig. 19 The prediction performance comparison in testing set



(c) R^2



(d) MAPE

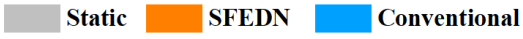


Fig. 19 The prediction performance comparison in testing set (Continued)

clustering using the SFEDN method, compared to static prediction, reduced RMSE and CV by 11.9%, achieved an R^2 of 0.890 and reduced MAPE by 19.9%. Compared to the dynamic prediction of single-attribute clustering modeling using the SFEDN method, both RMSE and CV decreased by 12.6%, R^2 increased by 4.0%, and MAPE decreased by 26.3%.

When cluster number is 3, the dynamic prediction of multiple-attribute clustering using SFEDN method reduced RMSE and CV by 9.1%, achieving an R^2 of 0.883 and reducing MAPE by 14.3% compared to static prediction. On the other hand, the dynamic prediction using conventional clustering reduced RMSE and CV by 0.5%, with an R^2 of 0.859 and a reduction of MAPE by 6.0% compared to static prediction. However, the performance of the dynamic prediction of single attribute (HVAC system energy consumption) clustering was not as good as static prediction.

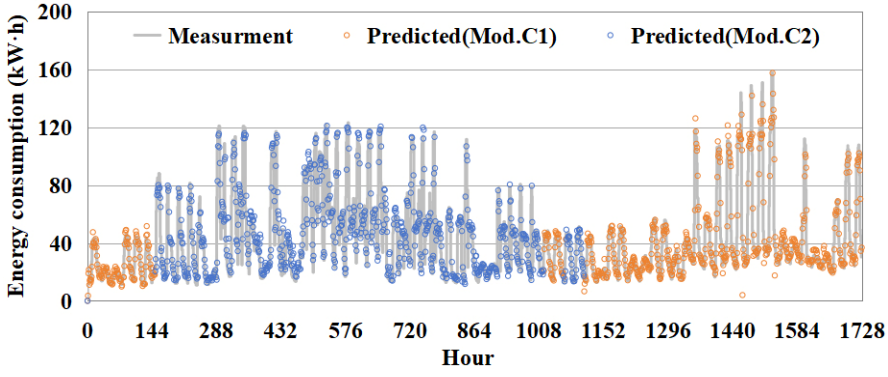
In summary, in the optimal number of cluster (K = 2), the dynamic prediction performance demonstrated that the SFEDN clustering method surpassed conventional clustering method, and multi-attribute clustering modeling outperformed single-attribute modeling. The real curves of dynamic prediction in testing sets were shown in the Figure 20 and Figure 21.

5 Discussion, limitations and future direction

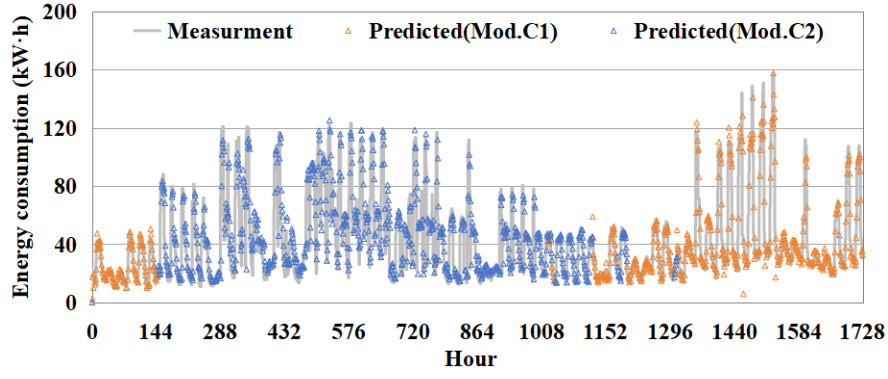
In this paper, we integrated meteorological data, indoor environmental data, and historical energy consumption data to construct multi-attribute and high-dimensional vectors for clustering. We employed a method of statistical feature extraction to reduce dimensions while preserving the original multi-attribute information. Subsequently, based on the principles of Euclidean distance clustering, the normalization was performed according to vector dimension instead of the conventional method based on attribute. This approach aimed to ensure the retention of original information while mitigating the impact of data attribute differences and correlations on clustering effects.

To enhance the training adaptability of the deep learning model, we employed the PSO to optimize the key parameters of the GRU model. Additionally, k-fold cross-validation was introduced to determine the sub-model architecture with the best prediction performance. Finally, to achieve dynamic predictions, we extracted historical 24-hour feature data to assess the uncertainty of future conditions and matched the best sub-model for dynamic prediction using multiple models.

The performance indicators of the dynamic prediction (Figure 19) showed that the dynamic prediction using the

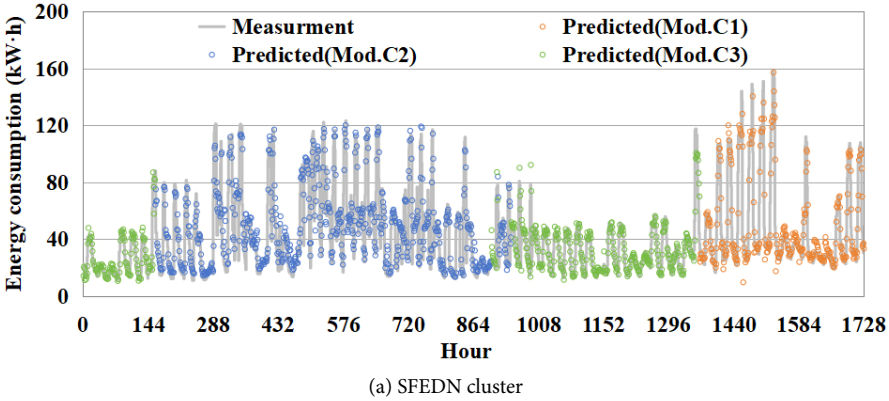


(a) SFEDN cluster

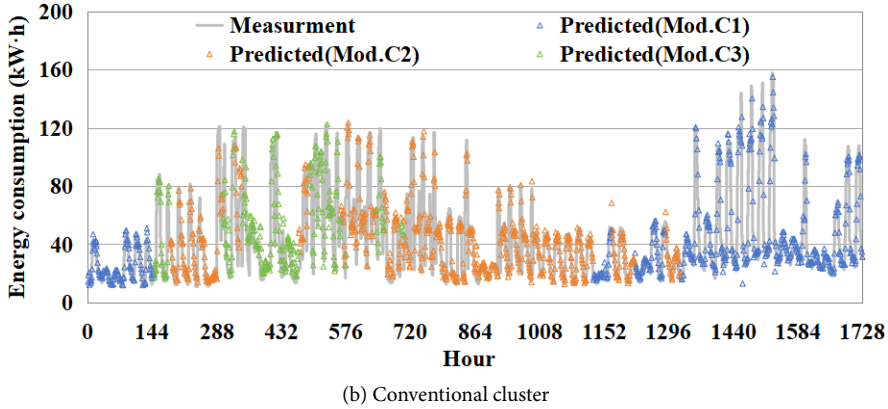


(b) Conventional cluster

Fig. 20 The real curve of dynamic prediction in testing set (Case 1, $K = 2$)



(a) SFEDN cluster



(b) Conventional cluster

Fig. 21 The real curve of dynamic prediction in testing set (Case 1, $K = 3$)

SFEDN clustering method reduced RMSE and CV by 11.9%, achieving an R^2 of 0.890 and reducing MAPE by 19.9% compared to static prediction. While the dynamic prediction performance based on multi-attribute clustering modeling (Case 1) exhibits robust predictive capabilities, the dynamic prediction based on single-attribute clustering modeling falls short of static prediction. However, to improve the performance of prediction models, these are some issues that should be discussed:

1) The configuration of attributes and dimensions in the cluster matrix

This article delved into the impact of multi-attribute and single-attribute clustering scenarios on prediction outcomes. The results indicated that clustering analysis effectively decomposed the data, generating subsets with significant characteristic differences, thereby providing robust training datasets for sub-models adaptable to various feature conditions. Therefore, relying solely on energy consumption clustering is insufficient for achieving optimal clustering effects, and the incorporation of relevant data for multi-attribute clustering is deemed essential (Bourdeau et al. 2021). Future research should further refine the prediction of heating, cooling, lighting, and socket energy consumption. Determining the attributes and dimensions of the clustering vector for maximizing accurate prediction is crucial. Constraints extend beyond the type of predicted energy, encompassing correlations between model inputs and outputs, input intercorrelations, and algorithmic mechanisms, all of which may influence clustering vector construction.

2) The importance of the testing set in multi-model predictions.

In this paper, the testing set is treated as an independent set, simulating a virtual dataset from an uncertain future, and is excluded from clustering to prevent information leakage. The objective is to ensure that prediction conditions for the testing set remain unknown during model training, preventing any potential bias. Our approach involves assessing these conditions before inputting them into the models during forecasting.

Evaluating the uncertainty of prediction conditions involves calculating the Euclidean distance between the statistical features of the past 24 hours' data and the clustering centers of each training subset. We focused on the matching prediction trained prediction models to match the prediction conditions instead of matching training subset with testing set. The optimal subset model is then matched for dynamic prediction through multi-model blending.

The model matching mechanism, relying on statistical features of the past 24 hours' data to prevent information leakage, enhances robustness and provides safeguards multi-modal prediction. However, dependence on historical information entails risks, and unforeseen situations may arise in future conditions. Therefore, accurate historical information and a stable energy system operation status are prerequisites for matching the optimal subset model accurately. Robustness can be further improved through measures such as enhancing online prediction frameworks, integrating online feedback attention mechanisms, self-learning integration, and reinforcement learning integration.

In summary, this paper successfully implements multi-model dynamic prediction based on multi-attribute feature extraction and clustering modeling, yielding commendable prediction results. Nevertheless, there remains a gap in defining the optimal attribute combination and dimension settings for clustering vectors associated with predicted energy consumption types. The mechanism of judging future conditions based on historical information has inherent limitations. To enhance the implementation of the proposed dynamic prediction framework, robust system operation and effective historical data are indispensable. Furthermore, when employing more advanced prediction models such as transfer learning, we can explore the generalization ability of model predictions on training and testing sets with entirely different distributions (Xu et al. 2024) and compare these results to those obtained from training and testing sets with the same distribution. This approach may yield new and valuable insights for forecasting energy consumption in increasingly complex future energy systems.

6 Conclusion

This paper integrated meteorological parameters and indoor environmental parameters to construct multi-attribute high-dimensional clustering vectors, and the clustering effect of the *K*-means algorithm was enhanced by the SFEDN method. Within the framework of cluster modeling, PSO was employed to implement adaptive training and optimization of the GRU deep learning model across different feature subsets. The optimal number of clusters was determined using the silhouette coefficient method, and *k*-fold cross-validation was introduced to assess the robustness and stability of the submodels. Simultaneously, in the testing set with completely unknown prediction conditions, ensuring no information leakage and suitability for training-validation sets, the optimal sub-model was determined by calculating the statistical features of the past 24 hours' data and the Euclidean distance between the clustering centers of each training subset. The dynamic prediction model based on multi-attribute clustering with SFEDN achieved a 11.9% reduction in both RMSE and CV,

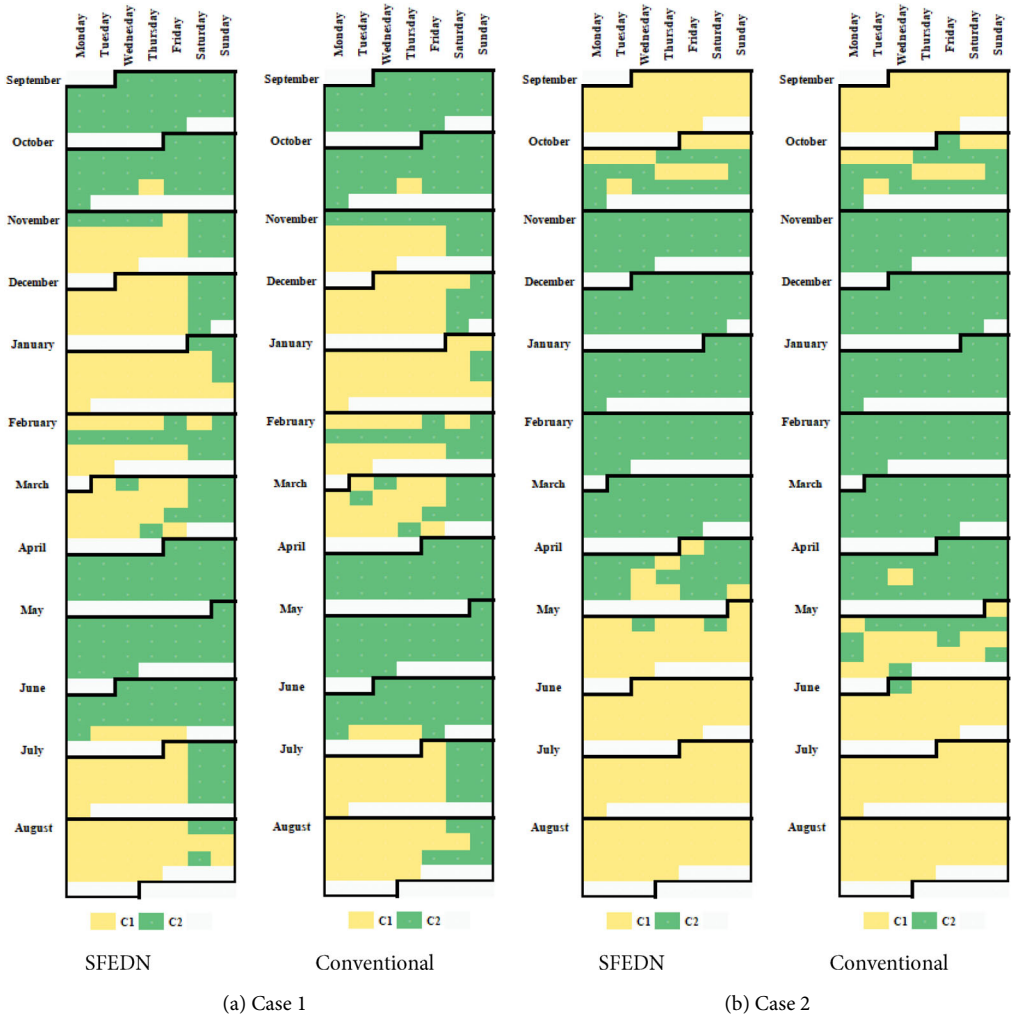
a 3.8% increase in R^2, and a 19.9% reduction in MAPE compared to static prediction. When compared to dynamic prediction based on single-attribute clustering modeling, both RMSE and CV decrease by 12.6%, R^2 increases by 4.0%, and MAPE decreases by 26.3%. In summary, the dynamic prediction performance demonstrated that the SFEDN clustering method surpasses conventional clustering method, and multi-attribute clustering modeling outperforms single-attribute modeling in HVAC systems. Lastly, when integrating clustering and prediction models, further discussion is needed to determine the optimal number of attributes and dimensions for clustering. Further research is needed on how to better pair the training and testing sets, or how to use a well-trained model.

Appendix

Table A The size of the dataset (in days) corresponding to each cluster

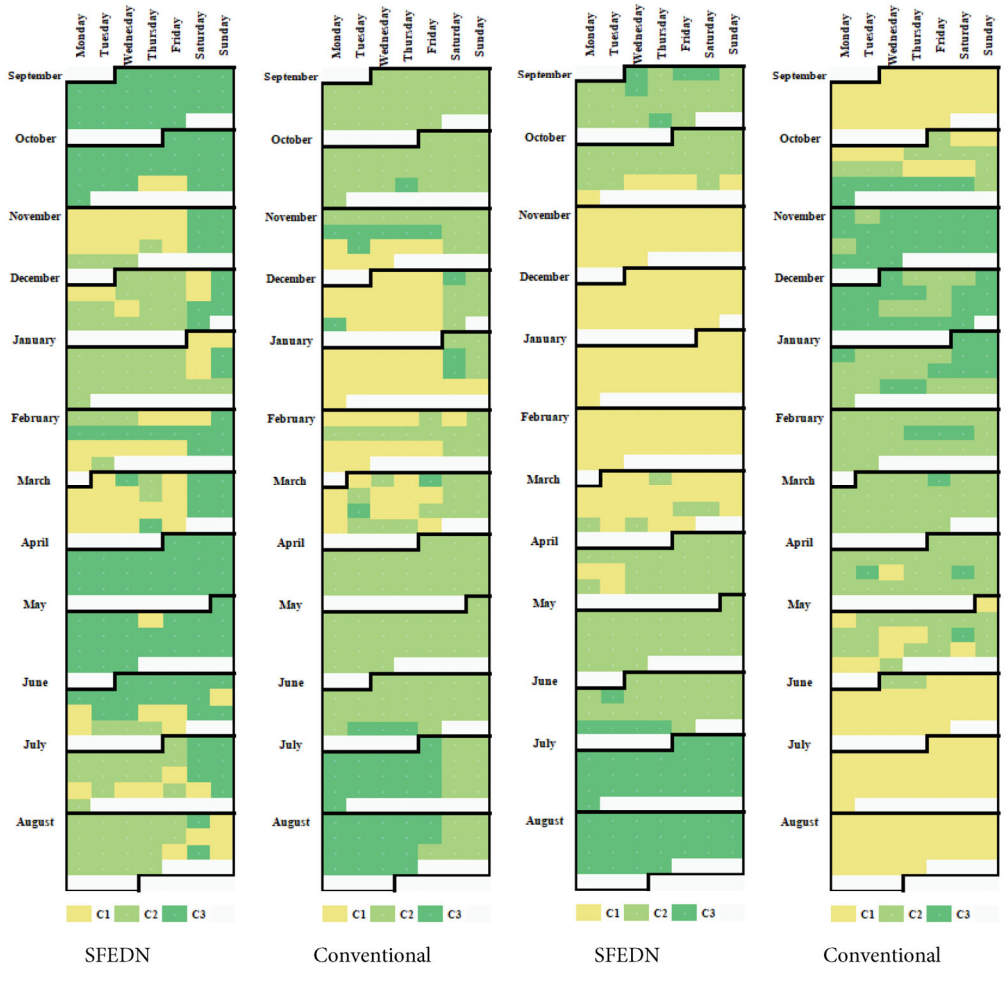
Cluster method	Case	$K = 2$		$K = 3$		
		C1	C2	C1	C2	C3
SFEDN	1	137	156	124	109	60
	2	124	169	66	77	150
Conventional	1	121	172	113	116	64
	2	121	172	65	177	51*

*Note: the minimum dataset comprises 51 days (1224 hours/points) of data.



(a) Case 1 (b) Case 2

Fig. A1 The mapping cluster distribution on calendar ($K = 2$)



(a) Case 1 (b) Case 2

Fig. A2 The mapping cluster distribution on calendar ($K = 3$)

Acknowledgements

This study is supported by the National Natural Science Foundation of China (No. 52108074) and the National Natural Science Foundation of China (No. 52078144).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article. Yu Huang is a Subject Editor of *Building Simulation*.

Author contribution statement

All authors contributed to the study conception and design. Huiheng Liu: conceptualization; formal analysis; investigation; methodology; writing—original draft, review & editing. Yanchen Liu: conceptualization; formal analysis; investigation; visualization; review & editing. Huakun Huang: investigation; writing—original draft, review & editing. Huijun Wu: writing—review & editing. Yu Huang: writing—review & editing. All authors read and approved the final manuscript.

References

Abdallah M, Abu Talib M, Hosny M, et al. (2022). Forecasting highly fluctuating electricity load using machine learning models based on multimillion observations. *Advanced Engineering Informatics*, 53: 101707.

Afram A, Janabi-Sharifi F (2014). Review of modeling methods for HVAC systems. *Applied Thermal Engineering*, 67: 507–519.

Arbelaitz O, Gurrutxaga I, Muguerza J, et al. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46: 243–256.

Atalay SD, Calis G, Kus G, et al. (2019). Performance analyses of statistical approaches for modeling electricity consumption of a commercial building in France. *Energy and Buildings*, 195: 82–92.

Boithias F, El Mankibi M, Michel P (2012). Genetic algorithms based optimization of artificial neural network architecture for buildings' indoor discomfort and energy consumption prediction. *Building Simulation*, 5: 95–106.

Bourdeau M, Basset P, Beauchêne S, et al. (2021). Classification of daily electric load profiles of non-residential buildings. *Energy and Buildings*, 233: 110670.

Chen Y, Tan H, Berardi U (2017). Day-ahead prediction of hourly electric demand in non-stationary operated commercial buildings: a clustering-based hybrid approach. *Energy and Buildings*, 148: 228–237.

Chen S, Wang LL, Li J, et al. (2022). A training pattern recognition algorithm based on weight clustering for improving cooling load prediction accuracy of HVAC system. *Journal of Building Engineering*, 52: 104445.

Chinese Society for Urban Studies (2022). 2021 Annual Development Research Report on Building Energy Efficiency in China. Beijing: China Architecture & Building Press. (in Chinese)

Chitalia G, Pipattanasomporn M, Garg V, et al. (2020). Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. *Applied Energy*, 278: 115410.

Cholewa T, Siuta-Olcha A, Smolarz A, et al. (2022). An easy and widely applicable forecast control for heating systems in existing and new buildings: first field experiences. *Journal of Cleaner Production*, 352: 131605.

Chung J, Gulcehre C, Cho K, et al. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv: 1412.3555.

Czétány L, Vámos V, Horváth M, et al. (2021). Development of electricity consumption profiles of residential buildings based on smart meter data clustering. *Energy and Buildings*, 252: 111376.

Elnour M, Himeur Y, Fadli F, et al. (2022). Neural network-based model predictive control system for optimizing building automation and management systems of sports facilities. *Applied Energy*, 318: 119153.

Erfani A, Jafarinejad T, Roels S, et al. (2024). In search of optimal building behavior models for model predictive control in the context of flexibility. *Building Simulation*, 17: 71–91.

Fan C, Xiao F, Zhao Y (2017). A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy*, 195: 222–233.

Fan C, Yan D, Xiao F, et al. (2021). Advanced data analytics for enhancing building performances: from data-driven to big data-driven approaches. *Building Simulation*, 14: 3–24.

Fan G, Zheng Y, Gao W, et al. (2023). Forecasting residential electricity consumption using the novel hybrid model. *Energy and Buildings*, 290: 113085.

Gad AG (2022). Particle swarm optimization algorithm and its applications: a systematic review. *Archives of Computational Methods in Engineering*, 29: 2531–2561.

Global Alliance for Buildings and Construction (2020). 2020 Global Status Report for Buildings and Construction. Paris: Global Alliance for Buildings and Construction.

Hsu D (2015). Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. *Applied Energy*, 160: 153–163.

Jota PRS, Silva VRB, Jota FG (2011). Building load management using cluster and statistical analyses. *International Journal of Electrical Power & Energy Systems*, 33: 1498–1505.

Kang X, An J, Yan D (2023). A systematic review of building electricity use profile models. *Energy and Buildings*, 281: 112753.

Kim TY, Cho SB (2019). Particle swarm optimization-based CNN-LSTM networks for forecasting energy consumption. In: Proceedings of 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand.

Kim J, Kim KI (2020). Data-driven hybrid model and operating algorithm to shave peak demand costs of building electricity. *Energy and Buildings*, 229: 110493.

Kim DD, Suh HS (2021). Heating and cooling energy consumption prediction model for high-rise apartment buildings considering design parameters. *Energy for Sustainable Development*, 61: 1–14.

Ko JH, Kong DS, Huh JH (2017). Baseline building energy modeling of cluster inverse model by using daily energy consumption in office buildings. *Energy and Buildings*, 140: 317–323.

Kohli GS, Kaur P, Singh A, et al. (2022). TransLearn: a clustering based knowledge transfer strategy for improved time series forecasting. *Knowledge-Based Systems*, 249: 108889.

Li Z, Dai J, Chen H, et al. (2019). An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage. *Building Simulation*, 12: 665–681.

Li P, Pye S, Keppo I (2020). Using clustering algorithms to characterise uncertain long-term decarbonisation pathways. *Applied Energy*, 268: 114947.

Li Y, Liu C, Zhang L, et al. (2021a). A partition optimization design method for a regional integrated energy system based on a clustering algorithm. *Energy*, 219: 119562.

Li K, Tian J, Xue W, et al. (2021b). Short-term electricity consumption prediction for buildings using data-driven swarm intelligence based ensemble model. *Energy and Buildings*, 231: 110558.

Li Z, Ye H, Liao N, et al. (2022). Impact of COVID-19 on electricity energy consumption: a quantitative analysis on electricity. *International Journal of Electrical Power & Energy Systems*, 140: 108084.

Li F, Wan Z, Koch T, et al. (2023). Improving the accuracy of multi-step prediction of building energy consumption based on EEMD-PSO-Informer and long-time series. *Computers and Electrical Engineering*, 110: 108845.

Li G, Wu Y, Yan C, et al. (2024). An improved transfer learning strategy for short-term cross-building energy prediction using data incremental. *Building Simulation*, 17: 165–183.

Liu H, Liu Y, Guo X, et al. (2023). An energy consumption prediction method for HVAC systems using energy storage based on time series shifting and deep learning. *Energy and Buildings*, 298: 113508.

Luo XJ, Oyedele LO, Ajayi AO, et al. (2020). Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings. *Renewable and Sustainable Energy Reviews*, 131: 109980.

Luo XJ, Oyedele LO (2021). Forecasting building energy consumption: adaptive long-short term memory neural networks driven by genetic algorithm. *Advanced Engineering Informatics*, 50: 101357.

Mohammed A, Alshibani A, Alshamrani O, et al. (2021). A regression-based model for estimating the energy consumption of school facilities in Saudi Arabia. *Energy and Buildings*, 237: 110809.

Pawar P, TarunKumar M, Vittal KP (2020). An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation. *Measurement*, 152: 107187.

Peng Y, Liu H, Li X, et al. (2020). Machine learning method for energy consumption prediction of ships in port considering green ports. *Journal of Cleaner Production*, 264: 121564.

Risbeck MJ, Bazant MZ, Jiang Z, et al. (2021). Modeling and multiobjective optimization of indoor airborne disease transmission risk and associated energy consumption for building HVAC systems. *Energy and Buildings*, 253: 111497.

Ruiz LGB, Pegalajar MC, Arcucci R, et al. (2020). A time-series clustering methodology for knowledge extraction in energy consumption data. *Expert Systems with Applications*, 160: 113731.

Sala J, Li R, Christensen MH (2021). Clustering and classification of energy meter data: a comparison analysis of data from individual homes and the aggregated data from multiple homes. *Building Simulation*, 14: 103–117.

Sala-Cardoso E, Delgado-Prieto M, Kampouropoulos K, et al. (2018). Activity-aware HVAC power demand forecasting. *Energy and Buildings*, 170: 15–24.

Seo J, Kim S, Lee S, et al. (2022). Data-driven approach to predicting the energy performance of residential buildings using minimal input data. *Building and Environment*, 214: 108911.

Shen Y, Hu Y, Cheng K, et al. (2024). Utilizing interpretable stacking ensemble learning and NSGA-III for the prediction and optimisation of building photo-thermal environment and energy consumption. *Building Simulation*, 17: 819–838.

Somu N, Raman GMR, Ramamritham K (2021). A deep learning framework for building energy consumption forecast. *Renewable and Sustainable Energy Reviews*, 137: 110591.

Vijayalakshmi K, Vijayakumar K, Nandhakumar K (2023). An ensemble learning model for estimating the virtual energy storage capacity of aggregated air-conditioners. *Journal of Energy Storage*, 59: 106512.

Wei Y, Zhang X, Shi Y, et al. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82: 1027–1047.

Xiao F, Fan C (2014). Data mining in building automation system for improving building operational performance. *Energy and Buildings*, 75: 109–118.

Xu H, Chen Y, Zhang D (2024). Worth of prior knowledge for enhancing deep learning. *Nexus*, 1: 100003.

Zhang Q, Tian Z, Ding Y, et al. (2019). Development and evaluation of cooling load prediction models for a factory workshop. *Journal of Cleaner Production*, 230: 622–633.

Zhang C, Li J, Zhao Y, et al. (2020). A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process. *Energy and Buildings*, 225: 110301.

Zhou G, Moayedi H, Bahiraei M, et al. (2020). Employing artificial bee colony and particle swarm techniques for optimizing a neural network in prediction of heating and cooling loads of residential buildings. *Journal of Cleaner Production*, 254: 120082.