

Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings

X.J. Luo¹, Lukumon O. Oyedele^{1,*}, Anuoluwapo O. Ajayi¹, Olugbenga O. Akinade¹, Hakeem A. Owolabi¹ and Ashraf Ahmed²

¹Big Data Enterprise and Artificial Intelligence Laboratory (Big-DEAL)
University of the West of England (UWE), Frenchay Campus, Bristol, United Kingdom

²Department of Civil and Environmental Engineering
Brunel University London, London, United Kingdom

*Corresponding author: L.Oyedele@uwe.ac.uk

Abstract

Accurate forecast of energy consumption is essential in building energy management. Owing to the variation of outdoor weather condition among different seasons, year-round historical weather profile is needed to investigate its feature thoroughly. Daily weather profiles in the historical database contain various features, while different architecture of deep neural network (DNN) models may be identified suitable for specific featuring training datasets. In this study, an integrated artificial intelligence-based approach, consisting of feature extraction, evolutionary optimization and adaptive DNN model, is proposed to forecast week-ahead hourly building energy consumption. The DNN is the fundamental forecasting engine of the proposed model. Feature extraction of daily weather profile is accomplished through clustering techniques. Genetic algorithm is adopted to determine the optimal architecture of each DNN sub-model. Namely, each featuring cluster of weather profile, along with corresponding time signature and building energy consumption, is adopted to train one DNN sub-model. Therefore, the structure, activation function and training approach of DNN sub-models are adaptive to diverse featuring datasets in each cluster. To evaluate the effectiveness of the proposed predictive model, it is implemented on a real office building in the United Kingdom. Mean absolute percentage error of the training and testing cases of the proposed predictive model is 2.87% and 6.12%, which has a 24.6% and 11.9% decrease compared to DNN model with a fixed architecture. With the latest weather forecast, the devised adaptive DNN model can provide accurate week-ahead hourly energy consumption prediction for building energy management system.

Keywords

Energy consumption prediction; Feature extraction; Clustering; Adaptive; Deep neural network; Genetic algorithm; Data-driven model.

Nomenclature

C	Dataset of cluster centroid
\mathbf{C}	Database of cluster centroid
d	Day of the week
e	Original measurement of energy consumption
e	Energy consumption after clustering
\hat{e}	Predicted energy consumption
E	Energy consumption in each cluster
\mathbf{E}	Original Energy consumption database
$\hat{\mathbf{E}}$	Predicted energy consumption database
m	Month of the year
$MAPE$	Mean absolute percentage error
MPE	Mean percentage error
N	Quantity
r^2	Coefficient of determination
$RMSE$	Root mean square error
t	Original time variables
\mathbf{t}	Time variables after clustering
\mathbf{T}	Time signature in each cluster
\mathbf{T}	Database of time signature
u	Clustering result
w	Weight of the connection in DNN
x'	Original weather data at each hour
x	Normalized weather data at each hour
X	Daily profile of weather data
\mathbf{X}	Weather data in each cluster
\mathbf{X}	Database of weather data
y	Dataset for DNN training
$\ \cdot \ $	Euclidean distance

Subscripts

i	Day of the year
k	Hour of the day
m	Cluster number
opt	Optimal quantity of clusters

Abbreviations

ADADELTA	Adaptive learning rate method
ADAGRAD	Adaptive gradient algorithm
ADAM	Adaptive moment estimation
ANN	Artificial neural network
CNN	Convolutional neural network
elu	Exponential linear function
DNN	Deep neural network
GA	Genetic algorithm
LSTM	Long short-term memory
NADAM	Nesterov-accelerated Adaptive Moment Estimation
relu	Rectified linear unit function
RMSProp	Root mean square propagation
RNN	Recurrent neural network
SGD	Stochastic gradient descent approach

1. Introduction

Owing to the fact that people spend over 90% of their time indoors, buildings have become the largest energy consumer worldwide. Increasing energy consumption has forced people to use energy in a more efficient and effective way. Long-term energy consumption prediction is important in building design, while short-term prediction is vital in effective energy resources management. Since it is difficult to accumulate energy at a considerable amount, building energy supply should be calibrated to match the consumption immediately [1, 2]. As a result, it is significant that the prediction of building energy consumption shall achieve as high accuracy as possible. However, there are a range of affecting factors on energy performance, including building thermal properties, type of installed energy equipment units, outdoor weather condition, as well as energy usage behaviour of the occupants [3]. Owing to the comprehensive and nonlinear relationship among these affecting factors and building energy performance, it is a highly sophisticated task to predict week-ahead hourly building energy consumption accurately.

1.1 Related works

First-principle models and data-driven models are two major types of building energy prediction [4-6]. First-principle models are generally comprised of thermodynamic equations. According to the detailed environmental and building design parameters, for instance, weather profile of outdoor environment, floor layout, thermal properties of building envelopes, and operating schedules of energy equipment units, building energy consumption can be estimated. However, these operating parameters are generally difficult to obtain. Moreover, the complicated building layouts, especially in modern buildings, will make the first-principle models computationally expensive [4]. On the other hand, data-driven models, especially various neural networks, are widely adopted in predicting building energy consumption. The historical measurement of outdoor environment and energy consumption is generally adopted as database to train data-driven models. The training process is to determine the related parameters of data-driven models and investigate the complicated nonlinear relationship among various input and output datasets.

Various neural networks have been constructed for building energy performance prediction, most of which use one week's, several months' or entire year's meteorological weather profile to train one neural network with a fixed structure. Yang *et al.* [7] adopted an artificial neural network (ANN) with a fixed structure for hourly energy consumption prediction in buildings. Several months' simulation data of outdoor air dry-bulb and wet-bulb temperature as well as electrical energy consumption is adopted to train the ANN model. Ahmad *et al.* [8, 9] adopted the ANN with a fixed structure for hourly cooling demand prediction in buildings using the simulation data from one week. Wang *et al.* [10]

adopted a hybrid ANN and ensemble model for cooling demand prediction in buildings using one-month's simulation data. Katsatos et al. [11] adopted an ANN with a fixed structure for building energy consumption prediction using 3 years' measurement data. Biswas et al. [12] adopted an ANN with fixed structure for daily electrical consumption prediction of a residential building. Three months' measurement data is used as database. Singaravel et al. [13] proposed a deep learning neural network architecture for predicting monthly energy consumption of the office building at design stage. Year-round simulation data is adopted as database to the neural network architecture while trial and error process is used to determine its optimal structure. Ahamad et al. [14] proposed a random neural network model for predicting energy utilization for a campus building composing of 562 rooms. The whole-year database is adopted to train one single random neural network. Kim et al. [15] proposed a hybrid convolutional neural network (CNN) and long short-term memory (LSTM) prediction model for predicting electricity consumption in the household. Four years' measurement data is adopted to train one single CNN-LSTM model with a fixed structure. Cheng et al. [16] investigated the use of the recurrent neural network (RNN) for cooling load prediction in buildings. The measurement data from the entire year of 2015 is adopted as database.

However, owing to the variation of yearly meteorological weather profile, database within a short period (i.e. one week or several months) is not sufficient to indicate the year-round variation of building energy consumption. Meanwhile, there exist different features in year-round weather profile. Using the whole-year database to train one neural network may lead to inaccurate prediction. Kusiak et al. [17] proposed an ANN model to forecast the daily building thermal energy demands with the information of outdoor air dry-bulb temperature and humidity ratio. To represent the daily weather patterns, the average, deviation, maximum and minimum values of outdoor air dry-bulb temperature and humidity ratio are adopted as input datasets to predictive model. Owing to the reason that ANN parameters would be profoundly affected by seasonal weather profile, manual classification approach is adopted for data division and selection. To be more specific, the database is grouped according to the pre-set time duration and the average value of daily outdoor air dry-bulb temperature. Jovanovic et al. [18] integrated feed-forward backpropagation neural network, radial basis function network and adaptive neuro-fuzzy interference system for predicting thermal energy consumption of a campus building. The entire year is classified as cold, mild and warm periods, while the historical measurement during the cold period is adopted to train the proposed model. Deb et al. [19] proposed an ANN predictive model for thermal energy consumption in the institutional buildings. Owing to the high variation of annual energy consumption, the values of energy consumption values are assigned as very low, low, medium, high and very high. Each group of database is adopted to train one ANN predictive model. Bagnasco et al. [20] proposed an ANN model for predicting electricity loads of a hospital building in Italy. To reduce the forecast errors caused by different load shapes in every season, the yearly dataset is divided into four seasonal subsets according to its number of month. Mena et al. [21] proposed an ANN model to

predict the hourly electricity consumption of a public building. The year-round dataset is divided into four groups according to the natural seasons. It is found that peak demand and standard deviation is similar in the same season.

The structure of those above-mentioned neural networks is determined randomly or through trial and error processes. To enhance prediction accuracy, hybrid evolutionary optimization and neural networks is proposed. Li et al. [22] proposed a hybrid teaching-learning optimization and ANN for predicting building electrical energy consumption. Several months' energy data from two buildings is adopted as database, while teaching-learning optimization approach is used to determine the optimal weighting factors of ANN. Somu et al. [23] proposed a hybrid sine cosine optimization and LSTM networks for energy consumption prediction of an academic building. The year-round weather database is divided into summer, monsoon and winter periods. The hyper-parameters of the LSTM networks, including learning rate, weight decay, momentum, and quantity of hidden layers, are optimized using the sine cosine optimization algorithm. Ruiz et al. [24] proposed a hybrid genetic algorithm optimization and Elman neural network for predicting building energy consumption. The genetic algorithm is adopted to search the optimal weighting factors for Elman neural network. The entire one-year measurement data is adopted as the database to train one single Elman neural network model. Wu et al. [25] proposed a hybrid multi-objective cuckoo search algorithm and generalized regression neural network prediction model for predicting regional electricity consumption. The cuckoo search algorithm is adopted to find the optimal weights and thresholds of the neural network. Liang et al. [26] proposed a hybrid fruit fly optimization and ANN prediction model for predicting region-wide electricity consumption for the electric power market. Three years measurement data from a regional power system is adopted to train the ANN model, while the fruit fly optimization is adopted to determine the optimal weighting factors of ANN.

Through performance comparison with random forest, multiple regression, genetic programming and support vector machine for electricity consumption prediction, it is also demonstrated that neural networks have the best prediction accuracy [27, 28]. Furthermore, ANN has also been widely adopted in building energy modelling [29-31] and control optimization of building energy [32, 33].

1.2 Limitation in the available literature

From the comprehensive literature review, it is found that various neural networks have become the most prominent techniques for building thermal and electrical energy prediction. The most commonly used neural network is ANN, while RNN, CNN and LSTM neural network is also investigated. However, the following deficits of the existing neural network-based predictive models are identified:

- The single hidden layer is adopted in most of the existing neural network-based predictive models. However, the single hidden layer might not be sufficient to reveal the comprehensive relationship among various affecting factors and building energy performance.
- Most of the existing neural network-based prediction models use one week or several months' meteorological weather profile to train the predictive model. However, data within a short period might be insufficient to represent the variation of building energy consumption because the meteorological weather profile varies during the whole year [1-3].
- The entire-year database is adopted to train one single neural network with fixed structure. Although [17-21] mentioned the classification of year-round database according to time duration, manual classification approaches are adopted and may not be accurate to distinguish the features of weather profile adequately.
- Most of the existing neural networks are based on a fixed architecture, with pre-set or trial-and-error-determined quantity of neurons in the hidden layer, activation function and training approach.
- In some of the studies, noise-free simulation data is adopted. It provides an ideal scenario thus the variation of energy performance is more predicable than that from energy simulation results.

1.3 Novelty and contribution

From the literature review mentioned above, there is a lack of study regarding the adaptive, robust and effective predictive models for building energy consumption. Therefore, the aim of this study is to propose a feature extraction and genetic algorithm enhanced adaptive deep neural network (FE-GA-DNN) predictive model for building energy consumption. The novelty of the proposed prediction model is summarized as follows:

- Deep neural network (DNN), with multiple hidden layers, will be adopted to reveal the complex relationship among diverse affecting factors and building energy consumption.
- Clustering technique will be used to extract the featuring patterns of daily weather data and group the yearly profile into several clusters. Thus, the datasets in each cluster, with the distinguishing features, can be adopted to train one DNN sub-model.
- Genetic algorithm (GA) is adopted to determine the optimal architecture of DNN sub-models for each group of datasets. Therefore, the architecture (i.e. quantity of hidden layers, quantity of neurons in each hidden layer, activation function and training approach) of the proposed predictive model is adaptive to the featuring datasets.
- The measurement from the energy management system in a practical office building is adopted. It provides a real-life scenario and there might exist noise or faulty data.

With the adoption of integrated artificial intelligence-based approaches, including feature extraction, evolutionary optimization and DNN, the proposed predictive model will serve as an adaptive, effective,

accurate and robust prediction approach for building energy consumption in various practical applications.

The rest of the paper is structured like this: The next section illustrates the generation of historical database; the following section illustrates the details of the proposed FE-GA-DNN model; the fourth section evaluates the performance of the proposed FE-GA-DNN model; the fifth section details the implication for practice and future study while the last section points out the conclusion.

2. Preparing the historical database

To evaluate the performance of the proposed FE-GA-DNN prediction model, it is tested on a practical office building in the United Kingdom. The electricity consumption data is collected from the energy management system of Costain Headquarters building over the year of 2017 at the time step of 1 h.

2.1 Building description

The Costain Headquarters building is situated at Vanwall Business Park, Maidenhead, the United Kingdom. It is a modern, three-storey air-conditioned office suite with the floor area of 277 m². The floor-to-ceiling height is 3.5 m, while the window-to-wall ratio is 0.8. During peak hours, the occupational density is 8 m²/person.

2.2 Composition of database

From the literature review discussed in Section 1, building energy performance is influenced by various factors, including weather conditions, building thermal properties, operating schedules of lighting and air conditioning systems, as well as occupancy behaviour. In the office building, the operating schedules of lighting and air conditioning systems as well as occupancy behaviour is relatively stable and correlated with time signatures. Therefore, time signatures, including hour of the day, month of the year, as well as day of the week are adopted as input datasets to the prediction model. Meanwhile, weather conditions can be well-described using the outdoor air dry-bulb temperature T_{db} , outdoor air dew-point temperature T_{dew} and cloud cover ratio R_c . The historical outdoor weather data is collected through the local weather station at Maidenhead.

2.3 Data pre-processing

The weather data $x'_{i,j,k}$ at the k^{th} hour on the i^{th} day of the year can be formulated as: $x'_{i,1,k} = T_{db,i,k}$; $x'_{i,2,k} = T_{dew,i,k}$; $x'_{i,3,k} = R_{c,i,k}$. Since the value range of each parameter $x'_{i,j,k}$ in the weather profile is quite different, min-max normalization is implemented to normalize the values of weather data into the range between 0 and 1.

$$x_{i,j,k} = \frac{x'_{i,j,k} - \min_{1 \leq i \leq 365, 1 \leq k \leq 24} x'_{i,j,k}}{\max_{1 \leq i \leq 365, 1 \leq k \leq 24} x'_{i,j,k} - \min_{1 \leq i \leq 365, 1 \leq k \leq 24} x'_{i,j,k}} \quad (1)$$

To represent the cyclical nature of the time, the time signature, hour of the day $h_{i,k}$, is represented by its sine and cosine value as:

$$h_{s,i,k} = \sin \frac{2\pi h_{i,k}}{24} \quad (2)$$

$$h_{c,i,k} = \cos \frac{2\pi h_{i,k}}{24} \quad (3)$$

The public holidays are marked as an additional type in day of the week. For normalization purpose, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday and public holiday is represented by $d_{i,k} = 1/8, 1/4, 3/8, 1/2, 5/8, 3/4, 7/8$ and 1, respectively. Meanwhile, month of the year $m_{i,k}$ is assigned with the values of $1/12, 1/6, 1/4, 1/3, 5/12, 1/2, 7/12, 2/3, 3/4, 5/6, 11/12$ and 1, respectively, according to the order of month. Therefore, time signature consists of four time elements: $t_{i,1,k} = h_{s,i,k}$, $t_{i,2,k} = h_{c,i,k}$, $t_{i,3,k} = d_{i,k}$, $t_{i,4,k} = m_{i,k}$. $T_{i,1,k} = h_{s,i,k}$, $T_{i,1,k} = h_{s,i,k}$, $T_{i,1,k} = h_{s,i,k}$, $T_{i,1,k} = h_{s,i,k}$, $T_{i,1,k} = h_{s,i,k}$,

In summary, the pre-processed input and output datasets of the proposed predictive model is summarized in Table 1. $\mathbf{X} = \{x_{i,j,k} | i = 1, 2, \dots, 365; j = 1, 2, 3; k = 1, 2, \dots, 24\}$, $\mathbf{T} = \{t_{i,j,k} | i = 1, 2, \dots, 365; j = 1, 2, 3, 4; k = 1, 2, \dots, 24\}$, $\mathbf{E} = \{E_{i,k} | i = 1, 2, \dots, 365; k = 1, 2, \dots, 24\}$. For training and testing purposes of the proposed predictive model, 80% of the historical database is selected for training one the remaining 20% was adopted for testing purpose, respectively.

Table 1. Input and output datasets of the proposed predictive model.

Datasets		Source	
Input	Weather condition \mathbf{X}	Outdoor air dry-bulb temperature $x_{i,1,k}$	Weather profile from local weather station
		Outdoor air dew-point temperature $x_{i,2,k}$	
		Cloud cover ratio $x_{i,3,k}$	
	Time signature \mathbf{T}	Sine value of hour of the day $t_{i,1,k}$	Time information related to weather condition and building energy consumption
		Cosine value of hour of the day $t_{i,2,k}$	
		Day of the week plus public holiday $t_{i,3,k}$	
		Month of the year $t_{i,4,k}$	
Output \mathbf{E}	Building electricity consumption $E_{i,k}$	Energy management system	

3. Proposed FE-GA-DNN prediction model

The proposed FE-GA-DNN prediction model mainly comprises of three artificial intelligence approaches, including: feature extraction, GA optimization and DNN model. The outline of the proposed predictive model is illustrated in Fig. 2.

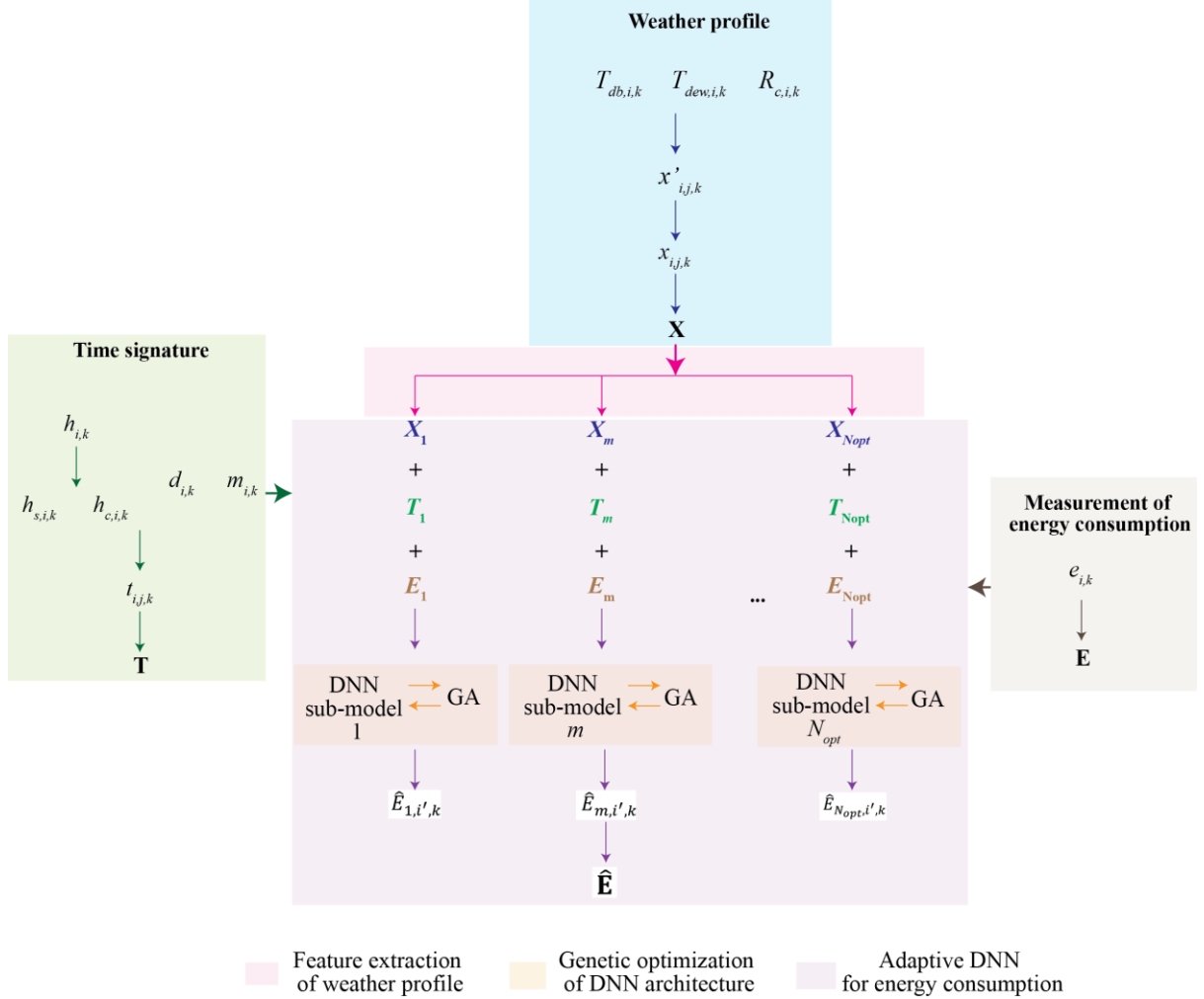


Fig. 2. Outline of the proposed clustering-enhanced adaptive artificial neural network model.

3.1 Clustering analysis in Feature extraction

To extract the features of daily weather profile, k -means clustering is adopted to classify the year-round daily weather profile into different featuring groups. The annual historical weather profile \mathbf{X} consists of 365 sets of daily weather profile the normalized weather data and $\mathbf{X} = [X_1; X_2; \dots; X_i; \dots; X_{365}]$; X_i is the daily profile of various weather data, and $X_i = [x_{i,1,1} \ x_{i,1,2} \ \dots \ x_{i,1,24} \ x_{i,2,1} \ x_{i,2,2} \ \dots \ x_{i,2,24} \ x_{i,3,1} \ x_{i,3,2} \ \dots \ x_{i,3,24}]$. Through k -means clustering, the daily weather profile X_i ($i = 1, 2, \dots, 365$) can be classified into certain number (m) of clusters ($m = 1, 2, \dots,$

N_c). The dataset of cluster centroid C_m is determined to minimize the sum of Euclidian distance among each dataset to its corresponding cluster centroid as described in Eq. (4) [36]

$$J = \sum_{m=1}^{N_c} \sum_{i=1}^{365} (\sum u_{i,m} \|X_i - C_m\|) \quad (4)$$

where,

$\| \ \|$: Euclidean distance;

m : number of cluster;

N_c : quantity of clusters;

$u_{i,m}$: clustering factor, $u_{i,m} = 1$ when X_i is belong to cluster m ; $u_{i,m} = 0$ otherwise.

The cluster centroid $C_m = [c_{m,1,1} \ c_{m,1,2} \ \dots \ c_{m,1,24} \ c_{m,2,1} \ c_{m,2,2} \ \dots \ c_{m,2,24} \ c_{m,3,1} \ c_{m,3,2} \ \dots \ c_{m,3,24}]$ is determined by:

$$C_m = \frac{\sum_{i=1}^{365} (X_i u_{i,m})}{\sum_{i=1}^{365} u_{i,m}} \quad (5)$$

The degree of variability of the weather profiles in each cluster is given by the within-cluster sum of squares (WCSS). The value of WCSS generally decreases as the quantity of clusters increases. The optimal quantity of clusters N_{opt} is determined using the elbow criterion as a heuristic, so as to minimize the value of WCSS while maintaining a low quantity of clusters [34, 35]:

$$WCSS = \sum_{m=1}^{N_c} \sum_{i=1}^{365} (\sum u_{i,m} \|X_i - C_m\|^2) \quad (6)$$

Based on the clustering result $u_{i,m}$, the time signature profile \mathbf{T} and year-round energy consumption profile \mathbf{E} are also grouped into corresponding clusters. The cluster centroid database \mathbf{C} can be formulated as $\mathbf{C} = [C_1; C_2; \dots; C_m; \dots; C_{N_{opt}}]$. The procedure of the data clustering (i.e. feature extraction) is shown in Fig. 3, where i' represents the i^{th} day while N_m stands for the quantity of days in cluster m .

In the m^{th} cluster, the weather profile, time signature and energy consumption can be formulated as

$$\mathbf{X}_m = \{\mathbf{x}_{m,i',j,k} | i' = 1, 2, \dots, N_m; j = 1, 2, 3; k = 1, 2, \dots, 24\};$$

$$\mathbf{T}_m = \{\mathbf{t}_{m,i',j,k} | i' = 1, 2, \dots, N_m; j = 1, 2, 3, 4; k = 1, 2, \dots, 24\};$$

$$\mathbf{E}_m = \{\mathbf{e}_{m,i',k} | i' = 1, 2, \dots, N_m; k = 1, 2, \dots, 24\}, \text{ respectively.}$$

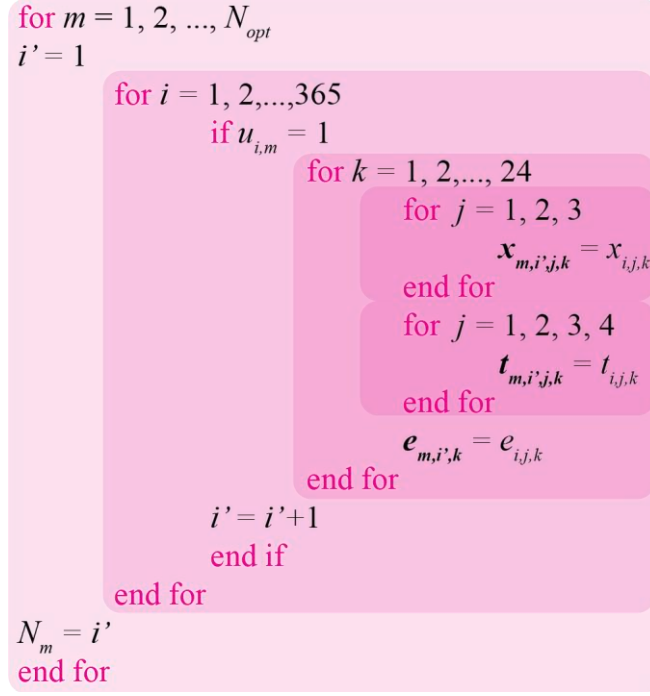


Fig. 3. Procedure of feature extraction.

3.2 DNN model in energy consumption prediction

$\mathbf{x}_{m,i',j,k}$, $\mathbf{t}_{m,i',j,k}$ and $\mathbf{e}_{m,i',k}$ at each time step is regarded as one training sample for the m^{th} DNN sub-model, and $Y_{m,j} = \{\mathbf{x}_{m,i',j,k} | i' = 1, 2, \dots, N_m; j = 1, 2, 3; k = 1, 2, \dots, 24\}$, $Y_{m,j+3} = \{\mathbf{t}_{m,i',j,k} | i' = 1, 2, \dots, N_m; j = 1, 2, 3, 4; k = 1, 2, \dots, 24\}$. The DNN models are generally comprised of three principle layers: an input layer, several hidden layers and the outlet layer [37, 38]. The quantity of hidden layers defines the depth of the architecture. The diagram of the DNN model is presented in Fig. 4.

$$z_{m,1,h_1} = f(\sum_{j=1}^{j=7} (w_{m,0,j,h_1} Y_{m,j})) \quad (7)$$

while the j_k^{th} neuron in the k^{th} ($n \geq k \geq 2$) hidden layer can be obtained as:

$$z_{m,i,h_i} = f(\sum_{h_{i-1}=1}^{h_{i-1}=N_{h_{i-1}-1}} (w_{m,i,h_{i-1},h_i} \cdot z_{m,i-1,h_{i-1}})) \quad (8)$$

The neuron z in the output layer is determined as:

$$\mathbf{E}_m = f(\sum_{h_n=1}^{h_n=N_{h_n}} (w_{m,n,h_n} \cdot z_{m,n,h_n})) \quad (9)$$

The aim of the training DNN model is to minimize the mean absolute percentage error $MAPE$ between the predicted energy consumption \hat{E}_m and the measured electricity consumption E_m by adjusting the group of weighting factors. Therefore, the year-round energy consumption can be predicted as $\hat{E} = \{\hat{e}_{m,i',k} | m = 1, 2, \dots, N_{opt}; i' = 1, 2, \dots, N_m; k = 1, 2, \dots, 24\}$

$$MAPE = \frac{1}{N_m \times 24} \sum_{i'=1}^{N_m} \sum_{k=1}^{24} \frac{|e_{m,i',k} - \hat{e}_{m,i',k}|}{e_{m,i',k}} \times 100\% \quad (10)$$

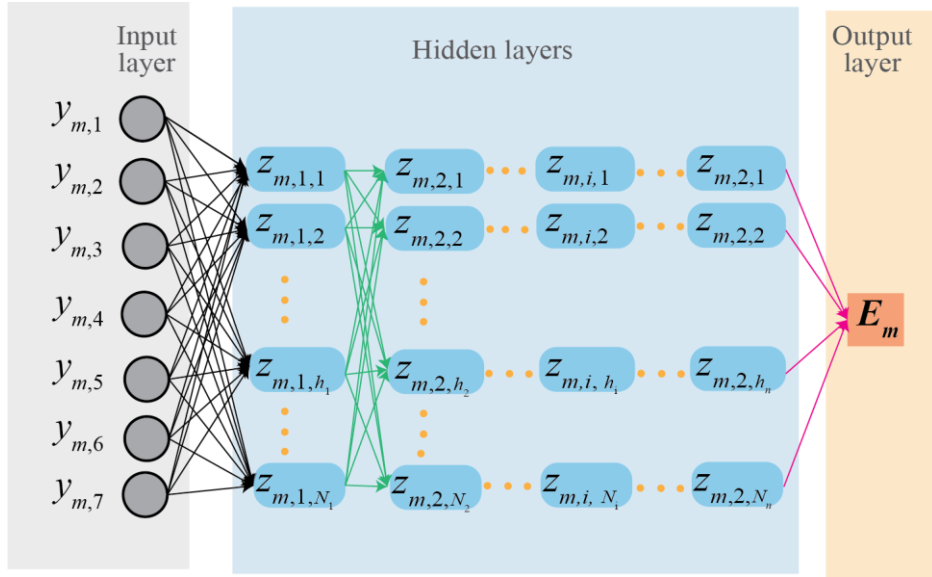


Fig. 4. Diagram of DNN model.

To investigate the effects of architecture on the prediction performance of DNN model, different quantity of hidden layers, different quantity of neurons in each layer (i.e. N_1, N_2, \dots, N_n), activation functions (f) and training approaches are investigated. The activation functions include sigmoid function, tanh function, Rectified Linear Unit (relu) function and Exponential Linear (elu) function. The training approaches include stochastic gradient descent (SGD), adaptive gradient algorithm (ADAGRAD), adaptive learning rate method (ADADELTA), root mean square propagation (RMSProp), adaptive moment estimation (ADAM), Nesterov-accelerated adaptive moment estimation (NADAM) and ADAMAX.

3.3 GA for determining optimal DNN architecture

To obtain the optimal architecture of DNN model, GA is applied to determine the optimal quantity of hidden layers, quantity of neurons in each hidden layer, activation function and training approach. The optimization variables are summarized in Table 2. Owing to the complicated nature of decision

variables, conventional optimization techniques such as gradient-based methods or linear programming are not suitable. However, GA is good at obtaining the global optimum in the infinite, complex, multimodal, and non-differentiable spaces to determine the optimal ANN architecture. The procedure of determining the optimal architecture of DNN is shown in Fig. 5. At the beginning, a population of 20 DNN models are generated while random decision variables are assigned to determine the architecture of each DNN model. Each DNN model is trained using the input and output datasets from historical database. The training error (Eq. (7)) of each DNN model is obtained and treated as the fitness value of the objective function in GA optimization. Based on the fitness value, selection, crossover and mutation operators of GA are conducted for the population of DNN models, respectively. Such procedure is repeated until the preset optimization criteria is met. Finally, the optimal DNN architecture can be determined. The parameters of GA are summarized in Table 3.

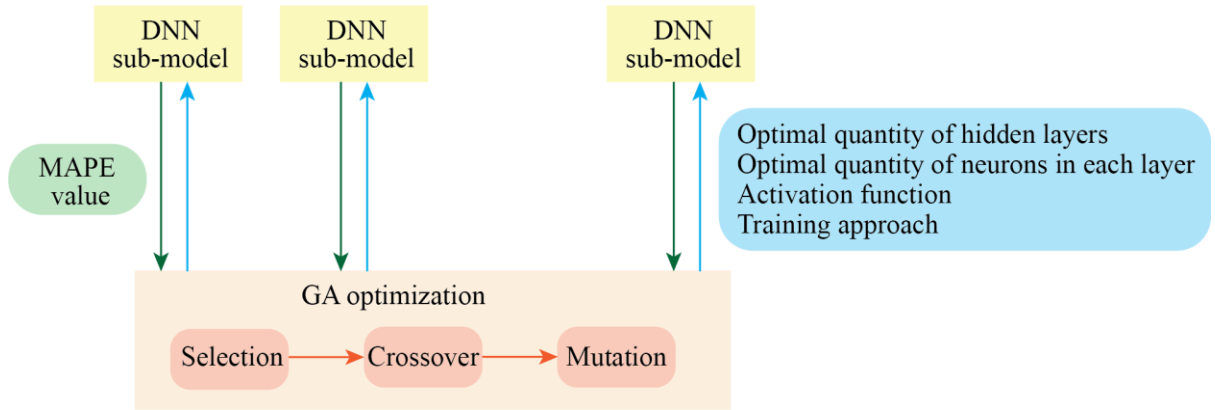


Fig. 5. Schematic flowchart of the proposed GA-enhanced DNN predictive model.

Table 2. Decision variables of GA for DNN.

Quantity of neurons in each hidden layer	{10, 20, 30, 40, 50, 60, 70, 80, 90, 100}
Quantity of hidden layers	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
Activation functions	{sigmoid, tanh, relu, elu}
Optimization approaches	{SGD, ADAGRAD, ADADELTA, RMSProp, ADAM, NADAM, ADAMAX}

Table 3. GA parameters

GA parameters	Value
Population	20
Maximum generations	50
Selection probability	0.4
Crossover probability	0.6
Mutation probability	0.3

3.4 Data post-processing

Through feature extraction, the year-round weather profile \mathbf{X} , along with time signature \mathbf{T} , can be divided into N_{opt} groups, as \mathbf{Y}_m and \mathbf{E}_m ($m = 1, 2, \dots, N_{opt}$). Through genetic algorithm, the optimal

architecture of each DNN sub-model can be determined. Through DNN model, the predicted energy consumption $\hat{\mathbf{E}}_m$ from each cluster can be determined. Thereafter, $\hat{\mathbf{E}}_m$ can be regrouped to formulate year-round predicted energy consumption $\hat{\mathbf{E}}$, and $\hat{\mathbf{E}} = \{\hat{\mathbf{E}}_m | m = 1, 2, \dots, N_{opt}\}$.

3.5 Performance evaluation index of DNN models

Apart from *MAPE*, three additional performance indicators are defined to evaluate the performance of DNN model, including coefficient of determination, mean percentage error and root mean square error:

- Coefficient of determination (r^2)

$$r^2 = \frac{\left[\sum_{i'=1}^{i'=Nm} \sum_{k=1}^{k=Nh} \left(E_{m,i',k} - \frac{\sum_{i'=1}^{i'=Nm} \sum_{k=1}^{k=24} e_{m,i',k}}{Nm \times Nh} \right) \left(\hat{E}_{m,i',k} - \frac{\sum_{i'=1}^{i'=Nm} \sum_{k=1}^{k=24} \hat{e}_{m,i',k}}{Nm \times Nh} \right) \right]^2}{\sum_{i'=1}^{i'=Nm} \sum_{k=1}^{k=24} \left(e_{m,i',k} - \frac{\sum_{i'=1}^{i'=Nm} \sum_{k=1}^{k=24} e_{m,i',k}}{Nm \times 24} \right) \sum_{i'=1}^{i'=Nm} \sum_{k=1}^{k=24} \left(E_{m,i',k} - \frac{\sum_{i'=1}^{i'=Nm} \sum_{k=1}^{k=24} e_{m,i',k}}{Nm \times 24} \right)} \quad (11)$$

- Mean percentage error (*MPE*)

$$MPE = \frac{1}{Nm \times 24} \sum_{i'=1}^{i'=Nm} \sum_{k=1}^{k=24} \frac{e_{m,i',k} - \hat{e}_{m,i',k}}{e_{m,i',k}} \quad (12)$$

- Root mean square error (*RMSE*)

$$RMSE = \sqrt{\frac{\sum_{i'=1}^{i'=Nm} \left(e_{m,i',k} - \hat{e}_{m,i',k} \right)^2}{Nm \times 24}} \quad (13)$$

4. Results and discussion

4.1. Feature extraction of weather profile

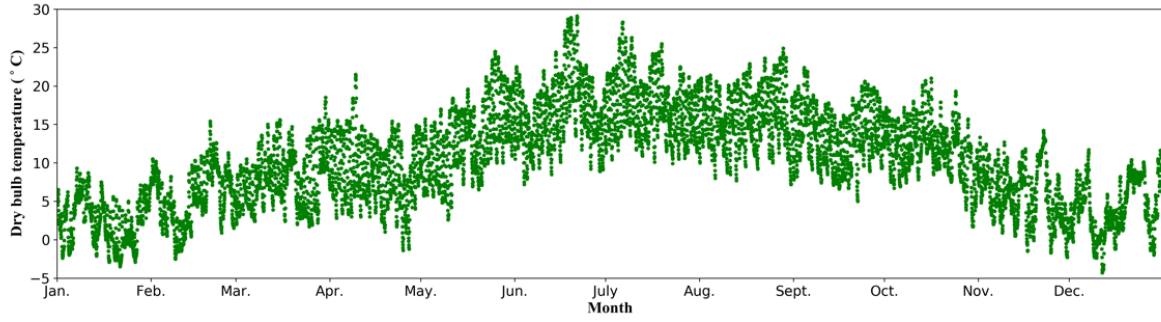
Feature extraction is one of the novelties of the proposed FE-GA-DNN model, and it plays a fundamental role. In this subsection, the weather profile is analysed first, followed by determination of optimal quantity of clusters. After that, feature identification and correlation coefficient analysis is conducted.

4.1.1 Analysis of weather profile

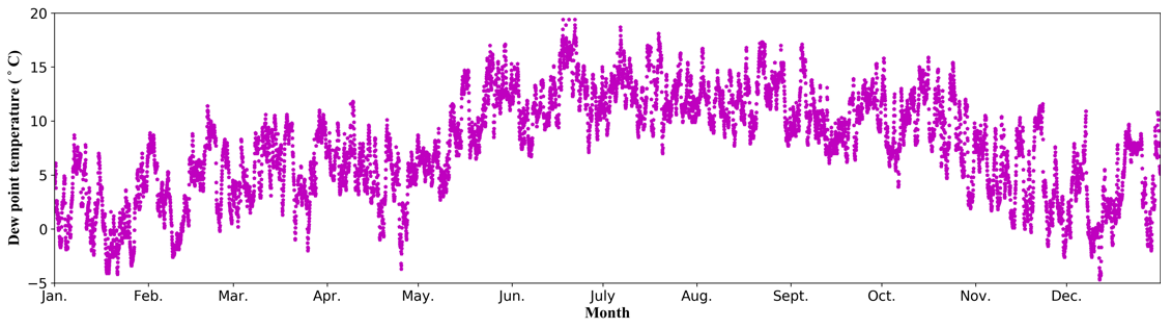
The outdoor air dry-bulb temperature, outdoor air dew-point temperature and cloud cover ratio are illustrated in Fig. 6. The maximum, minimum and average value of outdoor air dry-bulb temperature, outdoor air dew-point temperature and cloud cover ratio are summarized in Table 4. For outdoor air dry-bulb temperature and dew-point temperature, they reach the peak in June while drop to the valley in December. The average value of outdoor air dry-bulb and dew-point temperature has the largest and smallest value in July and January, respectively. For cloud cover ratio, it varies from 0 to 100 day-by-day. The average value of cloud cover ratio has the largest and smallest values in December and August, respectively.

Table 4. Brief comparison weather condition in each month.

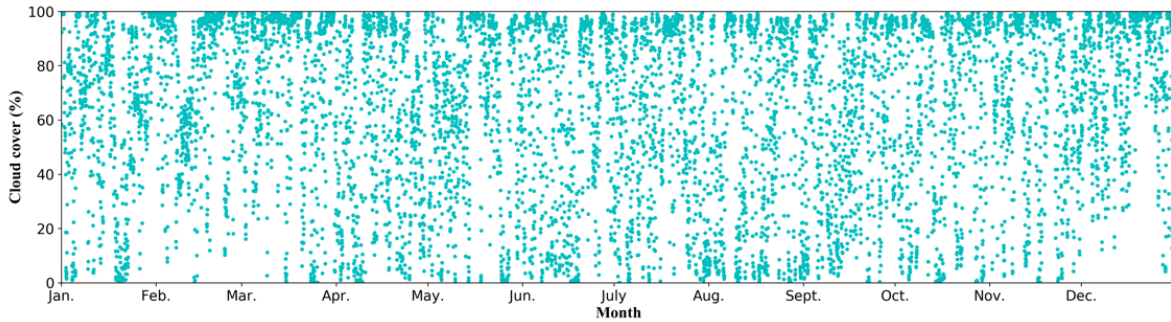
Weather condition		Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Dry-bulb temperature T_{db} (°C)	Maximum	9.3	15.4	18.5	21.5	24.5	29.1	28.3	24.9	22.4	21.0	14.2	11.7
	Minimum	-3.5	-2.5	1.7	-1.4	2.6	7.2	9.9	7.2	5.0	1.7	-2.3	-4.3
	Average	2.3	5.4	8.0	8.6	13.1	16.7	17.3	15.9	13.4	11.8	5.8	4.1
Dew-point temperature T_{dew} (°C)	Maximum	8.7	11.4	11.0	11.8	17.1	19.4	18.7	17.3	17.1	15.9	11.6	10.9
	Minimum	-4.2	-2.6	-2.0	-3.7	2.7	6.7	7.0	7.7	6.1	1.9	-2.6	-4.7
	Average	1.6	4.2	5.7	5.0	9.6	12.3	12.6	12.1	10.4	9.6	4.7	3.5
Cloud cover ratio R_c (%)	Maximum	100	99.9	99.9	99.6	99.6	99.6	99.6	99.5	99.5	99.7	99.9	100
	Minimum	0	0	0	0	0	0	0	0.1	0.1	0	0.2	4.1
	Average	63.4	74.7	75.1	55.8	62.6	59.6	58.4	54.6	58.5	63.6	69.4	78.5



(a) Outdoor air dry-bulb temperature



(b) Outdoor air dew-point temperature



(c) Cloud cover ratio

Fig. 6. Year-round historical data.

4.1.2 Determination of optimal quantity of clusters

To determine the optimal quantity of clusters N_{opt} for k -means clustering, the values of WCSS are tested among $N_c = 2-20$, as shown in Fig. 7. The decrease of WCSS is initially steep and then the decreasing rate slows down, which results in an elbow plot. The optimal quantity of clusters is considered to be located at the elbow point, and $N_{opt} = 4$.

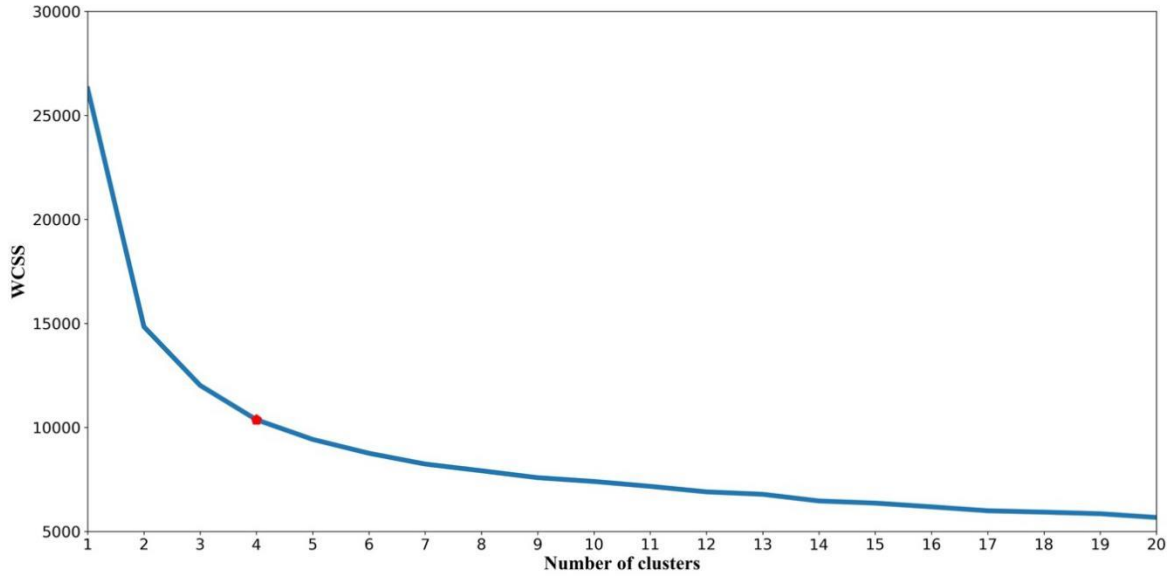


Fig. 7. WCSS value for different quantity of clusters.

The cluster distribution of each day over the whole year is shown in Fig. 8. It is found that clusters C1 and C3 are mainly distributed among the cold season, namely, from the beginning of January to middle of May, as well as from end of October to the end of the year. Meanwhile, clusters C2 and C4 are majorly distributed among the hot season, in other words, from the middle of May towards the end of October. However, there are also a few clusters C2 and C4 located in March, April and November, as well as some clusters C3 in June, August and September. It indicates that it might not be appropriate to simply divide the whole year into different groups manually according to the time period as introduced in references [8-10].

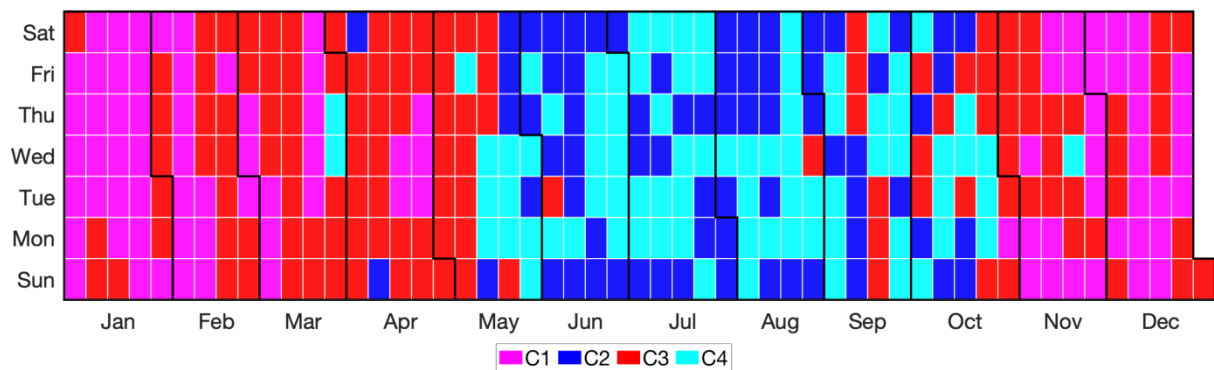
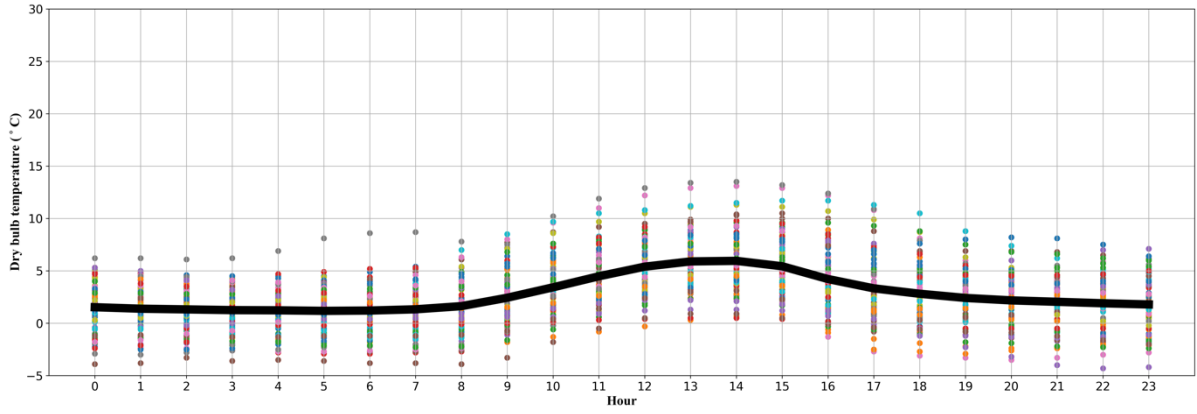
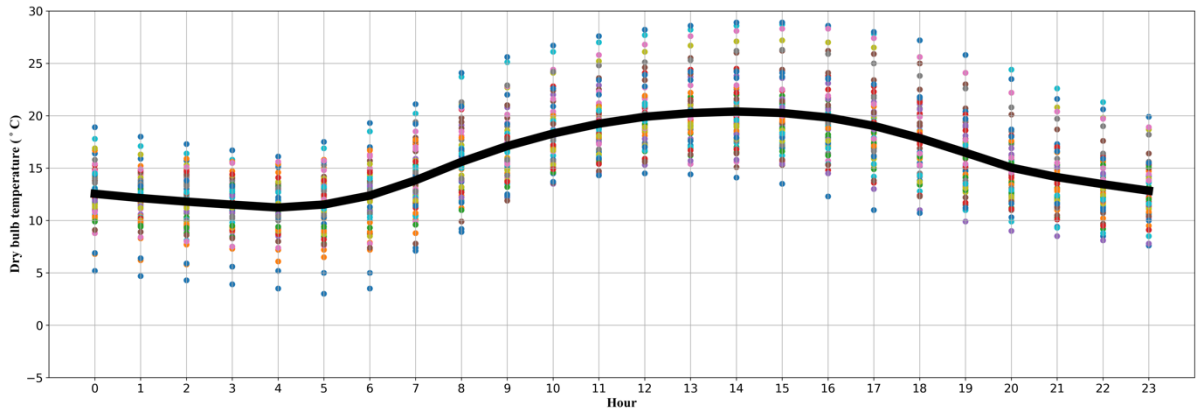


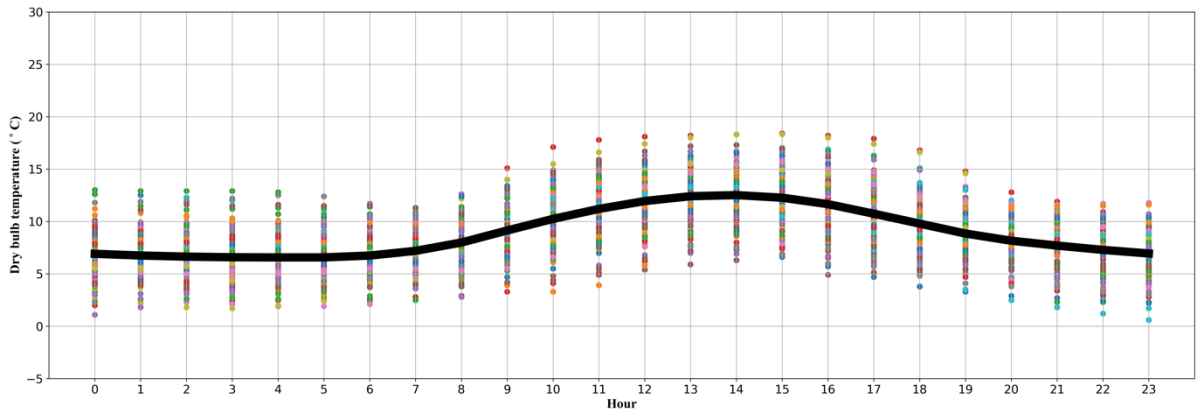
Fig. 8. Calendar view of cooling load distribution.



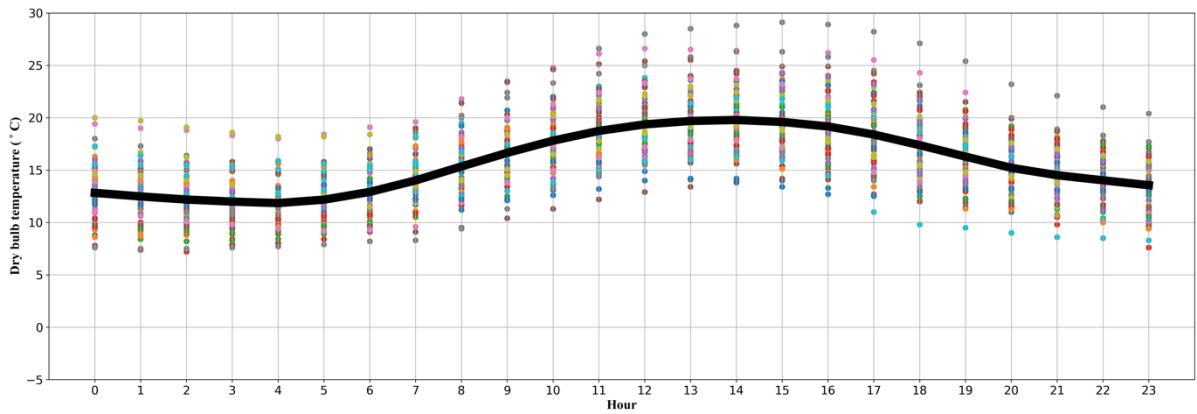
(a) C1



(b) C2

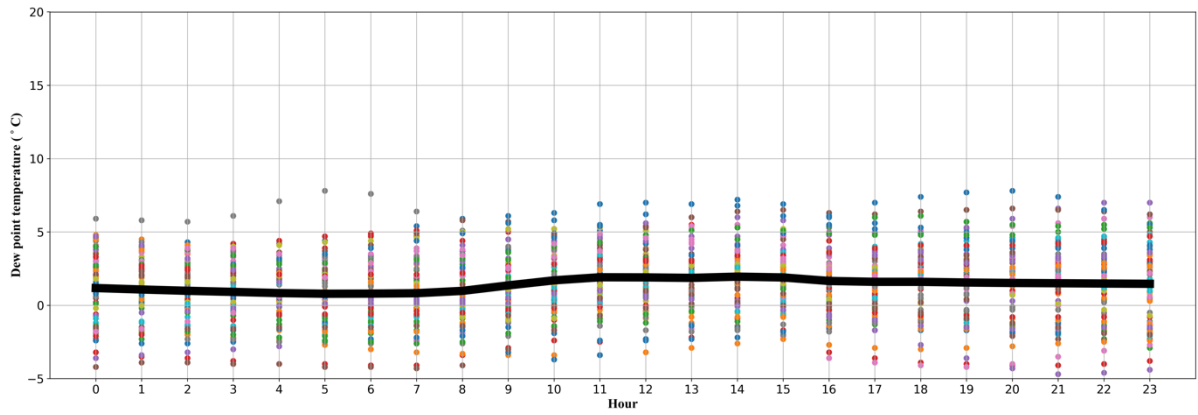


(c) C3

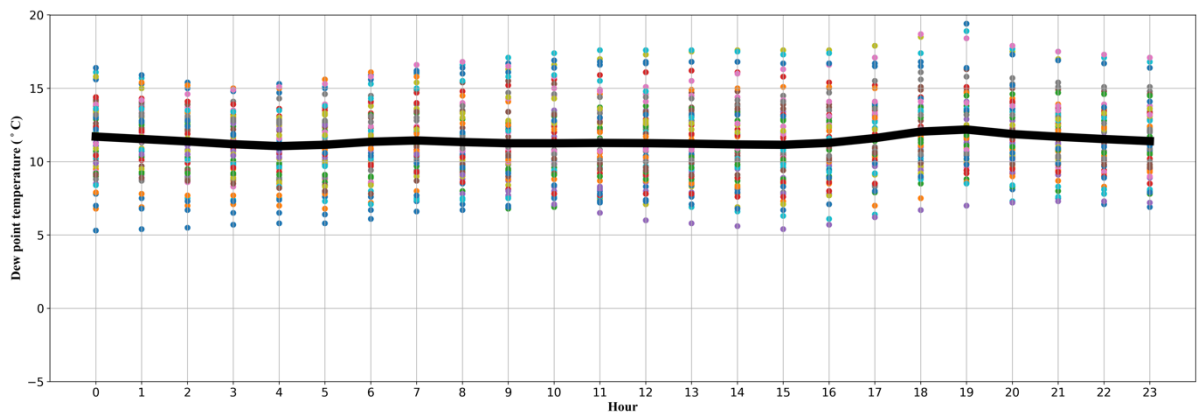


(d) C4

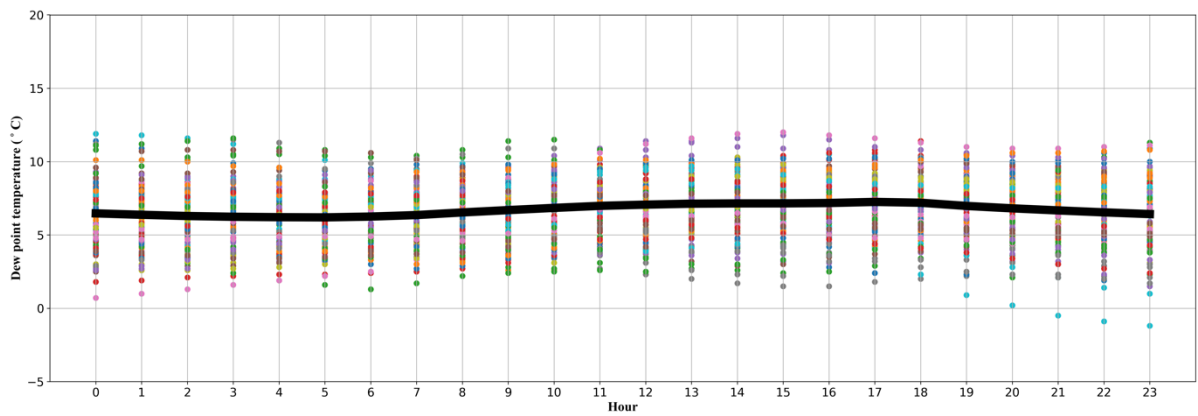
Fig. 9. Clustering result of outdoor air dry-bulb temperature.



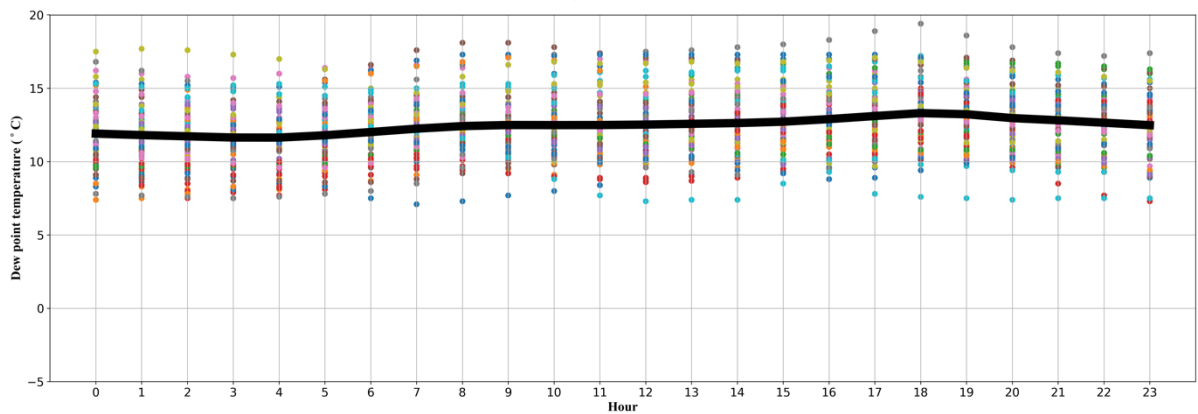
(a) C1



(b) C2

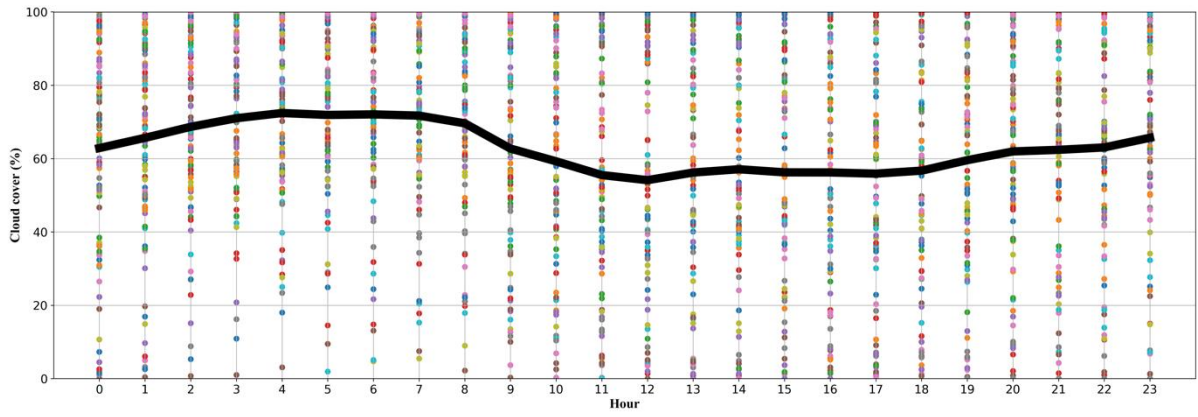


(c) C3

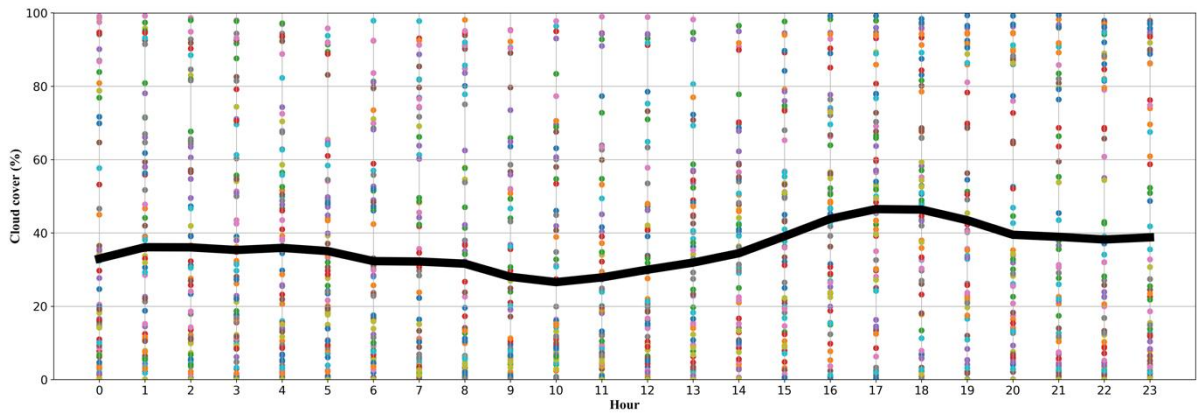


(d) C4

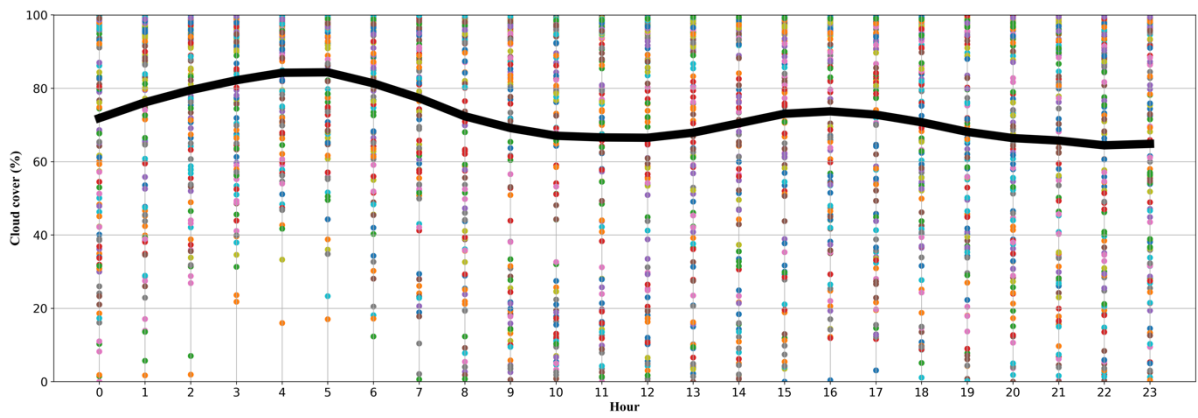
Fig. 10. Clustering result of outdoor air dew-point temperature.



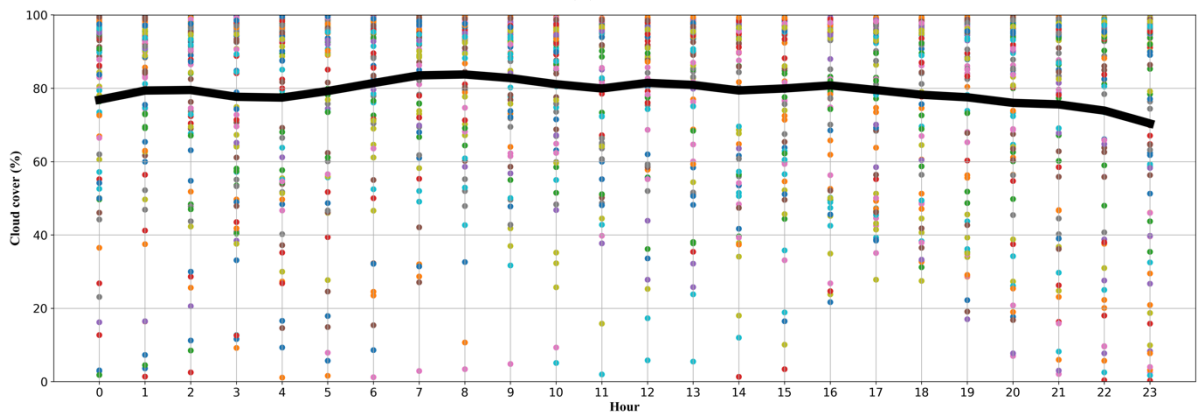
(a) C1



(b) C2

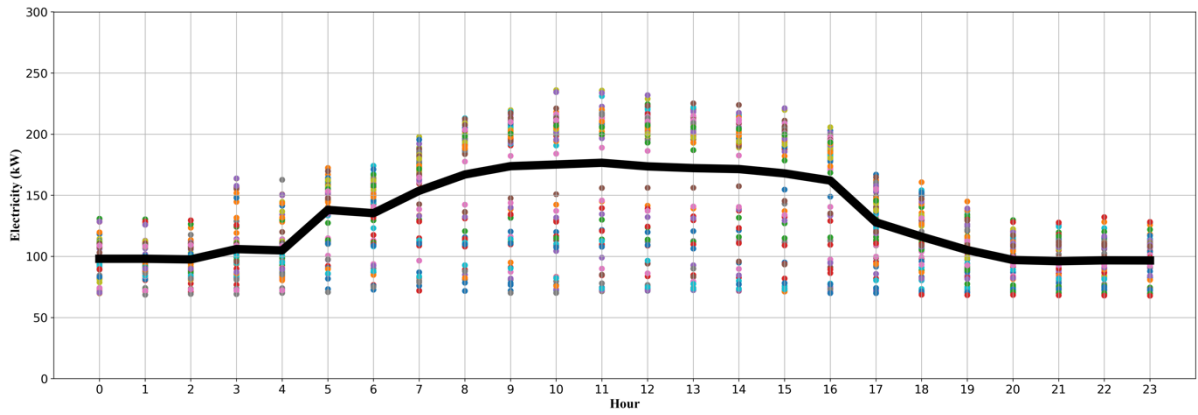


(c) C3

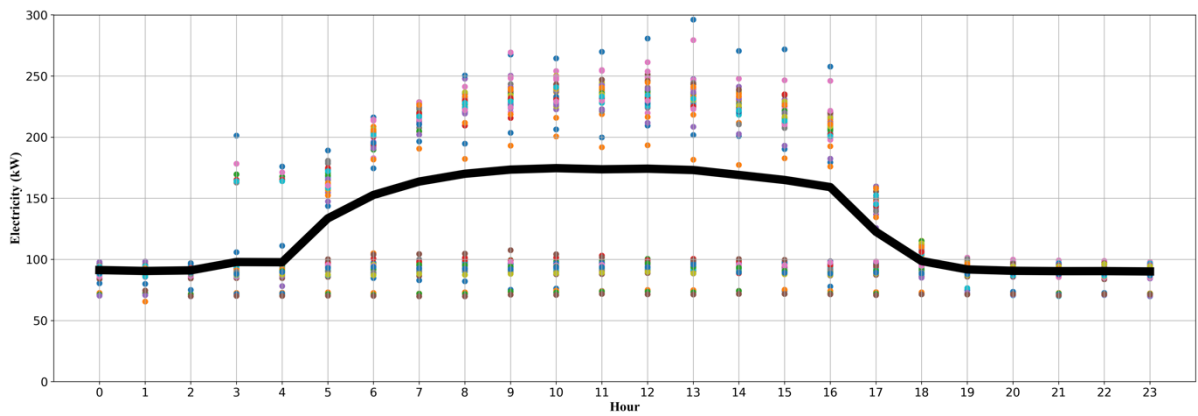


(d) C4

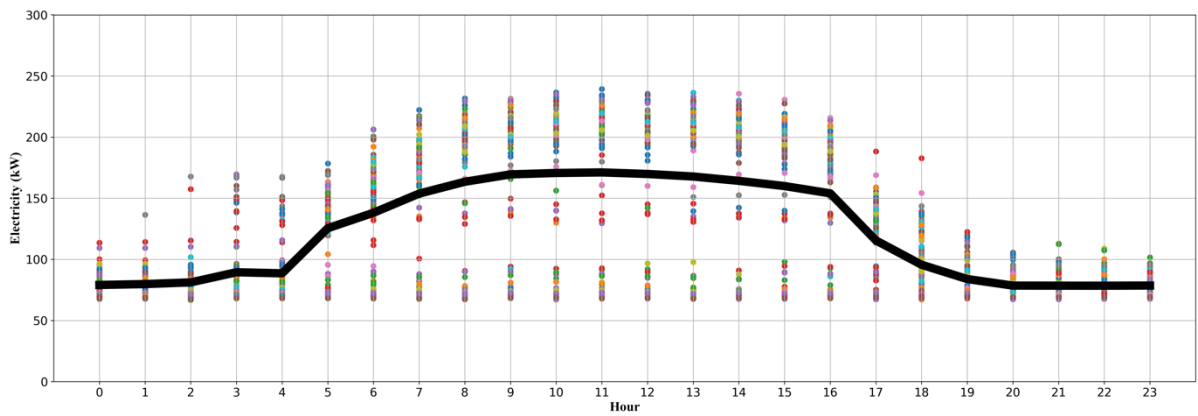
Fig. 11. Clustering result of cloud cover ratio.



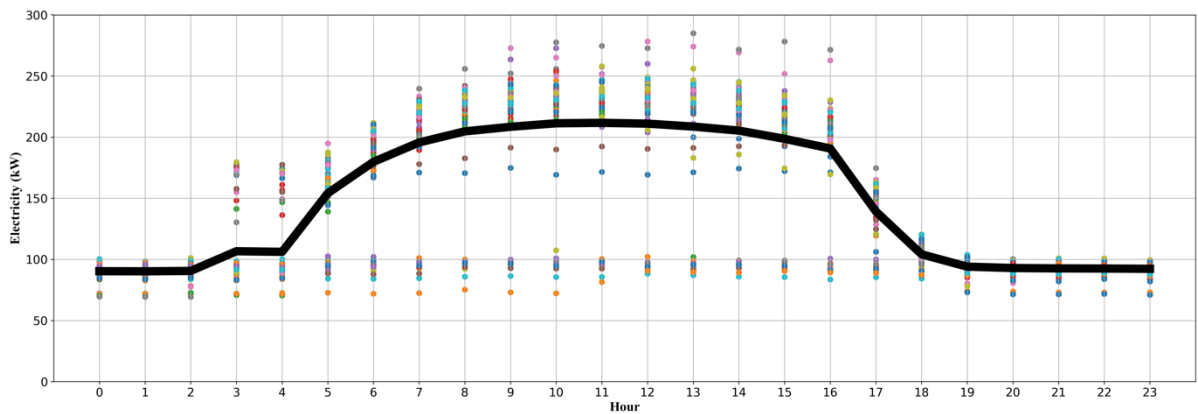
(a) C1



(b) C2



(c) C3



(d) C4

Fig. 12. Clustering result of daily energy consumption.

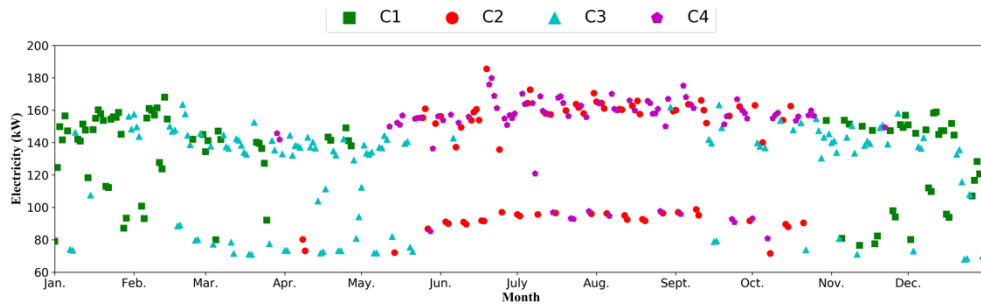


Fig. 13. Daily average energy consumption in each cluster.

4.1.3 Feature identification of weather profiles in each cluster

The featuring curve of outdoor air dry-bulb temperature, outdoor air dew-point temperature and cloud cover ratio during the 24-hour duration are summarized in Figs. 9-11, respectively. For both outdoor air dry-bulb temperature and dew-point temperature, the lowest value is identified in cluster C1, the middle value is found in cluster C3, while clusters C2 and C4 share the similar highest values. However, for cloud cover ratio, the lowest values are identified in cluster C2, while the highest average values are found in cluster C4. Moreover, the featuring daily energy consumption profiles are identified, as illustrated in Fig. 12. The cluster C3 has the lowest values while cluster C4 has the highest values.

The daily average energy consumption is presented in Fig. 13. It is seen that clusters C2 and C4 have relatively higher average daily energy consumption than clusters C1 and C3 do. The higher values in each cluster indicate the weekdays while the lower ones stand for the weekends and public holidays. To summarize the above discussion, the featuring patterns identified in each cluster is summarized in Table 5.

Table 5. Summary of featuring patterns in each cluster.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Dry-bulb temperature	Lowest value occurred time (h)	4			
	Valley value (°C)	1	11	7	12
	Highest value occurred time (h)	14			
	Peak value (°C)	6	20.5	12.5	20
Dew-point temperature	Lowest value occurred time (h)	5	4	4	3
	Valley value (°C)	1	11.5	6.5	12
	Highest value occurred time (h)	14	19	17	18
	Peak value (°C)	2.5	12.5	7.5	14
Cloud cover ratio	Lowest value occurred time (h)	12	10	22	23
	Valley value (%)	54	28	66	72
	Highest value occurred time (h)	4	17	5	8
	Peak value (%)	74	48	84	84
Energy consumption	Lowest value occurred time (h)	20-23	20-23	20-23	1
	Valley value (kW)	95	90	87.5	87
	Highest value occurred time (h)	11	10-12	10-11	10-12
	Peak value (kW)	180	176	170	212.5

4.1.4 Correlation coefficient analysis

The correlation coefficients of weather profile to energy consumption is summarized in Table 6. In most cases, the correlation coefficients in clusters C1-C4 are larger than those of year-round database. In Cluster C1, outdoor air dew-point temperature and cloud cover ratio have the largest effect on energy consumption. In Cluster C4, outdoor air dry-bulb temperature plays the predominate role in determining energy consumption.

Table 6. Correlation coefficient of weather profile to energy consumption.

Datasets	Dry-bulb temperature	Dew-point temperature	Cloud cover ratio
Cluster C1	0.170	-0.131	-0.132
Cluster C2	0.371	0.022	-0.057
Cluster C3	0.299	0.071	0.004
Cluster C4	0.448	0.120	0.045
Year-round database	0.244	0.078	-0.005

4.2 GA performance in determining the optimal DNN architecture

GA optimization is adopted to determine the optimal architecture of each DNN sub-model, including the quantity of hidden layers, quantity of neurons in each layer, activation function and training approach. The convergence performance of GA optimization is illustrated in Fig. 14. The convergence is achieved after 19, 19, 18 and 23 iterations for clusters C1, C2, C3 and C4, respectively. Meanwhile, the convergence of the reference GA-DNN model is achieved after 22 iterations.

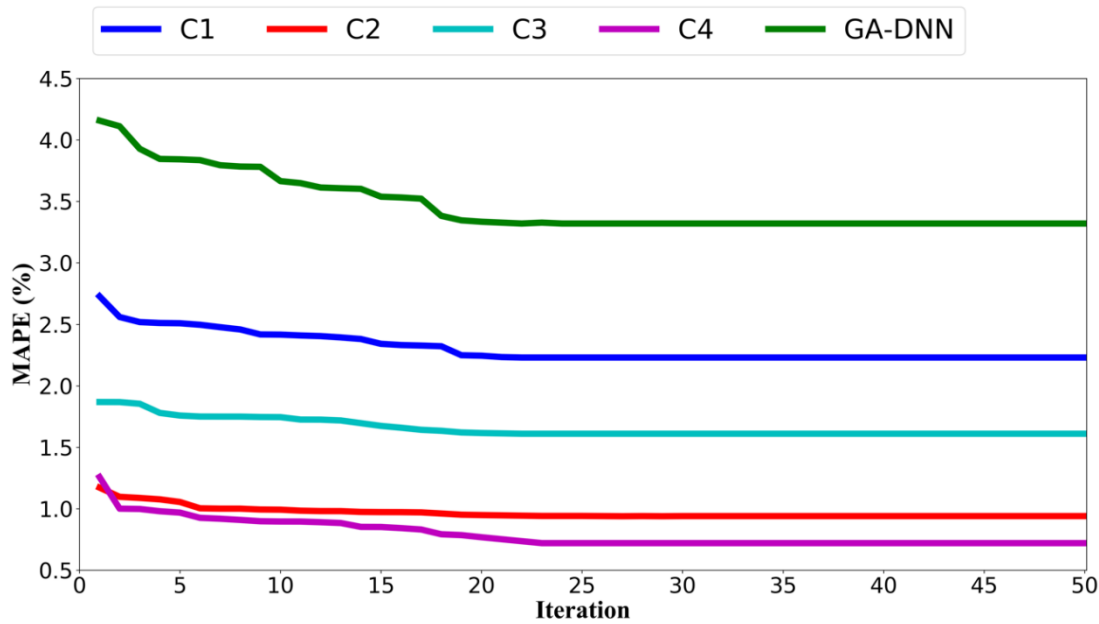


Fig. 14. Convergence performance of GA optimization based DNN model adjusting.

After GA optimization, the optimal architecture of each DNN model can be determined, as summarized in Table 7. The optimal quantity of hidden layers is 10, 10, 10, 9 and 10 for clusters C1, C2, C3, C4 in the proposed FE-GA-DNN model and the reference GA-DNN model, respectively, while different quantity of neurons are chosen in each layer. Rectified linear unit is chosen as the activation function for the proposed FE-GA-DNN models while exponential linear is selected for the reference GA-DNN model. For training approach in determining the weighting factors, NADAM, RMSPROP, ADAMAX, ADAM and ADAM are selected for clusters C1, C2, C3, C4 and the GA-DNN model, respectively.

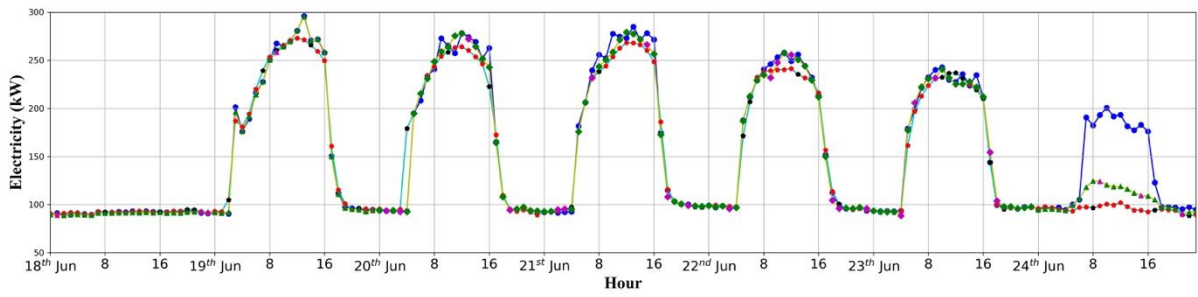
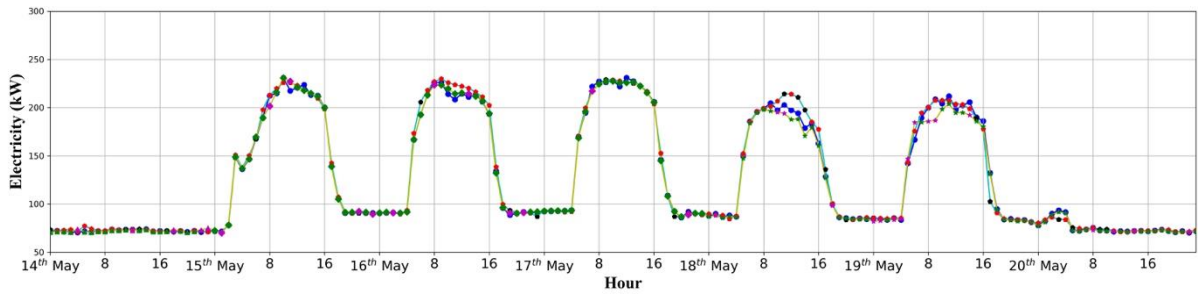
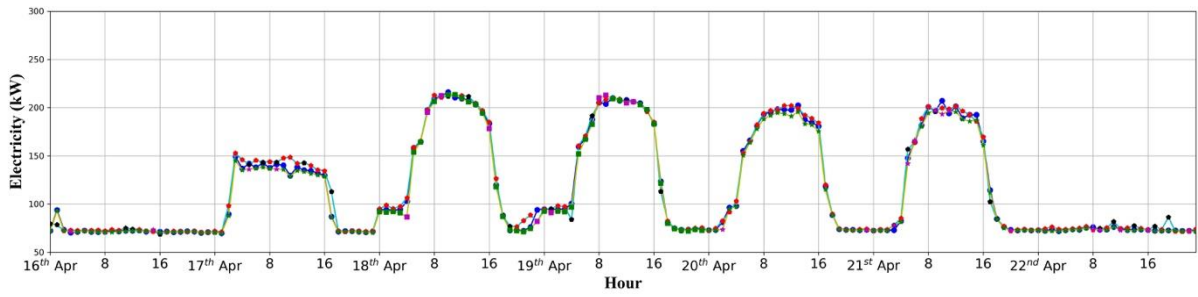
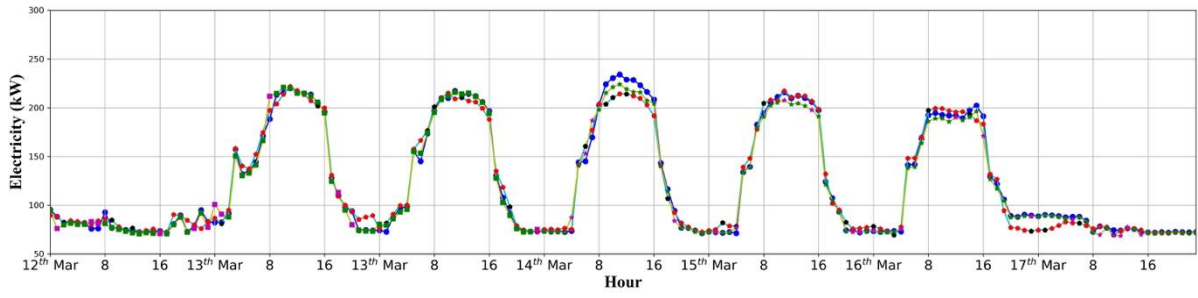
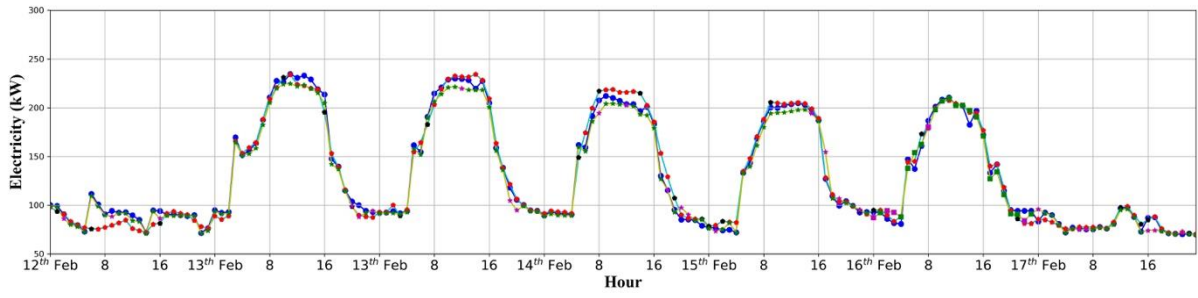
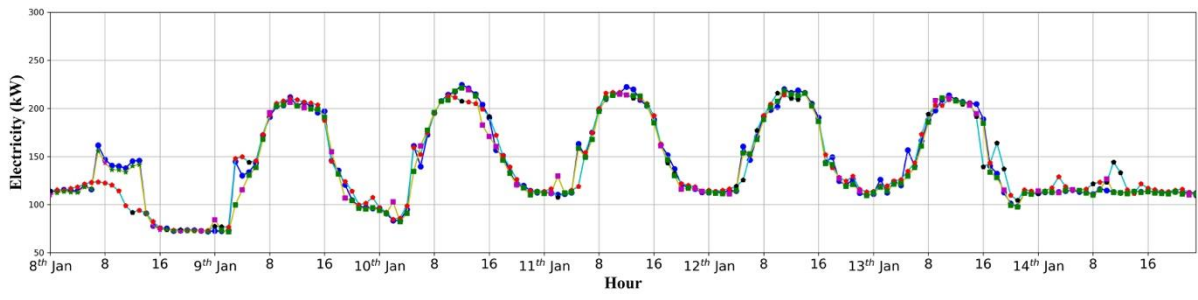
Table 7. Optimal architecture of each DNN model.

		Optimal quantity of neurons in each layer	Optimal layers	Activation function	Optimization approach
FE-GA-DNN model	C1	[70, 100, 100, 70, 100, 100, 90, 70, 90, 70]	10	elu	NADAM
	C2	[80, 100, 60, 80, 80, 80, 90, 20, 60, 60]	10	elu	rmsprop
	C3	[80, 100, 90, 70, 50, 40, 10, 100, 100, 100, 10]	10	elu	Adamax
	C4	[70, 70, 100, 90, 60, 40, 70, 100, 60, 80]	9	elu	ADAM
GA- DNN model		[70, 100, 80, 50, 100, 70, 40, 90, 90, 40]	10	relu	ADAM

4.3. Performance of energy consumption prediction

To take a deep look into performance of the weekly energy consumption prediction, the prediction results from 12 weeks in each month are presented. It is seen that the proposed FE-GA-DNN model is better at tracking the peak value, valley value, sudden variation as well as the off-day energy consumption profile than that of the reference GA-DNN model. For example:

- During the daytime on 13th Feb, 14th Feb, 18th May, 21st June, 22nd June, 10th July, 23rd Aug, 24th Aug, and 14th Sept, the predicted result shows that the proposed FE-GA-DNN model is better at tracking the peak value of energy consumption than that of the reference GA-DNN model;
- During the night-time on 14th Jan, 13th Mar, 19th Apr and 24th Nov, it is found that the proposed FE-GA-DNN model is better at tracking the valley value of energy consumption than that of the reference GA-DNN model;
- At the 6th h on 9th Jan, 10th Jan, 11th Jan, 12th Jan, at the 11th h on 16th May, at the 13th h on 19th Jun, during the 9-11th h on 9th Oct, the proposed FE-GA-DNN model is better at tracking the sudden change of energy consumption than that of the reference GA-DNN model;
- On 8th Jan 12th Feb, 25th to 29th Dec, which are either weekends or public holidays, the proposed FE-GA-DNN model is better at tracking the relatively constant but low energy consumption than that of the reference GA-DNN model.



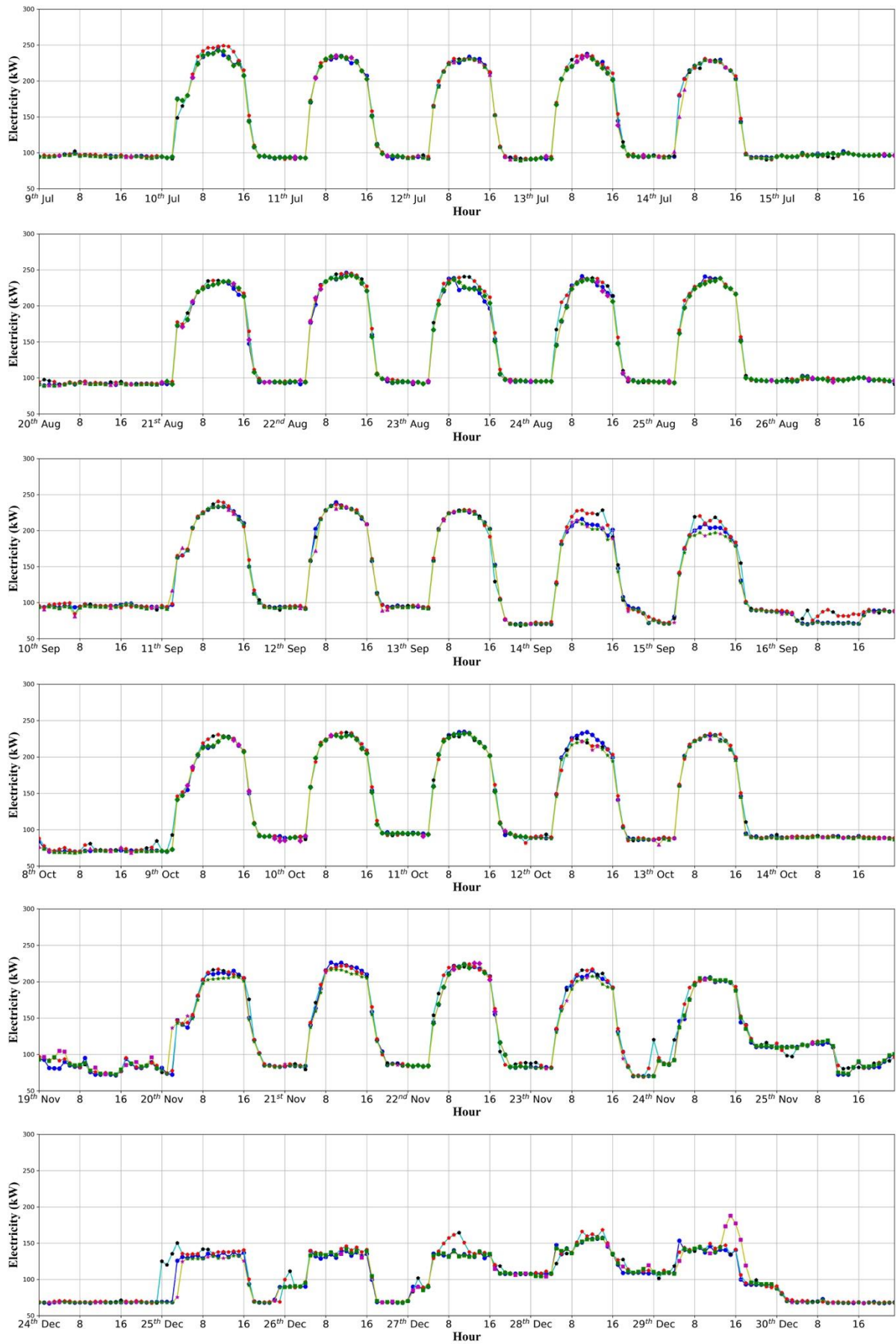
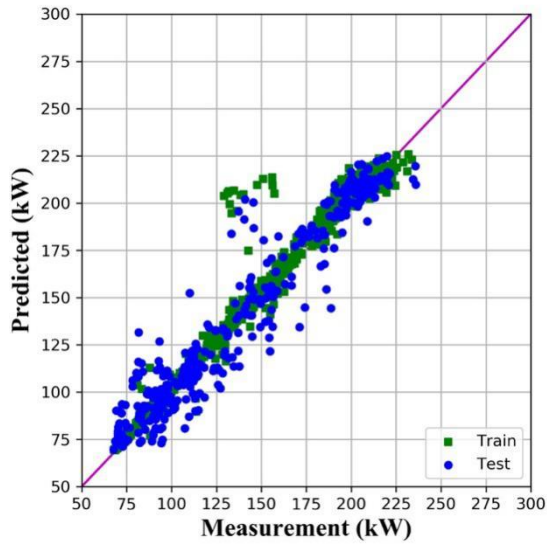
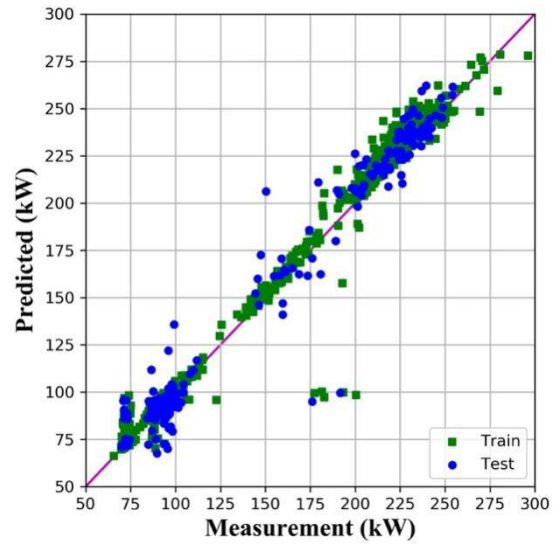


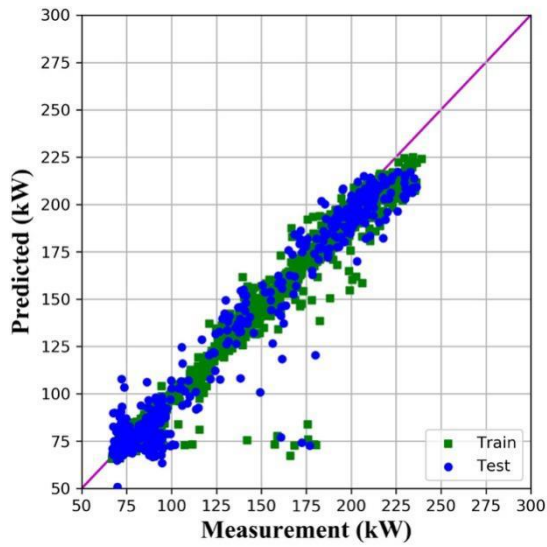
Fig. 15. Prediction results of weekly energy consumption in each month



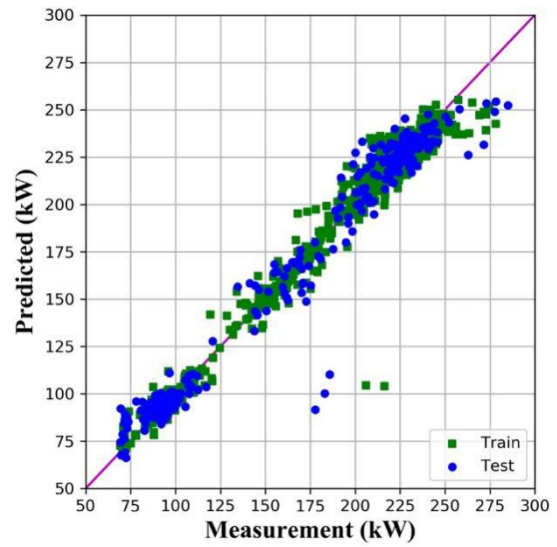
(a) C1 in FE-GA-DNN model



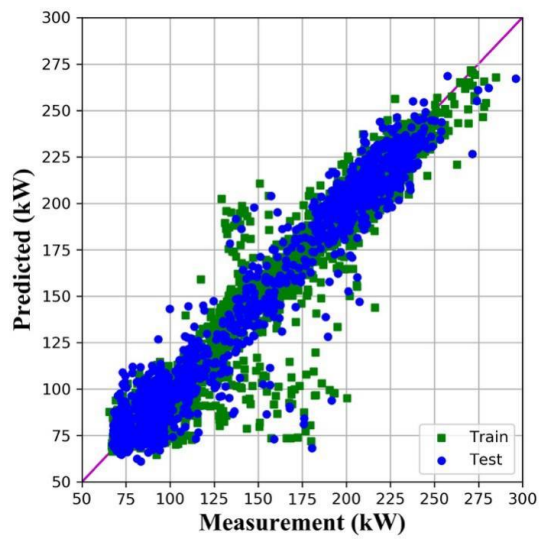
(b) C2 in FE-GA-DNN model



(c) C3 in FE-GA-DNN model



(d) C4 in FE-GA-DNN model



(e) GA-DNN

Fig. 16. Measurement and predicted value of energy consumption.

The measurement and predicted value of energy consumption in each DNN sub-model in the proposed FE-GA-DNN model and in the reference GA-DNN model is shown in Fig. 16. It is seen that the data points in clusters C1-C4 are more concentrated on the diagonal, while there exists a large quantity of data points on the right-hand side of the diagonal in the reference GA-DNN model.

To quantify the prediction accuracy of the proposed FE-GA-DNN model, the values of *MAPE*, R^2 , *RMSE* and *MPE* for both the training and testing cases are evaluated, as shown in Table 8. For both training and testing data, in each cluster of datasets, the proposed FE-GA-DNN model shows better prediction performance than that of the single GA-DNN model except for the testing case in cluster C1. Therefore, the overall performance of the proposed FE-GA-DNN model was better than that reference GA-DNN model, with 42.5%, 0.71%, 36.5 and 37.7% in *MAPE*, R^2 , *RMSE* and *MPE* in training case, while 15.9%, 0.21%, 13.0% and 9.28% in testing case, respectively.

Table 8. Prediction performance of the predictive models.

Performance index	FE-GA-DNN model										GA-DNN model	
	C1		C2		C3		C4		Overall		Train	Test
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test		
<i>MAPE</i> (%)	2.23	7.21	0.94	3.44	1.61	5.40	0.72	2.62	1.43	4.83	2.49	5.74
r^2	0.989	0.923	0.991	0.973	0.994	0.959	0.998	0.992	0.993	0.960	0.986	0.958
<i>RMSE</i> (kW)	5.05	13.27	5.70	10.05	4.22	11.43	2.54	5.54	4.33	10.29	6.82	11.83
<i>MPE</i> (%)	2.88	8.99	1.34	4.75	2.49	6.71	1.20	3.96	2.07	6.26	3.32	6.90

5. Implication for practice and future direction

To generate the database for training and testing cases for the proposed FE-GA-DNN model, the year-round historical weather data is collected from the local weather station while the electricity consumption data can be obtained from the building energy management system. After the proposed FE-GA-DNN model is well trained using at least one year's historical database, the week-ahead hourly electricity consumption can be predicted using the latest forecast of weather profile acquired from the weather reporting website [40-42]. Therefore, the proposed FE-GA-DNN model can be adopted in the building management system to forecast the week-ahead hourly electricity consumption thus to determine the operating schedules of various energy supply equipment units. The proposed FE-GA-DNN model can also be further refined to tackle the problem caused by the probable faulty measurement data owing to sensor and equipment faults.

In this study, *k*-means clustering technique is adopted to group the daily weather profile into different clusters to realise feature extraction purpose. The performance of other types of clustering algorithms in feature identification, such as connectivity-based clustering, centroid-based clustering and density-

based clustering should be evaluated. Moreover, it is worthwhile to investigate whether other evolutionary optimization algorithms, such as particle swarm optimization, ant colony optimization and artificial bee colony algorithm would show better performance than the GA optimization in determining the optimal architecture of each DNN sub-model. In addition, it is valuable to explore the abilities of other deep learning algorithms in electricity consumption prediction, such as RNN, CNN and LSTM neural networks.

6. Conclusion

To improve the accuracy, effectiveness and robustness of building energy consumption prediction, the feature extraction and genetic algorithm enhanced adaptive deep neural network predictive model is proposed in this study. The uniqueness of the proposed prediction model is its integration of three artificial intelligence approaches, including clustering technique for feature extraction, genetic algorithm for architecture optimization as well as deep neural network for energy prediction. In other words:

- *k*-means clustering is used to group the year-round daily weather profiles into different clusters thus to extract the representative features from each group. Thereafter, datasets in each cluster can be used to determine the optimal architecture and weighting factors of one DNN sub-model.
- GA is adopted to determine the optimal architecture of each DNN sub-model, including quantity of hidden layers, quantity of neurons in each layer, activation algorithm and training approach.
- DNN model, with multiple hidden layers and variable quantity of neurons in each layer, is adopted to reveal the comprehensive relationship among weather data, time signature and energy consumption.
- Since the architecture and weighting factors of each DNN sub-model is determined by the featuring patterns of daily weather profile in each cluster, the DNN sub-model is adaptive to the featuring input datasets. Therefore, the overall prediction performance can be improved compared to the single DNN model with fixed architecture. Furthermore, the measurement from energy management system in a practical office building is adopted, which provides a real-life scenario.

To test the effectiveness of the proposed FE-GA-DNN model, one-year measurement of weather data from the local weather station and the energy data from the building management system of a real office building are formulated as historical database. It can provide a real-life scenario. The main findings from this study are summarized below:

- According to the features of daily outdoor air dry-bulb temperature, outdoor air dew-point temperature and cloud cover ratio, the 365 sets of daily weather profiles can be grouped into four

clusters. The various featuring peak and valley values of dry-bulb temperature, dew-point temperature and cloud cover ratio are identified in each cluster, along with its occurring duration.

- Although similar optimal quantity of hidden layers (i.e. 9 or 10) and the same activation function (i.e. exponential linear) are identified for the four DNN sub-models, the optimal quantity of neurons in each layer and the optimal training approach for determining the weighting factors are different.
- Compared to the reference GA-DNN predictive model with fixed architecture, the proposed FE-GA-DNN model shows better performance at tracking the peak value, valley value and sudden variation of the energy consumption in working days as well as the relatively constant and low energy consumption during the weekend and public holidays.
- The overall *MAPE*, R^2 , *RMSE* and *MPE* is 1.43%, 0.993, 4.33 kW and 2.07 kW for training case, respectively; while 4.83 %, 0.960, 10.29 kW and 6.26 kW for testing case, respectively. Therefore, compared to the reference GA-DNN model, it has 42.5%, 0.71%, 36.5 and 37.7% improvement in *MAPE*, R^2 , *RMSE* and *MPE* in training case, while 15.9%, 0.21%, 13.0% and 9.28% in *MAPE*, R^2 , *RMSE* and *MPE* in testing case, respectively.

Acknowledgement

The authors would like to acknowledge and express their sincere gratitude to The Department for Business, Energy & Industrial Strategy through grant project number TEIF-101-7025. Opinions expressed and conclusions arrived at are those of the authors and are not to be attributed to BEIS.

References

- [1] Luo XJ and Fong KF. Development of multi-supply-multi-demand control strategy for combined cooling, heating and power system primed with solid oxide fuel cell-gas turbine. *Energy Conversion and Management*. 154(2017)538-561.
- [2] Luo XJ and Fong KF. Development of integrated demand and supply side management strategy of multi-energy system for residential building application. *Applied Energy*. 242(2019)570-587.
- [3] Luo XJ, Lukumon O, Anuoluwapo A, Chukwuka M, Olugbenga A and Lukman A. Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands. *Advanced Engineering Informatics*. 41(2019)100926.
- [4] Zhao HX and Frédéric M. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*. 16(2012)3586-3592.
- [5] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J and Zhao X. A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*. 82(2018)1027-1047.

- [6] Amasyali Kadir and Nora M. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*. 81(2018)1192-1205.
- [7] Yang J, Hugues R and Radu Z. On-line building energy prediction using adaptive artificial neural networks. *Energy and buildings*. 37(2005)1250-1259.
- [8] Ahmad T, Chen H, Shair J and Xu C. Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management. *International Journal of Refrigeration*. 98(2019)399-409.
- [9] Ahmad T and Chen HX. Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. *Energy and Buildings*. 166(2018) 460-476.
- [10] Wang L, Eric WML and Richard KKY. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. *Applied Energy*. 228(2018)1740-1753.
- [11] Katsatos AL, Moustris KP. Application of artificial neuron networks as energy consumption forecasting tool in the building of regulatory authority of energy, Athens, Greece. *Energy Procedia*. 157(159)851-861.
- [12] M.A.R. Biswas, M.D. Robinson, N. Fumo. Prediction of residential building energy consumption: a neural network approach. *Energy*, 117 (2016), pp. 84-92,
- [13] Singaravel, S., Suykens, J. and Geyer, P., 2018. Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction. *Advanced Engineering Informatics*, 38, pp.81-90.
- [14] Ahmad, J., Larijani, H., Emmanuel, R., Mannion, M., Javed, A. and Phillipson, M., 2017, April. Energy demand prediction through novel random neural network predictor for large non-domestic buildings. In *2017 Annual IEEE International Systems Conference (SysCon)* (pp. 1-6). IEEE.
- [15] Kim, T.Y. and Cho, S.B., 2019. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, 182, pp.72-81.
- [16] Fan, C., Wang, J., Gang, W. and Li, S., 2019. Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. *Applied energy*, 236, pp.700-710.
- [17] Kusiak A, Li MY and Zhang ZJ. A data-driven approach for steam load prediction in buildings. *Applied Energy*. 87(2010)925-933.
- [18] Jovanović RŽ, Sretenović AA and Živković BD. Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*. 94(2015)189-199.
- [19] Deb C, Eang LS, Yang J and Santamouris M. Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy and Buildings*. 121(2016)284-297.
- [20] A. Bagnasco, F. Fresi, M. Saviozzi, F. Silvestro, A. Vinci. Electrical consumption forecasting in hospital facilities: an application case. *Energy Build.*, 103 (2015), pp. 261-270.
- [21] R. Mena, F. Rodríguez, M. Castilla, M.R. Arahal. A prediction model based on neural networks for the energy consumption of a bioclimatic building. *Energy Build.*, 82 (2014), pp. 142-155.

- [22]K Li, X Xie, W Xue, X Dai, X Chen, X Yang. A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. *Energy Build*, 174 (2018)323-334.
- [23]Somu N, MR GR and Ramamritham K. A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy*, 261(2020)114131.
- [24]Ruiz LGB, Rueda R, Cuéllar MP and Pegalajar MC. Energy consumption forecasting based on Elman neural networks with evolutive optimization. *Expert Systems with Applications*, 92 (2018)380-389.
- [25] Z. Wu, X. Zhao, Y. Ma, X. Zhao. A hybrid model based on modified multi-objective cuckoo search algorithm for short-term load forecasting. *Appl Energy*, 237 (2019), pp. 896-909.
- [26]Y Liang, D Niu, WC Hong. Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy*, 166 (2019)653-663.
- [27]Ding Y, Zhang Q, Yuan T and Yang F. Effect of input variables on cooling load prediction accuracy of an office building. *Applied Thermal Engineering*. 128(2018)225-234.
- [28]D Zhao, M Zhong, X Zhang, X Su. Energy consumption predicting model of VRV (variable refrigerant volume) system in office buildings based on data mining. *Energy*, 102 (2016)660-668.
- [29]Gatt D, Yousif C, Cellura M, Camilleri L and Guarino F. Assessment of building energy modelling studies to meet the requirements of the new Energy Performance of Buildings Directive. *Renewable and Sustainable Energy Reviews*, 127(2020)109886.
- [30]Iddio E, Wang L, Thomas Y, McMorro G and Denzer A. Energy efficient operation and modeling for greenhouses: A literature review. *Renewable and Sustainable Energy Reviews*, 117(2020)109480.
- [31]Moradikazerouni A, Hajizadeh A, Safaei MR, Afrand M, Yarmand H and Zulkifli N. Assessment of thermal conductivity enhancement of nano-antifreeze containing single-walled carbon nanotubes: Optimal artificial neural network and curve-fitting. *Physica A: Statistical Mechanics and its Applications*, 521(2019)138-145.
- [32]Pallonetto F, Mattia DR, Francesco DE and Donal PF. On the assessment and control optimisation of demand response programs in residential buildings. *Renewable and Sustainable Energy Reviews*, 127(2020)109861.
- [33]Li W, Koo C, Hong T, Oh J, Cha SH and Wang S. A novel operation approach for the energy efficiency improvement of the HVAC system in office spaces through real-time big data analytics. *Renewable and Sustainable Energy Reviews*, 127(2020)109885.
- [34] Pappi IN, Paterakis NG, Catalão JP, Panapakidis I and Papagiannis G. Analysis of the energy usage in university buildings: The case of aristotle university campus. In 2015 Australasian Universities Power Engineering Conference (AUPEC) (pp. 1-5). IEEE.
- [35] AbuBaker M. Data Mining Applications in Understanding Electricity Consumers' Behavior: A Case Study of Tulkarm District, Palestine. *Energies*. 12(2019)4287.

- [36] Luo XJ, Fong KF, Sun YJ and Leung MKH. Development of clustering-based sensor fault detection and diagnosis strategy for chilled water system. Energy and Buildings. Energy and Buildings. 186(2019)17-36.
- [37] S. Haykin. Neural networks: a comprehensive foundation. Prentice Hall (1999)
- [38] K. Gurney. An introduction to neural networks. Routledge, London (1997)
- [39] Goldberg D. Genetic Algorithms in Search, Optimization, and Machine Learning Addison Wesley, Reading, Massachusetts, 1989.
- [40] <https://www.accuweather.com>
- [41] <https://weather.com/en-GB/>
- [42] <https://www.metoffice.gov.uk/>