Energy and AI 2 (2020) 100015

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.elsevier.com/locate/egyai

Genetic algorithm-determined deep feedforward neural network architecture for predicting electricity consumption in real buildings

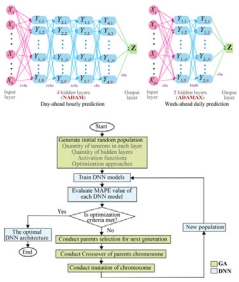
X.J. Luo[a], Lukumon O. Oyedele[a,*], Anuoluwapo O. Ajayi[a], Olugbenga O. Akinade[a], Juan Manuel Davila Delgado[a], Hakeem A. Owolabi[a], Ashraf Ahmed[b]

[a] Big Data Enterprise and Artificial Intelligence Laboratory (Big-DEAL), University of the West of England (UWE), Frenchay Campus, Bristol, United Kingdom
[b] Department of Civil and Environmental Engineering, Brunel University London, London, United Kindom

HIGHLIGHTS

- A GA-determined DFNN architecture for real building electricity consumption.
- Genetic algorithm to determine the optimal architecture of DFNN model.
- Both day-ahead hourly and week-ahead daily electricity consumption are considered.
- One year and six months of measurement data from a real campus building is used for training and testing the proposed GA-DFNN model.
- The proposed model has better performance than ANN, LSTM and TCN.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 30 April 2020
Received in revised form 16 June 2020
Accepted 30 June 2020
Available online 7 July 2020

Keywords:
Prediction
Deep learning
Feedforward neural network
Genetic algorithm
Electricity consumption

ABSTRACT

A genetic algorithm-determined deep feedforward neural network architecture (GA-DFNN) is proposed for both day-ahead hourly and week-ahead daily electricity consumption of a real-world campus building in the United Kingdom. Due to the comprehensive relationship between affecting factors and real-world building electricity consumption, the adoption of multiple hidden layers in the deep neural network (DFNN) algorithm would improve its prediction accuracy. The architecture of a DFNN model mainly refers to its quantity of hidden layers, quantity of neurons in the hidden layers, activation function in each layer and learning process to obtain the connecting weights. The optimal architecture of DFNN model was generally determined through a trial-and-error process, which is an exponential combinatorial problem and a tedious task. To address this problem, genetic algorithm (GA) is adopted to automatically design an optimal architecture with improved generalization ability. One year and six months of measurement data from a campus building is used for training and testing the proposed GA-DFNN model, respectively. To demonstrate the effectiveness of the proposed GA-DFNN prediction model, its prediction performance, including mean absolute percentage error, coefficient of determination, root mean square error and mean absolute error, was compared to the reference feedforward neural network models with single hidden layer, DFNN models with other architecture, random search determined DFNN model, long-short-term-memory model and temporal convolutional network model. The comparison results show that the proposed GA-DFNN predictive model has superior performance than all the reference prediction models, demonstrating the optimization effectiveness of GA and the prediction effectiveness of DFNN model with multiple hidden layers and optimal architecture.

* Corresponding author.
E-mail addresses: Xiaojun.Luo@uwe.ac.uk (X.J. Luo)., L.Oyedele@uwe.ac.uk, Ayolook2001@yahoo.co.uk (L.O. Oyedele)..

https://doi.org/10.1016/j.egyai.2020.100015

Nomenclature

d	Day of the week
E	Electricity consumption
f	Activation function
h	Hour of the day
m	Month of the year
$MAPE$	Mean absolute percentage error
MPE	Mean percentage error
n	Quantity of hidden layer
N	Quantity
R^2	Coefficient of determination
RMSE	Root mean square error
T	Total training samples or temperature
W	Set of weighting factors
x	Neuron in the input layer
X	Input dataset
Y	Neuron in the hidden layer
z	Neuron in the output layer
$\hat{Z}$	Estimated output through the DFNN model
$\mathbf{Z}$	Output dataset

Subscripts

d	Daily
db	Dry-bulb
dew	Dew-point
i	Number of neurons in the input layer
j	Number of neurons in the hidden layer
k	Number of hidden layers
h	Hourly
H	Hidden layer
IN	Input layer
O	Output layer
t	Number of samples, i.e. time step

Abbreviations

ADAM	Adaptive moment estimation
ANN	Artificial neuron network
elu	Exponential linear function
DFNN	Deep feedforward neural network
FNN	Feedforward neural network
GA	Genetic algorithm
GA-DFNN	Genetic algorithm determined deep feedforward neural network architecture
LSTM	Long-term-short-memory
NADAM	Nesterov-accelerated Adaptive Moment Estimation
PSO	Particle swarm optimization
relu	Rectified linear unit function
SGD	Stochastic gradient descent approach
TCN	Temporal convolutional networks

1. Introduction

The rapid economic and population growth are accelerating the consumption of electrical energy [1]. As indicated by World Energy Outlook 2017, the primary energy consumption all over the world was projected to grow at a compound annual growth rate of 1.0% over 2016–2040 [2]. Due to its physical characteristics, electrical energy must be consumed as soon as it is generated in the power plant. Therefore, accurate prediction of electricity demand is essential for stable power supply. On the other hand, due to the fact that people spend a substantial fraction of time in buildings, electricity consumption in buildings is increasing rapidly [3,4]. As a result, it is of significant importance to accurately predict electricity consumption and facilitate the building energy system management.

1.1. Literature review

With the advent of the era of big data, buildings have become not only energy-intensive but also data-intensive. Data-mining technologies have been widely utilised to investigate the values of massive amounts of building operation data with the aim of improving the operating performance of building energy systems [5]. As revealed by the review papers [6–9], various artificial intelligence techniques have been applied in forecasting building energy consumption, including multiple linear regression, statistical regression, decision tree, autoregressive integrated moving average model, support vector regression, neural network and their ensemble models. Due to its ability to approximate the complicated nonlinear relationship between the input and the output datasets of a complicated system with arbitrary and precision, neural networks have been widely adopted to investigate the performance of building energy.

The neural network models mainly include feedforward neural network (FNN), convolution neural network and recurrent neural network. The FNN with single hidden layer is the most popular type of neural networks in building energy analysis. Luo et al. [10,11] proposed an FNN prediction model for building heating and cooling demands. The quantity of neurons in the hidden layer was chosen through sensitivity analysis in the range of 60–80 while rectified linear unit was adopted as the activation function in both the hidden and output layer. Singh et al. [12] proposed an FNN predictive model with single hidden layer for electricity consumption in urban area. In the hidden layer, there were 20 neurons while the sigmoid activation function was adopted. Mena et al. [13] proposed two FNN predictive models for electricity consumption prediction in a bioclimatic building. The quantity of neurons in the input and hidden layer was 17 and 10 in the complete FNN model, while 6 and 15 in the particular FNN model. The hyperbolic tangent was adopted as activation function in both FNN models. Kusiak et al. [14] evaluated the performance of an FNN predictive model for building steam load prediction with different architecture. The quantity of neurons in the single hidden layer was tested in the range between 4 and 9. As for activation functions, logistic, hyperbolic tangent and exponential functions were tested in the hidden layer while sine, identity and logistic functions were tested at the output layer, respectively. Deb et al. [15] proposed a neural network predictive model for cooling energy consumption in an institutional building using the feedforward structure. The optimal quantity of neurons in the hidden layer, 20, was determined through a trial-and-error process. The sigmoid activation function was adopted in both the hidden and output layers. Ahmad et al. [16] proposed three FNN predictive models using the multiple linear regression, Gaussian process regression and Levenberg–Marquardt backpropagation approach, respectively. There was only one hidden layer in each ANN model, while the different quantities of neurons in the hidden layer were tested at 10, 15 and 20. Yang et al. [17] proposed an FNN model for energy consumption prediction in an office building in Canada. The quantity of neurons in the hidden layer was determined using the empirical equation in [18], which is one more than twice the quantity of input neurons. Sigmoid and linear activation function was adopted in the hidden and output layer, respectively. Wang et al. [19] proposed an FNN model for cooling load in a high-rise office building in Hong Kong. The quantity of neurons in the hidden layer was decided through the empirical equation in [20], which is the total value of quantity of input and output neurons and the square root of training samples. Hyperbolic tangent and linear activation function were adopted in the hidden and output layer, respectively. Bagnasco et al. [21] proposed an FNN model for predicting electricity consumption in a hospital building. To find out the optimal quantity of neurons in the hidden layer, it was tested in the range of 6–33, while the best value was found to be 10 and 30 in two different cases. Moreover, different activation functions were tested in the hidden and output layers, with the combination of logsig/tansig, tansig/purelin, logsig/purelin and tansig/tansig. The tansig activation function was found to be optimal in both the hidden and output layers. Muhammad et al. [22] compared the prediction performance between

an FNN and a random forest model for energy consumption of air conditioning system in a hotel. Sensitivity analysis of the FNN model was conducted for different quantity of neurons in the hidden layer in the range of 10–15. Meanwhile, sigmoid activation function was adopted in both hidden and output layers.

In the above-mentioned FNN predictive models with single hidden layer, Levenberg–Marquardt backpropagation was adopted to obtain the weighting factors in most of the studies. To achieve faster convergence and higher efficiency, evolutionary optimization algorithms, such as teaching-learning algorithm [23], particle swarm optimization (PSO) [24–26] and genetic algorithm (GA) [26], were adopted to automatically adjust weighting factors and threshold values of FNN models. Li et al. [23] proposed a FNN predictive model for building energy consumption and adopted it in two campus buildings located in the USA and China, respectively. The quantity of neurons in the hidden layer was fixed at 20 while hyperbolic tangent activation function was adopted in both the hidden and output layers. To evaluate the effects of channel length, depth, width and air mass flow rate on energetic performance of a building integrated photovoltaic/thermal system, Alnaqi et al. [24] proposed a FNN model. Sensitivity analysis of the FNN model was conducted for different quantity of neurons in the hidden layer within the range of 1–8, while 8 was found to be the optimal one. Hyperbolic tangent and sigmoid activation function were tested in both the hidden and output layers, respectively. PSO was adopted to adjust the weighting factors in the FNN to accelerate its convergence. Li et al. [25] proposed a FNN model for predicting electricity consumption in a library in East China. The quantity of neurons was chosen based on the empirical equation while hyperbolic tangent sigmoid was adopted in the hidden and output layers as activation function. Particle swarm optimization was adopted to search for the optimal weighting factors in FNN. Muralitharan et al. [26] proposed a FNN approach to predict energy demand at the consumer end. The weighting factors of the FNN were automatically adjusted by GA.

Deep feedforward neural network models (DFNN) have deeper architectures, which enable the input data being transformed multiple times before deriving the output. Owing to its multiple hidden layers, DFNN is more appropriate for comprehensive data [27,28]. Marijana et al. [29] adopted the DFNN models with different architecture for energy consumption prediction in public buildings. The architecture included 2 hidden layers with the combination of 75–59, 71–87, 57–63, as well as 3 hidden layers with the combination of 87–99–94, 14–31–28 and 61–70–69. Sigmoid function was adopted as the activation function while adaptive moment estimation optimization algorithm was adopted to obtain the weighting factors. Benda et al. [30] proposed a DFNN model for predicting electricity consumption in Cape Town Control center. There were two different architecture of DFNN model with two hidden layers, one with the combination of 56–15, and the other with 58–20. Torres et al. [31] proposed a DFNN prediction model for electricity consumption in Spain. Random search was adopted to tune the hyper-parameters, including the quantity of hidden layers, quantity of neurons in each hidden layer and the learning rate. The quantity of hidden layers was searched in the range of 1–5 while the quantity of neurons in the hidden layer in the range of 10–100. Park et al. [32] proposed a DFNN model for predicting energy production of a ground source heat pump. Through a trial-and-error process, the architecture of the DFNN model was determined to be 10–5–2. Hyperbolic tangent and linear activation function were adopted in the hidden and output layer, respectively. Lu et al. [33] proposed a DFNN-based predictive model for price and load forecasting in the smart grid. Through several accuracy tests, the selected DFNN model had three hidden layers, with 30, 20 and 10 neurons in each layer. Sigmoid function served as the activation function for each layer, while weight and bias adjustment in each layer was trained by Levenberg–Marquardt backpropagation algorithm.

The results from the above-mentioned research works indicated that the DFNN-based predictive model had higher prediction accuracy than the conventional FNN models with single hidden layer. However, the optimal DFNN architecture is contingent on the type of problem to be solved, such as the quantity of training samples as well as the quantity of input and output neurons. As a result, the full ability of the DFNN-based predictive model should be exploited by optimizing its architecture. Although the architecture of the DFNN models was designed by heuristic trial-and-error approaches in some of the previous research works [31–33], it is time-consuming and might reach the local-optimal architecture of DFNN.

1.2. Research gap and contribution

Based upon the literature review, it is found that the previous research works in building energy prediction have the following limitations:

- In the FNN models with single hidden layer, the quantity of neurons in the hidden layer was chosen by experience [12,13,23], using different empirical equations [17,19,25], through sensitivity tests [10,11,14,16,21,22] and [24] or via trial-and-error process [15]. Although some evolutionary algorithms were hybrid with FNN models to improve the prediction accuracy and speed up the computation convergence, they were mainly focusing on automatically adjusting the weighting factors.
- The DFNN models with multiple hidden layers would be able to reveal a more comprehensive relationship among datasets. However, the architecture of the DFNN model would become more complex with the increase of hidden layers. There is no empirical equations or rule-of-thumb to select the quantity of hidden layers, quantity of neurons in each layer, activation function in each layer as well as learning process for weighting factors. However, when the quantity of hidden layers is too small, the network could result in under-fitting and large prediction errors. Whereas, if the quantity of hidden layers is too large, the network might learn the noise in the training dataset and result in overfitting.
- In previous research on DFNN models, the quantity of hidden layers and the quantity of neurons in each hidden layer were selected based on experience [29,30] or trial-and-error process through enumeration [31–33], which might not be able to achieve the optimal performance or cause expensive computation load. Moreover, the authors did not mention the approach or criteria in selecting the activation functions and learning process.

Therefore, the objectives of this study are summarized as follows:

- Due to the complex affecting factors and building electricity consumption, the DFNN model with multiple hidden layers is adopted to construct the predictive model to improve the forecasting accuracy in building electricity consumption;
- To guarantee the effectiveness of the DFNN model, GA optimization is utilised to determine its optimal architecture, including the quantity of hidden layers, quantity of neurons in the hidden layers, activation function in both the hidden and output layer as well as the learning process for determining weighting factors;
- To facilitate building energy management, both day-ahead hourly and week-ahead daily electrical energy consumption are forecasted by the proposed GA-determined DFNN (GA-DFNN) prediction model.

The rest of this paper is organized as follows. Section 2 elucidates the proposed GA-DFNN predictive model. Section 3 illustrates the preparation of the historical database. Section 4 discusses the performance of GA optimization and the optimal architecture of DFNN predictive model. Section 5 presents the performance comparison against various reference prediction models. Section 6 clarifies the practical implication and future application, while Section 7 summarizes the conclusions.

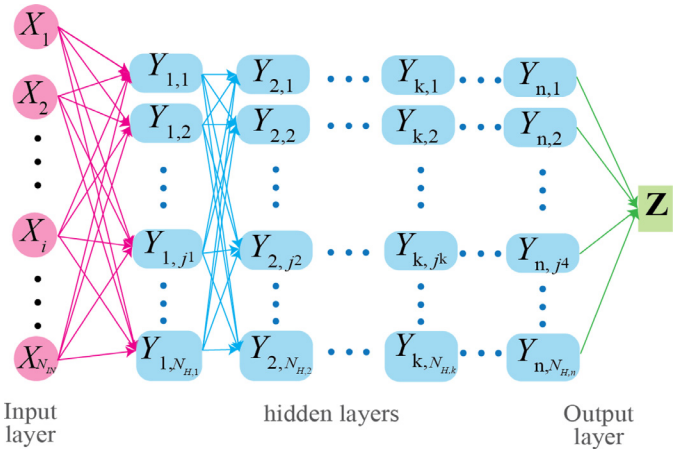


Fig. 1. Diagram of DFNN model.

2. Genetic algorithm-determined deep neuron network architecture

To improve the prediction accuracy of building electricity consumption, the GA-DFNN prediction model is proposed in this study. To be more specific, the GA is collaboratively hybrid with the DFNN algorithm in order to figure out the optimal DFNN architecture.

2.1. Brief theory of the DFNN

The DFNN is a mathematical model which was designed by mimicking the way the human brain processes information. The DFNN models are generally comprised of three principal layers: an input layer, several hidden layers and an outlet layer. The quantity of hidden layers defines the depth of the architecture. With large quantity of hidden layers and quantity of neurons in each hidden layer, DFNN can provide a multi-level representation of dataset. The diagram of a general DFNN model is presented in Fig. 1.

2.1.1. DFNN algorithm

The training database is obtained through the historical data, containing both input dataset $\mathbf{X} = \{X_i | i = 1, 2, \dots, N_{IN}\}$ and output dataset $\mathbf{Z} = \{Z_t | t = 1, 2, \dots, T\}$. The total types of input variables is N_{IN}, while the total training samples is T. For each type of input variables, X_i contains T time samples, and $X_i = \{x_{i,t} | t = 1, 2, \dots, T\}$. There are n hidden layers and $N_{H,k}$ neurons in the k^{th} hidden layer. Therefore, for the t^{th} ($t = 1, 2, \dots, T$) training sample, the j_1^{th} ($j_1 | j_1 = 1, 2, \dots, N_{H,1}$) neuron in the first hidden layer can be calculated as:

$$Y_{1,j_1} = f\left(\sum_{i=1}^{i=N_{IN}} \left(W_{IN,i,1,j_1} X_i\right)\right) \tag{1}$$

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

$$Y_{k,j_k} = f\left(\sum_{j_{k-1}=1}^{j_{k-1}=N_{H,k-1}} \left(W_{k-1,j_{k-1},k,j_k} Y_{k-1,j_{k-1}}\right)\right) \tag{2}$$

The neuron z in the output layer is determined as:

$$\mathbf{Z} = f\left(\sum_{j_n=1}^{j_n=N_{H,n}} \left(W_{n,j_n,O} Y_{n,j_n}\right)\right) \tag{3}$$

where w is the connection weighting factors between the two neurons, i is the number of neurons in the input layer, k is the number of hidden layers, j_k is the number of neurons in the k^{th} hidden layer, and f is the activation function.

2.1.2. Activation function

To achieve better prediction performance, four different activation functions are tested, including sigmoid function, hyperbolic tangent (tanh) function, rectified linear unit (relu) function and exponential linear (elu) function [34]. The four activation functions can be expressed as the follows.

- sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

- hyperbolic tangent function:

$$f(x) = \tanh(x) \tag{5}$$

- rectified linear unit function:

$$f(x) = \begin{cases} 0 \ for \ x < 0 \\ x \ for \ x \geq 0 \end{cases} \tag{6}$$

- exponential linear function:

$$f(x) = \begin{cases} e^x - 1 \ for \ x < 0 \\ x \ for \ x \geq 0 \end{cases} \tag{7}$$

2.1.3. Learning approach

The aim of the learning process of DFNN model is to minimize the mean absolute percentage error $MAPE$ between the produced $\hat{\mathbf{Z}}$ and desired output $\mathbf{Z}$ by adjusting the group of weighting factors $W = \{W_{IN,i,1,j_1} | N_{IN} \geq i \geq 1, \ N_{H,1} \geq j_1 \geq 1\} \cup \{W_{k-1,j_{k-1},k,j_k} | n \geq k \geq 2, \ N_{H,k-1} \geq j_k \geq 1\} \cup \{W_{n,j_n,O} | \ N_{H,k} \geq j_n \geq 1\}$.

In this study, seven different optimization approaches are adopted to train the proposed DFNN model, including stochastic gradient descent (SGD) [35], adaptive moment estimation (ADAM) [36], Nesterov-accelerated adaptive moment estimation (NADAM) [37] and a variant of ADAM based on the infinity norm (ADAMAX) [38].

2.1.4. Performance indicators

To evaluate the performance of the proposed model, four performance indicators are defined, including the mean absolute percentage error, coefficient of determination, mean percentage error as well as root mean square error:

- Mean absolute percentage error

$$MAPE = \frac{1}{T} \sum_{t=1}^{t=T} \frac{|z_t - \hat{z}_t|}{z_t} \times 100\% \tag{8}$$

- Coefficient of determination

$$R^2 = \frac{\left[\sum_{t=1}^{t=T} \left(z_t - \frac{\sum_{t=1}^{t=T} z_t}{T}\right) \cdot \left(\hat{z}_t - \frac{\sum_{t=1}^{t=T} \hat{z}_t}{T}\right)\right]^2}{\sum_{t=1}^{t=T} \left(z_t - \frac{\sum_{t=1}^{t=T} z_t}{T}\right) \cdot \sum_{t=1}^{t=L} \left(\hat{z}_t - \frac{\sum_{t=1}^{t=T} \hat{z}_t}{T}\right)} \tag{9}$$

- Root mean square error

$$RMSE = \sqrt{\frac{\sum_{t=1}^{t=T} \left(z_t - \hat{z}_t\right)^2}{T}} \tag{10}$$

- Mean absolute error

$$MAE = \frac{1}{T} \sum_{t=1}^{t=T} |z_t - \hat{z}_t| \tag{11}$$

2.2. Genetic algorithm

GA is capable of searching for the global optimum in the infinite, complex, multimodal, and non-differentiable search space to determine the optimal DFNN architecture [39,40]. In the beginning, a population of chromosomes in the solution space is randomly initialized. The quality of each chromosome is assessed using the objective function; thus the probability of adopting each of them in the next generation can be

Table 1
Summary of architecture of FNN and DFNN in literature review.

References	Quantity of neurons in input layer	Quantity of neurons in hidden layers	Quantity of neurons in output layer	Training samples
[10]	9	60–80	1	8760
[11]	6	15 or 20	1	8760
[12]	5	20	1	26,280
[13]	17	10	1	700,000
	6	15		
[14]	2	4~9	1	722
[15]	5	20	1	250
[16]	5	10 or 15 or 20	1	8928
[17]	12	25	1	8760
[19]	8	36	1	742
[21]	8	10 or 30	1	7200
[22]	10	10	1	10,972
[23]	3	20	1	2472
[24]	4	1~8	1	20
[25]	3	11	1	2472
[29]	82	75–59, 71–87, 57–63	2	17,000
		87–99–94, 14–31–28 and 61–70–69	3	
[30]	56	56–15	2	17,520
	58	58–20		
[31]	168	10~100	1~5	497,832
[32]	13	10–5–2	3	217,440
[41]	4	30–30	2	2700
[42]	7	512	3	19,704
[43]	4	12–12	2	279
[44]	784	500 per layer	9	600,000

determined. Generally, the new chromosome is generated through the three genetic operators: selection, crossover and mutation. Based on the fitness values of the objective function, some chromosomes are selected to remain while others are chosen for crossover and mutation. Chromosomes which are selected for the crossover operation are named as the parents. Through crossover operation, features of parents are exchanged between each other to produce the offspring. Meanwhile, the mutation operator implements the process of altering the features within the chromosome. Through conducting the three genetic operations, the initial population is updated with the offspring generated by crossover and mutation. At the final stage, the offspring with the best fitness values is returned to represent the optimum solution.

2.3. GA-DFNN predictive model

Although there are three different empirical equations in determining the optimal quantity of neurons in the single hidden layer of FNN, there is no rule-of-thumb for choosing the quantity of hidden layers, quantity of neurons in the hidden layers, activation functions in each layer as well as learning process for determining weighting factors. However, these architectural attributes are responsible for the performance of DFNN. The network with a small size can cause under-fitting while a large network may lead to overfitting. The activation function determines the connecting performance among each neuron, while the learning process affects the convergence of DFNN. These imply that the architectural design of DFNN model is very crucial and can be defined as an optimization problem.

To improve the prediction accuracy of DFNN predictive model, GA is applied to determine the optimal architecture of the DFNN model. The decision variables of GA include the total quantity of hidden layers, the quantity of neurons in each hidden layer, the activation function as well as the learning process to compute the weighting factors of the DFNN in the training phase. The optimization objective function is the *MAPE* value of the DFNN model. In other words, each set of decision variables determines one type of architecture of DFNN model. Through training the DFNN models using historical datasets, the *MAPE* value of the DFNN model, namely, optimization objective function, can be obtained.

The decision variables are summarized in Table 2. To generate the pool of decision variables regarding quantity of hidden layers and quantity of neurons in each hidden layer, an in-depth literature review on FNN and DFNN prediction models was conducted, as summarized in Table 1. As will be discussed in Section 3, there are 8760 training samples and 9 input neurons for day-ahead hourly prediction; while 365 training samples and 10 input neurons for week-ahead daily energy consumption prediction. In view of the hidden layers and neurons adopted in the previous FNN and DFNN models, the quantity of hidden layers was tested in the range of 1–5 while the quantity of neurons in each layer was tested in the range of 5–10. The procedure of the proposed predictive model is shown in Fig. 2.

Table 2
Decision variables of GA for DFNN.

Quantity of neurons in each hidden layer	{5, 6, 7, 8, 9, 10}
Quantity of hidden layers	{2, 3, 4}
Activation functions	{sigmoid, tanh, relu, elu}
Optimization approaches	{SGD, ADAM, NADAM, ADAMAX}

In the beginning, a population of 20 DFNN models are generated while random decision variables are assigned to determine the architecture of each DFNN model. Each DFNN model is trained using the input and output datasets from the historical database. Based on the fitness value of the objective function (i.e. MAPE value of DFNN), selection, crossover and mutation operators of GA are conducted for the population of DFNN models, respectively. Such a procedure is repeated until the preset optimization criteria are met. Finally, optimal DFNN architecture can be determined.

2.4. Research methodology

The proposed GA-DFNN predictive model is implemented in Python with the Keras and TensorFlow libraries as the backend. Keras is a library of open sources of the neural network developed in Python, which is good at the fast computation of DFNN and which is focused on minimization, modularity, and scalability [44]. Meanwhile, TensorFlow is an open-source software library which provides an interface for expressing and executing various machine learning algorithms [45].

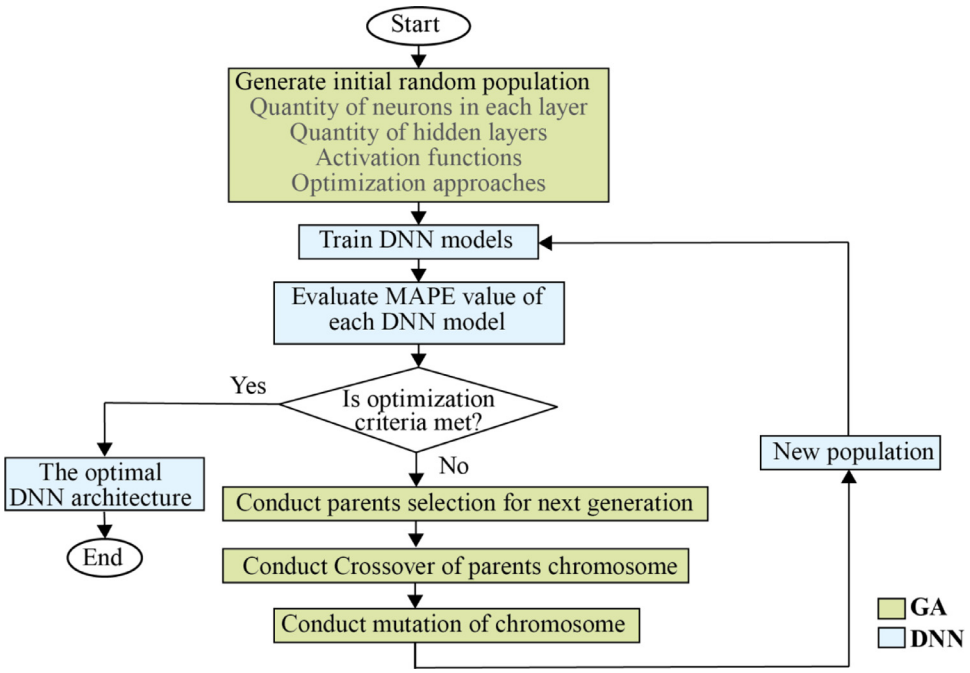


Fig. 2. Schematic flowchart of the proposed GA-DFNN predictive model.

Fig. 3. Distribution of training, validation and testing datasets.

3. Structure of historical database

To test the performance of the proposed GA-DFNN predictive model, it was implemented on a campus building in the United Kingdom. The energy data was collected from the Northavon House in University of the West of England, Bristol over the past 1 year and 6 months (1 July 2018 – 31 December 2019) at the time step of 1 h.

3.1. Composition of the historical database

Since the building envelop properties generally remain stable over long time, the affecting factors of building electricity consumption include weather conditions, occupancy and operating equipment in the building. Due to the difficulty in collecting information of occupancy and operating equipment, time indicators, such as hour of the day, day of the week, month of the year, are usually chosen as schedule-related inputs to represent occupancy and operating equipment scenarios. In addition, the historical electricity consumption during the previous day or week can indicate the trend of profile in a mathematical way. Consequently, three types of input variables, including weather conditions, time indicators and historical energy consumption, are selected for the proposed GA-DFNN predictive model. The weather condition can be well-described using outdoor air dry-bulb temperature, outdoor air dew-point temperature, wind speed and solar radiation. The historical outdoor weather data was collected through the weather station in Bristol [46]. Although solar radiation is accessible from the historical weather data, it is generally not available from weather forecasting websites [47–49]. Therefore, it is not considered in this study.

3.2. Data pre-processing

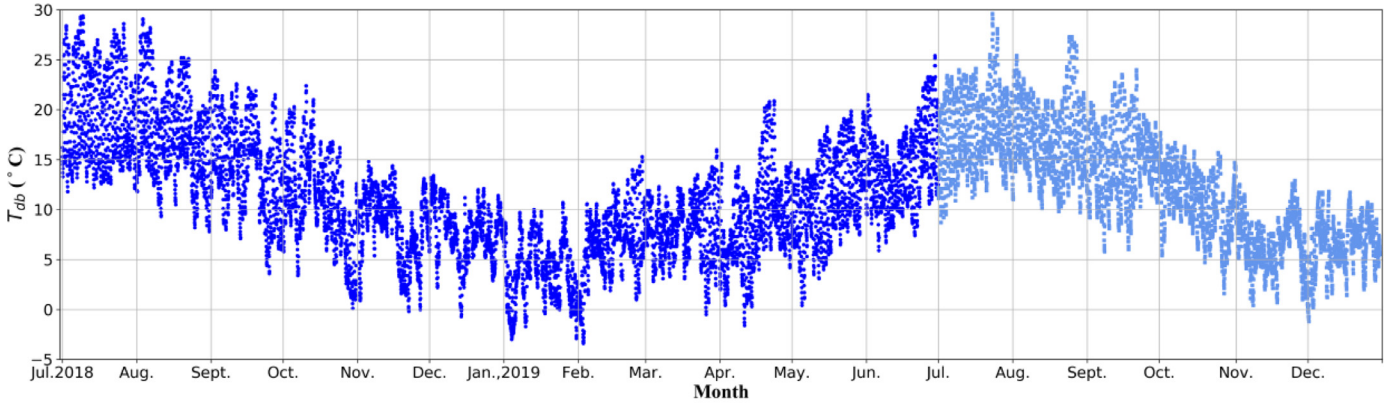
To facilitate the prediction of both day-ahead hourly and week-ahead daily electricity consumption, two sets of the historical database were formulated. As shown in Fig. 3, the random 80% of the datasets from the first one year, the remaining 20% of the datasets from the first one year, and datasets from the following 6 months were adopted for training, validation and testing purposes, respectively. To increase the prediction accuracy of DFNN while prevent it from overfitting, the average value of *MAPE* for training and testing cases was treated as the objective function in GA optimization.

3.2.1. Database for day-ahead hourly electricity consumption prediction

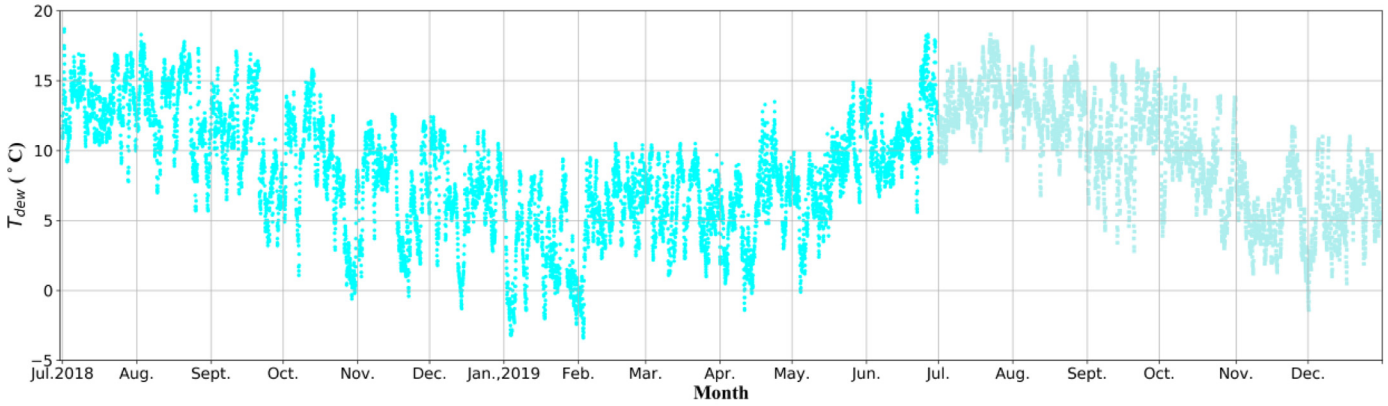
The training database for the day-ahead hourly electricity consumption prediction includes the hour of the day $h(t)$, the day of the week $d(t)$, the month of the year $m(t)$, hourly outdoor air dry-bulb temperature $T_{db}(t)$, hourly outdoor air dew-point temperature $T_{dew}(t)$, hourly wind speed $V_{wind}(t)$ as well as the hourly electricity consumption at the same hour of the previous day $E_h(t\text{-}24)$. To represent the cyclical nature of the time, $h(t)$ was represented by its sine and cosine value as [13,37]:

$$h_s(t) = \sin\frac{2\pi h(t)}{24} \; h_c(t) = \cos\frac{2\pi h(t)}{24} \tag{13}$$

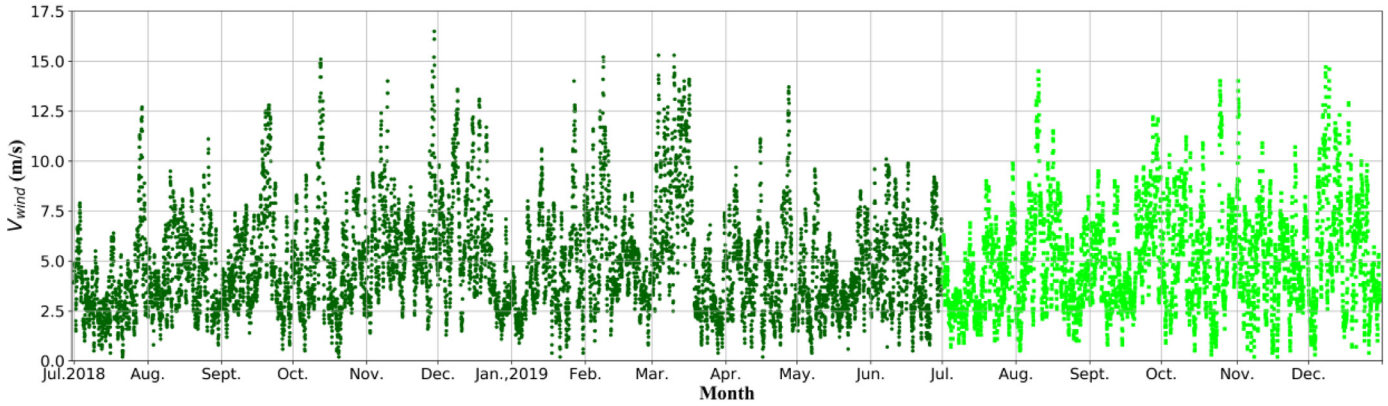
Owing to the different occupancy and operating equipment scenarios during the holidays, the holidays were marked as an additional day of the week. Meanwhile, the categorical data, including day of the week and month of the year were encoded using one-hot encoding approach. Day of the week was represented by a concatenation of 8 binary variables. For instance, bank holidays were expressed as [1 0 0 0 0 0 0



(a) Outdoor air dry-bulb temperature



(b) Outdoor air dew-point temperature



(c) Wind speed

Fig. 4. Hourly outdoor weather data.

0], Sunday was expressed as [0 1 0 0 0 0 0 0]. The first column corresponded to a flag whether or not the day was a bank holiday, while the second column referred to a flag whether or not the day was Sunday. Similarly, the remaining six columns corresponded to a flag for each of the six other days of the week [50]. Meanwhile, month of the year was referred by a concentration of 12 binary variables. For example, January was illustrated as [1 0 0 0 0 0 0 0 0 0 0 0], with the first column corresponded to a flag whether or not the month was January. In addition, the variables of outdoor weather data (i.e. $T_{db,\,h}$, $T_{dew,\,h}$ and $V_{wind,\,h}$) were normalized through the min-max scaling approach:

$$x' = \frac{x(t) - \min\limits_{1 \le t \le 365 \times 24} x(t)}{\max\limits_{1 \le t \le 365 \times 24} x(t) - \min\limits_{1 \le t \le 365 \times 24} x(t)} \tag{14}$$

The hourly outdoor air dry-bulb temperature, outdoor air dew-point temperature and wind speed are shown in Fig. 4. The peak dry-bulb

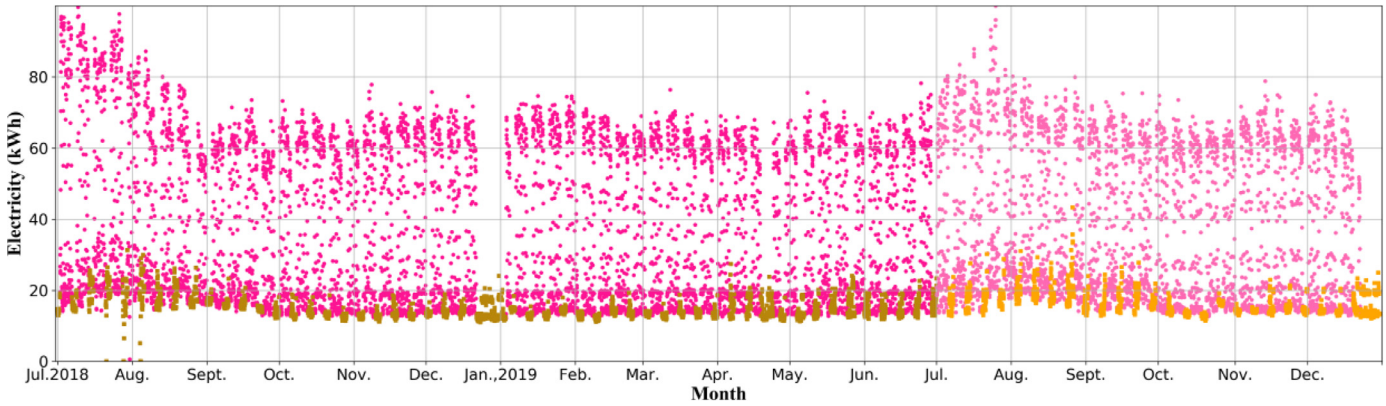


Fig. 5. Hourly electricity consumption.

and dew-point temperature happened in July while the valley occurred in January. The highest and lowest dry-bulb temperature is 30 °C and −3 °C, respectively. Meanwhile, the highest and lowest dry-bulb temperature is 29 °C and −3 °C, respectively. The overall trend of dry-bulb temperature and dew-point temperature is similar during the second-half year in 2018 and 2019. During most time of the year, wind speed is lower than 10 m/s, while it would reach 15 m/s on a few days.

The hourly electricity consumption is presented in Fig. 5. In Fig. 5, the pink dots stood for weekday electricity consumption while the orange dots represented the weekend and bank holidays. Owing to the normal operating schedule of the investigated office building, electricity consumption on weekdays was much higher than that during the weekends and bank holidays. Moreover, the peak electricity consumption was identified in July due to its high dry-bulb and dew-point temperature.

In conclusion, the input dataset at time step t is denoted by training sample $X_t = [x_{1,t}\ x_{2,t}\ x_{3,t}\ x_{4,t}\ x_{5,t}\ x_{6,t}\ x_{7,t}]$, where $x_{1,t} = h_s(t)$, $x_{2,t} = h_c(t)$, $x_{3,t} = d(t)$, $x_{4,t} = m(t)$, $x_{5,t} = T_{db,h}(t)$, $x_{6,t} = T_{dew,h}(t)$, $x_{7,t} = V_{wind,h}(t)$, while the output dataset is $Z_t = E_h(t)$.

3.2.2. Database for week-ahead daily prediction

The training database for the week-ahead daily electricity consumption prediction includes the day of the week $d(t)$, the month of the year $m(t)$, daily average outdoor air dry-bulb temperature $T_{db,ave,d}(t)$, daily maximum outdoor air dry-bulb temperature $T_{db,\max,d}(t)$, daily minimum outdoor air dry-bulb temperature $T_{db,\min,d}(t)$, daily average outdoor air dew-point temperature $T_{dew,ave,d}(t)$, daily maximum outdoor air dew-point temperature $T_{dew,\max,d}(t)$, daily minimum outdoor air dew-point temperature $T_{dew,\min,d}(t)$, daily average wind speed $V_{wind,d}(t)$ as well as the total daily electricity consumption $E_d(t)$. The variables of the daily outdoor weather data (i.e. $T_{db,\ ave,\ d}$, $T_{db,\ min,\ d}$, $T_{db,\ max,\ d}$, $T_{dew,\ ave,\ d}$, $T_{dew,\ min,\ d}$, $T_{dew,\ max,\ d}$ and $V_{wind,\ d}$) were also normalized:

$$x' = \frac{x(t) - \min\limits_{1\le t\le 365} x(t)}{\max\limits_{1\le t\le 365} x(t) - \min\limits_{1\le t\le 365} x(t)} \tag{15}$$

The daily average outdoor air dry-bulb temperature, outdoor air dew-point temperature and wind speed are shown in Fig. 6. The overall trends of daily-average dry-bulb and dew-point temperature is similar to those of hourly dry-bulb and dew-point, respectively. The peak value of daily-average dry-bulb temperature was in July at 23 °C, while the lowest value was in early January at −1 °C. Moreover, the peak value of daily-average dew-point temperature was in August at 17 °C, while the lowest value was in early January at −2 °C. The highest wind speed was found in October at 12.5 m/s.

The daily total electricity consumption is presented in Fig. 7. The pink dots stood for weekday electricity consumption while the orange dots represented the weekend and bank holidays. Owing to the normal operating schedule of the investigated office building, electricity consumption on weekdays was much higher than that during the weekends and bank holidays. The daily electricity consumption was between 700=1300 kWh on weekdays while 250–520 kWh on weekends and bank holidays. The relatively higher electricity consumption was occurred in January due to relatively higher dry-bulb and dew-point temperature. It is also noticed that the daily electricity consumption was 500 kWh on the 1st February, which was much lower than other weekdays. It is because that the campus was closed due to slow on that day. Therefore, this day was regarded as a holiday to increase the prediction accuracy of the proposed GA-DNN predictive model.

As a whole, the input dataset at time step t is denoted by $X_T = [x_{1,t}\ x_{2,t}\ x_{3,t}\ x_{4,t}\ x_{5,t}\ x_{6,t}\ x_{7,t} x_{8,t}\ x_{9,t}]$, where $x_{1,t} = d'(t)$, $x_{2,t} = m'(t)$, $x_{3,t} = T'_{db,ave,d}(t)$, $x_{4,t} = T'_{db,max,d}(t)$, $x_{5,t} = T'_{db,min,d}(t)$, $x_{6,t} = T'_{dew,ave,d}(t)$, $x_{7,t} = T'_{dew,max,d}(t)$, $x_{8,t} = T'_{dew,min,d}(t)$, $x_{9,t} = V'_{wind,d}(t)$, while the output dataset is $z_t = E_d(t)$.

4. Performance evaluation of the proposed GA-DFNN predictive model

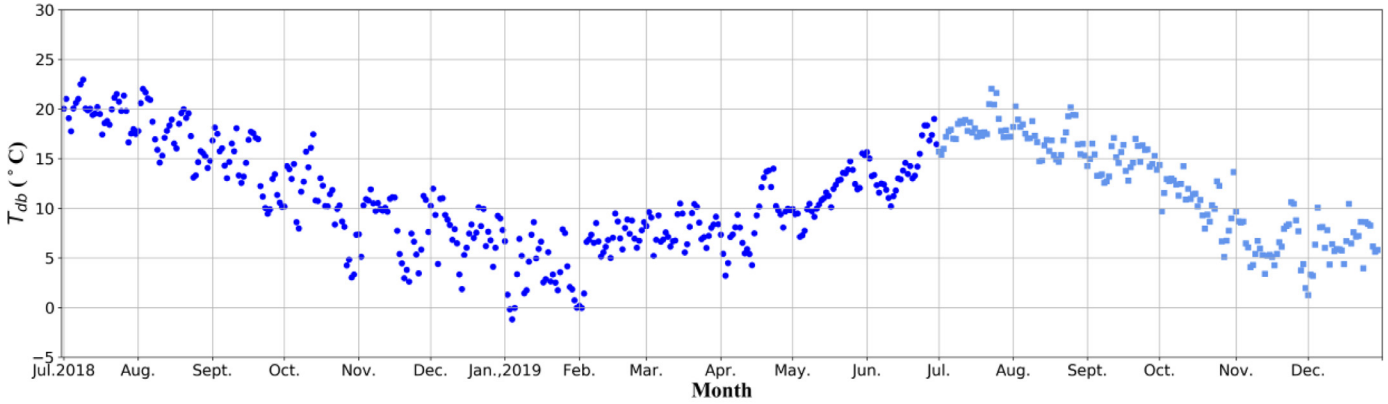
To prevent the optimization from being trapped in local optimum, the GA parameters were chosen using sensitivity analysis. The convergence performance of the GA optimization was assessed. Therefore, the optimal architecture of the DFNN prediction model can be determined.

4.1. Determination of GA parameters

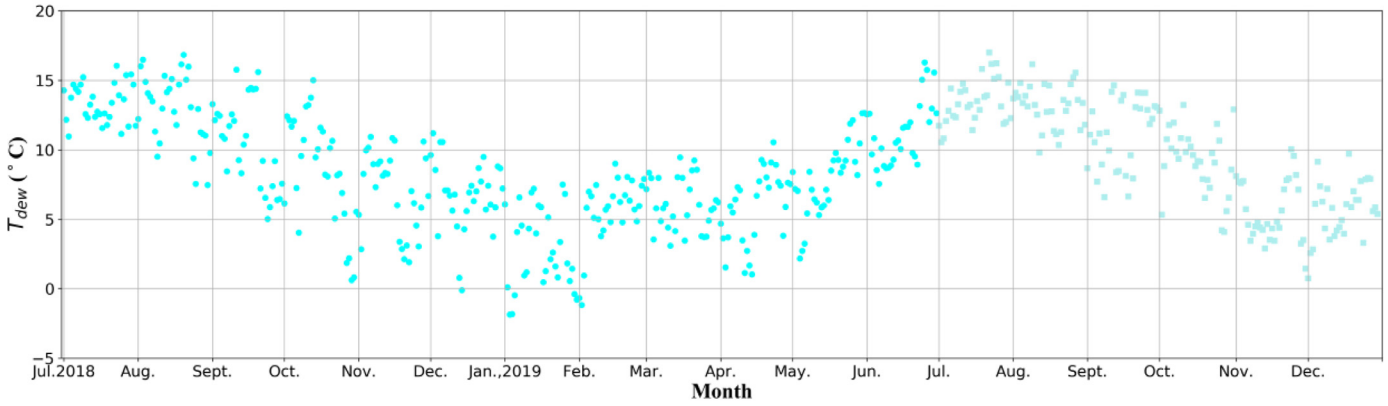
To prevent the optimization from being converged to local optimum, different GA parameters, including five *retaining probabilities* (i.e. 0.5, 0.6, 0.7, 0.8 and 0.9), four *selection probabilities* (i.e. 0.1, 0.2, 0.3 and 0.4) and four *mutation probabilities* (i.e. 0.1, 0.2, 0.3 and 0.4) were adopted in the optimization of DFNN architecture for week-ahead daily electricity consumption. The optimization results are shown in Fig. 8. For each variable bar, the average value of other variables is obtained. For example, when calculating the bar value of *retaining probability* = 0.5, the average value of 4 × 4 optimization results was adopted. According to the resulted *MAPE* value, the *retaining probability, selection probability* and *mutation probability* were chosen as 80%, 20%, and 20%, respectively (Fig. 9.).

4.2. Convergence performance of GA optimization

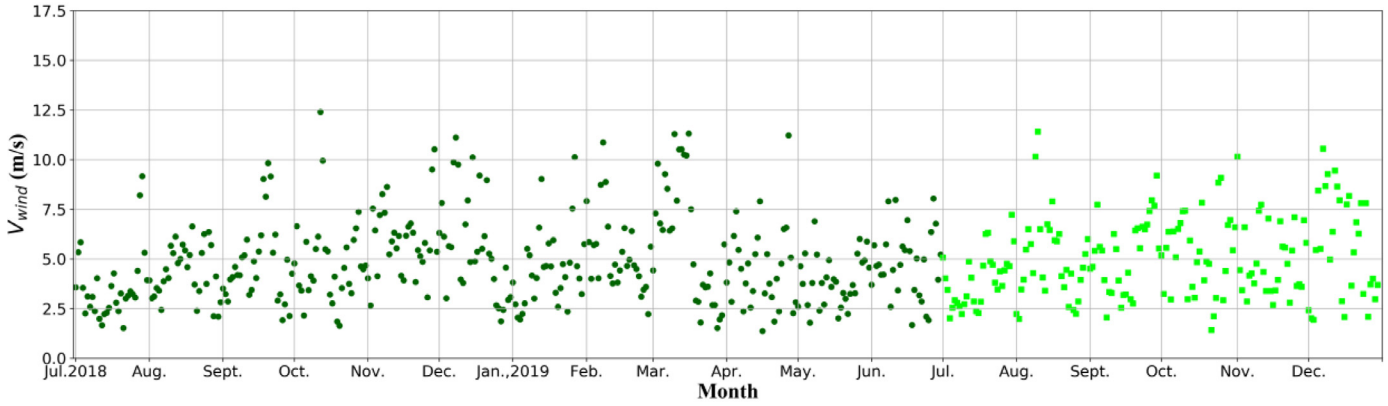
The convergence performance of the day-ahead hourly and the week-ahead daily electricity consumption prediction is shown in Fig. 10. The minimum value of day-ahead hourly and week-ahead daily electricity consumption reached convergent after 31 and 22 iterations, respectively. To ensure the convergence would be steady thereafter, the maximum iterations were chosen as 40 and 30 for day-ahead hourly and week-ahead daily electricity consumption, respectively.



(a) Outdoor air dry-bulb temperature



(b) Outdoor air dew-point temperature



(c) Wind speed

Fig. 6. Daily-average value of weather data.

4.3. Optimal architecture of the proposed GA-DFNN model

After the GA optimization, the optimal architecture of the DFNN model could be determined, as shown in Fig. 10. There are 4 and 2 hidden layers for day-ahead hourly and week-ahead daily electricity consumption, respectively. There were more hidden layers for day-ahead hourly electricity consumption prediction owing to the reason that there exists a higher number of training samples (i.e. 8736). Rectified linear unit and exponential linear activation functions were mainly adopted in hidden layers for day-ahead hourly prediction while exponential linear and hyperbolic tangent activation functions were used in week-ahead daily electricity consumption prediction. Exponential linear activation function was adopted in the output year for both prediction models. NADAM and ADAMAX were adopted as training process for day-ahead

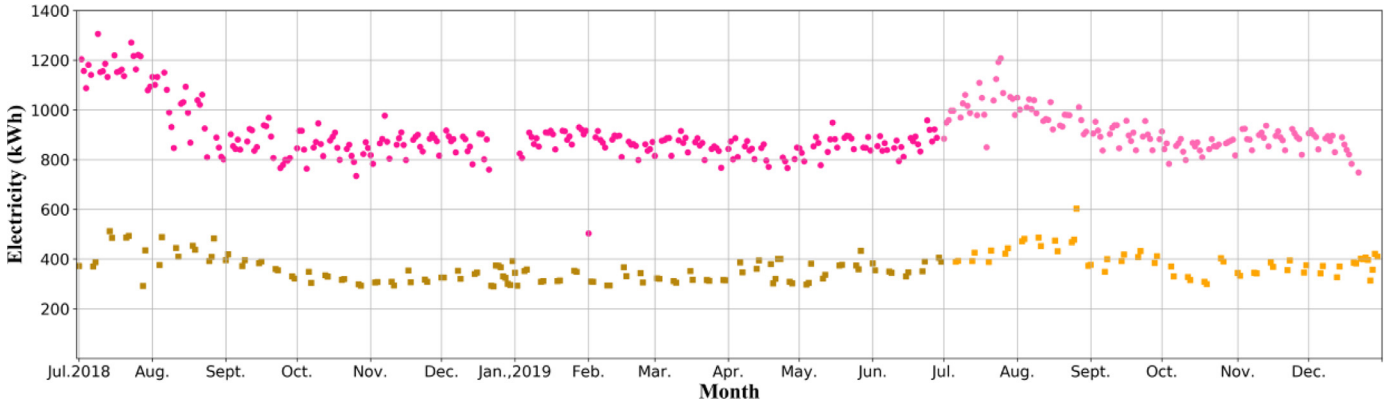


Fig. 7. Daily total electricity consumption.

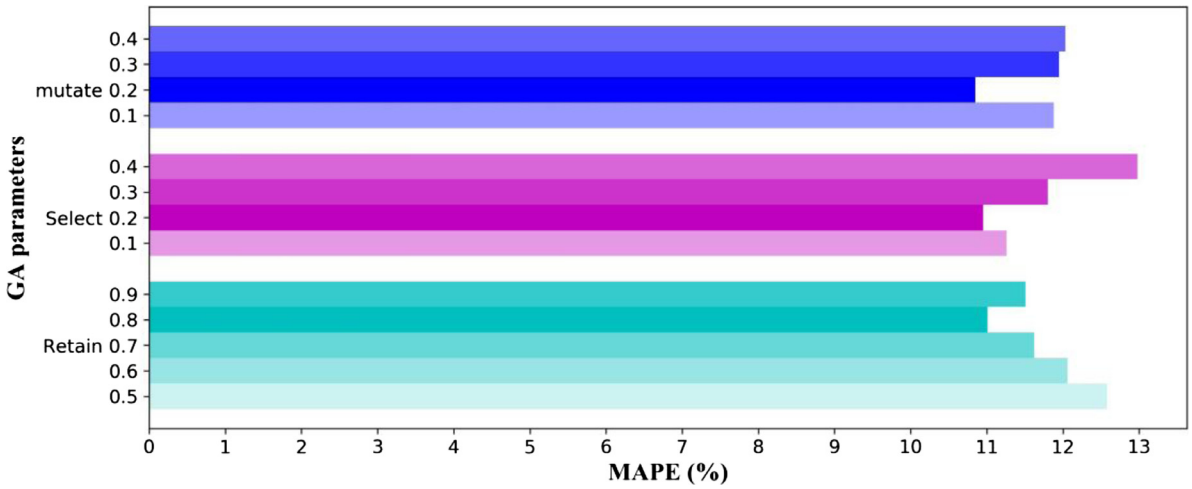
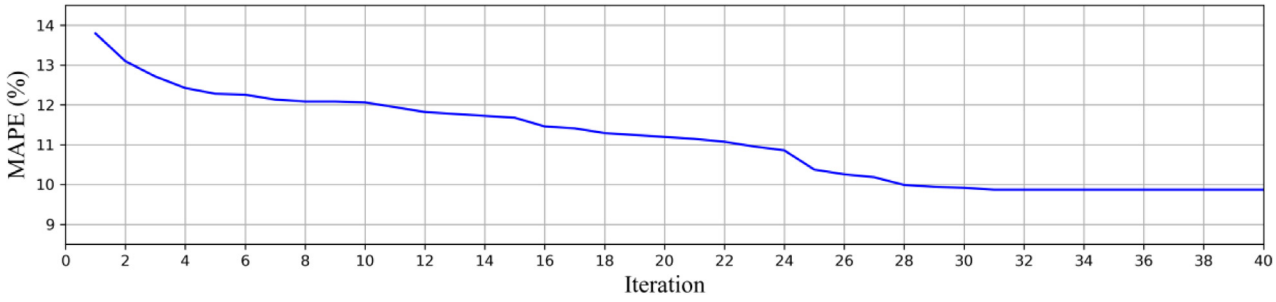
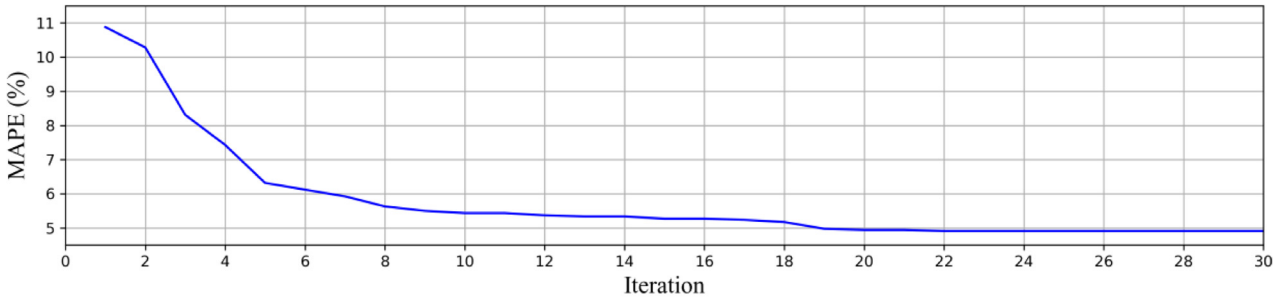


Fig. 8. Optimization result at different GA parameters.



(a) Day-ahead hourly prediction.



(b) Week-ahead daily prediction.

Fig. 9. Convergence of *MAPE* in GA optimization.

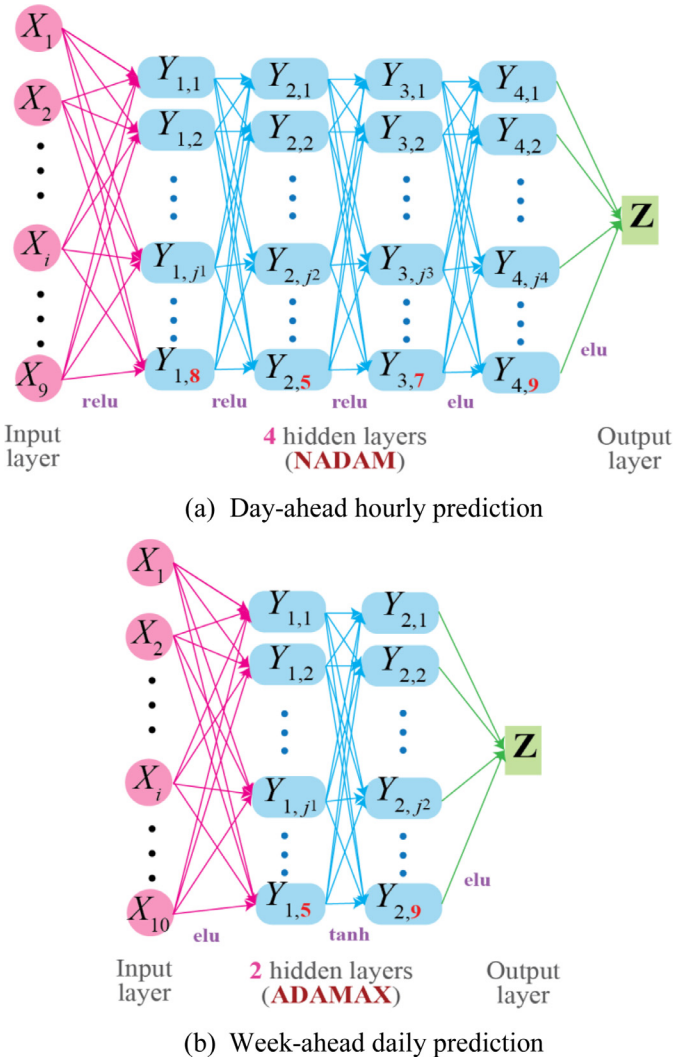


Fig. 10. Optimal architecture of the proposed GA-DFNN prediction model.

hourly and week-ahead daily electricity consumption prediction, respectively.

To reduce the computation time while prevent the proposed GA-DFNN predictive model from overfitting, early-stopping is adopted [51,52]. During the training phase, the *MAPE* value of the training and validation datasets are shown in Fig. 11. After 400 and 12,000 epochs, the *MAPE* value of daily and weekly prediction reached relatively constant. Therefore, the *MAPE* value for daily and weekly prediction was 9.8% and 4.9%, respectively.

5. Performance comparison with reference models

To assess the effectiveness of the proposed GA-DFNN predictive model, four groups of references cases were introduced. To investigate the prediction performance of the DFNN model with multiple hidden layers, the feedforward neural network model with single layer was adopted as reference. To assess the prediction performance improvement through DFNN model with optimal architecture, the DFNN models with different architecture are adopted as reference cases. To demonstrate the comprehensive relationship between affecting factors and electricity consumption, the time-series prediction approach including long-term-short-memory (LSTM) neural network model and temporal convolutional networks (TCN) model is adopted as reference. To exhibit the optimization performance of GA, the random search-based DFNN model is used as a reference case.

5.1. Performance comparison against FNN models with single hidden layer

Firstly, to investigate the prediction performance of the DFNN model with multiple hidden layers, the ANN model with single layer is adopted as reference. The quantity of neurons was selected according to the empirical equations in [18] and [20], respectively. Sigmoid activation function was adopted in the hidden layer while exponential linear activation function was used in the output layer. ADAM was adopted as the training process, as summarized in Table 3.

The performance of the proposed GA-DFNN model and two reference FNN models are summarized in Table 4.

- For day-ahead hourly prediction, although the prediction performance of GA-DFNN during training phase was a little worse than the reference FNN models with single layer, it showed better performance at testing phase. Compared to FNN1 and FNN2 models, there was 3.4% and 1.8% reduction in MAPE, 1.3% and 6.1% reduction in RMSE, 4.8 and 3.8% reduction in MAE, as well as 0.10% and 0.52% increase in R^2, respectively.
- For week-ahead daily prediction, the proposed GA-DFNN showed better performance in both training and testing phases. For testing cases of the proposed GA-DFNN model, the reduction in MAPE, RMSE and MAE was 34.5%, 38.9% and 29.3% while the increase in R^2 was 6.3% compared to FNN1 model; the reduction in MAPE, RMSE and MAE was 34.3%, 38.9% and 29.2% while the increase in R2 was 6.3% compared to FNN2 model.

The prediction results from two days in training, two days in testing, two weeks in training and two weeks in testing are shown in Figs. 12 and 13, respectively. The proposed GA-DFNN model showed better ability at tracking the variation of electricity consumption, such as at the 4th, 13th, 17th, 18th, 20th, 21st and 22nd h in Fig. 13(c); at the 1st, 2nd, 4th, 6th, 12th-14th, 17th, 21st h in Fig. 13(d); during Tuesday, Wednesday, Thursday and Friday in Fig. 14(c); as well as during Tuesday-Saturday in Fig. 14(d).

5.2. Performance comparison against DFNN models with other architecture

To verify the optimal architecture determined by GA optimization, the DFNN models with different neurons in each hidden layer, activation function and learning approach were adopted as reference. The selection of the reference architecture is based on the criteria that only one affecting factor was changed while other factors were kept the same as the GA-determined architecture. The architecture of the four reference DFNN models, along with the GA-determined DFNN, are summarized in Table 5.

The performance of the proposed GA-DFNN model and the four reference DFNN models are summarized in Table 6.

- For day-ahead hourly energy consumption prediction, although the prediction performance of GA-DFNN during training phase is a little worse than the reference FNN models with single layer, it showed better performance at testing phase. Compared to the reference DFNN1, DFNN2, DFNN3 and DFNN4 models, the proposed GA-DFNN model had 3.4%, 0.59%, 5.0% and 0.97% reduction in *MAPE* value; 0.21%, 0.31%, 0.52% and 0.31% increase in R^2; 1.4%, 0.25%, 0.34% and 0.27% reduction in RMSE; as well as 1.6%, 3.0%, 3.3% and 0.34% reduction in MAE.
- For week-ahead daily energy consumption prediction, the proposed GA-DFNN showed better performance in both training and testing phases. For the proposed GA-DFNN model, the reduction in MAPE value was 27.4%, 20.7%, 33.3% and 16.7%; the increase in R^2 was 2.9%, 4.2%, 5.8% and 1.9%; the reduction in RMSE was 25.3%, 32.0%, 37.4% and 18.9%; the reduction in MAE was 23.0%, 24.2%,

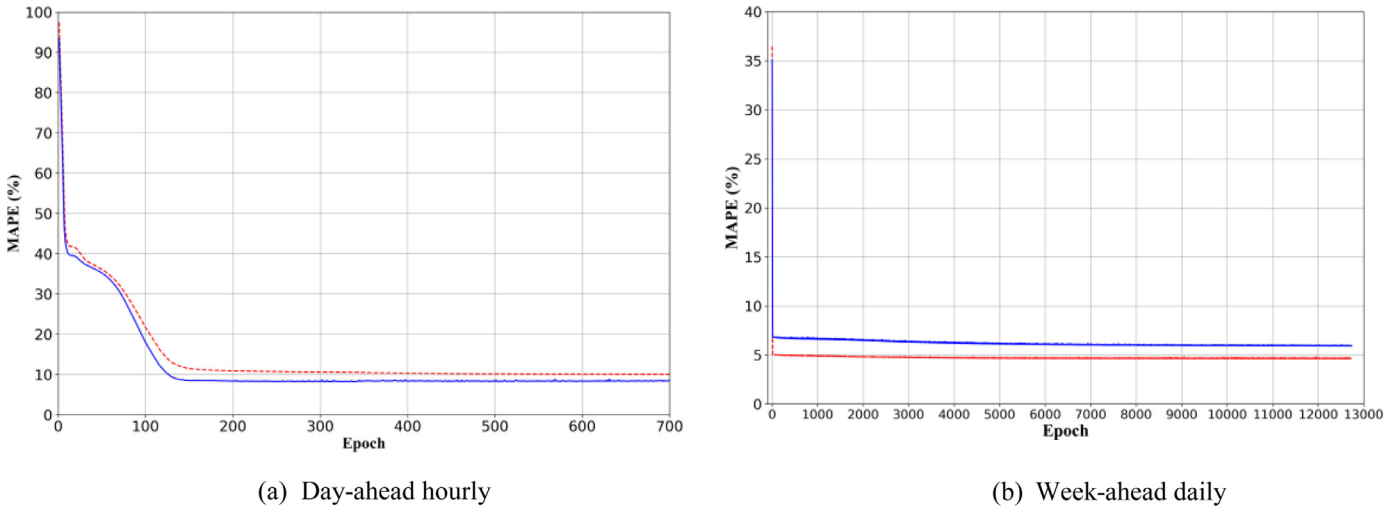


Fig. 11. *MAPE* convergence of the proposed GA-DFNN model.

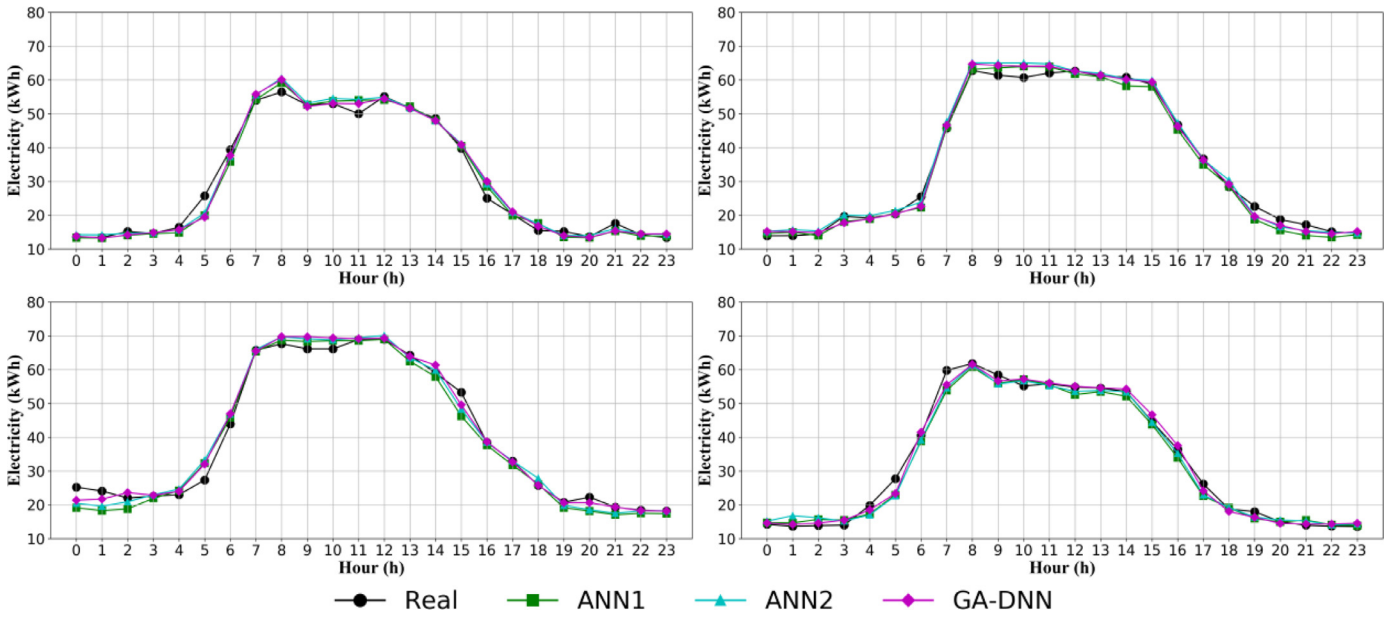


Fig. 12. Day-ahead hourly prediction of GA-DFNN and FNN models.

30.5% and 16.7%, compared to DFNN1, DFNN2, DFNN3 and DFNN4 models, respectively.

Since the architecture of reference models were chosen based on the GA-determined architecture by changing affecting factors one-by-one, the reduction in *MAPE*, RMSE and MAE as well as the increase in R^2 value verifies that the proposed GA-DFNN model was effective in finding its optimal architecture.

The prediction results from two days in training, two days in testing, two weeks in training and two weeks in testing are shown in Figs. 14 and 15, respectively. The proposed GA-DFNN model showed better ability at tracking the variation of electricity consumption, such as at 2nd, 11th, 12th, 13th and 16th −20th h in Fig. 15(c); at 0th-3rd, 8th, 11th-13th, 2nd, 4th, 6th, 12th-14th, 20th-22nd h in Fig. 15(d); on Tuesday, Wednesday and Friday in Fig. 16(c); and on Monday, Tuesday, Wednesday and Friday in Fig. 16(d).

5.3. Performance comparison against random search-based DFNN model

To assess the performance of GA in finding the optimal architecture, the DFNN model optimized by random search (RS-DFNN) was adopted as reference. The decision variables of random search also included

Table 3
Design parameters of reference DFNN models.

	Day-ahead hourly prediction		Week-ahead daily prediction	
Reference models	FNN1	FNN2	FNN1	FNN2
Neurons in hidden layer	17	98	21	25
Activation function	Sigmoid in hidden layer, exponential linear in output layer			
Optimization approach	ADAM			

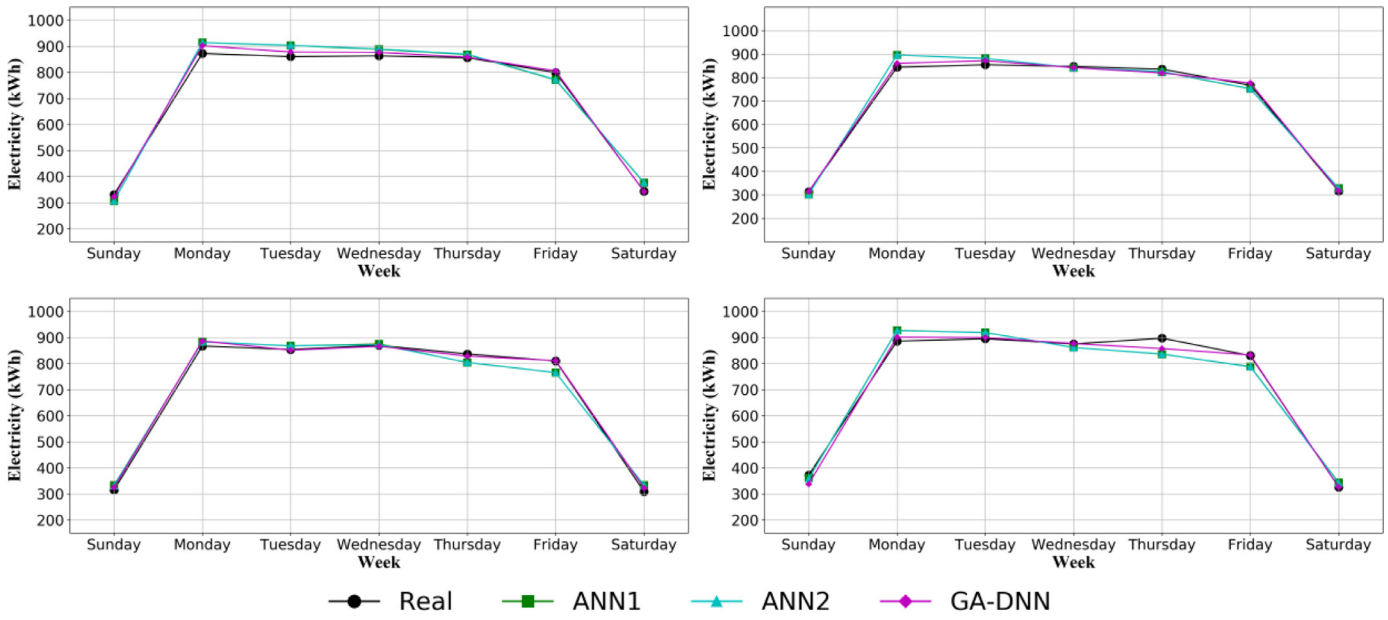


Fig. 13. Week-ahead daily prediction of GA-DFNN and FNN models.

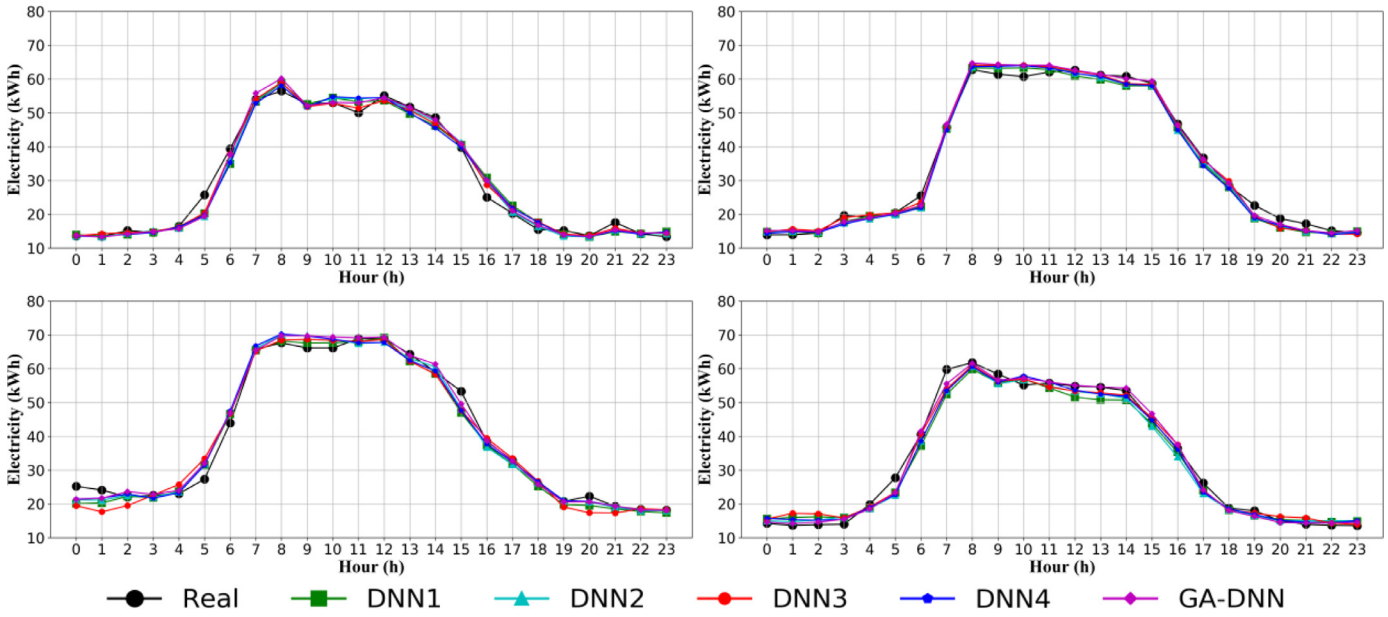


Fig. 14. Day-ahead hourly prediction of GA-DFNN and FNN models.

quantity of hidden layers, quantity of neurons in each layer, activation function and learning approach. The optimization objective was still the *MAPE* value between the predicted and actual measurement. The optimal architecture determined by random search is summarized in Table 7.

The performance of the proposed GA-DFNN model and the four reference DFNN models are summarized in Table 8.

- For day-ahead hourly prediction, although the prediction performance of GA-DFNN during training phase is a little worse than the reference FNN models with single layer, it showed better performance at testing phase. Compared to RS-DFNN model, the proposed GA-DFNN model has 9.7%, 6.8% and 8.2% reduction in *MAPE, RMSE* and *MAE*, while 0.52% increase in R^2, respectively.
- For week-ahead daily prediction, the proposed GA-DFNN showed better performance in both training and testing phases. Compared to RS-DFNN model, the reduction in *MAPE, RMSE* and *MAE* was 33.8%, 36.2% and 31.8% and 16.7%, respectively, while the increase in R^2 was 5.3%.

The worse performance of the RS-DFNN model may be owing to the fact that random search was trapped to local optimal rather than global optimal.

The prediction results from two days in training, two days in testing, two weeks in training and two weeks in testing are shown in Figs. 16 and 17, respectively. The proposed GA-DFNN model showed better ability at tracking the variation of electricity consumption in both training and testing cases during most of the time.

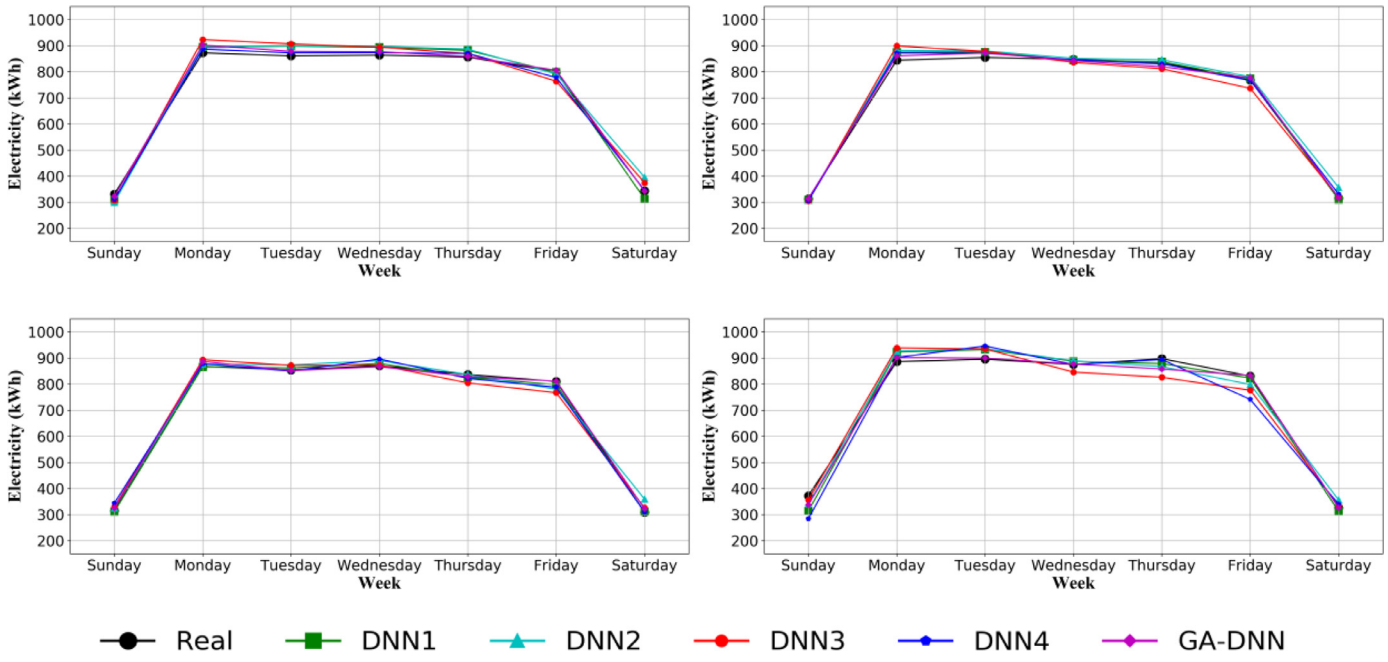


Fig. 15. Week-ahead daily prediction of GA-DFNN and DFNN models.

Table 4
Performance indicators of GA-DFNN and reference FNN models with single hidden layer.

		GA-DFNN		FNN1		FNN2	
		Train	Test	Train	Test	Train	Test
Daily	MAPE (%)	9.870	7.885	9.574	8.163	9.669	8.032
	R^2	0.958	0.970	0.961	0.969	0.963	0.965
	RMSE (kWh)	4.440	3.652	4.281	3.702	4.143	3.891
	MAE (kWh)	2.958	2.675	2.827	2.811	2.722	2.781
Weekly	MAPE (%)	4.908	5.326	10.04	8.13	10.04	8.11
	R^2	0.958	0.966	0.837	0.909	0.837	0.909
	RMSE (kWh)	55.25	47.24	108.6	77.29	108.7	77.27
	MAE (kWh)	32.49	35.52	60.26	50.30	60.30	50.17

5.4. Performance comparison against LSTM and TCN time-series prediction model

To demonstrate the comprehensive relationship between affecting factors and electricity consumption, the time-series prediction approach including LSTM [53] and TCN model is adopted as reference. The performance of the proposed GA-DFNN model and the four reference DFNN models are summarized in Table 9.

- For day-ahead hourly prediction, although the prediction performance of GA-DFNN during training phase was a little worse than the reference LSTM and TCN models, it showed better performance at testing phase. Compared to the reference LSTM model, the proposed GA-DFNN model has 12.2%, 40.1% and 32.6% reduction in *MAPE, RMSE* and *MAE*, while 5.9% increase in R^2, respectively. Compared to the reference TCN model, the proposed GA-DFNN model has 23.8%, 9.8% and 17.7% reduction in *MAPE, RMSE* and *MAE*, while 0.73% increase in R^2, respectively.
- For week-ahead daily prediction, the proposed GA-DFNN showed better performance in both training and testing phases. Compared to LSTM prediction model, the reduction in *MAPE, RMSE* and *MAE* was 32.7%, 46.4% and 26.4%, respectively, while the increase in R^2 is 9.6%. Compared to TCN prediction model, the reduction in *MAPE, RMSE* and *MAE* was 65.0%, 60.6% and 64.5%, respectively, while the increase in R^2 is 24.0%.

It demonstrated that the proposed GA-DFNN model has better accuracy than LSTM and TCN models for both day-ahead hourly prediction and week-ahead daily prediction.

The prediction results from two days in training, two days in testing, two weeks in training and two weeks in testing are shown in Figs. 18 and 19, respectively. The proposed GA-DFNN model showed better ability at tracking the variation of electricity consumption in both training and testing cases during most of the time than that of reference LSTM and TCN models.

Table 5
Architecture of the GA-DFNN and four reference DFNN models.

Models	GA-DFNN	DFNN1	DFNN2	DFNN3	DFNN4
Neurons in hidden layers	{8, 5, 7, 9}	{10, 10, 10, 10}	{5, 5, 5, 5}	{8, 5, 7, 9}	{8, 5, 7, 9}
Quantity of hidden layers	4				
Activation functions in hidden layers	relu, relu, relu, elu	relu, relu, relu, elu	relu, relu, relu, elu	sigmoid, sigmoid, sigmoid, sigmoid	relu, relu, relu, elu
Activation function in output layer	elu				
Learning approach	NADAM	NADAM	NADAM	NADAM	ADAM
Neurons in hidden layers	{5, 9}	{10, 10}	{5, 5}	{5, 9}	{5, 9}
Quantity of hidden layers	2				
Activation function in hidden layer	elu, tanh	elu, tanh	elu, tanh	Sigmoid, sigmoid	elu, tanh
Activation in output layer	elu				
Learning approach	ADAMAX	ADAMAX	ADAMAX	ADAMAX	ADAM

Table 6
Prediction performance of the GA-DFNN and four reference DFNN models.

		GA-DFNN		DFNN1		DFNN2		DFNN3		DFNN4	
		Train	Test	Train	Test	Train	Test	Train	Test		
Daily	MAPE (%)	9.870	7.885	9.574	8.163	9.669	7.932	9.521	8.303	9.751	7.962
	R^2	0.958	0.973	0.961	0.971	0.959	0.970	0.956	0.968	0.957	0.970
	RMSE (kWh)	4.440	3.652	4.277	3.602	4.379	3.643	4.520	3.782	4.483	3.662
	MAE (kWh)	2.958	2.675	2.843	2.718	2.897	2.683	2.926	2.765	2.955	2.684
Weekly	MAPE (%)	4.908	5.326	7.085	7.337	6.693	6.713	9.662	7.989	5.812	6.388
	R^2	0.958	0.966	0.890	0.939	0.860	0.927	0.851	0.913	0.942	0.948
	RMSE (kWh)	55.25	47.24	89.35	63.21	100.8	69.42	103.9	75.49	65.04	58.24
	MAE (kWh)	32.49	35.52	48.22	46.11	56.53	46.89	59.45	51.13	39.76	42.65

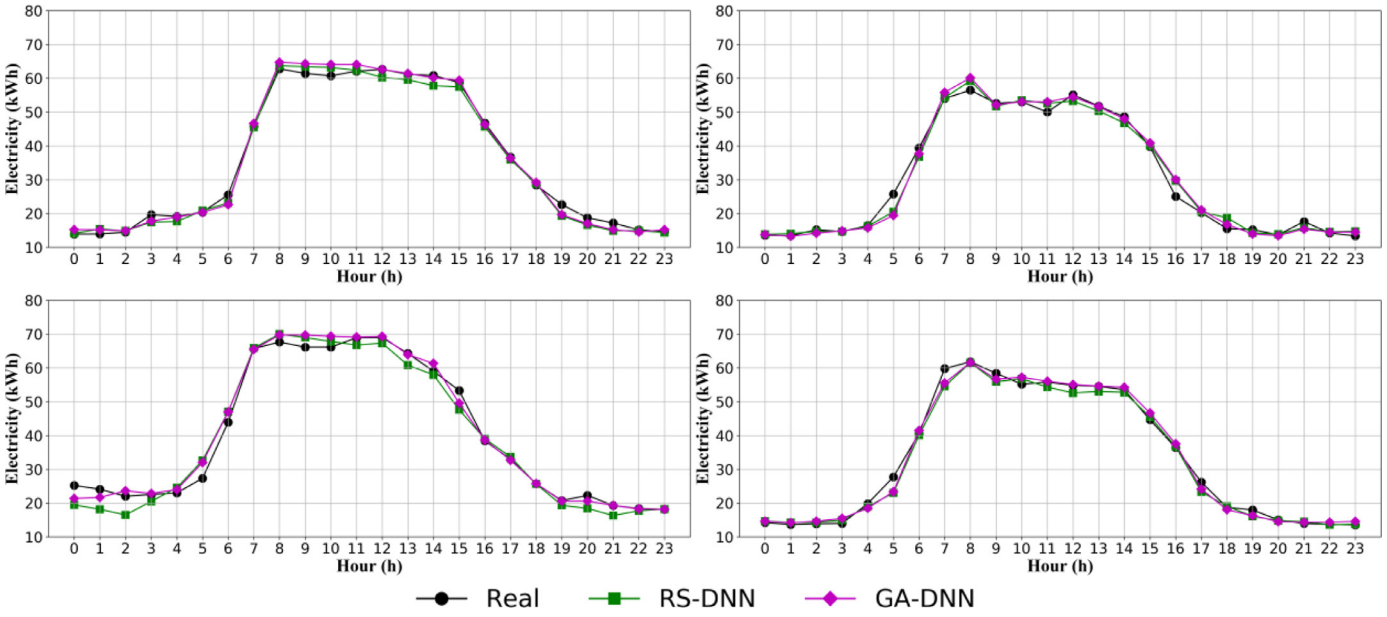


Fig. 16. Day-ahead hourly prediction of GA-DFNN and RS-DFNN models.

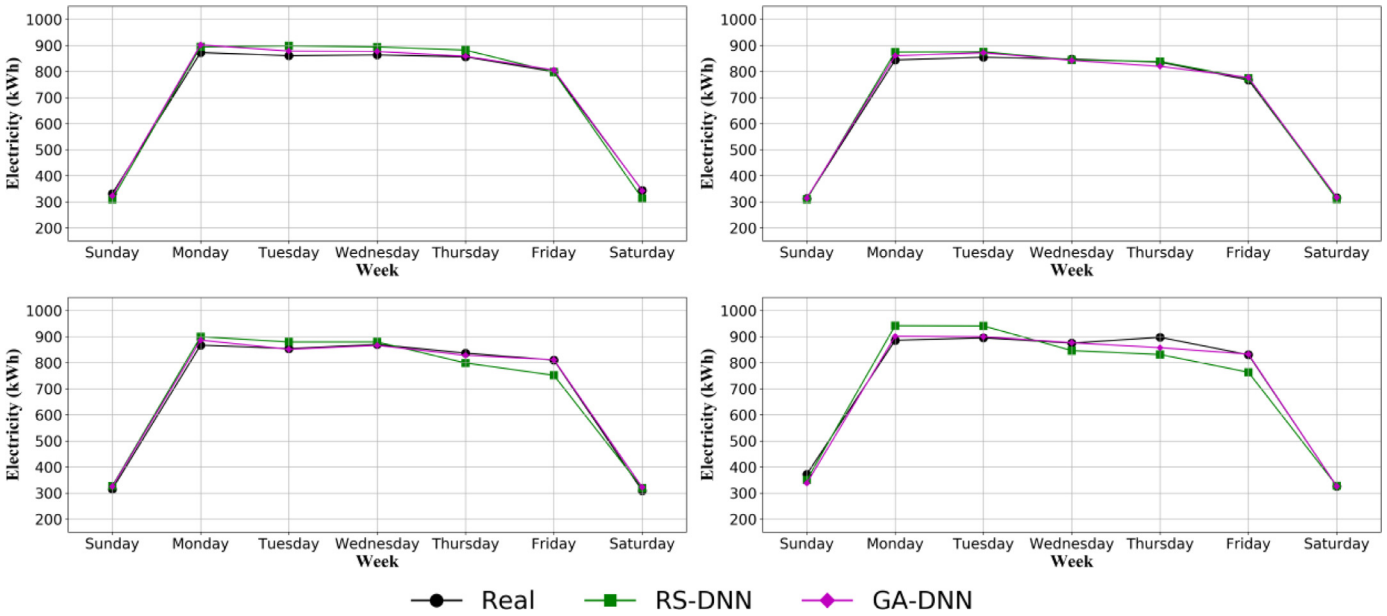


Fig. 17. Week-ahead daily prediction of GA-DFNN and RS-DFNN models.

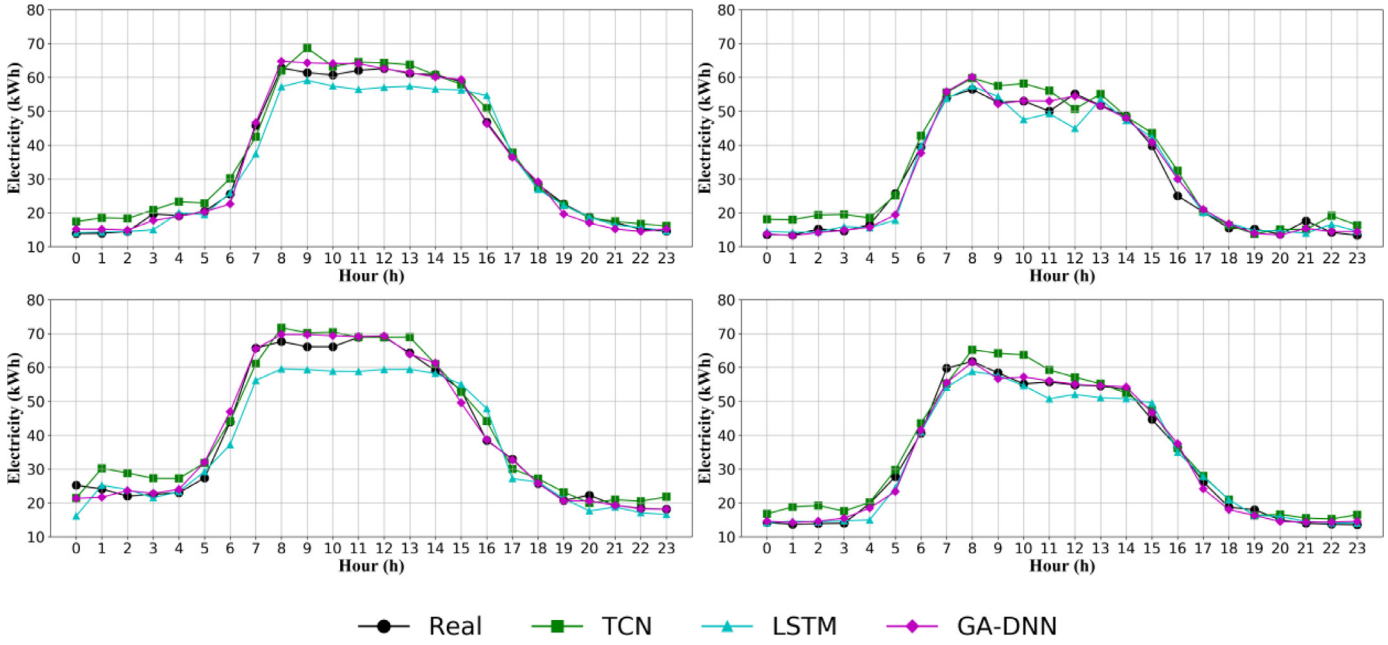


Fig. 18. Day-ahead hourly prediction of GA-DFNN, LSTM and TCN models.

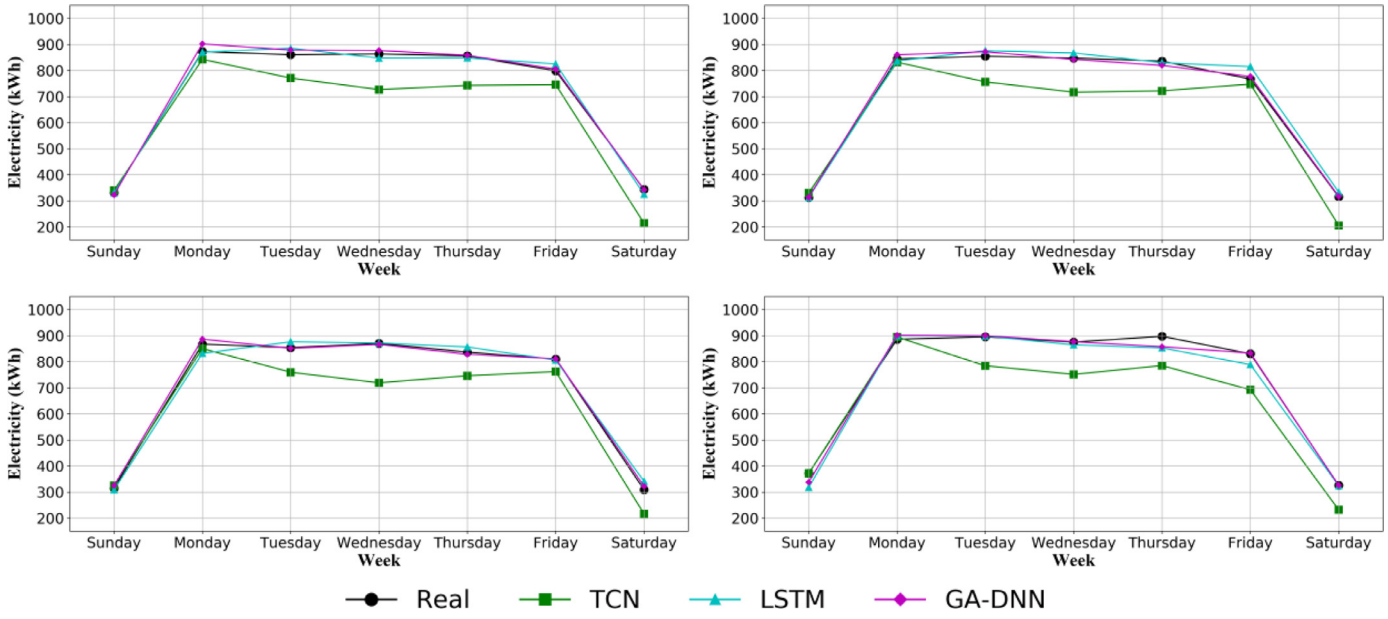


Fig. 19. Week-ahead daily prediction of GA-DFNN, LSTM and TCN models.

6. Practical implication and future application

To generate the database for training and testing the proposed GA-DFNN predictive model, more than one and a half year's historical outdoor weather data should be collected from the local weather station while the electricity consumption data ought to be obtained from the building energy system. Upon training and testing the propose GA-DFNN predictive model, it can be adopted in the building management system to predict the day-ahead hourly and week-ahead daily electricity consumption using the latest forecast of weather profile acquired from the weather reporting website [45–47]. It can also be further refined to tackle the problem caused by the probable faulty measurement data owing to sensor and equipment faults.

However, due to the fact that the campus building has relatively stable building operating schedule, the effectiveness of the proposed GA-DFNN predictive model should be further evaluated when adopted in residential, hospital, hotel or other types of buildings. 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 optimization performance than the GA algorithm in determining the optimal DFNN architecture.

Table 7
Optimal architecture of DFNN model determined by random search.

Models	Day-ahead hourly prediction	Week-ahead daily prediction
Neurons in hidden layers	{6, 8, 7, 9}	{7, 8}
Quantity of hidden layers	4	2
Activation functions in hidden layers	tanh, tanh, relu	elu
Activation function in output layer	elu	elu
Learning approach	ADAMAX	ADAM

Table 8
Prediction performance of the GA-DFNN and the reference RS-DFNN model.

		GA-DFNN		RS-DFNN	
		Train	Test	Train	Test
Daily	MAPE (%)	9.870	7.885	9.124	8.734
	R^2	0.958	0.970	0.960	0.965
	RMSE (kWh)	4.440	3.652	4.305	3.920
	MAE (kWh)	2.958	2.675	2.808	2.917
Weekly	MAPE (%)	4.908	5.326	9.592	8.049
	R^2	0.958	0.966	0.851	0.917
	MAE (kWh)	55.25	47.24	103.9	73.99
	RMSE (kWh)	32.49	35.52	60.34	52.10

Table 9
Prediction performance of the GA-DFNN and the reference LSTM and TCN model.

		GA-DFNN		LSTM		TCN	
		Train	Test	Train	Test	Train	Test
Daily	MAPE (%)	9.870	7.885	9.591	8.983	10.70	10.36
	R^2	0.958	0.970	0.910	0.916	0.970	0.963
	RMSE (kWh)	4.440	3.652	6.499	6.104	3.741	4.050
	MAE (kWh)	2.958	2.675	3.966	3.969	3.049	3.251
Weekly	MAPE (%)	4.908	5.326	7.013	7.909	16.12	15.21
	R^2	0.958	0.966	0.885	0.881	0.756	0.779
	RMSE (kWh)	55.25	47.24	88.79	88.11	130.8	119.9
	MAE (kWh)	32.49	35.52	44.13	48.26	101.7	100.1

6. Conclusion

To enhance the prediction accuracy of building electrical energy consumption, a hybrid predictive model based on the combination of GA optimization and DFNN algorithm is proposed. In conventional FNN or DFNN-based predictive models, the architecture of the FNN and DFNN network was selected based on the experience, intuition or trial and error process, which was time-consuming and lack of accuracy. However, in the proposed GA-DFNN predictive model, the quantity of hidden layers, the quantity of neurons in each hidden layer, the activation algorithm in each layer as well as the learning process for weighting factors are determined through GA optimization. As a result, the proposed GA-enhanced DFNN predictive model is a self-directed, systematic and well-behaved process to guarantee the optimal DFNN architecture with better prediction accuracy and less computation complexity.

The actual outdoor weather data from weather station and the measured electricity consumption data from the building management system of a real-world campus building from one entire year were adopted as historical database to train the proposed GA-enhanced DFNN predictive model, while the datasets from the other 6 months were used for testing purpose. The main findings are summarized as below:

- For day-ahead hourly electricity consumption prediction, the optimal quantity of hidden layer is 4, with 8, 5, 7 and 9 neurons in each hidden layer, and with the activation function of rectified linear unit in the first three hidden layers and exponential linear in the fourth hidden layer. The activation function in the output layer is exponential linear function while the optimal optimization approach is NADAM.
- For week-ahead daily electricity consumption prediction, the optimal quantity of hidden layer is 2. There are 5 neurons in the first hidden layer, while exponential linear function is adopted as the activation function. There are 9 neurons in the second hidden layer, while hyperbolic tangent is adopted. The activation function in the output layer is exponential linear while the optimal optimization approach is ADAMAX.
- For day-ahead hourly electricity consumption prediction, the *MAPE, R^2, RMSE* and *MAE* was 7.885%, 0.970, 3.652 kWh and 2.675 kWh for testing case, respectively; For week-ahead daily electricity consumption prediction, the *MAPE, R^2, MPE* and *RMSE* was 5.326%, 0.966, 47.24 kWh and 35.52 kWh for testing case, respectively.
- For day-ahead hourly and week-ahead daily electricity consumption prediction, the proposed GA-DFNN predictive model has 3.4% and 34.5% reduction in *MAPE* compared to the reference FNN models with single hidden layer, demonstrating that DFNN models with multiple layers can better reveal the complex relationship among various input datasets and output dataset.
- For day-ahead hourly and week-ahead daily electricity consumption prediction, the proposed GA-DFNN predictive model has 5.0% and 33.3% reduction in *MAPE* compared to the reference DFNN models with other architecture, demonstrating that GA optimization has the ability to find the optimal architecture for DFNN models.
- For day-ahead hourly and week-ahead daily electricity consumption prediction, the proposed GA-DFNN predictive model has 9.7% and 33.8% reduction in *MAPE* than the reference RS-DFNN model, demonstrating that GA optimization has better optimization performance than random search optimization in determining the optimal architecture for machine learning models.
- For day-ahead hourly and week-ahead daily electricity consumption prediction, the proposed GA-DFNN predictive model has 12.2% and 32.7% reduction in *MAPE* than the reference LSTM, while 23.8% and 65.0% reduction in *MAPE* than the reference TCN models, demonstrating that DFNN has better performance in building electricity consumption prediction than time-series based machine learning models.

Acknowledgment

The authors would like to acknowledge and express their sincere gratitude to The Department for Business, Energy & Industrial Strategy (BEIS) 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] Zhao GY, Liu ZY, He Y, Cao HJ, Guo YB. Energy consumption in machining: classification, prediction, and reduction strategy. Energy 2017;133:142–57.
[2] Sieminski A. International energy outlook. Energy Inf Adm 2017:5–30.
[3] Luo XJ, 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 Convers Manag 2017;154:538–61.
[4] Luo XJ, Fong KF. Development of integrated demand and supply side management strategy of multi-energy system for residential building application. Appl Energy 2019;242:570–87.
[5] Pérez-Chacón R, Luna-Romera JM, Troncoso A, Martínez-Álvarez F, Riquelme JC. Big data analytics for discovering electricity consumption patterns in smart cities. Energies 2018;11(3):683.
[6] Wang Z, Srinivasan RS. A review of artificial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models. Renew Sustain Energy Rev 2017;1:796–808.
[7] Zhao Y, Zhang C, Zhang Y, Wang Z, Li J. A review of data mining technologies in building energy systems: load prediction, pattern identification, fault detection and diagnosis. Energy Built Environ 2020;2:149–64.
[8] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, Han M, Zhao X. A review of data-driven approaches for prediction and classification of building energy consumption. Renew Sustain Energy Rev 2018;1:1027–47.
[9] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192–205.

[10] Luo XJ, Lukumon O, Anuoluwapo A, Chukwuka M, Olugbenga A, Lukman A. Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands. Adv Eng Inform 2019;41:100926.

[11] Luo XJ. A novel clustering-enhanced adaptive artificial neural network model for predicting day-ahead building cooling demand. J Build Eng 2020:101504.

[12] Singh S, Hussain S, Bazaz MA. Short term load forecasting using artificial neural network. 2017 Fourth International Conference on Image Information Processing (ICIIP), Shimla 2017:1–5. doi:10.1109/ICIIP.2017.8313703.

[13] Mena R, Rodríguez F, Castilla M, Arahal MR. A prediction model based on neural networks for the energy consumption of a bioclimatic building. Energy Build 2014;82:142–55.

[14] Kusiak A, Li MY, Zhang ZJ. A data-driven approach for steam load prediction in buildings. Appl Energy 2010;87:925–33.

[15] Deb C, Eang LS, Yang J, Santamouris M. Forecasting diurnal cooling energy load for institutional buildings using artificial neural networks. Energy Build 2016;121:284–97.

[16] Ahmad T, Chen H, Shair J, Xu C. Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management. Int J Refrig 2019;98:399–409.

[17] Yang J, Hugues R, Radu Z. On-line building energy prediction using adaptive artificial neural networks. Energy Build 2005;37:1250–9.

[18] Close R. Theory of the backpropagation neural network. In: Proceedings of IEEE international conference on neural networks, 1; 1989. p. 593–605.

[19] Wang L, Eric WML, Richard KKY. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. Appl Energy 2018;228:1740–53.

[20] . NeuroShell 2 manual. Frederick, MA: Ward System Group, Inc.; 1993.

[21] Bagnasco A, Fresi F, Saviozzi M, Silvestro F, Vinci A. Electrical consumption forecasting in hospital facilities: an application case. Energy Build 2015;103:261–70.

[22] Ahmad MW, Mourshed M, Rezgui Y. Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build 2017;147:77–89.

[23] Li K, Xie X, Xue W, Dai X, Chen X, Yang X. A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. Energy Build 2018;174:323–34.

[24] Alnaqi AA, Moayedi H, Shahsavar A, Nguyen TK. Prediction of energetic performance of a building integrated photovoltaic/thermal system thorough artificial neural network and hybrid particle swarm optimization models. Energy Convers Manag 2019;183:137–48.

[25] Li K, Hu C, Liu G, Xue W. Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. Energy Build 2015;108:106–13.

[26] Muralitharan K, Sakthivel R, Vishnuvarthan R. Neural network based optimization approach for energy demand prediction in smart grid. Neurocomputing 2018;273:199–208.

[27] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015;61:85–117.

[28] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;321:436–44.

[29] Zekić-Sušac M, Mitrović S, Has A. Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. Int J Inf Manag 2020:102074.

[30] Banda E, Folly KA. Short term load forecasting using artificial neural network. In: 2007 IEEE lausanne power tech. IEEE; 2007 Jul 1. p. 108–12.

[31] Torres JF, Gutiérrez-Avilés D, Troncoso A, Martínez-Álvarez F. Random hyper-parameter search-based deep neural network for power consumption forecasting. Advances in computational intelligence IWANN 2019 Lecture notes in computer science, 11506. Cham: Springer; 2019.

[32] Park SK, Moon HJ, Min KC, Hwang C, Kim S. Application of a multiple linear regression and an artificial neural network model for the heating performance analysis and hourly prediction of a large-scale ground source heat pump system. Energy Build 2018;165:206–15.

[33] Lu R, Hong SH. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. Appl Energy 2019;236:937–49.

[34] Schalkoff RJ, York New. Artificial neural networks. McGraw-Hill; 1997.

[35] John D, Elad H, Singer Yoram. Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 2011;12:2121–59.

[36] Kingma DP, Ba JL. Adam: a method for stochastic optimization. In: International conference on learning representations; 2015. p. 1–13. (2015).

[37] Dogo EM, Afolabi OJ, Nwulu NI, Twala B, Aigbavboa CO. A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. In: 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS). IEEE; December 2018. p. 92–9.

[38] Vani S, Rao TM. An experimental approach towards the performance assessment of various optimizers on convolutional neural network. In: 2019 3rd international conference on trends in electronics and informatics (ICOEI). IEEE; April 2019. p. 331–6.

[39] Goldberg D. Genetic algorithms in search, optimization, and machine learning. Reading, Massachusetts: Addison Wesley; 1989.

[40] Mitchell M. An introduction to genetic algorithms. MIT press; 1998.

[41] Oh BK, Glisic B, Park SW, Park HS. Neural network-based seismic response prediction model for building structures using artificial earthquakes. J Sound Vib 2020;468:115109.

[42] Ren T, Liu X, Niu J, Lei X, Zhang Z. Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network. J Hydrol (Amst) 2020:124783.

[43] Bahman AM, Ebrahim SA. Prediction of the minimum film boiling temperature using artificial neural network. Int J Heat Mass Transf 2020;155:119834.

[44] Chollet F. (2015) keras, GitHub. https://github.com/fchollet/keras.

[45] Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C., Zheng X. TensorFlow: large-scale machine learning on heterogeneous distributed systems. 2015.

[46] https://dpds.weatheronline.co.uk/historical_data/weather_stations_download.

[47] https://www.accuweather.com.

[48] https://weather.com/en-GB/.

[49] https://www.metoffice.gov.uk/.

[50] Berriel RF, Lopes AT, Rodrigues A, Varejao FM. and Oliveira-Santos T. Monthly energy consumption forecast: a deep learning approach. In 2017 international joint conference on neural networks (IJCNN), IEEE, 4283–4290

[51] Zur RM, Jiang Y, Pesce LL, Drukker K. Noise injection for training artificial neural networks: a comparison with weight decay and early stopping. Med Phys 2009;36(10):4810–18.

[52] Prechelt L. Early stopping-but when?. In: Neural networks: tricks of the trade. Berlin, Heidelberg: Springer; 1998. p. 55–69.

[53] Torres JF, Galicia A, Troncoso A, Martínez-Álvarez F. A scalable approach based on deep learning for big data time series forecasting. Integr Comput Aided Eng 2018;25(4):335–48.