Forecasting building energy consumption: Adaptive long-short term memory neural networks driven by Genetic Algorithm

X.J. Luo, Lukumon O. Oyedele*

Big Data Enterprise and Artificial Intelligence Laboratory (Big-DEAL)
University of the West of England (UWE), Frenchay Campus, Bristol, United Kingdom
*Corresponding author: L.Oyedele@uwe.ac.uk

The real-world building can be regarded as a comprehensive energy engineering system; its actual energy consumption depends on complex affecting factors, including various weather data and time signature. Accurate energy consumption forecasting and effective energy system management play an essential part in improving building energy efficiency. The multi-source weather profile and energy consumption data could enable integrating data-driven models and evolutionary algorithms to achieve higher forecasting accuracy and robustness. The proposed building energy consumption forecasting system consists of three layers: data acquisition and storage layer, data pre-processing layer and data analytics layer. The core part of the data analytics layer is a hybrid genetic algorithm (GA) and long-short term memory (LSTM) neural network model for accurate and robust energy prediction. LSTM neural network is adopted to capture the interrelationship between energy consumption data and time. GA is adopted to select the optimal architecture for LSTM neural networks to improve its forecasting accuracy and robustness. The hyper-parameters for determining LSTM architecture include the number of LSTM layers, number of neurons in each LSTM layer, dropping rate of each LSTM layer and network learning rate. Meanwhile, the effects of historical weather profile and time horizon of past information are also investigated. Two real-life educational buildings are adopted to test the performance of the proposed building energy consumption forecasting system. Experiments reveal that the proposed adaptive LSTM neural network performs better than the existing feedforward neural network and LSTM-based prediction models in accuracy and robustness. It also outperforms those LSTM networks whose hyper-parameters are determined by grid search, Bayesian optimisation and PSO. Such accurate energy consumption prediction can play an essential role in various areas, including daily building energy management, decision making of facility managers, building information model designs, net-zero energy operation, climate change mitigation and circular economy.

Keywords: Long-short term memory; Genetic algorithm; Building energy consumption; Energy forecast; Energy management system; Adaptive.

1. Introduction

The rapid growth of population and fast development of the economy has substantial impacts on global energy consumption and environmental concerns. The majority of people spend 90% of their daily lives in buildings, which further increases the building energy consumption to satisfy indoor activities and thermal comfort [1]. The building sector contributes towards 39% of global energy consumption and 38% of worldwide greenhouse gas emissions [2].

1.1 The importance of accurate building energy prediction

Building energy demand management is critical in decreasing primary energy consumption and mitigate climate change challenges [3, 4]. One of the essential parts of the energy management system is accurate and robust energy consumption forecasting. Effective day-to-day management of electric utility and energy devices relies on energy demand forecasting. Furthermore, an accurate and robust building energy consumption forecasting model can inspire energy-efficiency policies for energy consumption reduction, environmental pollution alleviation, and sustainable economic development. Building energy consumption depends on various influential factors, including weather condition, building properties, time, and occupancy. It is challenging to develop an accurate and robust energy consumption forecasting model owing to the non-linear, nonstationary and multi-seasonality features of the energy consumption data,

1.2 Existing approaches of building energy prediction

Energy consumption prediction has become a research problem since the early 1990s [5]. In general, energy consumption forecast techniques can be classified into conventional statistical forecasting methods and machine learning-based load forecasting techniques. Traditional load forecasting techniques include regression, multivariate adaptive regression [6], exponential smoothing and iterative reweighted least-squares technique. However, conventional forecasting methods have limited capability in representing non-linear, nonstationary and multi-seasonality features of datasets. These techniques are usually more complex in computational operations, require higher computational time, and result in lower prediction accuracy than machine learning-based forecasting techniques. On the other hand, machine learning-based forecasting techniques are generally based on artificial intelligence techniques such as *k*-nearest neighbour [7], autoregressive integrated moving average [6], fuzzy logic, neural networks, component-based machine learning method [8] and support vector regression [6]. Through big-data transformation, transmission and processing, machine learning models can effectively recognise various random and comprehensive function forms.

Among various machine learning-based building energy prediction models, neural networks have been demonstrated to be the most effective and accurate. Neural networks have the ability to adequately approximate the comprehensive non-linear relationship among the input and output datasets of a complex energy system with arbitrary and precision. Feedforward neural network models generally try to construct a direct mapping among comprehensive input historical data (i.e. weather condition, time signature and indoor sensor measurement) and output energy consumption data to achieve the forecasting purpose. However, due to the lack of time correlation in the data sequence, feedforward neural network models cannot capture the inter-relationship between energy data and time. Therefore, its capability in time series forecasting is limited. On the other hand, long short-term memory (LSTM) neural network is proposed to overcome such disadvantages. Through implementing recurring connections to neurons, LSTM neural network is able to represent the sequence-to-sequence mapping between input and output data. Since each time step's output is influenced by the input of the previous time step, the 'memory' characteristic can be recognised [9].

1.3 Existing approaches for hyper-parameters tuning

LSTM network has numerous hyper-parameters that must be modified by the developers, such as the number of layers, number of neurons in each layer, learning rate, dropping rate and activation function. Manual search and automatic search are two widely adopted methods for hyper-parameter optimisation. Manual search is largely dependent on the fundamental intuition and experience of users. Data science experts can identify the critical hyper-parameters which have superior impacts in determining the mathematical relationship between input and output datasets. However, it generally requires model developers to be equipped with background knowledge and practical experience. Thus it is challenging to be applied by non-expert users. Moreover, most of the hyper-parameter optimisation processes are not reproducible. Furthermore, with the increasing number and value range of hyper-parameters, it becomes increasingly challenging for manually data processing.

To overcome the drawbacks of manual search, automatic search algorithms have been proposed. The objective function of the hyper-parameter optimisation problem is like a black-box function. Thus, conventional optimisation techniques such as the Newton method or gradient descent cannot be applied. The existing automatic search algorithms include grid search, Bayesian optimisation and evolutionary optimisation. Grid search [10], also named exhaustive search, tests the machine learning model with all the possible combination of hyper-parameter values. Although grid search can automatically optimise and theoretically obtain the optimal global value of hyper-parameters, its costs large computational time consumption with the large number and wide value range of hyper-parameters.

To solve the problem of expensive computational cost in grid search, Bayesian optimisation is proposed. Using Bayesian formula, Bayesian optimisation obtains posterior information of the function distribution through integrating prior information of the unknown function with sample information [11]. The optimal value of the optimisation function can be estimated through the obtained posterior information. However, Bayesian optimisation is generally based upon the assumption that its optimisation function obeys the Gaussian distribution. Evolutionary optimisation [12] is inspired by the neural plasticity of the human cortex. It can also be adopted to estimate the hyper-parameters of neural networks automatically.

2. Related works

To investigate the development status of building energy prediction technologies and LSTM techniques, the literature review consists of three parts. Firstly, the conventional artificial neural network-based energy prediction models are investigated. Secondly, an overview of LSTM neural network-based energy prediction models is conducted. Thirdly, the LSTM neural network in other engineering applications is also explored. Through the comprehensive literature review, the significant research gaps are identified while the major research objective of this study is outlined.

2.1 Literature review on conventional artificial neural network

Various types of neural network-based energy prediction model have been proposed. Chang *et al.* [13] proposed a backpropagation neural network for country-wide electricity consumption prediction using a small dataset. The proposed neural network consists of one hidden layer, while there are two neurons in the hidden layer. Luo *et al.* [14] developed a feedforward neural network-based prediction model for cooling and heating loads in buildings. The inputs to the prediction model included multiple real-time temperature sensor readings, forecasted weather data profile and historical energy consumption data. Different numbers of neurons were tested and selected for various building sub-zones. Later on, Luo *et al.* [15] developed a clustering-enhanced feedforward neural network prediction model for building cooling demand. Several neural network-based sub-models were adopted for the year-round prediction, while a trial-and-error process determines the number of neurons in each sub-model. Luo *et al.* [16] also developed a hybrid genetic algorithm (GA) and feedforward neural network-based prediction model, in which GA was adopted to choose the optimal network architecture. The input datasets to the prediction model were composed of historical weather data and time signatures, while the historical energy consumption profile was not considered. The feature of training data input, the structure of the neural network, and performance indicator of those previously developed neural network-based building energy consumption prediction models are summarised in Table 1. From the comprehensive literature review, it is found that most of the ANN models are not able to reflect the inter-relationship between energy consumption and timeline.

2.2 *Literature review on LSTM neural network-based energy prediction model*

The selection of optimal LSTM neural network hyper-parameters varied largely in different pieces of literature. The feature of training data input, the structure of the LSTM neural network, and the performance indicator of the previously developed LSTM neural network-enabled building energy consumption prediction models are summarised in Table 2.

In most of the literature, the hyper-parameters were selected based on the authors' own experience. Jian *et al.* [31] adopted the LSTM neural network for the long-term energy consumption prediction of a cooling system. It is found that the proposed LSTM method resulted in 19.7% lower at root mean square error than the reference feedforward neural network. Khafaf *et al.* [32] proposed an LSTM neuron network model to forecast a 3-day ahead energy consumption of clusters of energy users. Wang *et al.* [33] adopted the LSTM neural network for power consumption prediction and grid anomalies detection. Khan *et al.* [34] adopted the hybrid convolutional neural network with LSTM autoencoder for energy consumption prediction in residential and commercial buildings. Wei *et al.* [35] proposed a hybrid singular spectrum analysis and LSTM model for daily natural gas consumption prediction. Singaravel *et al.* [36] tested the performance of LSTM for building energy consumption at the design stage. 201 design cases were evaluated with four different configurations of the LSTM model. It is found that LSTM models have higher accuracy and higher computation speed than ANN models. Zhou *et al.* [37] proposed an LSTM model to predict the energy consumption of air-conditioning systems. The constant experimentation was conducted to find the best hyper-parameters, including the number of training steps, the length of the time series entering the LSTM and the learning rate.

Trial-and-error process and numerical experiment are also generally adopted to select the optimal hyper-parameters of the LSTM neural network. Xue *et al.* [38] proposed an attention LSTM neural network to predict the heat load for the district heating system. It is found that the LSTM model can effectively memorise characteristics of the long-term historical heating load. A trial-and-error process is conducted to select the number of neurons and type of activation function in the hidden layer. Somu *et al.* [39] proposed a hybrid *k*-means clustering, convolutional neural network and LSTM model for building energy consumption prediction. The number of neurons in the LSTM layer was fixed at 5100. Wang *et al.* [40] demonstrated a multi-energy load prediction model based on an encoder-decoder LSTM neural network. Markvoic *et al.* [41] presented an approach for predicting the day-ahead energy plug-in loads using LSTM neural networks. Rahman *et al.* [42] adopted an LSTM-based deep recurrent neural network to predict electricity consumption of commercial and residential buildings. Arranz *et al.*

[43] proposed an LSTM-based predictor to forecast the day-ahead of power consumption of a heating, ventilation and air-conditioning system. The number of LSTM layers was fixed at 2, while both LSTM layers were assumed to have the same number of neurons. The number of neurons and the learning rate was selected by a trial-and-error process. Julio *et al.* [44] proposed an LSTM method for predicting the energy consumption of an educational building. The hyper-parameters, such as the number of neurons in each LSTM layer, number of epochs, number of batches, activation function in the hidden and output layer, were selected by numerous experiments. Laib *et al.* [45] proposed a 2-layer LSTM neural network for national-wide natural gas consumption prediction. Different experimental set-ups were conducted to find out the optimal architecture for LSTM neural network. Lu *et al.* [46] proposed an LSTM network model for regional thermal load prediction. The LSTM was adopted to process the intrinsic temporal relationships among input and output variables.

The Bayesian optimisation method is a recently adopted approach to select the appropriate hyper-parameters to achieve optimal prediction performance. Jin *et al.* [47] proposed an encoder-decoder architecture with a gated recurrent units recurrent neural network prediction model for short-term electric power load forecasting. The hyper-parameters include the number of network layers, number of network units, batch size, model learning approach and number of training epochs. Yang *et al.* [48] proposed an LSTM-attention-embedding model based on Bayesian optimisation to predict the day-ahead PV power output. The Bayesian optimisation is adopted to select the optimal values for the time window, number of statistical features and number of the combined features. Munem *et al.* [49] proposed a multivariate Bayesian optimisation based LSTM neural network to forecast the residential electric power load for the next hour. The Bayesian optimisation algorithm is conducted to select the best-fitted hyper-parameter values, including the number of LSTM cells, activation function, optimisation method, neurons in the hidden layer, dropout rate and batch size.

Meanwhile, the hybrid evolutionary optimisation algorithm, such as GA and PSO is also integrated with LSTM to improve its prediction accuracy. Most of the GA and PSO is adopted to select the optimal weight matrix or part of the LSTM hyper-parameters. He *et al.* [50] proposed a hybrid short-load forecasting method with variational mode decomposition and LSTM networks, while the hyper-parameters of the LSTM network is optimised using the Bayesian Optimisation algorithm. Kim *et al.* [51] proposed an LSTM network for residential energy consumption prediction. The PSO is adopted to find out the optimal hyper-parameters such as learning rate, layer size and dropout rate. Yang *et al.* [52] proposed a hybrid prediction model using extreme learning machine, recurrent neural network, and support vector machines. PSO is adopted to select the optimal weight matrix among these three networks. Bouktif *et al.* [53] adopted the LSTM network to construct forecasting models for short to medium term aggregate load forecasting. The optimal time lags and the number of layers of the LSTM network were determined by GA, while the number of neurons, activation function and learning approach was selected using a trial-and-error approach. Guo *et al.* [54] proposed a short-term forecasting model of LSTM neural network considering demand response. The improved GA is used to obtain the best weight matrix for LSTM. However, the structure of the LSTM network is fixed based on the developer's experience. Su *et al.* [55] proposed a hybrid wavelet transform and LSTM model for hourly natural gas demand forecasting. The number of LSTM layers is determined through trial-and-error, while the number of neurons in each LSTM layer is optimised through GA.

Table 1. Literature review of conventional artificial neural network-based energy prediction models.

Ref	Model	Length of datasets	Model input	Time step	Number of neurons in hidden layer	Number of hidden layer(s)	Learning approach	Activation function	Prediction objective	Performance evaluation	Year	Time correlation in the data sequence
13	Back propagation neural network	3 years	Electricity consumption data	Monthly	2	1	-	-	Electricity	MAPE	2016	×
14	Feedforward neural network	2 years	Weather data, indoor sensor data, previous energy data	Hourly	Tested 60-80	1	LM	ReLU	Heating Cooling	MAPE	2019	×
15	Feedforward neural network	2 years	Weather data Time index	Hourly	Tested 2-20	1	LM	ReLU	Cooling	MAPE	2020	×
16	Deep neural network	1 year and 6 months	Weather data Time signature	Hourly Daily	GA optimised	GA optimised	GA optimised	Tanh, Sigmoid, ReLU, ELU	Electrcity	MAPE RMSE R^2	2020	×
17	Feedforward neural network	1 year training 1 year testing	Weather data	Daily	Tested 4-9	1	LM	Trial-and-error on exponential, logistic, tanh	Heating Cooling	MAE MAPE	2010	×
18	ANN	1 year for training & testing	Previous energy data	Daily	20 (trial-and-error)	1	LM	Sigmoid	Cooling	R^2	2016	×
19	ANN	1 year for training & testing	Weather data	Hourly	10, 15, 20	1	LM	Gauss–Newton function	Cooling	MAE R^2	2019	×
20	ANN	1 year for training & testing	Weather data	Hourly	$2N_{in}$+1	1	LM	Logistic sigmoid	Chiller electric demand	MSE	2015	×
21	ANN and an ensemble	1 month for training	Weather data Previous energy data	Hourly	(N_{in} + N_{out})+ sqrt(N_{sample})	1	LM	Hyperbolic tangent	Cooling	MAE R^2	2018	×
22	ANN	1 year training 1 year testing	Weather data Time index Previous energy data	Half-hourly	Trial-and-error	Tested 1-10	LM	LReLU, PReLU, ELU, SELU	Electricity	RMSE MAPE R^2	2019	×
23	ANN	130 days for training & 21 days for testing	Weather data Occupancy level	Hourly	-	1	LM	-	Electricity	RMSE NMBE	2020	×
24	DNN	1 year for training 1 week for testing	Time index Previous load	Hourly	30, 20, 10	3	LM	Sigmoid	Electricity	MAE MAPE	2019	×
25	ANN	4 months for training & testing	Weather data Previous load	Hourly	sqrt(N_{in} + N_{out})+1-10	1	PSO	Sigmoid	Electricity	MAPE	2015	×
26	CNN	5 years for training& testing	Previous energy data	Hourly Daily	-	1	PSO GA	Sigmoid	Electricity	MSE	2018	×
27	Elman neural network	1 year for training	Weather data	Daily	Tested 2-20	1	GA	-	Electricity	MSE	2017	×
28	ANN	4 months training & testing	Weather data Time index	Hourly	20	1	Teaching-learning	-	Electricity	MAPE RMSE	2018	×
29	DNN	287 days training 124 days testing	Weather data Time signature	Hourly	8	2	ADAM	ReLU	Heating Cooling	MSE RMSE MAE R^2	2020	×
30	ANN	1 month training 1 month testing	Weather data Previous energy data	Hourly	2 N_{in} +1	1	LM	Sigmoid	Cooling	RMSE CV	2006	×

Table 2. Literature review of LSTM neural network-enabled energy prediction models.

Ref.	Model	Model input	Time step	Length of historical data	Number of neurons in hidden layer	Number of hidden layer(s)	Learning rate	Activation function	Learning approach	Prediction objective	Accuracy	Year	Hyperparameters selection
31	LSTM neural network	Historical compressor and condenser energy consumption	Hourly	1 month for training, 4 months for testing	100	2	0.0006	sigmoid	-	Cooling	MSE RMSE MAE MAPE	2020	Experience
32	LSTM	Previous energy data	Half-Hourly	1 year	-	1	-	-	-	Household energy consumption	RMSE	2019	Experience
33	LSTM	Weather data, previous energy data	-	-	50	1	-	Sigmoid Tanh	Adam	Electricity consumption	MAE	2021	Experience
34	Hybrid CNN with a LSTM-AE	Weather data, time index, previous energy data	Minutely	5 years	-	-	-	Sigmoid	Adam	Household electricity consumption	MSE MAE RMSE MAPE	2019	Experience
35	LSTM	Weather data, previous energy data	Daily	1 year	20	1	0.004	Tanh	Adam	Citywide natural gas consumption	MAPE MAE RMSE R^2	2019	Experience
36	LSTM	Building design and thermos property data	Monthly	Energy usage of different types of buildings	60 70 50, 20 50, 20, 20	1 1 2 3	0.001	Relu	Adam	Energy consumption of the entire building	R^2	2018	Experience
37	LSTM	Previous energy data	Hourly	1 year	-	-	-	Sigmoid Tanh		Energy consumption of HVAC system	RMSE MAPE MAE	2020	Experience
38	LSTM	Weather data, indoor sensor data, previous load	Hourly	1 year	Tested (5, 8, 10, 20, 30, 40)	1	0.005	Tested (sigmoid, relu, tanh, elu, selu)	-	District heating system	MAPE MAE RMSE R^2	2019	Trial-and-error
39	CNN-LSTM	Weather data Various types of previous building energy data	15 mins	1 year	5100	Tested 1-3	-	Tanh		Building electricity	MSE MAPE MAE MAPE	2021	Trial-and-error
40	RNN	Weather data	Hourly	1 year	Tested 4-9	1	-	Tanh	-	Electricity Cooling Or heating	RMSE MAPE	2020	Numerous experiment
41	LSTM	Previous energy data	Hourly	1.5 year	Tested 10-100	3	0.01	Tanh	Adam	Plug-in loads	MSE R^2	2021	Numerous experiment
42	Deep RNN with LSTM	Weather data, time signature, previous energy data	Hourly	1 year	Tested 10-100	5	-	Tested between ReLu and sigmoid	Adam	HVAC	RMSE	2018	Numerous experiment
43	LSTM	Weather data, time index, previous energy data	Hourly	-	Tested 5-50	2	Tested 0.0001-0.05	ReLU	Adam	Energy consumption of HVAC system	MSE RMSE	2020	Numerous experiment
44	LSTM	Previous energy data	0.5 min	2 days	Tested 1-36	1	-	Sigmoid	-	Building electricity	RMSE	2019	Numerous experiment
45	LSTM-RNN	Weather data, time signature, previous energy data	Hourly	1 year and 10 months	Tested 20-100 in 1st layer, 20-50 in 2nd layer	2	Tested 0.1-0.01	Sigmoid Tanh	-	National-wide natural gas consumption	MAPE	2019	Numerous experiment
46	LSTM	Weather data	Hourly		Tested 2-20	1	Tested 0.001-0.01	Sigmoid Tanh	Tested SGD MBGD AdagradAdam	Thermal load in regional energy system	MAPE	2021	Numerous experiment
47	LSTM	Previous electric load	Hourly	3 years	8-128 at step of 8	2	-	Sigmoid Tanh	ADAM NADAM SGD	Regional electric load	RMSE MAE R^2	2021	Bayesian Optimization
48	LSTM	Weather data Previous PV output	15 minutes	25 months	128-64-32	3	0.001	Sigmoid Tanh	ADAM	PV power	MSE MAE R^2	2019	Bayesian Optimization

49	LSTM	Time information Previous load	Hourly	4 years	4-512	1	-	ReLU Linear Sigmoid Tanh ELU	SGD ADAM NADAM Adamax Adadelta Adagrad RMPSprop	Residential house electricity consumption	MSE MAE RMSE	2020	Bayesian Optimisation
50	LSTM	Weather data Day type Previous load	Daily	1 year	-	1	-	Sigmoid Tanh	ADMM	Regional power load	RMSE MAE MAPE R2	2019	Bayesian Optimisation
51	LSTM	Time information Previous load	Hourly	1 year	-	-	10^{-5}-10^{-1}	Sigmoid Tanh	ADMM	Household electric power consumption	MSE	2019	Grid search
52	ELM RNN SVM	Previous electric load	Hourly	Several months	-	4	-	Sigmoid Tanh	-	Regional electric load	MAE MAPE RMSE	2020	Experience
53	LSTM	Previous energy data	Half-hourly	9 years	20-100 by GA	1	-	Tested sigmoid, tanh and ReLU	Tested SGD, RMSProp and ADAM	Energy demand	MAE RMSE	2018	Empirical and GA
54	LSTM	Weather data and electricity price	Daily	3 years training 1 year testing	-	-	-	Sigmoid Tanh	-	Electricity price	MAPE	2021	Empirical
55	LSTM	Previous energy data	Hourly	-	[100,110,120, 130,140,150]	2(trial and error)	-	Sigmoid Tanh	-	Natural gas demand	MAE RMSE	2020	GA for number of neurons

2.3 Literature review on LSTM approaches in other engineering applications

On the other hand, LSTM based regression and classification approaches have been adopted in various other engineering applications. For example, Pandya *et al.* [56] adopted the LSTM model to develop the acoustic event assistive framework for identification, detection and recognition of unknown acoustic events of a residential building. Cai *et al.* [57] proposed a context-augmented LSTM method to predict a sequence of target positions from a sequence of observation. Both individual movement and workplace contextual information, including movements of neighbouring entities, working group information, and potential destination information, were integrated into the LSTM network with an encoder-decoder architecture. Zhao *et al.* [58] proposed a convolutional LSTM recognition model for recognising construction workers' postures from motion data captured by wearable inertial measurement units. Yang *et al.* [59] proposed a bidirectional LSTM network to analyse masonry workers' lower body movement and identify physical loading conditions. The training data for the LSTM network was collected from an ankle-worn inertial measurement units sensor. Based on the characteristics of the knowledge expression of procedural construction constraints in Chinese regulations, Zhong *et al.* [60] proposed a hybrid bidirectional LSTM and conditional random field model for the automatic extraction of the qualitative construction procedural constraints. Sun *et al.* [61] proposed a hybrid LSTM neural network and bagging ensemble learning strategy for obtaining accurate results of exchange rates forecasting and to improve the profitability of exchange rates trading. Chen *et al.* [62] proposed an LSTM model to train the Time-Between-Failure prediction model for predictive maintenance, thus reducing maintenance cost and achieve sustainable operational management. Lee *et* *al.* [63] adopted the LSTM model for human activity recognition and daily activities classification. The activities such as personal hygiene, eating and mobility, can be categorised using the image data from various sensors. Amer *et al.* [64] proposed a hybrid LSTM and recurrent neural networks model to automatically learn construction knowledge from historical project planning and scheduling records. Yin *et al.* [65] proposed a bidirectional LSTM model for timely and accurate prediction of key quality characteristics of separated coal during the coal preparation stage. Rashid *et al.* [66] proposed an LSTM model for automated, real-time, and reliable equipment activity recognition at the construction site. The synthetic training data is generated by time-series data augmentation techniques. Yan *et al.* [67] proposed a hybrid machine learning model integrating the LSTM and Bayesian optimisation approach to predict the time-weighted average pressure of shield supporting cycles. The optimised hyper-parameters include the Adam optimiser parameter, dropout layer parameter, LSTM cell number and batch size. Lv *et al.* [68] presented an improved LSTM network to predict the price of the stock and price of nickel metal, respectively. The PSO algorithm is adopted to optimise the weight matrix of the LSTM network.

2.4 Identification of research gaps

With the extensive literature study on building energy forecasting, the identified research gaps are highlighted as follows:

A. As summarised in Table 1, ANN [13-30] was the most commonly adopted predictions for building energy consumption. However, it lacks the time correlation in data sequence; thus, its effectiveness and accuracy in time series forecasting is limited.
B. As illustrated in Table 2, most of the literature, like References [31-37], generally employed a pure LSTM neural network for building energy consumption prediction. There was no empirical equation for LSTM hyper-parameter selection.
C. Moreover, trial-and-error process, numerous experiments, like grid search [38-46], were conducted to determine the optimal hyper-parameters of neural network models. However, time and computation limitations in real-time energy consumption prediction make it impossible to sweep through a parameter space and find the optimal set of parameters.
D. In some of the previous works, Bayesian optimisation [47-49] is adopted to select the optimal hyper-parameters of LSTM. However, Bayesian optimisation is generally based upon the assumption that its optimisation function obeys the Gaussian distribution.
E. Although some of the literature mentioned the hybrid evolutionary algorithm and LSTM neural network [50-56], the main purpose of the adopted evolutionary algorithm optimisation is to select the optimal weight matrix of LSTM or only one or two hyper-parameters of the LSTM network.
F. Furthermore, LSTM neural network has been adopted in various engineering fields, such as acoustic [56], workplace [57], construction [58-60, 64, 66], exchange rates trading [61], predictive

maintenance [62], human daily activities recognition [63], coal preparation [65], shield supporting [67], and economic stock [68]. It was generally integrated with other machine learning methods, such as the convolutional neural network. However, the approach for determining the optimal LSTM architecture has seldom been mentioned.

G. Meanwhile, static data, simulation data or benchmark datasets were adopted in most of the research works to demonstrate the performance of LSTM prediction models. There is a lack of real-world testing of the prediction model on a practical operational building.

2.5 Research objectives of this study

The adaptive LSTM neural network, with its optimal architecture, can provide accurate and robust building energy forecast in a real-world application. The remarkable contributions of this paper lie in the following aspects:

- To overcome the shortage of Research gap A, LSTM neural network is adopted for building electricity consumption. LSTM neural network can represent the sequence-to-sequence mapping between input and output data through recurring connections among neurons. LSTM can recognise the ‘memory’ characteristics; thus, the previous time step can influence the output of the current time step.
- To overcome the disadvantage of Research gaps B, C, D, E and F, GA optimisation is adopted to determine the optimal architecture of LSTM networks, thus improving forecasting accuracy and robustness. The whole-set optimal hyper-parameters include the number of LSTM layers, number of neurons in each LSTM layer, dropping rate of each LSTM layer, and learning rate, are selected by GA.
- To overcome the shortage of Research gap G, two real-world educational buildings are selected as case studies to test the performance of the proposed building energy forecasting system. The effects of historical weather profile and time horizon of past information are also investigated.

3. **Adaptive long short-term memory networks**

To improve forecasting accuracy and robustness, GA is adopted to select the optimal hyper-parameters for LSTM neural networks. Therefore, the LSTM neural network is adaptive to the different features of the energy consumption of various buildings.

3.1 LSTM neural network

LSTM neural network is a special and improved architecture of the recurrent neural network (RNN). It employs gate units and ‘self-connected memory cells’ to extract the underlying complex temporal dependencies in time-series data. In the LSTM neural network, the memory block in the recurrent layer is adopted to overcome the vanishing and exploding gradient problems. Fig. 1 illustrates the structure of a memory block. The memory cells in the memory blocks are self-connected. Three gates are introduced to store temporal sequences, including the input gate, forget gate and output gate. The detailed description of LSTM neural network algorithm can be further found in [69]. The initial values of the weight matrices (i.e. $\boldsymbol{W}_i^x$, $\boldsymbol{W}_c^x$, $\boldsymbol{W}_f^x$, W_o^x, $\boldsymbol{W}_i^h$, $\boldsymbol{W}_c^h$, $\boldsymbol{W}_f^h$ and W_o^h) and bias (i.e. b_i, b_c b_f and b_o) are updated by standard gradient descent (SGD) method. However, the performance of the SGD depends on the architecture of the LSTM neural network.

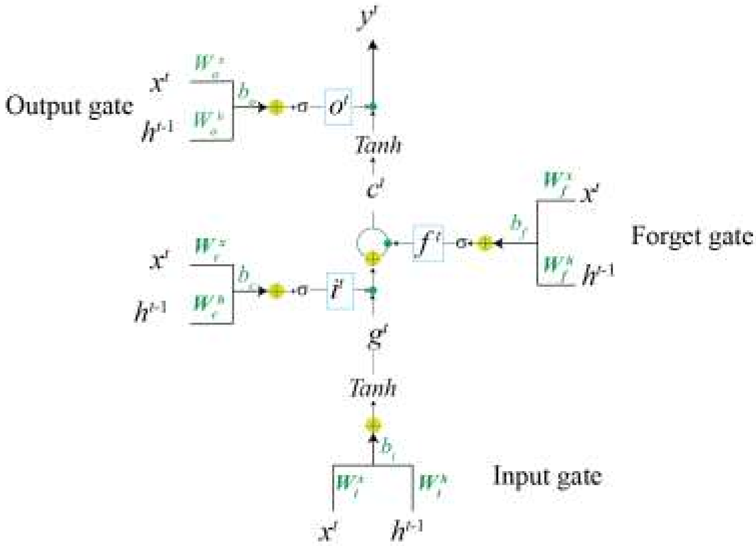


Fig. 1. The architecture of a memory block of LSTM.

3.2 GA enabled adaptive LSTM neural network

The architecture of the LSTM neural network has significant impacts on its prediction performance mainly due to the adoption of the backpropagation algorithm and SGD method in determining various weight matrices and bias. To automate the process and allow a multidimensional space of possible architectures, GA optimisation is adopted to determine the optimal values for the LSTM hyper-parameters, including the number of LSTM layers, the number of neurons in each LSTM layer, dropping rate of each LSTM layer and learning rate. GA is a powerful evolutionary algorithm based on the theory of natural selection and genetic mechanism [70]. The advantages of GA optimisation include no restrictions on the form of problems, strong searching performance and a high convergence rate [71]. The type of decision variables, along with feasible values, are summarised in Table 3. There is a total of $(10 \times 9 + 10^2 \times 9^2 + 10^3 \times 9^3) \times 4 = 2{,}948{,}760$ types of LSTM architectures.

Table 3. Decision variables of PSO for LSTM neural network.

Number of neurons in each hidden layer	{10, 20, 30, 40, 50, 60, 70, 80, 90, 100}
Number of hidden layers	{1, 2, 3}
Dropping rate	{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}
Learning rate	{0.001, 0.01, 0.05, 0.1}

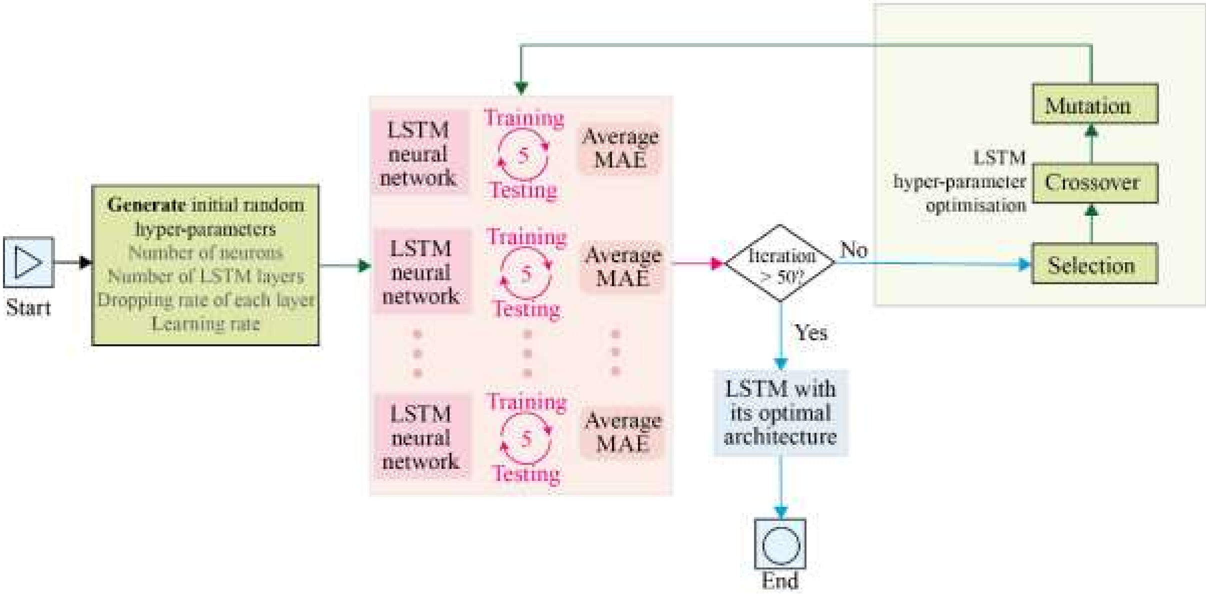


Fig. 2. Flowchart of the adaptive LSTM neural network.

The procedure of the proposed adaptive LSTM neural network model is illustrated in Fig. 2. The LSTM neural network model, with its hyper-parameters, is treated as a chromosome in the GA optimisation. The population size of GA optimisation is chosen at 20 [16]. In the beginning, n_{pop} = 20 LSTM neural networks are randomly assigned with different hyper-parameters. These LSTM neural networks are trained and tested using the same training datasets from the historical database. Since dataset training is a random process, there exists a slight difference in the resulted mean absolute error (MAE) when the training process is repeated. Therefore, each LSTM neural network is trained and tested 5 times using different initial weighting matrices and bias, while the average *MAE* value is adopted as the GA objective function. Based upon the average *MAE* value, selection, crossover and mutation processes are conducted to update each particle (i.e. hyper-parameters of LSTM neural network). The iterative process between GA and LSTM neural network is repeated 50 times to achieve the convergent optimal LSTM neural network architecture.

4. Energy consumption forecasting system

The proposed energy consumption forecasting system consists of three layers, namely, data acquisition and storage layer, data pre-processing layer and data analytics layer. The structure of the proposed energy consumption forecasting system is illustrated in Fig. 3.

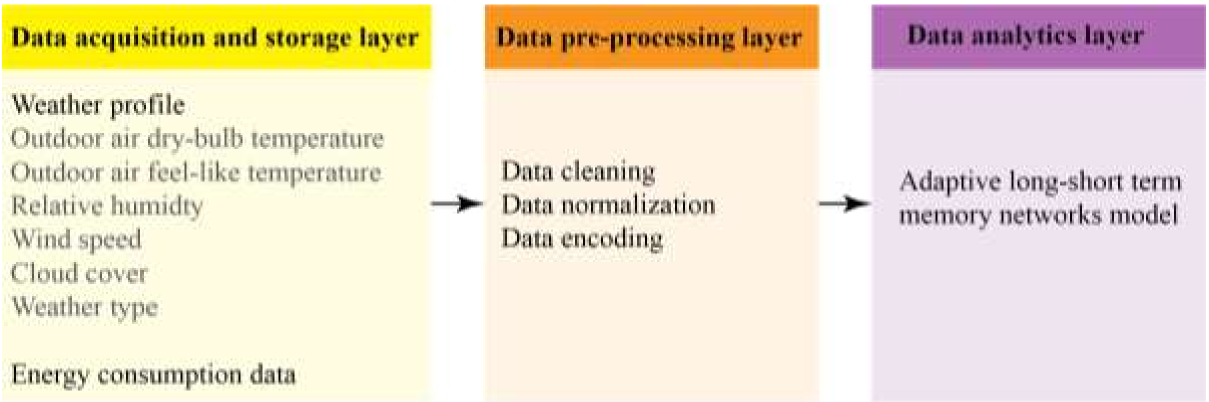


Fig.3. Structure of the proposed energy forecasting system.

4.1 Data acquisition and storage layer

The influential factors to energy consumption include weather conditions, occupancy behaviour and operating agendas of lighting and air conditioning. The occupancy behaviour and operating agendas is generally not available in a normal building. However, in the educational building, human behaviour and operating agendas of lighting and air conditioning is comparatively steady among working days and correlated with the hour of the day. As LSTM is time-correlated, the time signatures can be embedded in the LSTM network itself. On the other hand, historical weather profile from weather report websites and energy consumption data from energy management system formulates as input database for training and testing the adaptive LSTM neural network energy forecasting model. Meanwhile, forecasting data from weather forecast websites would be accessed to enable real-time energy prediction. Due to its comprehensive effect on building energy consumption, the historical weather data includes:

- outdoor air dry-bulb temperature $T_{db,i}$
- outdoor air feel-like temperature $T_{fl,i}$
- relative humidity RH_i
- wind speed V_i
- cloud cover θ_i
- weather type β_i

The weather type includes the sky is blue, few clouds, scattered clouds, broken clouds, overcast clouds, mist, fog, light rain, moderate rain, shower rain and drizzle.

4.2 Data pre-processing layer

Three data pre-processing steps are conducted to prepare the input dataset for the energy forecasting model, including data cleaning, data normalisation and data encoding.

4.2.1 Data cleaning

The acquired and stored data in the first layer generally contains outliers and missing data due to sensor faults and transmission errors. The outlier values are replaced using the theory of three-sigma rule of thumb [72].

$$x_t = \begin{cases} avg(X) + 2.std(X) \; if \; x_i > avg(X) + 2.std(X) \\ x_t \qquad\qquad otherwise \end{cases} \tag{1}$$

where X is the vector that consists of x_t, $avg(X)$ is the average value of X, while $std(X)$ is the standard deviation of X.

The missing values in the energy consumption dataset are imputed using a data interpolation method.

$$x_t = \begin{cases} \frac{x_{t-1}+x_{t+1}}{2} & x_t \in NaN;\; x_{t-1}, x_{t+1} \notin NaN \\ \frac{x_{t-1}+x_{t-2}}{2} & x_t, x_{t+1} \in NaN;\; x_{t-1}, x_{t-2} \notin NaN \\ x_t & x_t \notin NaN \end{cases} \tag{2}$$

4.2.2 Data normalisation

Due to the different magnitude among various types of weather data, the meteorological data, including outdoor air dry-bulb temperature, outdoor feel-like temperature, relative humidity, cloud cover ratio, wind speed and energy consumption, are normalised into the range of [0-1] using the min-max scaling approach [73]:

$$x_t^{'} = \frac{x_t - \min\limits_{1 \le t \le 365 \times 24} x_t}{\max\limits_{1 \le t \le 365 \times 24} x_t - \min\limits_{1 \le t \le 365 \times 24} x_t} \tag{3}$$

4.2.3. Data encoding

The one-hot encoding approach is adopted to pre-process weather type. The one-hot encoding can transform a single variable with δ observations and β distinct values to β binary variables with δ observations individually [74]. Each observation indicates the presence (i.e. 1) or absence (i.e. 0) of the dichotomous binary variable. The encoding results are summarised in Table 4.

Table 4. One-hot encoding of weather type

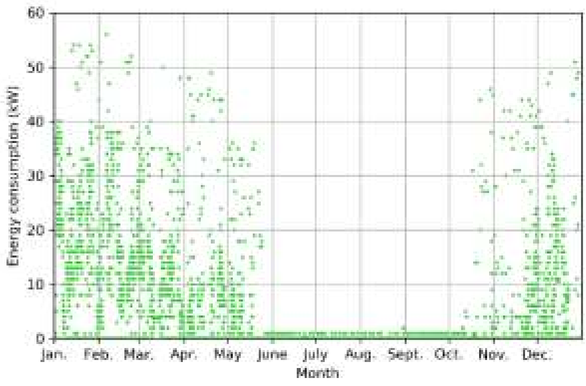
Weather type	Results of one-hot encoding
sky is blue	00000000001
few clouds	00000000010
scattered clouds	00000000100
broken clouds	00000001000
overcast clouds	00000010000
mist	00000100000
fog	00001000000
light rain	00010000000
moderate rain	00100000000
shower rain	01000000000
drizzle	10000000000

4.3 Data analytics layer

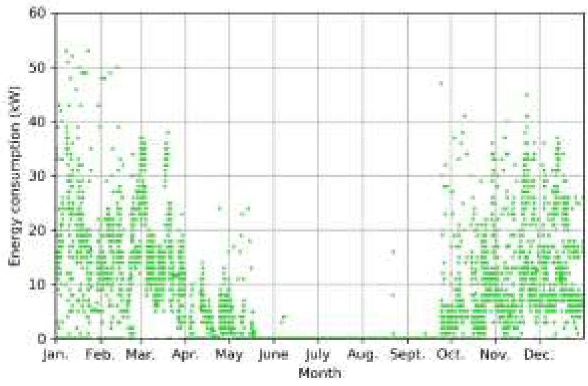
The input datasets **X** include six weather index and energy consumption over the past horizon h. $\mathbf{X} = \{X_t | t = h, h+1, \dots, 8760\}$, and $X_t = \{x_{t,j} \middle| j = 1, 2, 3, 4, 5, 6, 7, 8, \dots, 7 + H - 1\}$. $x_{t,1} = T_{db,t}, x_{t,2} = T_{fl,t}, x_{t,3} = RH_t, x_{t,4} = V_t, x_{t,5} = \theta_t, x_{t,6} = W_t, x_{t,7} = E_{t-1}, x_{t,8} = E_{t-2}, x_{t,6+H} = E_{t-H}$. The well-trained adaptive LSTM neural network, as described in Section 3 will be adopted in the data analytics layer to forecast building energy consumption. H is the time horizon, representing the number of hours used for the past energy data as the input to the LSTM neural network.

5. Case study on real-world buildings

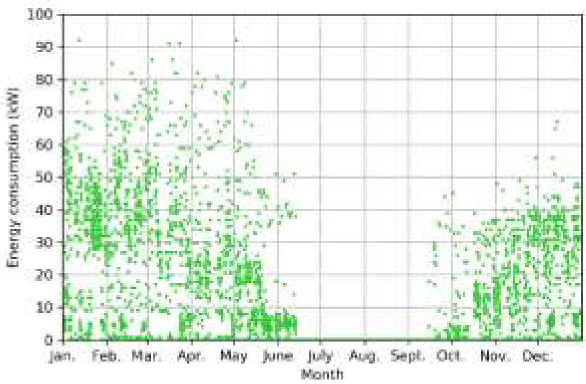
The proposed energy forecasting system is tested on two real-world campus buildings to evaluate its performance. The two buildings are Q block and S block of the Frenchay campus of University of the West of England, Bristol, respectively. These two blocks both contain classrooms for students and office rooms for university staff. The two different buildings at the same site are selected to further validate the effectiveness of the proposed adaptive LSTM neural network and energy consumption forecasting system. The building energy consumption profile $E(t)$ for the past two years (Jan 2018- Dec 2019) is collected from the energy management system at the time step of 1 h, as illustrated in Fig. 4. The electricity consumption is quite low during July and August for the S block and June to September for Q block. There also exist different features of energy consumption during other periods of the year.



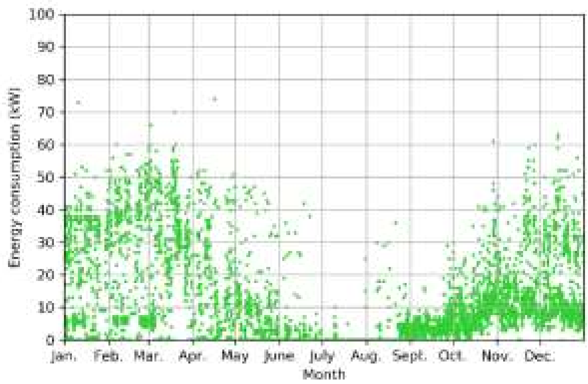
Q block, 2017



Q block, 2018



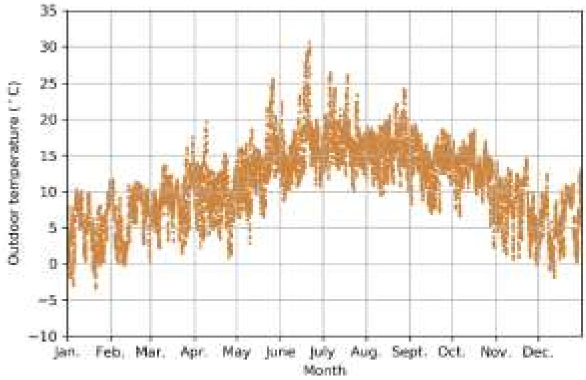
S block, 2017



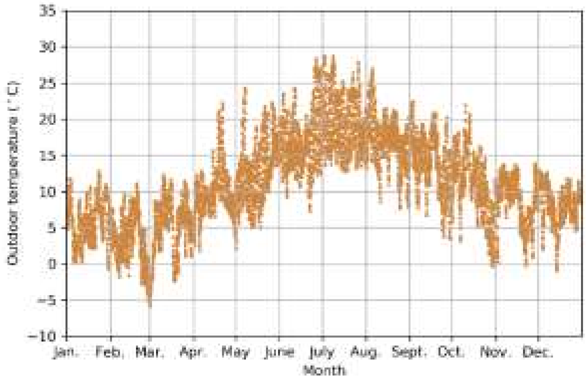
S block, 2018

Fig. 4. Energy consumption data

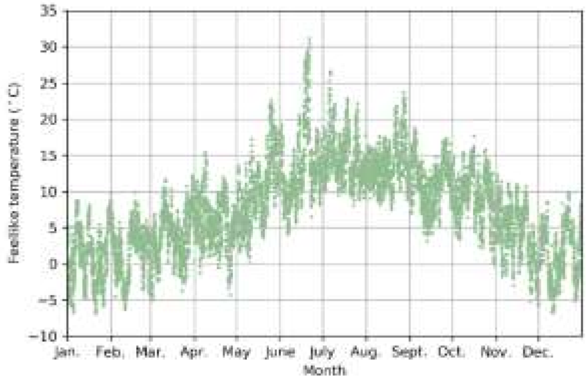
The weather data during the year 2017 and 2018 are collected from the local Bristol weather station near the campus building. The weather profile mainly includes outdoor air dry-bulb temperature, outdoor air feel-like temperature, relative humidity, wind speed, cloud cover and weather type are summarised in Fig. 5. The highest outdoor air dry-bulb temperature and the feel-like temperature is found in June for both 2017 and 2018, while the lowest temperatures are seen in January and March for 2017 and 2018, respectively. The lowest relative humidity is found in March and June for 2017 and 2018, respectively. Depending on the weather type, the range of cloud cover varies from 0 to 100% all over the year. The highest wind speed is about 35-40 m/s, respectively, found in September and January, respectively, in 2017 and 2018. During most of the time, the weather type is cloudy, while there are a few rainy and clear days. The various weather data, along with energy consumption data of 2017 and 2018, are adopted for training and testing purposes, respectively.



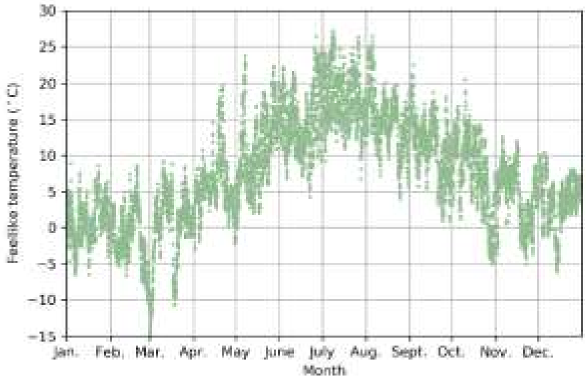
(a) $T_{db,i}$, 2017



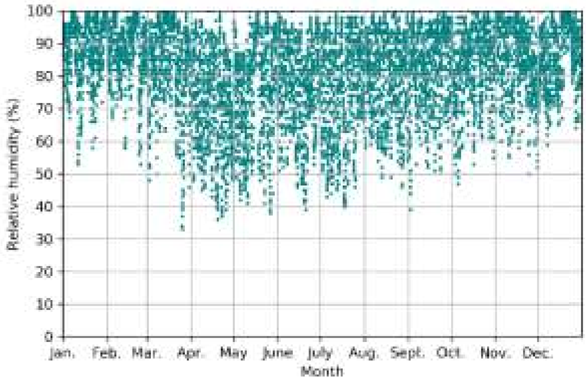
(b) $T_{db,i}$, 2018



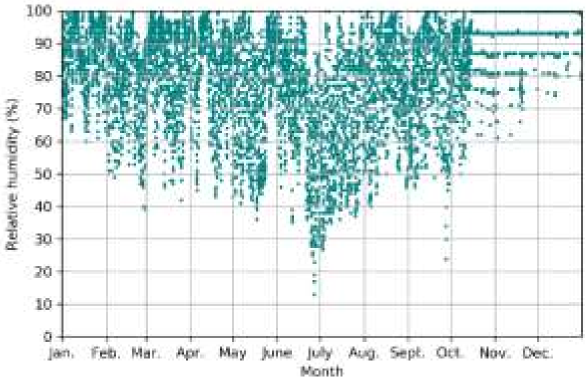
(c) $T_{fl,i}$, 2017



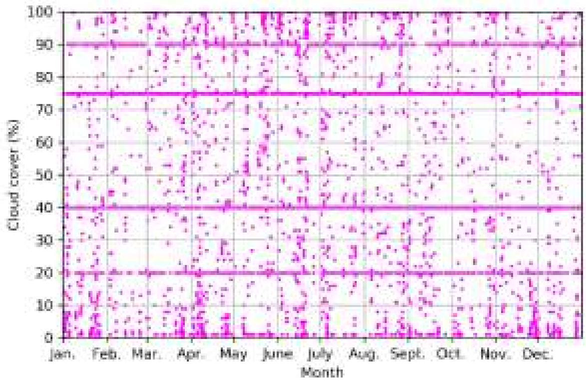
(d) $T_{fl,i}$, 2018



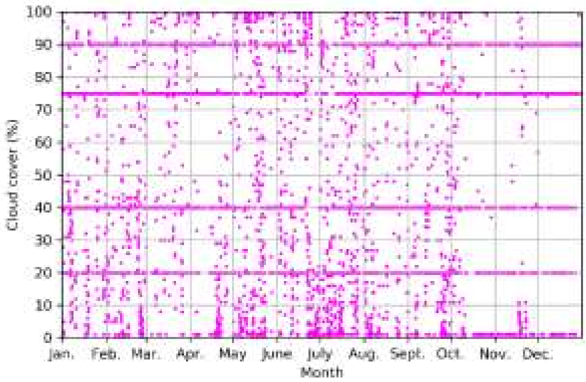
(e) RH_i, 2017



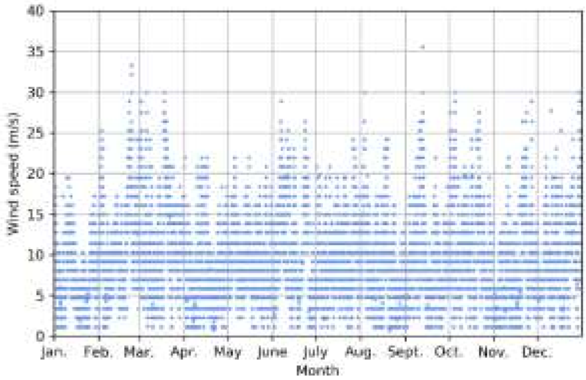
(f) RH_i, 2018



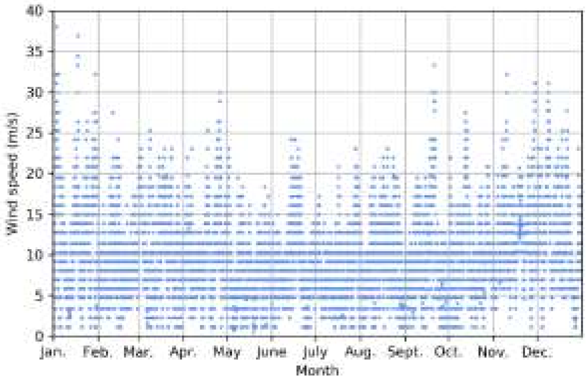
(g) θ_i, 2017



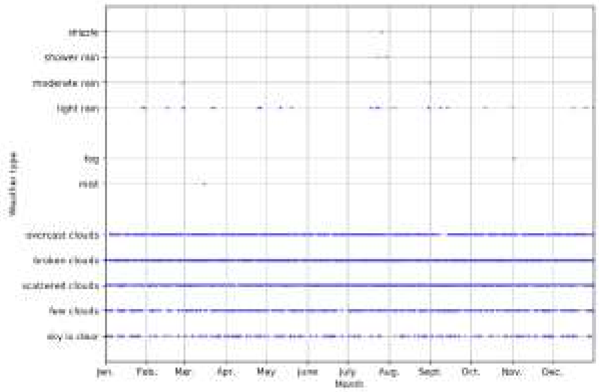
(h) θ_i, 2018



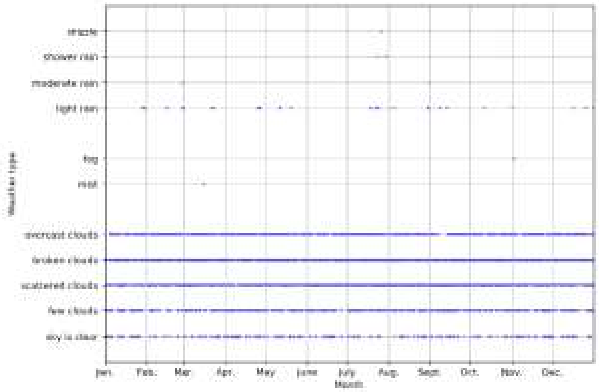
(i) V_i, 2017



(j) V_i, 2018



(k) W_i, 2017



(l) W_i, 2018

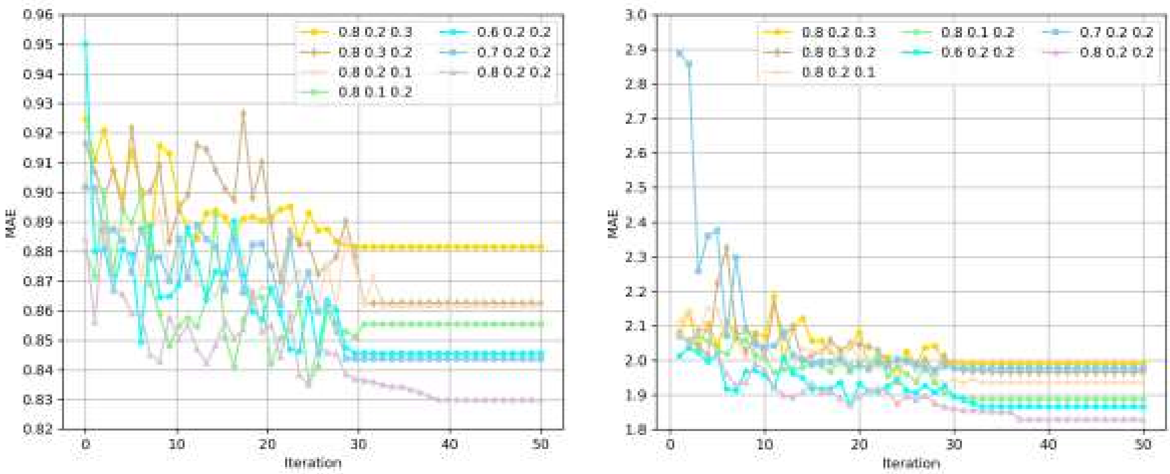
Fig. 5. Weather data

6. Results and analysis

The optimal parameters of GA are chosen based on the relevant prediction performance, including convergence performance and final MAE value. The effectiveness of weather profile as input dataset to the LSTM neural network and length of time horizon is also evaluated. The prediction performance from the selected optimal architecture of the LSTM neural network is evaluated with different reference parameters to verify the effectiveness of the adaptive LSTM neural network.

6.1 Performance evaluation of GA

The GA parameters include population size, maximum generation, selection probability, crossover probability and mutation probability. The population size represents the number of individuals at each iteration. The maximum generation represents the maximum iteration steps. The selection, crossover and mutation probability represents the likelihood of a selection, crossover and mutation operating occurring, respectively. Based on the complexity of the LSTM architecture optimisation problem, the population size is fixed at 20 while the maximum generation is set at 50. The convergence performance of GA under different selection, crossover and mutation probability is tested on two buildings, as shown in Fig. 6. There exists a large number of fluctuations of MAE at the beginning of optimisation; however, it becomes relatively steady after approximately 36 iterations. The smallest MAE of the LSTM neural network for the two buildings is identified with the same set of GA parameters. Therefore, selection, crossover and mutation probability is chosen as 0.8, 0.2 and 0.2, respectively.



Q Block S Block

Fig. 6. The convergence of *MAE* under different GA parameters.

The optimal architecture for the two LSTM neural networks is summarised in Table 5. The optimal number of layers and the learning rate is the same among the two LSTM neural network models. However, the number of neurons and the dropping rate in each LSTM layer is different among the two neural networks. It demonstrates that the architecture of the LSTM neural network needs to be optimised to make it adaptive and suitable for different features of building energy consumption.

Table 5. The optimal architecture of each LSTM neural network.

LSTM neural network architecture	Number of neurons in each LSTM layer		Number of layers	The dropping rate LSTM layer	Learning rate
	1	2		1	
Q Block	60	60	2	0.3	0.01
S Block	70	40	2	0.2	0.01

6.2 Comparison between LSTM and feedforward neural network

As summarised in Table 1, feedforward neural network models generally construct a direct mapping among comprehensive weather profile, time signature, indoor sensor measurement and energy consumption data. However, due to the lack of time correlation in the data sequence, feedforward neural network models cannot capture the inter-relationship between energy data and time. To investigate the capability of LSTM neural network in capturing the interrelationship between energy consumption data and time, the previously developed GA-optimised feedforward neural network [16] is adopted for comparison. The same sets of training and testing datasets are adopted for training and testing purposes, while the prediction performance is summarised in Table 6. It is found that the proposed adaptive LSTM neural network can achieve better accuracy and robustness than the reference GA-optimised feedforward neural network prediction model.

- For Q Block, if historical and forecasted weather profile is not available, there would be 23.53% and 21.20% increase in MAE and RMSE and a 3.39% decrease in R^2 for the training dataset. Meanwhile, there would be 21.74% and 20.36% increase in MAE and RMSE and 2.45% decrease in R^2 for the testing dataset.
- For S Block, if historical and forecasted weather profile is not available, there would be 22.58% and 20.99% increase in MAE and RMSE and 2.63% decrease in R^2 for the training dataset. Meanwhile, there would be 23.87% and 22.29% increase in MAE and RMSE, as well as 2.73% decrease in R^2 for the testing dataset.

It is seen that the prediction performance can be enhanced by selecting its hyper-parameters through GA optimisation. This also indicates that the proposed GA-determined LSTM prediction model can outperform most of those feedforward neural network prediction models presented in Table 1. This is due to the capability of the LSTM neural network in the time correlation of data sequence and its effectiveness in time series forecasting.

Table 6. Summary of prediction performance using adaptive LSTM and feedforward neural network.

Building and model		MAE (kW)		R^2 (%)		RMSE (kW)	
		Training	Testing	Training	Testing	Training	Testing
Q Block	GA-DNN	0.63	1.40	89.18	84.53	2.23	3.31
	Adaptive LSTM	0.51	1.15	92.31	86.65	1.84	2.75
	Change	↑23.53%	↑21.74%	↓3.39%	↓2.45%	↑21.20%	↑20.36%
S Block	GA-DNN	1.52	3.01	93.13	84.22	3.92	5.87
	Adaptive LSTM	1.24	2.43	95.65	86.58	3.24	4.80
	Change	↑22.58%	↑23.87%	↓2.63%	↓2.73%	↑20.99%	↑22.29%

6.3 Effects of weather data as input

To investigate the effects of weather condition as input training dataset, a reference model is adopted for the two buildings. In the reference model, the input dataset X only includes energy consumption at previous time steps. Namely, $x_{t,1} = E_{t-1}, x_{t,2} = E_{t-2}, \ldots, x_{t,H} = E_{t-H}$. The prediction performance with and without weather profile is summarised in Table 7.

- For Q Block, if historical and forecasted weather profile is not available, there would be 11.76% and 7.61% increase in MAE and RMSE, respectively, as well as 1.32% decrease in R^2 for the training dataset. Meanwhile, there would be 11.30% and 6.55% increase in MAE and RMSE, respectively, as well as 1.75% decrease for the testing dataset.
- For S Block, if historical and forecasted weather profile is not available, there would be 16.98% and 22.58% increase in MAE and RMSE, respectively, as well as 1.67% decrease in R^2 for the training dataset. Meanwhile, there would be 14.38% and 18.11% increase in MAE and RMSE, respectively, as well as 1.58% decrease for testing dataset.

Table 7. Summary of prediction performance with and without weather profile.

Building and model		MAE (kW)		R^2 (%)		RMSE (kW)	
		Training	Testing	Training	Testing	Training	Testing
Q Block	Without Weather	0.57	1.18	91.09	85.13	1.98	2.90
	With weather	0.51	1.15	92.31	86.65	1.84	2.75
	Change	↑11.76%	↑11.30%	↓1.32%	↓1.75%	↑7.61%	↑6.55%
S Block	Without weather	1.52	2.87	94.05	85.21	3.79	5.49
	With weather	1.24	2.43	95.65	86.58	3.24	4.80
	Change	↑22.58%	↑18.11%	↓1.67%	↓1.58%	↑16.98%	↑14.38%

6.4 Effects of time horizon on past information

The size of the input dataset to the LSTM forecasting model depends on the length of the time horizon. Short time horizon may result in inaccurate prediction while long time horizon could lead to large computational load. To identify the optimal length of time horizon, 4, 24, 48 and 72 hours are adopted for comparison. The MAE, R^2 and RMSE of different time horizon are summarised in Fig. 7. When H = 4h, the average value of MAE, RMSE and R^2 for training and testing datasets is 1.05 kW, 2.62kW and 86%, respectively. When H = 24h, the average value of MAE and RMSE for training and testing datasets decreases to 0.82 kW, 2.30 kW, respectively, while the corresponding R^2 increases to 89.5%. The MAE, RMSE and R^2 value are similar when H is further increased to 48h and 72h. Considering the computational load, the optimal time horizon is chosen as 24 hours.

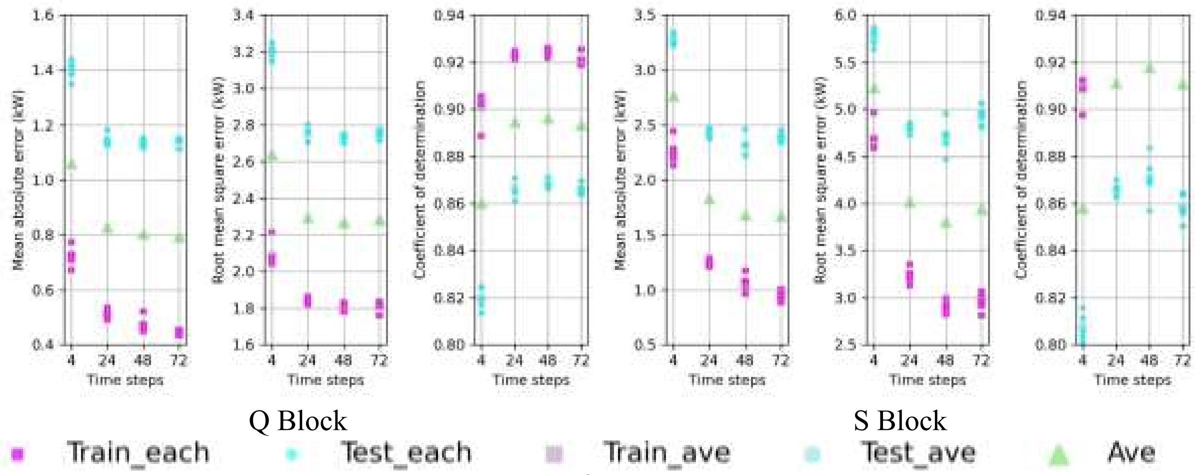


Fig. 7. MAE, RMSE and R^2 with different time horizons.

The energy consumption forecasting result at different time horizon is summarised in Fig. 8. Due to the lack of past information when the time horizon is 4 hours, there exists some discrepancy between the forecasted value and real measurement. However, there are also some fluctuations when the time horizon is 48 or 72 hours, owing to the possibility of over-convergence.

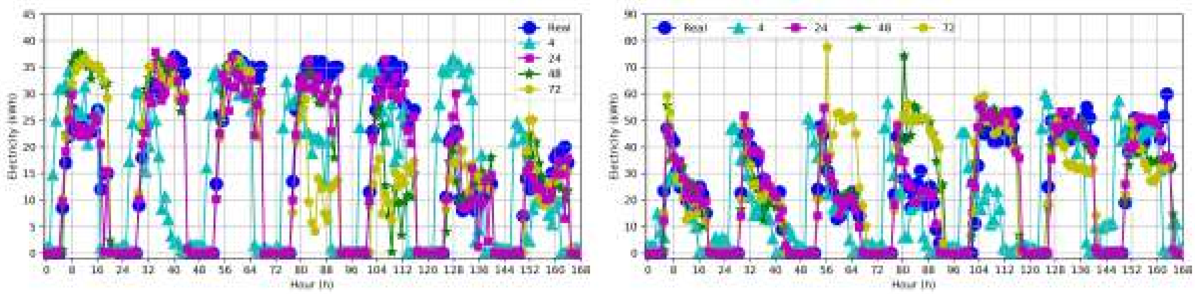


Fig. 8. Energy forecast using different time horizon.

6.5 *Effects of optimal LSTM architecture in prediction performance*

As illustrated in Table 2, the architecture of most of the previously developed LSTM models was generally based on experience, trial-and-error process or numerous experiments. The experienced-based LSTM hyper-parameters selection may result in low prediction accuracy, while the trial-and-error process and numerous experiments-based LSTM architecture selection required a large computational load. As the main LSTM hyper-parameters include the number of neurons in the LSTM layer, number of LSTM layers, dropping rate of LSTM layers and network learning rate, it would be difficult for the trial-and-error process and numerous experiments to find out the optimal hyper-parameter set. To demonstrate the effects of GA in determining the LSTM architecture, reference LSTM architectures are formulated by setting the value of each type of hyper-parameter different from the GA-determined optimal one. The MAE, RMSE and R^2 from these reference models are summarised in Table 8.

- For Q Block, the smallest average MAE, the smallest average RMSE and the largest average R^2 is identified when the number of neurons in both LSTM layers is 60. When the number of neurons in the first LSTM layer is 50 or 70, there would be a 0.18 kW or 0.20 kW increase in RMSE, 0.08 kW or 0.09 kW increase in MAE, and 0.9% or 1.1% decrease in R^2, respectively for the testing case. For Q Block, the smallest average MAE, the smallest average RMSE and the largest average R^2 is identified when the number of neurons in the first and section LSTM layer is 70 and 40, respectively. When the number of neurons in the first LSTM layer is 60 and 80, there would be a 0.25 kW and 0.17 kW increase in RMSE, 0.07 kW and 0.05 kW increase in MAE, as well as 1.1% and 1.0% decrease in R^2, respectively for the testing case.
- For both Q and S Blocks, the smallest average MAE, the smallest average RMSE and the largest average R^2 is identified when the number of LSTM layers is 2. When the number of LSTM layer is 1, there is 0.36 kW and 0.51 kW increase in MAE, 0.43 kW and 0.72 kW increase in RMSE, as well as 3.8% and 3.0% decrease in R^2 for Q Block and S Block, respectively. When the number of LSTM layer is 3, there is a 0.04 kW and 0.08 kW increase in MAE, 0.05 kW and 0.28 kW increase in RMSE, as well as 0.6% and 1% decrease in R^2 for Q Block and S Block, respectively. If there is only 1 LSTM layer, it is not sufficient to indicate the comprehensive relationship between historical weather profile and energy consumption data. On the other hand, it may cause over-convergence if there are 3 LSTM layers.
- For Q Block, the smallest average MAE, the smallest average RMSE and the largest average R^2 is identified when the dropping rate of LSTM layers is 0.3. When the dropping rate is larger or smaller than 0.3, there would be around 0.03-0.09 kW increase in MAE, 0.04-0.11 kW increase in RMSE, as well as 0.5%-1.2% decrease in R^2, respectively. For S Block, the smallest average MAE, the smallest average RMSE and the largest average R^2 is identified when the dropping rate of LSTM layers is 0.2. When the dropping rate is larger or smaller than 0.2, there would be around 0.10-0.26 kW increase in MAE, 0.25-0.37 kW increase in RMSE, as well as 0.9%-1.8% decrease in R^2, respectively.
- For both Q and S Block, the smallest average MAE, the smallest average RMSE and the largest average R^2 is identified when the learning rate is 0.01. For Q block, when the learning rate is larger or smaller than 0.01, there would be around 0.10-0.22 kW increase in MAE, 0.07-0.39 kW increase in RMSE, as well as 1.1%-4.2% decrease in R^2, respectively. For S Block, when the learning rate

is larger or smaller than 0.01, there would be around 0.20-0.92 kW increase in MAE, 0.35-1.12 kW increase in RMSE, as well as 1.9%-2.8% decrease in R^2, respectively.

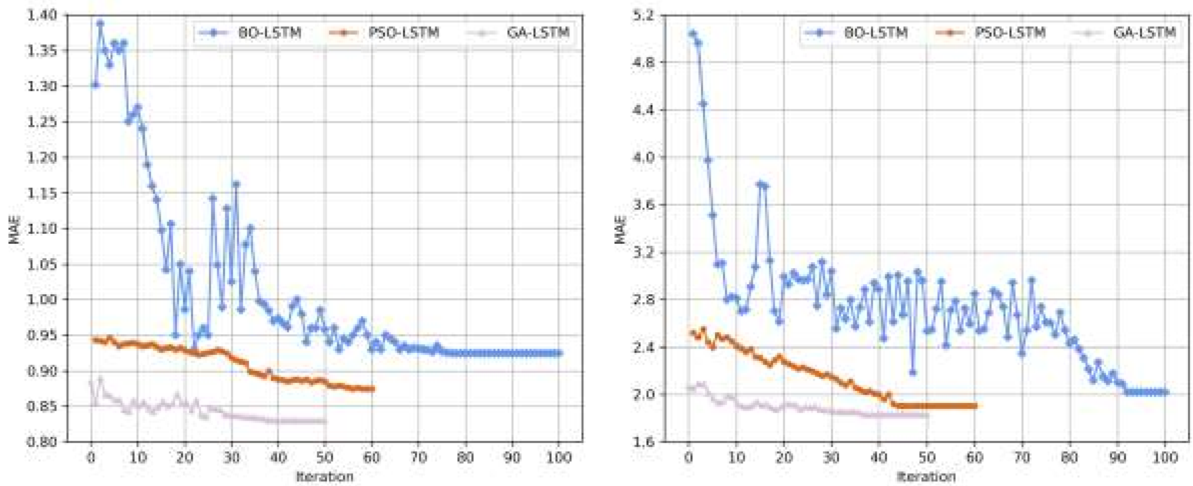
It is seen that the prediction performance can be enhanced by selecting its hyper-parameters through GA optimisation. This also indicates that the proposed GA-determined LSTM prediction model can outperform those LSTM models presented in Table 2. It is also seen that learning rate and dropping rate has the largest and smallest effects on prediction performance, while number of neurons and number of neurons in LSTM layers have moderate effects.

Table 8. Prediction performance at different LSTM hyper-parameters.

Building	Change of LSTM architecture	Number of LSTM layers	Number of neurons in each layer	Dropping rate	Learning rate	R^2 (%)		RMSE (kW)		MAE(kW)	
						Train	Test	Train	Test	Train	Test
Q block	Optimal	2	[60, 60]	[0.3, 0.5]	0.01	92.3	86.7	1.83	2.75	0.51	1.15
	Number of layers	3	[60,60,60]	[0.3, 0.5]	0.01	91.1	86.1	1.97	2.80	0.58	1.19
		1	60	0.3	0.01	90.7	82.9	2.03	3.11	0.72	1.58
	Dropping rate	2	[60, 60]	[0.1, 0.5]	0.01	91.9	85.9	1.93	2.82	0.55	1.20
				[0.2, 0.5]		91.9	86.1	1.89	2.80	0.54	1.19
				[0.4, 0.5]		92.0	86.2	1.91	2.79	0.56	1.18
				[0.5, 0.5]		91.4	85.5	1.95	2.86	0.59	1.24
	Learning rate	2	[60, 60]	[0.3, 0.5]	0.001	91.8	85.6	1.96	2.85	0.60	1.22
					0.1	85.1	82.5	2.56	2.97	0.94	1.54
					0.05	89.0	83.8	2.20	2.91	0.74	1.36
	Number of neurons	2	[50, 60]	[0.3, 0.5]	0.01	91.4	85.8	1.99	2.93	0.55	1.23
			[70, 60]			91.5	85.6	1.96	2.95	0.56	1.24
S block	Optimal	2	[70, 40]	[0.2, 0.2]	0.01	95.6	86.6	3.24	4.80	1.24	2.43
	Number of layers	3	[70,40,40]	[0.2, 0.2]	0.01	94.8	85.6	3.54	5.08	1.43	2.51
		1	70	0.2	0.01	94.0	83.6	3.46	5.31	1.62	3.15
	Dropping rate	2	[70, 40]	[0.1, 0.1]	0.01	94.6	85.7	3.36	5.12	1.30	2.53
				[0.3, 0.3]		95.2	85.4	3.38	5.05	1.39	2.54
				[0.4, 0.4]		95.0	85.1	3.49	5.08	1.45	2.57
				[0.5, 0.5]		94.5	84.8	3.64	5.17	1.55	2.69
	Learning rate	2	[70, 40]	[0.2, 0.2]	0.001	94.7	84.7	3.57	5.15	1.54	2.63
					0.1	89.4	83.8	5.04	5.92	2.53	3.35
					0.05	92.9	84.2	4.13	5.45	1.84	2.93
	Number of neurons	2	[60, 40]	[0.2, 0.2]	0.01	94.7	85.5	3.47	5.05	1.42	2.50
			[80, 40]			94.6	85.6	3.45	4.97	1.40	2.48

6.6 Performance comparison with other hyper-parameter selection approaches

To demonstrate the capability of GA in selecting optimal hyper-parameters of LSTM network, two reference prediction models are introduced for comparison. In other words, Bayesian optimisation and PSO is adopted to select the optimal hyper-parameters of the LSTM network. The search range of hyper-parameters is set the same as that in Table 3. The grid search approach can guarantee the global optimal hyper-parameter set. However, according to Section 2.2, if grid search is adopted for hyper-parameter selection, there would be 2,948,760 types of LSTM architecture trained, with each architecture trained and tested for 5 times. This costs too much computational time. Thus, it is not adopted as one of the reference models. The convergence of Bayesian optimisation and PSO is illustrated in Fig. 9, while the assessment of prediction performance is summarised in Table 9. The GA optimisation can reach convergent within 40 iterations, while it takes 60 and 90 iterations for PSO and Bayesian optimisation to get convergent. Although the same number of LSTM layers is identified by Bayesian optimisation and PSO as that from the proposed GA, the number of neurons in each LSTM layer, dropping rate of each layer and learning rate is different. As a result, the corresponding performance from Bayesian optimisation and PSO is also slightly worse than that from the proposed GA-LSTM prediction model. In other words, the LSTM network determined by Bayesian optimisation and PSO results in 0.42-2.50% decrease in R^2 value, 0.36-14.7% increase in RMSE and 2.47-31.4% increase in MAE. It is because that Bayesian optimisation is based upon the assumption that its optimisation function obeys Gaussian distribution, which might not be the truth for LSTM hyper-parameters optimisation. Moreover, PSO has better performance at continuous optimisation, while GA is better at discrete optimisation. According to Table 3, the searching range of the LSTM hyper-parameter is discrete. Therefore, the proposed GA-determined LSTM outperforms Bayesian optimisation and PSO.



(a) Q Block (b) S Block

Fig. 9. The convergence of MAE under Bayesian optimisation, PSO and GA.

Table 9. Prediction performance using different optimisation methods for hyper-parameter selection.

Building	Change of LSTM architecture	Number of LSTM layers	Number of neurons in each layer	Dropping rate	Learning rate	R^2 (%)		RMSE (kW)		MAE(kW)	
						Train	Test	Train	Test	Train	Test
Q block	Optimal	2	[60, 60]	[0.3, 0.5]	0.01	92.3	86.7	1.83	2.75	0.51	1.15
	BO	2	[40, 50]	[0.3, 0.7]	0.03	90.0	86.3	2.10	2.78	0.67	1.18
						↓2.50	↓0.46	↑14.7	↑1.09	↑31.4	↑2.61
	PSO	2	[70, 70]	[0.4, 0.5]	0.01	91.4	86.4	1.94	2.76	0.58	1.17
						↓0.98	↓0.35	↑6.01	↑0.36	↑13.7	↑1.74
S block	Optimal	2	[70, 40]	[0.2, 0.2]	0.01	95.6	86.6	3.24	4.80	1.24	2.43
	BO	2	[60, 40]	[0.1, 0.1]	0.001	94.9	85.1	3.52	4.92	1.53	2.51
						↓0.73	↓1.73	↑8.64	↑2.50	↑23.4	↑3.29
	PSO	2	[60, 30]	[0.2, 0.5]	0.01	95.2	85.8	3.32	4.87	1.32	2.49
						↓0.42	↓2.47	↑2.47	↑1.46	↑6.45	↑2.47

7. Implication of practical application in building energy performance prediction

Two years' historical outdoor weather profile and energy consumption data is collected from the local weather station and building energy management system to generate the database for the proposed energy consumption forecasting system. After the proposed energy forecasting model is well trained and tested using the historical database, it is expected to be adopted in the digital building management system for accurate energy consumption prediction using the latest weather forecast from weather reporting websites [74-76]. Such accurate energy consumption prediction plays a dominant role in various areas, including daily building energy management, decision making from facility managers, building information model designs, net-zero energy operation, climate change mitigation and circular economy.

First of all, accurate energy demand prediction plays a critical role in daily building energy management. The precise forecast of peak demand and hourly demand is important for efficient energy device scheduling and management, which can promote the building energy utilisation rate. An accurate building energy forecast can also assist building managers in making better decisions so as to reasonably control all types of energy devices [77].

Secondly, smart energy management and building energy efficiency retrofitting could be achieved on the basis of accurate energy consumption prediction. Building facility managers can gain an insight into future energy consumption, which allows them to calculate future energy costs and to determine whether to move to a more efficient pricing plan or modify future usage as appropriate [78].

Moreover, performance-based building requirements have become more predominant since it gives freedom in building design while maintain or even exceed the energy performance required by prescriptive-based requirements. However, in order to determine if building information model designs can reach targeted energy efficiency improvements, it is necessary to estimate the energy performance of a building using accurate building energy prediction models and different weather conditions.

Furthermore, accurate energy prediction also plays an important role in achieving net-zero energy operation of different types of buildings. With the transformation of the global energy industry, renewable energy is gradually replacing traditional fossil energy. With the accurate forecast of energy demand and various renewable energy generation, the desired output of other active energy devices can be determined. As such, the overall supply and demand of the building can be optimally coordinated in an effort to achieve its net-zero energy operation. The reduction of energy consumption also helps mitigate various climate-change related issues.

Last but not least, accurate electricity load forecasting at the power grid serves as a significant process in achieving a circular economy. Precise electricity forecast can help decrease energy consumption, reduce power generation costs, as well as improve social and economic benefits. With the development of power reform and the deepening of power marketisation, it is essential to improve power demand forecasting accuracy to enable stable and efficient operation of the power system. The strategy to reduce energy consumption can be regarded as a prominent factor that could influence economic growth since the energy demand from series of buildings remains at a high level [79].

8. Conclusion

In this study, an accurate and robust building energy forecasting system is proposed with the core of an adaptive LSTM neural network. The proposed energy forecasting approach has three distinct innovation over previous deep learning models in that:

- The LSTM neural network has recurring connections among neurons. Thus it can reveal the comprehensive time correlation in the data sequence.
- GA is good at discrete optimisation, and it's adopted to select the whole-set hyper-parameters of the LSTM network. The optimal hyper-parameters of the LSTM network include its number of neurons in each LSTM layer, the number of LSTM layers, dropping rate of each LSTM layer and learning rate of weighting matrices and bias.
- As the hyper-parameters of the LSTM network is selected according to the unique features of the energy dataset, it is adaptive to different types of energy data.

- The proposed building energy forecasting system is tested on two educational buildings using one-year historical weather and energy data for training the other year's profile for testing.

It is demonstrated that the proposed adaptive LSTM neural network is not only effective at reflecting the time correlation in data sequence but also powerful at investigating the complex relationship among various weather data and actual energy consumption. The major findings are listed as follows:

- The GA parameters, such as selection probability, crossover probability and mutation probability, are the same for the two testing buildings. It indicates that the GA algorithm is effective and robust in searching for the optimal LSTM neural network architecture.
- For both Q Block and S Block buildings, the optimal number of LSTM layers is 2, while the optimal learning rate is 0.01. However, the optimal number of neurons and dropping rate in each LSTM layer is different for Q Block and S Block buildings, indicating that the optimal LSTM neural network architecture would be different according to unique features of energy consumption data.
- The effect of available weather profile on energy consumption forecasting performance is assessed. When historical and forecasted weather profile is not available, there would be around 11.30-18.11% increase in MAE, 1.58-1.75% decrease in R^2, as well as 6.55-14.38% increase in RMSE for testing data.
- The effects of time horizon for past information is investigated while the optimal time horizon for past information is found to be 24 hours. When time horizon is 4 hours, the discrepancy between the forecasted value and real measurement is caused by insufficient past information. When time horizon is 48 or 72 hours, there exist some fluctuations in forecasted energy consumption owing to the over-convergence.
- The effectiveness of optimal architecture of the LSTM neural network is demonstrated by comparing its performance with different reference architectures. With the optimal architecture, Q Block results in 0.51 kW, 1.84 kW and 92.31% in MAE, RMSE and R^2 in the training dataset, respectively. It also results in 1.14 kW, 2.75 kW and 86.65% in MAE, RMSE and R^2 in the testing dataset. With the optimal architecture, S Block results in 1.24 kW, 3.24 kW and 95.65% in MAE, RMSE and R^2 in the training dataset, respectively; It also results in 2.43 kW, 4.80 kW and 86.58% in MAE, RMSE and R^2 in testing dataset.
- The proposed adaptive LSTM neural network outperforms the previously developed GA-optimised DNN model. Compared to the proposed adaptive LSTM neural network, the previously developed GA-optimised feedforward neural network prediction model has about 21.74%-23.87% increase in MAE, 2.45%-3.39% decrease in R^2, as well as 20.36%-22.29% increase in RMSE, respectively.
- GA is demonstrated to be the best optimisation approach in selecting LSTM hyper-parameters owing to its advantages of discrete optimisation. The GA-determined LSTM network has faster convergence with better optimisation results. The LSTM network determined by Bayesian

optimisation and PSO results in 0.42-2.50% decrease in R^2 value, 0.36-14.7% increase in RMSE and 2.47-31.4% increase in MAE.

9. New insights of the proposed computational method and future study

In this study, an adaptive LSTM neural network is proposed for an accurate and robust building energy forecasting system, which follows the long-standing research stream of using machine learning-based methods to predict the energy behaviour of buildings. Innovatively, the LSTM neural network is adaptive to different features of training data while GA is adopted to select the optimal hyper-parameters of the LSTM neural network, including its number of neurons in each LSTM layer, the number of LSTM layers, dropping rate of each LSTM layer and learning rate of weighting matrices and bias. Such innovation and adaptive LSTM neural network can also be adopted in other engineering fields for corresponding prediction tasks, such as construction, acoustic, workplace, exchange rates trading, predictive maintenance, human daily activities recognition and coal preparation.

This study can also provide new insights into how computational methods can be optimally and effectively adopted. In other words, the architecture of various machine learning models (i.e. artificial neural network, deep neural network, convolutional neural network and support vector machine) can be optimised by different optimisation approaches (i.e. Bayesian optimisation, PSO, GA and etc.) to enhance its accuracy and robustness. To better improve the performance of machine learning based prediction models, new evolutionary optimisation algorithms should also be proposed based on the unique characteristics of its application fields.

Acknowledgement

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.

Nomenclature

b	bias
f	forget gate
h	output of the same memory unit
i	input gate
o	output gate
R^2	coefficient of determination
RH	relative humidity

T Temperature

V wind speed

$\boldsymbol{W}$ weight matrix

x dataset

α coefficients of cubic interpolation

σ sigmoid activation

θ cloud cover

Tanh tangent activation

Superscripts

t time step

Subscripts

db dry-bulb

f forget gate

fl feel-like

i input gate

o output gate

Abbreviations

GA genetic algorithm

LSTM long-short term memory

MAE Mean absolute error

RMSE Root mean square error

SGD Standard gradient descent

References

[1] GlobalABC, IEA, UNE, 2019. Global status report for buildings and construction: towards a zero emissions, efficient and resilient buildings and construction sector.

[2] Sieminski A. International energy outlook. Energy Information Administration,(2017)5-30.

[3] 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.

[4] 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.

[5] G. Adams, P.G. Allen, B.J. Morzuch. Probability distributions of short-term electricity peak load forecasts. International Journal of Forecast, 7(1991)283-297
[6] Al-Musaylh MS, Deo RC, Adamowski JF and Li Y. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. Advanced Engineering Informatics, 35(2018)1-16.
[7] Saleh AI, Rabie AH and Abo-Al-Ez KM. A data mining based load forecasting strategy for smart electrical grids. Advanced Engineering Informatics, 30(2016)422-448.
[8] Singh MM, Singaravel S, Klein R and Geyer P. Quick energy prediction and comparison of options at the early design stage. Advanced Engineering Informatics, 46(2020)101185.
[9] YA LeCun, Y Bengio, GE Hinton. Deep learning. Nature, 521(2015)436-444.
[10]Bergstra J and Bengio Y. Random search for hyper-parameter optimization. Journal of machine learning research, 13(2012)281-305.
[11]Snoek J, Larochelle H and Adams RP. Practical bayesian optimization of machine learning algorithms. arXiv preprint:1206(2012)2944.
[12]Wieser E and Cheng G. EO-MTRNN: evolutionary optimization of hyperparameters for a neuro-inspired computational model of spatiotemporal learning. Biological cybernetics, 114(2020)363-387.
[13]Chang CJ, Lin JY and Chang MJ. Extended modeling procedure based on the projected sample for forecasting short-term electricity consumption. Advanced Engineering Informatics, 30(2016)211-217.
[14]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.
[15]Luo XJ. A novel clustering-enhanced adaptive artificial neural network model for predicting day-ahead building cooling demand. Journal of Building Engineering, 32(2020)101504.
[16]Luo XJ, Oyedele LO, Ajayi AO, Akinade OO, Delgado JMD, Owolabi HA and Ahmed A. Genetic algorithm-determined deep feedforward neural network architecture for predicting electricity consumption in real buildings. Energy and AI, 2(2020)100015.
[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]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.
[19]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 optimisation and management. International Journal of Refrigeration, 98(2019)399-409.
[20]Yang J, Hugues R and Radu Z. Online building energy prediction using adaptive artificial neural networks. Energy and Buildings, 37(2005)1250-1259.

[21]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.
[22]Moon J, Park S, Rho S and Hwang E. A comparative analysis of artificial neural network architectures for building energy consumption forecasting. International Journal of Distributed Sensor Networks, 15(2019)9.
[23]Kim MK, Kim YS and Srebric J. Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. Sustainable Cities and Society, 62(2020)102385.
[24]Lu R and Hong SH. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. Applied Energy, 236(2019)937-949.
[25]Li K, Hu C, Liu G and Xue W. Building's electricity consumption prediction using optimised artificial neural networks and principal component analysis. Energy and Buildings, 108(2015)106-113.
[26]Muralitharan K, Sakthivel R and Vishnuvarthan R. Neural network based optimisation approach for energy demand prediction in smart grid. Neurocomputing, 273(2018)199-208.
[27]Ruiz LGB, Rueda R, Cuéllar MP and Pegalajar MC. Energy consumption forecasting based on Elman neural networks with evolutive optimisation. Expert Systems with Applications, 92(2018)380-389.
[28]Li K, Xie X, Xue W, Dai X, Chen X and Yang X. A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. Energy and Buildings, 174(2018)323-334.
[29]Bünning F, Heer P, Smith RS and Lygeros, J. Improved day ahead heating demand forecasting by online correction methods. Energy and Buildings, 211(2020)109821.
[30]Yang J, Rivard H and Zmeureanu R. Online building energy prediction using adaptive artificial neural networks. Energy and buildings, 37(2005)1250-1259.
[31]Wang JQ, Du Y and Wang J. LSTM based long-term energy consumption prediction with periodicity. Energy, 197(2020)117197.
[32]Khafaf N, Jalili M and Sokolowski P. Application of deep learning long short-term memory in energy demand forecasting. In International Conference on Engineering Applications of Neural Networks. Springer, (2019)31-42.
[33]Wang XH, Zhao T, Liu H and He R. Power consumption predicting and anomaly detection based on long short-term memory neural network. IEEE 4th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), (2019)487-491.
[34]Khan ZA, Hussain T, Ullah A, Rho S, Lee M and Baik SW. Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework. Sensors, 20(2020) 1399.

[35]Wei N, Li C, Peng X, Li Y and Zeng F. Daily natural gas consumption forecasting via the application of a novel hybrid model. Applied Energy, 250(2019)358-368.

[36]Singaravel S, Suykens J and Geyer P. Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction. Advanced Engineering Informatics, 38(2018)81-90.

[37]Zhou C, Fang Z, Xu X, Zhang X, Ding Y and Jiang X. Using long short-term memory networks to predict energy consumption of air-conditioning systems. Sustainable Cities and Society, 55(2020)102000.

[38]Xue G, Qi C, Li H, Kong X and Song J. Heating load prediction based on attention long short term memory: A case study of Xingtai. Energy(2020)117846.

[39]Somu N, MR GR and Ramamritham K. A deep learning framework for building energy consumption forecast. Renewable and Sustainable Energy Reviews, 137(2021)110591.

[40]Wang SM, Wang SX, Chen HX and Gu Q. Multi-energy load forecasting for regional integrated energy systems considering temporal dynamic and coupling characteristics. Energy, 195(2020) 116964.

[41]Markovic R, Azar E, Annaqeeb MK and Frisch J. Day-Ahead Prediction of Plug-In Loads Using a Long Short-Term Memory Neural Network. Energy and Buildings, 234(1020)110667.

[42]Aowabin R, Vivek S and Amanda DS. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied Energy, 212(2018)372-385.

[43]Sendra-Arranz R and Gutiérrez A. A long short-term memory artificial neural network to predict daily HVAC consumption in buildings. Energy and Buildings, 216(2020)109952.

[44]Barzola-Monteses J, Espinoza-Andaluz M, Mite-León M and Flores-Morán M. Energy Consumption of a Building by using Long Short-Term Memory Network: A Forecasting Study. In 2020 39th International Conference of the Chilean Computer Science Society (SCCC), IEEE. (2020)1-6.

[45]Laib O, Khadir MT and Mihaylova L. Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks. Energy, 177(2019)530-542.

[46]Lu Y, Tian Z, Zhou R and Liu W. Multi-step-ahead prediction of thermal load in regional energy system using deep learning method. Energy and Buildings, 233(2021)110658.

[47]Jin XB, Zheng WZ, Kong JL, Wang XY, Bai YT, Su TL and Lin S. Deep-Learning Forecasting Method for Electric Power Load via Attention-Based Encoder-Decoder with Bayesian Optimization. Energies, 14(2021)1596.

[48]Yang T, Li B and Xun Q. LSTM-attention-embedding model-based day-ahead prediction of photovoltaic power output using Bayesian optimization. IEEE Access, 7(2019)171471-171484.

[49]Munem M, Bashar TR, Roni MH, Shahriar M, Shawkat TB and Rahaman H. Electric Power Load Forecasting Based on Multivariate LSTM Neural Network Using Bayesian Optimization. In 2020 IEEE Electric Power and Energy Conference (EPEC), (2020)1-6.

[50]He F, Zhou J, Feng ZK, Liu G and Yang Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. Applied Energy, 237(2019)103-116.
[51]Kim TY and Cho SB. Particle swarm optimization-based CNN-LSTM networks for forecasting energy consumption. In 2019 IEEE Congress on Evolutionary Computation (CEC) (2019)1510-1516.
[52]Yang Y, Shang Z, Chen Y and Chen YH. Multi-objective particle swarm optimization algorithm for multi-step electric load forecasting. Energies, 13(2020)532.
[53]Bouktif S, Fiaz A, Ouni A and Serhani MA. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. Energies, 11(2018)1636.
[54]Guo X, Zhao Q, Wang S, Shan D and Gong W. A Short-Term Load Forecasting Model of LSTM Neural Network considering Demand Response. Complexity, 2021.
[55]Su H, Zio E, Zhang J, Xu M, Li X and Zhang Z. A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. Energy, 178(2019)585-597.
[56]Pandya S and Ghayvat H. Ambient acoustic event assistive framework for identification, detection, and recognition of unknown acoustic events of a residence. Advanced Engineering Informatics, 47(2021)101238.
[57]Cai J, Zhang Y, Yang L, Cai H and Li S. A context-augmented deep learning approach for worker trajectory prediction on unstructured and dynamic construction sites. Advanced Engineering Informatics, 46(2020)101173.
[58]Zhao J and Obonyo E. Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units. Advanced Engineering Informatics, 46(2020)101177.
[59]Yang K, Ahn CR and Kim H. Deep learning-based classification of work-related physical load levels in construction. Advanced Engineering Informatics, 45(2020)101104.
[60]Zhong B, Xing X, Luo H, Zhou Q, Li H, Rose T and Fang W. Deep learning-based extraction of construction procedural constraints from construction regulations. Advanced Engineering Informatics, 43(2020)101003.
[61]Sun S, Wang S and Wei Y. A new ensemble deep learning approach for exchange rates forecasting and trading. Advanced Engineering Informatics, 46(2020)101160.
[62]Chen C, Liu Y, Wang S, Sun X, Di Cairano-Gilfedder C, Titmus S and Syntetos AA. Predictive maintenance using cox proportional hazard deep learning. Advanced Engineering Informatics, 44 (2020)101054.

[63]Lee H, Ahn CR and Choi N. Fine-grained occupant activity monitoring with Wi-Fi channel state information: Practical implementation of multiple receiver settings. Advanced Engineering Informatics, 46(2020)101147.

[64]Amer F and Golparvar-Fard M. Modeling dynamic construction work template from existing scheduling records via sequential machine learning. Advanced Engineering Informatics, 47 (2021)101198.

[65]Yin X, Niu Z, He Z, Li ZS and Lee DH. 2020. Ensemble deep learning based semi-supervised soft sensor modeling method and its application on quality prediction for coal preparation process. Advanced Engineering Informatics, 46(2020)101136.

[66]Rashid KM and Louis J. Times-series data augmentation and deep learning for construction equipment activity recognition. Advanced Engineering Informatics, 42(2019)100944.

[67]Yan W, Wang J, Cheng J, Wan Z, Xing K and Gao K. Long Short-Term Memory Networks and Bayesian Optimization for Predicting the Time-Weighted Average Pressure of Shield Supporting Cycles. Geofluids, 2021.

[68]Lv L, Kong W, Qi J and Zhang J. An improved long short-term memory neural network for stock forecast. In MATEC Web of Conferences, EDP Sciences, 232(2018)01024.

[69]Hochreiter S and Schmidhuber J. Long short-term memory. Neural computation, 9(1997)1735-1780.

[70]Goldberg D. Genetic Algorithms in Search, Optimisation, and Machine Learning Addison Wesley, Reading, Massachusetts, 1989.

[71]Mitchell M. An Introduction to Genetic Algorithms. MIT press, 1998.

[72]Best M and Neuhauser D. Walter A Shewhart, 1924, and the Hawthorne factory. BMJ quality & safety, 15(2006)142-143.

[73]Luo XJ, Oyedele LO, Ajayi AO, Akinade OO, Owolabi HA and Ahmed A. Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings. Renewable and Sustainable Energy Reviews, 131(2020)109980.

[74]Roman Cardell J. Python-based Deep-Learning methods for energy consumption forecasting. Bachelor's thesis, Universitat Politècnica de Catalunya, 2020.

[74]https://www.accuweather.com (last accessed 3 June 2021)

[75]https://weather.com/en-GB/ (last accessed 3 June 2021)

[76]https://www.metoffice.gov.uk/ (last accessed 3 June 2021)

[77]Li C, Ding Z, Zhao D, Yi J and Zhang G. Building energy consumption prediction: An extreme deep learning approach. Energies, 10(2107)1525.

[78]Nti IK, Teimeh M, Nyarko-Boateng O and Adekoya AF. Electricity load forecasting: a systematic review. Journal of Electrical Systems and Information Technology, 7(2020)1-19.

[79] Liu Z, Wu D, Liu Y, Han Z, Lun L, Gao J, Jin G and Cao, G. Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction. Energy Exploration and Exploitation, 37(2019)1426-1451.