IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 1, 2019, accepted October 16, 2019, date of publication October 23, 2019, date of current version November 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949065

A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households

KE YAN[1], (Member, IEEE), WEI LI[1], ZHIWEI JI[2], MENG QI[3], AND YANG DU[4], (Member, IEEE)

[1]Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, College of Information Engineering, China Jiliang University, Hangzhou 310018, China
[2]School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 311300, China
[3]School of Information Science and Engineering, Shandong Normal University, Jinan 250038, China
[4]College of Science and Engineering, James Cook University, Cairns, QLD 4870, Australia

Corresponding author: Zhiwei Ji (jzw18@hotmail.com)

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY19F020016, and in part by the National Natural Science Foundation of China under Grant 61602431, 61850410531 and 61902225.

ABSTRACT Irregular human behaviors and univariate datasets remain as two main obstacles of data-driven energy consumption predictions for individual households. In this study, a hybrid deep learning model is proposed combining an ensemble long short term memory (LSTM) neural network with the stationary wavelet transform (SWT) technique. The SWT alleviates the volatility and increases the data dimensions, which potentially help improve the LSTM forecasting accuracy. Moreover, the ensemble LSTM neural network further enhances the forecasting performance of the proposed method. Verification experiments were performed based on a real-world household energy consumption dataset collected by the 'UK-DALEąŕ project. The results show that, with a competitive training efficiency, the proposed method outperforms all compared state-of-art methods, including the persistent method, support vector regression (SVR), long short term memory (LSTM) neural network and convolutional neural network combining long short term memory (CNN-LSTM), with different step sizes at 5, 10, 20 and 30 minutes, using three error metrics.

INDEX TERMS Energy consumption, forecasting, long short term memory, wavelet transform.

I. INTRODUCTION

Data-driven energy consumption forecasting methods, which predicts the short-term future energy consumption values based on historical data, are important elements in the process of building information modeling (BIM) [1], [21], [36]. The short-term and very short-term energy consumption forecasting techniques are useful for residential energy demand-side management, electricity price market design, energy efficiency and maintenance scheduling of large-scale complex smart power grids [12], [28], [37]. Moreover, it provides extra reliability, security and protection for the smart grid to handle the increasing energy demand from residential households. Along with the fast development of artificial intelligence (AI) technology, the extended deep learning methods nowadays are capable of performing very short-term energy consumption forecasting results with spectacularly high prediction accuracy [7], [22], [29], [34].

The associate editor coordinating the review of this manuscript and approving it for publication was Lei Chen.

However, two difficulties remain as the major obstacles that hinder the existing data-driven forecasting strategies to be widely adopted in the smart grid development process. First, we doubt about the reliability of using pure data-driven methods on energy consumption prediction for residential households since the energy consumption patterns for individual households can be extremely irregular. Second, traditional deep learning neural networks, such as the convolutional neural networks (CNNs), require multi-dimensional inputs to achieve high prediction accuracy. Univariate time series data, e.g., the energy consumption data, forecasting is a challenging problem even for deep learning technologies.

In this study, we propose a hybrid deep learning forecasting framework that combines multiple LSTM neural networks with stationary wavelet transforms (SWT) to solve the irregular and univariate individual household energy consumption forecasting problem. The original energy consumption signal has been split into multiple more stationarized sub-signals; and each sub-signal is to be resolved by a corresponding LSTM neural network. The final forecasting result is a combination of multiple LSTM neural network forecasting results using inverse stationary wavelet transform. The SWT alleviates the volatility and increases the data dimensions, which potentially help improve the LSTM forecasting accuracy. Furthermore, the ensemble LSTM neural network enhances the forecasting performance of the proposed framework. Three main contributions of the current work to the literature are listed as follows:

1) **Hybrid deep learning framework combining wavelet transforms with LSTM neural networks.** A carefully designed hybrid deep learning forecasting framework combining wavelet transforms and LSTM neural networks is introduced. The original energy consumption data collected from individual households are stationarized by stationary wavelet transform. The number of wavelet transform is determined by Pearson correlation coefficient. In addition, we compared different types of base wavelets to select the most appropriate stationary wavelet transform to convert the original energy consumption data.
2) **Ensemble LSTM neural network structure integrating multiple LSTM neural networks.** An ensemble LSTM neural network structure is constructed integrating four LSTM neural networks to enhance the forecasting accuracy. Each LSTM neural network corresponds to one transformed signal in different levels. The stationarized signals potentially enhance the forecasting capability of each LSTM neural network, and consequently improve the overall forecasting accuracy.
3) **Remarkable margins comparing the proposed forecasting framework with the state-of-art methods.** Different error metrics, namely, root mean squared error(RMSE), mean absolute percent error (MAPE) and mean bias error (MBE), are employed to show the performance comparisons between the proposed algorithm and the existing methods in the literature. According to the experimental results, our method outperforms all compared methods with remarkable margins in terms of the three error metrics.

II. RELATED WORKS

The forecasting methodologies can be generally categorized into long-term (more than a year), mid-term (from a month to a year), short-term (from one day to a month) and very short-term (within 24 hours) forecasting methods. The fast development of artificial intelligence (AI) technology provides an important short-term forecasting solution for time series data compared to traditional methods, such as physical simulation models [4], [31], statistical analysis [2], [24] and regression models [17], [23]. Modern AI techniques, such as support vector machine (SVM), extreme learning machine (ELM), artificial neural networks (ANNs) and deep learning neural networks are employed to perform short-term forecasting tasks in various areas. Grigorievskiy *et al.* [11] introduced an optimally pruned extreme learning machine (OP-ELM) technique to perform long-term time series prediction. The original ELM was improved by internal pruning of inessential neurons. Hu *et al.* [14] compared modern machine learning techniques, including SVM, ELM and back-propagation neural network (BPNN), on the tunnel settlement forecasting problem. Comprehensive experimental results tested on real-world datasets were provided to show the effectiveness of AI technology applications for real-world industrial problems. Singh and Dwivedi [35] proposed a novel evolutionary algorithm based ANN framework to solve the short-term load forecasting problem. Benali *et al.* [3] compared smart persistence, ANN and random forest (RF) to forecasting solar irradiance components, including normal beam, horizontal diffuse and global components. The experimental results show that even with highly irregular solar irradiance data, such as solar irradiation in spring and autumn, the machine learning based forecasting model still has an obvious advantage compared to traditional forecasting models. Bouktif *et al.* [5] extended the traditional LSTM neural network by combining the genetic algorithm (GA) to forecast aggregate load in short to medium terms.

Modern civilization requires more precise and real-time forecasting techniques and has moved the short-term forecasting to very short-term forecasting, which perfectly matches the scope of machine learning and deep learning technologies [25]. Yan *et al.* [40] designed a hybrid deep learning methodology that integrates convolutional neural network (CNN) and long short term memory (LSTM) neural network to forecast power consumption values in every five minutes. The forecasting results were further extended to a time interval of half an hour through a multi-step forecasting strategy that was proposed by the same study. Kim *et al.* [20] proposed a hybrid power demand forecasting model, combining (c, l)-LSTM and CNN, for very short-term forecasting. The input sequence consists of multiple [Key, Context] pairs, where the key value is the power demand value; and context values include contextual information, such as temperature, humidity and season. Wang *et al.* [38] performed a thirty-minute energy consumption forecasting for a profile of small and medium enterprises. The pinball loss was used as a guideline for parameter tuning of the LSTM neural network, instead of the traditional mean square error (MSE). Imani and Ghassemian [15] combined discrete wavelet transform (DWT) and collaborative representation transform (CRT) to perform very short-term load forecasting at meter-level. Deng *et al.* [8] designed a multiscale convolutional neural network (MS-CNN) with time cognition to forecast load in steps of multiple hours. The proposed model was further strengthened by periodical coding. Kong *et al.* [21], [22] designed various LSTM neural network structures for different household loads, considering the resident behaviors. The forecasting accuracy using a single LSTM neural network was not satisfied enough. Ospina *et al.* [32] predicted power output for a PV plant in 30 minutes time interval using ensemble LSTM neural networks. Liu *et al.* [26] predicted wind power changes in very

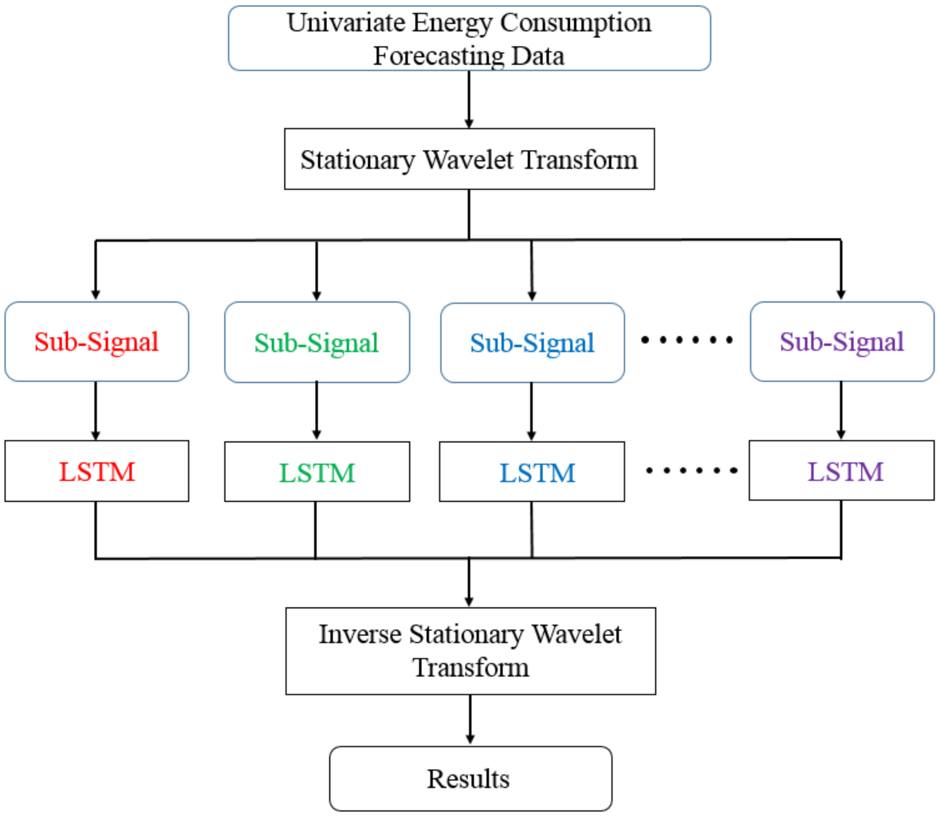


FIGURE 1. **The overall flowchart of the proposed ensemble deep learning framework for energy consumption forecasting targeting at individual households.**

short term (15 minutes) based on LSTM and discrete wavelet transform (DWT).

III. METHODOLOGY

A hybrid ensemble deep learning neural network framework is proposed to perform energy consumption forecasting for individual households. The original univariate energy consumption data or signal is split into several sub-signals using stationary wavelet transforms. For each stationarized sub-signal, exactly one LSTM neural network is employed to perform the forecasting operation. We believe that the stationarization of the wavelet transform potentially improves the LSTM forecasting results. Last, the forecasting results are integrated using inverse stationary wavelet transform. The overall flowchart is illustrated in Figure 1. A real-world dataset containing energy consumption data of five households located in London, UK is utilized for verification purposes.

A. DATA DESCRIPTION

The open-source energy consumption data that is used to verify the correctness and robustness of the proposed method is collected through remote sensors installed in five different family houses in London, United Kingdom (UK) with the project titled 'UK-DALE' [18]. In the original dataset, the data is collected in 6 seconds interval and only the energy consumption data in kilowatt-hour is available. We combine multiple entries of the original data and convert it to a dataset with time interval at 5 minutes. For each household, we collect 12,000 data samples. Approximately 66.7% of the data samples are used in to train the proposed hybrid ensemble deep learning framework. The remaining 33.3% of the collected data samples are used for testing purposes.

B. LONG SHORT TERM MEMORY NEURAL NETWORK

Long short term memory (LSTM) neural network is a special form of the recurrent neural network (RNN), which replaces the original low-cell neurons with cells consisting of more complex internal structures (Figure 2) [13]. First, it inherits the special property of RNN, which considers inputs as an inter-connected time series. Second, the complex internal structure of LSTM cell resolves the exploding and vanishing gradients problems [16]. There are four important elements in the flowchart of LSTM model: cell status, input gate, forget gate and output gate (Figure 2). The input, forget and output gates are used to control the update, maintenance and deletion of information contained in cell status. The forward computation process can be denoted as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{1}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{2}$$
$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{3}$$

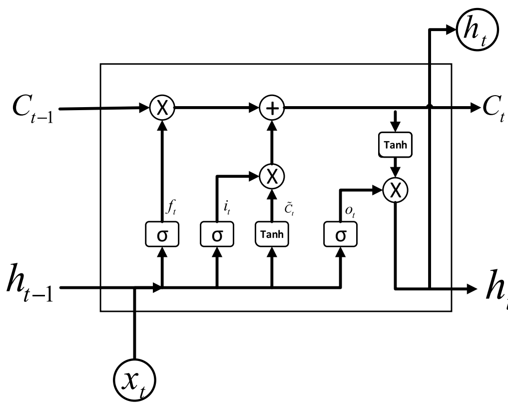


FIGURE 2. The internal structure of the proposed LSTM cell.

$$
\begin{aligned}
C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, && (4)\\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), && (5)\\
h_t &= tanh(C_t), && (6)
\end{aligned}
$$

where C_t, C_{t-1} and $\tilde{C}_t$ represent current cell status value, last time frame cell status value and the update for the current cell status value, respectively. The notations f_t, i_t and o_t represent forget gate, input gate and output gate, respectively. With proper parameter settings, the output value h_t is calculated based on $\tilde{C}_t$ and C_{t-1} values according to Eqs. (4) and (6). All weights, including: W_f, W_i, W_C and W_o, are updated based on the difference between the output value and the actual value following back-propagation through time (BPTT) algorithm [39].

C. STATIONARY WAVELET TRANSFORM

Wavelet transform (WT) is capable to split the original energy consumption data into more stationary sub-signals. The two most commonly used classes of WT are Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) [19].

The stationary wavelet transform (SWT), or discrete stationary wavelet transform, is an extension of the traditional DWT, which improves the traditional DWT using up-sampling instead of down-sampling in each signal subdivision process to prevent information loss [30]. More specifically, at each time of transformation, the SWT splits the original signal into two sub-signals through a high and low pass filter, respectively, with up-sampling (Figure 3). The original signal can be reconstructed using a combination of two sub-signals with inverse pass filters without loss of information.

Multi-step SWT split the original signal into n sub-signals through a multi-step transformation through $n - 1$ SWT processes (Figure 4). At each step of SWT, the original signal $F(t)$ is transformed to a more stationarized sub-signal A_n. It is evident that through multi-step SWT, the transformed sub-signal tends to be more stationarized [41]. The original signal $f(x)$ can be reconstructed through inverse SWT processes.

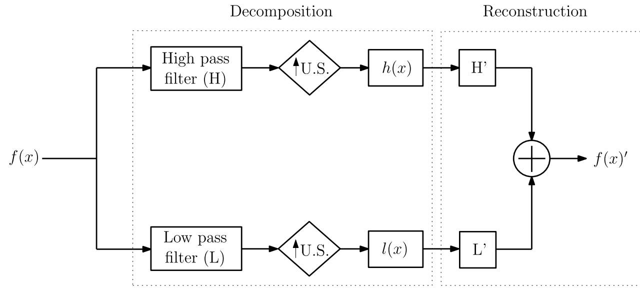


FIGURE 3. Standard stationary wavelet transform. The original signal $f(t)$ is split into two sub-signals through a high and low pass filter, respectively, with up-sampling (US).

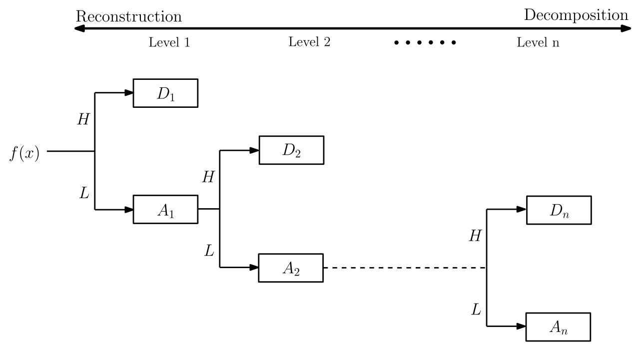


FIGURE 4. Multi-step stationary wavelet transform process. The original signal $f(t)$ is split into $n + 1$ sub-signals, $D_1, D_2, \ldots, D_n$ and A_n through n SWTs.

D. A HYBRID ENSEMBLE LSTM NEURAL NETWORK

The proposed hybrid ensemble LSTM neural network is depicted in Figure 1. As a necessary pre-processing step, the process of determining several important parameters is described in this subsection, including the selection of wavelet basis function and the number of SWT in the entire framework. These parameters are essential to guarantee the functionality of the proposed framework and the effectiveness of predicting energy consumption patterns for individual households.

1) SELECTING WAVELET BASIS FUNCTION

The entire UK-DALE dataset is divided into training and testing parts. For each household, the energy consumption data of a whole year in 2015 was collected. The first nine months of energy consumption data were treated as the training dataset. And the remaining three months' data was used in the testing phase.

There are a number of different wavelet basis functions available in the literature. The selection of a suitable wavelet basis function is crucial in the pre-processing stage. In this section, we compared most of the state-of-art wavelet basis functions, including Daubechies (db), Haar (haar), Symlets (sym), Coiflets (coif), Reverse biorthogonal (rbio), Biorthogonal (bior) and Discrete approximation of Meyer (dmey) wavelets, on the training dataset of the UK-DALE dataset. There is only one SWT performed. The mean absolute errors (MAE) [6] between the reconstructed signals $f(x)'$ and the

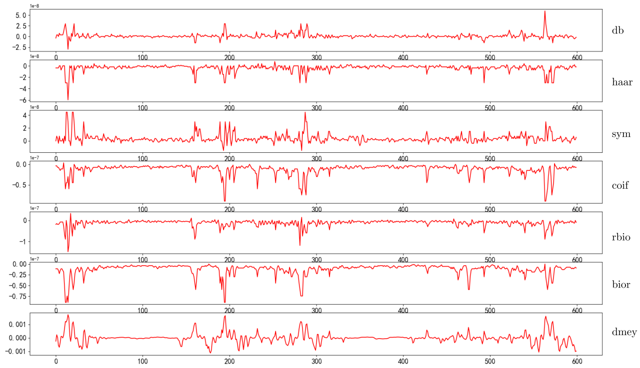


FIGURE 5. **The reconstruction errors using different wavelet basis functions using the UK-DALE dataset.**

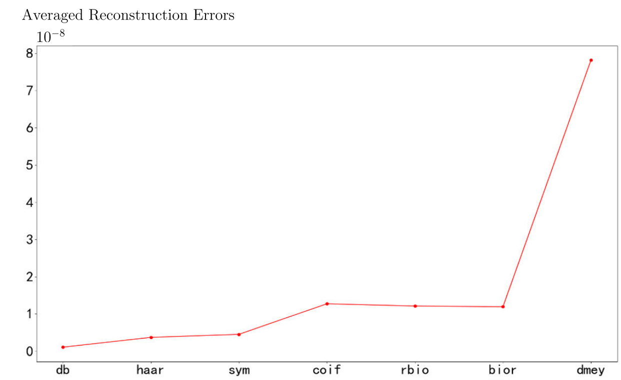


FIGURE 6. **Averaged reconstruction errors of all tested wavelet basis functions.**

original signals $f(x)$ are shown as the reconstruction errors in Figure 5. According to the averaged reconstruction errors (Figure 6), Daubechies (db) wavelet reaches the minimum, which is selected as the wavelet basis function in the proposed hybrid ensemble LSTM framework.

2) DETERMINING THE NUMBER OF SWTs

The second parameter that is crucial in the pre-processing phase is the number of SWTs that are carried out on the original energy consumption pattern. We employed the Pearson correlation coefficient (PCC) ($P_{f(x),f_d}$) to help determine the most appropriate number of SWTs according to Equation 7, where $f(x)$ is the original energy consumption pattern; $f_d = \Sigma_{j=1}^{N} D_j$ is the combination of all high pass sub-signals; and $cov(f(x), f_d)$ denotes the covariance between $f(x)$ and f_d. The PCC indicates the consistency before and after the SWTs.

$$P_{f(x),f_d} = \frac{cov(f(x), f_d)}{\sigma_x \cdot \sigma_y}, \tag{7}$$

$f(x) = f_d + A_N$, where A_N is the only low pass sub-signal, according to the multi-step SWT process shown in Figure 4.

The variance of PPC levels between f_d and $f(x)$, corresponding to the number of SWTs is shown in Figure 7. It is evident that after three times of SWTs, the f_d has the highest PPC value with the original signal. As a result, we adopt the three times SWTs in the comparative experiments in Section IV.

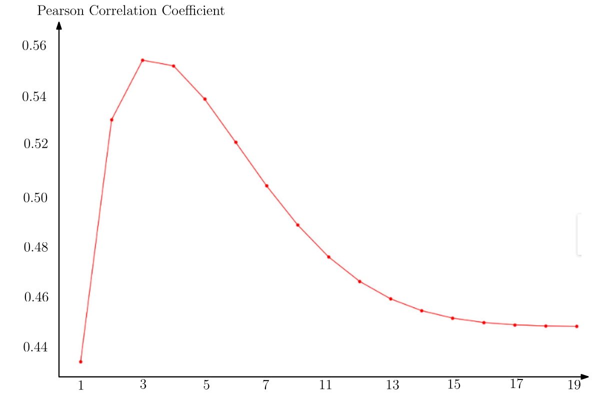


FIGURE 7. **Pearson correlation coefficient level versus the number of SWTs. The figure clearly show that three time SWTs are most appropriate for the pre-processing phase.**

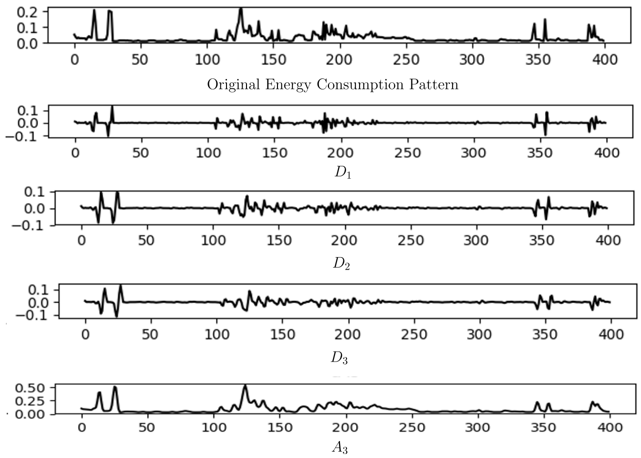


FIGURE 8. **With three times SWTs, the original energy consumption pattern is split into four sub-signals: D_1, D_2, D_3 and A_3.**

IV. RESULTS

The simulation results were performed with a personal computer with the configuration and environment: Intel Core(TM) i7-7700HQ CPU @ 2.8 GHZ, NVIDIA, GeForce GTX1050Ti graphics card, Python 3.6.5 (64-bit), Tensorflow with version 1.8.0 and Keras 2.6.1. There are in total five households energy consumption datasets collected by the UK-DALE project for the whole year of 2015. For each dataset, nine months of the energy consumption data were used as the training set. The remaining three months' data formed the testing set. The decomposition results of applying SWT three times on the household 1 training dataset is shown in Figure 8. With the three times SWTs, the original energy consumption pattern is decomposed into four sub-signals: D_1, D_2, D_3 and A_3.

The energy consumption forecasting results of the proposed hybrid ensemble deep learning framework are compared with four state-of-art methods, including the persistent method [9], support vector regression (SVR) [42], single

TABLE 1. The hyperparameters for all models that have been tested in the experiments, including the numbers of hidden layers, the numbers of neurons in each layer, the optimizer and the batch sizes.

Method	# layers	# neurons (each layer)	Optimizer	Batch size
LSTM	3	100	RMSprop	56
CNN-LSTM	2(CNN) + 1(LSTM)	100	RMSprop	72
SWT-LSTM	2	100	RMSprop	64

TABLE 2. A comparative study performed using the proposed method (Prop.) with four state-of-art methods, including the persistent model (Pers.), support vector regression (SVR), single LSTM neural network (LSTM) and the hybrid CNN-LSTM forecasting framework (CNN-LSTM) on 5 minutes step size data.

Data	RMSE					MAPE (%)					MBE				
	Pers.	SVR	LSTM	CNN-LSTM	Prop.	Pers.	SVR	LSTM	CNN-LSTM	Prop.	Pers.	SVR	LSTM	CNN-LSTM	Prop.
Hse 1	0.0227	0.0220	0.0217	0.0304	0.0104	17.563	15.046	18.279	18.126	8.1000	0.0084	0.0078	0.0085	0.0081	0.0040
Hse 2	0.0266	0.0229	0.0216	0.0016	0.0091	18.045	13.962	16.637	3.7647	7.0077	0.0089	0.0073	0.0077	0.0075	0.0035
Hse 3	0.0199	0.0202	0.0196	0.0117	0.0101	10.890	12.929	12.900	8.5512	7.3358	0.0055	0.0060	0.0064	0.0062	0.0032
Hse 4	0.0168	0.0149	0.0158	0.0064	0.0066	15.337	17.761	17.392	15.325	7.7487	0.0048	0.0051	0.0051	0.0053	0.0022
Hse 5	0.0247	0.0250	0.0237	0.0060	0.0099	7.8858	10.026	8.2243	5.0226	4.2021	0.0071	0.0083	0.0076	0.0076	0.0034
Aver.	0.0221	0.0210	0.0205	0.0112	**0.0092**	13.944	13.945	14.686	10.158	**6.8789**	0.0070	0.0069	0.0071	0.0069	**0.0033**

TABLE 3. Results of the comparative study performed on 10 minutes step-size data.

Data	RMSE					MAPE (%)					MBE				
	Pers.	SVR	LSTM	CNN-LSTM	Prop.	Pers.	SVR	LSTM	CNN-LSTM	Prop.	Pers.	SVR	LSTM	CNN-LSTM	Prop.
Hse 1	0.0476	0.0411	0.0398	0.0403	0.0173	21.157	17.693	22.016	20.278	8.7058	0.0182	0.0155	0.0168	0.0165	0.0072
Hse 2	0.0677	0.0533	0.0520	0.0506	0.0212	35.013	23.868	30.142	24.576	12.543	0.0262	0.0187	0.0203	0.0186	0.0088
Hse 3	0.0770	0.0678	0.0709	0.0721	0.0313	24.930	22.902	24.575	24.190	11.599	0.0327	0.0300	0.0333	0.0342	0.0145
Hse 4	0.0275	0.0241	0.0259	0.0250	0.0122	25.279	22.715	26.845	24.147	10.403	0.0131	0.0119	0.0135	0.0126	0.0059
Hse 5	0.0456	0.0434	0.0402	0.0413	0.0168	10.677	13.649	10.708	9.5062	4.7297	0.0152	0.0175	0.0152	0.0148	0.0065
Aver.	0.0531	0.0459	0.0458	0.0459	**0.0198**	23.411	20.165	22.857	20.539	**9.5963**	0.0211	0.0187	0.0198	0.0194	**0.0086**

TABLE 4. Results of the comparative study performed on 20 minutes step-size data.

Data	RMSE					MAPE (%)					MBE				
	Pers.	SVR	LSTM	CNN-LSTM	Prop.	Pers.	SVR	LSTM	CNN-LSTM	Prop.	Pers.	SVR	LSTM	CNN-LSTM	Prop.
Hse 1	0.0453	0.0379	0.0408	0.0393	0.0154	20.307	16.896	21.148	19.483	8.6134	0.0181	0.0146	0.0178	0.0166	0.0071
Hse 2	0.0878	0.0739	0.0737	0.0732	0.0292	36.581	24.637	33.744	33.468	14.661	0.0421	0.0331	0.0379	0.0385	0.0164
Hse 3	0.1371	0.1336	0.1299	0.1278	0.0619	29.868	28.641	28.565	30.854	16.054	0.0604	0.0612	0.0601	0.0613	0.0304
Hse 4	0.0751	0.0699	0.0668	0.0669	0.0304	30.662	26.097	28.914	29.769	12.666	0.0418	0.0384	0.0388	0.0389	0.0179
Hse 5	0.1217	0.1027	0.1034	0.1011	0.0400	16.957	15.076	16.782	14.878	6.4157	0.0436	0.0387	0.0410	0.0382	0.0160
Aver.	0.0934	0.0836	0.0829	0.0817	**0.0354**	26.875	22.270	25.830	25.691	**11.682**	0.0412	0.0372	0.0391	0.0387	**0.0175**

TABLE 5. Results of the comparative study performed on 30 minutes step-size data.

Data	RMSE					MAPE (%)					MBE				
	Pers.	SVR	LSTM	CNN-LSTM	Prop.	Pers.	SVR	LSTM	CNN-LSTM	Prop.	Pers.	SVR	LSTM	CNN-LSTM	Prop.
Hse 1	0.0968	0.0772	0.0782	0.0774	0.0347	25.377	18.724	24.513	22.109	11.205	0.0481	0.0366	0.0415	0.0398	0.0186
Hse 2	0.1291	0.1057	0.1027	0.1021	0.0441	29.396	23.430	31.464	22.952	12.035	0.0582	0.0453	0.0502	0.0441	0.0210
Hse 3	0.2284	0.2243	0.2151	0.2148	0.1065	36.541	32.953	35.272	35.454	17.758	0.1087	0.1112	0.1058	0.1074	0.0547
Hse 4	0.1089	0.0985	0.0981	0.0950	0.0420	33.695	25.821	32.778	29.066	15.190	0.0636	0.0538	0.0584	0.0551	0.0255
Hse 5	0.1600	0.1596	0.1427	0.1500	0.0610	18.224	16.194	17.215	16.508	7.9928	0.0836	0.0796	0.0764	0.0784	0.0341
Aver.	0.1446	0.1331	0.1274	0.1279	**0.0577**	28.647	23.424	28.248	25.218	**12.836**	0.0724	0.0653	0.0665	0.0650	**0.0308**

LSTM neural network [10] and the hybrid CNN-LSTM forecasting framework [40]. The hyperparameters of LSTM, CNN-SLTM and the proposed SWT-LSTM models are optimized and listed in Table 1. The C and γ values for the

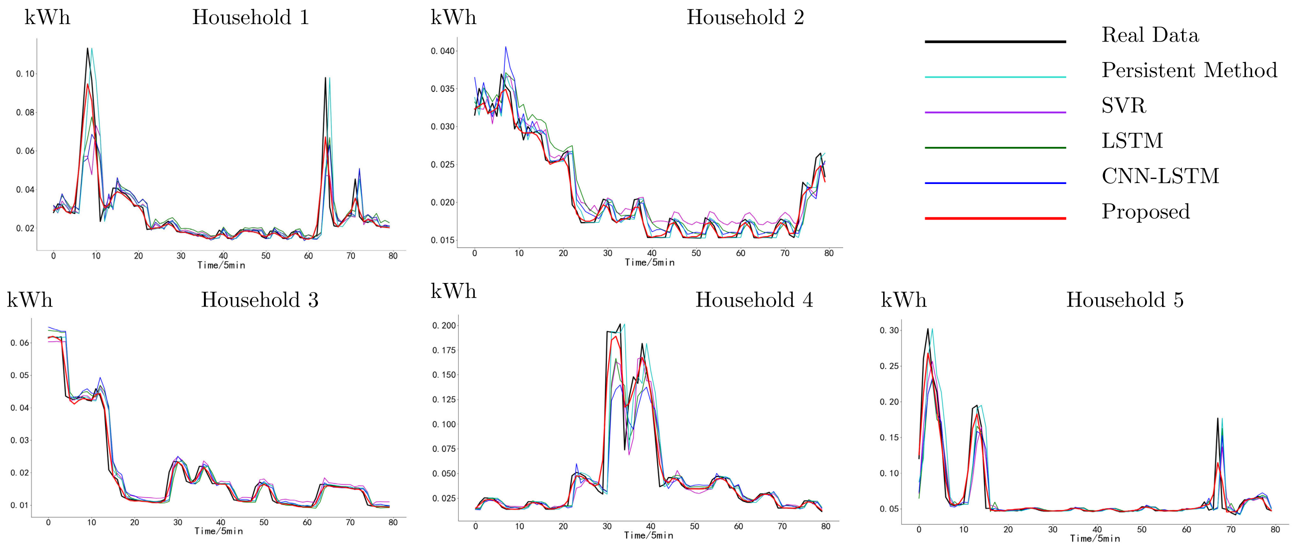


FIGURE 9. Forecasting results for five households energy consumption data collected by the UK-DALE project with time step length at 5 minutes.

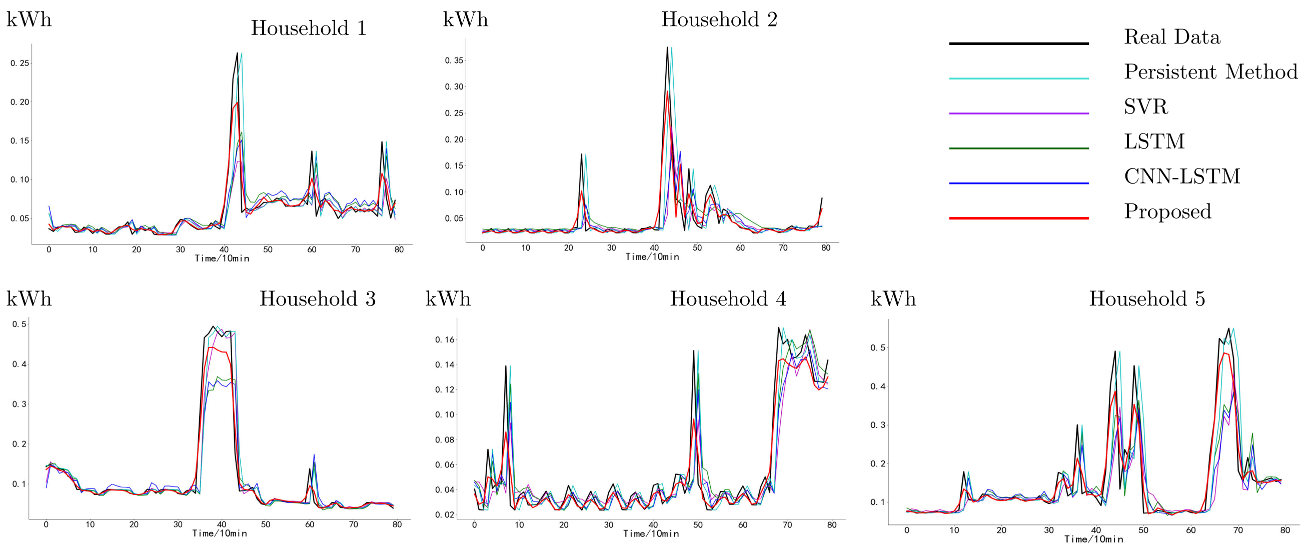


FIGURE 10. Forecasting results for five households energy consumption data collected by the UK-DALE project with time step length at 10 minutes.

SVR model are automatically optimized using grid search algorithm [27]. The forecasting results comparison is performed using three well-known error metrics that are root-mean-square error (RMSE), mean absolute percentage error (MAPE) [33] and mean bias error (MBE) with different step lengths. The detailed equations of the three error metrics are formulated in Equations 8, 9 and 10:

$$\mathrm{RMSE} = \sqrt{\frac{\Sigma_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}, \tag{8}$$

$$\mathrm{MAPE} = \frac{\Sigma_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100}{N}, \tag{9}$$

$$\mathrm{MBE} = \frac{\Sigma_{i=1}^{N}\left|\hat{y}_i - y_i\right|}{N}, \tag{10}$$

where y_i is an actual testing sample value; $\hat{y}_i$ is the prediction result of y_i; and N is the total number of testing samples. The forecasting results with step lengths of 5 minutes, 10 minutes; 20 minutes and 30 minutes are shown in Tables from 2 to 5, respectively. In addition, we illustrate the forecasting performance on the five households energy consumption data (partially) in Figures 9 - 12.

In average, the RMSE, MAPE and MBE values of the proposed method are lower than those of the other methods (Tables 2-5). In Table 2, while the step size is at 5 minutes, for the households 2, 4 and 5, the proposed method has slightly higher error rates compared to CNN-LSTM. According to the data description provided by the UK-DALE project, the data collected from households 2, 4 and 5 are less volatile. The energy consumption curves for households 1 and 3 are relatively more active. The results in Table 2 suggests that the proposed method is more suitable for active energy consumption data forecasting compared with existing approaches. Moreover, for step sizes at 10, 20 and 30 minutes

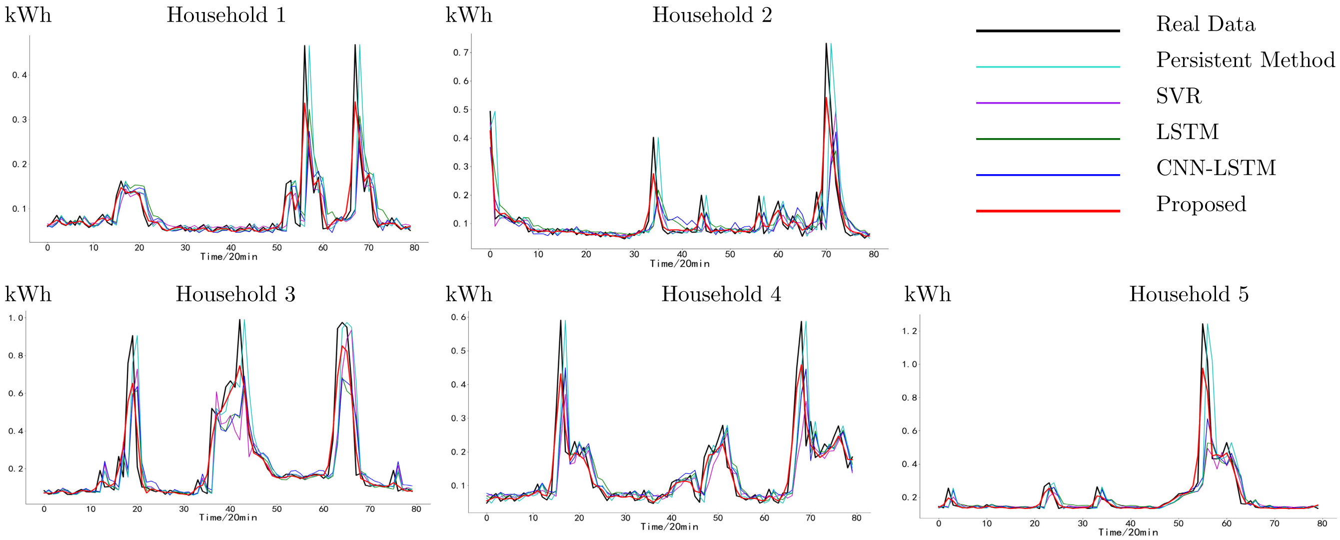


FIGURE 11. Forecasting results for five households energy consumption data collected by the UK-DALE project with time step length at 20 minutes.

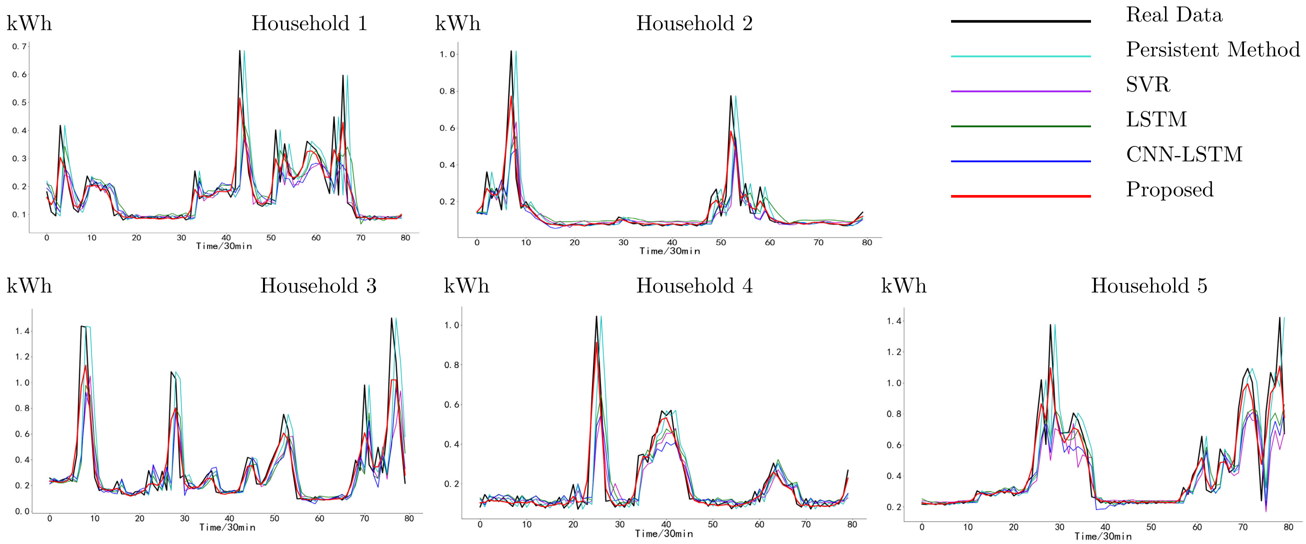


FIGURE 12. Forecasting results for five households energy consumption data collected by the UK-DALE project with time step length at 30 minutes.

TABLE 6. The averaged training time for all tested methods in Section IV.

Method	SVR	LSTM	CNN-LSTM	SWT-LSTM
Training time (sec.)	2.1330	50.6150	26.1312	40.4768

(Tables 3-5), the proposed method has more accurate forecasting results compared to the persistent method, SVR, LSTM and CNN-LSTM for all households datasets, which justifies that the performance of the proposed method becomes more stable while the step size increases.

V. CONCLUSION

This study proposed a hybrid ensemble deep learning framework aiming at the energy consumption forecasting problem of individual households. First, the original energy consumption data is stationarized using multiple stationary wavelet transform (SWT), which results in a decomposition of the original signal into multiple sub-signals. Second, each sub-signal is treated as an independent time series data. A long short term memory (LSTM) neural network is attached to each sub-signal and produces forecasting results. The SWT pre-processing phase is considered potentially improving the forecasting accuracy of LSTM neural networks. Last, the forecasting results are integrated using inverse SWT.

We carefully select the two parameters of the SWT pre-processing phase, namely, the wavelet basis function and the number of SWTs performed. The forecasting accuracy is compared with state-of-art methods, including SVR, LSTM and CNN-LSTM. The averaged training time for the four compared methods are listed in Table 6. Based on the results listed in Tables II-VI, SVR requires the shortest training time but performs the worst in term of forecasting accuracy. The proposed SWT-LSTM framework produces the most accurate forecasting results with reasonable training time on various time horizons, including 5, 10, 20 and 30 minutes. In the future, the SWT-LSTM will be tested with more real-world energy consumption data other than the UK-DALE dataset.

Another future work of this study is to employ the clustering technique to further improve the forecasting accuracy and apply the implemented hybrid ensemble deep learning framework to other application fields, such as the solar energy generation forecasting problem.

CONFLICT OF INTERESTS

All authors declare that there is no conflict of interest regarding the publication of this manuscript.

REFERENCES

[1] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1192–1205, Jan. 2018.

[2] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 498–505, Aug. 2001.

[3] L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene, "Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components," *Renew. Energy*, vol. 132, pp. 871–884, Mar. 2019.

[4] K. J. Beven, M. J. Kirkby, N. Schofield, and A. F. Tagg, "Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments," *J. Hydrol.*, vol. 69, nos. 1–4, pp. 119–143, 1984.

[5] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies* vol. 11, no. 7, p. 1636, Jun. 2018.

[6] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?" *Geosci. Model Develop. Discuss.*, vol. 7, pp. 1525–1534, Feb. 2014.

[7] H. Choi, S. Ryu, and H. Kim, "Short-term load forecasting based on ResNet and LSTM," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids*, Oct. 2018, pp. 1–6.

[8] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting," *IEEE Access*, vol. 7, pp. 88058–88071, 2019.

[9] J. R. Driscoll, N. Sarnak, D. D. Sleator, and R. E. Tarjan, "Making data structures persistent," *J. Comput. Syst. Sci.*, vol. 38, no. 1, pp. 86–124, 1989.

[10] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[11] A. Grigorievskiy, Y. Miche, A.-M. Ventelä, E. Séverin, and A. Lendasse, "Long-term time series prediction using OP-ELM," *Neural Netw.*, vol. 51, pp. 50–56, Mar. 2014.

[12] P. Guo, J. C. Lam, and V. O. Li, "Drivers of domestic electricity users' price responsiveness: A novel machine learning approach," *Appl. Energy*, vol. 235, pp. 900–913, Feb. 2019.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] M. Hu, W. Li, K. Yan, Z. Ji, and H. Hu, "Modern machine learning techniques for univariate tunnel settlement forecasting: A comparative study," *Math. Problems Eng.*, vol. 2019, Apr. 2019, Art. no. 7057612.

[15] M. Imani and H. Ghassemian, "Residential load forecasting using wavelet and collaborative representation transforms," *Appl. Energy*, vol. 253, Nov. 2019, Art. no. 113505.

[16] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2342–2350.

[17] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renew. Energy*, vol. 34, no. 5, pp. 1388–1393, 2009.

[18] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Sci. Data*, vol. 2, Mar. 2015, Art. no. 150007.

[19] C. H. Kim and R. Aggarwal, "Wavelet transforms in power systems. Part 1: General introduction to the wavelet transforms," *Power Eng. J.*, vol. 14, no. 2, pp. 81–87, 2000.

[20] M. Kim, W. Choi, Y. Jeon, and L. Liu, "A hybrid neural network model for power demand forecasting," *Energies*, vol. 12, no. 2, p. 931, 2019.

[21] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1087–1088, Jan. 2018.

[22] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Sep. 2017.

[23] Y.-S. Lee and L.-I. Tong, "Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming," *Knowl.-Based Syst.*, vol. 24, no. 1, pp. 66–72, 2011.

[24] C. Lim and M. McAleer, "Time series forecasts of international travel demand for australia," *Tourism Manage.*, vol. 23, no. 4, pp. 389–396, Aug. 2002.

[25] K. Liu, S. Subbarayan, R. R. Shoults, M. T. Manry, C. Kwan, F. I. Lewis, and J. Naccarino, "Comparison of very short-term load forecasting techniques," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 877–882, May 1996.

[26] Y. Liu, L. Guan, C. Hou, H. Han, Z. Liu, Y. Sun, and M. Zheng, "Wind power short-term prediction based on LSTM and discrete wavelet transform," *Appl. Sci.*, vol. 9, no. 2, p. 1108, 2019.

[27] R. G. Mantovani, A. L. D. Rossi, J. Vanschoren, B. Bischl, and A. C. P. L. F. de Carvalho, "Effectiveness of random search in SVM hyper-parameter tuning," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2015, pp. 1–8.

[28] N. Mazzi, J. Kazempour, and P. Pinson, "Price-taker offering strategy in electricity pay-as-bid markets," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2175–2183, Mar. 2018.

[29] N. Mohan, K. Soman, and S. S. Kumar, "A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model," *Appl. Energy*, vol. 232, pp. 229–244, Dec. 2018.

[30] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*. New York, NY, USA: Springer, 1995, pp. 281–299.

[31] A. H. Neto and F. A. S. Fiorelli, "Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption," *Energy Buildings*, vol. 40, no. 12, pp. 2169–2176, 2008.

[32] J. Ospina, A. Newaz, and M. O. Faruque, "Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model," *IET Renew. Power Gener.*, vol. 13, no. 7, pp. 1087–1095, May 2019.

[33] C. Prakash, A. Sujil, R. Kumar, and N. Mittal, "Linear prediction model for joint movement of lower extremity," in *Recent Findings in Intelligent Computing Techniques*. New York, NY, USA: Springer, 2019, pp. 235–243.

[34] W. Shen, V. Babushkin, Z. Aung, and W. L. Woon, "An ensemble model for day-ahead electricity demand time series forecasting," in *Proc. 4th Int. Conf. Future Energy Syst.*, May 2013, pp. 51–62.

[35] P. Singh and P. Dwivedi, "Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem," *Appl. Energy*, vol. 217, pp. 537–549, May 2018.

[36] L. Wang, R. Kubichek, and X. Zhou, "Adaptive learning based data-driven models for predicting hourly building energy use," *Energy Buildings*, vol. 159, pp. 454–461, Jan. 2018.

[37] W. Wang, H. Chen, B. Lou, N. Jin, X. Lou, and K. Yan, "Data-driven intelligent maintenance planning of smart meter reparations for large-scale smart electric power grid," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov.*, Oct. 2018, pp. 1929–1935.

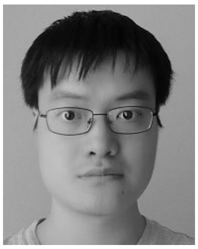
[38] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang, "Probabilistic individual load forecasting using pinball loss guided LSTM," *Appl. Energy*, vol. 235, pp. 10–20, Feb. 2019.

[39] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[40] K. Yan, X. Wang, Y. Du, N. Jin, H. Huang, and H. Zhou, "Multi-step short-term power consumption forecasting with a hybrid deep learning strategy," *Energies*, vol. 11, no. 11, p. 3089, Nov. 2018.

[41] P. Yu, A. Goldenberg, and Z. Bi, "Time series forecasting using wavelets with predictor-corrector boundary treatment," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0098300412003470

[42] F. Zhang, C. Deb, S. E. Lee, J. Yang, and K. W. Shah, "Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique," *Energy Buildings*, vol. 126, pp. 94–103, Aug. 2016.

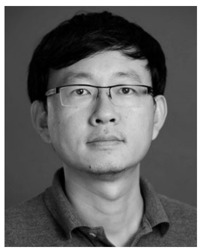


KE YAN received the bachelor's degree from the National University of Singapore (NUS), and the Ph.D. degree in computer science from the National University of Singapore (NUS), in 2012, under the supervision of Dr. H.-L. Cheng. From 2013 to 2014, he was a Postdoctoral Researcher with the Masdar Institute of Science and Technology, Abu Dhabi, UAE. He currently serves as an Associate Professor with China Jiliang University, Hangzhou, China. His main research interests include computer graphics, computational geometry, data mining, and machine learning.

WEI LI is currently pursuing the master's degree with the Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, College of Information Engineering, China Jiliang University. His main research interests include machine learning technology on industrial applications.

ZHIWEI JI received the bachelor's degree from Zhejiang A&F University (ZAFU), Linan, China, in 2003, the master's degree from Shanghai University, Shanghai, China, in 2009, and the Ph.D. degree from the School of Electronics and Information Engineering, Tongji University, Shanghai, in 2016. Since 2003, he has been with the School of Information Engineering, ZAFU. From 2016 and 2017, he was a Postdoctoral Research Fellow with Wake Forest University, USA. He currently serves as an Assistant Professor with Zhejiang Gongshang University, Hangzhou, China. His research interests mainly include pattern recognition, machine learning, computational biology, and bioinformatics.

MENG QI received the Ph.D. degree from the National University of Singapore, in 2014. She is currently a Lecturer with Shandong Normal University, China. Her research interests include interactive computer graphics, computational geometry, and computer graphics.



YANG DU received the Ph.D. degree in electrical engineering from The University of Sydney, Sydney, NSW, Australia, in 2013. From 2013 to 2014, he was with the Masdar Institute of Science and Technology, Abu Dhabi, UAE, as a Postdoctoral Research Fellow. From 2014 to 2018, he was a Lecturer with Xi'an Jiaotong-Liverpool University, Suzhou, China. He is currently with the College of Science and Engineering, James Cook University, Cairns, Australia, as a Lecturer. His research interests include photovoltaic power systems, power electronics, and smart grid.