Energy 182 (2019) 72–81

Contents lists available at ScienceDirect

Energy

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

# Predicting residential energy consumption using CNN-LSTM neural networks

Tae-Young Kim, Sung-Bae Cho*

*Department of Computer Science, Yonsei University, Seoul, South Korea*

A R T I C L E I N F O

*Article history:*
Received 28 December 2018
Received in revised form
29 May 2019
Accepted 31 May 2019
Available online 4 June 2019

*Keywords:*
Electric energy consumption
Deep learning
Convolutional neural network
Long short-term memory

A B S T R A C T

The rapid increase in human population and development in technology have sharply raised power consumption in today's world. Since electricity is consumed simultaneously as it is generated at the power plant, it is important to accurately predict the energy consumption in advance for stable power supply. In this paper, we propose a CNN-LSTM neural network that can extract spatial and temporal features to effectively predict the housing energy consumption. Experiments have shown that the CNN-LSTM neural network, which combines convolutional neural network (CNN) and long short-term memory (LSTM), can extract complex features of energy consumption. The CNN layer can extract the features between several variables affecting energy consumption, and the LSTM layer is appropriate for modeling temporal information of irregular trends in time series components. The proposed CNN-LSTM method achieves almost perfect prediction performance for electric energy consumption that was previously difficult to predict. Also, it records the smallest value of root mean square error compared to the conventional forecasting methods for the dataset on individual household power consumption. The empirical analysis of the variables confirms what affects to forecast the power consumption most.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Recent rapid economic growth and population growth are accelerating the increase in electric energy consumption [1]. According to the World Energy Outlook 2017, published by the International Energy Agency (IEA), the world's primary energy demand is projected to grow at a CAGR of 1.0% over the 2016–2040 period [2]. The residential sector also represents 27% of global energy consumption, and has a significant impact on overall energy consumption [3]. Electric energy must be consumed at the same time as it is generated in the power plant due to its physical characteristics [4]. Therefore, accurate power demand forecasting is required for stable power supply.

Electric energy consumption forecasting is a multivariate time series prediction problem [5]. Data collected from sensors is subject to uncertainty [6]. Using a specific window of several sensor signals, differentiated features can be extracted to forecast the power consumption by using the prediction model [7]. However, it is very difficult to predict electric energy consumption using classical prediction methods because power consumption has a regular seasonal pattern, while it contains irregular trend components [8]. Observed electric energy consumption can be decomposed into three components: trend factor, season factor, and random factor [9]. Fig. 1 represents three components of power consumption time series decomposition. Electric energy consumption represents a variety of patterns, and it is helpful to categorize some of the behaviors seen in the time series. Time series decomposition provides useful visualizations to better understand the problem of analyzing and predicting energy consumption in general. We refer to three types of time series patterns, and focus on the irregular trend in the energy consumption time series decomposition through Fig. 1.

The UCI machine learning repository provides electricity consumption dataset consisting of a total of 2,075,259 time-series and 12 variables (https://archive.ics.uci.edu/ml/datasets/). The dataset collected power consumption for four years (from December 16, 2006 to November 26, 2010) in a home in France [10].

Fig. 2 shows the impulse response function graph, which shows the influence of the variables making up the electrical energy consumption data. In the VAR model, the impulse response function can be used to determine how each variable affects other variables based on the correlation between several time series collected for power demand forecasting [11]. It is interesting to

* Corresponding author.
*E-mail addresses:* taeyoungkim@yonsei.ac.kr (T.-Y. Kim), sbcho@yonsei.ac.kr (S.-B. Cho).

https://doi.org/10.1016/j.energy.2019.05.230
0360-5442/© 2019 Elsevier Ltd. All rights reserved.

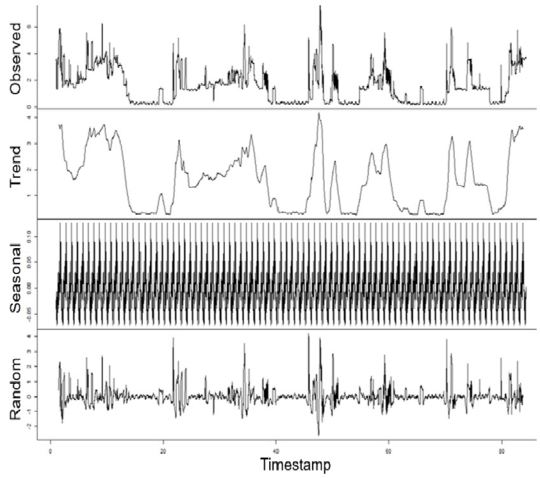


**Fig. 1.** Time-series decomposition to identify energy consumption components.

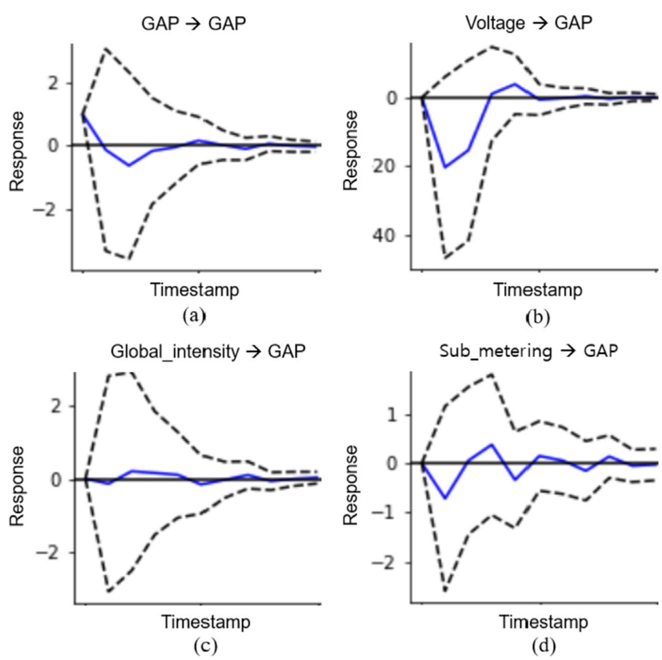


**Fig. 2.** Graph of impulse response function for multivariate correlation analysis.

know the response of a single variable to the impulse generated by other variables in the energy consumption forecasting system and is needed to find out the problem. If there is a response to a variable and the impulse of another variable, the latter can be called causal for the former case. We can see the causality between variables in Fig. 2. Fig. 2 (a) shows the effect of global active power (GAP) used in the previous time on power demand forecasting. Fig. 2 (b) shows the effect of the voltage used in the previous time on the power demand forecast. Fig. 2 (c) shows the effect of the intensity used in the previous time on the power demand forecast. Fig. 2 (d) shows the effect of electricity consumption on demand for electricity demand. According to Fig. 2, the influence among variables is large. So, it is very difficult to extract features among variables [12]. In addition, the behavior of the residents and the number of households have a significant impact on the fluctuation of electric energy consumption. Especially, household occupancy is the key factor which primarily drives the household energy consumption. R. Brown et al. formed the basis for the implementation of automated energy saving actions based on a unique household energy profile of human behavior and occupancy [13]. C. Beckel et al. used power consumption data collected from a smart meter to estimate the household occupancy [14].

The success of machine learning algorithms generally depends on the representation of the data. Among them, the deep learning method aims to create synthesis of several nonlinear transformations, useful expressions that can yield more abstract and ultimately more profit [15]. This deep learning method has been actively studied in various applications. Among the various areas of deep learning, convolutional neural network (CNN) is superior to existing methods in image recognition and recurrent neural network (RNN) achieves excellent performance in speech recognition and natural language processing. CNN learns the weights of the feature maps that consist of each layer, extracts abstract visual features such as points, lines, and faces of the input data, and preserves the relationships between pixels for the learning image [16]. The RNN stores the long sequential information in hidden memory for proper processing, representation, and storage. It also updates over time to ensure that the time information persists [17].

In recent years, many researchers have conducted to extract temporal and spatial features by combining CNN and LSTM models. In the field of natural language processing, Wang et al. analyzed the emotions using the regional CNN-LSTM model with text input [18]. In the field of speech processing, Sainath et al. demonstrated various voice search tasks using the CLDNN model combining CNN, LSTM, and DNN to show the robustness against noise [19]. In the field of video processing, Ullah et al. designed a model combining CNN and Bi-directional LSTM to recognize human action in video sequences [20]. In the medical field, Oh et al. used a model combining CNN and LSTM to accurately detect arrhythmias in the ECG [21]. In the industrial area, Zhao et al. designed the convolutional bi-directional long short-term memory network by mechanical health monitoring technique to predict tool wear [22].

The energy consumption prediction problem also deals with time series data applied in various domains. The conventional method uses a part of all the features and only models the temporal information for the time series prediction. We choose CNN-LSTM for energy consumption prediction, because CNN is used to remove noise and to take into account the correlation between multivariate variables, LSTM models temporal information and maps time series into separable spaces to generate predictions.

In this paper, we propose a CNN-LSTM neural network combining CNN and LSTM to predict residential energy consumption. Power consumption is multivariate time series that is recorded over time, including spatial information among variables and irregular patterns of temporal information [23]. Administrators can use CNN-LSTM model that learned through the spatial-temporal features of electrical energy consumption and can adjust supply appropriately through predicted consumption [24].

Our proposed CNN-LSTM method uses the CNN layers to extract features among several variables that affect energy consumption prediction. The outputs of these CNN layers are used as an input to the LSTM layer after extracting the important features of the power consumption and removing the noise [25]. The last output of the LSTM layer is passed to a fully connected layer, which easily generates a predicted time series of energy consumption [26]. The LSTM layer remembers irregular trend factor of electrical energy consumption. Finally, we change the time resolution to see if further improvements can be made using the CNN-LSTM method. We confirm that the CNN layer reduces the data spectrum and extracts the features in multiple variables. We also analyze which attributes are the most important selects in the CNN layers. This is the first time that CNN-LSTM has been designed, trained and analyzed to predict residential power consumption. The main contributions of this paper are as follows:

- We propose a novel deep learning model to stably predict the amount of electric energy consumption collected in an actual residential house.
- We achieve the highest performance in high resolution compared with the previous works.
- We analyze the variables of household appliances that influence the prediction of energy consumption.

The rest of this paper is organized as follows. In Section 2, we discuss the related work on electric energy consumption prediction. Section 3 details the proposed CNN-LSTM hybrid neural networks architecture. Section 4 presents experimental results and Section 5 concludes the paper.

## 2. Related works

A lot of studies have been conducted to extract features from energy consumption data and predict electrical energy consumption for stable supply. There are three categories used for energy consumption forecasting: statistical-based modeling, machine learning-based modeling, and deep learning-based modeling. Table 1 shows the studies related to power consumption prediction according to three categories.

N. Fume predicted the residential energy used in the home using multiple linear regression [27]. They confirmed that the time resolution of the observed data affects the performance of the predictive model. K. P. Amber predicted daily electricity consumption by applying genetic programming to multiple linear regression [28]. They used genetic programming to integrate five important independent variables to predict. These methods mainly used a general linear regression. This model design is straightforward and eliminates unnecessary variables to improve predictive performance stability. However, the correlation between independent variables used for prediction can lead to Multicollinearity problems. In addition, there is a disadvantage that it is difficult to obtain explanatory variables by using the linear regression.

Y. Chen used support vector regression to predict the electricity loads that occur in buildings [31]. They tried to improve performance by adding ambient temperature as well as basic electricity loads. A. Bogomolov attempted to predict energy consumption using a random forest regression method [33]. They used human dynamics analysis to predict next week energy consumption. These methods can generate complex decision boundaries even if there are not many features of the data [39]. However, the existing machine learning methods have severe overfitting when the complicated correlation of variables becomes complicated or the amount of data increases. When overfitting occurs, it is difficult to predict long-term consumption.

W. Kong used a sequence to sequence method to predict power loads in a group of general individual families [35]. They achieved higher performance than the existing prediction methods using the parameters measured by the real residential smart meter. C. Li predicted building energy consumption using stacked autoencoders [36]. They use the deep learning method to reduce randomness and noisy disturbance from power consumption data and extract important features. These methods can extract and model important features automatically even if they have a lot of existing data and complex attributes [40]. However, the conventional deep learning methods have difficulty in modeling the spatial-temporal features of the power consumption together.

As mentioned above, there have been a lot of attempts to predict energy consumption using various of modeling techniques, such as statistical-based, machine learning-based, and deep learning-based methods. However, there has been no attempt to predict the spatial correlation of variables with time information in power consumption, multivariate time series data. Most studies select only some of the important variables in energy consumption data and then model time information to predict energy consumption. Therefore, proper learning method is needed to model irregular time information of power consumption and spatial information of all variables [41].

## 3. The proposed method

Fig. 3 shows the overall architecture for predicting residential power consumption using CNN-LSTM model. We use actual residential energy consumption data composed of several housing variables. The CNN-LSTM model learns the input data preprocessed every 60 min by the sliding window algorithm [42]. The spatial characteristics of a multivariate time series variable are

**Table 1**
Related works on electric energy consumption prediction.

| Category | Author | Year | Data | Method | Description |
|---|---|---|---|---|---|
| Statistical based modeling | N. Fumo [27] | 2015 | Electricity consumption | Linear regression | Analysis of prediction performance according to time resolution |
| | K.P. Amber et al. [28] | 2015 | Electricity consumption | Multiple regression | Selecting critical variables using genetic algorithms |
| | D. H. Vu et al. [29] | 2015 | Electricity consumption | Multiple regression | Backward elimination processes to select the variables |
| | M. R. Braun et al. [30] | 2014 | Electricity consumption | Multiple regression | Using temperature and humidity records to improve performance |
| Machine learning based modeling | Y. Chen [31] | 2017 | Electricity load | Support vector regression | Predictions using electrical loads on typical office buildings |
| | Y. Yaslan [32] | 2017 | Electricity load | Support vector regression | Predictions using electrical loads on typical office buildings |
| | A. Bogomolov et al. [33] | 2016 | Electricity consumption | Random forest regression | Predicting energy consumption based on human dynamics analysis |
| | R. K. Jain et al. [34] | 2014 | Electricity consumption | Support vector regression | Predictive performance according to temporal impact |
| Deep learning based modeling | W. Kong et al. [35] | 2017 | Electricity load | Sequence to sequence | Data collection using real residential smart meter sensor |
| | C. Li et al. [36] | 2017 | Electricity consumption | Stacked Autoencoders | Feature extraction of power consumption using Autoencoders |
| | H. Shi [37] | 2017 | Electricity load | Pooling-based DRNN | Removing noise from electricity load using pooling mechanism |
| | D. L. Marino et al. [38] | 2016 | Electricity consumption | Sequence to sequence | Using sequence to sequence modified baseline LSTM |

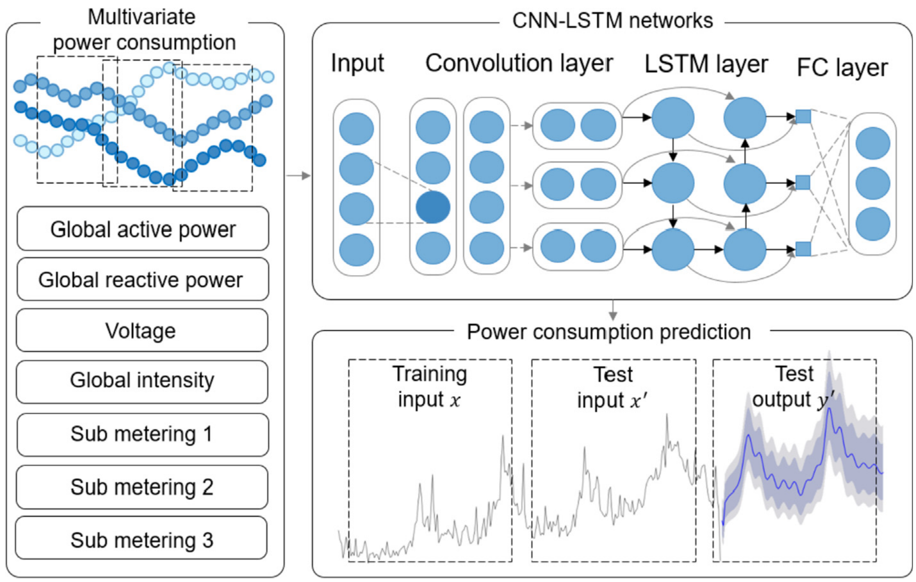


**Fig. 3.** The overall proposed CNN-LSTM structure.

extracted from the convolution and pooling layers of the CNN layer and passed to the LSTM layer with the noise removed. The LSTM layer models the irregular time information using the transmitted spatial features. Finally, the CNN-LSTM method can generate predicted electrical energy consumption in a fully connected hierarchy. The energy consumption values generated by the prediction model are evaluated and analyzed by several error metrics.

### 3.1. CNN-LSTM neural networks

The CNN-LSTM method for predicting electric energy consumption consists of a series connection of CNN and LSTM. CNN-LSTM can extract complex features among multiple sensor variables collected for electric energy demand forecasting and can store complex irregular trends. First, the upper layer of CNN-LSTM consists of CNN. The CNN layer can receive various variables that affect electric energy consumption such as voltage, intensity, and sub metering. In addition, household characteristics such as date, time, behavior of the residents, and household occupancy can also be modeled as meta information in the CNN layer. CNN consists of an input layer that accepts sensor variables as inputs, an output layer that extracts features to LSTMs, and several hidden layers. The hidden layer typically consists of a convolution layer, a ReLU layer, an activation function, and a pooling layer. The convolution layer applies the convolution operation to the incoming multivariate time series sequence and passes the results to the next layer. The convolution operation emulates the response of individual neurons to visual stimulation. Each convolution neuron processes energy consumption data only for the receptive field. Convolutional operation can reduce the number of parameters and make the CNN-LSTM network deeper. If $x_i^0 = \{x_1, x_2, \cdots, x_n\}$ is power consumption input vector and $n$ is the number of normalized 60 min unit per window. Equation (1) is the result of the vector $y_{ij}^1$ output from the first convolutional layer, $y_{ij}^1$ is calculated by output vector $x_{ij}^1$ of the previous layer, $b_j^1$ represents the bias for the $j^{th}$ feature map, $w$ is the weight of the kernel, $m$ is the index value of the filter, and σ is the activation function like ReLU. Equation (2) is the result of the vector $y_{ij}^l$ output from the $l^{th}$ convolutional layer.

$$y_{ij}^1 = \sigma\left(b_j^1 + \sum_{m=1}^{M} w_{m,\,j}^1 x_{i+m-1,\,j}^0\right) \tag{1}$$

$$y_{ij}^l = \sigma\left(b_j^l + \sum_{m=1}^{M} w_{m,\,j}^l x_{i+m-1,\,j}^0\right) \tag{2}$$

The convolution layer uses a pooling layer that combines the output of a neuron cluster in one layer into a single neuron in the next layer. The pooling layer reduces the space size of the representation to reduce the number of parameters and network computation costs. The max-pooling used for power demand prediction uses the maximum value from each neuron cluster in the previous layer. This also has the effect of adjusting overfitting. Equation (3) represents the operation of the max-pooling layer. $T$ is the stride that decide how far to move the area of input data, and $R$ is the pooling size of less than the size of the input $y$.

$$p_{ij}^l = \max_{r\,\in R} y_{i\times T+r,j}^{l-1} \tag{3}$$

LSTM, which is a lower layer of CNN-LSTM, stores time information about important characteristics of power demand extracted through CNN. LSTM provides a solution by preserving long-term memory by consolidating memory units that can update the previous hidden state. This function makes it easy to understand temporal relationships on a long-term sequence. The output values from the previous CNN layer are passed to the gate units. The LSTM network is well suited for predicting power demand by addressing explosive and vanishing gradient problems that can occur when learning traditional RNNs. The three gate units are a mechanism for determining the state of each individual memory cell through multiplication operations. The gate unit consists of input, output, and forget gate, depending on the function. The memory cells that make up the LSTM update their states with activation of each gate unit that are controlled to a continuous value between 0 and 1. The hidden state of the LSTM cell, $h_t$, is updated every $t$ step.

$$i_t = \sigma(W_{pi}p_t + W_{hi}h_{t-1} + W_{ci}\circ c_{t-1} + b_i) \tag{4}$$

$$f_t = \sigma\left(W_{pf}p_t + W_{hf}h_{t-1} + W_{cf} \circ c_{t-1} + b_f\right) \quad (5)$$

$$o_t = \sigma(W_{po}p_t + W_{ho}h_{t-1} + W_{co} \circ c_t + b_o) \quad (6)$$

Equations (4)–(6) show the operation of the input gate, the forget gate, and the output gate constituting the LSTM, and the output of each gate is represented by $i, f$, and $o$ notation. Equations (7) and (8) represent the cell states and the hidden states determined through the input, the forget, and the output gate. They are represented by c and h notation. σ is an activation function such as tanh. This activation function has non-linearity and similarly squash its inputs to [-1, 1] ranges. $W$ is the weight matrix of each gate unit and $b$ is the bias vector. The term $p_t$ contains the critical features of electric energy consumption as the output of the pooling layer at time $t$ and is used as input to the LSTM memory cell. CNN-LSTM networks using LSTM cells provide superior performance through time information modeling of signals and provide leading-edge results in predicting residential energy consumption.

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma(W_{pc}p_t + W_{hc}h_{t-1} + b_c) \quad (7)$$

$$h_t = o_t \circ \sigma(c_t) \quad (8)$$

The last layer of CNN-LSTM is made up of fully connected layers. This can be used to generate the power consumption over a certain period of time. The output of the LSTM unit is flattened to a feature vector. $h^l = \{h_1, h_2, \ldots h_l\}$ where $l$ is the number of units in LSTM. The output of the LSTM is used as input to the fully connected layer. We use CNN-LSTM to predict energy consumption in 60 min. Equation (9) shows the equation used in this layer. σ is a non-linear activation function, $w$ is the weight of the $i^{th}$ node for layer $l-1$ and the $j^{th}$ node for layer $l$, and $b_i^{l-1}$ represents a bias.

$$d_i^l = \sum_j w_{ji}^{l-1}\left(\sigma\left(h_i^{l-1}\right) + b_i^{l-1}\right) \quad (9)$$

### 3.2. Architecture

The design of CNN-LSTM can be variously modified according to the type and parameter adjustment of the layers constituting the network. The CNN-LSTM basically consists of Convolutional layer, Pooling layer, LSTM layer and Dense layer [43]. Each layer can adjust the number of filters, the kernel size, and the number of strides. Adjusting these parameters can affect learning speed and performance depending on the characteristics of the learning data [44]. We can confirm the performance change by increasing or decreasing the parameter. In order to adjust the parameters and build an optimal architecture for energy consumption prediction, we need to understand the characteristics of the input data. Energy consumption data is multivariate time series and is preprocessed into the 60-min window by the sliding window algorithm. We used a $2 \times 1$ kernel to minimize the loss of temporal information. The input of CNN-LSTM is $60 \times 10$ size data. There are a total of 10 variables consisting of a 60-min time-series. The data passes through the convolution layer and the pooling layer and then through the LSTM. We designed the parameters of the CNN-LSTM as shown in Table 2. This table shows the number of filters in each convolution layer, the size and stride number of the convolution layer, the kernel of the pooling layer, and the number of parameters of the entire layers including the LSTM layer.

**Table 2**
The proposed CNN-LSTM architecture.

| Type | # Filter | Kernel size | Stride | # Param |
|---|---|---|---|---|
| Convolution | 64 | (2, 1) | 1 | 192 |
| Activation (ReLu) | – | – | – | 0 |
| Pooling | – | (2, 1) | 2 | 0 |
| Convolution | 64 | (2, 1) | 1 | 8256 |
| Activation (ReLu) | – | – | – | 0 |
| Pooling | – | (2, 1) | 2 | 0 |
| TimeDistributed | – | – | – | 0 |
| LSTM (64) | – | – | – | 180,480 |
| Activation (tanh) | – | – | – | 0 |
| Dense (32) | – | – | – | 2080 |
| Dense (60) | – | – | – | 1980 |
| Total number of parameters | | | | 192,988 |

## 4. Experiments

### 4.1. Individual household electric power consumption dataset

In this paper, we have tested the individual household electric power consumption dataset provided by UCI machine learning repository [10]. To predict the validation of the proposed power consumption prediction architecture, we used several time series variables to predict the global active power (GAP). This dataset is displayed in 1-min units with actual power consumption data collected from a household in France. A total of 25,979 missing values were removed for preprocessing. Therefore, we applied the sliding window algorithm to use multivariate time series data as input. To test the proposed method, The sliding window algorithm used to create a 60-min window and label the next 60 min as a predicted result. The multivariate time series converted to a two-dimensional window for input to the CNN-LSTM neural network.

Table 3 lists the 9 variables that make up the power consumption data and the 3 variables collected from energy consumption sensors. In Table 4, we have provided a quantitative form for detailed description of the dataset on individual household electric power consumption [45]. The time variables include day, month, year, hour, and minute. The variables collected from the energy consumption sensors include sub_metering 1, sub_metering 2, and sub_metering 3. Sub_metering 1 corresponds to the kitchen including the dishwasher and microwave oven. Sub_metering 2 corresponds to the laundry room which includes washing machine, tumble-drier, and lighting. Sub_metering 3 corresponds to electric water heater and air conditioner.

### 4.2. Performance comparison with conventional machine learning methods

To verify the usefulness of the proposed CNN-LSTM power consumption prediction model, we conducted experiments using linear regression and machine learning algorithms. 10-fold cross validation was used for performance comparison experiments. The proposed CNN-LSTM for the prediction of residential power consumption achieved the lowest mean square error (MSE) compared to other machine learning methods as well as linear regression. Following the CNN-LSTM model, linear regression model (LR), random forest regression (RF), decision tree (DT) regression and multilayer perceptron (MLP) were obtained with low MSE. Fig. 4 is a box plot showing the MSE obtained by 10-fold cross validation experiments. We can see that the proposed method shows the best performance on average and reveals the significant difference of MSE compared with the existing method in Fig. 4. In addition, we can confirm that the proposed method guarantees a more stable performance through the distance of the variables from the average

**Table 3**
The features of individual household power consumption dataset.

| # | Attribute | Description |
|---|---|---|
| 1 | Day | An integer value between 1 and 31 |
| 2 | Month | An integer value between 1 and 12 |
| 3 | Year | An integer value between 2006 and 2010 |
| 4 | Hour | An integer value between 0 and 23 |
| 5 | Minute | An integer value between 1 and 60 |
| 6 | Global active power (GAP) | Household global minute-averaged active power (in kilowatt) |
| 7 | Global reactive power (GRP) | Household global minute-averaged reactive power (in kilowatt) |
| 8 | Voltage | Minute-averaged voltage (in volt) |
| 9 | Global intensity (GI) | Household global minute-averaged current intensity (in ampere) |
| 10 | Sub metering 1 (S1) | It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave, hot plates being not electric but gas powered (in watt-hour of active energy) |
| 11 | Sub metering 2 (S2) | It corresponds to the laundry room, containing a washing machine, a tumble-drier, a refrigerator and a light (in watt-hour of active energy) |
| 12 | Sub metering 3 (S3) | It corresponds to an electric water heater and an air conditioner (in watt-hour of active energy) |

**Table 4**
Quantitative form of the dataset on individual household power consumption.

| Attribute | Date | Time | GAP (kW) | GRP (kW) | Voltage | GI (A) | S1 (Wh) | S2 (Wh) | S3 (Wh) |
|---|---|---|---|---|---|---|---|---|---|
| Max | 26/11/2010 | 23:59 | 11.122 | 1.390 | 254.150 | 48.400 | 88.00 | 80.00 | 31.00 |
| Min | 16/12/2006 | 00:00 | 0.076 | 0.000 | 223.200 | 0.200 | 0.000 | 0.000 | 0.000 |
| Average | – | – | 1.089 | 0.124 | 240.844 | 4.618 | 1.117 | 1.289 | 6.453 |
| Std. Dev | – | – | 1.055 | 0.113 | 3.239 | 4.435 | 6.139 | 5.794 | 8.436 |

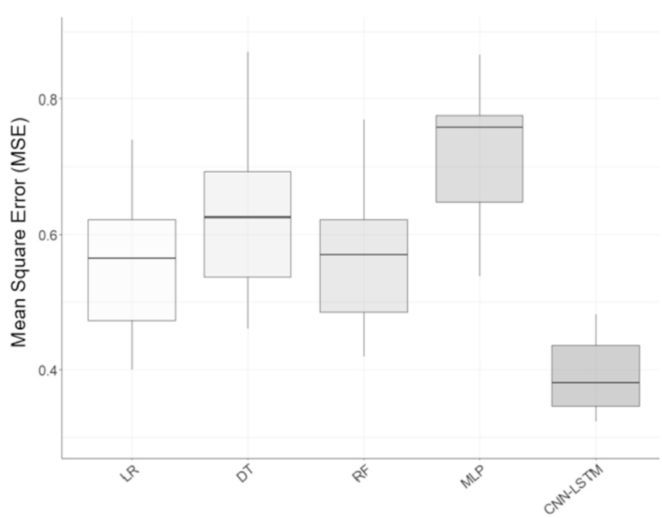


**Fig. 4.** The accuracy of 10-fold cross validation using machine learning.

value. Table 5 shows the settings of the machine learning method used in the experiment. We adjusted the hyperparameter to achieve high performance for each model to compare the existing methods with CNN-LSTM. We used the scikit-learn library to set up a hyperparameter for each model, and each hyperparameter was determined empirically. Equation (10) is used to calculate the MSE. If $\widehat{Y}_i$ is a vector of $n$ predictions generated from a sample of n energy consumption data points on all variables, and $Y_i$ is the vector of observed consumption values of the variable being predicted.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \tag{10}$$

### 4.3. Performance comparison with other deep learning-based models

We have conducted experiments to confirm that the proposed method is superior to other deep learning-based models. Table 6 includes the performance of deep learning models for predicting energy consumption. LSTM, GRU, Bi-LSTM, and Attention LSTM are adopted for time series prediction, and the results are evaluated in four error metrics like MSE, RMSE, MAE, and MAPE. Experimental results show that the proposed CNN-LSTM model achieves superior performance to the conventional deep learning methods for power

**Table 6**
Performance comparison of deep learning methods.

| Method | MSE | RMSE | MAE | MAPE |
|---|---|---|---|---|
| LSTM | 0.7480 | 0.8649 | 0.6278 | 51.45 |
| GRU | 0.7432 | 0.8620 | 0.6291 | 51.47 |
| Bi-LSTM | 0.7235 | 0.8505 | 0.6122 | 50.10 |
| Attention LSTM | 0.6984 | 0.8357 | 0.5911 | 48.36 |
| The proposed | **0.3738** | **0.6114** | **0.3493** | **34.84** |

**Table 5**
Parameters for machine learning techniques.

| # | Model | Description |
|---|---|---|
| 1 | Linear regression | Fit_intercept = True, normalize = False, n_jobs = None |
| 2 | Decision tree regression | Splitter = "best", maximum depth = None, min samples split = 2, min sample leaf = 1 |
| 3 | Random forest regression | No of estimators = 10, maximum depth = None, min samples split = 2, min sample leaf = 1 |
| 4 | Multilayer perceptron | Activation = relu, weight optimization = adam, batch size = 512, number of epochs = 100, learning rate = 0.01 |

demand forecasting. Attention LSTM and Bi-LSTM follow the proposed model. We have confirmed through experiments that the proposed CNN-LSTM is a competitive method for power consumption prediction.

### 4.4. Prediction results for CNN-LSTM network

Fig. 5 shows the predicted results of the CNN-LSTM network. We used linear regression to compare the results of the proposed method. This figure represents that the CNN-LSTM method predicts the global characteristics of power consumption better than the conventional method. It also predicts local characteristics well. The linear regression model shows that modeling performance for local characteristics is poor, while global characteristics are matched similarly. Fig. 5 shows that the proposed method well models the irregular trend of power consumption. We can see that the proposed method minimizes the error in all intervals compared to the linear regression. In addition, it can be confirmed that the complex time-series pattern is predicted almost similarly. Especially, it predicts the peak which occurs frequently in the electric energy consumption.

### 4.5. Comparison of prediction performance by time resolution

We confirmed the effect of changes in time resolution on the CNN-LSTM method. Fig. 6 shows the irregular trend of power consumption data according to time resolution. The electric energy consumption prediction system is divided into four stages

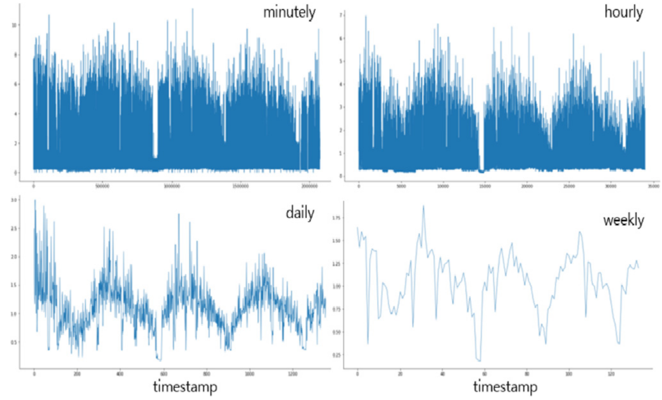


**Fig. 6.** Graph of electrical energy consumption with different time resolution.

including long-term, medium-term, short-term and real-time prediction. We have experimented by aggregating energy consumption by minutely, hourly, daily, and weekly units. The resolution decreases as the unit changes from minutely unit to weekly unit. We have evaluated the proposed model to be robust even in terms of resolution. We used linear regression and LSTM models to compare the experimental results. For the performance evaluation, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) were used. Table 7 shows the performance of each model according to time resolution. The lower the resolution, the lower the error rate of most models. The proposed method achieves higher performance than linear regression and LSTM at all resolutions. Thus, we have proven that it achieves robust performance even with time resolution changes.

Equation (11) is used to calculate the RMSE. If $\widehat{Y}_i$ is a vector of $n$ predictions generated from a sample of n data points on all power consumption variables, and $Y_i$ is the vector of observed consumption of the variable being predicted.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2} \quad (11)$$

Equation (12) is used to calculate the MAE. Assume $X_i$ and $Y_i$Y are variables of paired observations that express the same phenomenon.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (12)$$

Equation (13) is used to calculate the MAPE. MAPE is a measure

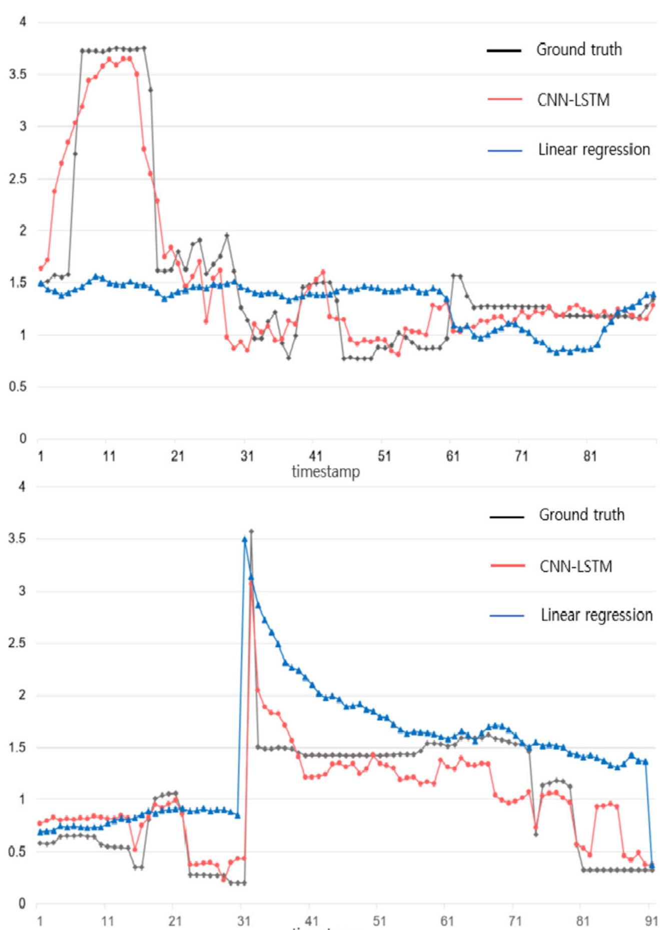


**Fig. 5.** Graph of power consumption prediction for performance comparison.

**Table 7**
Prediction performance with time resolution change.

| Method | Resolution | MSE | RMSE | MAE | MAPE |
|---|---|---|---|---|---|
| Linear Regression | Minutely | 0.4046 | 0.6361 | 0.4176 | 74.52 |
| | Hourly | 0.4247 | 0.6517 | 0.5022 | 83.74 |
| | Daily | 0.2526 | 0.5026 | 0.3915 | 52.69 |
| | Weekly | 0.1480 | 0.3847 | 0.3199 | 41.33 |
| LSTM | Minutely | 0.7480 | 0.8649 | 0.6278 | 51.45 |
| | Hourly | 0.5145 | 0.7173 | 0.5260 | 44.37 |
| | Daily | 0.2406 | 0.4905 | 0.4125 | 38.72 |
| | Weekly | 0.1049 | 0.3239 | 0.2438 | 35.78 |
| CNN-LSTM | Minutely | **0.3738** | **0.6114** | **0.3493** | **34.84** |
| | Hourly | **0.3549** | **0.5957** | **0.3317** | **32.83** |
| | Daily | **0.1037** | **0.3221** | **0.2569** | **31.83** |
| | Weekly | **0.0952** | **0.3085** | **0.2382** | **31.84** |

of prediction accuracy of a CNN-LSTM method in statistics, for example in trend prediction. Where $A_t$ is the actual consumption value and $F_t$ is the predicted value. The absolute value in this calculation is summed for every predicted point in time and divided by the number of fitted points $n$.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{13}$$

### 4.6. Performance comparison with competitive benchmarks

We compare the performance of the proposed method with competitive benchmarks using individual household electric power consumption dataset. Table 8 summarizes the performance of competitive benchmarks for test cases. We set up the experiment with the same time resolution of minutely, hourly, daily, and weekly units. Marino et al. analyzed a new energy load prediction methodology based on deep neural networks and an LSTM-based Sequence-to-Sequence (Seq2Seq) [46]. Mocanu et al. investigated two newly developed stochastic models for time series prediction of energy consumption, namely conditional restricted Boltzmann machine (CRBM) and factored conditional restricted Boltzmann machine (FCRBM) [47]. We confirm that the proposed method is useful by comparing the performance with competitive benchmarks.

### 4.7. CNN-LSTM model internal analysis

We have analyzed the inside of the model to confirm the operation principle of CNN-LSTM. We use the 60-min power consumption time window as input. The input window is reduced four times through two convolution and pooling layers [24]. The output result is used as an input to the LSTM layer. The power consumption through the convolution layer shows that the noise is reduced, but the temporal trend is almost the same. We have demonstrated through model internal analysis that CNN-LSTM performs modeling with minimal loss of information. Fig. 7 shows the kernel output of power consumption with reduced spatial spectrum by convolution and pooling operations.

We have visualized some of the multiple kernels in the convolution layer. The intermediate output of energy consumption data filtered by A and B kernels in Fig. 7 not only reduces noise, but also preserves local and global feature. Because energy consumption prediction contains irregular time information, it is important that the output of the kernel preserves its feature. This internal visualization is useful for understanding the behavior of the model by analyzing the intermediate output of CNN-LSTM.

### 4.8. Analysis of variables affecting power consumption prediction

Fig. 8 shows the activation region of the last convolutional layer visualized using the class activation map. We use the class activation map to obtain a weighted activation map generated for each electric energy consumption window [48]. The class activation map

**Table 8**
Competitive benchmark performance.

| Resolution | Minutely | Hourly | Daily | Weekly |
|---|---|---|---|---|
| CRBMs | 1.0619 | 0.4769 | – | 0.2434 |
| FCRBM | 0.8091 | 0.4396 | – | 0.2139 |
| Seq2Seq | 0.4448 | 0.3906 | – | – |
| CNN-LSTM | **0.3738** | **0.3549** | **0.1037** | **0.0952** |

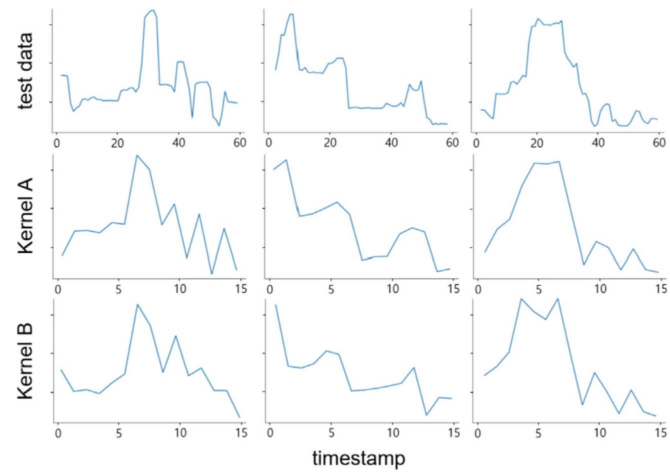


**Fig. 7.** Visualization of CNN-LSTM internal analysis.

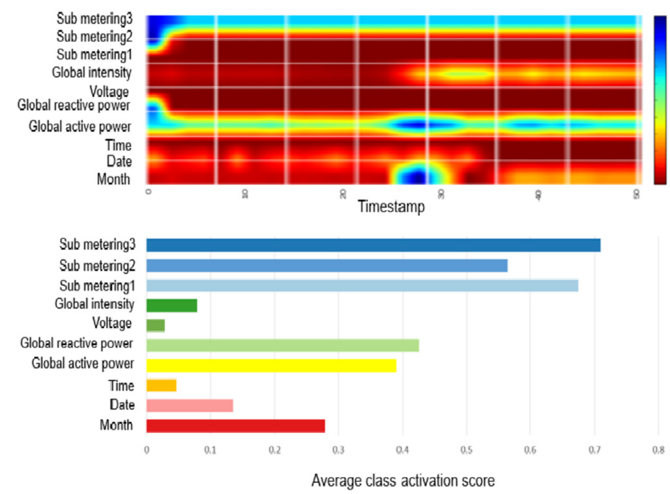


**Fig. 8.** Visualization of critical variables using class activation maps.

can also localize the receptive field by summing the weights of CNN's top-most feature maps. Fig. 8 represents the input of multivariate energy consumption test data divided into 60 min using the sliding window algorithm. We can see in Fig. 8 which features are activated. In the heatmap, the more blue, the more influence on prediction, and the more red, the less effect on prediction. We calculate the average of the class activation scores for all test data. We can see that sub_metering1, sub_metering2 and sub_metering3 are important variables through the average value of the class activation score. In particular, sub_metering3 corresponds to an electric water heater and an air conditioner. The use of heating and cooling systems has a significant impact on energy consumption demand and is an important parameter in predicting consumption. We explain the electric energy consumption variables of household appliances that have influenced the CNN-LSTM prediction model.

## 5. Discussions

We have proposed a CNN-LSTM network that models temporal features and the correlations of multivariate variables to predict residential power consumption. The reason for the difficulty in predicting residential energy consumption is that it has an irregular trend as shown in the time series decomposition factor of Fig. 1. In addition, since the residential power consumption prediction is a multivariate time series problem, as shown in the impulse response function graph in Fig. 2, several property variables affect one prediction value. We have improved the modeling performance by linearly combining CNN and LSTM to work out these difficulties and to model complex features of actual residential power consumption dataset.

We have compared the performance with the conventional machine learning methods in Fig. 4. The performance is evaluated using 10-fold crossover validation for reliable generalization performance measurements. We obtain the highest performance of 0.37 MSE in CNN-LSTM. Also, the standard deviation of cross-validation is small compared to other models. As can be seen in Table 6, CNN-LSTM show excellent performance compared to LSTM, GRU, Bi-LSTM and Attention LSTM, which are competitive deep learning methods for power demand forecasting. As low-resolution prediction of power demand causes inefficiency of energy production for operation, high-resolution prediction is indispensable for efficient forecasting. We see that the proposed method achieves high prediction performance at high resolution.

As shown in Fig. 6, we aggregate energy consumption on a time axis. Our method achieves high performance at all resolutions as shown in Table 7. We explain the energy consumption variables of household appliances that have influenced the model prediction. We have confirmed in Fig. 7 that the CNN-LSTM model removes noise from power demand. In order to identify the variables affecting the power demand forecasting of the model, we analyze the black-box model using the class activation score in Fig. 8.

## 6. Conclusions

We propose a CNN-LSTM model for robust and efficient forecasting of residential energy consumption. Our model quickly and accurately predicts electrical energy consumption by extracting features from variables that affect power consumption. The proposed model is compared to other machine learning methods to demonstrate its usefulness and excellence. We used CNN-LSTM to model decision variables and temporal information from multivariate time-series. We compare the predicted result graph with the linear regression to confirm that CNN-LSTM predicts well the local features of the time series. Also, we confirmed that it achieves better performance than existing models without dependence on time resolution. Finally, we have confirmed the process of reducing the noise of power consumption data through CNN-LSTM internal analysis and analyzed the variables that have important influence on energy consumption prediction using class activation map. The CNN-LSTM model proposed in this paper predicts irregular trends in electrical energy consumption that could not be predicted in existing machine learning methods. The results of this paper represent that the CNN-LSTM model predicts residential electricity demand efficiently and stably, and shows the highest performance compared with the existing competitive benchmarks. We have also attempted to explain the variables that affect the prediction.

1) We propose a CNN-LSTM network to estimate the electric energy consumption in real residential houses with a stable performance of 0.37 MSE.
2) Our model predicts complex electric energy consumption with the highest performance in all cases of minutely, hourly, daily, and weekly unit resolutions compared to other methods.
3) We have analyzed the proposed method and found the variables of electric water heater and air conditioner that have the greatest influence on the prediction model.

On the other hand, the potential limitation of the CNN-LSTM model lies in relatively large efforts by trial and error to determine the optimal hyperparameters. In order to work out this problem, we need to automate searching for the best hyperparameters, and now we are working on a genetic algorithm that can automatically search for the hyperparameter space of CNN-LSTM's. Also, we need to collect energy consumption data from a lot of houses to validate the model. Household characteristics such as occupancy and behavior have a large influence on predicting electric energy consumption. However, there is no information about the household characteristics in our current data. We should have collected the household characteristics, an important variable that indicates the number and behavior of residents, and verify the impact of the variables. In addition, further research is needed to confirm the modeling process of temporal feature through analysis of LSTM layer in CNN-LSTM. As a policy implication, electric energy prediction has a great influence on stable power supply, efficient operation management, and safety of power generation system. To manage increasing energy consumption, the policy will introduce a power consumption prediction method to minimize costs. Therefore, the proposed method can minimize energy wastage and economic loss caused by unplanned power plant operation.

## Acknowledgements

This research was supported by Korea Electric Power Corporation (Grant number: R18XA05).

## References

[1] Zhao GY, Liu ZY, He Y, Cao HJ, Guo YB. Energy consumption in machining: classification, prediction, and reduction strategy. Energy 2017;133:142–57.
[2] Sieminski A. International energy outlook. *Energy Information Administration (EIA)*; 2017. p. 5–30.
[3] Nejat P, Jomehzadeh F, Taheri MM, Gohari M, Majid MZA. A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries). Renew Sustain Energy Rev 2015;43:843–62.
[4] Ibrahim H, Ilinca A, Perron J. Energy storage systems-characteristics and comparisons. Renew Sustain Energy Rev 2008;12(5):1221–50.
[5] Deb C, Zhang F, Yang J, Lee SE, Shah KW. A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 2017;74:902–24.
[6] Kim K-H, Cho S-B. Modular Bayesian networks with low-power wearable sensors for recognizing eating activities. Sensors 2017;17(12):1–16.
[7] Arghira N, Hawarah L, Ploix S, Jacomino M. Prediction of appliances energy use in smart homes. Energy 2012;48(1):128–34.
[8] Ahmad MI. Seasonal decomposition of electricity consumption data. Review of Integrative Business and Economics Research 2017;6(4):271–5.
[9] Chujai P, Kerdprasop N, Kerdprasop K. Time series analysis of household electric consumption with ARIMA and ARMA models. Proc. of the Int. Multi-Conf. of Engineers and Computer Scientists 2013;1:295–300.
[10] Hébrail G, Bérard A. "Individual household electric power consumption data set," *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science; 2012. Retrieved from, http://archive.ics.uci.edu/ml.
[11] Koop G, Pesaran MH, Potter SM. Impulse response analysis in nonlinear multivariate models. J Econom 1996;74(1):119–47.
[12] Beckel C, Sadamori L, Staake T, Santini S. Revealing household characteristics from smart meter data. Energy 2014;78:397–410.
[13] Brown R, Ghavami N, Adjrad M, Ghavami M, Dudley S. Occupancy based household energy disaggregation using ultra wideband radar and electrical signature profiles. Energy Build 2017;141:134–41.
[14] Beckel C, Sadamori L, Staake T, Santini S. Revealing household characteristics from smart meter data. Energy 2014;78:397–410.
[15] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35(8):1798–828.
[16] Donahue J, Anne Hendricks L, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T. Long-term recurrent convolutional networks for visual recognition and description. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2015:2625–34.
[17] Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: a search space odyssey. IEEE Trans. on Neural Networks and Learning Systems 2017;28(10):2222–32.
[18] Wang J, Yu LC, Lai KR, Zhang X. Dimensional sentiment analysis using a regional CNN-LSTM model. Proc. of the Annual Meeting of the Association for Computational Linguistics 2016;2:225–30.

[19] Sainath TN, Vinyals O, Senior A, Sak H. Convolutional, long short-term memory, fully connected deep neural networks. IEEE Int. Conf. on Acoustics, Speech and Signal Processing 2015:4580–4.
[20] Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW. Action recognition in video sequences using deep bi-directional LSTM with CNN features. IEEE Access 2018;6:1155–66.
[21] Oh SL, Ng EY, San Tan R, Acharya UR. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. Comput Biol Med 2018;102:278–87.
[22] Zhao R, Yan R, Wang J, Mao K. Learning to monitor machine health with convolutional bi-directional LSTM networks. Sensors 2017;17(2):273–90.
[23] Wang CH, Grozev G, Seo S. Decomposition and statistical analysis for regional electricity demand forecasting. Energy 2012;41(1):313–25.
[24] Kim T-Y, Cho S-B. Predicting the household power consumption using CNN-LSTM hybrid networks. In: In int. Conf. On intelligent data engineering and automated learning; 2018. p. 481–90.
[25] Guo Z, Zhou K, Zhang X, Yang S. A deep learning model for short-term power load and probability density forecasting. Energy 2018;160:1186–200.
[26] Bouktif S, Fiaz A, Ouni A, Serhani M. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches. Energies 2018;11(7):1–20.
[27] Fumo N, Biswas MR. Regression analysis for prediction of residential energy consumption. Renew Sustain Energy Rev 2015;47:332–43.
[28] Amber KP, Aslam MW, Hussain SK. Electricity consumption forecasting models for administration buildings of the UK higher education sector. Energy Build 2015;90:127–36.
[29] Vu DH, Muttaqi KM, Agalgaonkar AP. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. Appl Energy 2015;140:385–94.
[30] Braun MR, Altan H, Beck SBM. Using regression analysis to predict the future energy consumption of a supermarket in the UK. Appl Energy 2014;130: 305–13.
[31] Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, Wang K. Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings. Appl Energy 2017;195: 659–70.
[32] Yaslan Y, Bican B. Empirical mode decomposition based denoising method with support vector regression for time series prediction: a case study for electricity load forecasting. Measurement 2017;103:52–61.
[33] Bogomolov A, Lepri B, Larcher R, Antonelli F, Pianesi F, Pentland A. Energy consumption prediction using people dynamics derived from cellular network data. EPJ Data Science 2016;5(1):1–13.
[34] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Appl Energy 2014;123:168–78.
[35] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans. on Smart Grid (Early Access) 2017:1–11.
[36] Li C, Ding Z, Zhao D, Yi J, Zhang G. Building energy consumption prediction: an extreme deep learning approach. Energies 2017;10(10):1525.
[37] Shi H, Xu M, Li R. "Deep learning for household load forecasting–a novel pooling deep RNN. IEEE Trans. on Smart Grid 2017;9(5):5271–80.
[38] Marino DL, Amarasinghe K, Manic M. Building energy load forecasting using deep neural networks. IEEE Annual Conf. on Industrial Electronics Society 2016:7046–51.
[39] Ronao CA, Cho S-B. Anomalous query access detection in RBAC-administered databases with random forest and PCA. Inf Sci 2016;369:238–50.
[40] Ronao CA, Cho S-B. Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models. Int J Distributed Sens Netw 2017;13(1):1–16.
[41] Koschwitz D, Frisch J, Van Treeck C. Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: a comparative study on district scale. Energy 2018;165:134–42.
[42] Kim T-Y, Cho S-B. Web traffic anomaly detection using C-LSTM neural networks. Expert Syst Appl 2018;106:66–76.
[43] Zhou C, Sun C, Liu Z, Lau F. A C-LSTM neural network for text classification. 2015. arXiv preprint arXiv:1511.08630.
[44] He K, Sun J. Convolutional neural networks at constrained time cost. IEEE Conf. on Computer Vision and Pattern Recognition 2015:5353–60.
[45] Kim J-Y, Cho S-B. Electric energy consumption prediction by deep learning with state explainable autoencoder. Energies 2019;12(4):739–52.
[46] Marino DL, Amarasinghe K, Manic M. Building energy load forecasting using deep neural networks. IEEE Conf. of the Industrial Electronics Society 2016: 7046–51.
[47] Mocanu E, Nguyen PH, Gibescu M, Kling WL. Deep learning for estimating building energy consumption. Sustainable Energy, Grids and Networks 2016;6:91–9.
[48] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition 2016:2921–9.