



Research article

BILSTM-SimAM: An improved algorithm for short-term electric load forecasting based on multi-feature

Mingju Chen^{1,2}, Fuhong Qiu^{1,*}, Xingzhong Xiong^{1,2}, Zhengwei Chang³, Yang Wei³ and JieWu³

¹ School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644002, China

² Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin 644002, China

³ State Grid Sichuan Electric Power Company Electric Power Scientific Research Institute, Chengdu 644002, China

* **Correspondence:** Email: 322085404203@stu.suse.edu.cn.

Abstract: With the growing number of user-side resources connected to the distribution system, an occasional imbalance between the distribution side and the user side arises, making short-term power load forecasting technology crucial for addressing this issue. To strengthen the capability of load multi-feature extraction and improve the accuracy of electric load forecasting, we have constructed a novel BILSTM-SimAM network model. First, the entirely non-recursive Variational Mode Decomposition (VMD) signal processing technique is applied to decompose the raw data into Intrinsic Mode Functions (IMF) with significant regularity. This effectively reduces noise in the load sequence and preserves high-frequency data features, making the data more suitable for subsequent feature extraction. Second, a convolutional neural network (CNN) mode incorporates Dropout function to prevent model overfitting, this improves recognition accuracy and accelerates convergence. Finally, the model combines a Bidirectional Long Short-Term Memory (BILSTM) network with a simple parameter-free attention mechanism (SimAM). This combination allows for the extraction of multi-feature from the load data while emphasizing the feature information of key historical time points, further enhancing the model's prediction accuracy. The results indicate that the R^2 of the BILSTM-SimAM algorithm model reaches 97.8%, surpassing mainstream models such as Transformer, MLP, and Prophet by 2.0%, 2.7%, and 3.6%, respectively. Additionally, the remaining error metrics also show a reduction, confirming the validity and feasibility of the method proposed.

Keywords: VMD; multi-feature; BILSTM; simple attention; short-term load forecasting

1. Introduction

Electricity has powered a global technological revolution and transformed people's lives, the stable operation of the electrical system holds significant importance for the orderly development of the social economy [1]. Short-term load forecasting [2,3] is important to maintain stable electrical system operation. It provides a strong foundation for power dispatch and is crucial for maintaining the balance between load supply and demand within the power grid [4]. Short-term load forecasting accuracy is greatly challenged by the diversity and complexity of influencing factors, such as meteorological factors [5], geographic factors [6], electricity prices [7,8], holidays [9], and more. This challenge arises from the high volatility and randomness [10,11] of short-term load sequences. Consequently, accurate electricity load forecasting is essential for the effective and smooth operation of the power system. There is already a wide variety of load forecasting techniques, which can be broadly categorized based on their level of development.

Prediction techniques based on mathematical models fall under the first group. Examples include the use of the Time series method [12], Regression analysis [13], Kalman filter [14], Exponential smoothing [15], Trend extrapolation [16], Singular spectrum analysis [17], and others. The benefit of these approaches is that the models can be built quickly and predictively, but they have poor performance when faced with nonlinear, stochastic load sequences and high smoothness requirements. Vlahović et al. [18] proposed a trend extrapolation approach to anticipate the load on the power system. This method has the benefit of using fewer load data samples, but it also has the drawback of increasing inaccuracy when the load volatility is high. Lekshmi et al. [19] proposed the use of an auto regressive moving average (ARMA) model in the time series method for load forecasting, with ambient temperature as an influencing factor. Although the forecasting accuracy has improved to some extent, as indicated by the reduced volume of load data required for modeling and the high forecasting accuracy for smooth data and forecasting, the majority of load sequences in the real world are not smooth, leaving room for improvement.

The second group consists of conventional machine learning prediction techniques. The most important ones include Support vector machines (SVM) [20], K-nearest neighbors [21], Decision trees [22], Random forests [23], Deep convolutional neural network (DCNN) [24], and others. Actual load data exhibits nonlinear and non-smooth characteristics [25], and the structure of machine learning prediction models has limitations, making it challenging to achieve the desired prediction accuracy, despite the many advantages of machine learning prediction methods over traditional ones. Sun et al. [26] employed a multi-objective regression model using K-means and KNN. They utilized the KNN algorithm to weigh the prediction points before applying various linear regression techniques for modeling. While regression analysis modeling is quick and easy to use on smooth load data, its predictive impact is less than ideal and cannot be generalized. Dong et al. [27] proposed a short-term electricity load forecasting method based on K-means and SVM. They employed the K-means clustering method to categorize seasonal load data according to temperature characteristics as input data to explore the effect of temperature on seasonal loads. Subsequently, they trained the SVM forecasting model, which improved accuracy and runtime to some extent. However, as the SVM relies on quadratic programming to solve support vectors, its predictive performance tends to diminish when

dealing with a large number of nonlinear prediction samples, which is not conducive to accurate prediction.

The combined prediction approach based on deep learning falls under the third category. Actual load data is often accompanied by irregular vacancies and noise interference, limiting the prediction accuracy of a single machine learning method when data volume is insufficient, even though both mathematical modeling methods and machine learning methods have their advantages [28]. Combination prediction approaches [29,30] have been proposed to enhance forecast accuracy. Zeng et al. [31] proposed a load forecasting method based on empirical mode decomposition (EMD) and long- and short-term time-series network (LSTNet). The results indicated that the hybrid method is beneficial for improving forecasting accuracy. However, EMD-like algorithms often encounter issues such as boundary effects, modal overlap, and sensitivity to noise, potentially impacting the final prediction results. As demonstrated in the experiments from [32], VMD outperforms EMD-like algorithms in decomposition. VMD reduces prediction difficulty and ensures high accuracy by decomposing the original sequence into linear and nonlinear components. Although Chen et al. [33] and Wan et al. [34] combined CNNs with GRUs and LSTMs, respectively. Their prediction performance significantly improved compared to individual GRUs and LSTMs. Later, ordinary attention mechanisms were incorporated into the networks to enhance the influence of important information by assigning different weights, leading to improved accuracy in load prediction. However, directly combining multiple networks can have certain disadvantages, increasing the likelihood of gradient vanishing, gradient explosion, and overfitting in neural networks. According to the experimental results from [35], Dropout has a positive effect on reducing the time complexity of neural networks, enhancing prediction accuracy, and improving the robustness of short-term load forecasting models. Ji et al. [36] proposed a Variational Mode Decomposition (VMD) improved whale optimization algorithm (IWOA) wavelet temporal convolutional network (WTCN) bidirectional gated recurrent unit (BIGRU) attention and Categorical Boosting (CatBoost) model, which achieves good prediction accuracy. However, its complexity and lack of consideration for the holiday factor diminish its practical applicability. Yao et al. [37] proposed a CNN-DBILSTM-Attention-based short-term electricity load prediction. The attention mechanism allows the neural network to focus on specific features through probability allocation, thereby improving prediction accuracy. Nevertheless, the ordinary attention mechanism itself has too many parameters. In combination with CNN and DBILSTM, this results in a significant increase in training parameters, making the entire model overly complex. This complexity is not conducive to improving prediction efficiency. Therefore, the careful selection of appropriate attention mechanisms is crucial to enhance the operational efficiency of complex models.

Building on the previous research results, we further analyze the relationship between the volatility and randomness of power load data and multi-features, and subsequently propose an improved multi-feature-based short-term forecasting algorithm, BILSTM-SimAM, for power load forecasting. To improve the stability of load forecasting, the load sequence is chosen to be partitioned into Intrinsic Mode Functions (IMF) using Variational Mode Decomposition [38] (VMD), thereby reducing the non-stationarity and complexity of the load signal. Next, we leverage the CNN model, augmented with Dropout [39] technology, to extract the key factors influencing load fluctuations, leading to more reliable prediction performance. Further refinement is achieved through load prediction using the BILSTM [40] neural network in combination with the SimAM [41] attention mechanism, resulting in improved predictive accuracy. The major contributions of this paper are as follows: 1) Through the utilization of Variational Mode Decomposition (VMD) in data processing, the original load data is separated into components with distinct frequencies, hence improving the

regularity and consistency of the data. This method is beneficial in the context of multi-feature prediction challenges. 2) The BILSTM-SimAM model framework is presented, demonstrating superior accuracy in power load forecasting while utilizing a small number of model parameters. This method can be readily applied to various power system equipment with adaptability and versatility. 3) The analysis of various factors that influence loads, such as holidays, helps address the discrepancy between the actual distribution side and the client side. This factor is crucial in guaranteeing the reliable functioning of urban electricity networks.

2. Short-term electric load forecasting algorithm

2.1. Implementation process

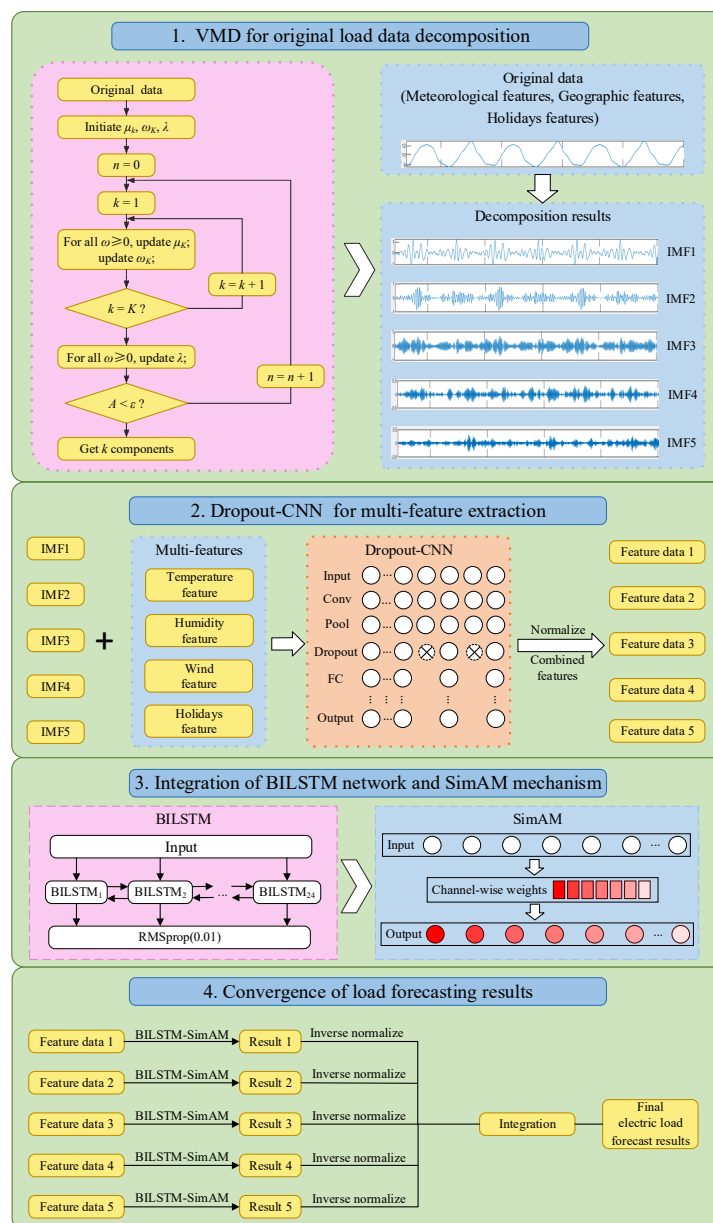


Figure 1. BILSTM-SimAM implementation process.

We propose an improved short-term forecasting model for electricity load base on multi-feature, BILSTM-SimAM, which is mainly divided into four parts:

- VMD for original load data decomposition;
- Dropout-CNN for feature extraction;
- Integration of BILSTM network and SimAM mechanism;
- Convergence of load forecasting results.

The following subsections describe the network modeling and improvement methods, respectively, and the implementation process is shown in Figure 1.

2.2. VMD for original load data decomposition

The time series prediction data utilized pertain to residential electricity demands, which incorporate external elements such as temperature, humidity, wind, and holidays. However, a challenge arises with many external components, such as climatic features, which exhibit low internal regularity. The core idea of the algorithm is to construct and solve the optimal solution of the variational problem in the following steps:

Establish the constrained variational problem. Assuming that the original signal is decomposed into K modal components with different frequency characteristics, the objective is to minimize the sum of the estimated bandwidths of each mode while ensuring that the sum of all modes equals that of the original signal. The constrained variational expression formula is established as follows:

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} \{ \sum_{k=1}^K \|\partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t}\|_2^2 \} \\ \text{s.t. } \sum_{k=1}^K u_k = f(t) \end{aligned} \quad (1)$$

In Eq (1), where $\{U_k\}$ represents the k modal component of the signal decomposition; where $\{\omega_k\}$ represents the frequency center of the k modal component; where $\delta(t)$ represents the Dirac component; where $*$ represents the convolution operator;

Solve the constrained variational problem optimally. The Lagrange transform is first performed to change the above problem into an unconstrained variational problem to be solved. The calculation formula is given as follows:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = \|f(t) - \sum_{k=1}^K u_k(t)\|_2^2 \\ + \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ + \langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \rangle \end{aligned} \quad (2)$$

In Eq (2), where represents the quadratic penalty term factor, where λ represents the Lagrange multiplier operator;

In order to find the optimal solution of the problem, the alternating direction multiplier method is used to update each modal component and its center frequency. The calculation formula is given as follows:

$$\begin{cases} \hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \\ \hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_{k=1}^K \hat{u}_k^{n+1}(\omega)) \\ \omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \end{cases} \quad (3)$$

In Eq (3), where n represents the number of iterations; where τ represents the up-date parameter; where \mathcal{A} represents the Fourier transform of the corresponding signal.

It is evident that the variational load data decomposition technique has processed the original load data, significantly improving its periodicity and smoothness. This enhancement enables better avoidance of the impact of the original load data's volatility and randomness characteristics on load forecasting.

2.3. Designing Dropout-CNN module for multi-feature extraction

Electricity load variation is influenced by multiple factors. Extracting features from these influencing factors is crucial for load variation prediction. Therefore, a Dropout-CNN module is designed to perform multi-feature extraction. CNN, also known as a convolutional neural network, is a powerful tool for extracting data features and Dropout makes CNN networks more stable. Consequently, Dropout-CNN can effectively extract feature quantities that have a significant impact on load variations.

The majority of CNN consists of the convolutional layer and the pooling layer. The convolutional layer employs convolutional kernels to efficiently extract nonlinear local feature from electricity load data, while the pooling layer compresses the extracted feature to generate more significant feature information, enhancing generalization capability. These features are transformed into vectors before being transmitted to the fully connected layer, which primarily maps the features' space to the sample labeling space to enhance network robustness. Finally, the output layer is responsible for data export.

We chose to utilize the well-known CNN LeNet model, renowned for its typical structure comprising 3-layer convolution, 2-layer pooling, and 1-layer complete connection. This model is specifically tailored to accommodate the volume and size of the collected data, characterized by its simplicity and broad applicability. However, in the actual local multi-feature extraction process of LeNet, employing a more complex model for training with relatively limited training data can lead to issues, such as an excessively large loss function in the test load data, resulting in overfitting. This phenomenon reduces the model's generalizability, hampering effective predictions. To address these challenges, we introduced a Dropout layer, incorporating a random deactivation function, to enhance the LeNet model. This modification streamlines training, effectively preventing overfitting and improving generalizability. The construction of the Dropout random deactivation function is illustrated in Figure 2.

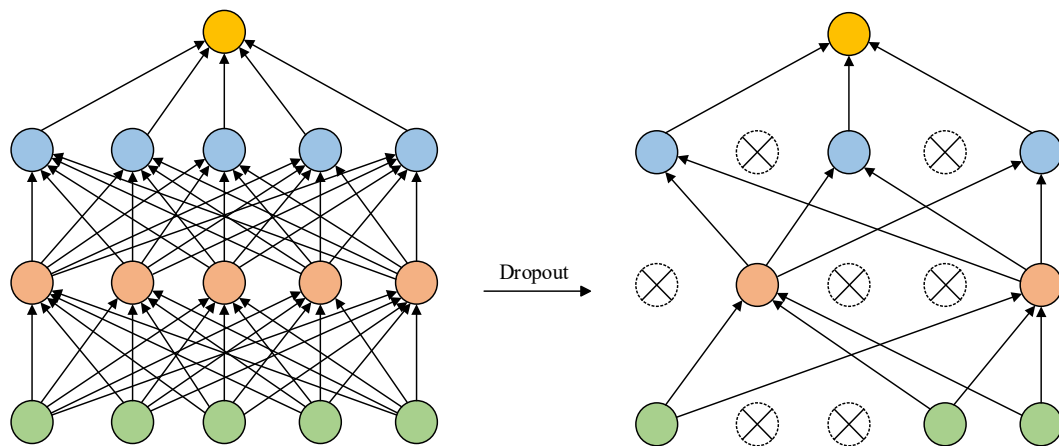


Figure 2. Dropout structure.

The convolutional layer extracts effective nonlinear local feature from the load data, while the pooling layer selects the optimal pooling method to compress the extracted feature and generate more essential feature information. Additionally, a Dropout function layer is introduced after the pooling layer, serving the dual purpose of preventing overfit-ting and improving generalization while reducing model complexity. After the LeNet model has extracted and flattened the load data, the inherent characteristics of the load data are automatically retrieved and then input into the BILSTM model for load prediction. Figure 3 illustrates the basic structure of the Dropout-CNN.

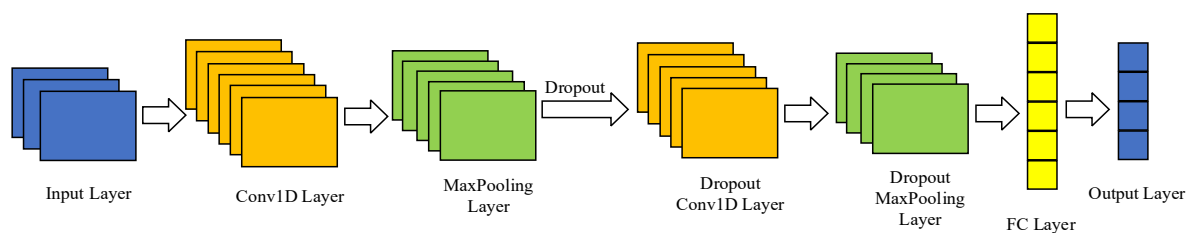


Figure 3. Dropout-CNN model structure.

2.4. Integration of BILSTM network and SimAM mechanism for load forecasting

LSTM stands for Long Short-Term Memory Neural Network. It is essentially a type of recurrent neural network designed to address the limitations of traditional RNN, specifically tailored to handle long-term dependencies. LSTM has made significant advancements and finds extensive use in various fields such as speech recognition, visual description, natural language processing, and more. Figure 4 illustrates the LSTM cell structure.

Multi-featured electric load forecasting is characterized by the requirement of training the network to capture the past and future of the input data and to learn the long-term dependencies in the data, through the analysis we selected BILSTM neural network for multi-featured electric load forecasting. BILSTM, an enhancement of the standard unidirectional LSTM, combines both a forward and a backward LSTM layer, both of which influence the output results. While LSTM fully utilizes historical load data information to prevent the generation of long-distance dependence situations, BILSTM incorporates both forward and backward sequence information inputs, considering past and

future information. This approach is conducive to further improving model prediction accuracy.

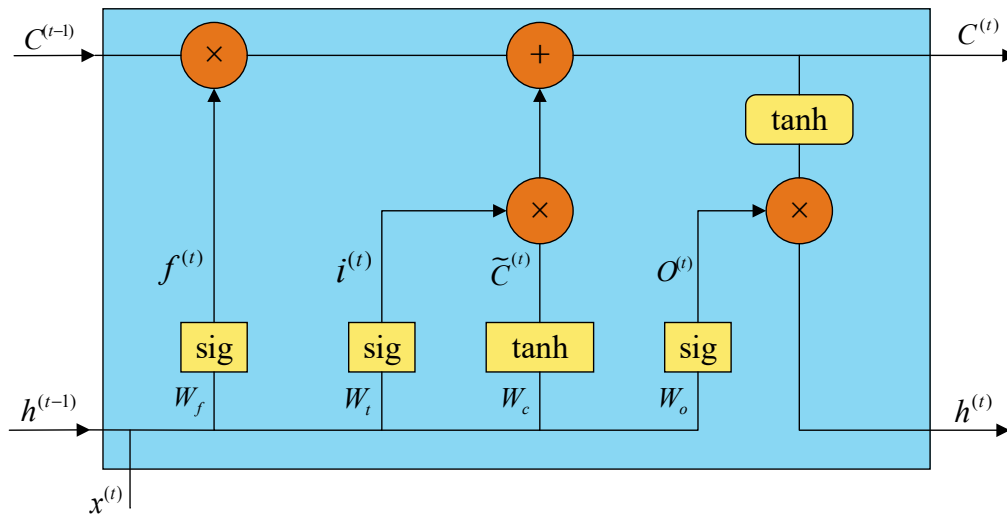


Figure 4. LSTM cell structure.

The structure of BILSTM is shown in Figure 5. Where, x_1, x_2, \dots, x_t denote the corresponding input data at each moment A_1, A_2, \dots, A_t , B_1, B_2, \dots, B_t denote the corresponding forward and backward iterative LSTM hidden states, respectively. y_1, y_2, \dots, y_t denote the corresponding output data.

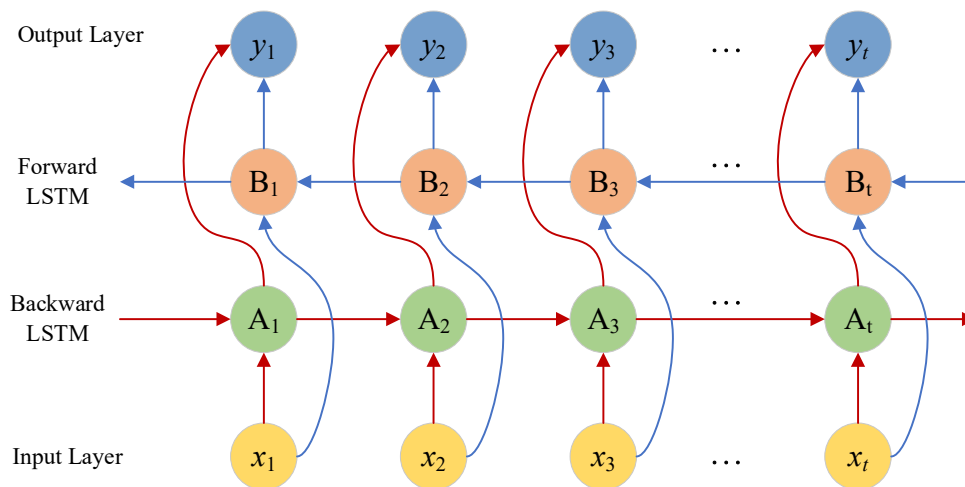


Figure 5. BILSTM neural network structure.

The forgetting gate determines which information should be forgotten from the cell state $C^{(t-1)}$ at time $t-1$, as depicted in Eq (4). The forgetting gate considers the hidden layer state $h^{(t-1)}$ at time $t-1$ and the input sequence $x(t)$ at time t . It produces an output value between 0 and 1, where 1 signifies the retention of complete information and 0 implies complete information discard. The calculation formula is as follows:

$$f^{(t)} = \sigma(W_f \cdot [h^{(t-1)}, x^{(t)}] + b_f) \tag{4}$$

In Eq (4), where $f(t)$ is the oblivious gate state at moment t ; W_f and b_f are the weight and bias of the oblivious gate; and sig is the bipolar sigmoid activation function.

The input gate reads the input $x^{(t)}$ at moment t and determines the information stored in the neuron. Then the temporary state $\tilde{C}^{(t)}$ of the memory cell at moment t is generated by the tanh layer. Finally, the cell state is then updated to obtain the new cell state $C^{(t)}$. The calculation formula is given as follows:

$$i^{(t)} = \sigma \left(W_i \cdot \left[\hat{h}^{(t-1)}, x^{(t)} \right] + b_i \right) \quad (5)$$

$$\tilde{C}^{(t)} = \tanh \left(W_c \cdot \left[\hat{h}^{(t-1)}, x^{(t)} \right] + b_c \right) \quad (6)$$

$$C^{(t)} = f^{(t)} \otimes C^{(t-1)} + i^{(t)} \otimes \tilde{C}^{(t)} \quad (7)$$

In Eqs (5)–(7), $i^{(t)}$ is the input gate state at moment t , which controls the amount of information passed from $x^{(t)}$ to $C^{(t)}$; W_i and b_i are the weight and bias of the input gate; W_c and b_c are the weight matrix and bias term of the cell state; \tanh is the hyperbolic tangent activation function; and \otimes is the Hadamard product.

The output gate selects crucial information from the current state for output. The sigmoid layer initially determines which part of the neuron state should be output. Then, the neuron state earmarked for output passes through the tanh layer and is multiplied by the output of the sigmoid layer to yield the output value $h^{(t)}$, which also serves as the input value for the subsequent hidden layer. The calculation formula is as follows:

$$O^{(t)} = \sigma \left(W_o \cdot \left[\hat{h}^{(t-1)}, x^{(t)} \right] + b_o \right) \quad (8)$$

$$h^{(t)} = O^{(t)} \otimes \tanh C^{(t)} \quad (9)$$

In Eqs (8) and (9), where $O^{(t)}$ represents the output gate state at moment t ; W_o , b_o represent the weight matrix and bias term of the output gate.

Temperature, humidity, and wind represent the three types of influencing factors in the original load data. In order to mitigate the extent to which the less-correlated wind factor impacts the load, we employ the feature generated by the BILSTM hidden layer as inputs to the simple parameter-free attention mechanism known as SimAM. SimAM automatically assesses the significance of temporal information extracted from the BILSTM hidden layer by comparing it to temporal data extracted from the original load dataset. Deep temporal correlations are discovered by utilizing the inherent time-series properties of the load data to systematically reduce the weights assigned to factors with low correlation with actual loads, and focusing attention on the temperature and humidity factors with higher correlation, thereby reducing the impact of redundant information on the load prediction results. This is achieved by constructing a BILSTM model that integrates the SimAM attention mechanism.

SimAM is a simple parameter-free attention module, which constructs an energy function based on neuroscience theory to identify key neurons. SimAM attention outperforms standard attention mechanisms by offering higher performance with fewer parameters. The SimAM algorithm initially evaluates the relevance of each neuron. In neuroscience, information-rich neurons often display distinct firing patterns compared to surrounding neurons. Moreover, active neurons frequently inhibit adjacent neurons, resulting in null-space inhibition. Therefore, neurons with null-space inhibitory

effects should be given priority. The following is the energy function used to identify key neurons by assessing the linear separability between neurons:

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (10)$$

In Eq (10), where $\hat{t} = w_t t + b_t$, $\hat{x}_i = w_t x_i + b_t$.

Minimizing the above equation is equivalent to training the linear separability between neurons within the unified channel and other neurons. By employing binary labeling and incorporating regularization terms, the final energy function is presented as follows:

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (w_t x_i + b_t))^2 + (1 - (w_t t + b_t))^2 + \lambda w_t^2 \quad (11)$$

Figure 6 illustrates the non-parametric attention mechanism SimAM module. It is evident that the attention mechanism assigns different weights based on the significance of the original data, while also considering the feature of various channels. This enables the model to focus more on the feature attributes that exert the greatest influence on load prediction.

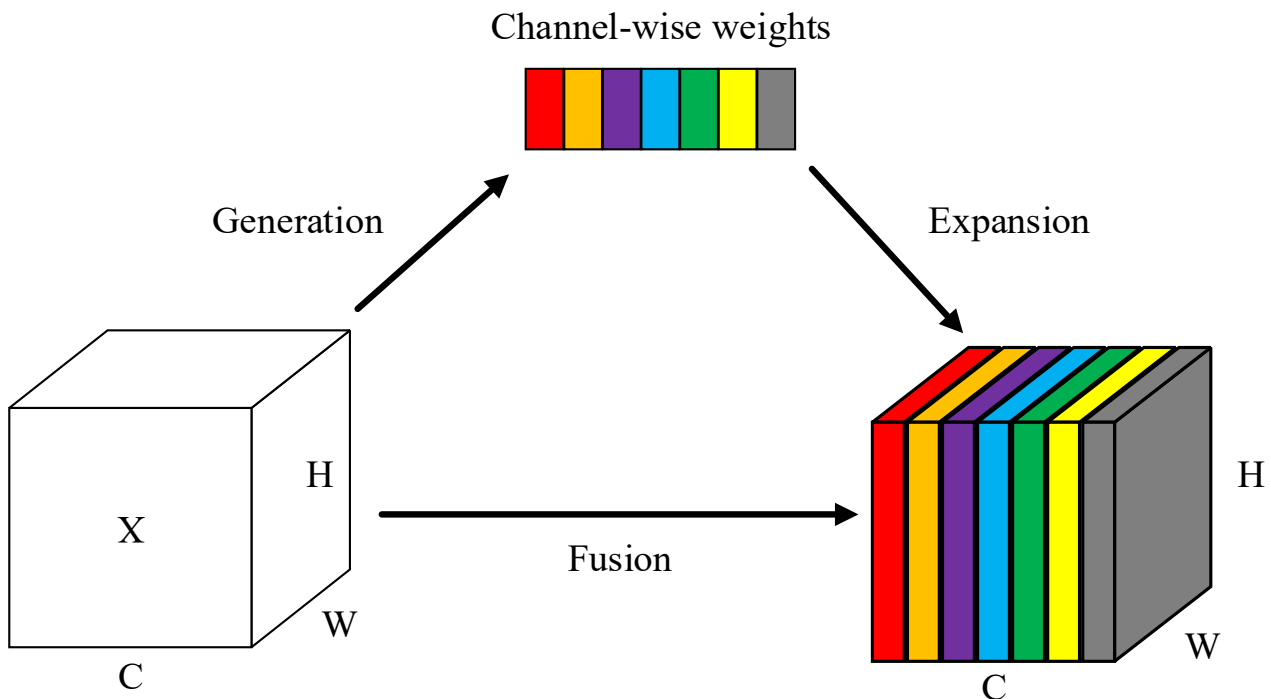


Figure 6. Simple attention mechanism.

3. Results

3.1. Data processing and correlation analysis

The experimental data in this paper comes from the Kaggle public dataset Smart Meters in London, which includes contains the energy consumption readings for a sample of 5567 London Households that took part in the UK Power Networks led Low Carbon London project between November 2011 and February 2014. The weather data is sourced from the official weather station,

collecting various weather metrics at a sampling period of 1 day, these include maximum temperature, dew point temperature, wind speed, pressure, visibility, humidity, moon phase, UV, and cloud cover. Additionally, holiday data is incorporated, accounting for important holidays that significantly impact residents' electricity consumption; these include New Years Day, Good Friday, Easter Monday, May Day Bank Holiday, Spring Bank Holiday, Summer Bank Holiday, Christmas Day, and Boxing Day. The experiment is a short-term load forecast, and the data range is selected from January 1, 2012 to December 31, 2013, with a time interval of one day, and the load forecast time scale is divided as shown in Table 1. Preprocessing operations must be carried out on the dataset before model training can commence, the main steps include outlier handling, screening for correlations, special value handling, data expansion, and normalization.

Table 1. Load forecasting time scale division.

| Load forecasting | Ultra-short-term | Short-term | Mid-term | Long-term |
|------------------|------------------|---------------------|---------------------|--------------------|
| Time scale | Within one hour | One day to one week | January to one year | More than one year |

Outlier handling: The load data may contain missing data, unexpected fluctuations, and other issues throughout the collection or transmission process due to the characteristics of measurement equipment, energy supply restrictions, and other variables. Using raw data directly would introduce too many intervening elements that could impact the prediction results. Therefore, we fill in the missing and aberrant values with the average values of the relevant time points before and after the point, ensuring the correctness and completeness of the data.

Screening for correlations: Included in the dataset are meteorological variables such as maximum temperature, dew point temperature, wind speed, pressure, visibility, humidity, moon phase, UV, and cloud cover. Utilizing the Pearson correlation coefficient, we computed the correlation coefficient matrix between each weather element and the load value. The analysis revealed that the load value is of particular interest due to its strong positive correlation with humidity and a significant negative correlation with temperature. The depth of color in the correlation matrix visually represents the degree of association. Pressure and moon phases exhibited the least correlation with energy consumption, leading to their exclusion from further analysis. While wind speed demonstrates a lower correlation with the load value, it is not significantly correlated with other factors, making it a viable variable for consideration. Conversely, dew point temperature and UV are highly correlated with temperature, leading to their exclusion. Similarly, cloud cover and visibility exhibit multicollinearity with humidity, resulting in their exclusion from further analysis. Figure 7 illustrates the correlation coefficient values obtained from this analysis.

Special value processing: In general, when a regular family has a holiday, the time spent at home is always longer than on a workday, and energy consumption increases. Therefore, the holiday becomes a particular node in the forecast process. Time series prediction algorithms that do not take into account holiday information will produce poor prediction results at these nodes, lowering the model's overall forecast accuracy. Hence, the dates corresponding to holidays in the dataset are explicitly eliminated. Table 2 displays some of the holiday statistics.

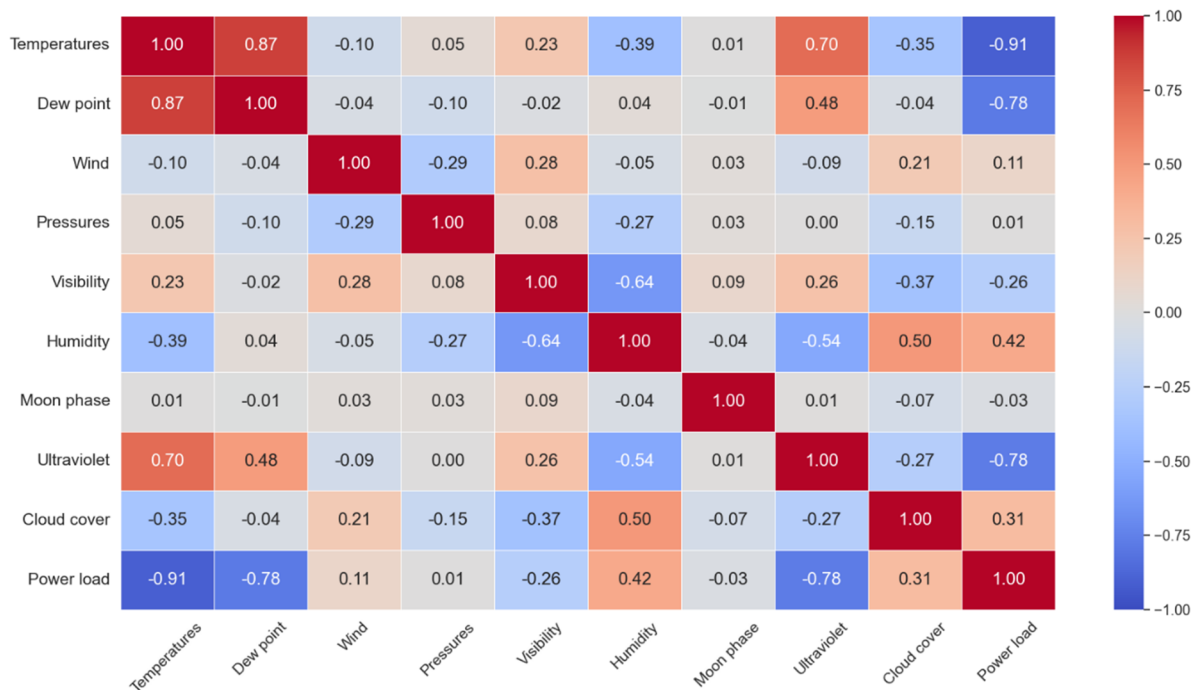


Figure 7. Weather factor correlation coefficient values.

Table 2. Data on selected holidays.

| First year | Holidays | Second year | Holidays |
|------------|----------------------|-------------|----------------------|
| 01/02 | New Years Day | 01/01 | New Years Day |
| 04/06 | Good Friday | 03/29 | Good Friday |
| 04/09 | Easter Monday | 04/01 | Easter Monday |
| 05/07 | May Day Bank Holiday | 05/06 | May Day Bank Holiday |
| 06/04 | Spring Bank Holiday | 05/27 | Spring Bank Holiday |
| 08/27 | Summer Bank Holiday | 08/26 | Summer Bank Holiday |
| 12/25 | Christmas Day | 12/25 | Christmas Day |
| 12/26 | Boxing Day | 12/26 | Boxing Day |

Data expansion: After the preliminary experimental analysis, it was found that relying only on a small amount of cyclical data from January 1, 2012, to December 31, 2013, in the dataset does not yield good prediction results. Therefore, we decided to use generative adversarial networks (GAN) to expand the original data volume by learning the stochastic statistical laws implicit in the original real load data. This approach allows us to generate corresponding period data, ensuring that the model can fully extract the load characteristics and thereby improving the generalization ability of the model. Taking the original data as the benchmark, the experimental data date is expanded to October 10, 2023, increasing the dataset from the original 714 data to 4284 data. The distribution of the original and generated data is shown in Figure 8.

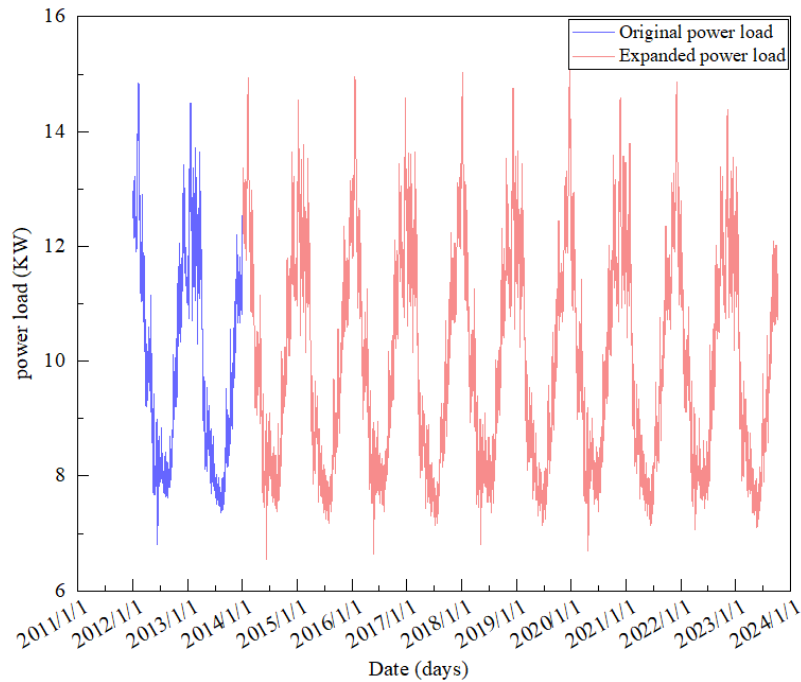


Figure 8. Before and after data expansion.

Processing for normalization: The normalization procedure serves to accelerate the convergence rate of the loss function, prevent gradient explosions, and improve computational accuracy. Normalizing the data is essential because, in real load forecasting, the model's input typically consists of data with varying scales. Removing the influence of different scales on the prediction results enhances both the model's accuracy and efficiency. Data normalization is achieved using the Min-Max method, scaling the data to the range $[0, 1]$, and the calculation formula is as follows:

$$X_i^* = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (12)$$

In Eq (12), X_i represents the original measured data of the i sampling point; X_i^* represents the value after normalization of X_i ; X_{\max} and X_{\min} represent the maximum and minimum values in the measured data.

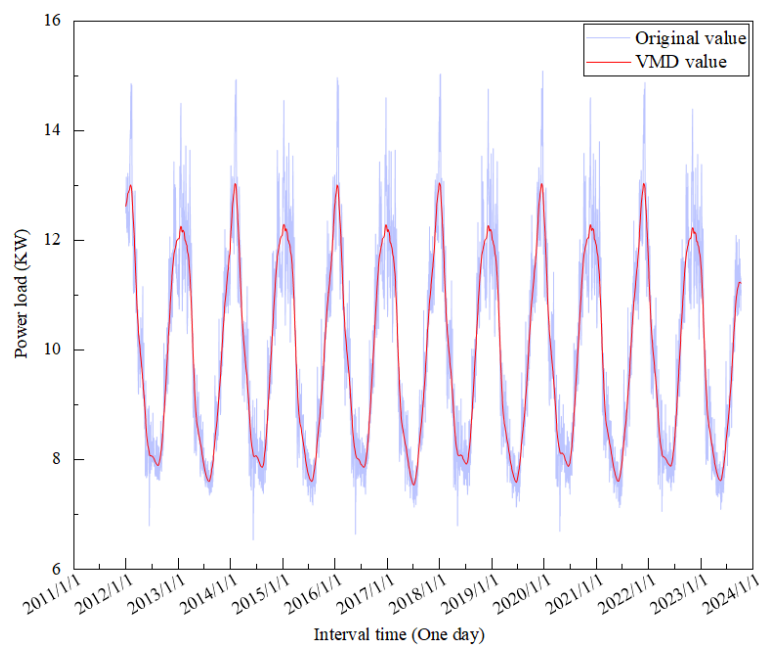
3.2. VMD data preprocessing

When the decomposition layer K value of VMD is too little, data under-decomposition occurs, decreasing prediction accuracy, and when the K value is too big, modal repetition occurs, introducing noise. Therefore, the center frequency method of Optimal Variational Mode Decomposition (OVMD) is used to determine the optimal number of decomposition layers K . The most suitable penalty factor, α , is determined using the Particle Swarm Optimization (PSO) method, while the remaining parameters are set to their default values. Table 3 lists the optimal values for each parameter of VMD.

Table 3. VMD parameter settings.

| VMD | K (Modal number) | α (Bandwidth constraint) |
|--------|--------------------|---------------------------------|
| Values | 5 | 1864 |

For regional electricity loads, meteorological parameters such as temperature and humidity influence both the magnitude and timing of the daily load curve peak, introducing uncertainty into the load data. When we analyze the five modal components obtained through VMD decomposition, IMF1 and IMF2 exhibit relatively low-frequency volatility but demonstrate noticeable periodicity. This periodicity can partially reflect the load data in cyclic patterns and is advantageous for forecasting. On the other hand, IMF3 to IMF5 display high-frequency and relatively abrupt changes, reflecting the stochastic nature of the load data to some extent, and they can also be valuable for forecasting purposes. Among them, the raw data components and the results of each modal decomposition have been demonstrated in Figure 1. However, the residual component, characterized by low-frequency volatility, lacks visible periodicity and can be safely disregarded. The time series load forecast data selected for this paper is displayed in Figure 9 after undergoing VMD.

**Figure 9.** Comparison of data waveforms before and after VMD processing.

3.3. Parameterization and evaluation metrics

Since deep learning inherently includes some randomness, this experiment employs multiple cross-validations to ensure the results are more accurate and reliable. The final experimental results are obtained by averaging the outcomes of 10 load prediction runs, which helps mitigate the impact of any particular outlier experiment. The model takes an input window of 24 days, predicts a single step for 1 day, is trained on data from day 0 to 3800, and predicts the range from day 3800 to 4284. The neural network model for Dropout-CNN is configured with 64 convolutional kernels, 3 convolutional layers, 2 pooling layers, 1 fully connected layer, a Dropout function layer with a parameter of 0.2, and the ReLU activation function. For the BILSTM model, both forward and reverse neuron counts are set

to 12, the optimizer used is RMSprop, the error evaluation employs the mean square error loss function, there are 300 training cycles, and a total of 256 training batches. To ensure fair comparison experiments and evaluate the learning performance of the proposed model, this study maintains consistent model parameters with other comparable models in Table 4.

Table 4. Model parameter settings.

| Name | Parameters | Name | Parameters |
|-------------|------------|---------------------|------------|
| Filters | 64 | Loss | MSE |
| Kernel_size | 3 | Epochs | 300 |
| Pool_size | 2 | Batcha_size | 256 |
| Strides | 1 | BILSTM (forward) | 12 |
| Activation | RELU | BILSTM (backward) | 12 |
| Dropout | 0.2 | BILSTM (12) | 0.01 |
| RMSprop | 0.01 | SimAM (λ) | 0.0001 |

The dataset is partitioned into training, testing, and validation sets with a distribution ratio of 7:2:1. The neural network model undergoes training and validation, followed by predictions using the validation set. Evaluation of the model relies on the coefficient of determination (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) as indices. R^2 , which falls within the range of 0 to 1, quantifies the percentage of variance in the dependent variable that can be attributed to the independent variable. A value closer to 1 signifies a stronger regression fit, indicating a closer relationship between the regression sum of squares and the total sum of squares, a better alignment of the regression line with the data, and a tighter association between the variations in y and x . For this reason, R^2 is often referred to as the ‘‘Goodness of Fit’’ statistic. Furthermore, smaller values of MAE, RMSE, and MAPE suggest reduced errors and improved model accuracy. The following formulas are used for calculation:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (13)$$

$$e_{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (14)$$

$$e_{RMSE} = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

$$e_{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (16)$$

In Eqs (13)–(16), where y_i represents the true value, \hat{y}_i represents the predicted value, where \bar{y} represents the sample mean; n represents the total number of test samples.

3.4. Ablation experiment

In order to verify the superiority of the BILSTM-SimAM algorithm proposed in load forecasting, the test set sampling points used included 3800 to 4284 for a total of 484 days, this paper adopts the ablation experimental method model with variables including:

- Original BILSTM: BILSTM;
- BILSTM-1: CNN+BILSTM+Attention;
- BILSTM-2: Dropout-CNN+BILSTM+Attention;
- BILSTM-3: VMD+Dropout-CNN+BILSTM+Attention;
- BILSTM-SimAM: VMD+Dropout-CNN+BILSTM+SimAM.

Table 5. Comparison of evaluation indexes of ablation experimental.

| Method | R^2 | MAE | RMSE | MAPE |
|--|--------------|--------------|--------------|--------------|
| Original BILSTM (BILSTM) | 0.927 | 0.393 | 0.469 | 0.042 |
| BILSTM-1 (CNN+BILSTM+Attention) | 0.935 | 0.355 | 0.441 | 0.029 |
| BILSTM-2 (Dropout-CNN+BILSTM+Attention) | 0.948 | 0.332 | 0.329 | 0.021 |
| BILSTM-3 (VMD+Dropout-CNN+BILSTM+Attention) | 0.965 | 0.243 | 0.302 | 0.018 |
| BILSTM-SimAM (VMD+Dropout-CNN+BILSTM+SimAM) | 0.978 | 0.192 | 0.245 | 0.016 |

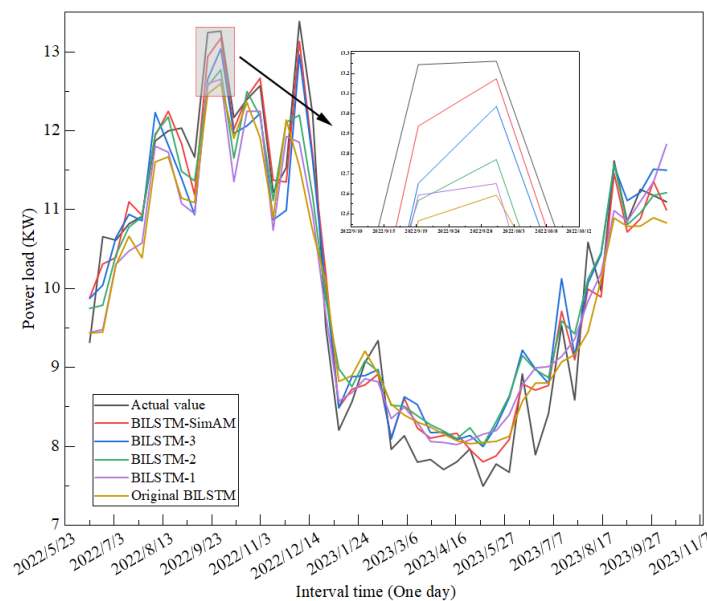


Figure 10. Decomposition predicted effect of ablation experiment.

The R^2 coefficient of the original BILSTM model is 92.7%. BILSTM-1 incorporates the traditional CNN model and the Attention mechanism, resulting in an improved R^2 of 93.5%. BILSTM-2 introduces the Dropout function to the CNN model, leading to an R^2 of 94.8% and highlighting the positive impact of Dropout on network accuracy. BILSTM-3 achieves an R^2 of 96.5% by utilizing VMD for modal decomposition of the original data. BILSTM-SimAM enhances the traditional Attention mechanism with SimAM, combining VMD, Dropout-CNN, and SimAM, which significantly improves accuracy, resulting in an R^2 of 97.8%, a 5.1% increase over the original BILSTM. Table 5 shows the results of the ablation experiment.

To provide a clearer demonstration of the improved model is efficiency, we tested it on the test dataset. The results, depicted in Figure 10, show that the prediction curve generated by the BILSTM-SimAM algorithm closely aligns with the actual value curve. This verification confirms the effectiveness and superiority of the algorithm proposed.

3.5. Comparison experiment

Figure 11 illustrates that the prediction curve generated by our proposed short-term load forecasting algorithm closely resembles the actual value curve. This demonstrates the superior prediction performance of the BILSTM-SimAM algorithm when compared to single models such as CNN-LSTM, CNN-DBILSTM-Attention, Prophet, MLP, and Transformer.

Table 6. Comparison of evaluation indexes of contrasting experimental.

| Method | R ² | MAE | RMSE | MAPE |
|-----------------------|----------------|--------------|--------------|--------------|
| CNN-LSTM | 0.923 | 0.385 | 0.467 | 0.046 |
| CNN-DBILSTM-Attention | 0.935 | 0.351 | 0.435 | 0.032 |
| Prophet | 0.942 | 0.336 | 0.412 | 0.028 |
| MLP | 0.951 | 0.298 | 0.378 | 0.024 |
| Transformer | 0.958 | 0.256 | 0.336 | 0.023 |
| BILSTM-SimAM | 0.978 | 0.192 | 0.245 | 0.016 |

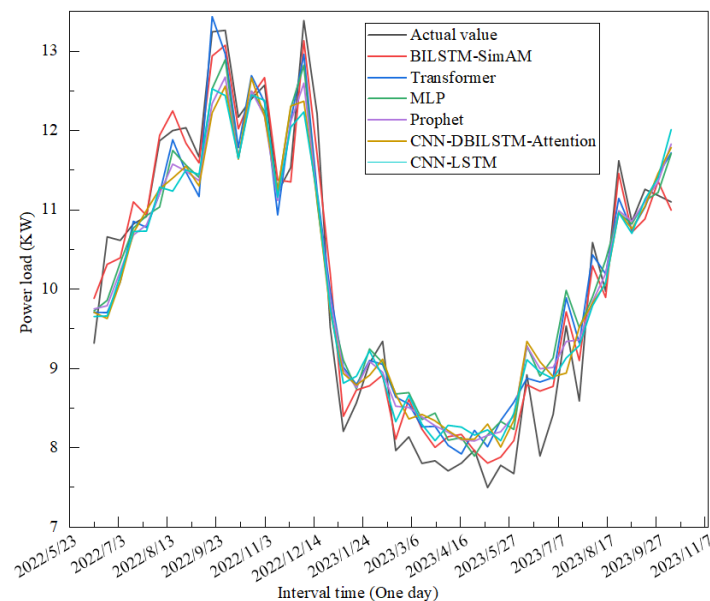


Figure 11. Comparison of experimental model prediction effects.

The fitting impact of the composite neural network model put forward in comparison to the Transformer, MLP, Prophet, CNN-DBILSTM-Attention, and CNN-LSTM models is then verified by a set of comparative tests. Table 6 displays the prediction results, comparing the predicted values to the actual values for each model. Compared with the baseline algorithm, the model presented demonstrates significant improvements: I average R² increases by 2.0%, 2.7%, 3.6%, 4.3%, and 5.5%, respectively; the average MAE decreases by 6.4%, 10.6%, 14.4%, 15.9%, and 19.3%; the average RMSE decreases

by 9.1%, 13.3%, 16.7%, 19.0%, and 22.2%; and the average MAPE decreases by 0.7%, 0.8%, 1.2%, 1.6%, and 3.0%.

The improved algorithm BILSTM-SimAM proposed shows a significant improvement in R^2 and a considerable decrease in MAE, RMSE, and MAPE. These results indicate that the model prediction performance is significantly enhanced, confirming the validity of the BILSTM-SimAM algorithm for short-term electric electricity load forecasting.

4. Discussion

In conclusion, our proposed algorithm takes into account multiple features influencing load changes on the customer side and demonstrates promising results in short-term electricity load forecasting. Notably, no other mainstream algorithms have achieved comparable accuracy based on the same experimental setup. Due to limitations in the dataset, we were unable to explore holiday load forecasting or study the influence of economic factors on household electricity load. In our future research, we plan to incorporate electricity prices as one of the features affecting load forecasting and investigate holiday load forecasting to achieve more accurate predictions. Additionally, the algorithms in this paper will exhibit excellent applicability in the context of a smart city energy internet.

5. Conclusions

In this paper, we propose an improved short-term forecasting algorithm for electric load based on multiple features. To address the volatility and randomness of the time load sequence, the original load sequence is first divided into IMF modal components with smoother features using VMD. Each IMF modal component is then input into the improved CNN network, which utilizes the stochastic deactivation function for multi-feature extraction. Finally, the feature change patterns are learned using the BILSTM model, incorporating the simple parameter-free attention mechanism. The validation and analysis of real load datasets demonstrate that the BILSTM-SimAM combined network proposed in this paper outperforms the BILSTM single model. It results in reductions of 20.1%, 22.2%, and 2.6% in the MAE, RMSE, and MAPE error evaluation indices, respectively, and an increase of 5.1% in the R^2 coefficient of determination. This confirms the validity and practicality of the model presented in this article, which can serve as a guide for load forecasting on the energy consumption side of the intelligent integrated energy internet systems, and provides a basis for the construction of smart cities.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded by the Natural Science Foundation of Sichuan, grant number 2023NSFSC1987, supported by the Opening Fund of Artificial Intelligence Key Laboratory of Sichuan Province (2023RYY07), the Opening Fund of Power Internet of Things Key Laboratory of Sichuan Province (PIT-F-202303), and the Sichuan University of Science and Engineering Postgraduate Innovation Fund Project, grant number Y2023302.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. I. S. Jahan, V. Snasel, S. Misak, Intelligent systems for power load forecasting: A study review, *Energies*, **13** (2020). <https://doi.org/10.3390/en13226105>
2. A. K. Singh, S. Khatoun, M. Muazzam, D. Chaturvedi, Load forecasting techniques and methodologies: A review, in *2012 2nd International Conference on Power, Control and Embedded Systems*, (2012), 1–10. <https://doi.org/10.1109/ICPCES.2012.6508132>
3. J. Zhu, H. Dong, W. Zheng, S. Li, Y. Huang, L. Xi, Review and prospect of data-driven techniques for load forecasting in integrated energy systems, *Appl. Energy*, **321** (2022). <https://doi.org/10.1016/j.apenergy.2022.119269>
4. N. Ahmad, Y. Ghadi, M. Adnan, M. Ali, Load forecasting techniques for power system: Research challenges and survey, *IEEE Access*, **10** (2022) 71054–71090. <https://doi.org/10.1109/access.2022.3187839>
5. R. Jiao, S. Wang, T. Zhang, H. Lu, H. He, B. B. Gupta, Adaptive feature selection and construction for day-ahead load forecasting use deep learning method, *IEEE Trans. Netw. Serv. Manage.*, **18** (2021), 4019–4029. <https://doi.org/10.1109/tnsm.2021.3110577>
6. H. L. Willis, J. E. Northcote-Green, Spatial electric load forecasting: A tutorial review, *Proc. IEEE*, **71** (1983), 232–253. <https://doi.org/10.1109/tnsm.2021.3110577>
7. V. Azarova, D. Engel, C. Ferner, A. Kollmann, J. Reichl, Exploring the impact of network tariffs on household electricity expenditures using load profiles and socio-economic characteristics, *Nat. Energy*, **3** (2018), 317–325. <https://doi.org/10.1038/s41560-018-0105-4>
8. A. Ghasemi, H. Shayeghi, M. Moradzadeh, M. Nooshyar, A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand-side management, *Appl. Energy*, **177** (2016), 40–59. <https://doi.org/10.1016/j.apenergy.2016.05.083>
9. F. Ziel, Modeling public holidays in load forecasting: A German case study, *J. Mod. Power Syst. Clean Energy*, **6** (2018), 191–207. <https://doi.org/10.1007/s40565-018-0385-5>
10. F. M. Butt, L. Hussain, A. Mahmood, K. J. Lone, Artificial Intelligence based accurately load forecasting system to forecast short and medium-term load demands, *Math. Biosci. Eng.*, **18** (2020), 400–425. <https://doi.org/10.3934/mbe.2021022>
11. S. R. Khuntia, J. L. Rueda, M. A. van Der Meijden, Forecasting the load of electrical power systems in mid- and long-term horizons: A review, *IET Gener. Transm. Distrib.*, **10** (2016), 3971–3977. <https://doi.org/10.1049/iet-gtd.2016.0340>
12. M. T. Hagan, S. M. Behr, The time series approach to short term load forecasting, *IEEE Trans. Power Syst.*, **2** (1987), 785–791. <https://doi.org/10.1109/TPWRS.1987.4335210>
13. T. Hong, and P. Wang, Fuzzy interaction regression for short term load forecasting. *Fuzzy Optimization and Decision Making*, **13** (2013) 91-103. <https://doi.org/10.1007/s10700-013-9166-9>
14. H. M. Al-Hamadi, S. A. Soliman, Fuzzy short-term electric load forecasting using Kalman filter, *IEE Proc. Gener. Transm. Distrib.*, **153** (2006), 217–227. <https://doi.org/10.1049/ip-gtd:20050088>

15. J. W. Taylor, Short-term electricity demand forecasting using double seasonal exponential smoothing, *J. Oper. Res. Soc.*, **54** (2017), 799–805. <https://doi.org/10.1057/palgrave.jors.2601589>
16. J. W. Taylor, R. Buizza, Neural network load forecasting with weather ensemble predictions, *IEEE Trans. Power Syst.*, **17** (2002), 626–632. <https://doi.org/10.1109/TPWRS.2002.800906>
17. W. Sulandari, S. Subanar, M. H. Lee, P. C. Rodrigues, Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks, *Energy*, **190** (2020). <https://doi.org/10.1016/j.energy.2019.116408>
18. V. Vlahović, I. Vujošević, Long-term forecasting: A critical review of direct-trend extrapolation methods, *Int. J. Electr. Power Energy Syst.*, **9** (1987), 2–8. [https://doi.org/10.1016/0142-0615\(87\)90019-6](https://doi.org/10.1016/0142-0615(87)90019-6)
19. M. Lekshmi, K. A. Subramanya, Short-term load forecasting of 400 kV grid substation using R-tool and study of influence of ambient temperature on the forecasted load, in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, (2019), 1–5. <https://doi.org/10.1109/ICACCP.2019.8883005>
20. M. Mohandes, Support vector machines for short-term electrical load forecasting, *Int. J. Energy Res.*, **26** (2002), 335–345. <https://doi.org/10.1002/er.787>
21. Y. Dong, X. Ma, T. Fu, Electrical load forecasting: A deep learning approach based on K-nearest neighbors, *Appl. Soft Comput.*, **99** (2021). <https://doi.org/10.1016/j.asoc.2020.106900>
22. Z. Xie, R. Wang, Z. Wu, T. Liu, Short-term power load forecasting model based on fuzzy neural network using improved decision tree, in *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, (2019), 482–486. <https://doi.org/10.1109/iSPEC48194.2019.8975070>
23. G. Dudek, Short-term load forecasting using random forests, in *Intelligent Systems '2014*, Springer, (2015), 821–828. https://doi.org/10.1007/978-3-319-11310-47_1
24. K. B. Lindberg, P. Seljom, H. Madsen, D. Fischer, M. Korpås, Long-term electricity load forecasting: Current and future trends, *Util. Policy*, **58** (2019), 102–119. <https://doi.org/10.1016/j.jup.2019.04.001>
25. Z. A. Khan, A. Ullah, I. Ul Haq, M. Hamdy, G. M. Mauro, K. Muhammad, et al., Efficient short-term electricity load forecasting for effective energy management, *Sustainable Energy Technol. Assess.*, **53** (2022). <https://doi.org/10.1016/j.seta.2022.102337>
26. X. Sun, Z. Ouyang, D. Yue, Short-term load forecasting based on multivariate linear regression, in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, (2017), 1–5. <https://doi.org/10.1109/EI2.2017.8245401>
27. X. Dong, S. Deng, D. Wang, A short-term power load forecasting method based on k-means and SVM, *J. Ambient Intell. Hum. Comput.*, **13** (2021), 5253–5267. <https://doi.org/10.1007/s12652-021-03444-x>
28. S. Fallah, M. Ganjkhani, S. Shamshirband, K. Chau, Computational intelligence on short-term load forecasting: A methodological overview, *Energies*, **12** (2019). <https://doi.org/10.3390/en12030393>
29. A. Heydari, M. M. Nezhad, E. Pirshayan, D. A. Garcia, F. Keynia, L. De Santoli, Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm, *Appl. Energy*, **277** (2020). <https://doi.org/10.1016/j.apenergy.2020.115503>

30. M. Chen, Z. Lan, Z. Duan, S. Yi, Q. Su, HDS-YOLOv5: An improved safety harness hook detection algorithm based on YOLOv5s, *Math. Biosci. Eng.*, **20** (2023), 15476–15495. <https://doi.org/10.3934/mbe.2023691>
31. W. Zeng, J. Li, C. Sun, L. Cao, X. Tang, S. Shu, et al., Ultra short-term power load forecasting based on similar day clustering and ensemble empirical mode decomposition, *Energies*, **16** (2023). <https://doi.org/10.3390/en16041989>
32. X. Yan, M. Jia, Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings, *Mech. Syst. Signal Process.*, **122** (2019), 56–86. <https://doi.org/10.1016/j.ymsp.2018.12.022>
33. J. Chen, J. Zhang, A dual attention-based CNN-GRU model for short-term electric load forecasting, in *The Proceedings of the 10th Frontier Academic Forum of Electrical Engineering (FAFEE2022)*, Springer, (2023), 715–725. https://doi.org/10.1007/978-981-99-3404-1_63
34. A. Wan, Q. Chang, K. Al-Bukhaiti, J. He, Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism, *Energy*, **282** (2023). <https://doi.org/10.1016/j.energy.2023.128274>
35. Q. Chen, W. Zhang, K. Zhu, D. Zhou, H. Dai, Q. Wu, A novel trilinear deep residual network with self-adaptive Dropout method for short-term load forecasting, *Expert Syst. Appl.*, **182** (2021). <https://doi.org/10.1016/j.eswa.2021.115272>
36. X. Ji, D. Liu, P. Xiong, Multi-model fusion short-term power load forecasting based on improved WOA optimization, *Math. Biosci. Eng.*, **19** (2022), 13399–13420. <https://doi.org/10.3934/mbe.2022627>
37. Z. Yao, T. Zhang, Q. Wang, Y. Zhao, R. Wang, Short-term power load forecasting of integrated energy system based on attention-CNN-DBILSTM, *Math. Probl. Eng.*, **2022** (2022), 1–12. <https://doi.org/10.1155/2022/1075698>
38. K. Dragomiretskiy, D. Zosso, Variational Mode Decomposition, *IEEE Trans. Signal Process.*, **62** (2014), 531–544. <https://doi.org/10.1109/tsp.2013.2288675>
39. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.*, **15** (2014), 1929–1958.
40. Z. Huang, W. Xu, K. Yu, Bidirectional LSTM-CRF models for sequence tagging, *arXiv preprint*, (2015), arXiv:1508.01991. <https://doi.org/10.48550/arXiv.1508.01991>
41. L. Yang, R. Y. Zhang, L. Li, X. Xie, In SimAM: A simple, parameter-free attention module for convolutional neural networks, in *International Conference on Machine Learning*, PMLR, (2021), 11863–11874. <https://icml.cc/virtual/2021/spotlight/8922>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)