

Research Article

A Novel Short-Term Power Load Forecasting Method Based on TSNE-EEMD-LSTM

Mingkun Jiang,^{1,2} He Jiang ,^{1,2} Yan Zhao ,^{1,2} Chenjia Hu,^{1,2} and Jian Xu^{1,2}

¹School of Renewable Energy, Shenyang Institute of Engineering, Shenyang 110136, China

²Key Laboratory of Regional Multi-energy System Integration and Control of Liaoning Province, Shenyang 110136, China

Correspondence should be addressed to He Jiang; jianghescholar@163.com

Received 29 March 2022; Revised 18 May 2022; Accepted 24 May 2022; Published 6 July 2022

Academic Editor: Qiuye Sun

Copyright © 2022 Mingkun Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, a novel power load forecasting model is proposed to fully extract the periodic characteristics of short-term load at various time scales and explore the potential correlations between influencing factors and characteristics of load components. Firstly, the t-distributed stochastic neighbor embedding algorithm is used to map sample points of high-dimensional load influencing factors to low-dimensional space, and the ensemble empirical mode decomposition algorithm is employed to split the historical load curve into multiple signal components with different frequencies. Then, several long short-term memory networks including nonlinear mapping and time series models are established to mine the relationship between low-dimensional comprehensive influencing factors and each intrinsic mode function component by utilizing different inputs. Finally, the effectiveness of the hybrid model is verified via using the short-term load dataset of 3-hour data granularity in a certain region, and the influence of key parameters of the model on the forecasting effect is discussed.

1. Introduction

Load forecasting in short term is one of the important daily tasks in the contemporary power system. The unit output plan and economic dispatch strategy based on good forecasting results can improve the operation stability and economy of the system to a certain extent. Because the short-term load is influenced by succession uncertain factors, the model can be expressed as the sum of nonlinear mapping relationships between short-term load and related influencing factors and a series of uncertain random loads.

Up to now, there have been some state-of-the-art works to explore the ways to enhance the accuracy of load forecasting from the aspect of uncertain random loads. Reference [1] used user behavior to reflect the fluctuation of some uncertain loads, used smart meters to extract user-level data, analyzed the similarity of user behavior, and introduced them into the forecasting model. Reference [2] proposed a multistep forecasting model containing three channels (load, time, and user behavior). The user behavior type was identified by combining convolution automatic encoder and

k-means, and it was used as a feature to combine with the feature information obtained by the other two channels for the comprehensive forecasting of short-term power load. Most of the existing user behavior feature extraction starts from the similarity of user energy consumption curve, classifies and forecasts users. However, the user's energy consumption law itself has strong randomness, and the error generated by the classification will be transmitted to the final forecasting results through the forecasting model, resulting in the instability of the model in practical application.

At present, mainstream load forecasting research still focuses on how to better and more comprehensively excavate the nonlinear mapping relationship between load and its influencing factors. This kind of research mainly adopts machine learning modeling methods [3–5]. Reference [6] used the Bat algorithm and Kalman filtering method to optimize the support vector machine (SVM) and combined the fuzzy combination weight and empirical mode decomposition to forecast the load in short term. Reference [7] established a combination forecasting model including SVM and generalized regression model and used the weight

determination theory to determine the final forecasting results. In Reference [8], the daily load curve was clustered, and a step-wise compound forecasting model was constructed based on a convolutional neural network (CNN) and long short-term memory (LSTM) neural network [8]. Reference [9] utilized a CNN to construct a deep learning model and carried out load forecasting for users with similar energy consumption. Under the comprehensive influence of many influencing factors, short-term power load still has certain periodic characteristics, and the aforementioned research methods to extract the periodic characteristics of load by influencing factors inevitably lead to incomplete feature extraction.

To solve this problem, empirical mode decomposition (EMD) [10–12] was introduced into load forecasting models. EMD can decompose the complex load curve into multiple intrinsic mode functions (IMFs) [10] with various frequencies and amplitudes and sole residual signal. In such studies, each sub-signal was often used for multiple modeling via the same forecasting model, such as the state-dependent autoregressive model [10] and CNN-LSTM model [11], ignoring the characteristics of each sub-signal. Appropriate and different ways of load modeling according to the characteristics of different IMFs can undoubtedly better and more comprehensively extract the load characteristics.

In this paper, a novel hybrid model for load forecasting in short term is proposed on the basis of previous studies. First, with the help of the time series memory capacity, the LSTM [13] model is divided into two categories by different model inputs, and different IMFs are modeled in an appropriate way to excavate the potential information of various IMFs. Second, the ensemble empirical mode decomposition (EEMD) [14] is applied to split the load signal, so as to overcome the mode mixing phenomenon that is easy to occur in the EMD and avoid its adverse impact on the forecasting effect. The nondestructive dimension reduction of power load influencing factors is carried out by t-distributed stochastic neighbor embedding (TSNE) [15], which reduces the amount of model calculation and the redundant features extracted from the model. Third, several LSTM models in two types that have different inputs are established by using the data obtained by TSNE and EEMD. This study introduces the algorithms used in the proposed hybrid model, and gradually elaborates the structure and construction process of the model. Finally, two key parameters of the model are discussed through examples and experiments, and the forecasting performance of the model is verified. As a result, the main contributions and highlights can be summarized as follows:

- (1) The EEMD algorithm is employed to decompose the power load curve into load components with different time-scale characteristics, and the appropriate modeling is carried out according to the characteristics of different components, which provides a new application form for the time series decomposition of load.
- (2) The nondestructive dimension reduction of power load influencing factors is realized through TSNE,

which solves the problems of long training time of hybrid model composed of multiple sub-models.

- (3) This paper provides a new short-term load forecasting scheme, which can also be applied to the forecasting fields of new energy generation and integrated loads.

2. Data Processing and Basic Forecasting Model

2.1. Decomposition of Load Curve Based on EEMD. EMD and EEMD belong to nonlinear signal decomposition algorithms, which decompose a nonstationary time series signal into several groups of IMFs with frequency from high to low and a group of residual signals representing the overall trend of the original signal. For the power load signal, the IMFs components of different frequencies contain the periodic characteristics of power load at different time scales. With the decrease of IMFs component frequency, the low-frequency component represented by the residual signal contains the load trend of power load for a time.

Due to this phenomenon of missing data and abnormal data in the acquisition process of power load data, this phenomenon may lead to the discontinuity of the original load signal or the existence of signal step-change on the time scale, which makes such data prone to modal aliasing when using EMD for signal decomposition. Modal aliasing refers to the characteristic signal that an IMF component contains different time scales at the same time [14], which affects the signal decomposition effect and load modeling. As an improved method of EMD, based on EMD, EEMD adds the process of adding white noise to the signal many times and calculating the mean value of each sub-signal, which can avoid the occurrence of modal aliasing.

The steps of EEMD signal splitting are as follows:

- (1) Set the number of adding white noise: M .
- (2) Add normal distribution white noise to the historical load curve to constitute the signal $f_{i,j}$, $i \in [1, M]$ and $j \in [1, N]$, N denotes the number of sub-signals obtained by decomposition after step (5).
- (3) Find out all the local maximum and minimum values contained in the signal, and fit the envelope curves with the maximum and minimum points, respectively.
- (4) The mean curve $g_{i,j}$ constructed by two envelope curves is obtained, and the difference $h_{i,j}$ between the signal $f_{i,j}$ and the curve $g_{i,j}$ is obtained. The mean curve $g_{i,j}$ will participate in the next iteration, and $f_{i,j+1} = g_{i,j}$.
- (5) Repeat steps (3) and (4) until the difference between signal $f_{i,j}$ and curve $g_{i,j}$ is small enough, that is, the signal $f_{i,j}$ cannot be decomposed again. At this time, the quantity of $h_{i,j}$ obtained by decomposition is N .
- (6) Repeat steps (2) to (5) until the original signal is processed M times by white noise and decomposed. At this time, the mean of all $h_{i,j}$ is obtained by calculating the M times signal decomposition,

namely, the signal components obtained by EEMD. The calculation formula is given by

$$I_j = \frac{1}{M} \sum_{i=1}^M h_{i,j}, \quad (1)$$

where I_j is each component of the original load signal obtained by EEMD, and each component satisfies

$$f = \sum_{j=1}^N I_j, \quad (2)$$

where f is the original power load signal, I_1, I_2, \dots, I_{N-1} are the IMF components, and I_N is the residual.

2.2. Dimension Reduction of Influencing Factors Based on TSNE. TSNE algorithm is improved from stochastic neighbor embedding (SNE). SNE deems that the distance between corresponding points in original dimension and converted dimension is also similar, and conditional probability is used to represent the similarity of this distance relationship [15].

In SNE algorithm, for the data points X_α and X_β of the influencing factors of power load in any high-dimensional space, the probability of X_β as the proximity point of X_α is set to be $P_{\beta|\alpha}$. After mapping to the low-dimensional space, the probability of low-dimensional mapping point x_β as the proximity point of x_α is set to be $Q_{\beta|\alpha}$, and the calculation formulas of $P_{\beta|\alpha}$ and $Q_{\beta|\alpha}$ are expressed as

$$P_{\beta|\alpha} = \frac{\exp\left[-(2\delta_\alpha^2)^{-1}\|X_\alpha - X_\beta\|^2\right]}{\sum_{k \neq \alpha} \exp\left[-(2\delta_\alpha^2)^{-1}\|X_\alpha - X_k\|^2\right]}, \quad (3)$$

$$Q_{\beta|\alpha} = \frac{\exp\left[-\|x_\alpha - x_\beta\|^2\right]}{\sum_{k \neq \alpha} \exp\left[-\|x_\alpha - x_k\|^2\right]}, \quad (4)$$

where δ_α is the standard deviation of Gaussian distribution with X_α as the center point, X_k and x_k are the arbitrary initial data points in original dimension and mapping points in converted dimension, respectively.

The cost function C is constructed and defined as

$$C = \sum_{\alpha} \sum_{\beta} P_{\beta|\alpha} \log \frac{P_{\beta|\alpha}}{Q_{\beta|\alpha}}. \quad (5)$$

At this time, the data dimension reduction problem is transformed into an optimal problem. Through iterative calculation, when the cost function C takes the minimum value, the optimal solution of the mapping point of original dimension samples in converted dimension can be obtained. The iterative process is given by

$$x^r = x^{r-1} + \eta \frac{\partial C}{\partial x} + \varepsilon(r)(x^{r-1} - x^{r-2}), \quad (6)$$

where x^r represents the sample points in lower dimension obtained in iteration r , r denotes the number of iterations, η is the learning rate, $\varepsilon(r)$ is the learning momentum in iteration r , $\partial C/\partial x$ is the gradient vector, and the gradient $\partial C/\partial x_\alpha$ at point x_α is written as

$$\frac{\partial C}{\partial x_\alpha} = 2 \sum_{\beta} (P_{\beta|\alpha} - Q_{\beta|\alpha} + P_{\alpha|\beta} - Q_{\alpha|\beta})(x_\alpha - x_\beta). \quad (7)$$

In the SNE algorithm, the samples in both the original dimension and converted dimension use Gaussian distribution to represent the similarity between data, which will lead to the inconvenience of data congestion in the converted dimension, which is laborious to distinguish [16], and increase the difficulty of model feature extraction. In addition, the probability of data points calculated by the SNE algorithm in original dimension and converted dimension is asymmetric and the calculation of gradient is complex.

Given these limitations of the aforementioned SNE algorithm, the TSNE algorithm has made the following improvements based on SNE:

- (1) In the TSNE algorithm, the symmetric SNE method is applied to calculate the probability of samples in original dimension and simplify the calculation of gradient formula. The probability of samples $P_{\beta|\alpha}$ and the joint probability density of data points $P_{\alpha\beta}$ in original dimension can be expressed as

$$P_{\beta|\alpha} = \frac{\exp\left[-(2\delta_\alpha^2)^{-1}\|X_\alpha - X_\beta\|^2\right]}{\sum_{k \neq \alpha} \exp\left[-(2\delta_\alpha^2)^{-1}\|X_\alpha - X_k\|^2\right]}, \quad (8)$$

$$P_{\alpha\beta} = \frac{P_{\alpha|\beta} + P_{\beta|\alpha}}{2n}, \quad (9)$$

where n is the number of data points of high-dimensional space load influencing factors.

- (2) In SNE, Gaussian distribution is used to represent the similarity between sample points in space with different dimensions, while TSNE improves the Gaussian distribution in converted dimension to t-distribution, which reduces the inconvenience of data crowded and difficulty in distinguishing after dimension reduction. So far, the joint probability density of low-dimensional space samples $Q_{\alpha\beta}$ can be defined as

$$Q_{\alpha\beta} = \frac{\left(1 + \|x_\alpha - x_\beta\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|x_k - x_l\|^2\right)^{-1}}, \quad (10)$$

where x_k and x_l are the mapping points of any low-dimensional space samples, respectively.

TSNE algorithm uses joint probability density to replace conditional probability in the SNE algorithm. For any joint probability density between samples, there are $P_{\alpha\beta} = P_{\beta\alpha}$ and $Q_{\alpha\beta} = Q_{\beta\alpha}$. The improved cost function C and gradient $\partial C/\partial x_\alpha$ can be written as

$$C = \sum_{\alpha} \sum_{\beta} P_{\alpha\beta} \log \frac{P_{\alpha\beta}}{Q_{\alpha\beta}}, \quad (11)$$

$$\frac{\partial C}{\partial x_\alpha} = 4 \sum_{\beta} (P_{\alpha\beta} - Q_{\alpha\beta}) (x_\alpha - x_\beta) \left(1 + \|x_\alpha - x_\beta\|^2\right)^{-1}. \quad (12)$$

Finally, the TSNE algorithm uses (6) to continuously iteratively obtain the optimal solution of mapping points in low-dimensional space.

2.3. LSTM Neurons. LSTM is an improved recurrent neural network (RNN) published by Hochreiter and Schmidhuber [13], which overcomes the difficulty of training RNN in practical applications. LSTM neural network has a strong storage capacity of time series features and has strong adaptability in feature extraction of data with such properties. LSTM unit can be divided into input subunit, output subunit, forgetting subunit, and alternative cell state, and their respective states are expressed as

$$D^t = \tanh(W_D X_{\text{lstm}}^t + U_D H_{\text{lstm}}^{t-1} + b_D), \quad (13)$$

$$I^t = \sigma(W_I X_{\text{lstm}}^t + U_I H_{\text{lstm}}^{t-1} + b_I), \quad (14)$$

$$O^t = \sigma(W_O X_{\text{lstm}}^t + U_O H_{\text{lstm}}^{t-1} + b_O), \quad (15)$$

$$F^t = \sigma(W_F X_{\text{lstm}}^t + U_F H_{\text{lstm}}^{t-1} + b_F) \quad (16)$$

where D^t is the alternative cell state at t moments, \tanh is the hyperbolic tangent activation function, W_D and U_D are the weights assigned to X_{lstm}^t and H_{lstm}^{t-1} in the calculation of the alternative cell state, X_{lstm}^t is the input of this unit at t moments, H_{lstm}^{t-1} is the output of this unit at $t-1$ moments, t represents time, b_D is a bias term for alternative cell state, I^t , O^t and F^t are the gating coefficients of the input subunit, output subunit, and forgetting subunit, σ is the Sigmoid function, W_I , W_O and W_F are the weights assigned to X_{lstm}^t in the three subunits. U_I , U_O and U_F are the weights assigned to H_{lstm}^{t-1} in the three subunits, and b_I , b_O and b_F are the bias terms of the three subunits.

Thus, the weight value and bias value of each door control the size of the corresponding gating coefficient. The gating coefficient of the input subunit is applied to control how much information in the alternative cell state can be written in the updated cell state, and the gating coefficient of the forgetting subunit is utilized to restrict how much information from the previous cell state can be inherited to updated cell state. The state of the cell of this LSTM unit at t moments can be written as

$$C^t = I^t D^t + F^t C^{t-1}, \quad (17)$$

where C^t and C^{t-1} are the states of this cell at t moments and $t-1$ moments, respectively.

Similarly, the function of the LSTM unit output gating coefficient is to control how many cells state information activated by activation function can be output from this unit, so the output of a single LSTM unit is given by

$$H_{\text{lstm}}^t = O^t \times \tanh(C^t), \quad (18)$$

where H_{lstm}^t is the output data of this unit at t moments.

LSTM network is composed of multiple memory units and even multi-layer LSTM unit layers. In the training, the weight value and bias value of all LSTM units are determined by iteration with the minimum loss function as the goal. At this time, the input-output mapping relationship of the whole network can represent the nonlinear connection between influencing ingredients and power load or the time series tendency from the historical load.

3. Hybrid Model Based on TSNE-EEMD-LSTM

According to the aforementioned introduction, the TSNE-EEMD-LSTM model proposed in this paper includes power load decomposition, influence factor dimension reduction, and load forecasting combination modeling. The proposed model structure is shown in Figure 1.

In step 1 of Figure 1, the influence factors dimension reduction based on TSNE algorithm and load signal decomposition based on EEMD algorithm are described, respectively.

There are six types of power load influencing factors in the original dataset, including time, temperature, pressure, humidity, wind speed, and precipitation. These influencing factors together constitute a high-dimensional sample space, and these points in the original dimension are reduced to a lower dimension by TSNE algorithm. It is assumed that the number of the target dimension is S , and the projection of sample points in S -dimensional space on each coordinate system is the value of each comprehensive influencing factor.

When the value of S is 2 or 3, the data after dimension reduction can be visually represented, which is convenient for intuitive analysis of the distribution relationship of points in lower dimension. The visualization effect is shown in Figure 2 and Figure 3. The mapping points in the low-dimensional space still maintain good dispersion after the dimension reduction of the power load influencing factor data by the TSNE algorithm.

According to relevant conclusions in reference [14], some essential parameters in EEMD are set as follows:

- (1) The standard deviation of white noises amplitude added to the original data is 0.2 times that of them.
- (2) The number of adding white noise M is set to 200.

The results of signal decomposition of power load application EEMD are shown in Figure 4.

Each IMF component has different frequencies and characteristics. The IMF1 component has the largest

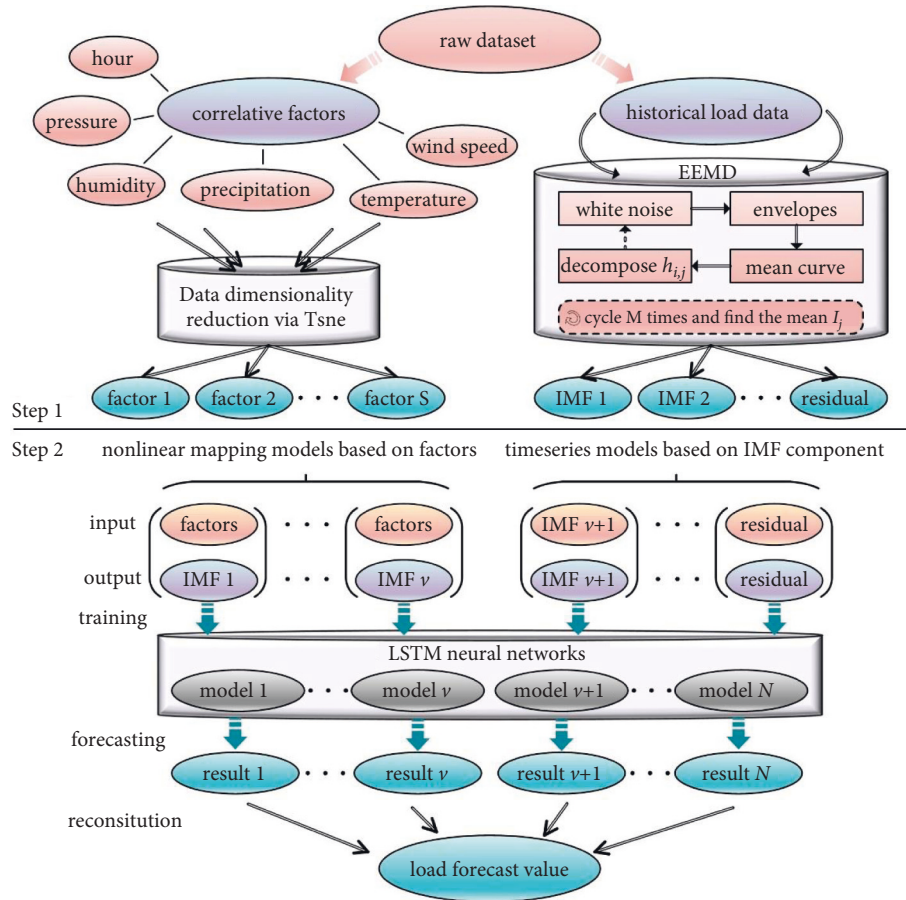


FIGURE 1: TSNE-EEMD-LSTM model.

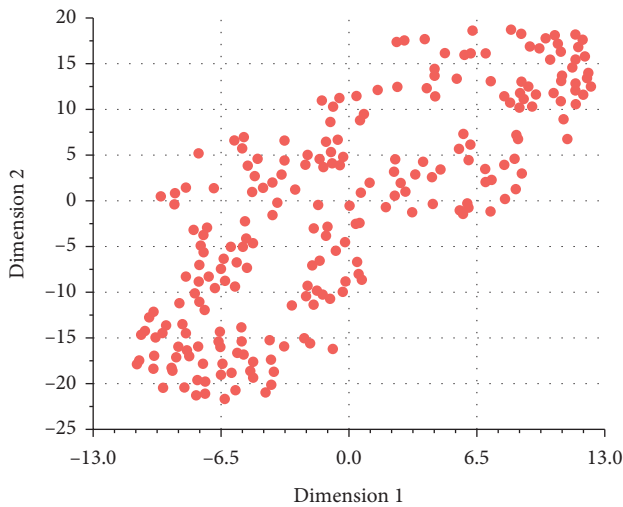


FIGURE 2: 2D spatial mapping points.

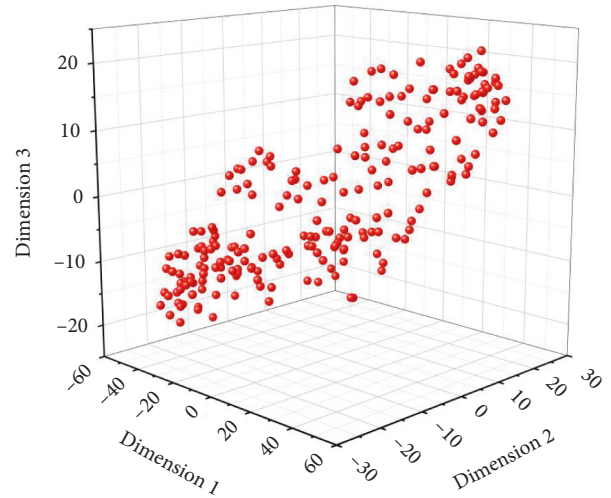


FIGURE 3: 3D spatial mapping points.

variation frequency and changes dramatically with the change of each sample number in the original data, which contains the load characteristics of the minimum particle size. The period of the IMF2 component is almost the same as the original signal, and the change of each extreme point has a similar trend with the peak and valley value of the original signal, which indicates that the IMF2 component

may contain characteristic information related to the peak and valley value of daily load. The frequency of the IMF3 component to the IMF6 component decreases gradually; these signals may contain several days, a week, or longer load periodic characteristics.

In step 2 of Figure 1, N forecasting models based on LSTM are established. Figure 4 shows that the original load

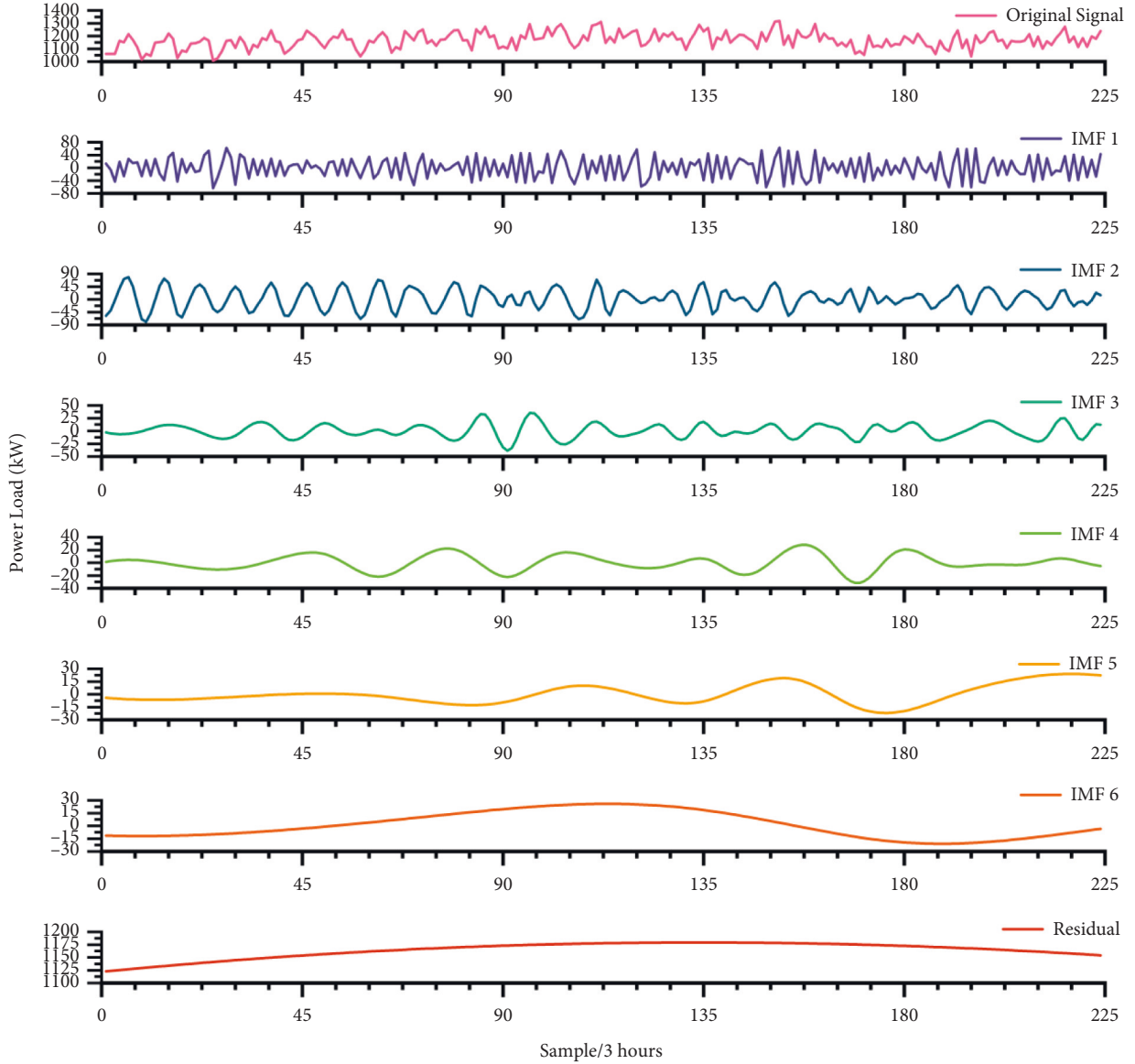


FIGURE 4: Original load signal and each IMF component.

signal is decomposed into six IMF components and one residual signal, so $N = 7$. The established seven models can be divided into two categories. One is the nonlinear mapping model based on the influencing factors, which is applied to explore the nonlinear connection between the influencing ingredients and corresponding load component data. The other is the time series model to explore the time series trend of certain load component curves on the basis of itself.

The nonlinear mapping model takes the comprehensive influencing factors of power load in low-dimensional space as the input of the model, and each IMF component or residual signal as the output of the model. Assuming that the target dimension of the TSNE algorithm is S , the data of comprehensive influencing factors in low-dimensional space is $[A_1, A_2, \dots, A_S]$, and then the output data of LSTM network participating in model supervised learning at τ

moments is y^τ , the corresponding input data of LSTM network is $[A_1^\tau, A_2^\tau, \dots, A_S^\tau]^T$.

Time series model, the input or output of the model is the same IMF component or residual signal sequence. If the output data of a LSTM network participating in model supervised learning at τ moments is y^τ , the corresponding input data of the LSTM network is $[y^{\tau-8}, y^{\tau-7}, \dots, y^{\tau-1}]^T$.

When applying to forecast, the execution process is shown in Figure 5. Trained models include ν nonlinear mapping models and $N - \nu$ time series models. Firstly, TSNE is used to reduce the data dimension of the power load influencing factors needed in the forecasting day, and the data points of the influencing factors in original dimension are mapped to the lower dimension. Then, influencing mapping factors in converted dimension are input into trained ν nonlinear mapping models, and the historical data

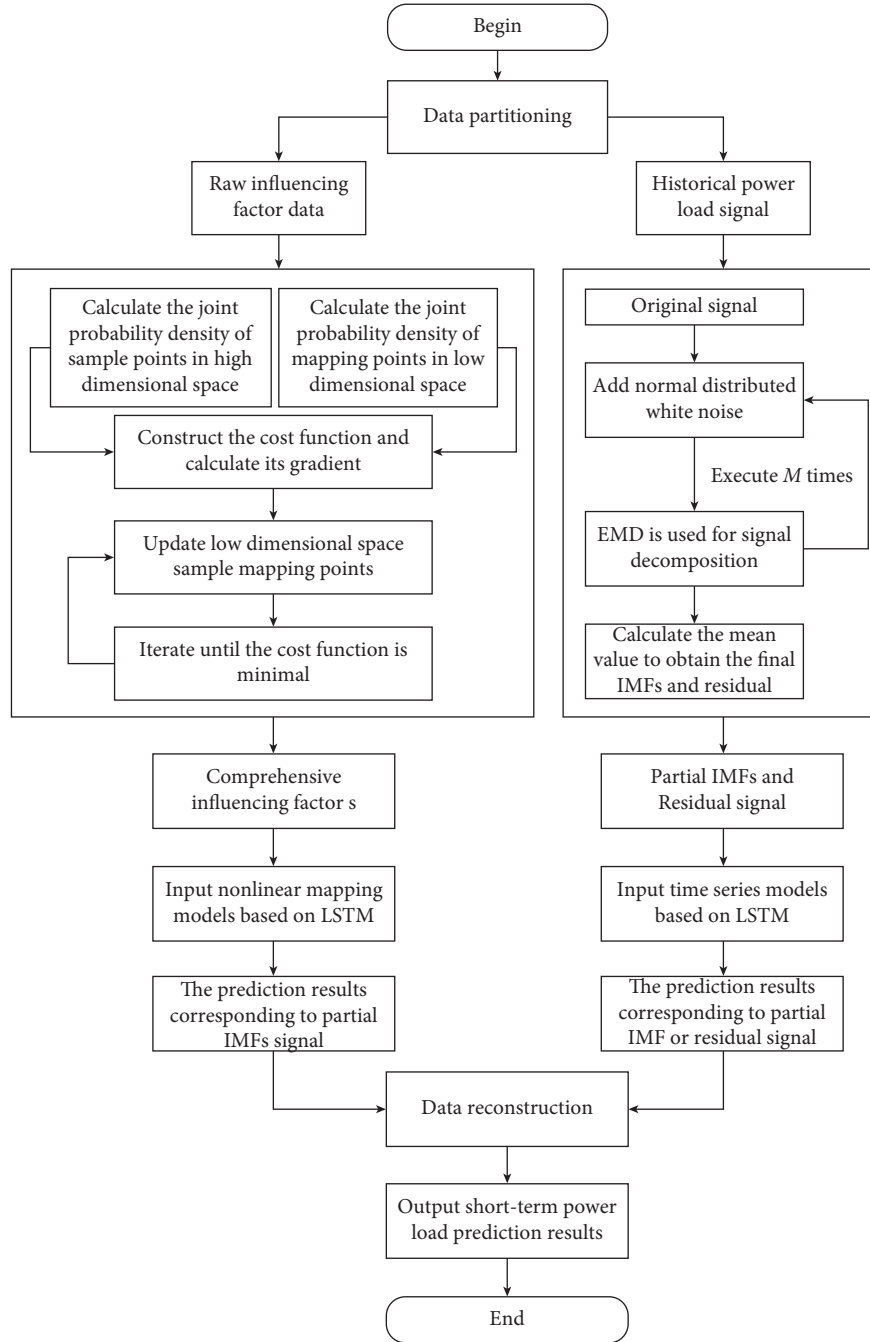


FIGURE 5: Forecasting process of the hybrid model.

of IMF components or residual signals corresponding to the models are input to $N - \nu$ time series models. Finally, a total of N model outputs are obtained, and the N model outputs are reconstructed according to Equation (2) to get the final forecasting results.

4. Experimental Verification and Discussion

In this section, the essential parameters of the model are discussed and the forecasting performance of the combined model is testified by using datasets including the 31-day 3-hour granular load data of a city and the corresponding time

temperature, pressure, and other six kinds of power load influencing factors data.

4.1. Discussion on Key Parameters of Model. Selecting appropriate parameters is advantageous to enhance the forecasting effect of the proposed model. This part will discuss the influence of low EEMD parameter S and the number ν of nonlinear mapping models based on LSTM on the forecasting efficiency of the model.

Different S will affect the calculation and forecasting speed of the nonlinear mapping model. To verify this

hypothesis, six IMFs and one residual signal gained via EEMD are used as the output of supervised learning models to establish seven nonlinear mapping models, and then the average total time required to complete the training and forecasting of these seven models is counted when S takes different values. In Figure 6, there is a tendency that when the input dimension of the model decreases, the total time required to complete the model training and forecasting is shorter. That is, the forecasting speed of the model increases with the decrease of S .

The forecasting speed of the model can be improved by mapping the influencing factors of power load to the low-dimensional space for model building. However, the data gained via TSNE have similar distribution relations in different dimensions. When the target dimension is too low, the projection process of raw data in lower dimension will become difficult, and even the phenomenon of feature loss occurs. This kind of phenomenon is explained as that when the low-dimensional space dimension is too low, the data mapping points cannot carry sufficient feature information.

Therefore, we compare the forecasting accuracy of the combined model with different S and v . Among them, each group of composite models is composed of seven models, namely, nonlinear mapping models or time series models. At this time, the number of nonlinear mapping models is v , and the number of time series models is $7 - v$. Except for different model structures, the parameters of the two models are all the same. In terms of model structure, the nonlinear mapping model consists of a layer of LSTM, which has 60 LSTM units, and the time series model consists of three LSTM layers, each of which has 60 LSTM units. The reason for this distinction is that the input of the time series model is the load information at the historical time, and the data itself has the characteristics of time series. Therefore, it is necessary to extract the relevant characteristics from more layers of the LSTM layer. The input of the nonlinear mapping model is the comprehensive influencing factor data in low-dimensional space, and it does not have a time-series relationship. Therefore, only one layer of the LSTM layer is set to simplify the calculation of the model.

The experimental results are shown in Figure 7. When $S = 4$ and $v = 6$, the combined model has the highest average forecasting accuracy.

Laterally, the combination model with the same S will have the best forecasting performance at $v = 6$, that is, the combination model using all IMF components to build nonlinear mapping model and using residual signal to build time series model has the best effect, indicating that the relatively flat residual signal is more suitable for time series forecasting method. When $v = 0$, the combined model consists of seven time series models, that is, all IMF components and residual signals are modeled by time series models. At this time, the average forecasting accuracy of the combined model is the lowest, indicating that each IMF component is not suitable for time series forecasting.

Vertically, the models with higher accuracy are concentrated in the middle part of Figure 7, indicating that too large or too small S will influence the efficiency of the proposed model. Part of the reason is that there is a certain

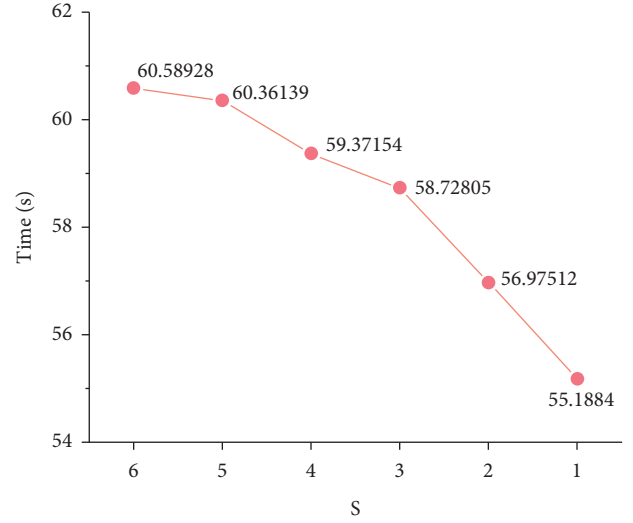


FIGURE 6: Total time to complete all model training and forecasting at different S .

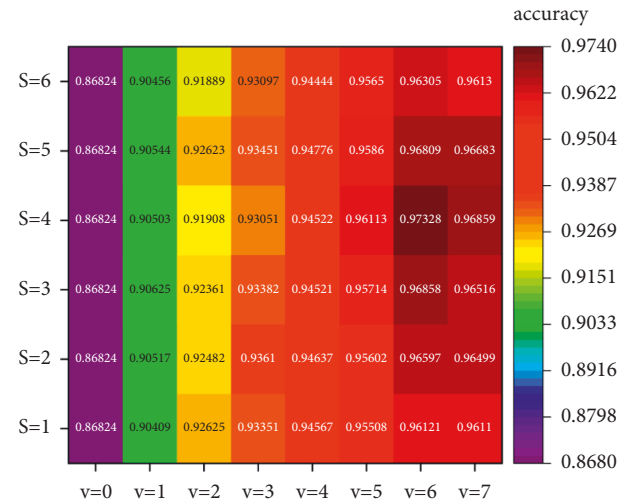


FIGURE 7: Average forecasting accuracy of combined models with different S and v .

potential connection between the influencing factors of high-dimensional space. If all the input forecasting models are not reduced by data dimension, it will make the model extract some redundant features, increase the calculation amount of the model, and even bring negative improvement to the forecasting accuracy of the combined model. On the contrary, if the dimension of influencing factors is reduced to 2 or lower, the data in low-dimensional space cannot carry all features of the raw data, thus reducing the generalization ability of the model.

4.2. Model Comparison. To test the performance of the proposed model in short-term load forecasting, this paper establishes multiple models for simulation and comparison, including the time series model based on LSTM (TLSTM), the nonlinear mapping model based on LSTM (MLSTM),

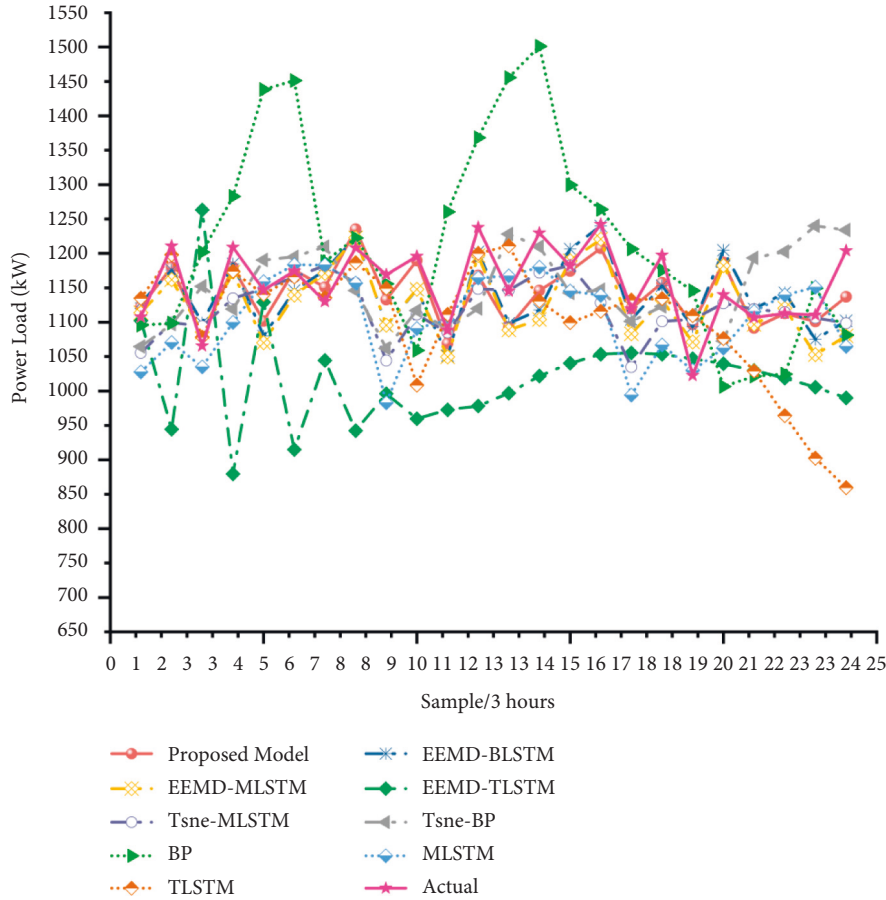


FIGURE 8: The performance of forecasting of each model.

the back propagation neural network (BP), the model combining TSNE and MLSTM (TSNE-MLSTM), the model combining TSNE and BP (TSNE-BP), the model combining EEMD and MLSTM (EEMD-MLSTM), the model combining EEMD and TLSTM (EEMD-TLSTM), and the model combining EEMD, MLSTM, and TLSTM (EEMD-BLSTM). Among them, BP uses three hidden layers, each hidden layer contains 60 neurons, and LSTM model parameters are set as above.

The forecasting results of each model are shown in Figure 8, and the detailed comparison information is demonstrated in Table 1, including the mean absolute percentage error (MAPE), the root mean square error (RMSE), and the mean absolute error (MAE).

In Figure 8, the load forecasting value of each sample point of the EEMD-TLSTM model deviates from the actual load value greatly, and the IMF component with high-frequency characteristics has a large error when using the time series model to forecast. The accuracy of TLSTM model is superior to that of EEMD-TLSTM, but its forecasting accuracy is still not remarkable in the comparison of all models, mainly due to the strong fluctuation of the original load signal, the effect of load forecasting directly using the time series model is not good. On the contrary, if the original load signal is decomposed by EEMD, the decomposed IMF component is used to build the MLSTM model, and the

TABLE 1: Error of forecasting results of each model.

Forecasting model	MAPE (%)	RMSE (kW)	MAE (kW)
Proposed model	2.6716	39.3329	31.0916
EEMD-BLSTM	3.6951	50.3015	42.7011
EEMD-MLSTM	3.8699	54.7858	45.1035
TSNE-MLSTM	4.4523	67.2027	52.2686
TSNE-BP	6.0121	77.4130	69.4224
MLSTM	5.7885	84.5392	68.0465
TLSTM	6.3646	108.5245	73.9725
BP	10.2852	147.7220	118.6361
EEMD-TLSTM	13.1761	177.7819	155.0436

residual signal is used to build the TLSTM model. Then, the appropriate models are used to forecast, respectively, such as EEMD-BLSTM and proposed model, and the forecasting efficiency will be significantly improved.

Ultimately, due to the fact that TSNE reduces the dimension of the influencing factors of power load in high-dimensional space, the potential redundant features in the comprehensive influencing factors in low-dimensional space are less, so that the forecasting effects of TSNE-MLSTM and TSNE-BP compared with MLSTM and BP models are improved by 1.9186% and 4.2731%, respectively. The proposed model also improves the accuracy of 1.0235% for the EEMD-BLSTM model with the EEMD algorithm. The forecasting

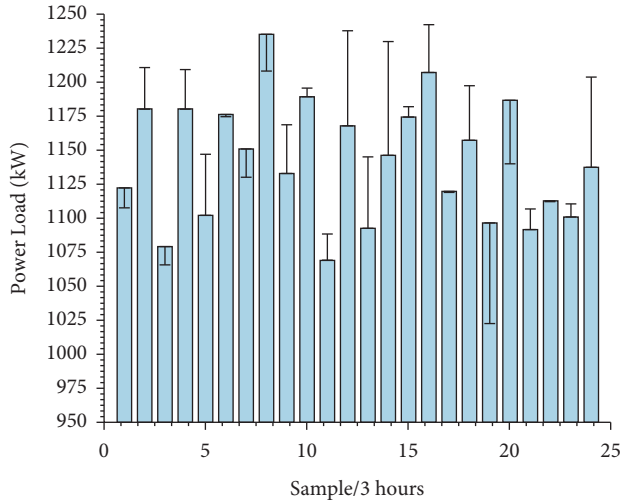


FIGURE 9: Forecasting results and errors of each sample point. The error bars represent absolute error between the forecasting value and actual value.

values of each sample point in proposed model are shown in Figure 9. The size of the error bars in Figure 9 indicates the absolute error. The position of the bar end is the actual load data of each sample point. The RMSE of the proposed model is 39.3329 kW, and the MAE is 31.0916 kW. The forecasting results are close to the actual load value.

5. Conclusions

This paper establishes a short-term power load forecasting model based on TSNE-EEMD-LSTM. TSNE is used to map the data of power load influencing factors in original dimension to the lower dimension, which reduces the calculation of redundant features. EEMD is used to split the power load curve into multiple IMF components with different frequencies and a flat residual signal. Based on the characteristics of IMF components and residual signals, appropriate models based on LSTM are established, including the nonlinear mapping model with low-dimensional space comprehensive influencing factors as model input and the time series model with historical data of signal components as model input. Finally, the important parameters of the combined model are determined by comparative experiments, and the feasibility of the proposed hybrid model in short-term load forecasting is verified. Several following conclusions are drawn:

- (1) There are redundant features between the original power load influencing factors. The multiple extractions of redundant features by the forecasting model will increase the calculation amount of the model, decrease the forecasting speed of the model, and even lead to poor generalization ability of the model. TSNE algorithm can greatly retain data structure and reduce redundant feature extraction. When the original influencing factors are time, temperature, pressure, humidity, wind speed and

rainfall, it is best to map the influencing factors to the four-dimensional space.

- (2) The IMFs signal and residual signal decomposed by EEMD have different characteristics. Using these characteristics to build several targeted sub-models can improve the accuracy of prediction. In this paper, the LSTM model is divided into nonlinear mapping model and time sequence model by changing the model input. All IMFs signals are suitable for modeling by nonlinear mapping model, and the residual signal is suitable for modeling by time series model.
- (3) The TSNE-EEMD-LSTM model proposed in this paper can avoid extracting redundant features as much as possible and fully tap the potential characteristics of influencing factors, which has an excellent performance in the comparison of multi-model forecasting performance.

The shortcomings of this study and follow-up work:

The models used in this paper are based on LSTM, and work types are divided by changing the input of the model. In the subsequent research, more advanced and more appropriate forecasting models will be taken into account to improve the performance of each sub-model in the hybrid model. In the future, it is expected that our proposed scheme can be applied to the forecasting fields of new energy generation [17,18] and integrated loads [19–21].

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by 2022 Innovation and Entrepreneurship Training Program for College Students of Shenyang Institute of Engineering, Scientific Research Project of Education Department of Liaoning Province (LJKQZ2021079), Doctoral Scientific Research Foundation of Liaoning Province (2020-BS-181) and Liaoning Revitalization Talents Program (XLYC1907138).

References

- [1] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, 2015.
- [2] X. Shao and C. S. Kim, "Multi-step short-term power consumption forecasting using multi-channel LSTM with time location considering customer behavior," *IEEE Access*, vol. 8, pp. 125263–125273, 2020.

- [3] T. H. Ouyang, Y. S. He, H. J. Li, Z. Y. Sun, and S. Baek, "Modeling and forecasting short-term power load with copula model and deep belief network," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 2, pp. 127–136, 2019.
- [4] R. Wang, Q. Sun, P. Tu, J. Xiao, Y. Gui, and P. Wang, "Reduced-order aggregate model for large-scale converters with inhomogeneous initial conditions in DC microgrids," *IEEE Transactions on Energy Conversion*, vol. 36, no. 3, pp. 2473–2484, 2021.
- [5] W. C. Kong, Z. Y. Dong, Y. W. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019.
- [6] Q. Liu, Y. Shen, L. Wu, J. Li, L. Zhuang, and S. Wang, "A hybrid FCW-EMD and KF-BA-SVM based model for short-term load forecasting," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 2, pp. 226–237, 2018.
- [7] H. Bo, Y. Nie, and J. Wang, "Electric load forecasting use a novelty hybrid model on the basic of data preprocessing technique and multi-objective optimization algorithm," *IEEE Access*, vol. 8, pp. 13858–13874, 2020.
- [8] C. Shang, J. Gao, H. Liu, and F. Liu, "Short-term load forecasting based on PSO-KFCM daily load curve clustering and CNN-LSTM model," *IEEE Access*, vol. 9, pp. 50344–50357, 2021.
- [9] K. Aurangzeb, M. Alhussein, K. Javaid, and S. I. Haider, "A pyramid-CNN based deep learning model for power load forecasting of similar-profile energy customers based on clustering," *IEEE Access*, vol. 9, pp. 14992–15003, 2021.
- [10] W. Xu, H. Hu, and W. Yang, "Energy time series forecasting based on empirical mode decomposition and FRBF-AR model," *IEEE Access*, vol. 7, pp. 36540–36548, 2019.
- [11] Z. Kong, C. Zhang, H. Lv, F. Xiong, and Z. Fu, "Multimodal feature extraction and fusion deep neural networks for short-term load forecasting," *IEEE Access*, vol. 8, pp. 185373–185383, 2020.
- [12] S. Das, B. R. Prusty, and K. Bingi, "Review of adaptive decomposition-based data preprocessing for renewable generation rich power system Applications," *Journal of Renewable and Sustainable Energy*, vol. 13, no. 6, Article ID 062703, 2021.
- [13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [14] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 01, no. 01, pp. 1–41, 2009.
- [15] J. Liu, Q. Li, H. Yang, Y. Han, S. Jiang, and W. Chen, "Sequence fault diagnosis for PEMFC water management subsystem using deep learning with t-SNE," *IEEE Access*, vol. 7, pp. 92009–92019, 2019.
- [16] M. A. Da Silva Lopes, A. D. Dória Neto, and A. De Medeiros Martins, "Parallel t-SNE applied to data visualization in smart cities," *IEEE Access*, vol. 8, pp. 11482–11490, 2020.
- [17] W. Liu, C. Ren, and Y. Xu, "PV generation forecasting with missing input data: a super-resolution perception approach," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1493–1496, 2021.
- [18] M. Li, M. Yang, Y. Yu, and W.-J. Lee, "A wind speed correction method based on modified hidden markov model for enhancing wind power forecast," *IEEE Transactions on Industry Applications*, vol. 58, no. 1, pp. 656–666, 2022.
- [19] W. Y. Liu, "Cooling, heating and electric load forecasting for integrated energy systems based on CNN-LSTM," in *Proceedings of the 2021 6th International Conference On Power And Renewable Energy (ICPRE)*, Shanghai, China, 2021.
- [20] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7866–7879, 2021.
- [21] E. Liu, Y. Wang, and Y. Huang, "Short-term forecast of multi-load of electrical heating and cooling in regional integrated energy system based on deep LSTM RNN," in *Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, Wuhan, China, 2020.