Energy Engineering

Tech Science Press

https://doi.org/10.32604/ee.2026.077923

ARTICLE

Load Forecasting Method Considering Distributed Photovoltaics and Multi-Dimensional Temporal Characteristics

Zhengji Meng[1], **Lei Wang**[1], **Xuekai Hu**[1], **Zijian Wang**[2,*] **and Linjun Shi**[2]

[1]State Grid Hebei Electric Power Company Electric Power Research Institute, Shijiazhuang, China
[2]School of Electrical and Power Engineering, Hohai University, Nanjing, China

*Corresponding Author: Zijian Wang. Email: wangzijian20011015@163.com

Received: 19 December 2025; Accepted: 26 February 2026; Published: 12 November 2026

ABSTRACT: With the rapid popularization of distributed photovoltaic (PV) systems, traditional load forecasting is facing more and more challenges due to various problems caused by distributed PV grid connection. Distributed PV significantly affects the net load distribution and weakens the effectiveness of traditional forecasting methods. Therefore, improving the accuracy of load forecasting under high PV penetration has become a key issue for the safe and efficient operation of modern power systems. To solve this problem, a load forecasting method based on Particle Swarm Optimization and Bi-directional Long Short-Term Memory (PSO-BiLSTM) network with distributed PV identification and multi-dimensional temporal characteristics is proposed. Firstly, the multi-dimensional temporal characteristics are constructed by combining the lagged characteristics, cyclical characteristics and calendar characteristics, which enhances the ability of the model to capture the time correlation between load data. At the same time, using the net load characteristics under different weather conditions, the distributed PV identification characteristic is constructed to effectively identify whether users install distributed PV. Then, PSO algorithm is used to automatically optimize the key hyperparameters of BiLSTM, so as to improve the robustness of the model. Finally, the effectiveness of the method is verified by the actual load data. The experimental results show that the PSO-BiLSTM model is superior to other models in many evaluation indexes. Especially in the case of distributed PV grid connection, the prediction accuracy of the model has been significantly improved, and the time relationship and load changes caused by distributed PV have been effectively captured. These results confirm the effectiveness and practicability of the proposed method in power system load forecasting with high distributed PV penetration.

KEYWORDS: Load forecasting; BiLSTM; temporal characteristics; distributed photovoltaics; particle swarm optimization

1 Introduction

Electric power is a foundational industry vital to national economic and social development. The rapid expansion of China's economy has consistently driven increased electricity demand across all sectors. As a crucial tool for the operation and management of modern power systems, load forecasting holds significant importance in supporting the energy transition, optimizing resource allocation, and facilitating the achievement of the "dual carbon" goals [1]. Concurrently, the development and widespread adoption of new energy technologies have substantially increased the complexity and uncertainty of load profiles on the consumer side of power systems, posing significant challenges to power dispatch operations [2]. Unlike traditional load forecasting scenarios, the net load in contexts involving distributed photovoltaics (PV) is influenced by both PV output and user consumption patterns. In this case, it often leads to the distortion

Copyright © 2027 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

of load curve, the weakening of time correlation and the effectiveness of traditional statistical and machine learning forecasting models. Therefore, accurately capturing the complex time dependence and identifying the impact of distributed PV has become a key challenge for modern load forecasting, especially in the case of high PV penetration. In 2024, the total electricity consumption of society reached 9852.1 billion kWh, representing a year-on-year growth of 6.8%. Within this total, the generation of electricity by industries above designated size amounted to 9418.1 billion kWh. The primary sector consumed 135.7 billion kWh, representing a 6.3% year-on-year increase. The secondary sector consumed 6387.4 billion kWh, marking a 5.1% year-on-year increase. Meanwhile, the tertiary sector consumed 1834.8 billion kWh, reflecting a 9.9% year-on-year increase and the residential electricity consumption reached 1494.2 billion kWh, indicating a 10.6% year-on-year increase [3].

In virtual power plants (VPPs), accurate load forecasting considering distributed PV can effectively quantify and hedge against financial risks arising from renewable generation and price volatilities, enabling them to fulfill their crucial role in the transition to a sustainable, reliable, and efficient power system [4]. Therefore, accurate power load forecasting is of critical importance in mitigating operational risks within power systems, supporting the development of energy storage and distributed energy resources (DER), achieving supply-demand balance, and ensuring the secure and stable operation of power systems [5,6].

At present, the methods for load forecasting can mainly be divided into traditional methods and artificial intelligence (AI) methods. The traditional methods are mainly based on statistical analysis methods [7], such as multiple linear regression [8], Kalman filtering [9], etc. Methods based on machine learning and deep learning have gradually become mainstream with advances in artificial intelligence. Models such as Random Forest (RF) [10], Support Vector Machines (SVM) [11], and Extreme Gradient Boosting (XGBoost) [12] can handle non-linear and multidimensional data well, but are sensitive to outliers. Deep learning models such as Recurrent Neural Networks (RNN) [13], Long Short-Term Memory (LSTM) [14], and Convolutional Neural Networks (CNN) [15] can extract high-level features and perform well in handling complex nonlinear problems.

Researchers worldwide have conducted extensive studies on electricity load forecasting. Reference [16] introduced a novel optimized hybrid model combining LSTM and Transformer deep learning architectures designed for power load forecasting. It leverages the strengths of both LSTM and Transformer models. Reference [17] proposed a short-term power load forecasting model called PowerTimeMixer, which enhances feature extraction capability through time series decomposition and introduces convolution. Finally, it drives an independent date pattern prediction module to generate prediction results. Reference [18] proposed the use of LSTM with the attention mechanism to perform the time series forecasting, considering a sequence-to-sequence (Seq2Seq) time series data of the electricity load in Brazil. The proposed Seq2Seq-LSTM with attention mechanism outperforms other well-established models. Reference [19] proposed an electric load prediction framework based on attention mechanism time series deep separable convolutional neural network, which combines seasonal decomposition algorithm with manual feature engineering to extract periodic and stochastic features of electric loads. At the same time, Bayesian algorithm also optimizes model hyperparameters to achieve optimal model performance. Reference [20] proposed a multi model fusion prediction method based on complementary ensemble empirical mode decomposition, genetic algorithm LSTM, radial basis function fusion autoencoder, and PSO-SVM. Reference [21] used Spearman correlation coefficient method to analyze the correlation between various influencing factors and loads, and selected the factors with higher correlation as inputs for the model. The existing research mainly focuses on various load forecasting methods, but there are few studies that consider the integration of multidimensional temporal characteristics and distributed photovoltaic identification characteristics. Therefore, this paper constructs

multidimensional temporal characteristics and distributed photovoltaic identification characteristics, and combines them with traditional load forecasting methods.

The organization of this manuscript is as follows:

In Section 2, by integrating multiple time lagged characteristics, cyclical characteristics and calendar characteristics, a characteristic space with strong temporal representation ability is constructed. It also identifies and determines whether distributed photovoltaics are installed in the load. Section 3 details how the BiLSTM model achieves intelligent optimization of its key parameters through the Particle Swarm Optimization (PSO) algorithm, employing an adaptive inertia weight adjustment strategy. In Section 4, through experimental data comparison, the improved PSO-BiLSTM model that takes into account distributed photovoltaics and multidimensional temporal characteristics has higher prediction accuracy, verifying the superiority of the model in capturing complex temporal patterns of power loads. Section 5 summarizes the research conclusions and outlines directions for future work.

2 Characteristic Construction

2.1 Multi-Dimensional Temporal Characteristics

The formation and variation of load data are closely related to factors such as weather conditions, geographical location, regional type, and electricity consumption patterns of users. Regional loads exhibit distinct periodicity, including daily, weekly, and annual cycles. In order to identify the most relevant time points for constructing temporal features, Pearson correlation, Spearman correlation, and Maximal Information Coefficient (MIC) analyses were used to investigate the correlation coefficients at different lag time points. The analysis utilized all load data from January to April 2025 (as detailed in the Case Study Analysis section), with resulting coefficients shown in Fig. 1.

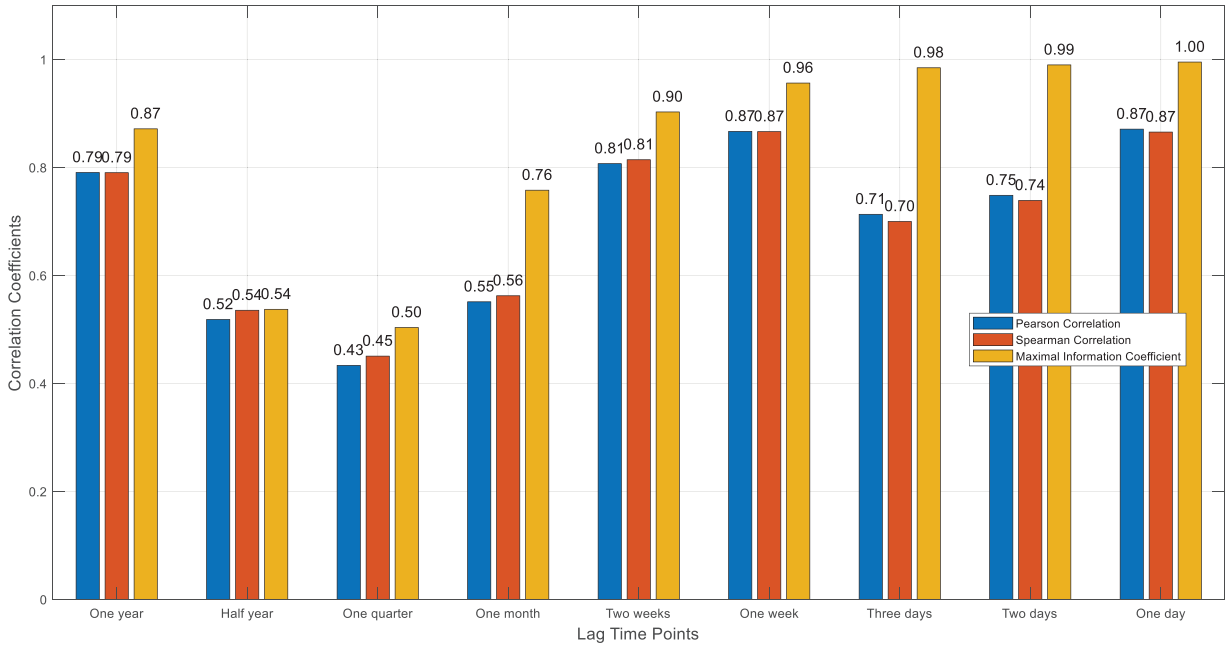


Figure 1: Correlation coefficients at different lag time points.

The Pearson correlation coefficient is a statistical measure that is employed to ascertain linear relationships between two variables. It is calculated using the concept of covariance and the standard deviations of

both variables. The Spearman correlation coefficient is a non-parametric, rank-based method for assessing monotonic relationships. It first converts raw data into ranks before computing the Pearson correlation coefficient of these ranks. The MIC is a non-parametric method measuring non-linear relationships by partitioning the data space into grids and calculating mutual information within each grid. Comprehensive analysis of correlation coefficients across lag times revealed that current load data exhibited high correlation with load data from one year prior, one week prior, and one day prior. The respective correlation coefficients were 0.79, 0.79, 0.87 (one year prior); 0.87, 0.87, 0.96 (one week prior); and 0.87, 0.87, 1.00 (one day prior). Therefore, the forecasting model was developed on the basis of load data from one year prior, one week prior, and one day prior. This data was selected for predicting the target day's load, forming the lagged characteristics within the multi-dimensional temporal characteristics:

$$L_t = \left[L_{1y}, L_{1w}, L_{1d}\right] \tag{1}$$

where L_t represents load data matrix comprising lag characteristics; L_{1y} is historical load data from one year prior; L_{1w} is historical load data from one week prior; L_{1d} represents historical load data from one day prior.

Power loads demonstrate clear intra-day fluctuations and seasonal variations. To explicitly represent these periodic characteristics, trigonometric functions were adopted:

$$t_{d1} = \sin\left(\frac{2\pi h}{24}\right) \tag{2}$$

$$t_{d2} = \cos\left(\frac{2\pi h}{24}\right) \tag{3}$$

$$t_m = \sin\left(\frac{2\pi m}{12}\right) \tag{4}$$

where t_{d1} and t_{d2} are intra-day periodic characteristics; h represents hour of the day; t_m represents monthly periodic characteristics; m is month of the year.

In view of the considerable variations in load patterns between workdays and weekends, calendar characteristics were constructed for all dates by incorporating three distinct labels:

$$D = \left[D_w, I_w, D_y\right] \tag{5}$$

where D represents calendar characteristic matrix; D_w is day-of-week label, $D_w \in [1,7]$; I_w is weekend indicator parameter ($I_w = 0$ for weekdays, $I_w = 1$ for weekends); D_y represents day-of-year index ($D_y \in [1,365]$; $D_y \in [1,366]$ for leap years).

Additionally, special consideration was given to leap years (e.g., 2024). For 29 February, where corresponding data from the previous year is absent, the following mapping function was applied:

$$L(t) = \begin{cases} datetime\,(y-1, 3, 1, H, Min) \\ datetime\,(y-1, M, D, H, Min) \end{cases} \tag{6}$$

where $L(t)$ is load data from the corresponding date in the previous year; y represents the year, e.g., 2024; M represents the month, e.g., 3; D is the day, e.g., 1; H is the hour, e.g., 14; Min represents the minute, e.g., 30; For 29 February, the load data from 1 March of the previous year is used.

In summary, the present study constructed multi-dimensional temporal characteristics encompassing lagged characteristics, cyclical characteristics, and calendar characteristics, while accounting for leap year anomalies through explicit mapping rules.

2.2 Distributed PV Identification Characteristics

First, according to Basic Terminology of Weather Forecast (GB/T 35663-2017) issued by SAC/TC345 on 29 December 2017 [22], sky conditions are classified into four types based on total cloud cover: clear sky, partly cloudy, cloudy, and overcast. Since precipitation (e.g., rain/snow) impacts PV output similarly to overcast conditions, this study simplifies weather types (WT) into four categories: WT-S (clear sky), WT-L (partly cloudy), WT-C (cloudy), and WT-R (overcast). By analyzing historical net load data, net load curves (NLCs) under these four weather types can be derived. The user's net load is defined as:

$$P_{\text{net}}(t) = P_{\text{real}}(t) - P_{\text{pv}}(t) \tag{7}$$

where $P_{\text{net}}(t)$ is net load at time t; $P_{\text{real}}(t)$ represents actual power consumption at time t; $P_{\text{pv}}(t)$ represents output of distributed PV at time t, if no PV system is installed, $P_{\text{pv}}(t) = 0$.

The NLCs are obtained by averaging net load data under each weather type:

$$\text{NLC}_X(t) = \frac{1}{T_X}\sum P_{\text{net},X}(t) \tag{8}$$

where $\text{NLC}_X(t)$ represents net load curve function for weather type X at time t ($X \in \{\text{S, L, C, R}\}$); T_X is total number of days with weather type X; $P_{\text{net},X}(t)$ represents net load at time t on the i-th day of type X. Fig. 2 illustrates the NLC of a certain user.

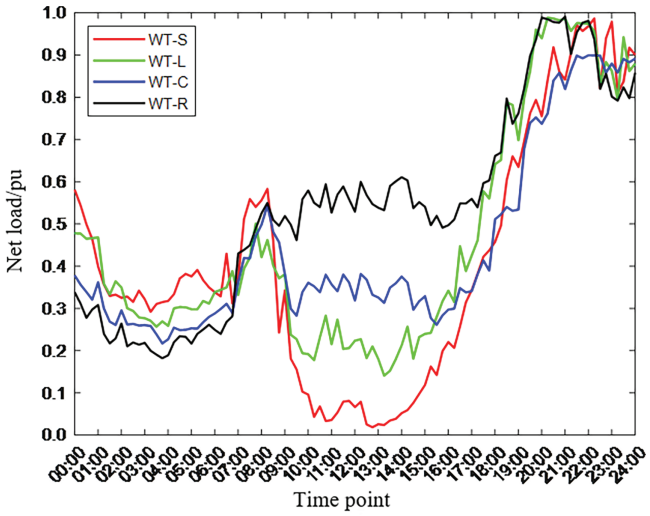


Figure 2: Net load curve of a certain user under four different weather types.

Given seasonal variations in sunrise/sunset times and PV generation being active only during daylight, this study focuses on the NLC between 09:00 and 16:00. Define the start point $S(t_s, \text{NLC}(t_s))$, and end point $E(t_e,\text{NLC}(t_e))$.

(1) Clear-to-Cloudy Net Load Power Ratio (C_1)

The ratio of total net load power under clear sky (WT-S) and cloudy (WT-C) conditions:

$$C_1 = \frac{\sum_{t=t_s}^{t=t_e} |\text{NLC}_S(t)|}{\sum_{t=t_s}^{t=t_e} |\text{NLC}_C(t)|} \tag{9}$$

where $|\cdot|$ represents the absolute value of '·'; t_s = 09:00; t_e = 16:00; $\mathrm{NLC}_S(t)$ is net load curve function for weather type WT-S at time t; $\mathrm{NLC}_C(t)$ represents net load curve function for weather type WT-C at time t.

(2) Concavity Metric (C_2)

Quantifies curve concavity by connecting S and E with a straight line:

$$y_L(t) = k_L \times (t - t_s) + NLC_S(t_s) \tag{10}$$

$$k_L = \frac{\mathrm{NLC}_S(t_e) - \mathrm{NLC}_S(t_s)}{t_e - t_s} \tag{11}$$

where $y_L(t)$ is the expression for the line L; k_L represents the slope of the line L.

Define two sets: $A = \{y_L(t) | t \in [t_s, t_e]\}$ (points on line L) and $B = \{\mathrm{NLC}_S(t) | \mathrm{NLC}_S(t) < y_L(t), t \in [t_s, t_e]\}$ (points below line L).

The concavity metric (C_2) is:

$$C_2 = \frac{|A|}{|B|} \tag{12}$$

where $|\cdot|$ is the number of elements in the sets A and B.

(3) Concavity Index (C_3)

Addresses limitations of C_2 by comparing slopes around noon (12:00). For clear sky conditions:

$$k_{L1} = \frac{\mathrm{NLC}_S(t_{\mathrm{noon}}) - \mathrm{NLC}_S(t_A)}{t_{\mathrm{noon}} - t_A} \tag{13}$$

$$k_{L2} = \frac{\mathrm{NLC}_S(t_B) - \mathrm{NLC}_S(t_{\mathrm{noon}})}{t_B - t_{\mathrm{noon}}} \tag{14}$$

$$k_S = |k_{L1}| + |k_{L2}| \tag{15}$$

where t_A = 08:00; t_B = 18:00; t_{noon} = 12:00; k_{L1} and k_{L2} are the slopes of lines L_1 and L_2.

Similarly for cloudy conditions:

$$k_{L3} = \frac{\mathrm{NLC}_C(t_{\mathrm{noon}}) - \mathrm{NLC}_C(t_A)}{t_{\mathrm{noon}} - t_A} \tag{16}$$

$$k_{L4} = \frac{\mathrm{NLC}_C(t_B) - \mathrm{NLC}_C(t_{\mathrm{noon}})}{t_B - t_{\mathrm{noon}}} \tag{17}$$

$$k_C = |k_{L3}| + |k_{L4}| \tag{18}$$

where k_{L3} and k_{L4} are the slopes of lines L_3 and L_4.

The concavity index (C_3) is:

$$C_3 = \frac{k_S}{k_C} \tag{19}$$

(4) Load Change Rate (C_4)

As night falls, the output of distributed PV power will decay to 0, and at the same time, due to the arrival of peak electricity consumption by users, the net load curve will rise rapidly. Characterizes evening load surge after PV generation ceases:

$$\lambda_S = \frac{\mathrm{NLC}_S\left(t_{\mathrm{night}}\right) - \mathrm{NLC}_S\left(t_{\mathrm{noon}}\right)}{t_{\mathrm{night}} - t_{\mathrm{noon}}} \tag{20}$$

$$\lambda_C = \frac{\mathrm{NLC}_C\left(t_{\mathrm{night}}\right) - \mathrm{NLC}_C\left(t_{\mathrm{noon}}\right)}{t_{\mathrm{night}} - t_{\mathrm{noon}}} \tag{21}$$

where t_{night} = 21:00; λ_S and λ_C represent the slopes under sunny and cloudy conditions, respectively.

The load change rate (C_4) is:

$$C_4 = \frac{\lambda_S}{\lambda_C} \tag{22}$$

These four features form the input matrix for PV identification:

$$\boldsymbol{C} = [C_1, C_2, C_3, C_4] \tag{23}$$

A RF classifier model is trained using $\boldsymbol{C}$ to determine whether distributed PV is installed.

3 Development of the Improved BiLSTM Model

3.1 Principles of LSTM and BiLSTM

LSTM is an enhanced variant of RNN. While sharing a similar basic structure with RNN, LSTM incorporates a long-term memory module within its hidden layer. This innovation enables LSTM to retain early time-series features and mitigates issues of gradient vanishing and explosion during backpropagation [23].

The LSTM cell comprises three core components: the input gate, forget gate, and output gate. Its computational workflow is defined as follows:

$$\begin{cases} f_t = \sigma\left(W_f \bullet [h_{t-1}, x_t] + b_f\right) \\ i_t = \sigma\left(W_i \bullet [h_{t-1}, x_t] + b_i\right) \\ \tilde{c}_t = \tanh\left(W_c \bullet [h_{t-1}, x_t] + b_c\right) \\ c_t = f_t \bullet c_{t-1} + i_t \bullet \tilde{c}_t \\ o_t = \sigma\left(W_o \bullet [h_{t-1}, x_t] + b_o\right) \\ h_t = o_t \bullet \tanh\left(c_t\right) \end{cases} \tag{24}$$

where f_t is the vector for forget gate at time step t; σ represents sigmoid activation function; W_f and b_f represent the weight matrix and bias vector for forget gate; h_{t-1} and h_t are hidden state outputs at time steps $t-1$ and t; x_t represents input vector at time step t; i_t is the vector for input gate at time step t; W_i and b_i represent the weight matrix and bias vector for input gate; c_{t-1} and c_t are cell state vectors at time steps $t-1$ and t; W_o and b_o represent the weight matrix and bias vector for the candidate cell state; o_t is the vector for output gate at time step t; W_o and b_o represent the weight matrix and bias vector for output gate.

BiLSTM extends the LSTM architecture by incorporating a reverse hidden layer. This allows sequence information to propagate in both forward and backward directions [24]. Consequently, the model captures contextual information from past and future states when generating outputs, significantly enhancing its generalization and representational capacity.

The BiLSTM structure consists of a forward LSTM layer and a backward LSTM layer. Its computations are:

$$\begin{cases} \vec{h}_t = \vec{\text{LSTM}}\left(x_t, \vec{h}_{t-1}\right) \\ \overleftarrow{h}_t = \overleftarrow{\text{LSTM}}\left(x_t, \overleftarrow{h}_{t+1}\right) \\ H_t = \left[\vec{h}_t, \overleftarrow{h}_t\right] \end{cases} \tag{25}$$

where $\vec{h}_t$ represents hidden state of the forward LSTM at time step t; $\vec{\text{LSTM}}$ is a forward LSTM unit; $\vec{h}_{t-1}$ represents forward hidden state at time step $t-1$; $\overleftarrow{h}_t$ is hidden state of the backward LSTM at time step t; $\overleftarrow{\text{LSTM}}$ represents a backward LSTM unit; $\overleftarrow{h}_{t+1}$ is backward hidden state at time step $t+1$; H_t represents combined hidden state of BiLSTM at time step t.

3.2 PSO Optimized BiLSTM Algorithm

Due to the significant impact of hyperparameter setting on BiLSTM, and the tedious task of determining the optimal value of hyperparameters, manual adjustment attempts are obviously not realistic. Therefore, this paper selects PSO to iteratively optimize the hyperparameters of BiLSTM and obtain the optimal value.

PSO algorithm is an evolutionary computation method that originated from simulating bird flocks foraging. A group of birds randomly search for food, and each bird flies at a certain speed, which is dynamically adjusted based on its own flying experience and the flying experience of its companions, ultimately reaching the optimal flying state.

In this paper, a PSO hybrid is used to optimize BiLSTM hyperparameters, including the number of BiLSTM units, learning rate, dropout rate, and batch size. The PSO optimization procedure is as follows:

(1) Define the PSO parameters; the parameters used in this paper are listed in Table 1; the selection of PSO parameters referred to recent research on load forecasting and deep learning, and the rationality of parameter selection was proven through multiple value verifications;

Table 1: PSO parameters.

Parameters	Numeric Value
Population size	15
Number of iterations	25
Optimize dimensions	4

(2) Define the hyperparameter search space, specify which hyperparameters to optimize and their ranges (BiLSTM units, learning rate, dropout rate, batch size), as shown in Table 2; the selection of hyperparameter ranges also drew reference from recent research and was verified through experiments;

(3) Train the model with a given hyperparameter set and return validation metrics;

(4) Initialize the swarm's particle positions and velocities; each particle represents one hyperparameter configuration;

(5) Iteratively update the swarm, record results at each iteration, and continuously update the global best solution;

(6) Train the final model using the best hyperparameter set found.

Table 2: Hyperparameters and ranges.

Hyperparameters	Numeric Range
BiLSTM units	64–256
Learning rate	0.0001–0.01
Dropout rate	0.1–0.5
Batch size	32–128

3.3 Forecasting Model

The workflow of the proposed load forecasting model, which integrates multi-dimensional temporal characteristics and employs PSO for hyperparameter optimization, is illustrated in Fig. 3.

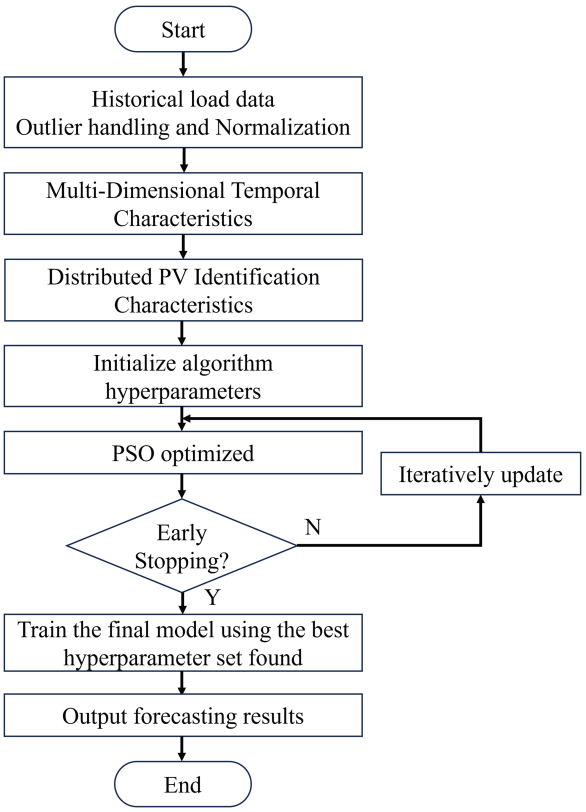


Figure 3: Flowchart of the load forecasting model.

The key steps of the forecasting model are as follows:

(1) Input historical load data and perform preprocessing steps including normalization and outlier handling;

(2) Construct multi-dimensional temporal characteristics comprising lagged characteristics, cyclical characteristics, and calendar characteristics. Assign corresponding data labels to all historical load records;

(3) Initialize PSO parameters and iteratively optimize algorithm hyperparameters within predefined search spaces. Record the global optimal solution throughout iterations;

(4) Train the final model using the optimized hyperparameters and output forecasting results.

3.4 Model Evaluation Metrics

This paper employs four evaluation metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). Their calculations are defined as follows:

$$\delta_{\mathrm{MAE}} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{26}$$

$$\delta_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{27}$$

$$\delta_{\mathrm{MAPE}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \tag{28}$$

$$\delta_{\mathrm{R^2}} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2} \tag{29}$$

where δ_{MAE} is the value of MAE; n represents total number of prediction points; $\hat{y}_i$ represents the i-th forecasted value; y_i is the i-th actual value; δ_{RMSE} represents the value of RMSE; δ_{MAPE} is the value of MAPE; $\delta_{\mathrm{R}}{}^2$ represents the value of R^2; $\overline{y}$ is mean of actual values.

If the prediction accuracy of the model is higher, MAE, RMSE and MAPE are smaller, and R^2 is closer to 1.

4 Case Study Analysis

4.1 Data Overview

The data used in this study are from a publicly available dataset for Brussels, Belgium, obtained from a certain website: https://opendata.elia.be/pages/home/. This paper selected all historical load data for 2023–2024 and January–April, 2025 with a sampling interval of 15 min. Load data for the full year of 2023 and January–November 2024 were used as the training set, December 2024 was used as the validation set, and January–April 2025 were used as the test set.

Because there is a large volume of historical load data, data preprocessing is necessary to handle possible data issues. The preprocessing mainly includes outlier handling and normalization.

For missing data, linear interpolation is used to fill missing values. Linear interpolation is a common method appropriate when data are continuous in time or space. It uses the known values on both sides of a missing point and assumes linear variation between them to estimate the missing value. Suppose the known data points are (x_0, y_0) and (x_1, y_1), and $x_0 < x < x_1$ is the independent variable corresponding to the missing value. The interpolated missing value y is computed by linear interpolation:

$$y = y_0 + \frac{(y_1 - y_0)}{(x_1 - x_0)} \times (x - x_0) \tag{30}$$

To avoid the influence of different units and large numerical differences among input variables that could introduce prediction errors, all inputs are normalized using min-max normalization.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{31}$$

where $x^{'}$ is the normalized input value; x represents the original input value; $x_{\min}$ is the minimum value of the input variable; $x_{\max}$ represents the maximum value of the input variable.

4.2 Model Prediction Results

First, this study employed a single-layer LSTM model for training. The model parameters are listed in Table 3.

Table 3: Parameters of the single-layer LSTM model.

Hyperparameters	Numeric Range
LSTM units	192
Learning rate	0.001
Dropout rate	0.3
Batch size	64
Maximum training epochs	30
Learning rate reduction period	20
Learning rate reduction ratio	50%

The prediction results are shown in Fig. 4.

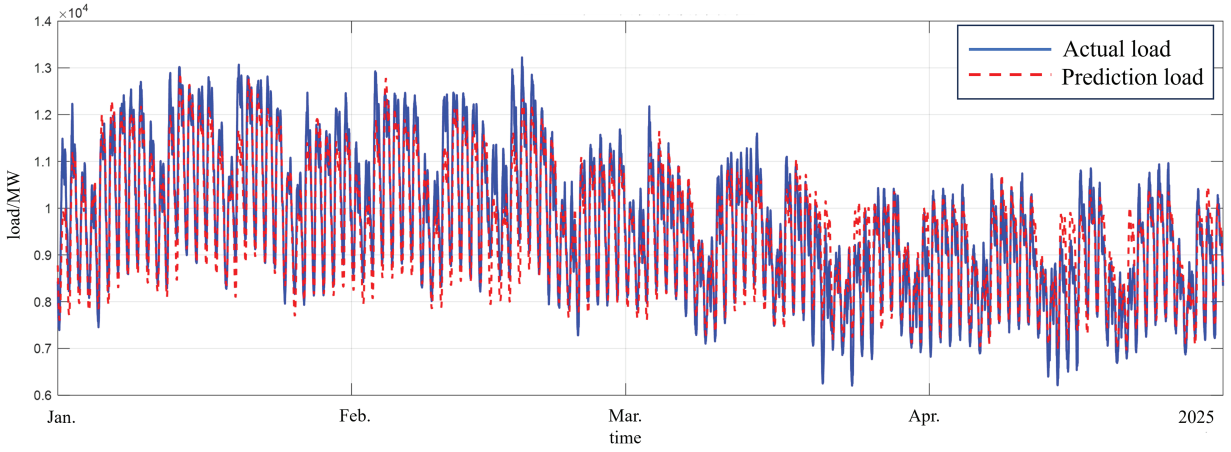


Figure 4: Prediction results of single-layer LSTM model (all data).

For the BiLSTM model, hyperparameter settings were identical to the LSTM model. The overall prediction results and the one representative week are shown in Figs. 5 and 6, respectively.

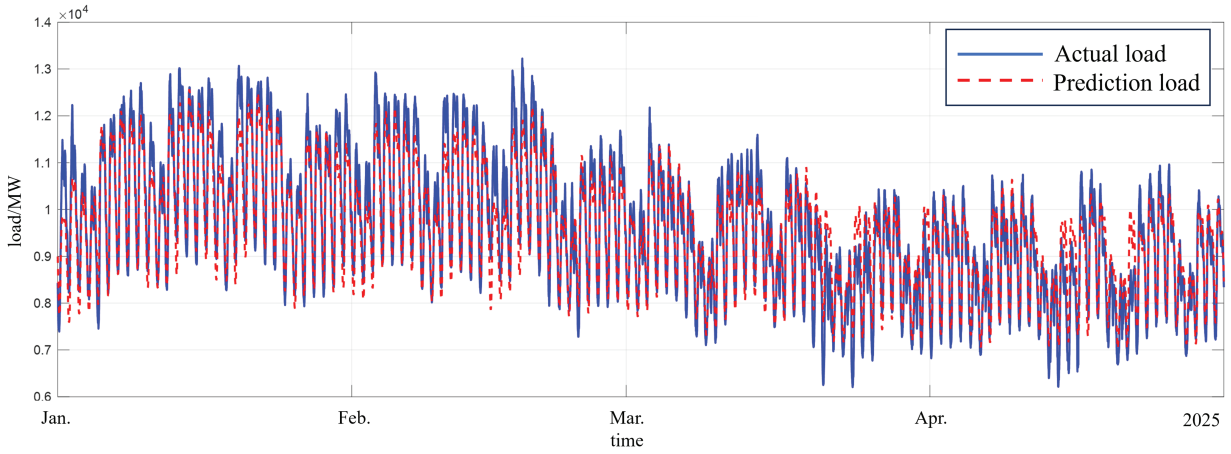


Figure 5: Prediction results of single-layer BiLSTM model (all data).

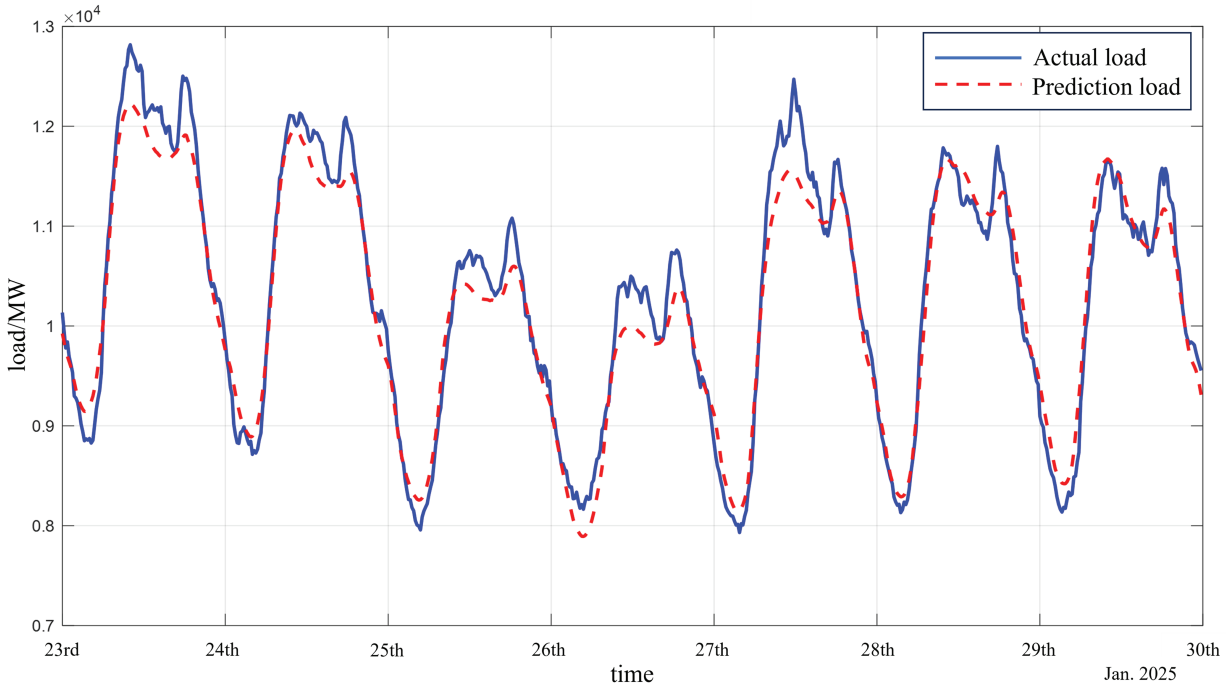


Figure 6: Prediction results of single-layer BiLSTM model (one week).

Subsequently, all single-layer models were replaced with double-layer LSTM and BiLSTM architectures for retraining. Model parameters are provided in Table 4, with results shown in Figs. 7–9.

Finally, hyperparameter optimization via PSO yielded optimal parameters (Table 5). The optimization process is visualized in Fig. 10. The retrained models produced results shown in Figs. 11–13.

Table 4: Parameters of the double-layer LSTM model.

Hyperparameters	Numeric Range
LSTM units	128
Learning rate	0.001
Dropout rate	0.3
LSTM units	64
Dropout rate	0.2
Maximum training epochs	30
Batch size	64
Learning rate reduction period	20
Learning rate reduction ratio	50%

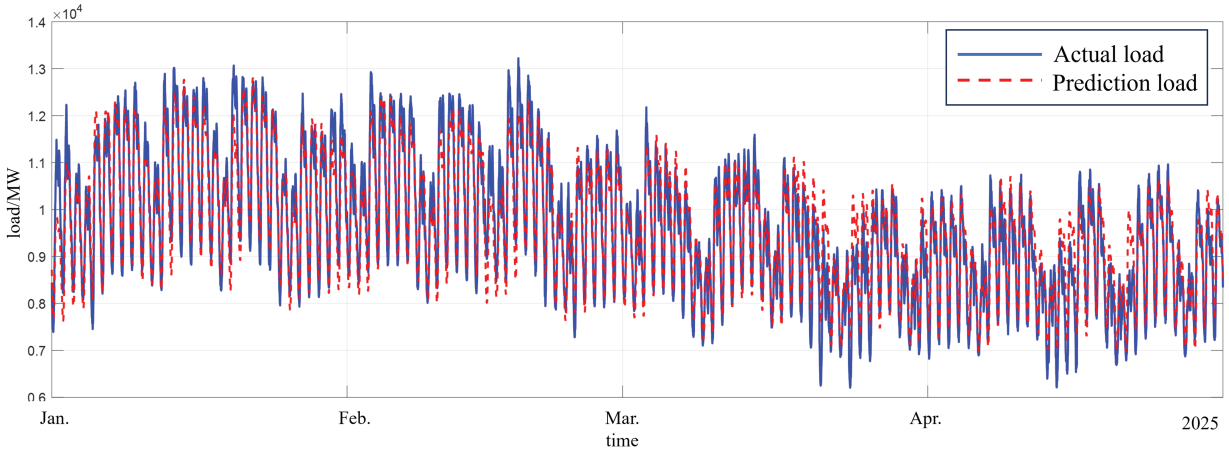


Figure 7: Prediction results of double-layer LSTM model (all data).

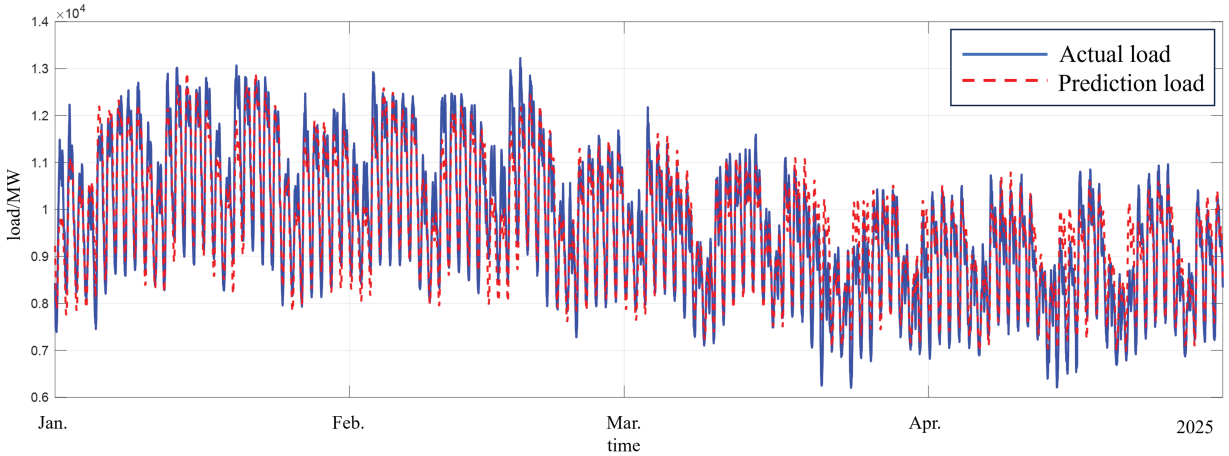


Figure 8: Prediction results of double-layer BiLSTM model (all data).

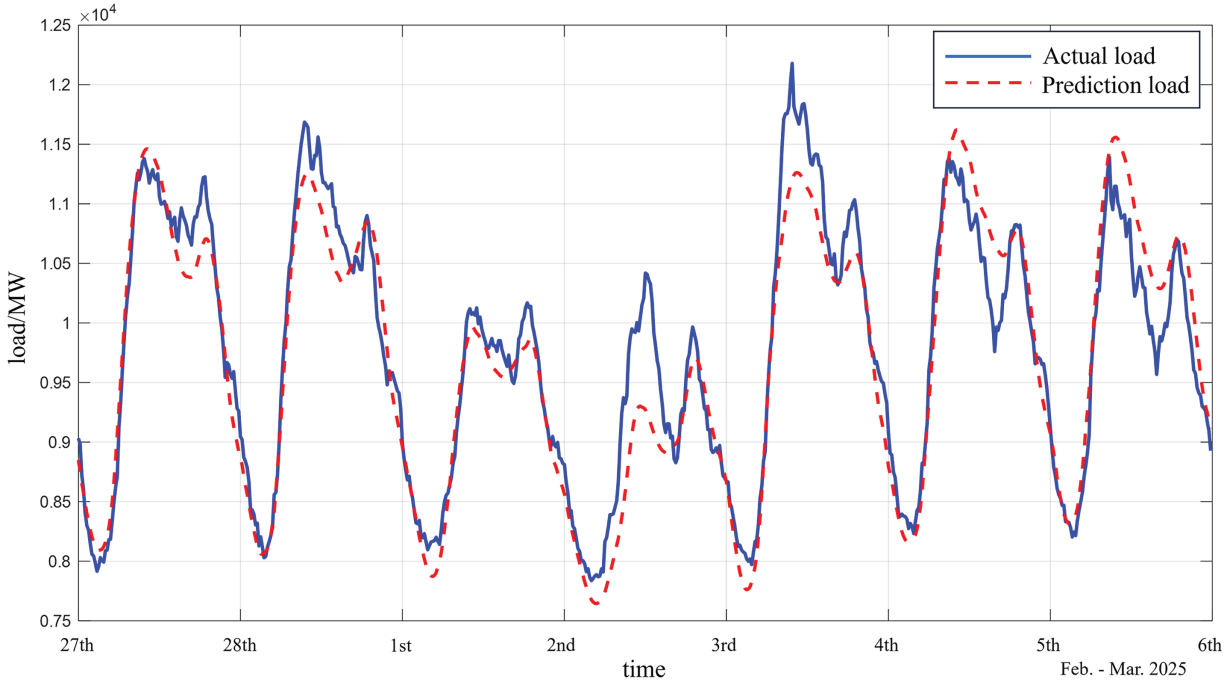


Figure 9: Prediction results of double-layer BiLSTM model (one week).

Table 5: Parameters of the PSO-LSTM model.

Hyperparameters	Numeric Value
LSTM units	84
Learning rate	0.00054
Dropout rate	0.12
Maximum training epochs	50
Batch size	123
Learning rate reduction period	15
Learning rate reduction ratio	70%

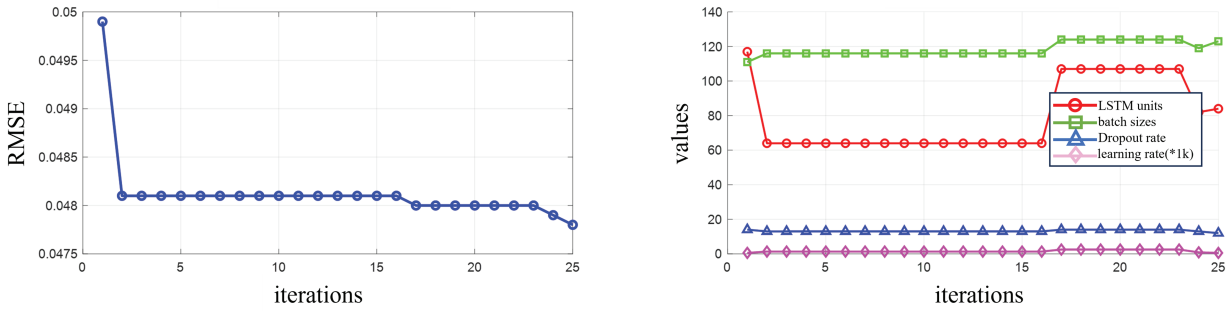


Figure 10: (Continued)

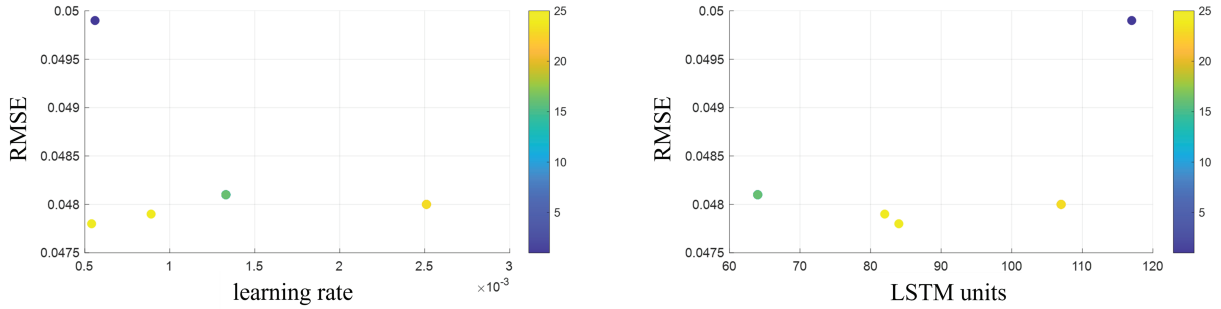


Figure 10: PSO optimization process.

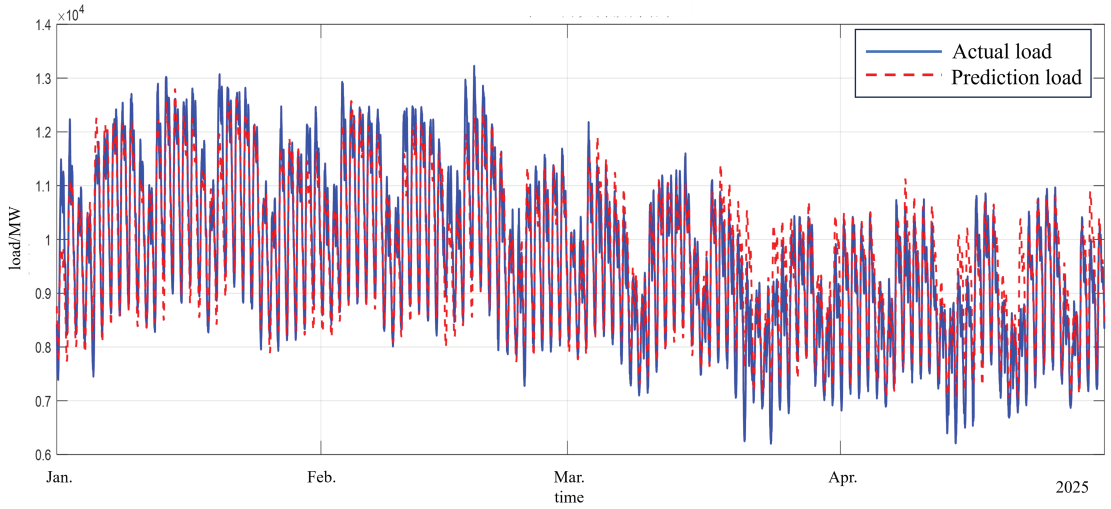


Figure 11: Prediction results of PSO-LSTM model (all data).

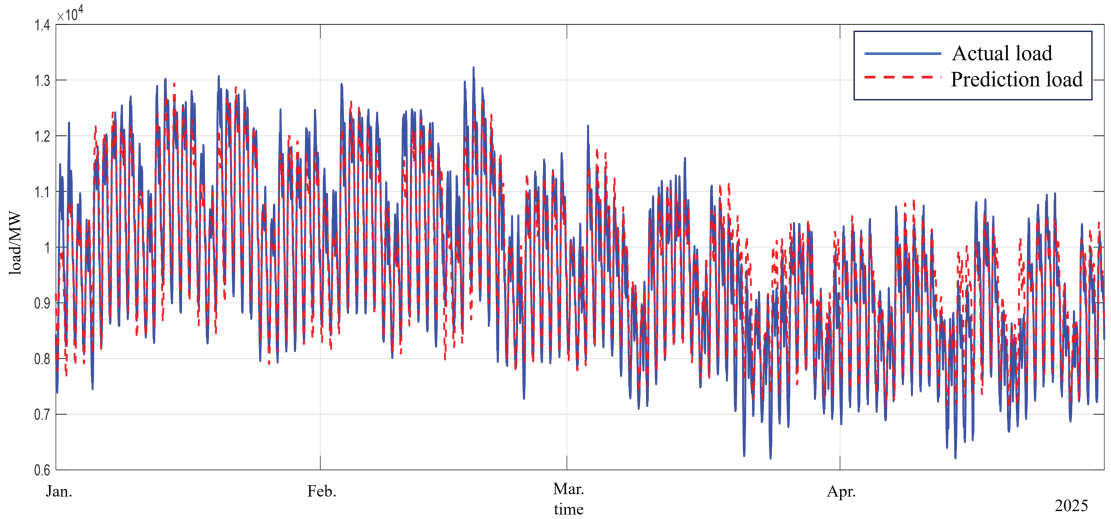


Figure 12: Prediction results of PSO-BiLSTM model (all data).

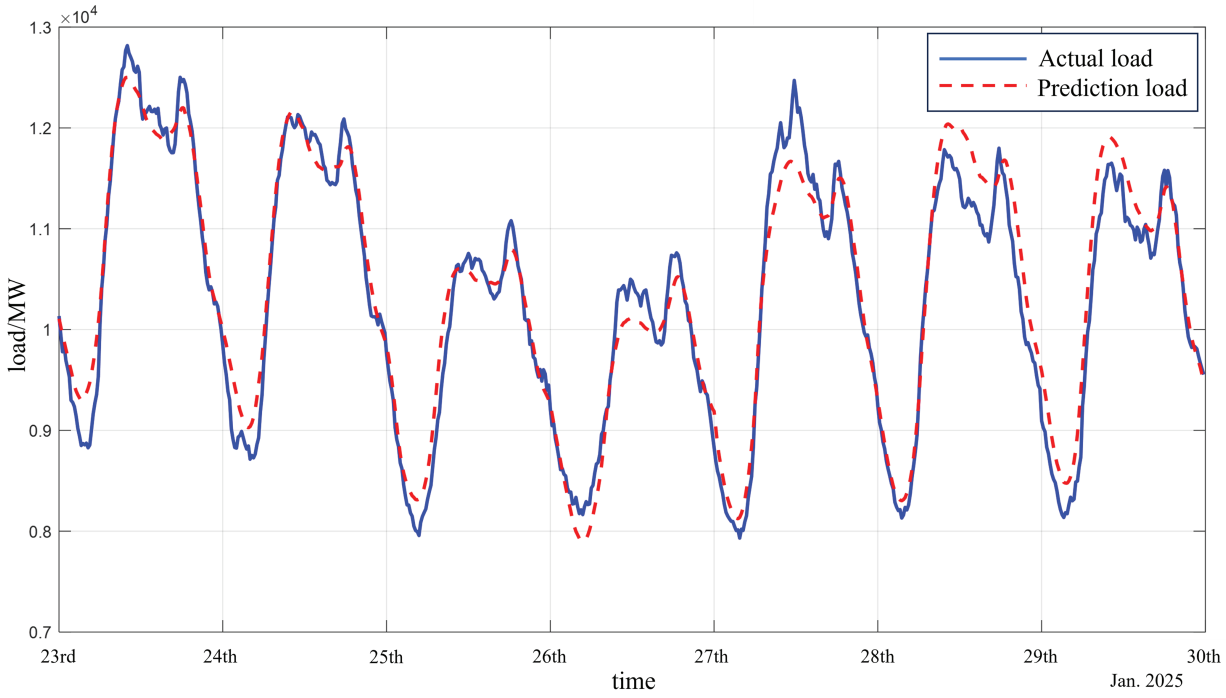


Figure 13: Prediction results of PSO-BiLSTM model (one week).

During the experiment, if the training epochs of the model are adjusted to varying degrees, there may be differences in the prediction results. Therefore, this paper sets different training epochs for various models that have not undergone PSO optimization. In addition, this article also used some classic and advanced algorithms in the field of load forecasting to conduct comparative experiments, and the evaluation metrics obtained are shown in Table 6.

Table 6: Evaluation metrics of all models.

Name of Models	Epochs	δ_{MAE}/MW	δ_{RMSE}/MW	δ_{MAPE}/%	δ_R^2
Single-layer-LSTM	5	389.25	488.49	3.8701	0.87445
	10	370.74	478.02	3.7858	0.88911
	30	359.1	467.37	3.6132	0.89375
Single-layer-BiLSTM	5	385.59	479.96	3.814	0.88922
	10	367.75	469.91	3.7632	0.90265
	30	356.34	458.86	3.549	0.90311
Double-layer-LSTM	5	385.98	487.59	3.7759	0.88093
	10	361.97	465.69	3.6858	0.89118
	30	352.01	453.21	3.5437	0.90553
Double-layer-BiLSTM	5	374.39	478.52	3.7498	0.88988
	10	355.79	463.29	3.6626	0.89227
	30	348.16	451.78	3.5035	0.90684
XGBoost	50	315.79	420.78	3.3871	0.92007

(Continued)

Table 6 (continued)

Name of Models	Epochs	δ_{MAE}/MW	δ_{RMSE}/MW	δ_{MAPE}/%	δ_{R}^2
PSO-SVM	50	302.08	412.24	3.2729	0.92165
PSO-LSTM	50	302.08	412.24	3.2729	0.92165
PSO-BiLSTM	50	295.51	395.54	3.2322	0.92877

By comparing the evaluation metrics of each group of models, it can be found that with the increase of training epochs, the predictive performance of the models also improves slightly; When the training epochs are the same, the double-layer model shows some improvement compared to the single-layer model. And due to its better feature capture performance for complex time series, BiLSTM usually outperforms LSTM in terms of performance. After optimizing the algorithm hyperparameters using PSO, the MAPE of the PSO-BiLSTM model is 3.2322%, and the R^2 is 0.92877, which is 8.93% and 2.84% higher than the single-layer BiLSTM model's 3.549% and 0.90311, respectively; Compared to the 3.5035% and 0.90684% of the double-layer BiLSTM, it has increased by 7.74% and 2.42% respectively. For the XGBoost model and PSO-SVM model, the MAPE of PSO-BiLSTM increased by 5.49% and 4.57%, respectively, and the R^2 increased by 2.05% and 0.95%, respectively. It can be seen that the PSO-BiLSTM models optimized by PSO have excellent predictive performance.

4.3 Characteristic Impact Analysis

The lagged characteristics within the multi-dimensional temporal characteristics constructed in this paper encompass three temporal dimensions: one year prior, one week prior, and one day prior. Each dimension significantly influences the load forecasting results. To investigate the importance of individual lag characteristics, this paper conducted removal experiments by sequentially removing each temporal dimension.

Three comparative experiments were established by removing the one-year-prior, one-week-prior, and one-day-prior dimensions respectively. The evaluation metrics provided in Table 7.

Table 7: Evaluation metrics of models after removing different characteristics.

Name of Models	Removing Characteristics	δ_{MAE}/MW	δ_{RMSE}/MW	δ_{MAPE}/%	δ_{R}^2
PSO-LSTM	One-year-prior	354.16	466.37	3.8529	0.87076
	One-week-prior	456.85	571.86	4.9225	0.82951
	One-day-prior	496.59	624.12	5.3217	0.81273
PSO-BiLSTM	One-year-prior	334.03	432.35	3.6005	0.9101
	One-week-prior	424.95	528.87	4.4834	0.8417
	One-day-prior	529.81	655.35	5.6557	0.7935

In the comparative experiment of removing lagged characteristics from different time dimensions, the prediction accuracy of the model was affected to varying degrees. Among them, removing the time dimension from one-year-prior had a relatively small impact on the results. The MAPE and R^2 of the PSO-LSTM and PSO-BiLSTM models were 3.8529%, 0.87076 and 3.6005%, 0.9101, respectively, which were reduced by 17.72%, 5.52% and 11.39%, 1.25% compared to the normal model; Excluding the time dimension

from one-week-prior had a relatively significant impact on the results. The MAPE and R^2 of the two models were 4.9225%, 0.82951 and 4.4834%, 0.8417, respectively. Compared with the normal model, the MAPE and R^2 decreased by 50.40%, 9.99% and 38.71%, 9.37%, respectively; Excluding the time dimension from one-day-prior had the greatest impact on the results, with MAPE and R^2 of the two models being 5.3217%, 0.81273, and 5.6557%, 0.7935, respectively. Compared to the normal model, MAPE and R^2 decreased by 62.60%, 12.49%, and 74.98%, 14.56%, respectively.

4.4 Load Forecasting with Distributed PV

The identification of distributed photovoltaics using the RF algorithm is shown in the Table 8 below for various evaluation indicators.

Table 8: Evaluation indicators for identification models based on RF.

Evaluation Indicators	Values
Accuracy	95.77%
Error	4.23%
Recall	94.44%

By identifying the characteristics of distributed PV, it is possible to distinguish whether the load includes PV output. For loads containing PV output, this paper will conduct special training and analysis, still using two models: PSO-LSTM and PSO-BiLSTM. The weekly prediction results of PSO-LSTM and PSO-BiLSTM with PV are shown in Figs. 14 and 15.

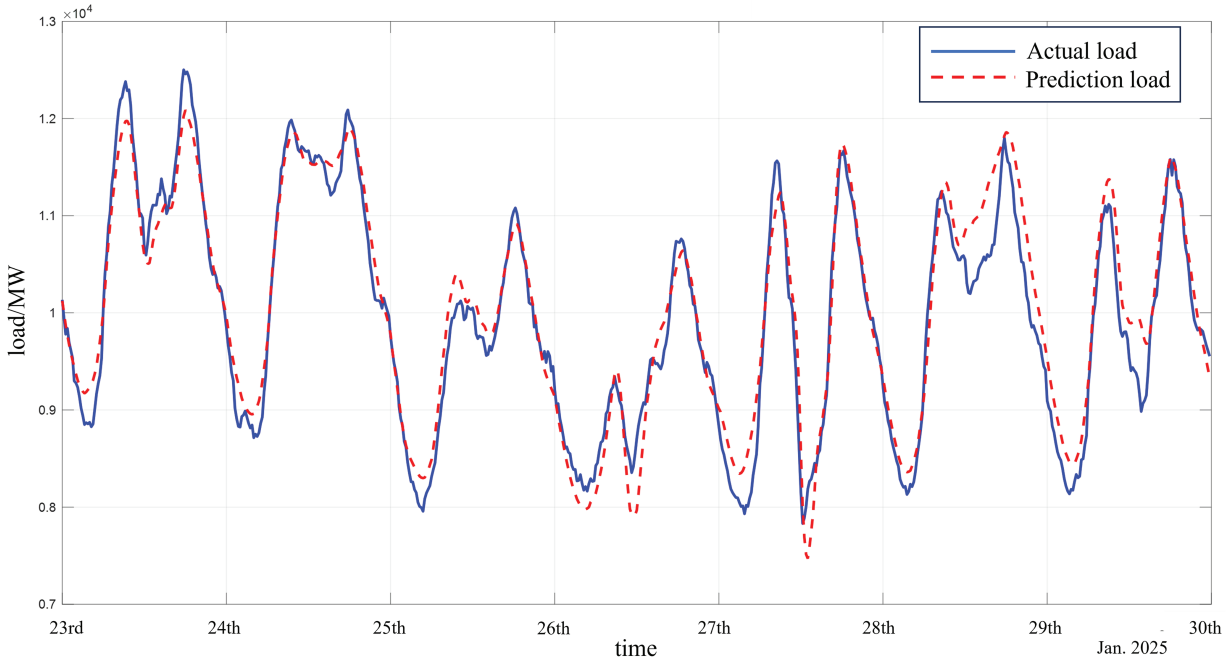


Figure 14: Prediction results of PSO-LSTM model with PV (one week).

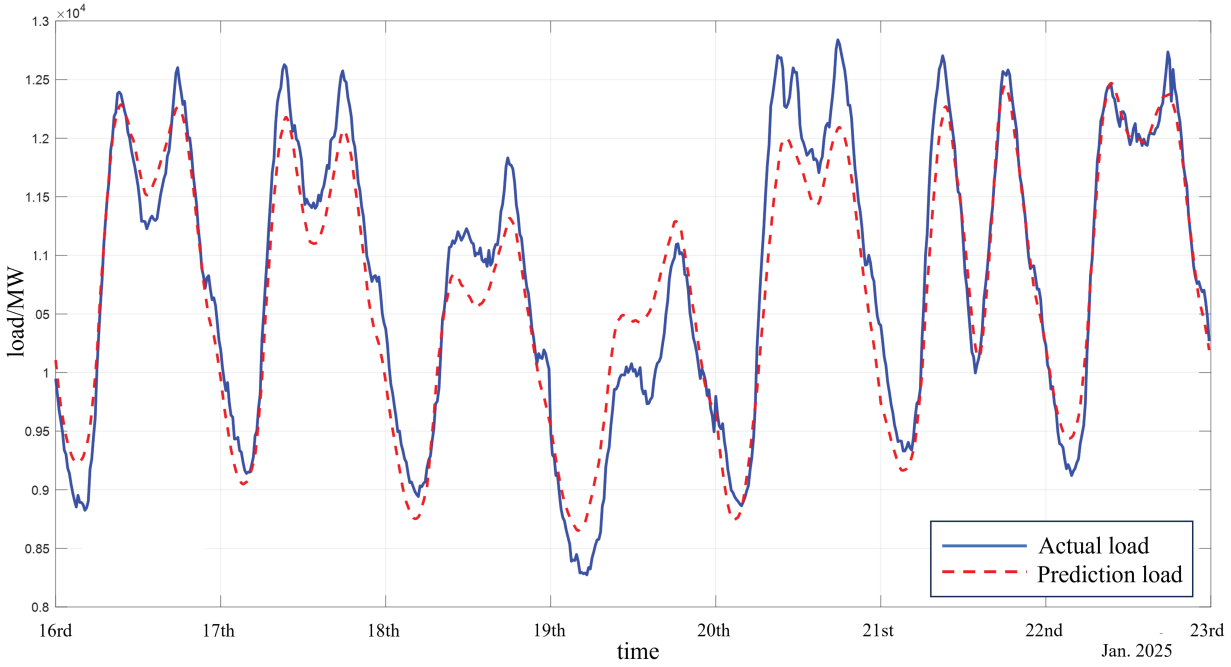


Figure 15: Prediction results of PSO-BiLSTM model with PV (one week).

Similarly, this paper uses XGBoost and PSO-SVM models for load forecasting in the case of distributed PV, and the evaluation metrics are shown in Table 9.

Table 9: Evaluation metrics of models with PV.

Name of Models	δ_{MAE}/MW	δ_{RMSE}/MW	δ_{MAPE}/%	$\delta_R{}^2$
XGBoost-pv	373.31	478.91	5.2378	0.9212
PSO-SVM-pv	359.03	461.25	5.0219	0.9354
PSO-LSTM-pv	352.26	458.28	5.1321	0.9558
PSO-BiLSTM-pv	325.19	424.67	4.3905	0.9621

According to the evaluation metrics, the MAPE and R^2 of PSO-BiLSTM are 4.3905% and 0.9621. Compared with the above three models, PSO-BiLSTM has improved by 16.18%, 12.57%, 14.45%, 4.44%, 2.85%, and 0.66%, respectively. Due to the input of PV related characteristics into the model, the accuracy of the model prediction has also increased to a certain extent, and the PSO-BiLSTM model has relatively better predictive performance.

5 Conclusion

With the rapid development of economy and technology, the proportion of new energy, especially distributed PV, on the user side is growing rapidly. This makes it difficult for the grid to obtain relevant data in a timely and accurate manner, posing a major challenge to the accuracy of traditional load forecasting. In order to address these challenges, this paper proposes a PSO-BiLSTM load forecasting model, which takes into account the characteristics of distributed PV power generation and multi-dimensional temporal characteristics. The main contributions of this paper can be summarized as follows:

By integrating lagged characteristics, cyclical characteristics and calendar characteristics, multi-dimensional temporal characteristics are constructed, which significantly enhance the ability of the model to capture complex time dependencies in load data. We propose a distributed PV identification method, which is only based on the net load behavior under different weather conditions, and can effectively identify whether users have installed distributed PV. PSO is used to automatically optimize the hyperparameters of the BiLSTM model, so as to improve the accuracy and robustness of prediction while avoiding manual optimization. Extensive experiments based on real load data show that the performance of the proposed PSO-BiLSTM model is consistently better than other models, especially in the distributed PV grid-connected scenario.

In subsequent studies, characteristics with higher correlation can be integrated into multi-dimensional temporal characteristics. Furthermore, meteorological data such as weather and temperature can be associated with temporal characteristics to further improve the correlation of characteristics at different dates. Additionally, when weather factors are considered, it is necessary to further study the impact of different climate conditions in multiple regions on the prediction results to ensure the universality of the method. More advanced and industry-leading prediction algorithms can be adopted to enhance the predictive performance of the model.

Acknowledgement: This paper was completed with the hard help of every author.

Funding Statement: This study was sponsored by Science and Technology Project of State Grid Corporation of China (kj2024-027).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Zhengji Meng, Zijian Wang, Linjun Shi; data collection: Zhengji Meng, Lei Wang, Xuekai Hu, Zijian Wang; analysis and interpretation of results: Zijian Wang; draft manuscript preparation: Zijian Wang, Linjun Shi. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu T, Li F, Zhang D, Shan L, Zhu H, Du P, et al. Intelligent load forecasting technologies for diverse scenarios in the new power systems. Renew Sustain Energy Rev. 2026;226(24):116366. doi:10.1016/j.rser.2025.116366.
2. Khan SU, Iqbal E, Khan N, Zweiri Y, Abdulrahman Y. Towards net zero energy building: AI-based framework for power consumption and generation prediction. Energy Build. 2025;331(2):115311. doi:10.1016/j.enbuild.2025.115311.
3. National Energy Administration. The year-on-year growth of electricity consumption in the whole society in 2024 is 6.8%; Jan 2025 [cited 2025 Jan 1]. [In Chinese]. Available from: https://www.nea.gov.cn/20250120/4f7f249bac714e7693adecac996d742f/c.html.
4. Xiao D. A review on risk-averse bidding strategies for virtual power plants with uncertainties: resources, technologies, and future pathways. Technologies. 2025;13(11):488. doi:10.3390/technologies13110488.
5. Zhang M, Xiao G, Lu J, Liu Y, Chen H, Yang N. Robust load feature extraction based secondary VMD novel short-term load demand forecasting framework. Electr Power Syst Res. 2025;239(7):111198. doi:10.1016/j.epsr.2024.111198.
6. Zhou H, Ai Q, Li R. Short-term multi-energy load forecasting method based on transformer spatio-temporal graph neural network. Energies. 2025;18(17):4466. doi:10.3390/en18174466.
7. Kim TG, Kwon BS, Yoon SG, Song KB. Very short-term load forecasting for large power systems with kalman filter-based pseudo-trend information using LSTM. Energies. 2025;18(18):4890. doi:10.3390/en18184890.

8. Urošević V, Savić AM. Temporal clustering for accurate short-term load forecasting using Bayesian multiple linear regression. Appl Intell. 2024;55(1):19. doi:10.1007/s10489-024-05887-z.
9. ElMenshawy MS, Massoud AM. Short-term load forecasting in active distribution networks using forgetting factor adaptive extended kalman filter. IEEE Access. 2023;11:1. doi:10.1109/ACCESS.2023.3315591.
10. Al-Musaylh MS, Deo RC, Adamowski JF, Li Y. Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF Reanalysis atmospheric predictors in southeast Queensland. Australia Renew Sustain Energy Rev. 2023;186:113702. doi:10.1016/j.rser.2023.113702.
11. Panigrahi R, Patne NR, Surya VBV, Khedkar M. Short-term load analysis and forecasting using stochastic approach considering pandemic effects. Electr Eng. 2024;106(3):3097–108. doi:10.1007/s00202-023-02135-4.
12. Beloev HI, Saitov SR, Filimonova AA, Chichirova ND, Babikov OE, Iliev IKS. Short-term electrical load forecasting based on XGBoost model. Energies. 2025;18(19):5144. doi:10.3390/en18195144.
13. Danish MU, Grolinger K. Kolmogorov–Arnold recurrent network for short term load forecasting across diverse consumers. Energy Rep. 2025;13:713–27. doi:10.1016/j.egyr.2024.12.038.
14. Lv L, Wu Z, Zhang J, Zhang L, Tan Z, Tian Z. A VMD and LSTM based hybrid model of load forecasting for power grid security. IEEE Trans Ind Inf. 2022;18(9):6474–82. doi:10.1109/tii.2021.3130237.
15. Sekhar C, Dahiya R. Robust framework based on hybrid deep learning approach for short term load forecasting of building electricity demand. Energy. 2023;268(1):126660. doi:10.1016/j.energy.2023.126660.
16. Vasileios P, Spyros T, Jason W, Nasser K. A hybrid LSTM-Transformer model for power load forecasting. IEEE Trans Smart Grid. 2025;16(3):1. doi:10.1109/TSG.2025.3535407.
17. Li Y, Li H, Li X, Cao Y, Xie Y, Zhang C. Short-term electricity load forecasting based on PowerTimeMixer. Power Syst Technol. 2025;49(10):4216–27. [In Chinese]. doi:10.13335/j.1000-3673.pst.2024.2208.
18. Buratto WG, Muniz RN, Nied A, Gonzalez GV. Seq2Seq-LSTM with attention for electricity load forecasting in Brazil. IEEE Access. 2024;12:30020–9. doi:10.1109/ACCESS.2024.3365812.
19. Xu H, Hu F, Liang X, Zhao G, Abugunmi M. A framework for electricity load forecasting based on attention mechanism time series depthwise separable convolutional neural network. Energy. 2024;299(1):131258. doi:10.1016/j.energy.2024.131258.
20. Shi J, Teh J. Load forecasting for regional integrated energy system based on complementary ensemble empirical mode decomposition and multi-model fusion. Appl Energy. 2024;353(1):122146. doi:10.1016/j.apenergy.2023.122146.
21. Zulfiqar M, Kamran M, Rasheed MB, Alquthami T, Milyani AH. A hybrid framework for short term load forecasting with a navel feature engineering and adaptive grasshopper optimization in smart grid. Appl Energy. 2023;338(1):120829. doi:10.1016/j.apenergy.2023.120829.
22. GB/T 35663-2017. Basic terminology of weather forecast. Beijing, China: Standards Press of China; 2017. [In Chinese].
23. Mu Y, Kong L, Zheng G, Su Z, Wang G. A short-term load forecasting method considering multiple feature factors based on long short-term memory and an improved temporal convolutional network. Eng Appl Artif Intell. 2025;159(C):111649. doi:10.1016/j.engappai.2025.111649.
24. Hussain A, Kumar K, Ravikumar G. Bayesian-optimized bidirectional long-short-term memory network for wind power forecasting with uncertainty quantification. Electr Power Syst Res. 2026;251:112261. doi:10.1016/j.epsr.2025.112261.