Sensors and Materials, Vol. 37, No. 7 (2025) 2971–2988
MYU Tokyo

S & M 4097

Novel Fluctuation- and Smooth-channel-based Deep Learning Model Driven by Multidimensional Information-aware Sensor for Electric Load Forecasting

Wan-Qin Ding,[1] Yun-Xin Zhou,[1] Wen-Dong Wang,[1] and Li-Bo Han[2*]

[1]Changdian (Zhangye) Energy Development Co., Ltd., Gansu Zhangye 734034, China
[2]Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing 100190, China

(Received April 25, 2025; accepted June 17, 2025)

Keywords: load forecasting, deep learning, evolutionary algorithm, power system, abnormal value detection, sensor

The accurate prediction of power load is a prerequisite for maintaining the supply-demand stability of the power system. To improve the accuracy of load forecasting, we proposed a novel fluctuation- and smooth-channel-based deep learning model for load forecasting. First, the multidimensional information-aware sensor technology is used to collect wind speed, temperature, and humidity information through the analog-to-digital converter (ADC) digital-information-acquisition, filter-circuit-noise-elimination, and amplification circuits. The microcontroller in the sensor is employed to process the output of the characteristics of the impact of the load data information. Then, the interquartile range was employed to detect the abnormal values in the load data, and the missing values caused by their removal were filled and eliminated by cubic spline interpolation to enhance the quality of the load data. Second, the long short-term memory (LSTM) model based on a fluctuation channel and a smooth channel was constructed, which can autonomously distinguish the fluctuation period from the smooth period in the data, fully exploiting the fluctuation information in the fluctuation period and the subtle changes in the smooth period. Additionally, an improved catch fish optimization algorithm was specifically designed to optimize the hyperparameters of the prediction model, enhancing its ability to characterize complex load sequences. Finally, the proposed method and model were validated through case studies. The results demonstrated that compared with existing load prediction models, the proposed model achieved a mean absolute percentage error below 3% and a goodness-of-fit exceeding 98%, effectively capturing the fluctuation trend of complex load sequences.

1. Introduction

Under the background of large-scale landscape grid connection, enhancing the power load forecasting accuracy has evolved into a critical technical requirement for ensuring grid stability.(1,2) The high-precision data acquisition system based on multidimensional sensor

*Corresponding author: e-mail: hanlibo@mail.iee.ac.cn
https://doi.org/10.18494/SAM5711

ISSN 0914-4935
https://myukk.org/

technology enables the real-time capture of pivotal meteorological parameters, which, when processed through advanced analytics and deep learning models, provide decision-makers with actionable load predictions.(3,4) Constructing an integrated framework that unifies environmental sensing, data analytics, and predictive modeling has become pivotal for achieving dynamic supply–demand equilibrium in contemporary power systems.(5)

Currently, load prediction can be categorized into four types: physical models, traditional statistical models, artificial intelligence models, and hybrid models. First, physical models are typically constructed by exploring the intrinsic relationships between historical data and physical parameters to predict load.(6) Typical physical forecasting methods include the unit consumption, elasticity coefficient, and load density methods.(7) For example, Wang *et al.*(8) employed the unit consumption and elasticity coefficient methods to forecast and analyze future long-term electricity consumption trends, although they require a substantial amount of data. However, Liu *et al.*(9) argued that physical models are heavily dependent on expert knowledge, and obtaining the relevant data can be time-consuming.

In contrast to physical models, statistical models mainly analyze historical data to reveal the relationship between load fluctuations and time.(10) For example, Li *et al.*(11) combined a genetic algorithm with the auto-regression and moving average (ARMA) model to develop a more robust model with a prediction error lower than that of the ARMA model. In addition, Bikcora *et al.*(12) integrated the ARMA model with the generalized autoregressive conditional heteroskedasticity framework, with experimental results indicating that this model outperformed other traditional ARMA models. Rendon-Sanchez and de Menezes(13) proposed two types of structural combination using seasonal exponential smoothing as the base model and applied them to forecast short-term electricity demand. However, they ignored that this model might rely on seasonal patterns in historical data.

Models such as random forest,(14) extreme learning machine,(15) support vector machine,(16) and deep learning fall into the category of computational intelligence models. For example, Chaturvedi *et al.*(17) noted that load forecasting typically uses periodic time series information as input sequences and applied a recurrent neural network (RNN) specifically designed for handling sequence information in load prediction. To improve the generalization and effectiveness of the forecasting model, Fang *et al.*(18) proposed a short-term electric load forecasting method based on the Fourier multi-layer perception (FMLP)-Transformer model. It is suitable for scenarios with the large-scale integration of intermittent renewable energy into the power grid and demonstrates excellent adaptability and generalization ability.

Experimental results showed that hybrid models often have better prediction performance than single models.(19) For example, Li *et al.*(20) proposed a prediction model based on bidirectional long short-term memory (BiLSTM)-Transformer model and experimentally demonstrated its superiority over single prediction models. Guo *et al.*(21) proposed a new short-term load forecasting method for the power system based on the graph convolutional network and long short-term memory (LSTM). Their experimental results showed that their proposed method can fully utilize the effect of multidimensional data and effectively improve the load forecasting accuracy and training efficiency.

On the basis of the issues identified, deep-learning-based power load forecasting methods still face several urgent challenges. First, current models struggle to effectively capture and predict significant fluctuations in power system load. Second, integrating data from different sources and time scales, as well as dynamically optimizing model structures to adapt to various forecasting scenarios, requires further research and exploration.

To address these challenges, we propose a multidimensional information-aware sensor technology combined with a hybrid power load forecasting model that integrates multichannel deep learning, attention mechanisms, and evolutionary algorithms to enhance the forecasting performance. Initially, the proposed sensor technology incorporates acquisition circuits, operational amplifiers, and filtering modules to accurately capture temperature, humidity, and wind speed data. A sophisticated data preprocessing pipeline is implemented to eliminate outliers and smooth the raw dataset, ensuring robust input for subsequent predictive analytics. Subsequently, a multichannel LSTM network processes data across different time scales. Attention mechanisms are introduced to enhance the model's ability to capture critical information, and an improved fishing optimization algorithm is employed for parameter optimization. The main innovations and contributions of this study are reflected in the following aspects.

- A multidimensional information-aware sensor technology for power load forecasting was constructed, enabling the precise and stable acquisition of temperature, humidity, and wind speed datasets.
- A deep learning network architecture containing fluctuation and smooth channels was proposed, which, after the parallel processing of power load data with different volatilities, uses attention mechanisms to effectively capture and integrate critical information.
- An anomaly correction technique that improves the quality of model input data by detecting, removing, and completing anomalies in the original data, followed by smoothing, was introduced, thereby providing more accurate data support for the model.

The remaining structure of this paper is as follows: in Sect. 2, the details of the data preprocessing methods are given; the fundamental principles and model construction of power load forecasting are explored in Sect. 3; in Sect. 4, we describe the entire power load forecasting process; in Sect. 5, the superiority and effectiveness of the proposed model and algorithm are validated using multiple examples; and the main research findings, contributions, and limitations of this study are summarized in Sect. 6.

2. Interquartile Range and Cubic Spline Interpolation-based Data Preprocessing Driven by Multidimensional Information-aware Sensor

In the field of electrical load forecasting, data preprocessing is a crucial step to ensuring the effectiveness and accuracy of the models. However, if the abnormal data is not properly addressed, it may negatively impact the model's predictive performance. Therefore, we employed the interquartile range (IQR) and cubic spline interpolation (Spline) to correct anomalies in the electrical load data. For this study, power load data collected at 15-min intervals from a specific region in southern China during February were selected as the input dataset for analysis. The steps for anomaly correction are as follows.

2.1 Outlier detection and removal with the following mathematical model

Outliers are detected using I_{QR}.

$$\begin{cases} I_{QR} = Q_3 - Q_1 \\ L_{OW} = Q_1 - 1.5 I_{QR} \\ U_P = Q_3 + 1.5 I_{QR} \end{cases} \tag{1}$$

Q_1 is the minimum value of the lowest 25% of the data points in the dataset, known as the first quartile; Q_3 is the maximum value of the highest 25% of the data points, known as the third quartile. L_{OW} and U_P are the distinct lower and upper boundaries, respectively. The sample is considered an outlier and is removed from the dataset when it is below L_{OW} or above U_P.

2.2 Missing value completion

After the removal of outliers, some missing values may appear in the dataset. To address these missing values, we employ spline cubic interpolation as follows.

(1) Define the set of interpolation points $\{x_0, x_1, ..., x_n\}$ and their corresponding function values $\{y_0, y_1, ..., y_n\}$. Here, x_i represents the x-coordinate of the i_{th} known data point and y_i denotes the y-coordinate corresponding to x_i.

(2) For each interval, define a cubic polynomial:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, \tag{2}$$

where $S_i(x)$ represents the vertical coordinate on the spline curve corresponding to x_i. The spline curve must pass through each data point (x_i, y_i), meaning that for each i, $S_i(x)=y_i$; a_i, b_i, c_i, d_i are the coefficients to be determined.

(3) To ensure the continuity of the curve, each x_i must satisfy the following conditions: $S_{i-1}(x_i) = S_i(x_i)$, $S'_{i-1}(x_i) = S'_i(x_i)$ and $S''_{i-1}(x_i) = S''_i(x_i)$.

(4) To determine the coefficients c_i and d_i, boundary conditions must also be applied. We used natural boundary conditions, where the second derivatives at the endpoints are 0.

$$S''(x_0) = 0 \wedge S''(x_n) = 0 \tag{3}$$

(5) The spline cubic interpolation function $S(x)$ is the combination of all these cubic polynomials, with each polynomial $S_i(x)$ defined over its respective interval [x_i, x_i+1]:

$$S(x) = \sum_{i=0}^{n} c_i (x - x_i)^3 + S_i(x). \tag{4}$$

Given the known missing value on the *x*-coordinate, the corresponding $S(x)$, i.e., the input value for the missing data, is obtained. Outlier handling using IQR and Spline provides high-quality and highly relevant data inputs for the power load forecasting model.

However, it is not enough to use previous load data for load forecasting; some indirect features are also needed to assist the improvement of the forecasting accuracy. For example, temperature, humidity, and wind speed show a correlation with electrical load, and in this study, we will use three types of advanced sensor for data collection. The process of using multidimensional information-aware sensor technology to collect weather and electric load characteristics is illustrated in Fig. 1.

The selection and parameters of each component in the circuit diagram are shown in Table 1.

3. Formulation of Fluctuation-smoothing Feature-fusion-based Dual-channel LSTM

LSTM is a special type of RNN structure. It has three gates, namely, the Forget gate, Input gate, and Output gate.

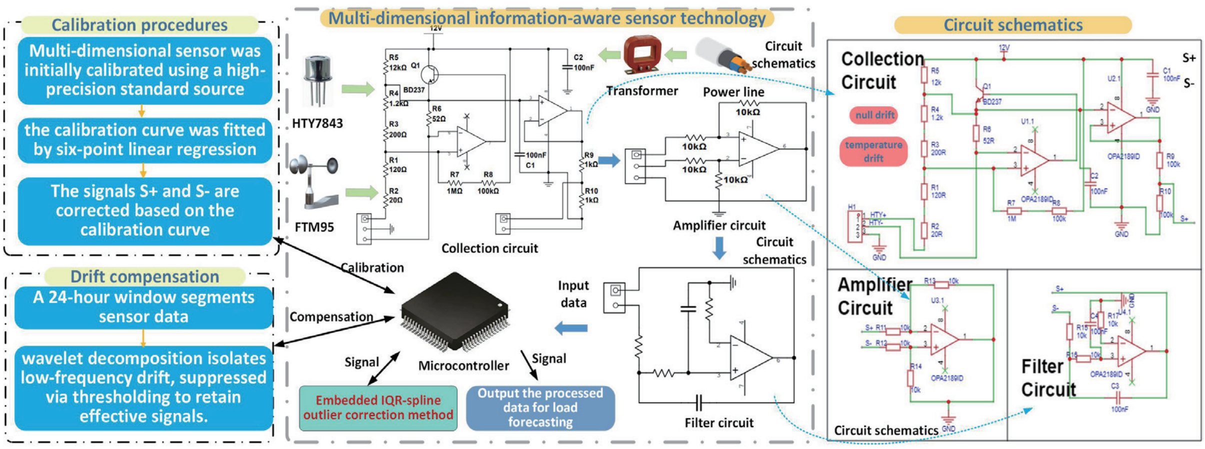

Fig. 1. (Color online) Multidimensional information-aware sensor technology process.

Table 1
Selected components and their parameters.

Component type	Type	Parameter
Resistance	RC0603	±1% accuracy
Audion	BD237	NPN type, rated voltage 80 V, rated current 2 A, rated power 25 W, DC current gain 40 times
Operational amplifier	OPA2188	Low noise, 1.7 μVpp with 0.1–10 Hz noise
Capacitance	C0603	±10% accuracy, X7R medium
Temperature and humidity sensor	HTY7843	±0.3 °C, 1–10 Hz sampling frequency
Wind sensor	FTM95	±1.5% F.S., 10–100 Hz sampling frequency

The Forget gate determines which information to discard. Through a sigmoid function, it calculates a vector between 0 and 1 on the basis of the previous hidden state and current input, indicating the extent of forgetting for each unit state, where 0 means complete forgetting and 1 means complete retention.

The Input gate determines which new information to store. It consists of a sigmoid function and a tanh function. The sigmoid function decides which information to update, while the tanh function generates candidate new memory content.

The Output gate determines which information to output. Similarly, it utilizes a sigmoid function and a tanh function. First, the sigmoid function decides which part to output, then the tanh function processes the cell state to bring its value between −1 and 1, and finally, the two are multiplied to obtain the final output.

However, the precision of LSTM decreases when dealing with highly volatile data. Therefore, in this study, the LSTM is divided into two channels on the basis of the volatility rate η. The calculation principle of the volatility rate is as follows.

The index η, used to characterize the volatility of load data, is defined as

$$\eta = \frac{\int x(t)dt}{(x_0 + x_N) \times N / 2} - 1, \tag{5}$$

where x_0 represents the starting point of the first data and x_N represents the endpoint of the last data. When the volatility of the interval is low, η tends to 0 and *vice versa*.

The volatility channel is mainly used to process segments of load data with high volatility and outputs the hidden layer state h_t^F containing volatility information.

$$\begin{cases} f_t^F = \sigma\left(W_{fx}^F \cdot x_t^F + W_{fh}^F \cdot h_{t-1}^F + b_f^F\right) \\ i_t^F = \sigma\left(W_{ix}^F \cdot x_t^F + W_{ih}^F \cdot h_{t-1}^F + b_i^F\right) \\ o_t^F = \sigma\left(W_{ox}^F \cdot x_t^F + W_{oh}^F \cdot h_t^F + b_o^F\right) \end{cases} \tag{6}$$

Here, F represents the parameters of the fluctuation channel; the control of the weight parameters and bias parameters selectively forgets, inputs, or outputs features related to the fluctuation at different time steps.

The smooth channel works the same as the fluctuation channel.

The mathematical principle of the fusion gate mechanism is

$$W = Fullyconnect\left(W^F h_t^F + W^S h_t^S + b\right), \tag{7}$$

where h_t^{FSD} represents the global feature hidden state; W^F and W^S are the weight parameters of the fusion gate mechanism; b_o is the bias parameter; W is the final prediction result output through a fully connected layer. W^F and W^S undergo joint optimization using Forget gate parameters and other modular components through gradient-descent-based algorithms and backpropagation. The proposed fusion gate mechanism dynamically calibrates these weighting factors on the basis of real-time input characteristics, enabling the model to adaptively harmonize feature contributions from dual-channel encoding pathways. This parametric interaction ensures optimal feature fusion by striking a data-driven balance between temporal continuity and abrupt variations, thereby significantly enhancing the forecasting accuracy in complex load scenarios.

The fusion gate achieves the integration of fluctuation and smooth features by adaptively adjusting the relevant weights. The FSD-LSTM-Attention model is illustrated in Fig. 2.

4. Catch Fish Optimization Algorithm Based on Leader-guided Capture Strategies

The catch fish optimization algorithm (CFOA) simulates fishermen catching fish in a pond to optimize algorithm search and parameters. CFOA consists of two main phases: exploration and exploitation. During the exploration phase, the algorithm searches in reference to individual and collective fishing behaviors. In the exploitation phase, it simulates fishermen surrounding fish schools and employs collective fishing strategies for exploitation.

4.1 Exploration phase

Each fisherman possesses a certain level of fishing experience, enabling them to adjust their exploration direction and position in accordance with the fishing situations of other fishermen. If

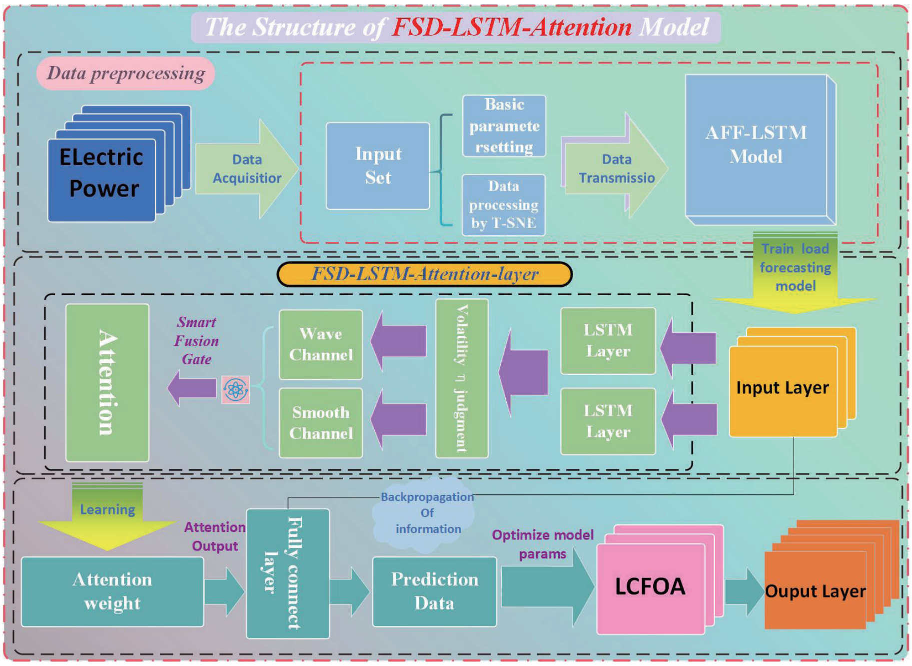

Fig. 2. (Color online) FSD-LSTM-Attention model.

their own catching situation is more favorable, they will continue searching in the original direction; if the reference catching situation is more advantageous, they will alter their original direction, thereby moving towards a better catching position.

$$\begin{cases} Fisher_i^{T+1} = Fisher_i^T + \left(Fisher_p^T - Fisher_i^T\right) \times Exp + r_s \times s \times Dis \times \sqrt{|Exp|} \times \left(1 - \dfrac{EFs}{MaxEFs}\right) \\ Exp = \dfrac{fit_i - fit_p}{fit_{worst} - fit_{best}} \end{cases} \tag{8}$$

Exp is the experience analysis value obtained by the fisherman with any reference position p; fit_{worst} and fit_{best} are the worst and best fitness values at the T_{th} iteration, respectively; Dis is the Euclidean distance between the i_{th} and p_{th} populations; r_s is a random number; s is a random vector of dimension d; R is the exploration range; EFs represents the current number of iterations; $MaxEFs$ denotes the maximum number of iterations for the algorithm.

Fishermen use fishing nets to expand their fishing capabilities and cooperate with other fishermen, randomly forming groups of three to four individuals who collaborate. Leveraging each person's unique mobility, they explore the area more precisely. The position update formula is

$$Fisher_i^{T+1} = Fisher_i^T + r_2 \times \left(mean(Fisher_c^T) - Fisher_i^T\right) + \left(1 - \frac{2 \times EFs}{MaxEFs}\right)^2 \times r_3, \tag{9}$$

where c is a group of three to four individuals; $Centre_c$ is the target area surrounded by group c; r_2 is the speed at which fishermen approach the central area, with values varying among individuals in the range from 0 to 1; r_3 is the offset of movement, with values in the range of (−1, 1).

4.2 Exploitation phase

During the development phase, all fishermen search under a unified strategy, purposefully driving and concentrating the fish for capture. In the process of luring, with the fish as the center, fishermen gradually gather from the center to the periphery, leading to decreasing density and range as they move outward. This distribution pattern is simulated using a Gaussian distribution, with the update formula

$$\begin{cases} Fisher_i^{T+1} = Gbest + GD\left(0, \dfrac{r_4 \times \sigma \times \left|mean\left(Fisher\right) - Gbest\right|}{3}\right), \\ \sigma = \sqrt{\left(2\left(1 - \dfrac{EFs}{MaxEFs}\right) \middle/ \left(\left(1 - \dfrac{EFs}{MaxEFs}\right)^2 + 1\right)\right)}, \end{cases} \tag{10}$$

where *GD* is the Gaussian distribution function with a mean of 0 and a variance decreasing from 1 to 0 as the number of iterations increases; *Gbest* is the global best position; r_4 is a random number in the range of {1, 2, 3}.

CFOA struggles in early exploration as it is prone to falling into local optima, hindering later development. Improved fishing optimization algorithm (LCFOA) incorporates two improvements to address this problem: chaotic mapping for better population initialization and a leader-guided strategy for enhanced search.

4.2.1 Initialization of population using logistic-tent chaotic mapping

First, the first population of fishermen is initialized with the following formula:

$$Fisher_{1,j} = \left(ub_j - lb_j\right) \cdot r + lb_j, \tag{11}$$

where $Fisher_{1,j}$ represents the position of the first population in dimension *j*th; ub_j and lb_j are the upper and lower bounds of dimension *j*th, respectively; *r* is a random number between 0 and 1.

To ensure that subsequent populations are evenly distributed in the solution space, the logistic-tent algorithm is introduced to generate chaotic sequences for population initialization:

$$Fisher_i = \begin{cases} \mathrm{mod}\left(r_x \cdot Fisher_{i-1} \cdot \left(1 - Fisher_{i-1}\right) + (4 - r_x) \cdot \dfrac{Fisher_{i-1}}{2} \right), & mean\left(Fisher_{i-1}\right) < 0.5 \\ \mathrm{mod}\left(r_x \cdot Fisher_{i-1} \cdot \left(1 - Fisher_{i-1}\right) + (4 - r_x) \cdot \dfrac{1 - Fisher_{i-1}}{2} \right), & mean\left(Fisher_{i-1}\right) \geq 0.5 \end{cases} \tag{12}$$

where *i* is the population sequence number with a range of $i \in \{2, 3, 4 \cdots N\}$; r_x is a control parameter with values in the range of (0, 4).

4.2.2 Leader-guided fishing strategy

The leader-guided fishing strategy enhances the original algorithm by using the top three fishermen in terms of fishing efficiency to guide position updates, replacing random reference selection. This accelerates convergence speed and precision, offering numerous potential optimal solution regions. It boosts later-stage development and improves the global optimal solution accuracy. The modified mathematical model is

$$Leader^T = \left(Fisher_{best}^T + Fisher_{best-1}^T + Fisher_{best-2}^T\right) / 3, \tag{13}$$

$$Fisher_i^{T+1} = Fisher_i^T + \left(Leader^T - Fisher_i^T\right) \times Exp + r_s \times s \times R, \tag{14}$$

where $Leader^T$ is the position of the leader in the *T*th iteration; $Fisher_{best}^T$, $Fisher_{best-1}^T$ and $Fisher_{best-2}^T$ are the top three individuals ranked by fishing efficiency.

The structure and pseudocode of the LCFOA algorithm are shown in Fig. 3.

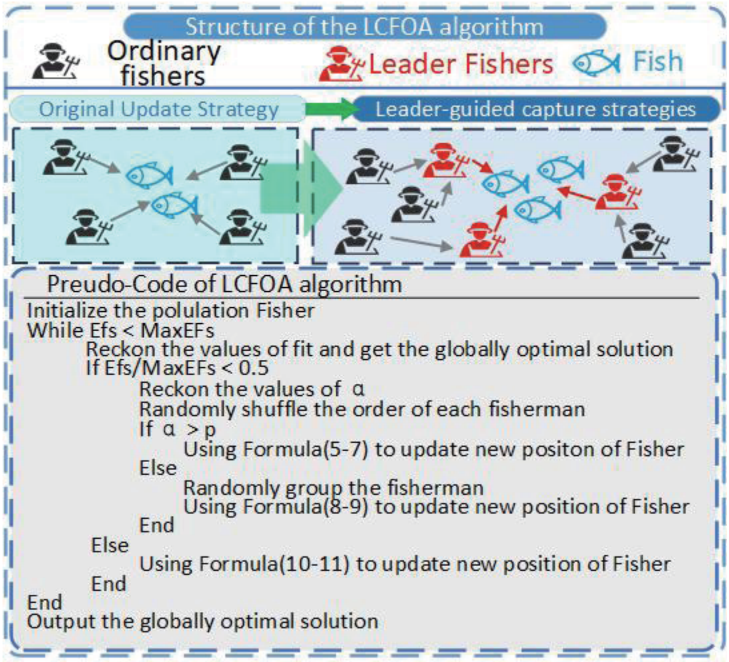

Fig. 3. (Color online) Structure and pseudocode of the LCFOA algorithm.

5. Case Studies

The power load forecasting process is illustrated in Fig. 4. The load forecasting process of the fluctuation- and smooth-channel-based deep learning model is as follows.

(1) Data preprocessing. Remove outlier data from the power load data set using the interquartile range and complete the data using cubic spline interpolation.

(2) Initialize model and algorithm parameters. Initialize the parameters of the FSD-LSTM-Attention and LCFOA models, as shown in Table 2.

(3) Optimize model hyperparameters with LCFOA. Use the LCFOA algorithm to optimize the hyperparameters of the FSD-LSTM-Attention model to enhance the accuracy of power load prediction.

(4) Predict load results. With the best hyperparameters, use the FSD-LSTM-Attention model to forecast the load for the next day.

5.1 Case 1: Power load forecasting using different algorithms

In Case 1, under the condition of keeping the prediction model unchanged, the original CFOA algorithm, Grey Wolf Optimization (GWO), and Great Wall Construction Algorithm (GWCA) were selected as comparative optimization algorithms to search for the hyperparameters of the LSTM-Attention model to examine the effectiveness and feasibility of the LCFOA algorithm in power load forecasting. Furthermore, a comparative analysis was conducted between the

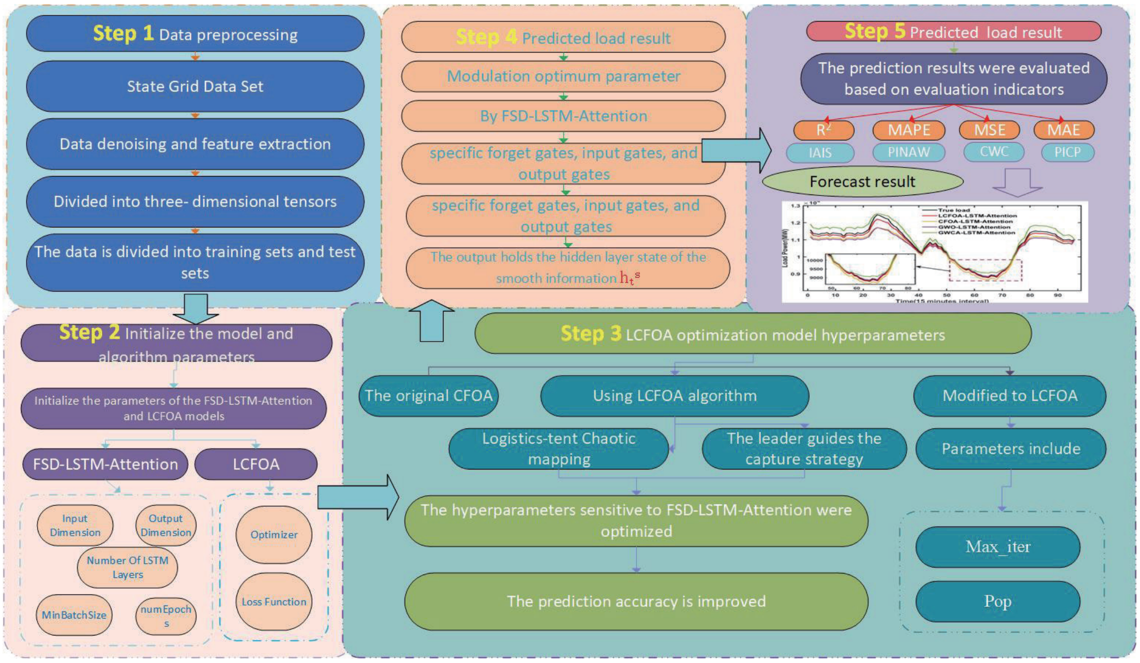

Fig. 4. (Color online) Power load forecasting process.

Table 2
FSD-LSTM-Attention parameter settings.

Parameter		Value
FSD-LSTM-Attention	Input dimension	5
	Output dimension	1
	Epochs	40
	Min batch size	24
	Number of LSTM layers	2
	Optimizer	Adam
	Loss function	*MAE*

classical grid search and random search methodologies. To ensure a rigorous evaluation of each method's performance, the optimization algorithm configuration was standardized as follows: the maximum number of iterations was set to 50, the population size was set to 20, the initial learning rate range was defined as [0.001, 0.5], and the hidden units range was specified as [20, 200]. Since random optimization and grid optimization differ among iteration methods, the maximum number of iterations was set to 1000 for both random search and grid search, which is the maximum number of iterations multiplied by population size.

The power load prediction results for Case 1 obtained using each model are shown in Fig. 5. As presented in Fig. 5(a), specifically, the deviation between the values predicted using the LCFOA-based model and the actual values was notably lower. In contrast to the CFOA-based, GWO-based, and GWCA-based models, the load prediction curve obtained using the LCFOA-

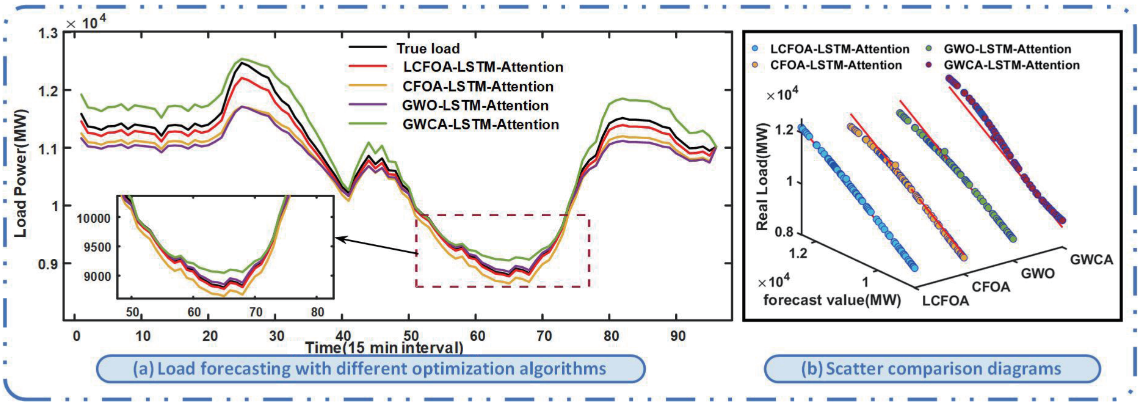

Fig. 5. (Color online) Predicted results obtained from different optimization algorithms for Case 1.

based model showed remarkable consistency with the actual load curve. In the region with the largest prediction error, where the original load was approximately 1.24×10^4 MW, the LCFOA-based model predicts a value of around 1.22×10^4 MW, resulting in a maximum prediction error of 1.6%. This outcome highlights the superiority and competitiveness of the LCFOA algorithm in load forecasting. Figure 5(b) reveals that, compared with other algorithms, the prediction results obtained with the LCFOA-based model mostly aligned along the 1:1 line, indicating predictions closest to the true values.

The evaluation results of power load forecasting for Case 1 are detailed in Table 3. Table 3 shows that the R^2 value of LCFOA is 99.90%, which is 1.9, 1.1, and 0.8% higher than the results obtained respectively by CFOA, GWO and GWCA. Its optimization time only increases by 128.3, 189.24, and 76.93 s compared with those of the CFOA-, GWO-, and GWCA-based algorithms, respectively. This indicates a very high degree of fit between the predicted curve and the actual load curve. Additionally, the LCFOA-based model reduced *MSE* by approximately 50% compared with the other algorithms, demonstrating higher prediction accuracy. For grid search, its search effect is poor, and various indicators have a large gap compared with the prediction results of the parameters obtained with the metaheuristic algorithm. This is because grid search is a traversal method with extremely high time cost. However, with the maximum number of searches set to 1000 in this study, grid search cannot traverse the hyperparameter space, resulting in poor returns. Random search, on the other hand, is slightly inferior to the proposed LCFOA because its completely random sampling makes it difficult to effectively capture the optimal combination of hyperparameters of the model.

5.2 Case 2: Power load forecasting using different models

Case 2 primarily validates the improvement effect of the proposed FSD-LSTM-Attention model. Therefore, under the optimization algorithm of LCFOA, the load forecasting results are compared with those of CNN and bidirectional gated recurrent unit (BIGRU).

The power load prediction results obtained by each model for Case 2 are illustrated in Fig. 6. Figure 6(a) shows the prediction results of each model. Specifically, the FSD-LSTM-Attention

Table 3
Load forecasting evaluation indexes of each algorithm.

Model	Assessment index				Wall-clock time (s)
	R^2 (%)	*MAE*	*MAPE* (%)	*MSE*	
LCFOA-LSTM-Attention	99.90	91.4	0.82	11221	1981.47
CFOA-LSTM-Attention	98.00	249	2.25	81074	1853.17
GWO-LSTM-Attention	98.80	245.6	2.15	97918	1792.23
GWCA-LSTM-Attention	99.10	217.9	1.99	60640	1904.54
Grid Search-LSTM-Attention	94.26	433.7	3.30	133726	2031.24
Random Search-LSTM-Attention	98.22	219.4	3.05	77104	1742.19

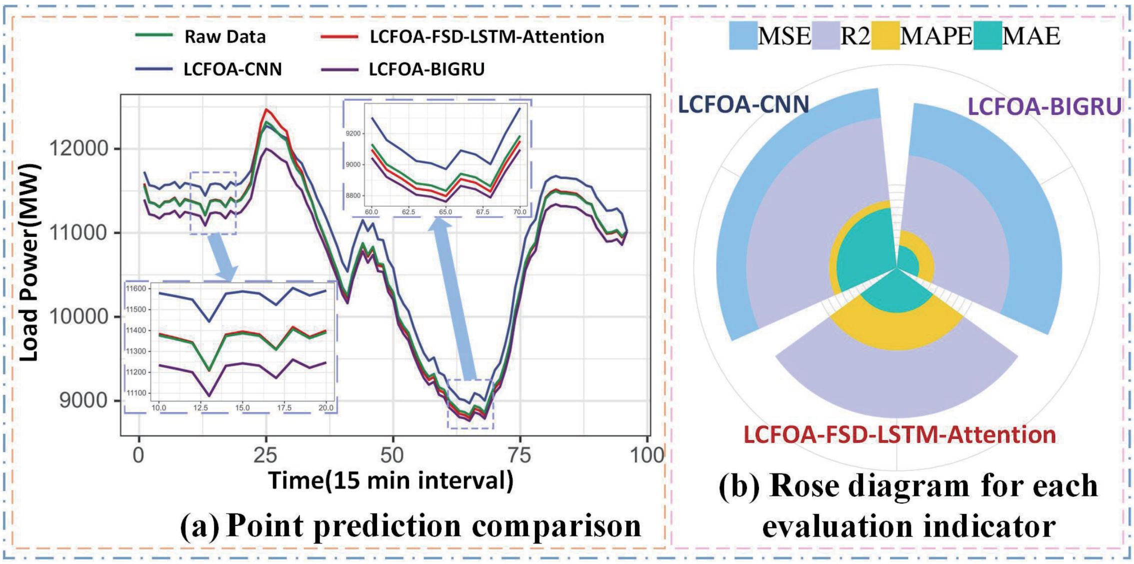

Fig. 6. (Color online) Predicted results using different optimization algorithms.

model demonstrated outstanding predictive capabilities. During phases of relatively stable power load, the proposed model's prediction curve captured subtle variations in the actual load curve, with the prediction curve almost aligning with the actual curve. In Fig. 6(b), the radar chart shows the relative evaluation metrics of the three models. The arc radius corresponding to the R^2 metric for the FSD-LSTM-Attention model was the longest, indicating the best fitting effect; the arcs for *MAE* and *RMSE* are the shortest, especially with *MSE* being nearly zero, suggesting minimal prediction errors for the proposed model. Table 4 provides the evaluation results of each model for Case 2.

The data in Table 4 indicate that the FSD-LSTM-Attention model demonstrated the most competitive performance in terms of prediction error and directional accuracy. The LCFOA-FSD-LSTM-Attention model outperformed the other models across all evaluation metrics, showcasing strong competitiveness. Specifically, the R^2 metric of the LCFOA-FSD-LSTM-

Table 4
Prediction indicators of different models.

Prediction model	Assessment criterion			
	MSE	R^2 (%)	*MAPE* (%)	*MAE*
LCFOA-FSD-LSTM-Attention	1686.1	99.93	0.26	28.1
LCFOA-CNN	51916	98.3	2.05	216.43
LCFOA-BIGRU	24990	99.59	1.13	127.1

Attention model exceeded 99%, indicating robust data-fitting capabilities.

To further validate the predictive performance of the proposed LCFOA-FSD-LSTM-Attention model across diverse scenarios, we employ hourly resolved datasets collected over consecutive 10-day periods spanning four distinct seasons (March, June, September, and December) in South China. Leveraging a temporal sliding window approach, the data of the first six days is utilized to forecast values for the subsequent four days. The benchmark models include TCN,[22] Autoformer,[23] and GRU-Attention,[24] with hyperparameters of all models optimized using LCFOA while maintaining consistency with parameter configurations reported in their respective reference papers. Table 5 presents the forecasting performance metrics of different models across the four seasonal datasets.

As shown in Table 5, the proposed LCFOA-FSD-LSTM-Attention model demonstrates superior predictive performance across all four seasonal datasets when evaluated using the four forecasting metrics. Specifically, during spring, the model achieves R^2 score improvements of 6.90, 4.49, and 2.20% compared with LCFOA-TCN, LCFOA-Autoformer, and LCFOA-GRU-Attention, respectively. This performance advantage persists across other seasons, with the proposed model consistently outperforming baseline architectures under diverse climatic conditions. Notably, the R^2 metric of the LCFOA-FSD-LSTM-Attention model remains above 85% throughout all seasonal evaluations. These results collectively validate the model's robust adaptability and practical applicability across temporal and environmental variations.

Figure 7 shows the prediction curves of different models in the four different seasons. As illustrated in Fig. 7, the LCFOA-GRU-Attention model demonstrates the closest alignment with actual load profiles across all four seasons. During spring, summer, and winter, the proposed model exhibits remarkable fitting accuracy at most time points, with predicted load trends showing strong consistency with observed patterns. Although the proposed model exhibits relatively large errors in autumn compared with other seasons, it still maintains the smallest prediction error among the three traditional models.

5.3 Case 3: Proof of effectiveness of the proposed dual-channel models

Case 3 employs hourly industrial load data collected over a consecutive 10-day period in September from South China. The initial 6-day subset serves as the training set, while the subsequent 4-day period constitutes the forecasting horizon. A comparative evaluation was conducted among three proposed architectures: single-channel smooth, single-channel fluctuation, and combined dual-channel models. Furthermore, a rolling forecast mechanism was implemented to iteratively refine the pretrained model by incorporating sequential feedback from both predicted values and actual load observations, thereby enhancing practical

Table 5
Prediction indicators of different seasons.

Season	Prediction model	Assessment criterion			
		R^2 (%)	*MSE*	*MAPE* (%)	*RMSE*
Spring	LCFOA-TCN	87	232661	2.92	482
	LCFOA-Autoformer	89	234155	2.82	483
	LCFOA-GRU-Attention	91	183282	2.33	428
	LCFOA-FSD-LSTM-Attention	93	154600	2.32	393
Summer	LCFOA-TCN	93	296802	2.91	544
	LCFOA-Autoformer	92	398935	3.58	631
	LCFOA-GRU-Attention	93	267363	2.68	517
	LCFOA-FSD-LSTM-Attention	98	82071	1.53	286
Autumn	LCFOA-TCN	75	450588	4.27	671
	LCFOA-Autoformer	74	476246	4.54	690
	LCFOA-GRU-Attention	76	433194	4.19	658
	LCFOA-FSD-LSTM-Attention	85	50	4.27	84
Winter	LCFOA-TCN	89	326616	3.73	571
	LCFOA-Autoformer	85	449963	4.47	670
	LCFOA-GRU-Attention	88	327085	3.73	571
	LCFOA-FSD-LSTM-Attention	89	287592	3.58	536

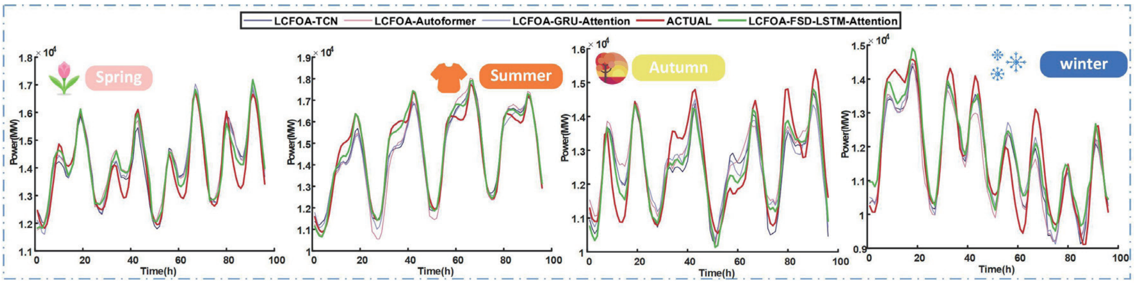

Fig. 7. (Color online) Predicted results for different seasons.

applicability in operational forecasting scenarios.

To further validate the efficacy of the proposed interquartile range and cubic spline interpolation-based data preprocessing, Table 6 shows the results of a comparative analysis of forecasting performance across the four model configurations: original LSTM model, single-channel smooth, single-channel fluctuation, combined dual-channel models before data preprocessing, and combined dual-channel models after data preprocessing.

As shown in Table 6, single-channel models (smooth and fluctuation variants) demonstrate suboptimal performance when trained independently, with *MSE* values exceeding 2×10^5. This limitation arises from the inability of isolated channels to capture holistic load power characteristics. The comparative analysis of pre-preprocessing and post-preprocessing performance reveals severely degraded accuracy in raw data scenarios, where outliers impede the model's ability to discern critical patterns within the dataset. After implementing preprocessing and fusing smooth/fluctuation channel information, the model exhibits substantial

Table 6
Prediction indicators under different improvement strategies.

Prediction Model	Assessment criteria			
	R^2 (%)	*MSE*	*MAPE* (%)	*RMSE*
LSTM	86	245766	3.72	495
Single-channel smooth	87	232661	2.92	482
Single-channel fluctuation	83	234155	3.55	483
Combined dual-channel models before data preprocessing	84	658859	5.31	811
Combined dual-channel models after data preprocessing	92	162583	2.24	403

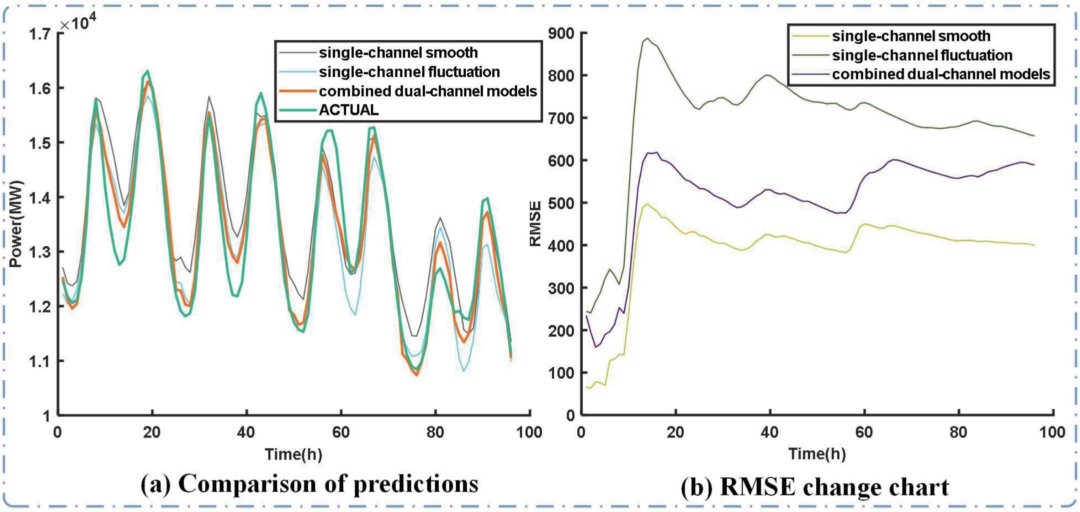

Fig. 8. (Color online) Predicted results under different improvement strategies.

performance gains: the R^2 metric exceeds 90%, while the other evaluation indices show marked reductions. These results confirm the superior performance of combined dual-channel architectures in industrial load forecasting contexts.

Figure 8 illustrates the forecasting outcomes under various enhancement strategies, along with the dynamic evolution of *RMSE* metrics during rolling window evaluations. As depicted in Fig. 8(a), both single-channel models exhibit suboptimal forecasting performance with significant prediction errors observed across most time points. These errors peak during abrupt changes in the load profile. In contrast, the combined dual-channel model demonstrates superior feature extraction capability, accurately capturing load curve characteristics while maintaining acceptable prediction accuracy even during volatile periods. Figure 8(b) presents the dynamic evolution of *RMSE* metrics across 96 rolling window evaluations. Notably, the combined dual-channel model maintains *RMSE* values below 500 throughout the testing period, exhibiting a decreasing trend as the sliding window progresses iteratively. These findings confirm the practical effectiveness of our proposed model in operational forecasting scenarios, addressing critical requirements of industrial load modeling applications.

6. Conclusions

In response to the uncertainty and variability of power load brought about by renewable energy integration, we proposed a novel deep learning model with a two-channel LSTM and attention mechanism driven by multidimensional information-aware sensor technology. Additionally, an improved LCFOA algorithm was proposed to significantly enhance the algorithm's global search capability and convergence performance. The main findings of this study are summarized as follows.

- In this study, we integrated the multidimensional information-aware sensor technology, built the sensor ADC acquisition, filter, and amplification circuits, and finally processed the information through a microcontroller chip to realize the accurate acquisition and processing of temperature, humidity, and wind speed.
- The LCFOA algorithm showed significant advantages in load forecasting accuracy. In Case 1, the load forecasting curve obtained by LCFOA closely matched the actual load curve with a maximum prediction error of only 1.6%, indicating its competitiveness in load forecasting.
- The LCFOA-FSD-LSTM-Attention model exhibited notably favorable performance in terms of load forecasting error and directional accuracy. In Case 2, the proposed model achieved an R^2 value exceeding 99.9%, *MAPE* below 0.3%, and competitive *MSE* and *MAE* values.

While we have made significant achievements in power load forecasting in this study, there are limitations such as the lack of real-time application and validation in actual large-scale power systems. At the same time, the fused sensor technologies are also subject to hardware sampling errors and sampling real-time effects, making the sampling results subject to inherent sensor acquisition errors. Our future research will be focused on addressing these limitations.

Acknowledgments

This work was supported by the project of Research and Application of Health Status Perception in the Construction Power Supply System of Zhangye Pumped Storage Power Station in Gansu Province (No. JG-ZY-0123001).

References

1 J. X. Hu, J. X. Hu, W. H. Hu, D. Cao, X. W. Sun, J. J. Chen, Y. H. Huang, Z. Chen, and F. Blaabjerg: Renew. Energy **225** (2024) 13. https://doi.org/10.1016/j.renene.2024.120253
2 Z. F. Liu, Y. H. Huang, S. R. Zhang, X. F. Luo, X. R. Chen, J. J. Lin, Y. Tang, L. Guo, and J. X. Li: Appl. Energy **377** (2025) 29. https://doi.org/10.1016/j.apenergy
3 X. L., Wang, H. Wang, B. Bhandari, and L. M. Cheng: Int. J. Precis. Eng. Manuf. Green Technol. **11** (2024) 963. https://doi.org/10.1007/s40684-023-00537-0
4 Z. F. Liu, S. X. Zhao, X. F. Luo, Y. H. Huang, R. Z. Gu, J. X. Li, and L. L. Li: Appl. Energy **379** (2025) 31. https://doi.org/10.1016/j.apenergy.2024.124918
5 L. L. Li, B. X. Ji, Z. T. Li, M. K. Lim, K. Sethanan, and M. L. Tseng: Appl. Energy **378** (2025) 20. https://doi.org/10.1016/j.apenergy.2024.124853
6 Y. Q. Jiang, T. L. Gao, Y. X. Dai, R. Q. Si, J. Hao, J. Zhang, and D. W. Gao: Appl. Energy **328** (2022) 13. https://doi.org/10.1016/j.apenergy.2022.120120

7 H. R. Zhao, X. Y. Han, and S. Guo: Neural Comput. Appl. **30** (2018) 1811. https://doi.org/10.1007/s00521-016-2799-1
8 J. Z. Wang, L. Y. Zhang, and Z. W. Li: Appl. Energy **305** (2022) 18. https://doi.org/10.1016/j.apenergy.2021.117911
9 S. Liu, Z. M. Kong, T. Huang, Y. Du, and W. Xiang: Neural Networks **173** (2024) 12. https://doi.org/10.1016/j.neunet.2024.106150
10 D. M. Bui, P. D. Le, M. T. Cao, T. T. Pham, and D. A. Pham: Int. J. Green Energy **17** (2020) 382. https://doi.org/10.1080/15435075.2020.1761810
11 Y. T. Li, K. Wu, and J. Liu: Knowledge-Based Syst. **269** (2023) 13. https://doi.org/10.1016/j.knosys.2023.110489
12 C. Bikcora, L. Verheijen, and S. Weiland: Sustainable Energy Grids Networks **13** (2018) 148. https://doi.org/10.1016/j.segan.2018.01.001
13 J. F. Rendon-Sanchez and L. M. de Menezes: Eur. J. Oper. Res. **27** (2019) 916. https://doi.org/10.1016/j.ejor.2018.12.013
14 I. A. Ibrahim and T. Khatib: Energy Convers. Mange. **138** (2017) 413. https://doi.org/10.1016/j.enconman.2017.02.006
15 M. Guermoui, K. Bouchouicha, N. Bailek, and J. W. Boland: Energy Convers. Mange. **245** (2021) 21. https://doi.org/10.1016/j.enconman.2021.114569
16 C. F. Shang and P. C. Wei: Renew. Energy **127** (2018) 269. https://doi.org/10.1016/j.renene.2018.04.067
17 S. Chaturvedi, E. Rajasekar, S. Natarajan, and N. McCullen: Energy Policy **168** (2022) 13. https://doi.org/10.1016/j.enpol.2022.113097
18 B. L. Fang, L. Xu, Y. J. Luo, Z. X. Luo, and W. Li: Energy Rep. **12** (2024) 3405. https://doi.org/10.1016/j.egyr.2024.09.023
19 Q. Yan, Z. Y. Lu, H. Liu, X. T. He, X. H. Zhang, and J. L. Guo: Energy Build. **297** (2023) 16. https://doi.org/10.1016/j.enbuild.2023.113396
20 Z. Y. Li, X. W. Zhang, and W. Gao: Energy **311** (2024) 13. https://doi.org/10.1016/j.energy.2024.133418
21 Y. X. Guo, Y. Li, X. B. Qiao, Z. Y. Zhang, W. F. Zhou, Y. J. Mei, J. J. Lin, Y. C. Zhou, and Y. Nakanishi: IEEE Trans. Smart Grid **13** (2022) 3481. https://doi.org/10.1109/tsg.2022.3173964
22 H. T. Han, J. S. Peng, J. Ma, S. L. Liu, and H. Liu: Sci. Rep. **15** (2025) 14530. https://doi.org/10.1038/s41598-025-98224-4
23 Y. Q. Jiang, T. L. Gao, Y. X. Dai, R. Q. Si, J. Hao, J. Zhang, and D. W. Gao: Appl. Energy **328** (2022) 120120. https://doi.org/10.1016/j.apenergy.2022.120120
24 Q. Tao, J. Lv, and J. Wu: Int. J. Low-Carbon Technol. **19** (2024) 798. https://doi.org/10.1093/ijlct/ctae034