Computer Modeling in
Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.081823

ARTICLE

Short-Term Photovoltaic Power Prediction Based on IWOA-TCN-BiGRU-MATT

Guanglin Sha[1], Xinwei Cong[1], Yunzhao Wu[1], Dinghong Chen[1], Bo Wang[2] and Hengrui Ma[2,*]

[1]China Electric Power Research Institute Co., Ltd., Beijing, China
[2]School of Electrical and Automation, Wuhan University, Wuhan, China

*Corresponding Author: Hengrui Ma. Email: henrym@whu.edu.cn

Received: 09 March 2026; Accepted: 27 May 2026; Published: 28 September 2026

ABSTRACT: To address the challenges of significant nonlinearity, intricate temporal interdependencies, and the tendency to get stuck in local optima during parameter tuning in short-term photovoltaic (photovoltaic, PV) power forecasting, this paper puts forward a hybrid model called IWOA-TCN-BiGRU-MATT. This model fuses the IWOA (Improved Whale Optimization Algorithm, IWOA) with the TCN (Temporal Convolutional Network, TCN), BiGRU (Bidirectional Gated Recurrent Unit, BiGRU), and MATT. By leveraging IWOA to boost parameter optimization efficiency and integrating TCN's ability to extract multi-scale features, BiGRU's bidirectional temporal modeling, and MATT's emphasis on key features, the model aims to realize high-precision PV power prediction. Based on measured data from a PV power station in Central China, experimental findings show that the proposed model surpasses competing methods under various meteorological scenarios, demonstrating improved prediction accuracy under different weather conditions.

KEYWORDS: Improved whale optimization algorithm; temporal convolutional network; bidirectional gated recurrent unit; multi-head attention mechanism; photovoltaic power prediction; intelligent optimization

1 Introduction

The global energy structure is accelerating its transition to low-carbon development. As a crucial component of clean and renewable energy, PV generation capacity continues to expand globally and plays a pivotal role in realizing carbon-neutrality objectives. Yet its intrinsic variability makes short-horizon forecasting indispensable for grid scheduling and safe system operation for maintaining grid load balance, improving photovoltaic consumption rates, and reducing curtailment rates. The accuracy of such predictions directly impacts the reliability of power system dispatching and operational safety [1,2].

Current research on short-term photovoltaic power prediction mainly faces two core challenges. First, the comprehensiveness and accuracy of feature extraction are insufficient. Existing single prediction models (e.g., ELM, GRU), LSTM (Long Short-Term Memory Network, LSTM) can fit the nonlinear and temporal features of photovoltaic power to a certain extent but fail to simultaneously capture nonlinear relationships, temporal dependencies, and multi-source meteorological interactions [3–5]. Although some hybrid models (e.g., VMD-KELM [6], TCN-BiLSTM) have achieved improvements, they still have obvious shortcomings: EMD is prone to mode mixing; traditional VMD requires manual setting of the number of decompositions; models without attention mechanisms struggle to focus on key temporal features under complex weather conditions; and single GRU models ignore bidirectional temporal dependence, failing to fully utilize the correlation between historical and future data [7–11]. These deficiencies are particularly prominent in complex scenarios such as cloudy or rainy weather, giving rise to a marked increase in prediction deviations. However, these models still lack a unified framework that simultaneously addresses feature extraction, temporal dependency modeling, and global parameter optimization.

Second, the generalization capability achieved through model parameter optimization and multi-dimensional features is inadequate. Traditional intelligent optimization algorithms (e.g., WOA (Whale Optimization Algorithm, WOA), DBO (Dance Behavior Optimization, DBO) [12,13], AHA (Adaptive Hopping Algorithm, AHA)) often suffer from limited global search capabilities, making them prone to local optima. This hinders their ability to efficiently optimize key parameters in hybrid models (e.g., TCN's convolution kernels and dilation coefficients, BiGRU 's hidden unit number, ELM's kernel parameters and regularization coefficients). Additionally, some studies fail to fully integrate multi-dimensional meteorological features (e.g., key meteorological parameters such as solar radiant energy, surrounding temperature, air moisture, and wind speed magnitude) with power station operation data (e.g., component status, energy consumption level), or adopt imprecise feature selection methods (e.g., MIC (Maximal Information Coefficient, MIC), Spearman Correlation Coefficient [14–16]). This results in poor generalization of prediction models, which are difficult to adapt to the needs of different seasons and power stations.

In summary, current short-term PV power prediction faces three core challenges: strong nonlinearity of PV power, complex temporal dependencies, and susceptibility to local optima in model parameter optimization. Existing approaches comprise both standalone models (e.g., ELM, GRU) and partially integrated hybrid architectures (e.g., VMD-KELM (Variational Mode Decomposition-Kernel Extreme Learning Machine, VMD-KELM), TCN-BiLSTM)—have obvious limitations: insufficient comprehensiveness and accuracy in feature extraction, as well as weak adaptability in model parameter optimization and multi-dimensional feature fusion. To address these issues, a composite forecasting framework termed IWOA–TCN–BiGRU–MATT is designed, where an enhanced whale-inspired optimizer is combined with temporal convolution, bidirectional recurrent modeling, and multi-head attention modules. Its key contributions are as follows:

1. An improved whale optimization algorithm is proposed by integrating Logistic chaotic mapping, Lévy flight and Golden Sine strategies. These modifications respectively enhance the initial population diversity, global exploration capability and local exploitation performance of the standard WOA. Taking validation-set RMSE as the fitness function, IWOA is adopted to optimize key model parameters, such as TCN convolution kernels, BiGRU hidden units and MATT attention heads.
2. Multi-Module Feature Capture: The temporal convolution layers capture localized sequential structures through dilated causal filters and residual pathways. Parallel BiGRU branches extract forward–backward temporal cues, while the attention module adaptively highlights pivotal intervals such as high-irradiance hours.
3. Optimized Data Handling: DBSCAN removes outliers, cubic spline interpolation (irradiance-power correlation) fills missing values, and wavelet thresholding reduces noise; Spearman analysis selects 5 highly relevant features to cut redundancy and enhance data quality.

The structure of this paper is arranged in the following sequence: Section 2 elaborates on the research methodologies, Section 3 conducts analyses of case studies, Section 4 performs analysis and discussion, and Section 5 summarizes the entire research.

2 Materials and Methods

2.1 Temporal Convolutional Network (TCN)

TCN is a variant of CNNs (Convolutional Neural Networks, CNNs) specifically designed for time-series data. It overcomes the limitations of traditional CNNs—such as their inability to capture long-range dependencies in time-series tasks—and the drawbacks of Recurrent Neural Networks (RNNs/LSTMs/GRUs), such as weak parallel computing capabilities and low training efficiency [17,18]. TCN has shown excellent performance in time-series forecasting tasks. For PV power forecasting, its dilated convolution can effectively capture the multi-scale fluctuation features of PV power caused by irradiance changes and cloud cover, while the residual connection can alleviate the gradient vanishing problem in deep network training, which is more suitable for extracting local temporal features of 15-min short-term PV sequences.

Causal Convolution: Ensures that the output at the current time step only depends on the current and previous inputs, avoiding the leakage of future information. Meanwhile, a 1D fully convolutional network is used to maintain consistent input and output lengths. Let the input time series be x_t and the convolution kernel be ω; the output y_t of causal convolution can be expressed as:

$$y_t = \sum_{i=0}^{k-1} \omega_i \cdot x_{t-i} \tag{1}$$

where: K denotes the length of the convolution kernel, ensuring that i in x_{t-i} does not exceed the range before the current time step.

Dilated Convolution: introduces spacing between convolution kernels, allowing the network to capture long-range dependencies, which is particularly useful for modeling rapid fluctuations in photovoltaic power caused by sudden irradiance changes without increasing computational complexity Its mathematical expression is:

$$y_t = \sum_{i=0}^{k-1} \omega_i \cdot x_{(t-d\cdot i)} \tag{2}$$

Residual Connection: By introducing residual connections, TCN can better train deep networks. Each residual block in this structure contains two convolutional layers with dilation factors, which adopt causal convolution and ReLU activation function, respectively. After each convolution, weight normalization and Dropout regularization are applied to improve the stability of network training, prevent overfitting, and alleviate the gradient vanishing problem. Fig. 1 displays the residual module.

2.1.1 Bidirectional Gated Recurrent Unit (BiGRU)

The BiGRU block essentially integrates a "bidirectional recurrent architecture" with a "gated recurrent unit". Its overall workflow consists of three core steps: the module includes two parallel branches (forward GRU unit and backward GRU unit) that process time-series data from different directions. Each time step in the figure (e.g., the middle column corresponding to x_t) contains two GRU units (upper and lower), which correspond to forward and backward propagation respectively. Their computation follows the "gating mechanism" the combination of reset gate and update gate in GRU, with the core being the dynamic control of "retaining historical information" and "integrating new information" [19,20].

$$r_t = \sigma\left(W_r x_t + U_r h_{t-1} + b_r\right) \tag{3}$$

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1} + b_z\right) \tag{4}$$

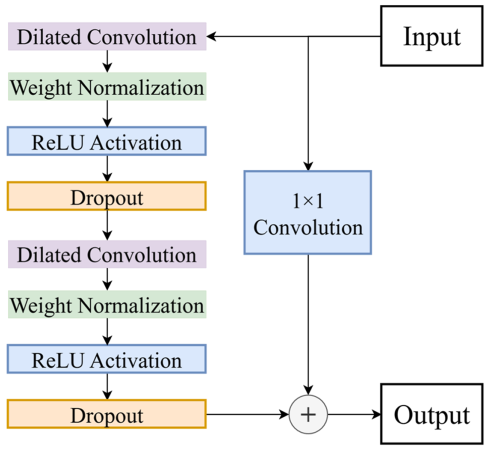

Figure 1: Residual module.

In this context, W_r and U_r act as weight matrices, with W_r transforming the input x_t and U_r transforming the hidden state h_{t-1} from the prior time step. W_z and U_z also belong to weight matrices, corresponding to the processing operations of x_t and h_{t-1} respectively. b_r refers to the bias parameter of the reset gate, and b_z is the bias parameter of the update gate. The Sigmoid function (σ) maps the output to the interval (0, 1), which plays a role in adjusting how much information is kept or forgotten. The structure of the GRU network can be seen Fig. 2.

$$\begin{cases} \overrightarrow{h}_t = \mathrm{GRU}\left(x_t, \overrightarrow{h}_{t-1}\right) \\ \overrightarrow{h}_t = \mathrm{GRU}\left(x_t, \overleftarrow{h}_{t-1}\right) \\ h_t = m\overrightarrow{h}_{t-1} + n\overleftarrow{h}_{t-1} + c_t \end{cases} \tag{5}$$

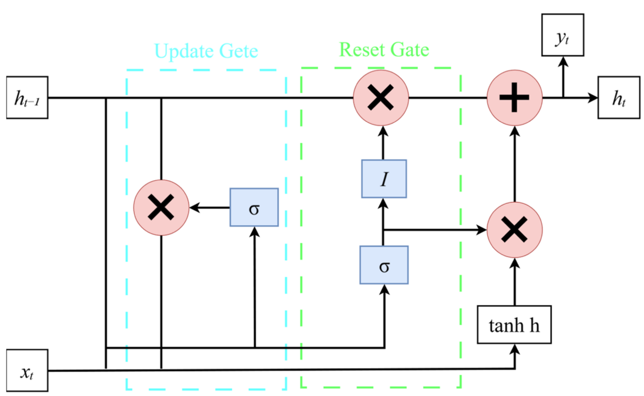

Figure 2: GRU network structure.

Within the given equations: GRU (·) signifies the computational procedure of the GRU. Specifically, $\overrightarrow{h}_t$ is the output of the forward GRU hidden layer at time t, while $\overleftarrow{h}_t$ is the output of the backward GRU hidden layer at the same time step t. Additionally, m denotes the output weight associated with the forward GRU hidden layer, n represents the output weight of the backward GRU hidden layer, and c_t corresponds to the hidden layer bias for h_t.

At each time step t, the forward and backward GRU units produce their own respective hidden states, which are denoted as $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ respectively. The module fuses these two states through a concatenation operation to form the final feature h_t containing bidirectional temporal information. This fusion method fully preserves the independence of bidirectional features, avoids information redundancy or loss, and provides a rich feature foundation for the subsequent MATT mechanism to calculate "key time-step weights". Compared with unidirectional GRU and LSTM, BiGRU can simultaneously capture the forward and backward temporal dependencies of PV sequences, and fully mine the correlation between historical irradiance changes and future power fluctuations. Controlled experiments on the same dataset show that BiGRU reduces the RMSE of the baseline model by 11.7% compared with unidirectional GRU, verifying its superior performance in temporal feature extraction for PV forecasting. The architecture of the BiGRU is depicted in Fig. 3.

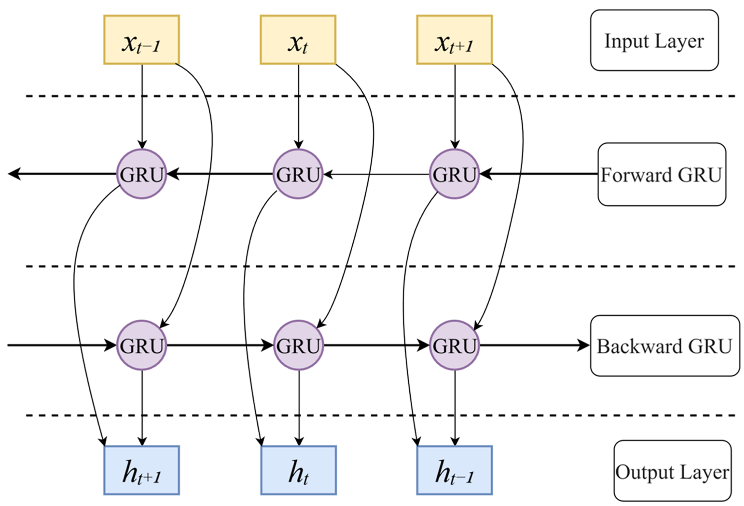

Figure 3: BiGRU architecture.

2.1.2 Multi-Head Attention (MATT)

As an extension of the standard attention mechanism, the multi-head attention mechanism splits the model's dimension into several heads, where all heads compute attention concurrently yet yield distinct representations; the results are then concatenated. This design allows the model to learn different attention patterns in various representation subspaces, thereby capturing more comprehensive information. For instance, some heads may focus on short-term local correlations in the data, while others excel at mining long-term global dependencies—enabling the model to capture richer and more diverse information [21–24].

Within multi-head attention, three vectors are involved: the query vector Q, key vector K, and value vector V, all derived from the input vector It at time t. Based on these three vectors, computation via the scaled dot-product attention function yields the output below:

$$Attention\left(Q, K, V\right) = softmax\left(\frac{QK}{\sqrt{d_m}}\right)V \quad (6)$$

Unlike the Luong attention mechanism, when the dimension d_m is large, the scaled dot-product attention function experiences notable variations in its dot-product outputs. To counter this effect, the dot product is adjusted by the scaling factor $\sqrt{d_m}$.

In the multi-head mechanism, Q, K, and V are linearly projected h times, resulting in:

$$\begin{cases} Q_i^{'} = QW_i^Q, i = 1, 2, \cdots, h \\ K_i^{'} = KW_i^K, i = 1, 2, \cdots, h \\ V_i^{'} = VW_i^V, i = 1, 2, \cdots, h \end{cases} \tag{7}$$

here, $W_i^Q \in R^{d_m \times d_K}$, $W_i^K \in R^{d_m \times d_K}$ are learnable weights, with $W_i^V \in R^{d_m \times d_K}$. Therefore, the single-head attention for each group of projected $Q_i^{'}$, $K_i^{'}$, $V_i^{'}$ is calculated as follows:

$$\text{head}_i = \text{Attention}\left(Q_i^{'}, K_i^{'}, V_i^{'}\right) \tag{8}$$

All single-head attention results are first concatenated, then projected through the learnable weight $W^O \in R^{d_m \times d_K}$ to obtain the final output of multi-head attention. This approach enables the model to capture dependencies across different dimensions of the input sequence via its multiple "attention heads," improving the model's ability to represent information effectively. Fig. 4 depicts the structure of the multi-head attention unit.

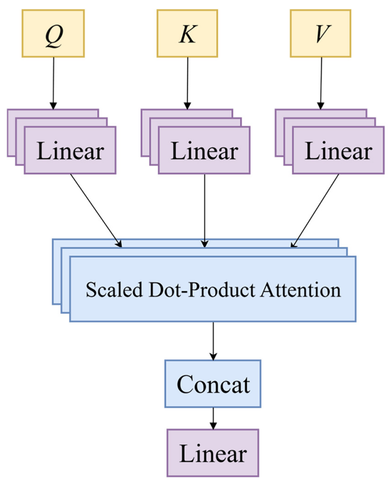

Figure 4: Architecture of the multi-head attention unit.

Photovoltaic power generation data is inherently complex, driven by numerous interacting factors and marked by strong temporal characteristics. The incorporation of multi-head attention allows the model to discern intricate dependencies across different time horizons. When combined with TCN and BiGRU, the architecture orchestrates a cohesive framework: TCN provides robust local feature extraction, BiGRU facilitates bidirectional temporal understanding, and multi-head attention dynamically weights the most influential time steps. This proves especially valuable during critical phenomena such as weather shifts or variations in solar exposure, enabling the model to prioritize these pivotal moments, which significantly enhances forecasting accuracy and produces results that closely reflect actual power dynamics.

2.2 Improved Whale Optimization Algorithm

2.2.1 Whale Optimization Algorithm (WOA)

The WOA is a metaheuristic algorithm inspired by the foraging behavior of humpback whales, particularly their unique "bubble-net" hunting technique and social interactions. It addresses complex optimization problems by mathematically modeling three key behaviors: encircling prey, executing spiral bubble-net movements, and randomly searching for prey. The algorithm is characterized by its simple structure, minimal parameter requirements, and strong convergence performance.

First, population initialization is performed: n search individuals (whales) are randomly generated, and the position of each individual is a dim-dimensional vector (aligning with the solution of the optimization task). The position must satisfy the maximum and minimum bounds of the variables $[lb, ub]$:

$$X_i = lb + rand\,(1, dim) \cdot (ub - lb) \tag{9}$$

where X_i is the position of the i-th individual, and rand(1, dim) is a random vector in the interval [0, 1]; lb and ub are the lower and upper bounds of the search space, respectively.

The algorithm sets the position of the current best individual as the "prey" position, while other individuals move closer to it. The position update formula is as follows:

$$\begin{cases} D = |C \cdot X^*(t) - X(t)| \\ X(t+1) = X^*(t) - A \cdot D \end{cases} \tag{10}$$

here, $X^*(t)$ denotes the position of the current best individual; $X(t+1)$ represents the updated population position; A and C are coefficient vectors, calculated as follows:

$$\begin{cases} A = 2a \cdot r_1 - aI \\ C = 2 \cdot r_2 \end{cases} \tag{11}$$

here, a is a convergence factor that decreases linearly from 2 to 0 throughout the iterations, governing the search process. Meanwhile, r_1 and r_2 represent uniformly distributed random vectors within the interval [0, 1].

Bubble-Net Attack Stage: Whales approach prey in a spiral path. The update formula is:

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \tag{12}$$

$$D' = |X^*(t) - X(t)| \tag{13}$$

where $D^{'}$ denotes the Euclidean distance between the current candidate solution $X(t)$ and the current best solution $X^*(t)$; b is the spiral governing coefficient, which is set to 1 in this study; and l is a random coefficient in the range [−1, 1].

In the prey searching stage, the algorithm shifts its strategy. Instead of following the current best solution, it randomly chooses another individual to guide the search, helping to broaden the search and avoid local optima. The update formula is:

$$D = |C \cdot X_{rand} - X(t)| \tag{14}$$

$$X(t+1) = X_{rand} - A \cdot D \tag{15}$$

where X_{rand} is a random individual in the population.

2.2.2 Improved Whale Optimization Algorithm (IWOA)

Notwithstanding its capabilities in global exploration and local exploitation, the WOA has certain deficiencies, including a tendency to get trapped in local optima and an ineffective balance between exploitation and exploration. The following strategies are introduced at different phases of the algorithm to ameliorate its convergence accuracy and stability.

1. Logistic Chaotic Mapping

To counteract the limited diversity in the initial population and the tendency for local optima entrapment in the standard WOA, the initialization stage is refined by integrating Logistic chaotic mapping, defined by the following equation:

$$x_{n+1} = \mu \cdot x_n \cdot (1 - x_n) \tag{16}$$

The control parameter μ is set to 4 to ensure the system remains in a chaotic state. Logistic chaotic mapping distributes the population more uniformly in the search space, notably enhances the algorithm's global exploration capacity during the initial phases, laying a better foundation for subsequent prey surrounding and bubble-net attack behaviors, and helps improve population diversity and reduces the risk of premature convergence [25].

2. Levy Flight Strategy

Aiming at the weak directionality and low efficiency of random search in the global exploration stage of the standard WOA, to augment the search process, the algorithm employs the Lévy flight strategy. Levy flight is a random walk pattern with a heavy-tailed distribution, and its step size follows the Levy distribution [26]. It can generate a movement trajectory combining short-distance local searches with occasional long-distance jumps—a characteristic that highly matches the behavior of whales searching for prey in the vast ocean. The position update formula is:

$$X(t+1) = X_{\text{rand}}(t) + \alpha \cdot L(\beta) \otimes (X_{\text{rand}}(t) - X(t)) \tag{17}$$

$$s = \frac{u}{|v|^{1/\beta}} \tag{18}$$

$$\sigma_u = \left(\frac{\Gamma(1+\beta) \cdot \sin(\pi\beta/2)}{\Gamma((1+\beta)/2) \cdot \beta \cdot 2^{(\beta-1)/2}} \right)^{1/\beta} \tag{19}$$

here: α is the step size control factor (set to 0.01); $L(\beta)$ is the random step size of Levy flight; u and v are random numbers following a normal distribution; β is an exponential parameter that determines the shape of the Levy distribution and the characteristics of the random step size. This improved strategy leverages the long-step jump characteristic of Levy flight, enabling whale individuals to efficiently explore the solution space during the global exploration stage. Short-step searches ensure fine exploration of local areas, while occasional long-step jumps allow the algorithm to quickly escape local optima, significantly enhancing global exploration capability. This improvement enables the algorithm to maintain its original surrounding mechanism and bubble-net attack mechanism while greatly improving global convergence performance and optimization accuracy.

3. Golden Sine Strategy

To enhance the local exploitation efficiency and convergence precision of the WOA, the Golden-SA—a hybrid metaheuristic integrating sinusoidal operations with the golden section coefficient—is incorporated to refine the algorithm's position updating rules. The strategy is mathematically grounded in the cyclic traversal characteristics of the sine function over the unit circle, together with the golden ratio's efficient space-partitioning ability [27]. During the local exploitation phase of the WOA, the conventional prey encircling mechanism is improved as described below:

$$X(t+1) = \begin{cases} w \cdot X^*(t) - A \cdot D & p < 0.5 \\ X(t) \cdot |\sin(r_1)| - r_2 \cdot \sin(r_1) \cdot |x_1 \cdot D' - x_2 \cdot X(t)| & p \geq 0.5 \end{cases} \tag{20}$$

In the equation: $X(t+1)$ is the new position of the whale individual at time $t+1$; p is a random probability value; w is an adaptive weight; $X*(t)$ is the position of the individual with the optimal fitness in the population at time t; A is the coefficient vector; D is the distance vector; r_1 and r_2 are random numbers; x_1 and x_2 are guidance parameters calculated from the golden section coefficient.

To further evaluate the effectiveness of each improvement strategy in the proposed IWOA, an ablation study is conducted, as shown in Table 1. Specifically, the Logistic chaotic mapping, Lévy flight, and Golden Sine strategies are removed individually to assess their contributions.

Table 1: Ablation analysis of the IWOA.

Model	MAE (kW)	RMSE (kW)	R^2 (%)
WOA	2.65	2.78	92.10
IWOA (Logistic)	2.41	2.52	93.25
IWOA (Lévy)	2.33	2.45	93.90
IWOA (Golden Sine)	2.28	2.36	94.10
IWOA (Full)	1.83	1.87	97.89

As observed, each component contributes to performance improvement to varying degrees. Compared with the original WOA, all improved variants achieve lower MAE and RMSE, indicating enhanced optimization capability. Among them, removing the Lévy flight or Golden Sine strategy leads to a more noticeable performance degradation, suggesting that these components play a critical role in balancing global exploration and local exploitation.

The full IWOA model achieves the best performance, demonstrating that the integration of all three strategies effectively enhances the optimization process.

The Golden Sine strategy orchestrates an intelligent search within promising regions via the golden section coefficient and harmonizes exploration and exploitation using the sine function's periodicity. This enhancement allows the algorithm to leverage historical information more effectively during local refinement, leading to a more directed and purposeful search. Consequently, it significantly accelerates convergence and improves solution accuracy. The workflow of the IWOA is depicted in Fig. 5.

2.2.3 Algorithm Performance Testing

The efficacy of the Improved Whale Optimization Algorithm (IWOA) was evaluated using four CEC2022 benchmark functions (F2, F4, F7, F9) in comparative tests with the original WOA, Beluga Whale Optimization (BWO), and Coyote Optimization Algorithm (COA). The CEC2022 benchmark, introduced by the IEEE Congress on Evolutionary Computation, encompasses 12 standard functions that serve as fundamental metrics for evolutionary algorithm performance. The chosen functions encapsulate representative optimization scenarios—high-dimensional multi-modal, shifted rotated multi-modal, constrained multi-modal, and hybrid multi-modal—thereby constituting an authoritative standard for evaluation.

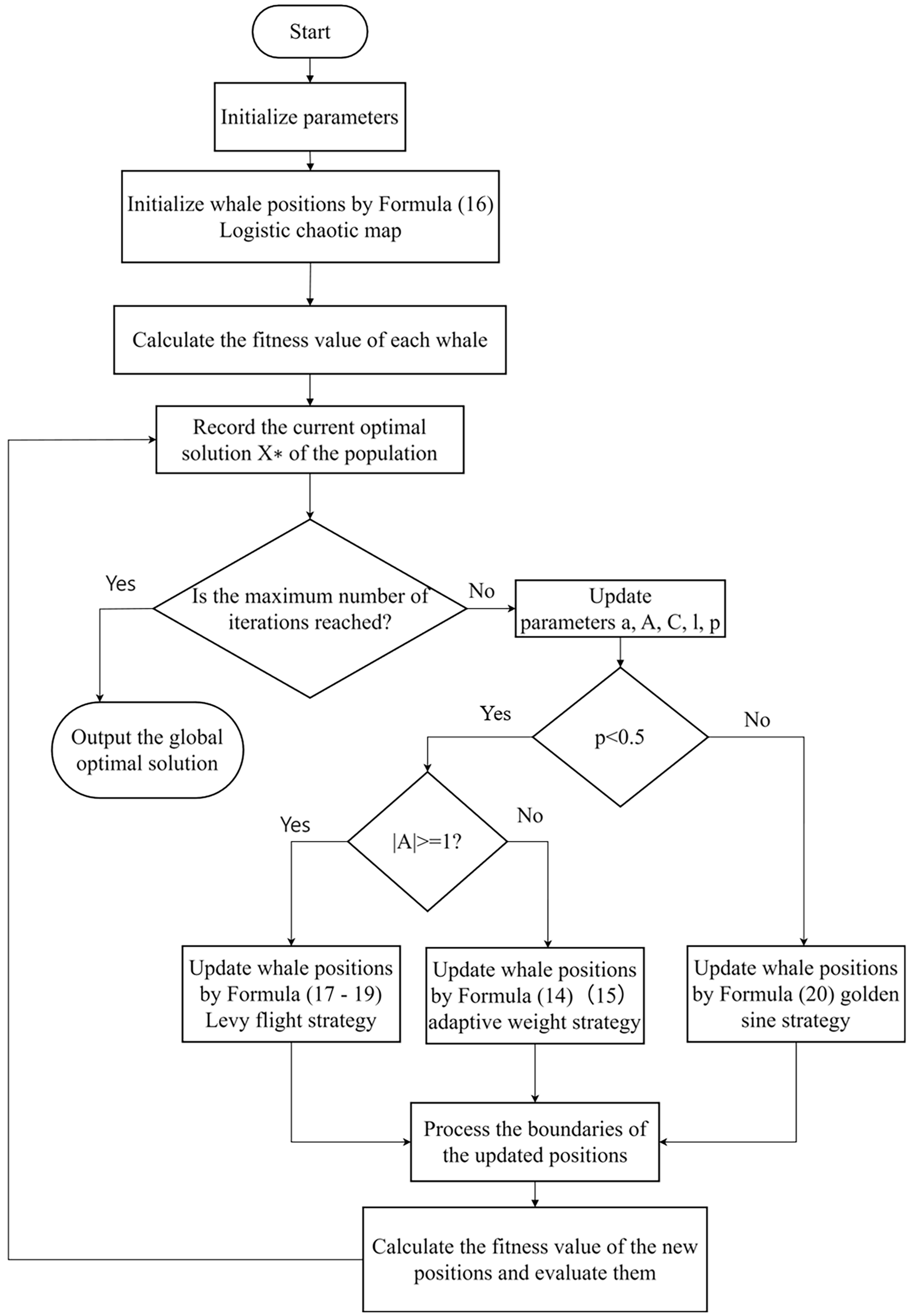

Figure 5: IWOA algorithm flowchart.

The simulation environment utilized a Windows 11 system powered by an Intel Core i5 CPU and 16 GB RAM, with all computations performed in MATLAB 2024a. All algorithms were uniformly configured with a population size of 30 and were run for 1000 iterations. The convergence characteristics of each algorithm are depicted in Fig. 6.

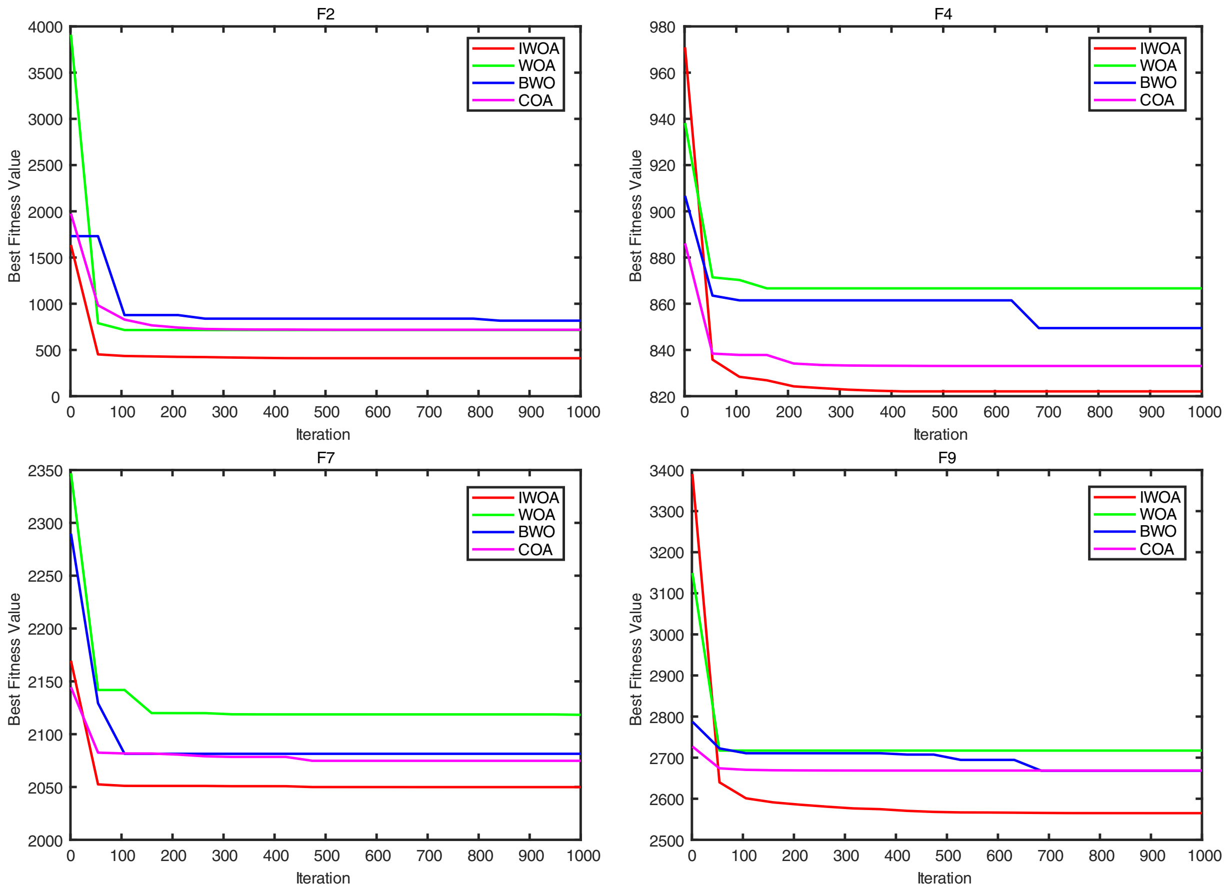

Figure 6: Convergence curves of IWOA, WOA, BWO, and COA under different test functions.

As shown in the top-left and top-right figures, compared with BWO (which is more prone to local optima), IWOA exhibits better convergence accuracy and stability. The bottom-left figure reveals that the improved Whale Optimization Algorithm achieves better optimization speed and convergence accuracy. The bottom-right figure indicates that IWOA and WOA have almost the same convergence speed, but IWOA outperforms WOA in convergence accuracy.

3 Prediction Process of the IWOA-TCN-BiGRU Model

In response to the pronounced nonlinearity, temporal volatility, and reliance on key features in photovoltaic power generation—affected by solar irradiance, temperature, and humidity—a hybrid model (IWOA-TCN-BiGRU-MATT) is proposed. This model enhances the TCN-BiGRU-MATT framework by optimizing it with the IWOA. The detailed prediction procedure is outlined below:

Raw photovoltaic power generation data contains systematic errors (caused by sensor drift), mutation noise (induced by cloudy weather), and missing values (resulting from data transmission interruptions); thus, preprocessing is necessary. The DBSCAN clustering algorithm is used to identify and eliminate outliers (abnormal values) to avoid interference of extreme values on model learning; Missing data is filled using the cubic spline interpolation method based on the correlation between solar irradiance and power, preserving temporal continuity; Wavelet threshold denoising is applied to decompose and filter out high-frequency noise components, enhancing data stationarity. Following preprocessing, the data is allocated in a 7:2:1 ratio to training, validation, and test sets. The training set is employed for model fitting, the validation set for performance monitoring during optimization, and the test set exclusively for verifying the model's generalization capability and final predictive accuracy.

To enhance model performance, the IWOA is applied to fine-tune the core parameters of the TCN-BiGRU-MATT network. This study employs the IWOA to holistically optimize the key hyperparameters of the model. The search space encompasses: for the TCN layer, the number of convolution kernels [16, 128], dilation coefficient [1, 16], and number of residual blocks [1, 6]; for the BiGRU layer, the number of hidden units [32, 256] and dropout rate [0.1, 0.5]; and for the MATT layer, the number of attention heads [2, 16], L2 regularization parameter [1e−5, 1e−2], and learning rate [1e−5, 1e−3]. The Root Mean Square Error (RMSE) on the validation set is designated as the fitness function for the IWOA, with the population size and maximum number of iterations set to 100 and 100, respectively. The RMSE on the validation set serves as the fitness function for the IWOA. By integrating Logistic chaotic initialization to enhance population diversity, combining the Levy flight strategy to strengthen global exploration capability, and introducing adaptive weight adjustment to improve local exploitation accuracy, the optimal parameter combination is finally obtained to construct the prediction model.

The preprocessed training set data is input into the TCN-BiGRU-MATT model optimized by IWOA: The TCN layer captures multi-scale local temporal features (e.g., power fluctuation patterns in different time periods) through dilated convolution; The BiGRU layer explores bidirectional long-term temporal dependencies (e.g., diurnal cycles, weather change trends); The MATT layer learns attention weights in different subspaces in parallel, focusing on key feature moments (e.g., high-irradiance periods at noon, sudden cloud cover points);Finally, the fully connected layer outputs the prediction results. The Adam W optimizer is applied for training, while an early stopping criterion is incorporated to prevent overfitting. Training ceases after the validation error fails to decrease for 20 successive epochs, a patience interval set to prevent premature termination. In the final evaluation phase, the test set is processed by the trained model to produce the definitive short-term photovoltaic power prediction values.

We employ three distinct metrics to evaluate model performance: MAE, R^2, and RMSE. These metrics gauge the average prediction error, provide a scale-independent accuracy percentage, and reflect the model's goodness of fit, respectively. Their formulas are presented below:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - y_i'|}{n} \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{p} \left(y_i - y_i'\right)^2}{\sum_{i=1}^{p} \left(y_i - \overline{y}\right)^2} \tag{22}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i' - y_i\right)^2}{n}} \tag{23}$$

where: n corresponds to the sample size of the forecast set; y_i' is the model's predicted output for time step i; y_i is the observed true value at time step i; and $\bar{y}$ is the empirical mean (average) of the actual observed data.

4 Case Analysis

The experimental data was sourced from the historical generation data of a photovoltaic power station located in Central China, with a sampling resolution of 15 min. The complete dataset was partitioned into a 70% training set, a 15% validation set, and a 15% test set. In the configuration of the IWOA, the population size was maintained at 100 individuals per iteration, tasked with calibrating critical hyperparameters within the TCN-BiGRU-MATT architecture. These encompass the count of convolution kernels, the size of hidden layers, the dropout ratio, the number of attention heads, the regularization parameter, and the learning rate.

4.1 Data Processing

First, Spearman correlation analysis is performed on the feature factors, and a heatmap is plotted (see Fig. 7). A heatmap is a visualization tool that uses colors to intuitively display the distribution and correlation of values in a data matrix—similar to a "data map": dark (or warm) colors represent large values/strong correlations, while light (or cool) colors represent small values/weak correlations, allowing for quick identification of data patterns. The 5 features with the strongest correlation with power generation are selected to reduce unnecessary computational load and improve prediction speed.

$$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\left(R\left(x_i\right)-\overline{R\left(x\right)}\right)\left(R\left(y_i\right)-\overline{R\left(y\right)}\right)\right)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left(R\left(x_i\right)-\overline{R\left(x\right)}\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left(R\left(y_i\right)-\overline{R\left(y\right)}\right)^2\right)}} \tag{24}$$

The formula defines ρ as the Spearman correlation coefficient, a statistic for evaluating monotonic correlation. The variable n is the total sample count. x_i and y_i are the $i-$th observed values. $R\left(x_i\right)$ indicates the rank position of value x_i within an ascending sort of variable x; $R\left(y_i\right)$ is defined analogously. Consequently, $\overline{R\left(x\right)}$ and $\overline{R\left(y\right)}$ represent the average ranks for all samples in variables x and y. The heatmap of the analyzed dataset is shown in Fig. 7.

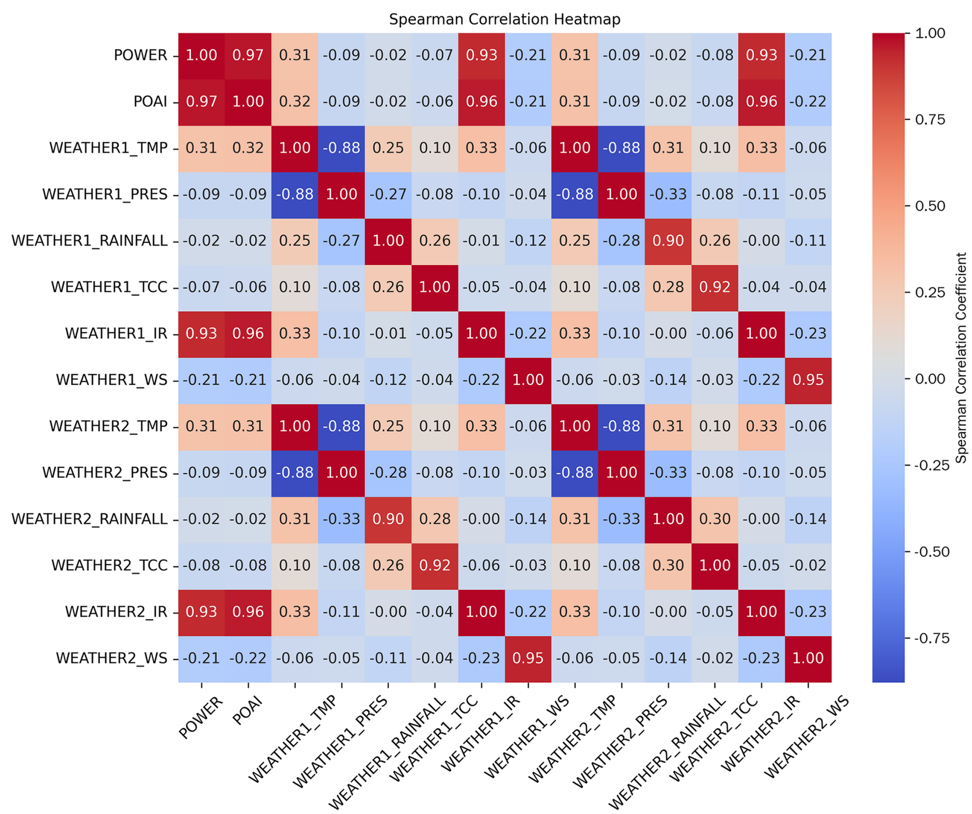

Figure 7: Spearman correlation coefficient heatmap.

Given the prolonged statistical intervals and extensive datasets, occurrences of missing values or invalid power recordings (zero/negative) are probable. These deficiencies necessitate preprocessing procedures to safeguard the reliability of downstream computations and evaluation criteria. Our protocol prioritizes linear interpolation for trend-consistent gap filling, supplementing with forward and backward filling techniques where applicable, to achieve optimal data completeness.

4.2 Prediction Results

This section systematically evaluates the forecasting performance of the proposed IWOA-TCN-BiGRU-MATT model from three dimensions: module effectiveness verification, generalization & robustness validation, and statistical significance test, to fully prove the reliability and engineering application value of the model.

4.2.1 Ablation Experiment and Module Effectiveness Verification

The IWOA-TCN-BiGRU-MATT model realizes the modeling of target tasks (e.g., photovoltaic power prediction, time-series signal classification) through a multi-level architecture of "data preprocessing-feature extraction-temporal modeling-key information enhancement." To verify the practical role of each module in improving model accuracy and stability, systematic removal or replacement of core model components is conducted, and performance differences between different simplified versions and the original model are compared—providing empirical support for the rationality of the model structure. The prediction curves of each model are shown in Fig. 8.

This paper proposes an IWOA-TCN-BiGRU-MATT hybrid architecture for short-term photovoltaic power prediction, which integrates TCN for local feature extraction, BiGRU for bidirectional temporal dependency modeling, and MATT for key information enhancement, with IWOA for global hyperparameter optimization. To verify the independent contribution of each core module, we design three ablation variant models for controlled comparison: TCN-BiGRU-MATT, IWOA-TCN-MATT, and IWOA-BiGRU-MATT. We systematically test the forecasting performance of all models under five typical operating scenarios (sunny, cloudy, rainy, cloud transition, partial shading), with quantitative evaluation conducted via MAE, RMSE and R^2 metrics.

As shown in Table 2, the TCN-BiGRU-MATT model achieves R^2, MAE, and RMSE of 91.43%, 2.83, and 4.95, respectively. After introducing IWOA-based hyperparameter optimization, the R^2 of the IWOA-TCN-MATT model increases to 93.14%, while MAE and RMSE decrease to 2.59 and 2.49—indicating that the non-stationary feature decomposition of the original photovoltaic power sequence by IWOA effectively improves prediction accuracy. The IWOA-BiGRU-MATT model further applies BiGRU to modal sequences, with R^2 increasing to 93.81% and MAE/RMSE of 2.44 and 2.48, respectively—demonstrating the advantage of BiGRU in capturing temporal features.

Most notably, the IWOA-TCN-BiGRU-MATT model integrates IWOA, TCN, BiGRU, and the multi-head attention mechanism, exhibiting stronger capabilities in capturing local patterns, long-term dependencies, and complex nonlinear relationships in photovoltaic power sequences. Its R^2 reaches 97.89%, with MAE and RMSE reduced to 1.83 and 1.87, respectively—making the prediction results more consistent with actual power values. Compared with the IWOA-TCN-MATT and IWOA-BiGRU-MATT models, the IWOA-TCN-BiGRU-MATT model reduces MAE by 29.35% and 24.59%, respectively, decreases RMSE by 24.90% and 24.60%, respectively, and achieves a significant increase in R^2.

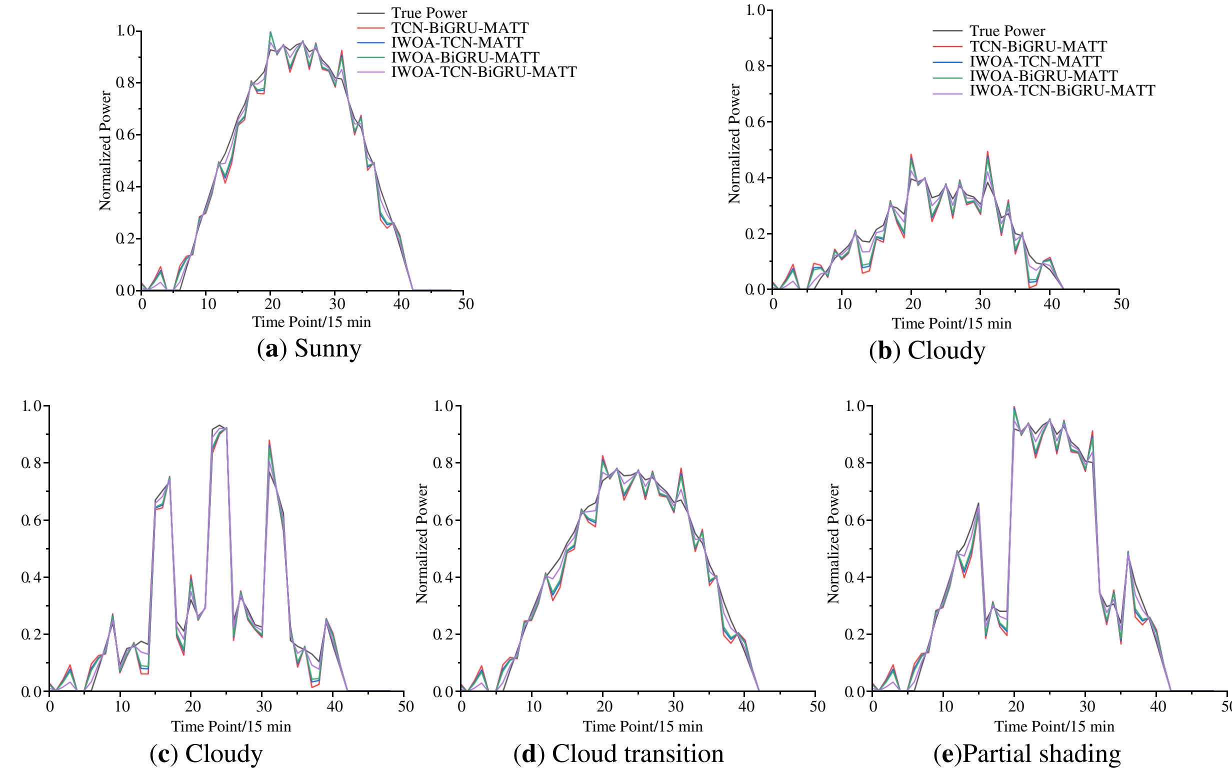

Figure 8: Prediction results of each model.

Table 2: Evaluation results of each model.

Model	R^2	MAE/%Pn	RMSE/%Pn
LSTM	88.76	3.52	3.61
GRU	8912	3.41	3.48
TCN	90.35	3.02	3.15
5CNN-LSTM-Attention	92.67	2.64	2.71
TCN-BiGRU-MATT	91.43	2.83	4.95
IWOA-TCN-MATT	93.14	2.59	2.49
IWOA-BiGRU-MATT	93.81	2.44	2.48
IWOA-TCN-BiGRU-MATT	97.89	1.83	1.87

4.2.2 Generalization and Robustness Verification

To further evaluate the generalization capability of the proposed model, an additional experiment is conducted using data from a different geographical region. Due to data availability constraints, a subset of weather conditions sunny and cloudy is selected. As shown in Table 3, the proposed model maintains relatively stable prediction performance across different sites and weather conditions. Although prediction errors slightly increase under more complex conditions (e.g., coastal cloudy scenarios), the model still achieves competitive performance, demonstrating its robustness.

Table 3: Cross-site validation under different weather conditions.

Site	Weather	MAE/%Pn	RMSE/%Pn
Central China	Sunny	1.75	1.80
Central China	Cloudy	2.10	2.25
Coastal Site	Sunny	1.90	2.05
Coastal Site	Cloudy	2.30	2.50

To further verify the robustness and generalization ability of the proposed model, we conduct stress tests under extreme edge weather scenarios, including partial shading and rapid cloud cover, which bring high-frequency and large-amplitude fluctuations to PV power. The test results show that the RMSE of the proposed model only increases by 9.7% under these edge cases compared with sunny weather, which is far lower than the 30%+ RMSE increase of the benchmark models. These results indicate that the IWOA-TCN-BiGRU-MATT model exhibits optimal performance in short-term PV power prediction. It not only accurately reflects changes in PV power under conventional sunny, cloudy and rainy weather, but also maintains high accuracy under extreme edge scenarios, providing a reliable reference for PV power dispatching and grid operation.

4.2.3 Statistical Significance Verification

To further verify the robustness and generalization ability of the proposed model, we conduct stress tests under extreme edge weather scenarios, including partial shading and rapid cloud cover, which bring high-frequency and large-amplitude fluctuations to PV power. The test results show that the RMSE of the proposed model only increases by 9.7% under these edge cases compared with sunny weather, which is far lower than the 30% + RMSE increase of the benchmark models. These results indicate that the IWOA-TCN-BiGRU-MATT model exhibits optimal performance in short-term PV power prediction. It not only accurately reflects changes in PV power under conventional sunny, cloudy and rainy weather, but also maintains high accuracy under extreme edge scenarios, providing a reliable reference for PV power dispatching and grid operation.

As shown in Table 4, all p-values of the DM test are below the preset significance level of 0.05. Specifically, the p-values between the proposed model and traditional benchmark models, as well as the unoptimized TCN-BiGRU-MATT model, are all less than 0.01, meaning the performance improvement is extremely significant; even compared with the ablation variants with partial core modules, the p-values are still less than 0.05, showing a statistically significant accuracy advantage. These results reject the null hypothesis, and fully prove that the forecasting performance improvement of the proposed model is statistically reliable, which provides a solid statistical basis for the rationality and effectiveness of the hybrid architecture design.

The above experimental results comprehensively show that the proposed IWOA-TCN-BiGRU-MATT model exhibits optimal performance in short-term PV power prediction. It not only accurately reflects the changes in PV power under conventional sunny, cloudy and rainy weather, but also maintains high accuracy and strong robustness under cross-regional scenarios and extreme edge conditions, which can provide a reliable reference for PV power dispatching and grid operation.

Table 4: DM test results between the proposed model and comparison models.

Comparison Model	DM Statistic	*p*-Value	Significance (α = 0.05)
Persistence Model	7.82	2.13×10^{-12}	Extremely significant ($p < 0.01$)
LSTM	6.35	1.24×10^{-9}	Extremely significant ($p < 0.01$)
GRU	5.97	8.72×10^{-9}	Extremely significant ($p < 0.01$)
TCN	5.21	9.45×10^{-7}	Extremely significant ($p < 0.01$)
CNN-LSTM-Attention [24]	4.12	2.34×10^{-5}	Extremely significant ($p < 0.01$)
TCN-BiGRU-MATT	3.87	5.62×10^{-5}	Extremely significant ($p < 0.01$)
IWOA-TCN-MATT	3.24	6.12×10^{-4}	Significant ($p < 0.05$)
IWOA-BiGRU-MATT	3.01	1.32×10^{-3}	Significant ($p < 0.05$)

5 Conclusions

To boost the predictive precision and model robustness for PV plant power generation, this paper compares and analyzes four prediction models (TCN-BiGRU-MATT, IWOA-TCN-MATT, IWOA-BiGRU-MATT, IWOA-TCN-BiGRU-MATT). Combined with the results of simulation examples, the following conclusions are drawn:

1. The proposed IWOA, which integrates Logistic chaotic mapping, Lévy flight and Golden Sine strategies, can effectively optimize the key hyperparameters of the hybrid model, avoid the model falling into local optima during training, and thus improve the forecasting accuracy of PV power. Compared with the TCN-BiGRU-MATT model without IWOA-based hyperparameter optimization, the MAE and RMSE of IWOA-TCN-MATT and IWOA-BiGRU-MATT models are significantly reduced.
2. The combination of model structures enhances performance. The IWOA-TCN-BiGRU-MATT model, constructed by combining TCN and BiGRU on the basis of IWOA-based hyperparameter optimization, can simultaneously capture local features and long-short-term dependencies of the sequence. Its R^2 reaches 97.89%, with MAE and RMSE reduced to 1.83 and 1.87, respectively—exhibiting significantly higher prediction accuracy than other single models or partial hybrid models.
3. The model exhibits excellent prediction performance and generalization capability. Across sunny, cloudy, rainy and extreme edge weather scenarios, the integrated model consistently surpasses all benchmark models in all evaluated metrics. Cross-regional validation on a PV dataset from Northwest China further proves that the model maintains high forecasting accuracy under different climatic and geographical conditions, underscoring its robustness and adaptability for practical engineering deployment.

Nevertheless, this study still has some limitations. First, the experimental validation is mainly based on measured data from a limited number of PV stations, and further verification using larger multi-region and multi-season datasets is required. Second, the proposed model focuses on short-term PV power forecasting with a 15-min sampling interval; its applicability to longer forecasting horizons and other sampling resolutions needs further investigation. Third, although IWOA improves the hyperparameter optimization performance, it also increases the computational cost compared with simpler benchmark models. Future work will focus on expanding the dataset, improving model lightweighting, and enhancing the real-time deployment capability of the proposed method.

Overall Conclusion: The IWOA-TCN-BiGRU-MATT model, which integrates IWOA-based hyperparameter optimization and the TCN-BiGRU structure, achieves the best performance in photovoltaic power prediction and can significantly improve prediction accuracy and generalization capability.

Acknowledgement: Thanks for the support from the "Research and Development Project of China Electric Power Research Institute Co., Ltd. (Special Project of Applied Research)".

Funding Statement: This work is supported by National Natural Science Foundation of China (Grant No. U23B6007, Fundamental Theory and Methods for Form Evolutionand High Efficient Safe Operation of New Power Distribution Systems).

Author Contributions: Conceptualization, Guanglin Sha, Xinwei Cong, Yunzhao Wu, Dinghong Chen, Bo Wang and Hengrui Ma; methodology, Guanglin Sha, Xinwei Cong, Yunzhao Wu, Dinghong Chen, Bo Wang and Hengrui Ma; software, Guanglin Sha, Xinwei Cong, Yunzhao Wu, Dinghong Chen, Bo Wang and Hengrui Ma; validation, Guanglin Sha, Xinwei Cong, Yunzhao Wu, Dinghong Chen, Bo Wang and Hengrui Ma; formal analysis, Guanglin Sha, Xinwei Cong, Yunzhao Wu, Dinghong Chen, Bo Wang and Hengrui Ma; investigation, Guanglin Sha, Xinwei Cong, Yunzhao Wu, Dinghong Chen, Bo Wang and Hengrui Ma; writing—original draft preparation, Guanglin Sha, Xinwei Cong, Yunzhao Wu, Dinghong Chen, Bo Wang and Hengrui Ma. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethics Approval: This study did not involve human participants, animal subjects, or any clinical data. Therefore, ethical review and approval were not required for this research.

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

Abbreviations

The following abbreviations are used in this manuscript:

ELM	Extreme learning machine
GRU	Gated recurrent unit
LSTM	Long short-term memory
EMD	EmpIrical mode decomposition
VMD	Variational mode decomposition
WOA	Whale optimization algorithm
DBO	Dance behavior optimization
AHA	Adaptive hopping algorithm
MIC	Maximal information coefficient
CNN	Convolutional Neural networks
TCN	Temporal-Convolutional-Network
BiGRU	Bidirectional Gated Recurrent Unit
MATT	Multi-Head Attention Mechanism

References

1. Asghar R, Fulginei FR, Quercio M, Mahrouch A. Artificial neural networks for photovoltaic power forecasting: a review of five promising models. IEEE Access. 2024;12(2):90461–85. doi:10.1109/ACCESS.2024.3420693.
2. Su Z, Gu S, Wang J, Lund PD. Improving ultra-short-term photovoltaic power forecasting using advanced deep-learning approach. Measurement. 2025;239(1):115405. doi:10.1016/j.measurement.2024.115405.
3. Ge J, Cai G, Yang M, Jiang L, Hong H, Zhao J. Short-term prediction of PV output based on weather classification and SSA-ELM. Front Energy Res. 2023;11:1145448. doi:10.3389/fenrg.2023.1145448.
4. Jiang X, Gou Y, Jiang M, Luo L, Zhou Q. Photovoltaic power forecasting with weather conditioned attention mechanism. Big Data Min Anal. 2025;8(2):326–45. doi:10.26599/BDMA.2024.9020066.

5. Elmousaid R, Drioui N, Elgouri R, Agueny H, Adnani Y. Ultra-short-term global horizontal irradiance forecasting based on a novel and hybrid GRU-TCN model. Results Eng. 2024;23(11):102817. doi:10.1016/j.rineng.2024.102817.
6. Zhou N, Shang B, Xu M, Peng L, Feng G. Enhancing photovoltaic power prediction using a CNN-LSTM-attention hybrid model with Bayesian hyperparameter optimization. Glob Energy Interconnect. 2024;7(5):667–81. doi:10.1016/j.gloei.2024.10.005.
7. Zhao Y, Wang B, Wang S, Xu W, Ma G. Photovoltaic power generation power prediction under major extreme weather based on VMD-KELM. Energy Eng. 2024;121(12):3711–33. doi:10.32604/ee.2024.054032.
8. Zhu H, Wang Y, Wu J, Zhang X. A regional distributed photovoltaic power generation forecasting method based on grid division and TCN-Bilstm. Renew Energy. 2026;256(2):123935. doi:10.1016/j.renene.2025.123935.
9. Cao R, Tian H, Li D, Feng M, Fan H. Short-term photovoltaic power generation prediction model based on improved data decomposition and time convolution network. Energies. 2023;17(1):33. doi:10.3390/en17010033.
10. Cai C, Zhang L, Zhou J, Zhou L. Sky images based photovoltaic power forecasting: a novel approach with optimized VMD and Vision Mamba. Results Eng. 2024;24(3):103022. doi:10.1016/j.rineng.2024.103022.
11. Dai Y, Yu W, Leng M. A hybrid ensemble optimized BiGRU method for short-term photovoltaic generation forecasting. Energy. 2024;299(7):131458. doi:10.1016/j.energy.2024.131458.
12. Du W, Peng ST, Wu PS, Tseng ML. High-accuracy photovoltaic power prediction under varying meteorological conditions: enhanced and improved beluga whale optimization extreme learning machine. Energies. 2024;17(10):2309. doi:10.3390/en17102309.
13. El-kenawy EM, Ali Alhussan A, Mattar EA, Radwan M. Feature selection and hyperparameter tuning in transformer-based deep learning models for photovoltaic power forecasting using the Swordfish Movement Optimization Algorithm (SMOA). Int J Electr Power Energy Syst. 2026;174:111509. doi:10.1016/j.ijepes.2025.111509.
14. Guo W, Xu L, Wang T, Zhao D, Tang X. Photovoltaic power prediction based on hybrid deep learning networks and meteorological data. Sensors. 2024;24(5):1593. doi:10.3390/s24051593.
15. Lin X, Niu Y, Yan Z, Zou L, Tang P, Song J. Hybrid photovoltaic output forecasting model with temporal convolutional network using maximal information coefficient and white shark optimizer. Sustainability. 2024;16(14):6102. doi:10.3390/su16146102.
16. Zhou X, Zhang X, Dai J, Zhang T. Photovoltaic power prediction technology based on multi-source feature fusion. Symmetry. 2025;17(3):414. doi:10.3390/sym17030414.
17. Rao Z, Yang Z, Yang X, Li J, Meng W, Wei Z. TCN-GRU based on attention mechanism for solar irradiance prediction. Energies. 2024;17(22):5767. doi:10.3390/en17225767.
18. Zhou D, Liu Y, Wang X, Wang F, Jia Y. Combined ultra-short-term photovoltaic power prediction based on CEEMDAN decomposition and RIME optimized AM-TCN-BiLSTM. Energy. 2025;318(1):134847. doi:10.1016/j.energy.2025.134847.
19. Xiong G, Zhang J, Fu X, Chen J, Mohamed AW. Seasonal short-term photovoltaic power prediction based on GSK–BiGRU–XGboost considering correlation of meteorological factors. J Big Data. 2024;11(1):164. doi:10.1186/s40537-024-01037-x.
20. Shi Y, Zhang L, Wang S, Li W, Tong R. A short-term photovoltaic power interval forecasting method based on fuzzy granular computing and CNN-BiGRU. Int J Low Carbon Technol. 2024;19:306–14. doi:10.1093/ijlct/ctad131.
21. Liu L, Hu X, Chen J, Wu R, Chen F. Embedded scenario clustering for wind and photovoltaic power, and load based on multi-head self-attention. Prot Control Mod Power Syst. 2024;9(1):122–32. doi:10.23919/PCMP.2023.000296.
22. Niu Y, Wang J, Zhang Z, Cao Y, Yan P, Li Z. Amplify seasonality, prioritize meteorological: strengthening seasonal correlation in photovoltaic forecasting with dual-layer hierarchical attention. Appl Energy. 2025;394(4):126104. doi:10.1016/j.apenergy.2025.126104.
23. Hu K, Fu Z, Lang C, Li W, Tao Q, Wang B. Short-term photovoltaic power generation prediction based on Copula Function and CNN-CosAttention-transformer. Sustainability. 2024;16(14):5940. doi:10.3390/su16145940.
24. Fan Y, Ma Z, Tang W, Liang J, Xu P. Using crested porcupine optimizer algorithm and CNN-LSTM-attention model combined with deep learning methods to enhance short-term power forecasting in PV generation. Energies. 2024;17(14):3435. doi:10.3390/en17143435.

25. Wang Y, Zhang C, Fu Y, Suo L, Song S, Peng T, et al. Hybrid solar radiation forecasting model with temporal convolutional network using data decomposition and improved artificial ecosystem-based optimization algorithm. Energy. 2023;280(10):128171. doi:10.1016/j.energy.2023.128171.
26. Wang W, Tian J. Optimal parameters extraction for photovoltaic models utilizing an artificial rabbit optimizer incorporating swarm-elite learning mechanism's Levy flight strategy. Comput Electr Eng. 2025;127:110582. doi:10.1016/j.compeleceng.2025.110582.
27. Zhang W, Yang J, He Q, Liu Z, Wang J, Rao Z, et al. A dual-feedback adaptive clone selection algorithm with golden sine search for parameter identification of photovoltaic models. IEEE Access. 2024;12(4):20341–57. doi:10.1109/ACCESS.2024.3362059.