Progress In Electromagnetics Research C, Vol. 166, 27–40, 2026

PIER C

(Received 24 December 2025, Accepted 9 January 2026, Scheduled 9 February 2026)

Photovoltaic Power Prediction Model Based on Fuzzy Entropy Clustering and Self-Attention Mechanism Combined with ICEEMDAN-WOA-CNN-BiLSTM

Zhongan Yu, Faneng Wu*, Zhiwei Huang, Zihao Deng, and Feng Zhang

School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, Jiangxi, China

ABSTRACT: To address the randomness and nonlinearity of photovoltaic (PV) power caused by meteorological factors, this paper proposes an ICEEMDAN-WOA-CNN-BiLSTM prediction model integrated with fuzzy entropy clustering and a self-attention mechanism. First, the original PV power sequence is decomposed into multiple multi-scale intrinsic mode function (IMF) components and residuals via the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN). Subsequently, components with similar complexity are merged using fuzzy entropy clustering to simplify the calculations. Then, the Whale Optimization Algorithm (WOA) is adopted to optimize the hyperparameters of the CNN-BiLSTM model, and the self-attention mechanism is integrated into the model to enhance the weights of key features. Comparative experiments demonstrate that the proposed model significantly outperforms single and traditional hybrid models in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). This can effectively improve the accuracy of short-term PV power prediction and provide support for power station dispatching and power grid stability.

1. INTRODUCTION

Against the backdrop of global green and low-carbon development, the global energy structure is accelerating its clean transition. At the 2023 COP28 Conference, over 100 countries signed commitments to triple global renewable energy installed capacity by 2030, demonstrating their resolve for this shift [1]. Thanks to its abundant, clean, and pollution-free advantages, global cumulative PV installed capacity has surged from 714 GW in 2020 to 1865 GW in 2024 [2]. However, PV power output is susceptible to meteorological factors like irradiance, temperature, and cloud cover, showing significant nonlinearity, non-stationarity, and randomness. This challenges power system stability and restricts the efficient integration and large-scale deployment of PV energy [3, 4]. Therefore, developing a high-precision and robust PV power prediction model is highly practical for optimizing power dispatching, cutting curtailment rates, and enhancing grid economy [5, 6].

To address PV power volatility challenges to power system stability, researchers have extensively studied high-precision, robust PV power prediction models. Relevant methods have advanced from traditional statistical models to machine learning and deep learning algorithms, with prediction accuracy continuously improving [7–9]. Traditional methods have clear limitations: persistence models are simple but only suitable for ultra-short-term prediction and are weather-sensitive; autoregressive moving average (ARMA) models have a concise structure, yet only applied to stationary sequences, failing to capture nonlinearity and sudden weather changes; regression analysis is interpretable but relies on linear assumptions inconsistent with actual PV power variations, leading to limited applicability and accuracy [10, 11]. Despite their strengths in interpretability and low computational cost for low-data, low-precision scenarios, traditional methods cannot meet practical demands in accuracy and robustness when dealing with PV power's inherent nonlinearity, non-stationarity, and randomness.

Photovoltaic power prediction methods based on machine and deep learning have gradually become mainstream approaches in recent years. In the field of photovoltaic power prediction using machine learning, Ref. [12] adopted a stacked machine learning model (integrating Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Multiple Linear Regression (MLR)) to analyze the input historical electrical and meteorological data, thereby achieving photovoltaic power prediction. In the research field of photovoltaic power prediction based on deep learning, Ref. [13] proposed a novel hybrid model by combining Neural Prophet (NP), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) for photovoltaic power prediction. Ref. [14] employed a dual-branch fine-grained segmentation network that integrates a CNN and Transformer architecture to extract local texture features and global contextual representations from ground-based cloud images, enabling high-precision photovoltaic power prediction.

To further improve prediction accuracy, the research on photovoltaic power prediction has gradually shifted toward directions such as decomposing and clustering input data and optimizing the hyperparameters of the models. To address the

* Corresponding author: Faneng Wu (6120240693@mail.jxust.edu.cn).

problem of low photovoltaic power prediction accuracy under different weather conditions, Ref. [15] proposed a secondary decomposition method combining Variational Mode Decomposition (VMD) and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) by considering the decomposition of original signals, which successfully reduces signal volatility and the complexity of feature mapping for photovoltaic data. Ref. [16] adopted the Fuzzy C-Means (FCM) clustering method to conduct clustering analysis on historical meteorological data, classifying the historical data into four categories (sunny, cloudy, rainy, and extreme weather) to enhance the accuracy of photovoltaic power prediction.

To date, despite remarkable advances in photovoltaic power prediction, existing models still have three key shortcomings: (1) Coarse sequence processing — most models directly model the original power sequence or use simple decomposition methods, which cannot effectively mitigate its nonlinearity and volatility and thus impairs feature extraction; (2) Incomplete parameter optimization — single network parameters are often optimized separately, while key parameters, such as the self-attention mechanism, are excluded from the optimization process, limiting overall model performance; (3) Insufficient feature focus — these models poorly capture long-term dependencies and lack differentiated weighting for key features, resulting in large prediction errors under complex conditions.

To address the aforementioned shortcomings of existing photovoltaic power prediction models, this paper proposes a deep learning-based photovoltaic power prediction model that integrates multi-scale feature extraction of photovoltaic power data and intelligent optimization. Firstly, key meteorological factors are screened through correlation analysis. Subsequently, the ICEEMDAN signal decomposition method is introduced to initially decompose the photovoltaic power data used in this study, yielding 12 intrinsic mode functions (IMFs) of different scales. Meanwhile, the fuzzy entropy calculation method is adopted to compute the fuzzy entropy values of the 12 decomposed IMFs. Based on the calculated fuzzy entropy values, the decomposed modal components are further classified into five categories of signals via K-means clustering, enabling refined extraction of multi-scale features from the power sequence. These features, together with the key meteorological factors, are input into a CNN-BiLSTM model based on the self-attention mechanism, and the WOA algorithm is introduced to optimize the model's hyperparameters for improved prediction performance. This study aims to provide a high-precision and reliable solution for photovoltaic power prediction under complex meteorological conditions, thereby offering technical support for the sustainable development of the photovoltaic industry.

2. RESEARCH MODEL

2.1. Correlation Analysis

The Pearson's correlation coefficient is a widely used statistical measure across multiple fields, as proposed by Karl Pearson. Its core purpose is to quantify the strength and direction of the linear correlation between two continuous variables, with a value range of $[-1, 1]$. The closer the absolute value is to 1, the stronger the correlation is [17]. In photovoltaic power prediction, photovoltaic output is affected by various factors such as irradiance, temperature, and humidity, resulting in complex types and large volumes of measured data. To accurately screen key influencing factors, simplify the model's computational complexity, and improve prediction accuracy, this study introduces this coefficient to conduct a correlation analysis on photovoltaic-related data. Finally, features highly correlated with photovoltaic power are selected for inclusion in the subsequent prediction model. The formula for calculating the Pearson correlation coefficient r is as follows:

$$r_{x,y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \tag{1}$$

where $r_{x,y}$ is the correlation between X and Y, and E represents the mathematical expectation.

2.2. ICEEMDAN Algorithm

ICEEMDAN (Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise), first proposed by Colominas et al. in 2014, is a novel signal decomposition method with notable improvements over EEMD (Ensemble Empirical Mode Decomposition) and CEEMDAN [18]. Unlike these two methods, ICEEMDAN employs a new noise injection and extraction strategy: when extracting each mode, it calculates the local mean of the sum of the raw signal and a specific modal component derived from EMD-decomposed white noise, instead of directly applying EMD (Empirical Mode Decomposition) to the signal. These upgrades nearly eliminate residual noise interference in decomposition results, while enhancing modal completeness and accuracy. Thus, this study uses ICEEMDAN to decompose PV power prediction data for improved model accuracy and precision. The detailed steps of ICEEMDAN are as follows:

Step 1: Assume that the original signal to be decomposed is $x(n)$, define the noise as $w^{(i)}(n)$, and initialize the residual as $r_0(n) = x(n)$. $E_k(x)$ denotes the kth IMF obtained by performing EMD on the signal x.

Step 2: Calculate the first-order IMF, denoted as IMF_1. First, compute $E_1(w^{(i)}(n))$: perform EMD on each set of white noise, and take the first IMF component of each set. Then, add white noise to generate a noise-added signal. The specific formula is as follows:

$$x^{(i)}(n) = x(n) + \sigma \cdot E_1\left(w^{(i)}(n)\right) \tag{2}$$

Here, σ denotes the noise standard deviation coefficient. Using the noise-added signal, compute the local mean of this signal, and then calculate the average value. The corresponding formula is as follows:

$$\begin{aligned} IMF_1(n) &= \left\langle M(x^{(i)}(n)) \right\rangle \\ &= \frac{1}{\xi}\sum_{i=1}^{\xi} M\left(x(n) + \sigma \cdot E_1(w^{(i)}(n))\right) \end{aligned} \tag{3}$$

Here, $M(x)$ represents the local mean of signal x; ξ denotes the number of noise addition times; and $\langle\cdot\rangle$ indicates taking the

average of the expression inside. Therefore, the calculation formula for solving the IMF in the first stage is:

$$IMF_1(n) = x(n) - IMF_1(n) \tag{4}$$

The formula for updating the residual is as follows:

$$r_1(n) = x(n) - IMF_1(n) \tag{5}$$

Step 3: Compute IMF_2. Calculate $E_2(w^{(i)}(n))$: perform EMD on each group of white noise, and extract the second IMF component from each group. Then, generate the noise-added residual for the second-order IMF. The specific formula is as follows:

$$r_1^{(i)}(n) = r_1(n) + \sigma \cdot E_2(w^{(i)}(n)) \tag{6}$$

The specific formula for calculating the local mean of the noise-added residual of the second-order IMF and then taking the average value is as follows:

$$\begin{aligned} IMF_2(n) &= \left\langle M(r_1^{(i)}(n)) \right\rangle \\ &= \frac{1}{\xi}\sum_{i=1}^{\xi} M\left(r_1(n) + \sigma \cdot E_2(w^{(i)}(n))\right) \end{aligned} \tag{7}$$

Finally, the formula for solving the second-order IMF is given as follows:

$$IMF_2(n) = r_1(n) - IMF_2(n) \tag{8}$$

The formula for updating the residual is:

$$r_2(n) = r_1(n) - IMF_2(n) \tag{9}$$

Step 4: Compute IMF_k. For the kth order IMF, calculate $E_k(w^{(i)}(n))$: perform EMD on each group of white noise, and extract the kth IMF component from each group. The formula for generating the noise-added residual is:

$$r_{k-1}^{(i)}(n) = r_{k-1}(n) + \sigma \cdot E_k(w^{(i)}(n)) \tag{10}$$

The specific formula for calculating the local mean of the noise-added residual of the kth order IMF and then taking the average value is as follows:

$$IMF_k(n) = \frac{1}{\xi}\sum_{i=1}^{\xi} M\left(r_{k-1}(n) + \sigma \cdot E_k(w^{(i)}(n))\right) \tag{11}$$

Finally, the formula for solving the kth order IMF is given as follows:

$$IMF_k(n) = r_{k-1}(n) - IMF_k(n) \tag{12}$$

The formula for updating the residual is:

$$r_k(n) = r_{k-1}(n) - IMF_k(n) \tag{13}$$

Step 5: Repeat the calculations in Step 4 until the residual $r_k(n)$ becomes a monotonic function, at which point the IMF decomposition can no longer proceed, and the solution is terminated. Finally, the original signal $x(n)$ is decomposed into:

$$x(n) = \sum_{k=1}^{K} IMF_k(n) + r_K(n) \tag{14}$$

Here, K denotes the total number of IMFs obtained from the final decomposition.

2.3. Fuzzy Entropy Algorithm

Fuzzy entropy, a key metric for fuzzy set uncertainty, quantifies the uncertainty of membership relations between elements and reflects the dispersion of their membership degree distribution. In fuzzy systems, entropy increases as element membership degrees approach 0.5 (higher fuzziness) and decreases as they approach 0 or 1 (lower fuzziness) [19].

For the PV power prediction in this paper, the above analysis confirms that ICEEMDAN is adopted to decompose the PV power data. However, the number of derived IMFs is excessively large. To improve model prediction accuracy and reduce computational complexity, this paper uses fuzzy entropy to calculate the entropy values of these IMFs, then reclassifies them via K-means clustering to form new IMF groups for subsequent modeling. The algorithm flow and calculation formula of fuzzy entropy are presented as follows:

Step 1: Let the time series for fuzzy entropy calculation be $\{u(i),\ i = 1, 2, ..., N\}$, with a length of N. The parameters required for fuzzy entropy calculation are the embedding dimension α, similarity tolerance β, and boundary gradient γ.

Step 2: Perform phase space vector reconstruction. The specific formula is as follows:

$$X(i) = [u(i), u(i+1), ..., u(i+\alpha-1)],\ i = 1, 2, ..., N-\alpha \tag{15}$$

Step 3: Calculate the distance between vectors. The specific formula is as follows:

$$d_{ij} = \max_{k=0}^{\alpha-1} |u(i+k) - u(j+k)| \tag{16}$$

Step 4: Compute the fuzzy similarity degree. The calculation formula is as follows:

$$D_{ij} = \exp\left(-(d_{ij})^{\gamma}/\beta\right) \tag{17}$$

Step 5: Calculate the average similarity degree. First, solve the average similarity degree of a single vector: for each α-dimensional vector $X(i)$, compute its average similarity degree with all other vectors. The specific formula is as follows:

$$\phi_i^{\alpha} = \frac{1}{N-\alpha-1}\sum_{j=1, j\neq i}^{N-\alpha} D_{ij} \tag{18}$$

After calculating the average similarity degree of individual vectors, compute the average of the average similarity degrees of all vectors. The specific formula is as follows:

$$\Phi^{\alpha} = \frac{1}{N-\alpha}\sum_{i=1}^{N-\alpha} \phi_i^{\alpha} \tag{19}$$

Step 6: Repeat the calculation steps from Step 2 to Step 5, and solve for $\alpha+1$ under the $\alpha+1$-dimensional condition.

Step 7: Through the above steps, finally calculate the fuzzy entropy value. The calculation formula is as follows:

$$FuzzyEn = \ln \Phi^{\alpha} - \ln \Phi^{\alpha+1} \tag{20}$$

2.4. Whale Optimization Algorithm

The Whale Optimization Algorithm (WOA) is a meta-heuristic optimization algorithm proposed by Mirjalili and Lewis in

2016, which simulates the predation behavior of humpback whales [20]. The WOA is designed to mimic the hunting strategies of humpback whales during prey encirclement: swimming around the prey and blowing bubbles to form a net-like structure. The WOA derived from these whale hunting strategies exhibits strong global search capability and high convergence performance.

The main purpose of introducing the WOA in this paper is to utilize its strong global search capability and high convergence performance to optimize the hyperparameters of the deep learning model during the subsequent photovoltaic power prediction process using the deep learning model. The specific solution model and formulas of the WOA are as follows:

(1) Strategy of the humpback whale's prey encircling phase

When humpback whales encircle their prey, they converge toward the current optimal position. The specific position update formula is as follows:

$$X(t+1) = X_p(t) - A \cdot |C \cdot X_p(t) - X(t)| \quad (21)$$

Herein, $X(t)$ and $X(t+1)$ denote the positions of the current and next-generation humpback whales, respectively; $X_p(t)$ represents the position of the current optimal individual; A and C are control coefficients, and their calculation formulas are as follows:

$$A = 2\lambda \cdot r_1 - \lambda \quad (22)$$

$$C = 2 \cdot r_2 \quad (23)$$

Herein, r_1 and r_2 are random numbers within the range of [0, 1]; λ is a convergence factor, whose value decreases linearly with the number of iterations. The corresponding calculation formula is as follows:

$$a = 2 - 2t/T_{\max} \quad (24)$$

Herein, t denotes the current number of iterations, and $T_{\max}$ represents the maximum number of iterations.

(2) Strategy of the humpback whale's bubble-net attacking phase

The bubble-net hunting behavior is simulated by approaching the prey along a spiral path, which enables local precise optimization. The specific formula is as follows:

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X_p(t) \quad (25)$$

Herein, $D' = |X_p(t) - X(t)|$ denotes the distance between the humpback whale and the prey; b is a constant used to control the shape of the spiral; l is a random number within the range of $[-1, 1]$, which determines the direction of the spiral.

(3) Strategy of the humpback whale's random swimming phase for searching prey

When the prey's position is unclear (i.e., $|A| \geq 1$), the humpback whale will swim randomly to explore new areas, thereby enhancing the global exploration capability. The specific calculation formula is as follows:

$$X(t+1) = X_{rand}(t) - A \cdot |C \cdot X_{rand}(t) - X(t)| \quad (26)$$

Herein, $X_{rand}(t)$ represents the position of a randomly selected individual in the current whale population.

2.5. Self-Attention Mechanism

Self-attention mechanism, a specialized attention mechanism variant, calculates interrelationships between sequence positions and captures internal dependencies via dynamic weighting, making it ideal for addressing long-range dependency issues [21]. Thus, this paper incorporates the mechanism, which enables global comparison of each input feature (e.g., meteorological, illumination, irradiance data) in PV power prediction-related information. It automatically adjusts attention weights for different features to capture global information, thereby enhancing model accuracy. The implementation process of the self-attention mechanism is as follows:

Step 1: Define the input sequence $X = [x_1, x_2, ..., x_n] \in R^{n \times d}$, where n denotes the length of the input sequence, and d represents the feature dimension of the input sequence.

Step 2: Generate Query (denoted as Q), Key (denoted as K), and Value (denoted as V) via linear transformation. The specific calculation formulas are as follows:

$$Q = XW^Q \quad (27)$$

$$K = XW^K \quad (28)$$

$$V = XW^V \quad (29)$$

Herein, W^Q, W^K, and W^V are learnable weight matrices.

Step 3: Calculate the attention weights. The calculation formula is as follows:

$$A = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (30)$$

Herein, $\sqrt{d_k}$ is a scaling factor, which is used to prevent the gradient vanishing problem caused by excessively large dot products.

Step 4: Perform a weighted summation to obtain the final output of the self-attention mechanism. The specific calculation formula is as follows:

$$Z = A \cdot V \quad (31)$$

2.6. Convolutional Neural Network

Convolutional Neural Networks (CNNs), deep learning architectures tailored for grid-like topological data, are widely used in deep learning [22]. Comprising convolutional, pooling, and fully connected layers, CNN reduces model complexity effectively through local connectivity, weight sharing, and hierarchical feature extraction, while ensuring invariance to translation, scaling, and distortion.

First, CNN uses convolution kernels in convolutional layers to perform sliding convolution on local input data regions, extracting local spatial or temporal features of the data. The discrete convolution operation expression for the input data and convolution kernel is as follows:

$$Y(t) = \sum_{i=0}^{k-1} \sum_{j=1}^{C} K(i,j) \cdot X(t+i,j) + b \quad (32)$$

Herein, $Y(t)$ denotes the output feature value at time t, k the size of the convolution window, C the feature dimension, K

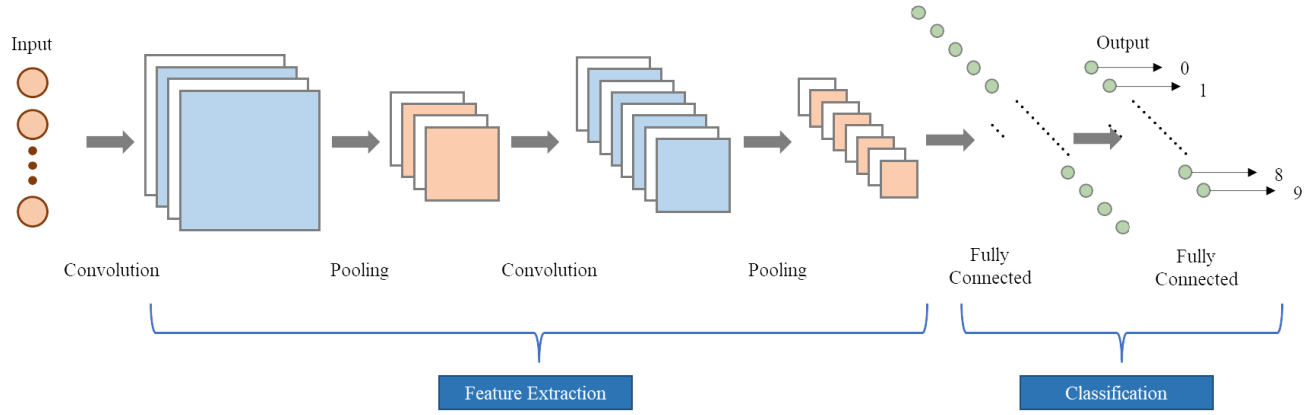

FIGURE 1. Structure diagram of CNN network.

stands for the convolution kernel, X the input data for convolution calculation, and b the bias term in the calculation. By inputting the processed photovoltaic power prediction data in this paper into the multi-channel convolution kernel and performing parallel computation, CNN can capture the fluctuation of the photovoltaic power prediction data at different frequencies.

The pooling layer in CNN, a key component for feature dimensionality reduction, mainly compresses CNN output parameters via max pooling and average pooling. This drastically reduces parameter count and dimensions, thereby lowering model computational complexity.

The fully connected layer of CNN realizes feature fusion and outputs results via full neuron connectivity, with distinct advantages in processing multi-dimensional and large-scale data. Applied to PV power prediction, it extracts features from PV power data; these multi-source features are then fed into the Bidirectional Long Short-Term Memory (BiLSTM) Neural Network model, enhancing prediction accuracy and overall model applicability. Figure 1 shows the CNN structure diagram.

2.7. Bidirectional Long Short-Term Memory

BiLSTM, an enhanced variant of a traditional Long Short-Term Memory Neural Network (LSTM), boosts sequence modeling capability by simultaneously capturing contextual information from both forward and backward directions [23]. LSTM integrates a gating mechanism into a Recurrent Neural Network (RNN), adding forget, input, and output gates to tackle the RNN's long-term dependency issue. BiLSTM improves on this with two independent LSTM layers (forward and backward) that process input sequences in opposite directions, thus comprehensively capturing long-term dependencies in sequences. Figure 2 shows the BiLSTM network structure.

This paper introduces the BiLSTM model: forward units capture the feature information on how historical meteorological data in the PV prediction dataset affects PV power, while backward units learn future relevance (e.g., PV power variations under different weather conditions). The output results are then rationally combined to enhance the overall PV prediction model's multivariate time-series modeling capability. The calculation formulas of the LSTM model are as follows:

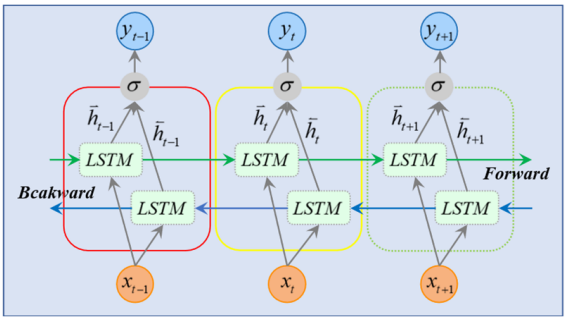

FIGURE 2. BiLSTM structure.

$$\begin{cases} f_t = \sigma\left(W_f[h_{t-1}; x_t] + b_f\right) \\ i_t = \sigma\left(W_i[h_{t-1}; x_t] + b_i\right) \\ \widetilde{C}_t = \tanh\left(W_C[h_{t-1}; x_t] + b_C\right) \\ C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \\ o_t = \sigma\left(W_o[h_{t-1}; x_t] + b_o\right) \\ h_t = o_t \odot \tanh(C_t) \end{cases} \tag{33}$$

where f_t denotes the calculation formula of the forget gate; i_t represents the calculation formula of the input gate; $\widetilde{C}_t$ and C_t refer to the candidate memory cell and the memory update cell respectively; o_t stands for the calculation formula of the output gate; h_t is the calculation formula of the hidden state output gate; W_f, W_i, W_C, W_o are the parameter matrices shared by the LSTM; b_f, b_i, b_C, b_o are the shared biases.

3. THE MODEL IN THIS PAPER

In this paper, a modeling solution is proposed for the problem of photovoltaic power prediction, and a photovoltaic power prediction model named ICEEMDAN-WOA-CNN-BiLSTM based on fuzzy entropy clustering and a self-attention mechanism is established. The overall solution process of this study is as follows:

First, for the various weather parameters contained in the photovoltaic dataset, this study adopts the Pearson correla-

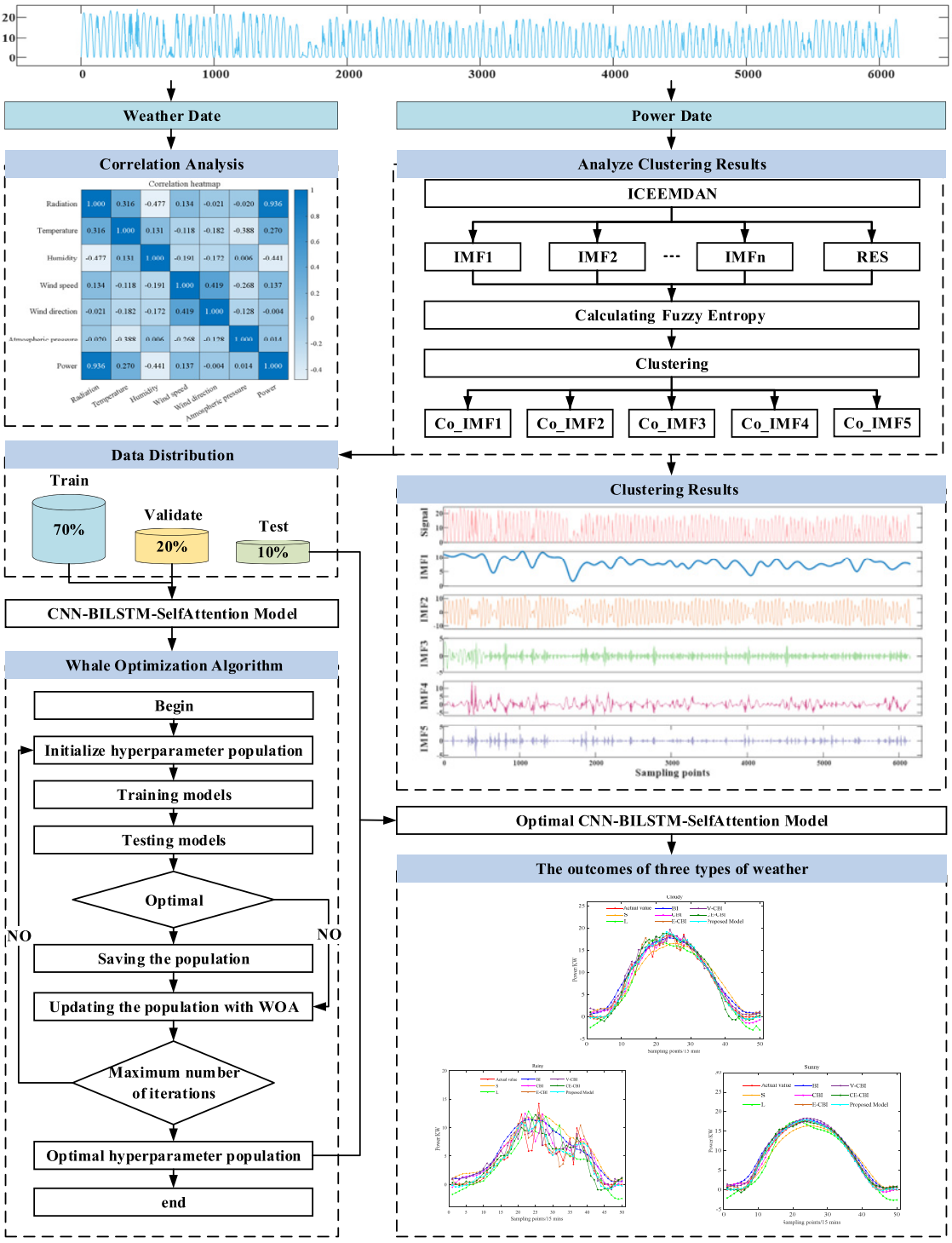

FIGURE 3. Comprehensive framework of the prediction model.

tion coefficient to conduct a correlation analysis. By calculating the correlation coefficients between features and setting a reasonable threshold, the key weather influencing factors that are strongly correlated with photovoltaic power are accurately screened out, thus providing high-quality input features for subsequent modeling.

Subsequently, the ICEEMDAN signal decomposition method is introduced to perform preliminary decomposition on the photovoltaic power time-series data adopted in this study, yielding a series of intrinsic mode functions (IMFs) with different scales. To further classify the obtained intrinsic mode functions of varying scales, this study adopts the fuzzy entropy calculation method to compute the fuzzy entropy value of each derived intrinsic mode function, and the decomposed mode functions are further categorized into five types of signals via

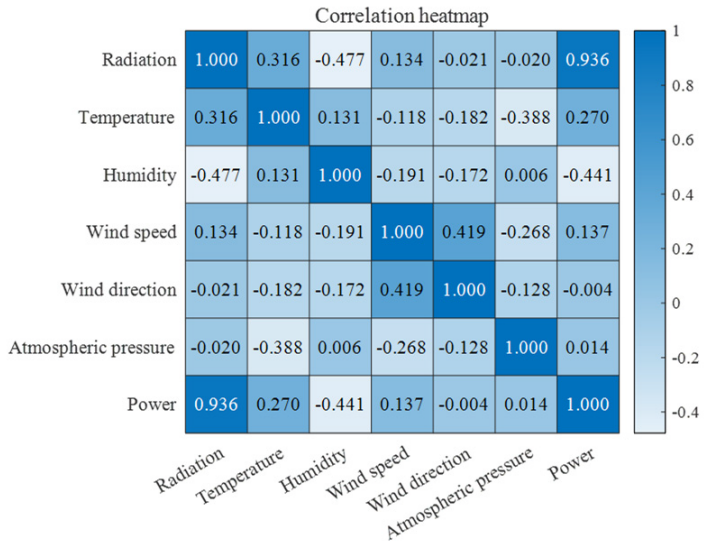

FIGURE 4. Results of correlation analysis.

K-means clustering, thus achieving a refined extraction of multi-scale features of the power sequence.

Finally, the multi-scale feature signals obtained via ICEEMDAN and fuzzy entropy clustering are fused with the key weather factor data screened in the earlier stage, which are jointly fed as input variables into the CNN-BiLSTM model based on the self-attention mechanism for training. During the model training process, the WOA is introduced to perform global optimization of the hyperparameters of the self-attention-based CNN-BiLSTM model. The optimal combination of hyperparameters is obtained through iterations, thereby constructing a photovoltaic power prediction model with superior prediction performance. Figure 3 shows the comprehensive architecture of the photovoltaic power prediction model proposed in this study.

4. CASE ANALYSIS

4.1. Data

The comprehensive dataset adopted in this study is derived from Hebei Province, China. This dataset integrates detailed atmospheric monitoring data and photovoltaic system operation performance data, specifically including seven dimensions of indicator information: solar irradiance, air temperature, relative humidity, wind speed, wind direction, atmospheric pressure, and photovoltaic output power. Given that the photovoltaic power is close to zero during nighttime periods, which has no practical significance for research and analysis, this study excludes all nighttime observation data. Meanwhile, targeted processing is performed on the abnormal values and missing data points existing in the dataset, and 6,150 valid sample data are finally obtained. To screen out the meteorological factors with significant impacts on photovoltaic power, the study adopts the Pearson correlation analysis method to test each meteorological data feature, and the results of the correlation analysis are shown in Figure 4.

The results indicate that among the various factors affecting photovoltaic power output, global solar irradiance, air temperature, wind speed, and atmospheric pressure are positively correlated with photovoltaic power, while wind direction and relative humidity are negatively correlated. From the perspective of the degree of correlation, global solar irradiance exhibits the highest correlation with photovoltaic power, followed by relative humidity and air temperature. Based on these analysis results, global solar irradiance, relative humidity, and air temperature are identified as the core factors restricting photovoltaic power output, and these key variables are incorporated into the prediction model as basic input features.

4.2. Data Decomposition and Reconstruction

The original photovoltaic power data is decomposed via the ICEEMDAN method, yielding 10 intrinsic mode functions (IMF1-IMF10) and one residual component (Res). The decomposition results are shown in Figure 5. As observed from the figure, the high-frequency IMFs (e.g., IMF1-IMF3) exhibit frequent waveform oscillations, which correspond to the short-time-scale fluctuations of the photovoltaic power. Such fluctuations are mainly induced by factors such as rapid cloud drift and local shading on the surface of photovoltaic panels, characterized by fast power variations and strong randomness. The intermediate-frequency IMFs (e.g., IMF4-IMF6) feature moderate oscillation frequencies, reflecting medium-time-scale variations in photovoltaic power. In contrast, the low-frequency IMFs (e.g., IMF7-IMF10) and residual component (Res) show smooth waveforms, which are associated with the long-term trends of photovoltaic power.

The complexity differences between the components can be clearly observed after calculating the fuzzy entropy for each IMF and Res. IMF3 has the highest fuzzy entropy value, indicating that it contains the most disordered photovoltaic power fluctuations. From IMF1 to IMF10 and Res, the fuzzy entropy generally shows a decreasing trend, which reflects the complex-

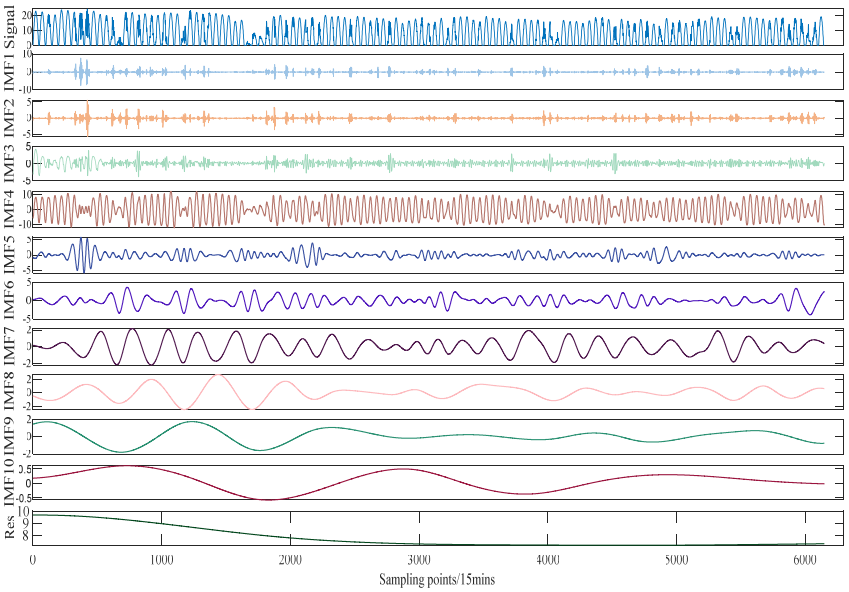

FIGURE 5. Results of ICEEMDAN decomposition.

ity of photovoltaic power decreasing with an increase in the time scale. High-frequency components are more complex due to their "rapid and random variations," while low-frequency components and residual trends are simpler due to their "slow and stable variations."

Normalization is required to eliminate the dimensional differences in the fuzzy entropy of different components. After normalization, the complexity characteristics of each component can be compared on a unified scale, providing a fair feature dimension for subsequent clustering. The fuzzy entropy ranking results for each component are shown in Figure 6.

A clustering algorithm is adopted to cluster the normalized fuzzy entropy into five categories, with the core objective of grouping components with similar fuzzy entropy (complexity) into the same category. The results of the component clustering are presented in Figure 7.

Through data decomposition and reconstruction, a multi-scale and complexity-stratified deconstruction of the photovoltaic power signal is achieved. This process not only preserves the full-scale information of photovoltaic power, ranging from "high-frequency random fluctuations" to "low-frequency stable trends," but also realizes the effect that "components in the same category have similar characteristics while those in different categories show distinct feature differentiation" via clustering, laying an accurate input foundation for the subsequent analysis of photovoltaic power.

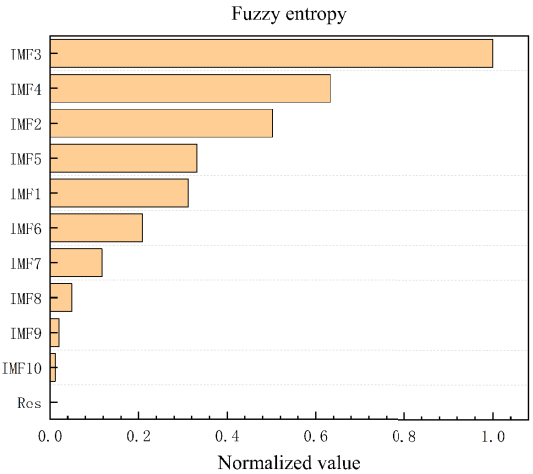

FIGURE 6. Ranking of fuzzy entropy.

4.3. Model Evaluation Indicators

To verify the effectiveness of the model, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) are selected as evaluation metrics. The specific formulas are as follows:

$$\mathrm{MAE} = \frac{1}{M}\sum_{j=1}^{M}\left|y_{\mathrm{true},j} - y_{\mathrm{fore},j}\right| \tag{34}$$

$$\mathrm{MSE} = \frac{1}{M}\sum_{j=1}^{M}\left(y_{\mathrm{true},j} - y_{\mathrm{fore},j}\right)^2 \tag{35}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(y_{\mathrm{true},j} - y_{\mathrm{fore},j}\right)^2} \tag{36}$$

TABLE 1. Main parameters of each model.

Model	Parameters
SVM	C = 1.0, Kernel = RBF, Gama = 0.1,
LSTM	Learning rate =1e-4,
BiLSTM	Number of neurons = 50,
CNN-BiLSTM	Regularization parameter = 1.00e-5

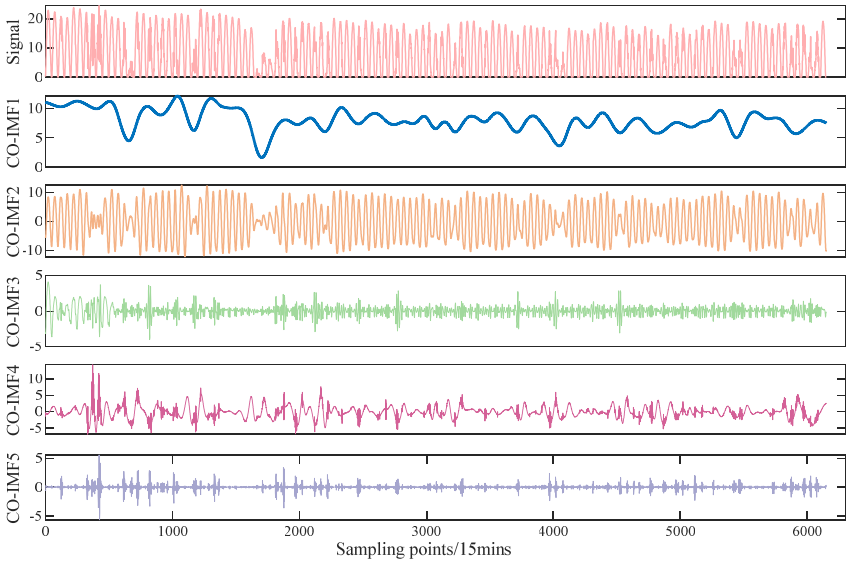

FIGURE 7. Clustering results.

TABLE 2. Optimal hyperparameter.

CO-IMF	Learning rate	Number of neurons	Key dimension	Regularization parameter
1	0.001871	10	47	0.000173
2	0.00194	66	15	0.000446
3	0.001488	69	14	0.000429
4	0.001291	18	3	0.00027
5	0.001129	41	8	0.000114

$$\mathbf{R}^2 = 1 - \frac{\sum_{j=1}^{M} \left(y_{\text{true},j} - y_{\text{fore},j}\right)^2}{\sum_{j=1}^{M} \left(y_{\text{true},j} - y_{\text{mean},j}\right)^2} \tag{37}$$

where M is the number of predicted points, $y_{\text{true},j}$ the actual value of power generation, and $y_{\text{fore},j}$ the predicted value of the power generation.

4.4. Analysis of Prediction Results

To verify the performance advantages of the model proposed in this study for photovoltaic power prediction, eight typical prediction models are selected for comparison. They include the traditional machine learning model SVM, basic deep learning models LSTM and BiLSTM, hybrid deep learning model CNN-BiLSTM, and four decomposition-integration models: EMD-CNN-BiLSTM, VMD-CNN-BiLSTM, and CEEMDAN-CNN-BiLSTM. The abbreviations of each model are shown in Appendix A. Table 1 lists the main parameter configurations of each model. The optimal hyperparameters of the model in this paper are presented in Table 2, and the prediction performance is displayed in Table 3.

From the perspective of the overall trend, the prediction accuracy of decomposition-integration models is generally superior to that of single models, whereas the model proposed in this paper achieves the best performance across all metrics.

TABLE 3. Results of different models.

Model	MAE	MSE	RMSE	R^2
S	1.632424	3.784133	1.945285	0.879956
L	1.435246	3.42396	1.850394	0.928726
BI	1.273387	2.724824	1.650704	0.929893
CBI	0.971451	1.994037	1.412104	0.949959
E-CBI	0.913335	1.748736	1.322398	0.955557
V-CBI	0.993061	1.447189	1.202992	0.965141
CE-CBI	0.866932	1.280264	1.131487	0.968181
Proposed Model	**0.468358**	**0.585213**	**0.764992**	**0.986442**

At the level of single models, as a traditional machine learning model, SVM struggles to capture complex temporal dependencies when processing high-dimensional, nonlinear time-series data, such as photovoltaic power, thus yielding the largest prediction errors. LSTM mitigates the gradient problem in long sequences through its gating mechanism, but it can only model temporal features unidirectionally and lacks sufficient capability to capture the bidirectional dependencies of photovoltaic power, achieving higher accuracy than SVM. BiLSTM incorporates backward LSTM and enables simultaneous bidi-

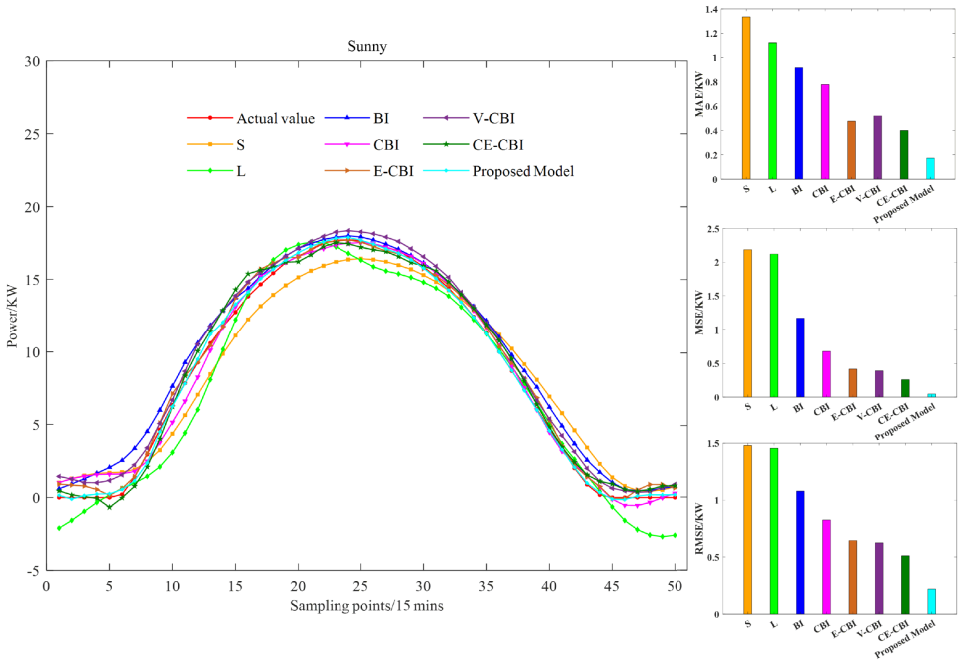

FIGURE 8. Results of sunny day forecast.

TABLE 4. Results under different weather conditions.

Weather	Model	MAE	MSE	RMSE
Sunny	S	1.334104	2.184809	1.47811
	L	1.120237	2.120001	1.456022
	BI	0.919037	1.163465	1.07864
	CBI	0.778891	0.680161	0.824719
	E-CBI	0.476661	0.416512	0.645378
	V-CBI	0.52131	0.390454	0.624863
	CE-CBI	0.399628	0.261262	0.511138
	Proposed Model	**0.17253**	**0.047502**	**0.21795**
Cloudy	S	1.517867	3.008864	1.734608
	L	1.22561	2.787269	1.669512
	BI	1.029906	1.711299	1.308166
	CBI	1.036036	1.552426	1.245964
	E-CBI	0.752877	1.025776	1.012806
	V-CBI	0.938001	1.306638	1.143083
	CE-CBI	1.023823	1.63801	1.279848
	Proposed Model	**0.593793**	**0.686937**	**0.828817**
Rainy	S	2.337861	8.306885	2.882167
	L	2.034842	7.450644	2.729587
	BI	2.037813	5.996526	2.44878
	CBI	1.695751	4.791903	2.189041
	E-CBI	1.014547	1.926506	1.387986
	V-CBI	1.360033	2.753552	1.659383
	CE-CBI	1.169082	2.337725	1.528962
	Proposed Model	**0.858772**	**1.704592**	**1.3056**

rectional modeling of photovoltaic data. However, none of these three models specifically address the multi-scale fluctuation characteristics of photovoltaic power, resulting in limited overall accuracy.

At the level of hybrid deep learning models, CNN-BiLSTM combines the local feature extraction capability of the CNN with the temporal modeling capability of BiLSTM. It can capture the local fluctuation patterns and bidirectional temporal dependencies of photovoltaic power, achieving higher accuracy than standalone BiLSTM. However, this model directly performs modeling on raw non-stationary signals without addressing the issue of multi-scale feature coupling, leaving room for improvement in fitting high-frequency random fluctuations.

At the level of decomposition-integration models, EMD-CNN-BiLSTM decomposes photovoltaic power signals via EMD. However, owing to the mode mixing problem inherent in EMD, some high-frequency and low-frequency components interfere with each other, leading to a limited improvement in prediction accuracy. VMD-CNN-BiLSTM is based on variational mode decomposition, which further enhances the accuracy. CEEMDAN-CNN-BiLSTM suppresses mode mixing through complete ensemble empirical mode decomposition with adaptive noise.

At the level of the proposed model in this paper, the model achieves more accurate multi-scale decomposition via ICEEMDAN and groups components with similar complexity through fuzzy entropy clustering to reduce model redundancy. The CNN-BiLSTM-Attention module focuses on local features, bidirectional temporal dependencies, and key nodes, while the WOA is employed to optimize hyperparameters for enhanced generalization capability. Ultimately, the model delivers optimal prediction accuracy, and the results fully verify the effectiveness of the full-process optimization strategy.

To further verify the effectiveness of the proposed model, the prediction results under sunny, cloudy, and rainy conditions are extracted for comparison, with the prediction performance details presented in Table 4. For an intuitive demonstration of the prediction performance, the visualization results corresponding

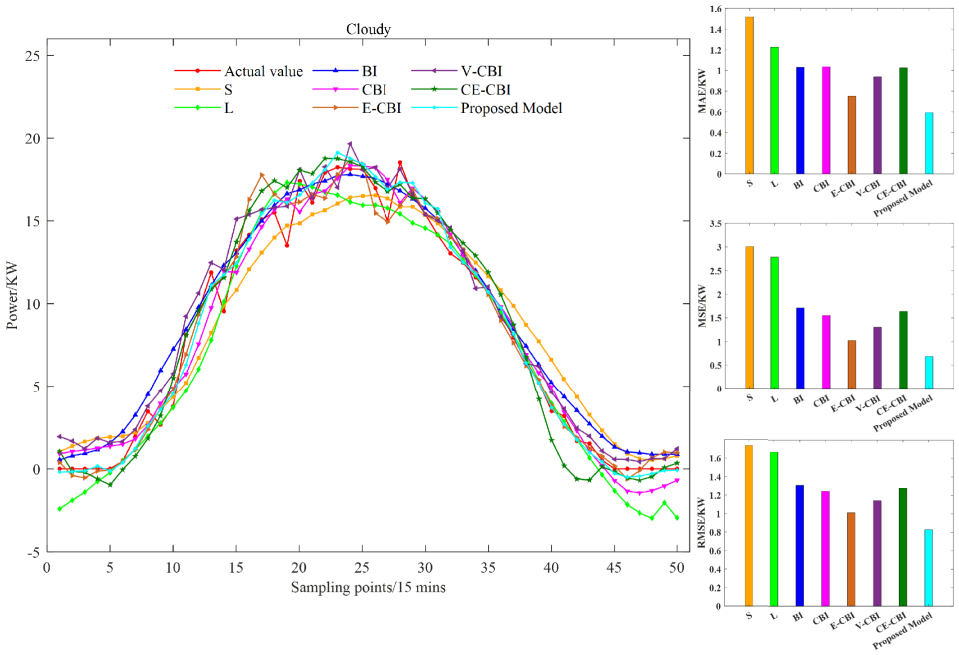

FIGURE 9. Results of cloudy day forecast.

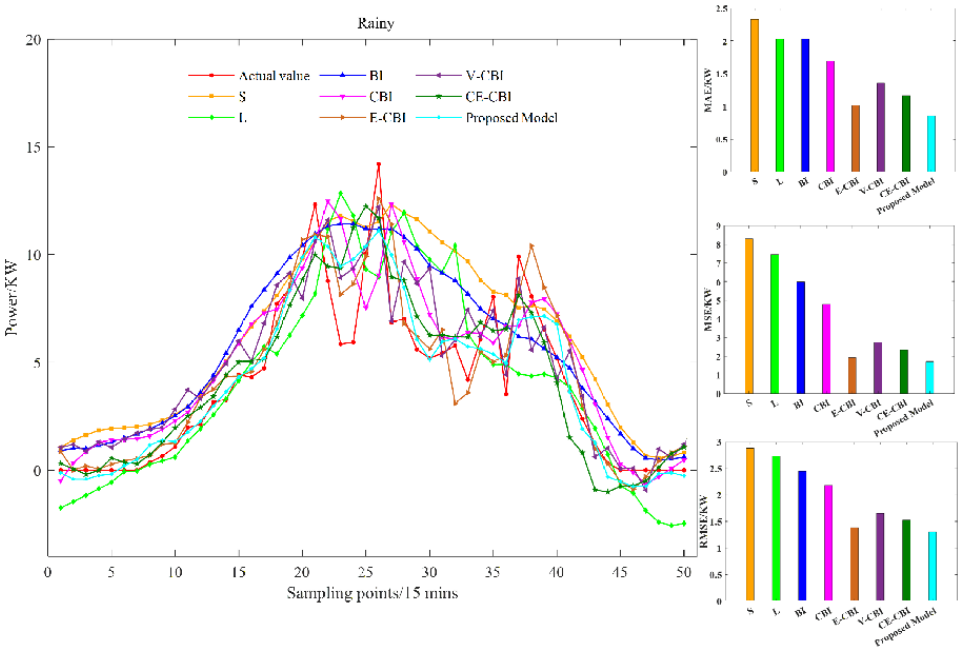

FIGURE 10. Results of rainy day forecast.

to sunny, cloudy, and rainy conditions are illustrated in Figures 8 to 10.

On sunny days, the photovoltaic power sequence exhibits strong regularity. Decomposition-integration models outperform single models as they can effectively capture stable trends and weak fluctuation characteristics. The proposed model in this paper achieves the accurate decomposition of low-frequency trend components and slight high-frequency fluctuations via ICEEMDAN, and eliminates redundant information through fuzzy entropy clustering, further improving prediction accuracy and achieving extreme fitting of stable sequences.

The core challenge of cloudy weather lies in the power fluctuations caused by sudden changes in irradiance. The self-attention mechanism of the proposed model can accurately identify key nodes of fluctuations and strengthen the weight assignment for mutation features. Meanwhile, the hyperparameter combination optimized by WOA enhances the model's adaptability to dynamic changes. In contrast, some decomposition-integration models (e.g., CE-CBI) overfocus on

high-frequency components, which instead introduce redundant information and lead to performance degradation.

On rainy days, irradiance is low and fluctuates drastically, resulting in raw sequences with prominent nonlinear characteristics that are difficult for traditional models to capture complex dependencies from. The proposed model decomposes the highly stochastic sequences into multiple stationary IMF components via ICEEMDAN, and merges components with similar fluctuation characteristics through fuzzy entropy clustering, thereby reducing nonlinear complexity. Meanwhile, CNN extracts local fluctuation features; BiLSTM captures bidirectional temporal dependencies; and the self-attention mechanism strengthens the response to key factors, such as precipitation intensity and cloud cover changes. Ultimately, the model maintains excellent performance even in this high-difficulty scenario, which verifies its robustness.

4.5. Comparative Experiment of Fuzzy Entropy Clustering

To further verify the unique value of fuzzy entropy clustering in multi-scale feature extraction, comparative experiments are designed with frequency-based K-means clustering and variance-based K-means clustering as baseline methods. The experiment focuses on analyzing the impact of different clustering criteria on the final prediction performance, with the same ICEEMDAN decomposition, CNN-BiLSTM-Attention model, and WOA hyperparameter optimization applied to all groups to eliminate interference from other factors. The prediction performance of the three clustering methods is shown in Table 5.

TABLE 5. Results of different clusterings.

Clustering Method	MAE	MSE	RMSE
Frequency	0.905132	1.595609	1.263174
Variance	0.845762	1.258361	1.121767
Proposed Model	**0.468358**	**0.585213**	**0.764992**

Fuzzy entropy clustering outperforms frequency/variance-based methods in photovoltaic power feature extraction. It captures the intrinsic dynamic complexity of power sequences, enabling more meaningful component grouping and laying a high-quality foundation for subsequent model training. This superiority is particularly prominent under complex weather conditions, further confirming the rationality of integrating fuzzy entropy clustering into the proposed prediction framework.

4.6. External Dataset Validation

To further verify the model's generalization ability, measured data from a photovoltaic power station in Ningxia Hui Autonomous Region in January 2020 is selected as the external validation dataset. Located in the arid northwest region, Ningxia has distinct climatic characteristics (low humidity, strong irradiance volatility) compared with the original Hebei dataset, which can effectively test the model's cross-regional adaptability. The results are shown in Table 6.

TABLE 6. Validation results with external dataset.

Model	MAE	MSE	RMSE
V-CBI	0.8127	1.2228	1.1058
CE-CBI	0.7532	1.0870	1.0426
Proposed Model	**0.4128**	**0.4751**	**0.6893**

The proposed model still maintains optimal performance on the Ningxia dataset, with an MAE of only 0.4128. This confirms that the model can effectively adapt to the climatic characteristics and irradiance fluctuation patterns in the arid regions of Northwest China.

4.7. Optimizer Selection Analysis

To verify the rationality and superiority of WOA in hyperparameter optimization of the model, two mainstream metaheuristic algorithms Particle Swarm Optimization (PSO) and Grey Wolf Optimizer (GWO) are selected for comparison. The comparison results of the three optimization algorithms are presented in Table 7.

TABLE 7. Comparison results of different algorithms.

Optimizer	MAE	MSE	RMSE
PSO	0.5984	0.7667	0.8756
GWO	0.5102	0.6780	0.8234
WOA	**0.4683**	**0.5852**	**0.7650**

WOA outperforms PSO and GWO in optimization performance. Its unique predation strategy enables it to accurately match the multi-scale features and complex network structure of the proposed model, effectively improving the prediction accuracy and stability of the model. Therefore, it is fully reasonable to select WOA as the hyperparameter optimization algorithm in this study.

5. CONCLUSION

PV power generation shows significant randomness and nonlinearity due to meteorological factors, hindering accurate power output prediction and challenging PV plant dispatching and grid stability. To resolve this issue, this study proposes an ICEEMDAN-WOA-CNN-BiLSTM integrated prediction model with targeted technical innovations. The main conclusions are as follows:

1. The ICEEMDAN-fuzzy entropy clustering preprocessing provides a high-quality data foundation. ICEEMDAN adaptively decomposes the original power sequence into multi-frequency IMF components and residuals, fully preserving multi-scale fluctuation characteristics; fuzzy entropy clustering further merges similar components, removes redundant information, simplifies calculations, avoids signal aliasing interference from single decomposition, and enhances preprocessing pertinence and efficiency.

2. The self-attention mechanism enhances the feature focusing capability. On the basis of CNN extracting local spatial features and BiLSTM capturing long-term temporal dependencies, the self-attention mechanism can adaptively calculate the importance weights of different features and moments, focusing on key meteorological features and temporal inflection points that influence power changes. This mechanism addresses the issue of insufficient capture of key information by traditional models and improves the prediction sensitivity of power fluctuations under complex working conditions.

3. The WOA realizes the global optimization of hyperparameters. Aiming at the problems of low efficiency and easy falling into local optimum in manual debugging of CNN-BiLSTM hyperparameters, WOA adaptively optimizes hyperparameters by virtue of its global search and convergence speed advantages, enabling the model structure to accurately match the preprocessed data features. While ensuring prediction accuracy, it effectively improves the stability and generalization ability of the model.

We objectively summarize areas for improvement, including certain dependence on data volume, insufficient integration of PV module status impacts, and slightly high computational complexity. Future work will explore few-shot learning, integrate equipment status data, optimize model lightweight, and further dig into the intrinsic connection between electromagnetic propagation of irradiance and PV power, providing more references for related research.

ACKNOWLEDGEMENT

This work was supported by the Postgraduate Innovation Special Fund of Jiangxi Province (No. YC2025-S589).

APPENDIX A.

Full form	Abbreviations
SVM	S
LSTM	L
BiLSTM	BI
CNN-BiLSTM	CBI
EMD-CNN-BiLSTM	E-CBI
VMD-CNN-BiLSTM	V-CBI
CEEMDAN-CNN-BiLSTM	CE-CBI

REFERENCES

[1] Shi, J. and Y. Zhao, "Challenges and pathways for tripling renewable energy capacity globally by 2030," *Chinese Journal of Urban and Environmental Studies*, Vol. 13, No. 1, 2550004, 2025.

[2] Liu, Y. and Z. Wang, "Robust and efficient multihorizon photovoltaic power forecasting with a dilated multi-scale transformer," *Solar Energy*, Vol. 302, 114077, 2025.

[3] Li, J. and Q. Liu, "Forecasting of short-term photovoltaic power generation using combined interval type-2 Takagi-Sugeno-Kang fuzzy systems," *International Journal of Electrical Power & Energy Systems*, Vol. 140, 108002, 2022.

[4] Erdener, B. C., C. Feng, K. Doubleday, A. Florita, and B.-M. Hodge, "A review of behind-the-meter solar forecasting," *Renewable and Sustainable Energy Reviews*, Vol. 160, 112224, 2022.

[5] Wang, K., X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Applied Energy*, Vol. 251, 113315, 2019.

[6] Gaboitaolelwe, J., A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod, and A. O. Salau, "Machine learning based solar photovoltaic power forecasting: A review and comparison," *IEEE Access*, Vol. 11, 40 820–40 845, 2023.

[7] Gigoni, L., A. Betti, E. Crisostomi, A. Franco, M. Tucci, F. Bizzarri, and D. Mucci, "Day-ahead hourly forecasting of power generation from photovoltaic plants," *IEEE Transactions on Sustainable Energy*, Vol. 9, No. 2, 831–842, 2018.

[8] Colbu, S.-C., D.-M. Bancila, and D. Popescu, "Long-term power generation prediction in photovoltaics using machine learning-based models," *Romanian Journal of Information Science and Technology (ROMJIST)*, Vol. 28, No. 1, 39–50, 2025.

[9] Rao, Z., Z. Yang, J. Li, L. Li, and S. Wan, "Prediction of photovoltaic power generation based on parallel bidirectional long short-term memory networks," *Energy Reports*, Vol. 12, 3620–3629, 2024.

[10] Kumar, A., Y. Kashyap, and R. Nasar, "Enhancing high-frequency PV power forecast using optimal hyperparameter setting in LSTM," in *International Conference on Sustainable Power and Energy Research*, 165–174, Warangal, India, 2024.

[11] Iheanetu, K. J., "Solar photovoltaic power forecasting: A review," *Sustainability*, Vol. 14, No. 24, 17005, 2022.

[12] Ait Abdelmoula, I., S. Elhamaoui, O. Elalani, A. Ghennioui, and M. E. Aroussi, "A photovoltaic power prediction approach enhanced by feature engineering and stacked machine learning model," *Energy Reports*, Vol. 8, 1288–1300, 2022.

[13] Xiao, Z., X. Huang, J. Liu, C. Li, and Y. Tai, "A novel method based on time series ensemble model for hourly photovoltaic power prediction," *Energy*, Vol. 276, 127542, 2023.

[14] Su, Z., C. Shi, K. Zhang, X. Xie, X. Zhang, and J. Xiao, "TransFCloudNet: A dual-branch feature fusion ground-based cloud image fine-grained segmentation method for photovoltaic power prediction," *Energy Conversion and Management*, Vol. 348, 120647, 2026.

[15] Zhang, R., Z. Xu, S. Liu, K. Fu, and J. Zhang, "Prediction of ultra-short-term photovoltaic power using BiLSTM-informer based on secondary decomposition," *Energies*, Vol. 18, No. 6, 1485, 2025.

[16] Fang, X., S. Han, J. Li, J. Wang, M. Shi, Y. Jiang, C. Zhang, and J. Sun, "A FCM-XGBoost-GRU model for short-term photovoltaic power forecasting based on weather classification," in *2023 5th Asia Energy and Electrical Engineering Symposium (AEEES)*, 1444–1449, Chengdu, China, Mar. 2023.

[17] Yu, Z., F. Wu, L. Chen, S. Zhu, and J. Zhang, "Photovoltaic power prediction model based on K-shape-NGO-CNN-BiLSTM with secondary decomposition," *Progress In Electromagnetics Research C*, Vol. 160, 183–195, 2025.

[18] Colominas, M. A., G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: A suitable tool for biomedical signal processing," *Biomedical Signal Processing and Control*, Vol. 14, 19–29, 2014.

[19] Zheng, J., H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mechanical Systems and Signal Processing*, Vol. 85, 746–759, 2017.

[20] Mirjalili, S. and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, Vol. 95, 51–67, 2016.

[21] Han, H., J. Peng, J. Ma, S. L. Liu, and H. Liu, "Research on load forecasting based on CEEMDAN SE VMD and SelfAttention TCN fusion model," *Scientific Reports*, Vol. 15, No. 1, 14530, 2025.

[22] Agga, A., A. Abbou, M. Labbadi, Y. El Houm, and I. H. O. Ali, "CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production," *Electric Power Systems Research*, Vol. 208, 107908, 2022.

[23] Liang, J., L. Yin, Y. Xin, S. Li, Y. Zhao, and T. Song, "Short-term photovoltaic power prediction based on CEEMDAN-PE and BiLSTM neural network," *Electric Power Systems Research*, Vol. 246, 111706, 2025.