(Received 14 October 2025, Accepted 3 December 2025, Scheduled 28 December 2025)

Short-Term Photovoltaic Power Prediction Based on SCC-CEEMDAN-HO-BiLSTM

Jianwei Liang[1], Jie Yue[1], Yanli Xin[2,*], Shuxin Pan[1], Jiaming Tian[1], and Jingxuan Sun[1]

[1] *School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, Jiangxi, China*
[2] *School of Automattion, Guangdong Polytechnic Normal University, Guangzhou 510665, Guangdong, China*

ABSTRACT: To address the challenge of high prediction difficulty caused by the random volatility of photovoltaic (PV) power output, this paper proposes a hybrid forecasting model that deeply integrates multi-scale feature analysis with an intelligent optimization algorithm. First, the spearman correlation coefficient (SCC) is used to select influencing factors as model inputs, and the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is applied to extract multi-scale features from the power data across four seasons. Second, the hippopotamus optimization (HO) algorithm is introduced in order to overcome the randomness and inefficiency of manual hyperparameter tuning and to optimize the hyperparameters of the bidirectional long short-term memory (BiLSTM) network. Through multi-seasonal case studies, the proposed SCC-CEEMDAN-HO-BiLSTM model outperforms conventional models. Specifically, it shows significant improvements in both prediction accuracy and robustness compared to benchmark methods such as the standalone BiLSTM model and the unoptimized CEEMDAN-BiLSTM model. The model effectively handles the multi-scale fluctuations in PV power sequences and meets the requirements for short-term photovoltaic power forecasting.

1. INTRODUCTION

In recent years, the PV power industry has experienced rapid growth, leading to a continuous expansion of installed capacity. According to data from the National Energy Administration of China, as of December 2022, the cumulative installed capacity of PV power in China had reached 55.76 GW, with a newly added installed capacity of 16.50 GW. PV power has now become a crucial component of the current power energy structure [1]. However, PV energy is affected by a variety of meteorological variables, resulting in volatility and nonlinearity; therefore, the integration of large-scale PV energy into the power grid may significantly impact the reliability and security of the grid [2].

At present, PV power prediction methods are mainly categorized into three types: physical models, statistical models, and hybrid models [3–5]. Physical models are based on the mechanism of solar radiation transmission and the photoelectric/thermodynamic characteristics of PV modules. They establish deterministic equations to predict power using numerical weather prediction (NWP) data and the physical parameters of PV power plants. Statistical models typically employ traditional methods to process historical photovoltaic power generation data, such as Kalman filtering and Bayesian regression. Due to their linear characteristics, statistical models are unsuitable for handling nonlinear and non-stationary data. In recent years, hybrid models have gained increasing popularity, as individual methods can be integrated within them to deliver superior performance in photovoltaic power forecasting compared to any single method alone.

Hybrid models generally consist of two main components: data preprocessing and forecasting. Ref. [6] utilized the SCC to select factors with significant impacts on PV output power as inputs for the prediction model, achieving favorable prediction performance. To mitigate the uncertainty and stochasticity inherent in photovoltaic power generation sequences, various signal decomposition methods have been widely adopted [7]. Common signal decomposition techniques include empirical mode decomposition (EMD) [8], ensemble empirical mode decomposition (EEMD) [9], and variational mode decomposition (VMD) [10]. In [8], a short-term PV power forecasting model was proposed based on EMD and extreme learning machine (ELM), which demonstrated high prediction accuracy. In [9], photovoltaic power was decomposed into low-frequency, mid-frequency, and high-frequency components. These components were then separately predicted using a variable-weight combined forecasting model, which further improved the forecasting accuracy. Similarly, in [10], the data were decomposed via VMD, and a hybrid model that combined the grey wolf optimizer (GWO) with long short-term memory (LSTM) networks was employed, leading to a significant enhancement in prediction accuracy. In [11], a deep temporal convolutional network (DeepTCN), complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) decomposition, and a multi-verse optimizer (MVO) were proposed for PV power forecasting. A case study utilizing real-time PV data from Alice Springs, Australia, demonstrated that the proposed method outperformed benchmark approaches across four conventional

* Corresponding author: Yanli Xin (yanlixin@gpnu.edu.cn).

performance metrics and two statistical tests, thereby validating its effectiveness for photovoltaic power generation prediction. Therefore, signal decomposition serves as a crucial data preprocessing step for enhancing forecasting performance. Although EMD and EEMD can improve prediction accuracy, they suffer from end-effects, which limit their performance. While VMD decomposition can effectively address the end-effect issue, it lacks adaptive capability. In contrast, CEEMDAN, as an advanced noise-assisted adaptive decomposition method, achieves significant improvements in the completeness of decomposition and the purity of its components. Thus, this study selects SCC and CEEMDAN for data preprocessing.

In [12], a combined analysis and forecasting model integrating graph convolutional networks (GCN) and LSTM networks was proposed and demonstrated robust performance. However, LSTM only considers unidirectional data information, neglecting the influence of reverse temporal dependencies. To address this, Ref. [13] employed a forecasting model based on fuzzy C-Means (FCM) and bidirectional long short-term memory (BiLSTM) networks, achieving improved prediction results. Although BiLSTM captures sequential patterns from both forward and backward directions, its prediction outcomes are relatively sensitive to parameter settings. The use of optimization algorithms can mitigate this sensitivity in BiLSTM parameter selection. It is noteworthy that different optimization algorithms may impact the final performance of photovoltaic power forecasting. Ref. [14] employed the sparrow search algorithm (SSA) to improve convergence speed, and Ref. [15] used the particle swarm optimization (PSO) algorithm to avoid the problem of premature convergence by leveraging its inherent advantages. While these algorithms have achieved certain results, they still suffer from issues such as easy trapping in local optima, slow convergence speed with the increase of search space dimension, and insufficient optimization accuracy. The HO algorithm is an emerging intelligent algorithm that exhibits the advantages of fast convergence speed and high optimization accuracy in addressing optimization problems.

Based on the above analysis, this study proposes a short-term PV power prediction model based on SCC-CEEMDAN-HO-BiLSTM, with the specific steps as follows: First, the SCC is used to screen the input features of the prediction model; second, CEEMDAN is applied to decompose the PV power data; next, the HO algorithm is used to optimize the hyperparameters of the BiLSTM neural network; finally, the HO-BiLSTM model is adopted to predict each subsequence individually, and the prediction results are superimposed to obtain the final PV power prediction value.

2. RESEARCH METHODOLOGY

2.1. Spearman Correlation Coefficient

In PV power prediction, including excessive non-critical factors increases model training time and compromises convergence speed and accuracy. Feature selection is therefore employed to reduce input variables, lowering computational load and enhancing prediction performance [16]. This study employs the spearman correlation coefficient (SCC) to identify the key factors influencing PV power, with the formula expressed as follows:

$$R = \frac{\sum_{i}^{N}(R_i - R)(S_i - S)}{\sum_{i}^{N}(R_i - R)^2 \sum_{i}^{N}(S_i - S)^2} \tag{1}$$

where R_i and S_i represent the ranks of the i-th observed values for two sorted variables; R and S denote the average ranks of the two variables; and N indicates the number of observations for each variable.

2.2. CEEMDAN

The decomposition process of CEEMDAN involves a series of iterative steps. First, specific Gaussian white noise is added to the original signal multiple times. Next, EMD is performed on each noise-augmented signal to obtain its first component, and all these components are ensemble-averaged to generate the first-order intrinsic mode function (IMF). This IMF is then subtracted from the original signal to produce the first-order residual. The process is iteratively repeated: during each iteration, new adaptive white noise is added to the current residual, and the next-order IMF is extracted via the steps described above while the residual is updated. Iterations continue until the residual can no longer be decomposed. Consequently, the original signal is decomposed into a series of IMFs and a final residual component [17].

2.3. HO

HO algorithm is a novel nature-inspired optimization algorithm proposed in 2024 [18]. By randomly generating initial candidate solutions and adaptively adjusting the resolution of the search space as well as the search speed, it can quickly and accurately find the optimal solution, featuring fast convergence speed and high solution accuracy. The main steps of HO are as follows:

(1) Initialization. Random initial solutions are generated, where each "hippopotamus" represents a candidate solution and is denoted by a vector. In this step, the following formula is used to generate the vector of decision variables:

$$\begin{aligned} &X_i : x_{ij} = lb_j + r\,(ub_j - lb_j) \\ &i = 1, 2, \ldots, \quad N, j = 1, 2, \ldots, m \end{aligned} \tag{2}$$

where the vector X_i represents the i-th candidate solution, i.e., the position of the i-th hippopotamus; x_{ij} is the position value of the i-th hippopotamus in the j-th decision variable; r is a random number within the range of $[0, 1]$; lb_j, ub_j denote the lower and upper bounds of the j-th variable to be optimized, respectively; N represents the number of hippopotamus individuals in the population; and m represents the number of variables to be optimized in the problem.

(2) The first phase (exploration phase). The positions of hippopotamuses in the river are updated. This phase ensures that

the solutions effectively explore the search space, as shown below:

$$X_i^{MH}: x_{ij}^{MH} = x_{ij} + r_1(D_{\text{hippo}} - I_1 x_{ij}) \quad i = 1, 2, \ldots, \left[\frac{N}{2}\right], j = 1, 2, \ldots, m \tag{3}$$

where X_i^{MH} represents the position of the male hippopotamus; x_{ij}^{MH} is the component of the male hippopotamus vector i in the j-th dimension; D_{hippo} is the hippopotamus with the optimal function value in the current iteration; and I_1 is equal to 1 or 2.

$$h = \begin{cases} I_2\overrightarrow{r_1} + (\sim Q_1) \\ 2\overrightarrow{r_2} - 1 \\ \overrightarrow{r_3} \\ I_1\overrightarrow{r_4} + (\sim Q_2) \\ r_5 \end{cases} \tag{4}$$

$$T = \exp\left(-\frac{t}{T}\right) \tag{5}$$

$$E = \begin{cases} x_{ij} + h_2\left(MG_i - D_{\text{hippo}}\right), r_6 > 0.5 \\ lb_j + r_7\left(ub_j - lb_j\right), else \end{cases} \tag{6}$$

$$X_i^{FH}: x_{ij}^{FH} = \begin{cases} x_{ij} + h_1\left(D_{\text{hippo}} - I_2 MG_i\right), T > 0.6 \\ E, else \end{cases} \tag{7}$$

where $\overrightarrow{r_1} \sim \overrightarrow{r_4}$ are random vectors within $[0, 1]$; $r_5 \sim r_7$ are random numbers within $[0, 1]$; Q_1 and Q_2 are integer random numbers taking 1 or 0; t represents the current iteration number; T denotes the maximum number of iterations; X_i^{FH} is the position of a female or immature hippopotamus; MG_i is the average value of hippopotamuses randomly selected from the group; and h_1 and h_2 are randomly selected from h.

The position update involves male, female, or immature hippopotamuses, where F_i represents the fitness function value, as shown in Equations (15)–(16).

$$X_i = \begin{cases} X_i^{\text{MH}}, F_i^{\text{MH}} < F_i \\ X_i, \text{else} \end{cases} \tag{8}$$

$$X_i = \begin{cases} X_k^{\text{FH}}, F_i^{\text{FH}} < F_i \\ X_k, \text{else} \end{cases} \tag{9}$$

(3) The second phase (exploration phase): Hippopotamuses defend against predators. In this phase, the algorithm utilizes a defense strategy inspired by the protective behavior of hippopotamuses to prevent premature convergence and enhance robustness in optimization.

$$\vec{D} = |P_j - x_{ij}| \quad i = \left[\frac{N}{2}\right] + 1, \left[\frac{N}{2}\right] + 2, \ldots, N, j = 1, 2, \ldots, m \tag{10}$$

$$X_i^R: x_{ij}^R = \begin{cases} \overrightarrow{RL} \oplus P_j + K \cdot \left(\frac{1}{\vec{D}}\right), F_{Pr_j} < F_i \\ \overrightarrow{RL} \oplus P_j + K \cdot \left(\frac{1}{2 \times \vec{D} + \overrightarrow{r_9}}\right), F_{Pr_j} \geq F_i \end{cases} \tag{11}$$

$$X_i = \begin{cases} X_i^{\text{R}}, F_i^{\text{R}} < F_i \\ X_i^{\text{R}}, F_i^{\text{R}} \geq F_i \end{cases} \tag{12}$$

where X_i^R represents the position of the hippopotamus facing the predator; P_j is the position of the predator in the solution space generated by the parameters to be optimized; $\vec{\text{D}}$ is the distance between the hippopotamus and the predator; $\overrightarrow{r_9}$ is an m-dimensional random vector; and $\overrightarrow{RL}$ is a random vector with a *Levy* distribution.

(4) The third phase (exploitation phase): It simulates the behavior of hippopotamuses escaping from predators. In this phase, when hippopotamuses evade threats, the algorithm adopts an escape strategy. When encountering a suboptimal region, it dynamically adjusts its position, explores other regions, and promotes global exploration.

$$X_i^E: x_{ij}^E = x_{ij} + r_{10} \cdot \left(\frac{lb_j}{\text{t}} + s_1\left(\frac{ub_j}{\text{t}} - \frac{lb_j}{\text{t}}\right)\right) \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, m \tag{13}$$

$$s = \begin{cases} 2 \times \overrightarrow{r_{11}} - 1 \\ r_{12} \\ r_{13} \end{cases} \tag{14}$$

$$X_i = \begin{cases} X_i^{\text{E}}, F_i^{\text{E}} < F_i \\ X_i^{\text{E}}, F_i^{\text{E}} \geq F_i \end{cases} \tag{15}$$

where X_i^E is the position of the hippopotamus searching for the nearest safe location, and s_i is randomly selected from s.

2.4. BiLSTM

LSTM can effectively model the long-term dependencies of sequential data, but its unidirectional structure only enables learning of historical context. However, information from future time steps is crucial for decision-making regarding the current state. To address this, the BiLSTM network was proposed. Its structure is shown in the Fig. 1. By combining two LSTM layers (one forward and one backward), BiLSTM achieves global context encoding of the sequence, thus exhibiting better performance in photovoltaic prediction tasks. The calculation process is as follows:

$$\begin{cases} h_t^{(1)} = \text{LSTM}\left(x_t, h_{t-1}^{(1)}\right) \\ h_t^{(2)} = \text{LSTM}\left(x_t, h_{t+1}^{(2)}\right) \\ p_t = \sigma\left(W_y \cdot \left[h_t^{(1)}, h_t^{(2)}\right] + b_y\right) \end{cases} \tag{16}$$

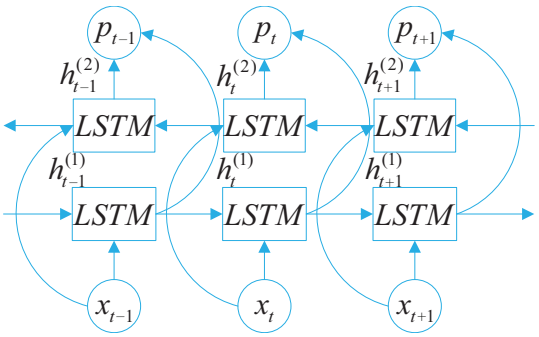

FIGURE 1. Structure of BiLSTM.

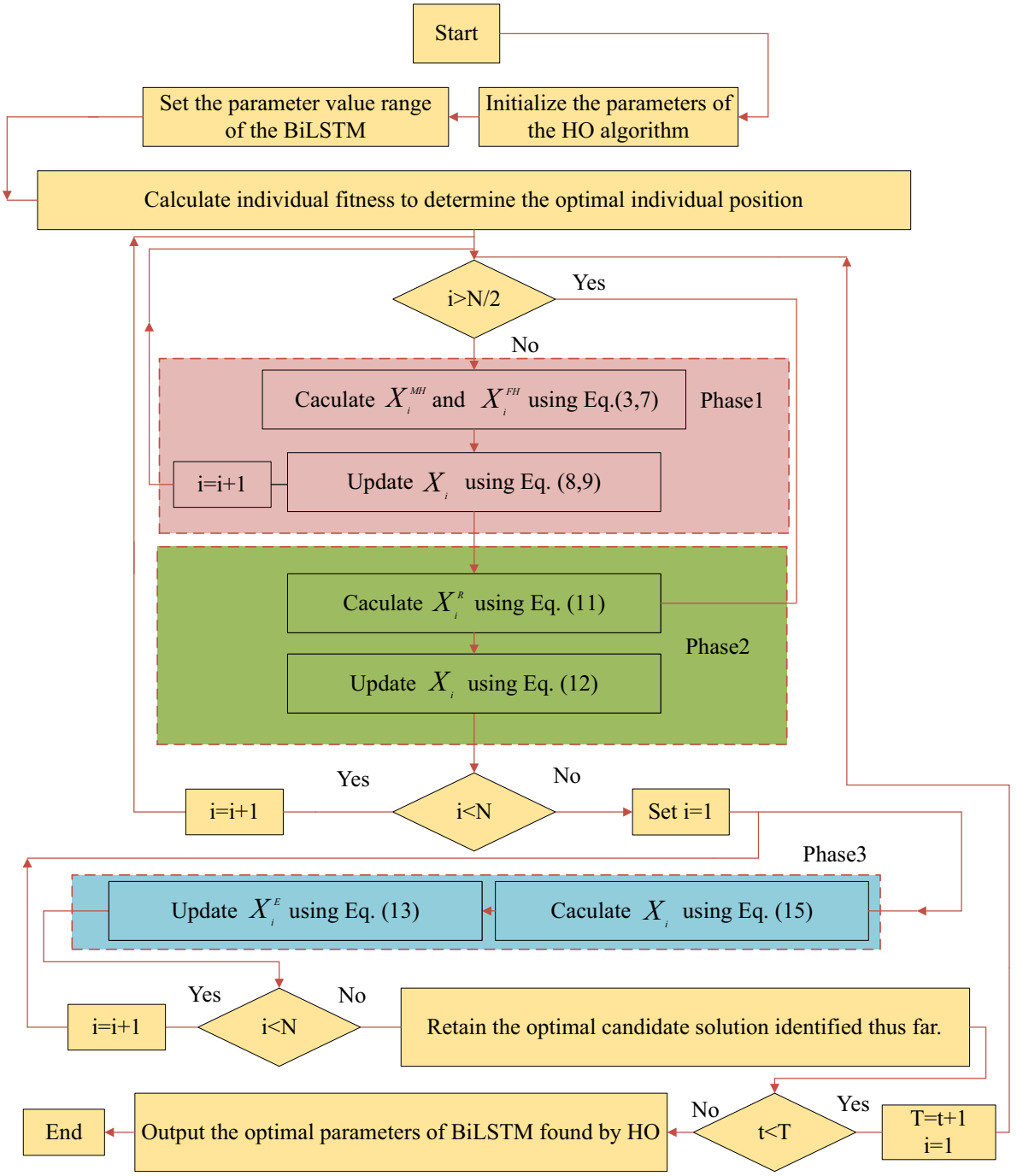

FIGURE 2. Flowchart of the HO-BiLSTM model.

where LSTM$(\cdot)$ denotes the computational process of a unidirectional LSTM; $h_t^{(1)}$ and $h_t^{(2)}$ represent the forward and backward hidden states at time step t, respectively; W_y and b_y correspond to the weight term and bias term, respectively.

2.5. HO-BiLSTM

The HO-BiLSTM hybrid model constructed in this paper combines HO and BiLSTM. Numerous studies have shown that although BiLSTM neural networks exhibit good fitting performance for time-series data, their convergence speed and generalization ability are still constrained by network hyperparameters such as the maximum number of iterations, the number of hidden layer neurons, and the learning rate [19]. Therefore, by utilizing the HO algorithm to perform global iterative optimization on key hyperparameters of the BiLSTM neural network (including the maximum number of iterations, the number of hidden layer neurons, and the learning rate), the fitting accuracy and prediction performance of the hybrid model can be effectively improved. The complete construction process of the HO-BiLSTM hybrid model is shown in Fig. 2.

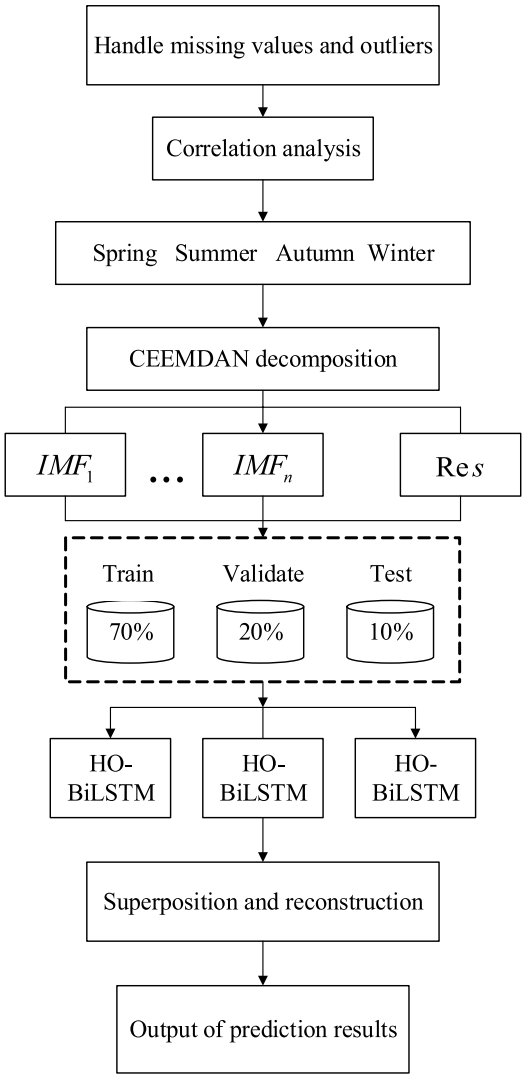

FIGURE 3. Prediction flow of PV power generation based on SCC-CEEMDAN-HO-BiLSTM.

3. SHORT-TERM PV POWER PREDICTION BASED ON SCC-CEEMDAN-HO-BILSTM

The flowchart of the short-term PV power prediction based on SCC-CEEMDAN-HO-BiLSTM is shown in Fig. 3.

First, missing values and outliers in photovoltaic-related data are processed to provide a high-quality data foundation for subsequent analysis and modeling.

Next, SCC is conducted to explore the degree of association between various influencing factors and PV output, and factors that have a significant impact on PV prediction are screened out.

Then, time is divided into four seasons (spring, summer, autumn, and winter) to account for the differential impacts of factors such as climate on photovoltaics across different seasons. CEEMDAN is applied to the PV data of each season, decomposing the data into several intrinsic mode functions (IMFs) and a residual component.

The decomposed data are divided into a training set (70%), a validation set (20%), and a test set (10%) in proportion, which are used for model training, parameter adjustment, and performance testing, respectively.

HO algorithm is employed to optimize the hyperparameters of the BiLSTM network, including the maximum number of iterations, the number of hidden layer neurons, and the learning rate. Subsequently, the model is trained and used for prediction on the training set, validation set, and test set, respectively.

Finally, the prediction results of each IMF and the residual component are superimposed and reconstructed to obtain the complete PV power prediction results.

4. CASE SIMULATION AND ANALYSIS

4.1. Data Set Source

The experimental data in this paper are the power generation data of a PV power station in Xinjiang, China, from January 1 to December 31, 2019. Each record includes module temperature, temperature, air pressure, humidity, global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI), where the data from March 1 to 31, June 1 to 30, September 1 to 30, and December 1 to 31 are selected to represent spring, summer, autumn, and winter, respectively. The sampling interval is 15 minutes. All experimental simulations are conducted on MATLAB R2024.

4.2. Feature Selection

Meteorological conditions are key factors affecting PV power generation, and it is essential to identify the primary meteorological factors that influence the power output of PV systems. In this study, SCC was conducted between meteorological data and historical power data, with the results presented in the Table 1.

TABLE 1. Correlation coefficients between PV power generation and environmental factors.

Weather characteristics	R
Module temperature	0.69
Temperature	0.29
Atmospheric pressure	−0.23
Humidity	−0.29
GHI	0.87
DNI	0.86
DHI	0.61

The spearman correlation coefficient R ranges from $[-1, 1]$, where a negative value indicates a negative correlation between two variables, while a positive value indicates a positive correlation; the absolute value of the coefficient is used to measure the strength of the correlation between variables. The commonly accepted criteria in the industry are as follows: an absolute value in the range of 0.8–1.0 indicates an extremely strong correlation between variables; a range of 0.6–0.8 indicates a strong correlation; a range of 0.4–0.6 indicates a moderate correlation; a range of 0.2–0.4 indicates a weak correlation; and a range of 0.0–0.2 indicates an extremely weak or almost no cor-

TABLE 2. Hyperparameter optimization results of HO for spring components.

Decomposition results	Epoch	Learning rate	Hidden dim	Decomposition results	Epoch	Learning rate	Hidden dim
IMF1	344	0.0014	33	IMF7	198	0.0018	20
IMF2	321	0.0068	26	IMF8	350	0.00053	94
IMF3	314	0.0064	81	IMF9	244	0.00015	80
IMF4	251	0.01	74	IMF10	227	0.0001	100
IMF5	350	0.0014	23	Res	314	0.00082	62
IMF6	345	0.00029	53				

relation between variables [20, 21]. Considering both calculation speed and prediction accuracy, the remaining input features after screening are module temperature, GHI, DNI, and DHI.

4.3. Evaluation Indexes

Model performance is evaluated using three key metrics, including root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) as defined below.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}[y_1(i) - y_2(i)]^2}{n}} \tag{17}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_1(i) - y_2(i)|}{n} \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[y_2(i) - y_1(i)]^2}{\sum_{i=1}^{n}[\overline{y}_1(i) - y_1(i)]^2} \tag{19}$$

where $y_1(i)$ represents the actual power of the i-th sampling instance; $y_2(i)$ represents the predicted power of the i-th sampling instance; $\overline{y}_1(i)$ represents the mean value of the actual power; and n represents the total number of samples in the test set.

4.4. CEEMDAN Decomposition Result Analysis

To deeply reveal the intrinsic multi-time-scale fluctuation characteristics of the PV power sequence, this study employs the CEEMDAN algorithm to analyze historical PV power data. The algorithm parameters are set as follows: noise intensity of 0.2, number of realizations of 100, and maximum number of iterations of 200. Taking spring as an example, the decomposition results are presented in Fig. 4, where the original power signal is adaptively decomposed into 11 IMFs and 1 Res. These components are strictly arranged in descending order of frequency, fully presenting the full-spectrum information on PV power, ranging from instantaneous disturbances to long-term trends. Through observation of the decomposition results, IMF1–IMF4 exhibit high fluctuation frequencies and strong randomness; IMF5–IMF7 show lower fluctuation frequencies and significant periodicity; while IMF8–IMF10 and Res have the smoothest fluctuations, reflecting the long-term variation trend of the power sequence.

TABLE 3. Comparison of model errors for different season types.

Season	Model	MAE	RMSE	R^2
Spring	LSTM	2.579	3.497	0.934
	BiLSTM	2.451	3.247	0.943
	SSA-BiLSTM	2.167	3.005	0.951
	HO-BiLSTM	2.111	2.860	0.956
	SCC-HO-BiLSTM	1.683	2.726	0.960
	CEEMDAN-BiLSTM	1.618	2.057	0.977
	CEEMDAN-HO-BiLSTM	1.184	1.469	0.988
	SCC-CEEMDAN-HO-BiLSTM	0.799	1.132	0.993
Summer	LSTM	3.429	4.653	0.867
	BiLSTM	2.665	4.470	0.878
	SSA-BiLSTM	2.458	4.410	0.881
	HO-BiLSTM	2.287	4.294	0.887
	SCC-HO-BiLSTM	2.196	4.239	0.890
	CEEMDAN -BiLSTM	1.838	2.598	0.958
	CEEMDAN-HO-BiLSTM	1.434	2.183	0.970
	SCC-CEEMDAN-HO-BiLSTM	1.291	2.130	0.977
Autumn	LSTM	3.457	4.500	0.880
	BiLSTM	2.846	3.967	0.907
	SSA-BiLSTM	2.670	3.853	0.913
	HO-BiLSTM	2.646	3.716	0.919
	SCC-HO-BiLSTM	2.163	3.241	0.938
	CEEMDAN -BiLSTM	2.070	2.920	0.949
	CEEMDAN-HO-BiLSTM	1.521	2.169	0.972
	SCC-CEEMDAN-HO-BiLSTM	1.006	1.391	0.991
Winter	LSTM	4.052	5.421	0.852
	BiLSTM	3.818	5.255	0.860
	SSA-BiLSTM	3.579	4.844	0.881
	HO-BiLSTM	3.280	4.575	0.894
	SCC-HO-BiLSTM	2.410	3.407	0.941
	CEEMDAN -BiLSTM	2.293	3.007	0.954
	CEEMDAN-HO-BiLSTM	1.475	1.871	0.982
	SCC-CEEMDAN-HO-BiLSTM	1.171	1.427	0.991

The core advantage of the components obtained by CEEMDAN decomposition lies in providing inputs with pure features and clear physical meanings for subsequent modeling. This

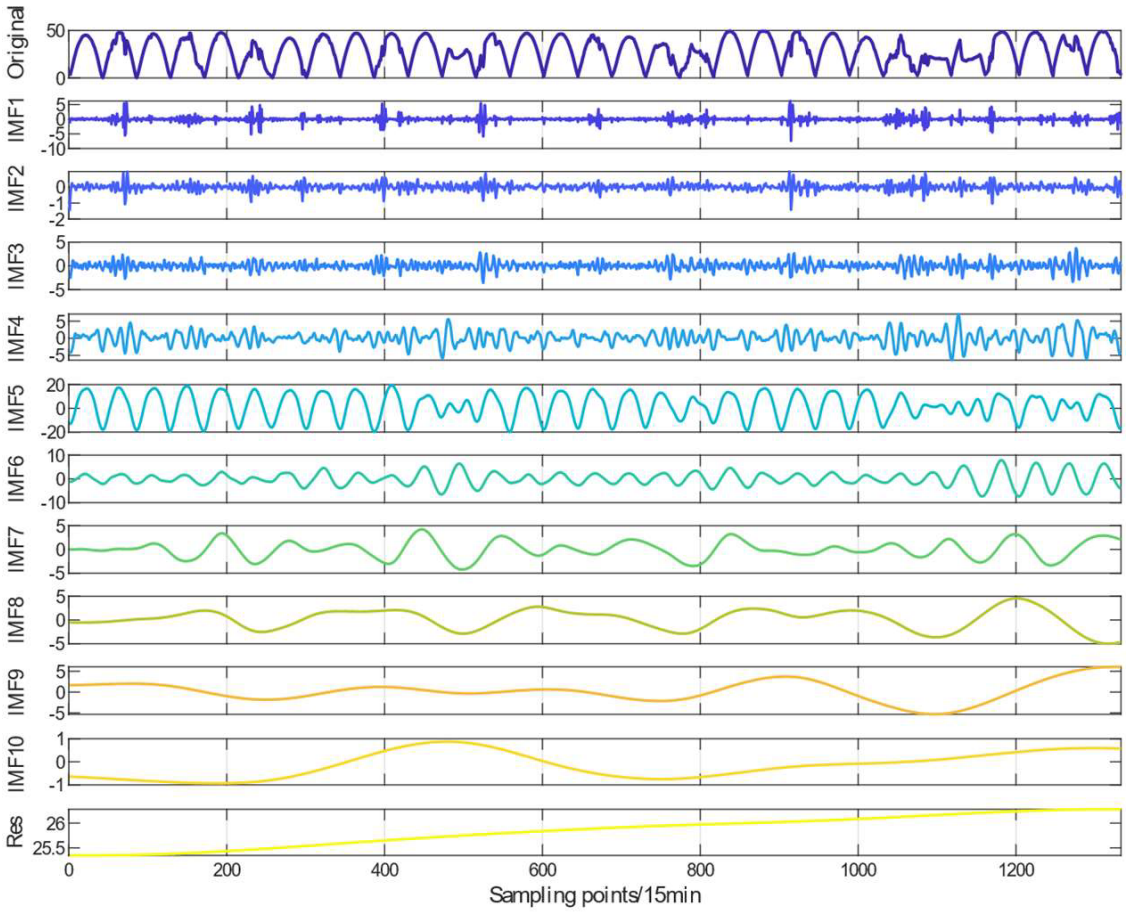

FIGURE 4. CEEMDAN decomposition of PV power generation signals under spring conditions.

inherent mode separation property enables the model to accurately characterize the physical mechanisms dominated by different frequency components, thereby laying a solid foundation for the construction of high-precision prediction models.

4.5. Hyperparameter Optimization Results

The HO algorithm mainly includes population size, maximum number of iterations, and dimension, where the population size and maximum number of iterations are set to 20, and the dimension is set to 3. In the process of using HO to optimize the hyperparameters of the BiLSTM network, the range of the maximum number of iterations (a hyperparameter of BiLSTM) is set between 100 and 350, the range of the number of hidden layer neurons set between 20 and 100, and the search range for learning rate update set between 0.0001 and 0.01.

Hyperparameter optimization was performed separately for the four seasons, and the optimization processes and results are presented in Fig. 5 and Table 2, with spring taken as an example.

4.6. Analysis of Prediction Results

To verify the effectiveness of the SCC-CEEMDAN-HO-BiLSTM photovoltaic power prediction model proposed in this paper, this section compares the overall performance of all models. Eight models, namely LSTM, BiLSTM, SSA-BiLSTM, HO-BiLSTM, SCC-HO-BiLSTM, CEEMDAN-BiLSTM, CEEMDAN-HO-BiLSTM, and the proposed SCC-CEEMDAN-HO-BiLSTM, were respectively applied to the four datasets corresponding to spring, summer, autumn, and winter. The MAE, RMSE, and R^2 of the models are presented in Table 3.

Across the four seasons, the LSTM model exhibits the worst prediction performance, while the CEEMDAN-HO-BiLSTM model achieves relatively good performance — but its prediction accuracy is still lower than that of the proposed SCC-CEEMDAN-HO-BiLSTM model. For the spring dataset: the LSTM model yields an MAE of 2.579, an RMSE of 3.497, and R^2 of 0.934; the CEEMDAN-HO-BiLSTM model obtains an MAE of 1.184, an RMSE of 1.469, and an R^2 of 0.988. Compared with the LSTM model (the worst-performing benchmark), the SCC-CEEMDAN-HO-BiLSTM model reduces MAE and RMSE by 69% and 67.6%, respectively, while increasing R^2 by 6.3%. Compared with the CEEMDAN-HO-BiLSTM model (a better-performing benchmark), it decreases MAE and RMSE by 32.5% and 22.9%, respectively, and improves R^2 by 0.5%. For the summer dataset: Relative to the LSTM model, the SCC-CEEMDAN-HO-BiLSTM model reduces MAE and RMSE by 62.3% and 54.2%, respectively, and increases R^2 by 14.9%. In contrast to the CEEMDAN-HO-BiLSTM model, it lowers MAE and RMSE by 9.9% and 2.4%, respectively, and raises R^2 by 0.7%. For the autumn dataset: Compared with the

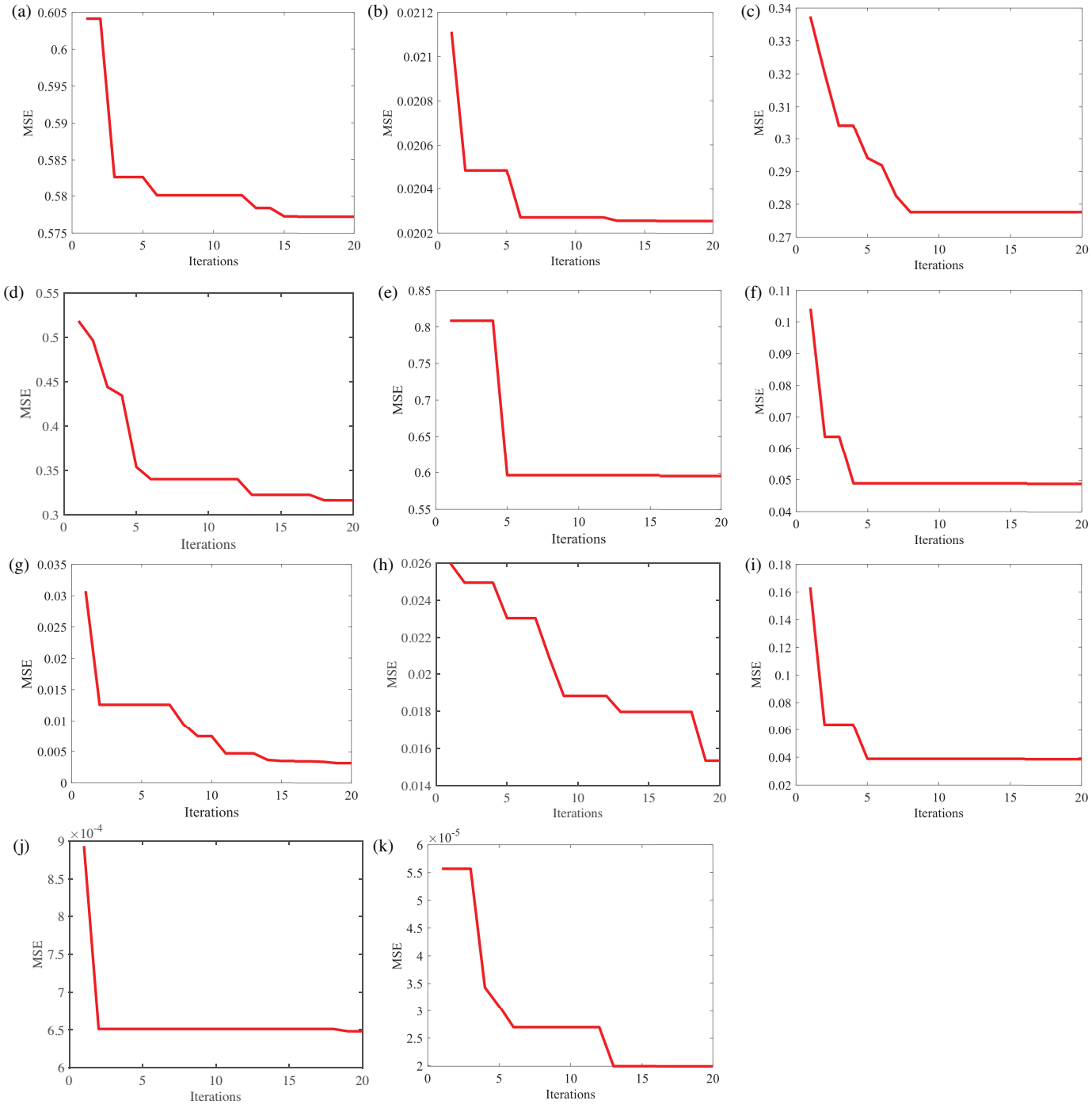

FIGURE 5. IMF1-IMF10 and Res optimization process in spring days. (a) IMF1. (b) IMF2. (c) IMF3. (d) IMF4. (e) IMF5. (f). IMF6. (g) IMF7. (h) IMF8. (i) IMF9. (j) IMF10. (k) Res.

LSTM model, the SCC-CEEMDAN-HO-BiLSTM model achieves a 70.8% reduction in MAE, a 69% reduction in RMSE, and a 12.6% increase in R^2 . When compared with the CEEMDAN-HO-BiLSTM model, it decreases MAE and RMSE by 33.8% and 35.8%, respectively, and improves R^2 by 1.9%. For the winter dataset: Relative to the LSTM model, the SCC-CEEMDAN-HO-BiLSTM model reduces MAE and RMSE by 71.1% and 73.6%, respectively, and increases R^2 by 16.3%. In comparison with the CEEMDAN-HO-BiLSTM model, it lowers MAE and RMSE by 20.6% and 23.7%, respectively, and raises R^2 by 0.9%.

The data in Table 3 fully confirm that hybrid models integrating optimization algorithms and feature engineering — such as the proposed models — outperform single neural networks like LSTM and BiLSTM in non-stationary PV power prediction, with better capture of volatile and seasonal features and higher adaptability to random and nonlinear fluctuations. HO-BiLSTM outperforms both BiLSTM and SSA-BiLSTM

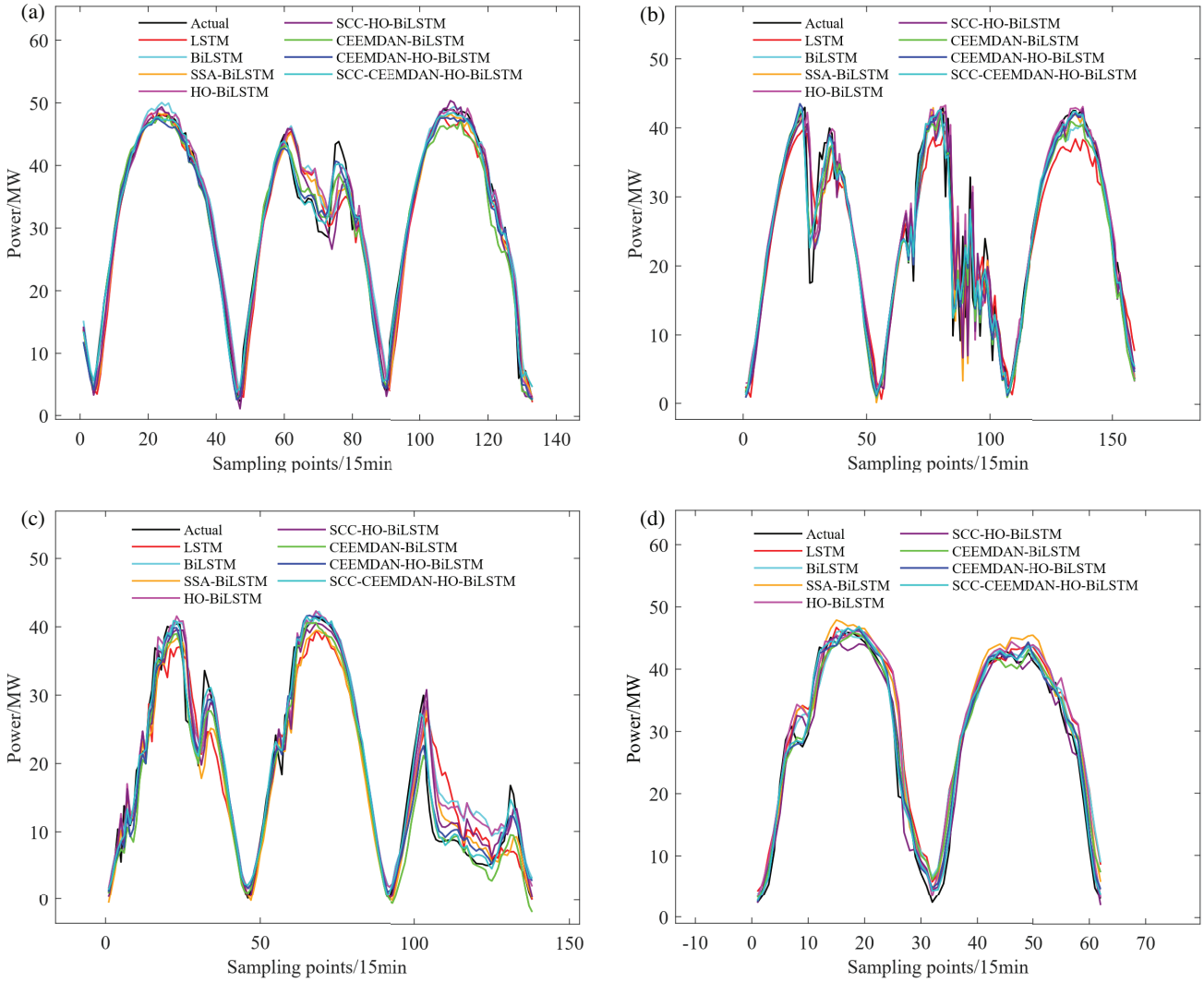

FIGURE 6. Comparison of prediction results of eight models. (a) Spring. (b) Summer. (c) Autumn. (d) Winter.

across all seasons, attributable to the HO algorithm's superior convergence efficiency and global optimization capability in hyperparameter tuning, which enhances feature extraction. CEEMDAN-BiLSTM and CEEMDAN-HO-BiLSTM surpass non-decomposition models including LSTM, BiLSTM, SSA-BiLSTM, HO-BiLSTM, and SCC-HO-BiLSTM, as CEEMDAN reduces complexity and noise by decomposing non-stationary data into stationary sub-sequences. Notably, SCC-CEEMDAN-HO-BiLSTM outperforms the other seven models: SCC correlation analysis combined with CEEMDAN-decomposed data optimizes input features by eliminating redundancy, cuts computational complexity and noise, and boosts prediction accuracy and stability. Comparisons of predicted vs. actual values in Fig. 6 further show that the SCC-CEEMDAN-HO-BiLSTM model has the curve closest to actual values, the smallest error variation, and the highest fitting degree across the three errors metrics, confirming its optimal performance.

To verify the generalizability of the method proposed in this paper, data from a site in Zhejiang Province, China, for October 2019 were selected. The proposed SCC-CEEMDAN-HO-BiLSTM model was compared with the CEEMDAN-HO-BiLSTM model. The training, validation, and test sets were divided in a ratio of 7 : 2 : 1, with data collected every 15 minutes. The corresponding evaluation metrics are presented in Table 4.

TABLE 4. Calculation of evaluation indicators.

Model	MAE	RMSE	R^2
CEEMDAN-HO-BiLSTM	2.299	3.173	0.980
SCC-CEEMDAN-HO-BiLSTM	1.155	1.3928	0.995

Analysis of Table 4 shows that, based on one month of data from another site, the prediction accuracy of the proposed SCC-CEEMDAN-HO-BiLSTM model has been improved.

4.7. Performance and Complexity Evaluation of the Model

For the statistical verification of performance differences, a quantitative analysis was conducted based on normalized error metrics using the Friedman test. The Friedman test is a nonparametric statistical method for comparing multiple forecasting models. When the test rejects the null hypothesis ($p <$

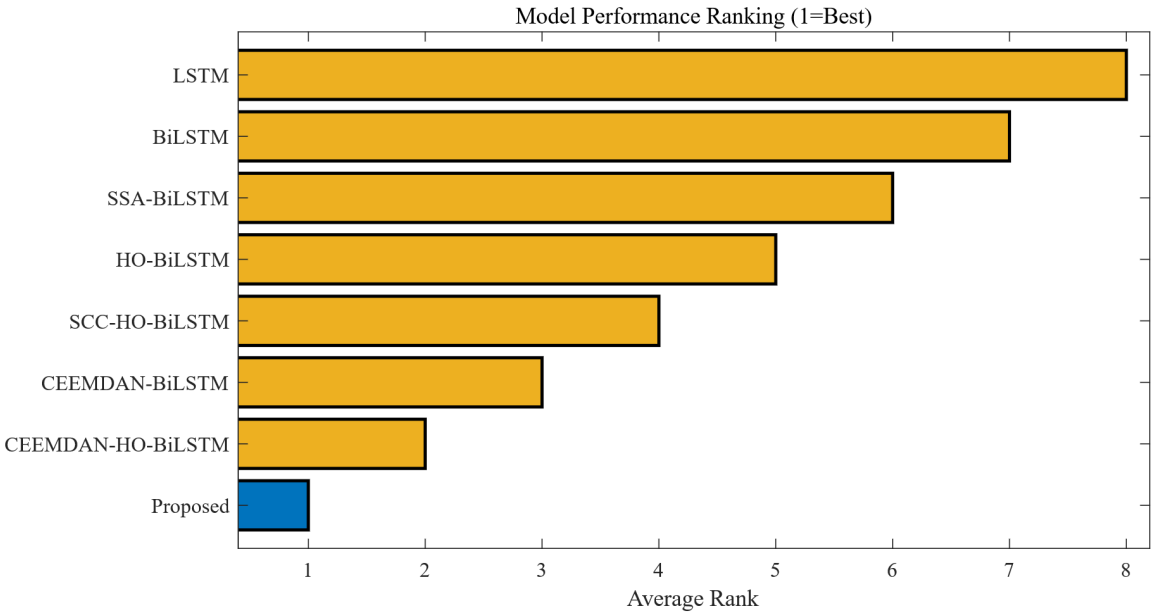

FIGURE 7. Ranking plot of model performance.

0.05), post hoc pairwise comparisons are typically performed to identify the specific models responsible for the observed differences. In this study, the result of the Friedman test was statistically significant, with $p = 0.00026 < 0.05$, indicating a notable performance difference among the evaluated models and necessitating post hoc analysis to determine the specific pairwise differences. As shown in Fig. 7, the model proposed in this study achieves the lowest average rank of 1, confirming its superior performance among all models. Furthermore, the proposed model exhibits an average RMSE of 1.591, demonstrating its excellent predictive accuracy and validating that its performance advantage is statistically significant.

Taking the spring data as an example, this study requires decomposing the photovoltaic power generation signal using CEEMDAN, which increases the training workload to 11 times of a single BiLSTM model. Consequently, the total training time increases to 166.09 seconds. Although the adoption of the HO algorithm considerably extends the training duration, its efficient optimization capability leads to a marked improvement in prediction accuracy. Finally, by utilizing feature selection to identify and incorporate key factors as inputs, the proposed model significantly reduces computational costs while enhancing prediction performance.

5. CONCLUSION

During the grid-connected operation of large-scale PV power plants, the stochastic and intermittent nature of PV power generation can compromise grid security and stability. Therefore, the accurate prediction of PV power output is of great significance. This paper proposes a combined deep learning model based on SCC-CEEMDAN-HO-BiLSTM for short-term PV power prediction. Simulated and experimental results indicate that, compared with traditional models lacking feature optimization, SCC correlation analysis enhances the quality of input features by screening key features strongly correlated with PV power and eliminating redundant information, which effectively reduces noise interference while lowering the computational complexity of the model. Meanwhile, the CEEMDAN method thoroughly decomposes the PV power series, thereby improving the prediction accuracy and stability of the HO-BiLSTM model. By leveraging the complementary advantages of different algorithmic components, the proposed hybrid model outperforms other benchmark models in terms of prediction accuracy and consistency, demonstrating considerable practical value for research in the field of PV power forecasting.

ACKNOWLEDGEMENT

This work is supported by the Guangdong Basic and Applied Basic Research Foundation (2022A1515110650) and by Science and Technology Projects in Guangzhou (2024A04J4760).

REFERENCES

[1] Zhu, H., Y. Sun, H. Zhou, Y. Guan, N. Wang, and W. Ma, "Intelligent clustering-based interval forecasting method for photovoltaic power generation using CNN–LSTM neural network," *AIP Advances*, Vol. 14, No. 6, 065329, 2024.

[2] Wang, Y., W. Li, H. Chen, Y. Ma, B. Yu, and Y. Yu, "Short-term photovoltaic power forecasting based on an improved zebra optimization algorithm — Stochastic configuration network," *Sensors*, Vol. 25, No. 11, 3378, 2025.

[3] Limouni, T., R. Yaagoubi, K. Bouziane, K. Guissi, and E. H. Baali, "Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model," *Renewable Energy*, Vol. 205, 1010–1024, 2023.

[4] Wang, X. and W. Ma, "A hybrid deep learning model with an optimal strategy based on improved VMD and transformer for short-term photovoltaic power forecasting," *Energy*, Vol. 295, 131071, 2024.

[5] Huang, S., Q. Zhou, J. Shen, H. Zhou, and B. Yong, "Multistage spatio-temporal attention network based on NODE for short-term PV power forecasting," *Energy*, Vol. 290, 130308, 2024.

[6] Wang, Q. and H. Lin, "Ultra-short-term PV power prediction using optimal ELM and improved variational mode decomposition," *Frontiers in Energy Research*, Vol. 11, 1140443, 2023.

[7] Zhang, C., T. Peng, and M. S. Nazir, "A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables," *Electric Power Systems Research*, Vol. 213, 108796, 2022.

[8] Khelifi, R., M. Guermoui, A. Rabehi, A. Taallah, A. Zoukel, S. S. M. Ghoneim, M. Bajaj, K. M. AboRas, and I. Zaitsev, "Short-term PV power forecasting using a hybrid TVF-EMD-ELM strategy," *International Transactions on Electrical Energy Systems*, Vol. 2023, No. 1, 6413716, 2023.

[9] Wang, H., J. Sun, and W. Wang, "Photovoltaic power forecasting based on EEMD and a variable-weight combination forecasting model," *Sustainability*, Vol. 10, No. 8, 2627, 2018.

[10] Xu, Z., K. Xiang, B. Wang, and X. Li, "Study on PV power prediction based on VMD-IGWO-LSTM," *Distributed Generation & Alternative Energy Journal*, Vol. 39, No. 03, 507–530, 2024.

[11] Huang, Y., A. Wang, J. Jiao, J. Xie, and H. Chen, "Short-term PV power forecasting based on CEEMDAN and ensemble DeepTCN," *IEEE Transactions on Instrumentation and Measurement*, Vol. 72, 1–12, 2023.

[12] Liao, W., B. Bak-Jensen, J. R. Pillai, Z. Yang, and K. Liu, "Short-term power prediction for renewable energy using hybrid graph convolutional network and long short-term memory approach," *Electric Power Systems Research*, Vol. 211, 108614, 2022.

[13] Cao, W., J. Zhou, Q. Xu, J. Zhen, and X. Huang, "Short-term forecasting and uncertainty analysis of photovoltaic power based on the FCM-WOA-BILSTM model," *Frontiers in Energy Research*, Vol. 10, 926774, 2022.

[14] Ma, W., L. Qiu, F. Sun, S. S. M. Ghoneim, and J. Duan, "PV power forecasting based on relevance vector machine with sparrow search algorithm considering seasonal distribution and weather type," *Energies*, Vol. 15, No. 14, 5231, 2022.

[15] Li, Y., L. Zhou, P. Gao, B. Yang, Y. Han, and C. Lian, "Short-term power generation forecasting of a photovoltaic plant based on PSO-BP and GA-BP neural networks," *Frontiers in Energy Research*, Vol. 9, 824691, 2022.

[16] Amer, H. N., N. Y. Dahlan, A. M. Azmi, M. F. A. Latip, M. S. Onn, and A. Tumian, "Solar power prediction based on Artificial Neural Network guided by feature selection for Large-scale Solar Photovoltaic Plant," *Energy Reports*, Vol. 9, 262–266, 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] Amiri, M. H., N. M. Hashjin, M. Montazeri, S. Mirjalili, and N. Khodadadi, "Hippopotamus optimization algorithm: A novel nature-inspired optimization algorithm," *Scientific Reports*, Vol. 14, No. 1, 5032, 2024.

[19] Gao, B., J. Xu, Z. Zhang, Y. Liu, and X. Chang, "Marine diesel engine piston ring fault diagnosis based on LSTM and improved beluga whale optimization," *Alexandria Engineering Journal*, Vol. 109, 213–228, 2024.

[20] Liang, J., L. Yin, S. Li, X. Zhu, Z. Liu, and Y. Xin, "Photovoltaic power prediction based on K-means++-BiLSTM-transformer," *Progress In Electromagnetics Research C*, Vol. 154, 191–201, 2025.

[21] 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.