



Multi-Scale Forecasting of Photovoltaic Power Based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Hybrid Neural Network

Qiming Yang¹, Daiyu Pang^{2*}, Wei Hu²

¹ Distribution Network Management Department, State Grid Suzhou Power Supply Company, 215031 Suzhou, China

² School of Economics and Management, Shanghai Electric Power University, 200090 Shanghai, China

* Correspondence: Daiyu Pang (pangdaiyu1211@163.com)

Received: 05-01-2025

Revised: 06-29-2025

Accepted: 07-14-2025

Citation: Q. M. Yang, D. Y. Pang, and W. Hu, “Multi-scale forecasting of photovoltaic power based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and hybrid neural network,” *J. Sustain. Energy*, vol. 4, no. 3, pp. 206–222, 2025. <https://doi.org/10.56578/jse040301>.



© 2025 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

Abstract: To address the challenge of limited photovoltaic (PV) power forecasting accuracy, which is primarily attributed to the significant impacts of abrupt weather changes and the strong non-stationarity of PV power time series, this paper proposes a multi-scale PV power forecasting model based on modified Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) and a hybrid neural network. First, key meteorological features including solar irradiance and ambient temperature are screened via the Pearson correlation coefficient (PCC), and the K-means clustering algorithm is adopted to construct three weather scenario datasets for sunny, cloudy, and rainy days, which effectively mitigates cross-scenario data distribution discrepancy. Second, the noise standard deviation and number of decomposition layers of the ICEEMDAN are dynamically optimized using the Dream Optimization Algorithm (DOA), achieving optimal modal decomposition and stationarization reconstruction of PV time series features. Subsequently, the Long Short-Term Memory (LSTM) network is utilized to deeply extract the periodic and trend characteristics embedded in the time series, which is combined with the multi-head attention mechanism from the Transformer architecture to effectively capture dynamic correlation information in the global time dimension. Finally, extensive experimental results demonstrate that the proposed PV forecasting method exhibits significant outperformance in both computational efficiency and forecasting accuracy under various weather conditions compared with state-of-the-art methods.

Keywords: Photovoltaic power forecasting; Dream Optimization Algorithm; Hybrid neural network; Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

1 Introduction

With the ongoing transition of the global energy mix and the advancement of global carbon neutrality targets, photovoltaic (PV) power generation, as a core pillar of clean renewable energy, has witnessed a continuous rise in its penetration in modern power systems [1]. However, PV power generation is highly susceptible to meteorological conditions, exhibiting strong stochasticity and intermittency, which leads to significant fluctuations in its power output and poses substantial challenges to the stable operation of power grids and power system dispatch [2]. Therefore, improving the accuracy of PV power forecasting is of great significance for enhancing the accommodation capacity of renewable energy, optimizing power system dispatch, reducing reserve capacity requirements, and ensuring the safe and stable operation of power grids [3].

Existing forecasting methods are mainly categorized into three classes: physical models, statistical models, and artificial intelligence-based machine learning models [4]. Among them, physical models perform forecasting primarily based on the physical characteristics of PV cells and meteorological parameters (e.g., solar irradiance, ambient temperature, wind speed), combined with radiative transfer theory, the power output characteristics of PV modules, and Numerical Weather Prediction [5]. Nevertheless, the forecasting accuracy of physical models is heavily dependent on high-precision meteorological data, and they are susceptible to errors in input parameters in complex environments, which inevitably leads to non-negligible forecasting deviations [6]. Statistical models mainly conduct time series modeling based on historical PV power data, and fit the data distribution to implement trend

forecasting through mathematical methods [7]. Typical statistical methods include the Autoregressive Integrated Moving Average model [8], the Generalized Autoregressive Conditional Heteroskedasticity model [9], and Support Vector Regression [10]. However, statistical models have inherent limitations when processing PV power data with strong nonlinearity and high stochasticity.

With the rapid development of deep learning, artificial intelligence-based methods have been extensively applied in PV power forecasting, mainly including neural network models, deep learning models, and hybrid modeling methods [11]. Typical neural network models include the Backpropagation Neural Network [12], Long Short-Term Memory (LSTM) network [13], Convolutional Neural Network (CNN) [14], and Transformer architecture [15]. The LSTM network exhibits outstanding performance in processing time series data and has been widely adopted in PV power forecasting. Wang et al. [16] proposed an LSTM-based PV power generation forecasting method, which verified the effectiveness of LSTM in capturing the inherent features of time series. However, the forecasting accuracy of a single LSTM model may be severely limited when dealing with complex non-stationary PV power series. To address this issue, Li et al. [17] combined Empirical Mode Decomposition (EMD) with LSTM to improve forecasting performance, and developed a PV power forecasting model based on EMD and LSTM. This model first decomposes the original data via EMD, and then uses LSTM to forecast each decomposed component, achieving favorable forecasting results. To further enhance the forecasting capability of the model, the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method has been introduced. Zhang et al. [18] proposed a PV power forecasting model based on CEEMDAN and LSTM, which decomposes the PV power series via CEEMDAN to reduce the non-stationarity of the original data, and then adopts LSTM to forecast each decomposed component, thus improving the forecasting accuracy. Nevertheless, CEEMDAN may still lose partial time series information when processing high-dimensional features. Accordingly, Sheng and Zhang [19] introduced the Transformer architecture to enhance the model's ability to capture dynamic correlation information. The Transformer can effectively process the correlation information in the global time dimension through its multi-head attention mechanism, and is particularly suitable for capturing the minute-level fluctuations of PV power.

However, existing methods still have notable shortcomings in dynamic parameter optimization and multi-modal feature fusion, which severely restrict their forecasting performance under complex and volatile weather conditions. In view of this, targeting the core challenges of PV power forecasting, this paper develops a high-precision multi-scale PV power forecasting model based on the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) and a hybrid neural network, which is tailored for different weather scenarios. The proposed model can provide technical reference for the dispatch optimization of power systems and renewable energy accommodation, and is of great practical significance for improving the utilization efficiency of renewable energy.

2 Data Preprocessing

2.1 Correlation Analysis

To optimize the performance of the forecasting model and reduce computational resource consumption, it is essential to perform rigorous screening of meteorological features for PV power forecasting to improve data quality. In the feature selection procedure, the Pearson Correlation Coefficient (PCC), a well-established statistical metric, is adopted to evaluate the linear correlation between variables. It characterizes the strength of the association between two variables by quantifying the ratio of their covariance to the product of their standard deviations. The specific mathematical expression is given as follows:

$$P_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y) \quad (2)$$

$$\sigma_X \sigma_Y = \sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)} \quad (3)$$

where, $\text{cov}(X,Y)$ denotes the covariance between X and Y ; σ_X and σ_Y represent the standard deviations of X and Y , respectively; $E(\cdot)$ is the mathematical expectation; $P_{X,Y}$ is the PCC with a value range of $[-1,1]$. A positive value of $P_{X,Y}$ indicates a positive correlation, while a negative value indicates a negative correlation, and the larger the absolute value of $P_{X,Y}$, the stronger the correlation between the variables.

In this paper, the PV operation data of a PV power station aggregated by a virtual power plant in January, April, July, and October 2019 are taken as the research object. The dataset covers a total of 123 days with a 15-minute sampling interval, containing 11808 sets of sample data. The aforementioned months correspond to the core periods of winter, spring, summer, and autumn, respectively, which can fully cover the seasonal differences in annual solar radiation intensity, ambient temperature, sunshine duration, and typical weather patterns throughout the year, thus

ensuring sufficient diversity and representativeness of the training data in terms of seasonal characteristics and key meteorological elements. This data selection strategy can not only enable the model to effectively learn the annual time series variation law of PV power generation, but also greatly reduce the computational load caused by full-year complete data, and improve the feasibility of model training and experimental iteration.

The meteorological features of the dataset include ambient temperature, azimuth angle, Cloud Opacity (CLOp), Dew Point Temperature (Td), Diffuse Horizontal Irradiance (DHI), Direct Normal Irradiance (DNI), Global Horizontal Irradiance (GHI), Global Tilted Irradiance (GTI), Tracked Tilted Irradiance (TTI), Precipitable Water Vapor (PWV), relative humidity, snow depth, surface atmospheric pressure, 10-m wind direction, 10-m wind speed, zenith angle, and actual PV power. The above influencing factors for PV power forecasting are numbered sequentially from 1 to 17, and the corresponding PCC calculation results are shown in Table 1 and Figure 1.

Table 1. Correlation analysis results

Features	Ambient Temperature	Azimuth Angle	CLOp	Td	DHI	DNI	GHI	GTI
r	0.398	-0.035	-0.261	0.113	0.722	0.873	0.998	0.970
Features	TTI	PWV	Relative Humidity	Snow Depth	Surface Atmospheric Pressure	10-m Wind Direction	10-m Wind Speed	Zenith Angle
r	0.968	0.099	-0.418	-0.178	0.001	0.028	0.125	-0.826

Note: CLOp = Cloud Opacity; Td = Temperature; DHI = Diffuse Horizontal Irradiance; DNI = Direct Normal Irradiance; GHI = Global Horizontal Irradiance; GTI = Global Tilted Irradiance; TTI = Tracked Tilted Irradiance; PWV = Precipitable Water Vapor.

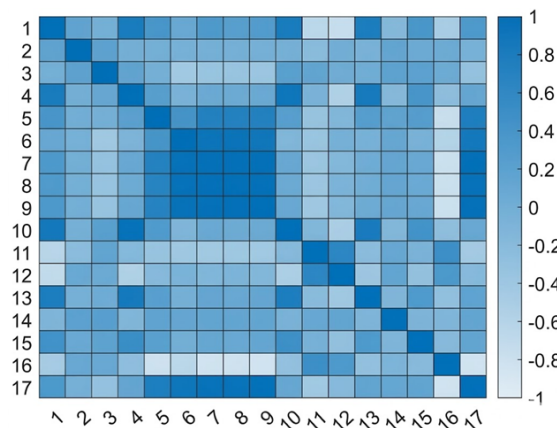


Figure 1. Pearson correlation coefficient (PCC) calculation results

As shown in Table 1, GHI presents a highly significant positive correlation with the output power of the PV system, with a correlation coefficient of 0.998, and its statistical significance ranks first among all meteorological parameters. This is followed by GTI and TTI, with correlation coefficients of 0.970 and 0.968, respectively, both showing strong correlations. In addition, the zenith angle, DNI and DHI have correlation coefficients of -0.826, 0.873 and 0.722, respectively, which also exhibit significant correlations. Relative humidity and ambient temperature show moderate negative correlation and weak positive correlation, respectively. Among them, although ambient temperature has a weak correlation, it still has engineering value due to its compensation effect on PV module temperature. Parameters such as 10-m wind speed, snow depth, surface atmospheric pressure, PWV and wind direction all show weak correlations. Therefore, ambient temperature, DHI, DNI, GHI, GTI, TTI, relative humidity and zenith angle are selected as the input parameters of the model.

PV time series data are prone to problems such as data missing and outlier noise, which require systematic preprocessing before being input into the forecasting model. For the measured data of the virtual power plant-integrated PV power station, this paper implements missing value imputation and outlier correction, to ensure the accurate correspondence between each variable and the time series, effectively improve data reliability, and lay a foundation for subsequent data processing and forecasting modeling.

In view of the diversity of data types, differences in data dimensions may lead to bias in clustering results, and different weight allocations will also affect the accuracy of the forecasting algorithm. Therefore, the min-max normalization method is adopted in the experiment of this paper to normalize the data, which converts the data of

each variable into dimensionless values in the interval of 0 to 1. This method not only retains the variation trend of the original data, but also reduces the computational complexity of the program. Its mathematical expression is as follows:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

where, x_i^* denotes the normalized data; x_i represents the input feature variables or output power in the PV dataset.

2.2 K-Means Clustering

To meet the demand of PV power forecasting, this study proposes a feature-oriented weather category integration method based on a comprehensive evaluation of existing meteorological classification systems, which combines the energy conversion characteristics of PV power generation systems and the requirements of forecasting modeling. This method classifies complex meteorological conditions into three categories: sunny, rainy, and cloudy days. The classification is based on the following criteria: different clusters show significant differentiation in key indicators such as irradiance and humidity, and the time series curves of PV output power under the same category have high morphological similarity. To better characterize the weather features of a single day, quantitative indicators are constructed using the maximum and average values within a unit time interval. Finally, CIOp and DHI are selected in this chapter to construct feature variables for clustering.

The clustering results are presented in Figure 2, which yields 49 sunny days, 32 rainy days, and 42 cloudy days, indicating favorable differentiation among the three weather clusters. It can be observed from the figure that sunny days have a relatively low maximum DHI and low cloud cover, rainy days have a high maximum DHI and high cloud cover, while cloudy days fall between the two. This significant distribution difference verifies the effectiveness of the selected feature variables and the rationality of the clustering results.

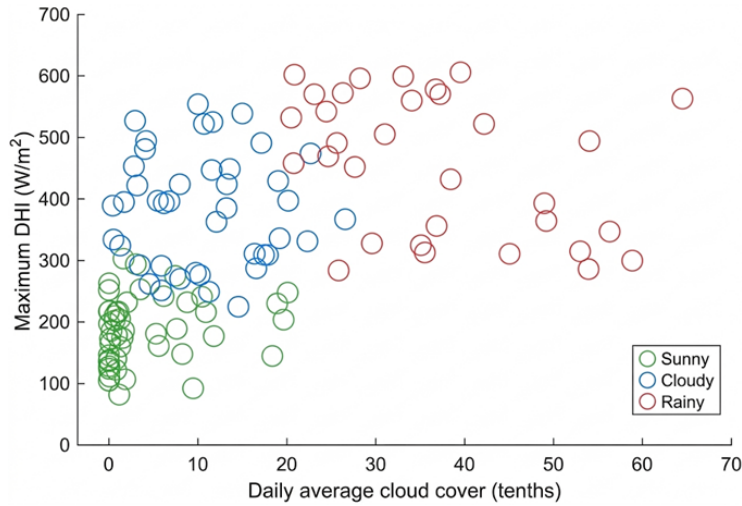


Figure 2. K-means clustering results

The PV power generation characteristics under different weather types are significantly distinct. On sunny days, the power generation is relatively stable and maintains a high level. Thick cloud cover on rainy days leads to a sharp reduction in DNI; despite the high DHI on rainy days, the overall power generation is low with significant fluctuations. Under cloudy weather conditions, the cloud cover changes frequently, so the stability and overall level of power generation are between those of sunny and rainy days. This is consistent with the actual PV power generation situation, which further demonstrates the reliability of the clustering results.

In summary, the weather clustering method proposed in this paper can effectively adapt to the actual demand of PV power forecasting scenarios. By establishing a hierarchical data support system for weather categories, it provides a structured data architecture for the training and validation of subsequent forecasting models. For the three classification results, a stratified random sampling strategy is adopted for sample division. Under each type of meteorological condition, 30% of the data from the total samples are taken as an independent test set, and the remaining 70% are used for the supervised learning training phase. This allocation mechanism not only ensures category balance, but also guarantees the objectivity of the evaluation on the model's generalization ability.

3 Dream Optimization Algorithm-Optimized Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

3.1 Dream Optimization Algorithm

The Dream Optimization Algorithm (DOA) is a novel metaheuristic algorithm inspired by human dream features [20]. It simulates the optimization process effectively by mimicking human memory retention, forgetting and logical self-organization behaviors during dreaming. The DOA delivers excellent performance in handling complex optimization problems, with unique strengths in balancing global and local search. Its iterative process consists of four phases:

(1) Initialization Phase

The initial solution set is established through random sampling, with its spatial distribution satisfying the mathematical description in Eqs. (5)–(6), where the dimension of the solution matrix is determined by the optimization problem:

$$X_i = X_l + \text{rand} \times (X_u - X_l), i = 1, 2, \dots, N \quad (5)$$

where, N denotes the number of individuals, i.e., the population size; X_i represents the i -th individual in the population; X_l and X_u are the lower bound and upper bound of the search space, respectively; rand is a Dim -dimensional vector, with each dimension being a random number between 0 and 1. The obtained population can be expressed as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,j} & \dots & x_{1,\text{Dim}} \\ x_{2,1} & x_{2,2} & \dots & x_{2,j} & \dots & x_{2,\text{Dim}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i,1} & x_{i,2} & \dots & x_{i,j} & \dots & x_{i,\text{Dim}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,j} & \dots & x_{N,\text{Dim}} \end{bmatrix} \quad (6)$$

where, $x_{i,j}$ denotes the position of the i -th individual in the j -th dimension, and Dim represents the dimension of the optimization problem.

(2) Exploration Phase

In this phase, a grouped collaborative search mechanism is adopted, and the following three steps are performed:

a) Memory Inheritance

For individuals in group q , they retain the position information of the best individual within the group prior to the dreaming process, and reset their own position information to that of the best individual in the group:

$$X_i^{t+1} = X_{\text{best}q}^t \quad (7)$$

where, X_i^{t+1} denotes the i -th individual at iteration $t + 1$; $X_{\text{best}q}^t$ represents the best individual of group q at iteration t .

b) Dynamic Forgetting

The dynamic forgetting strategy integrates global and local search capabilities. Building upon the memory inheritance strategy, this strategy enables individuals to forget and self-organize the position information within the forgotten dimensions. The specific mathematical formulation is given as follows:

$$x_{ij}^{t+1} = x_{\text{best}q,j}^t + (x_{l,j} + \text{rand} \times (x_{u,j} - x_{l,j})) \times \frac{1}{2} \times \left(\cos \left(\pi \times \frac{t + T_{\max} - T_d}{T_{\max}} \right) + 1 \right) \quad (8)$$

$$j = K_1, K_2, \dots, K_{k_q}$$

where, x_{ij}^{t+1} denotes the position of the i -th individual in the j -th dimension at iteration $t + 1$; $x_{\text{best}q,j}^t$ represents the position of the best individual of group q in the j -th dimension at iteration t ; $x_{l,j}$ and $x_{u,j}$ are the lower bound and upper bound of the search space in the j -th dimension, respectively; t is the current iteration number; $T_{\max}$ is the maximum iteration number.

c) Information Sharing

The information sharing strategy in the DOA enhances the capability of escaping from local optima. Implemented in parallel with the dynamic forgetting strategy and executed subsequent to the memory inheritance strategy, this strategy allows individuals to randomly acquire the position information of other individuals within the forgotten dimensions. The specific mathematical formulation is given as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{m,j}^{t+1}, & m \leq i \\ x_{m,j}^t, & i < m \leq N \end{cases} \quad j = K_1, K_2, \dots, K_{k_q} \quad (9)$$

where, $x_{i,j}^{t+1}$ denotes the position of the i -th individual in the j -th dimension at iteration $t + 1$; m is a natural number randomly selected from the range $[1, N]$ during the update of each dimension.

(3) Exploitation Phase

In the exploitation phase (iteration count from T_d to $T_{\max}$), grouping is no longer performed. Prior to each dreaming phase, the best dream from the previous iteration of the entire population (i.e., the best individual from the previous iteration) is presented to the population. Then, the position of each individual in the forgotten dimensions is updated. All individuals in the population have the same number of forgotten dimensions, denoted as $K_r \cdot k_r$. forgotten dimensions are randomly selected from the D dimensions, denoted as $K_1, K_2, \dots, K_k$, and the positions in these dimensions are updated. This phase mainly implements two steps:

a) Global Memory Convergence

$$X_i^{t+1} = X_{\text{best}}^t \quad (10)$$

where, X_i^{t+1} denotes the i -th individual at iteration $t + 1$; X_{best}^t represents the best individual of group q at iteration t .

b) Directional Dimension Optimization

$$x_{i,j}^{t+1} = x_{\text{best},j}^t + (x_{l,j} + \text{rand} \times (x_{u,j} - x_{l,j})) \times \frac{1}{2} \times \left(\cos \left(\pi \times \frac{t}{T_{\max}} \right) + 1 \right) \quad (11)$$

$$j = K_1, K_2, \dots, K_{k_r}$$

where, $x_{i,j}^{t+1}$ denotes the position of the i -th individual in the j -th dimension at iteration $t + 1$; $x_{\text{best},j}^t$ represents the position of the best individual of the entire population in the j -th dimension at iteration t ; $x_{l,j}$ and $x_{u,j}$ are the lower bound and upper bound of the search space in the j -th dimension, respectively; rand is a random number between 0 and 1; t is the current iteration number; $T_{\max}$ is the maximum iteration number of the algorithm.

(4) Parameter Adaptive Mechanism

To ensure the stability and applicability of the algorithm, the parameters of the DOA are set as follows in this paper:

$$T_d = \frac{9}{10} \times T_{\max} \quad (12)$$

where, T_d denotes the maximum iteration number of the exploration phase; $T_{\max}$ represents the total maximum iteration number of the algorithm.

$$k_q = \text{randi} \left(\left\lceil \frac{\text{Dim}}{8 \times q} \right\rceil, \max \left\{ 2, \left\lceil \frac{\text{Dim}}{3 \times q} \right\rceil \right\} \right), q = 1, 2, 3, 4, 5 \quad (13)$$

where, $\text{randi}(a, b)$ denotes a random integer selected from the range a to b ; k_q represents the number of forgotten dimensions of group q during the exploration phase, and Dim denotes the dimension of the problem.

$$k_r = \text{randi} \left(2, \max \left\{ 2, \left\lceil \frac{\text{Dim}}{3} \right\rceil \right\} \right) \quad (14)$$

where, k_r denotes the number of forgotten dimensions in the exploitation phase; Dim represents the dimension of the problem.

In the exploration phase, the parameter u is used to adjust the ratio between the dynamic forgetting strategy and the information sharing strategy. When $\text{rand} < u$, the dynamic forgetting strategy is executed; otherwise, the information sharing strategy is implemented. In addition, u is set to 0.9 in this paper.

3.2 Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Algorithm

To tackle the nonlinear and non-stationary characteristics of PV power generation, the PV power sequence can be decomposed into multiple components of different frequencies, with high- and low-frequency components reconstructed separately. Corresponding forecasting models are then built for each component to complete the prediction. This method can better extract signal features, mitigate data fluctuation, and enhance forecasting accuracy. The existing CEEMDAN algorithm solves the mode mixing defect of EMD, but tends to produce residual noise and spurious modes in the decomposition process. With the in-depth development of relevant research, the ICEEMDAN algorithm is thus put forward, whose main decomposition steps are listed below:

(1) The decomposition sequence $x_i(t)$ is constructed by adding a group of white noise to the original PV power sequence:

$$x_i(t) = x(t) + \theta_0 E_1(\delta_i(t)) \quad (15)$$

where, θ_0 denotes the signal-to-noise ratio of the initial decomposition; $x(t)$ represents the original PV power sequence; $E_k(\cdot)$ is the k -th component obtained via EMD decomposition (in Eq. (15), $k=1$); $\delta_i(t)$ is the i -th added white noise.

(2) Calculate the residual $r_1(t)$ and the intrinsic mode function IMF_1 of the first decomposition:

$$\begin{cases} r_1(t) = \text{average}[M(x_i(t))] \\ \text{IMF}_1(t) = x(t) - r_1(t) \end{cases} \quad (16)$$

where, $M(\cdot)$ denotes the operation for obtaining the local mean of the signal.

(3) Calculate the second residual component $r_2(t)$ and the intrinsic mode function IMF_2 :

$$\begin{cases} r_2(t) = \text{average}[M(r_1(t) + \theta_1(t)E_2(\delta_i))] \\ \text{IMF}_2(t) = r_1(t) - r_2(t) \end{cases} \quad (17)$$

where, $r_2(t)$ is the residual from the second decomposition; $\theta_1(t)$ is the signal-to-noise ratio coefficient of the second decomposition; $E_2(\delta_i)$ is the second intrinsic mode function of the white noise after EMD decomposition; $\text{IMF}_2(t)$ is the second intrinsic mode function.

(4) Similarly, the k -th order residual component $r_k(t)$ and the intrinsic mode function IMF_k are calculated as:

$$\begin{cases} r_k(t) = \text{average}[M(r_{k-1}(t) + \theta_{k-1}E_k(\delta_i))] \\ \text{IMF}_k(t) = r_{k-1}(t) - r_k(t) \end{cases} \quad (18)$$

where, $r_k(t)$ denotes the residual of the k -th decomposition; $r_{k-1}(t)$ represents the $(k-1)$ -th order residual component; θ_{k-1} is the noise coefficient of the $(k-1)$ -th order; $E_k(\delta_i)$ is the k -th intrinsic mode function obtained via EMD decomposition of the white noise; $\text{IMF}_k(t)$ is the k -th order intrinsic mode function.

(5) Repeat step (4) to obtain all intrinsic mode functions and residual components.

3.3 Implementation Path of Dream Optimization Algorithm Optimization

(1) Parameter Coding Mapping

To realize the parameter optimization of ICEEMDAN via DOA, the mapping between the parameter space and the algorithm solution space needs to be established. Let the set of parameters to be optimized be:

$$\Phi = \{\beta, K_{\max}, \theta\} \quad (19)$$

where, β is the noise amplitude adjustment coefficient; $K_{\max}$ is the maximum allowable mode number; θ is the correlation coefficient threshold for IMF screening.

Subsequently, it is encoded into a D -dimensional solution vector of the DOA ($D=5$ is adopted to enhance the search flexibility):

$$X_i = [\beta, K_{\max}, \theta, I, \epsilon] \in R^5 \quad (20)$$

where, I denotes the number of noise additions (integer type); ϵ represents the threshold of the decomposition stop condition.

The parameter constraints are defined by the boundary terms in Eq. (5):

$$\begin{cases} X_l = [0.05, 3, 0.2, 50, 0.001] \\ X_u = [0.2, 15, 0.8, 300, 0.01] \end{cases} \quad (21)$$

where, X_l and X_u denote the lower bound and upper bound of the search space, respectively.

The matrix form of the initialization process is given as follows:

$$X_{\text{init}} = \begin{bmatrix} 0.05 + r_1(0.2 - 0.05) & [3 + r_2(12)] & [0.2 + r_3(0.6)] & [50 + r_4(250)] & [0.001 + R_5(0.009)] \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (22)$$

(2) Fitness Function Construction

A multi-criteria integrated fitness evaluation function is designed as follows:

$$F(\Phi) = \underbrace{w_1 \cdot \text{MSE}}_{\text{Reconstruction Accuracy}} + \underbrace{w_2 \cdot \text{CE}}_{\text{Mode Purity}} + \underbrace{w_3 \cdot \text{PC}}_{\text{Over-decomposition Penalty}} \quad (23)$$

The components in the above equation are defined as follows:

a) Mean Square Reconstruction Error (MSE)

$$\text{MSE} = \frac{1}{L} \sum_{t=1}^L \left(X(t) - \sum_{k \in \mathcal{S}} \text{IMF}_k(t) \right)^2, \mathcal{S} = \{k \mid \rho_k > \theta\} \quad (24)$$

where, ρ_k denotes the PCC between IMF_k and the original signal.

b) Composite Entropy Index (CE)

$$\begin{cases} \text{CE} = \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} [\text{SE}(\text{IMF}_k) + \text{DE}(\text{IMF}_k)] \\ \text{SE} = -\sum_{m=1}^M p_m \ln p_m, p_m = \frac{|\text{IMF}_k^{(m)}|}{\sum |\text{IMF}_k|} \\ \text{DE} = -\sum_{\pi} p(\pi) \ln p(\pi) \end{cases} \quad (25)$$

where, π represents the phase space pattern, which reflects the sequence complexity.

c) Parameter Penalty Term (PC)

$$\text{PC} = \lambda_1 \left(\frac{K}{K_{\max}} \right)^2 + \lambda_2 \cdot \delta(\beta) + \lambda_3 \cdot \delta(\theta) \quad (26)$$

where, $\frac{K}{K_{\max}}$ is the actual mode proportion, which is controlled by Eq. (13); $\frac{K}{K_{\max}}$ denotes the actual mode proportion, which is controlled by Eq. (13) and $K_{\max}$; $\delta(\beta)$ and $\delta(\theta)$ denote the noise gain penalty and threshold overflow penalty, respectively, with their specific expressions given as follows:

$$\delta(\theta) = \begin{cases} 0, & 0.3 \leq \theta \leq 0.7 \\ |\theta - 0.5|, & \text{else} \end{cases} \quad (27)$$

$$\delta(\beta) = \begin{cases} 0, \beta \in [\beta_{\min}, \beta_{\max}] \\ |\beta - 0.125|, & \text{else} \end{cases} \quad (28)$$

where, θ is the correlation coefficient threshold for IMF screening; β is the noise amplitude adjustment coefficient.

In the method proposed in this paper, the key parameters of ICEEMDAN (including the noise gain coefficient, mode number threshold, etc.) are encoded into a multi-dimensional solution space, and the initial parameter population covering the feasible domain is generated based on uniform distribution, which lays a foundation for global search. In the early stage of optimization, the grouped memory inheritance and cross-dimensional information sharing mechanism drive the dynamic perturbation of parameter combinations in the exploration space. Among them, the nonlinear modulation of noise amplitude and the random dimension permutation strategy effectively balance the global exploration capability. When the iteration enters the late exploitation phase, the algorithm automatically converges to the neighborhood of the historically optimal parameters, and performs fine-grained adjustment with cosine attenuation for sensitive dimensions, so as to realize the collaborative optimization of the mode screening threshold and the decomposition stop condition. During this process, the dynamic boundary constraint mechanism ensures the rationality of the physical meaning of parameters and the stability of the decomposition process through random regeneration and threshold overflow penalty. This method innovatively maps the mode mixing suppression requirement of ICEEMDAN to the fitness function of DOA, and realizes the self-perception of parameter sensitivity via the dream-inspired memory evolution mechanism. Its core advantage lies in that the signal decomposition accuracy and feature extraction efficiency are simultaneously improved through the bidirectional dynamic coupling of noise injection and mode screening.

4 Multi-Scale Photovoltaic Power Prediction Model

4.1 Long Short-Term Memory Neural Network

In view of the temporal correlation of PV power data's feature distribution, the LSTM network is adopted as the base network for the power prediction model. The internal recurrent unit of the traditional Recurrent Neural Network (RNN) fails to capture and transfer the functional relationship between preceding and subsequent feature signals. To solve this problem, the LSTM network is proposed as an improved RNN variant, with its topology shown in Figure 3.

The core of the LSTM network is the dynamic management of long sequence information via the collaborative operation of memory cells and gating mechanisms. As the information storage carrier, the memory cell continuously retains the core features of historical data, laying a foundation for capturing dependencies across time steps. The input gate uses the sigmoid function to weight the current input and generates candidate values with the tanh function to precisely control the injection of new information. The forget gate evaluates the timeliness of information in the

memory cell through the sigmoid function and dynamically filters out redundant or invalid historical segments. The output gate further converts the filtered memory state into the effective output at the current moment. The formulas derived from the information flow are as follows:

$$\begin{cases} f_t = S(W_f [x_t, h_{t-1}]^T) \\ i_t = S(W_i [x_t, h_{t-1}]^T) \\ o_t = S(W_o [x_t, h_{t-1}]^T) \\ C'_t = \tanh(W_c \cdot h_{t-1} + W_f \cdot x_t) \\ C_t = f_t \odot C_{t-1} + i_t \odot C'_t \\ h_t = o_t \odot \tanh(C_t) \end{cases} \quad (29)$$

where, f_t is the output of the forget gate; i_t is the output of the input gate; o_t is the output of the output gate; W_f is the weight matrix of the forget gate; W_i is the weight matrix of the input gate; W_o is the weight matrix of the output gate; C'_t is the cell state input at time t ; S is the sigmoid activation function, and $\tanh$ is the tanh activation function; $\odot$ denotes matrix multiplication.

The internal state of the LSTM consists of the dual representation of the cell state and the hidden state. Among them, the cell state is updated at each time step through the joint action of the input gate, forget gate and candidate values, realizing the progressive accumulation and correction of time-series information. As the phased output of the network, the hidden state is not only transferred to the next time step to maintain temporal continuity, but also can be directly used for the final prediction task. This operation mode, driven by the gating mechanism and with dual-state linkage, enables the LSTM to stably preserve long-term pattern features while sensitively responding to the short-term dynamic changes of the sequence, thus exhibiting excellent modeling capability in scenarios such as time series prediction and natural language processing.

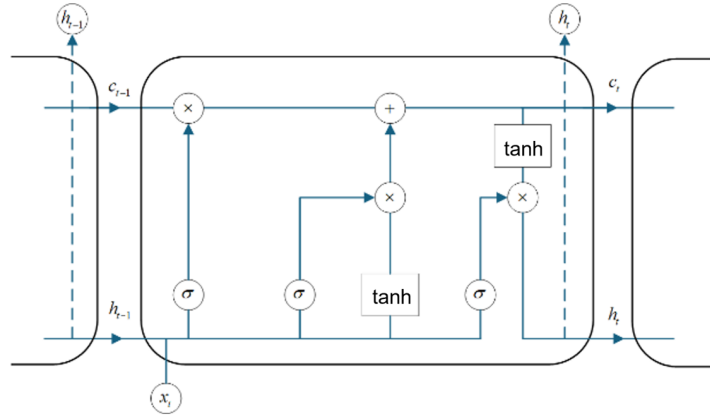


Figure 3. Network structure of Long Short-Term Memory (LSTM)

4.2 Transformer Network

Transformer is a deep learning model constructed based on the self-attention mechanism. With its superior capability of capturing the global correlation of data and support for efficient parallel computing, it has been extended from the field of natural language processing to tasks such as time series forecasting in recent years [21, 22]. As shown in Figure 4, the model is mainly composed of stacked encoders and decoders. Each layer internally contains a multi-head attention calculation unit and a feed-forward network component, and stable convergence of the training process is guaranteed through residual connections and layer normalization. This structural design enables it to exhibit significant advantages when processing data with temporal dependency characteristics such as PV power.

Compared with traditional RNNs, Transformer abandons the serial computing mode and adopts the self-attention mechanism to implement dynamic weight assignment for the long-range dependencies between

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (30)$$

The PV power forecasting accuracy is mainly restricted by the uncertainty and fluctuation characteristics of its output. Relying on the global modeling advantages of Transformer, this study constructs a joint feature representation for short-term PV power time series data and related multi-dimensional meteorological parameters, to simultaneously

capture the temporal dependency characteristics of the power sequence and the cross-variable correlation patterns between PV power and meteorological factors. In the model construction process, spatiotemporal encoding is first performed on the original power data and related meteorological parameters to form the initial input matrix. Then, the implicit patterns of PV power are mined through the multi-head attention mechanism of Transformer. Among them, the Query matrix (Q) integrates the temporal variation characteristics of the power sequence and the correlation parameters with meteorological variables, while the Key matrix (K) is associated with the dynamic characteristics of power output. The attention distribution is obtained through the interactive calculation of the two matrices, followed by weight normalization via the Softmax function, and the final predicted value of PV power is output.

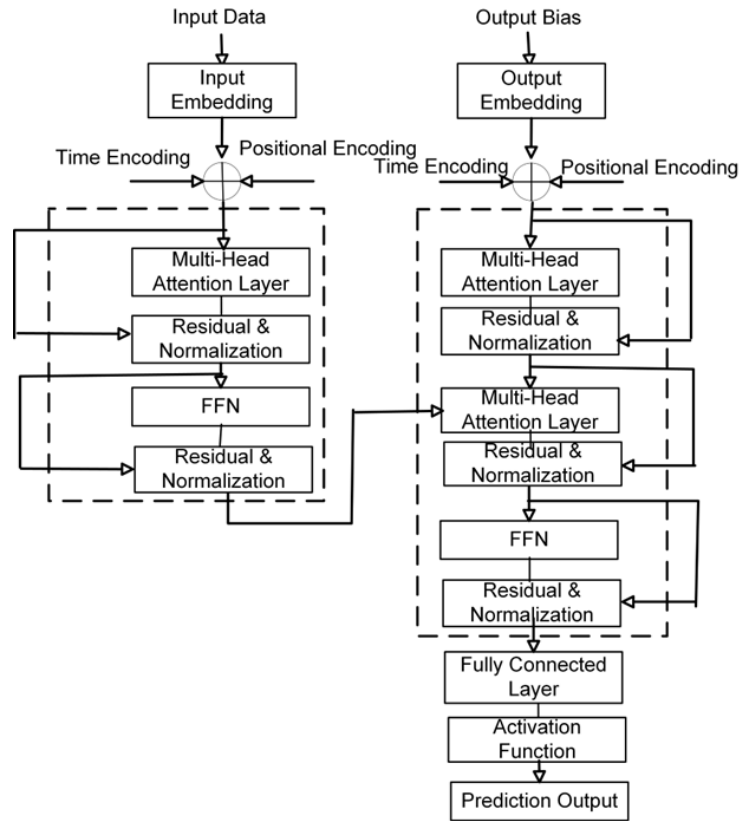


Figure 4. Structure of Transformer network

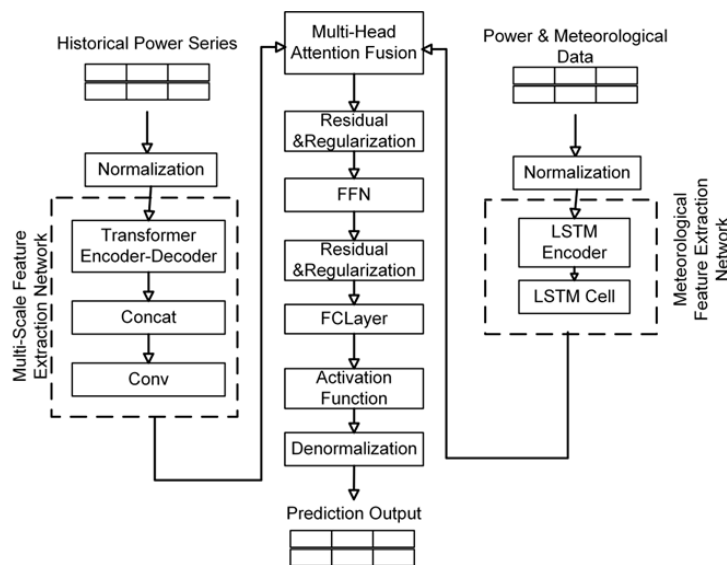


Figure 5. Photovoltaic (PV) multiscale prediction network structure

4.3 Multi-Scale Photovoltaic Power Prediction Model

The hybrid neural network model with multi-scale feature fusion proposed in this paper deeply integrates the improved ICEEMDAN decomposition and hybrid neural network architecture, aiming to address the key challenge of multi-scale feature mining for PV power sequences under multi-weather scenarios. Its network structure is shown in Figure 5.

The multi-scale PV power prediction model consists of three core modules: multi-modal decomposition, spatiotemporal feature fusion, and prediction output. In the feature extraction stage, the original PV power sequence is first decomposed into IMFs characterizing different physical processes via DOA-optimized ICEEMDAN, and key meteorological factors such as GHI and zenith angle are screened based on the PCC. Subsequently, an LSTM branch is adopted to perform memory modeling of the long-period trends of the IMF components, while a parallel Transformer channel is constructed to capture the cross-scale correlation patterns between abrupt meteorological events and power fluctuations using the multi-head attention mechanism. Compared with traditional single-scale prediction models, the hierarchical processing of decomposition, reconstruction and collaboration adopted in this paper significantly enhances the accuracy and robustness of the prediction results.

5 Case Study

5.1 Dataset Description and Preprocessing

In this section, air temperature, DHI, DNI, GHI, GTI, TTI, relative humidity, and zenith angle are selected as the input features of the prediction model according to the results of correlation analysis. Meanwhile, based on the analysis results of the K-means clustering algorithm, PV power prediction experiments are conducted separately for different weather conditions in the sample set, namely sunny days, cloudy days, and rainy days, to evaluate the prediction performance of the proposed model under various weather scenarios. The specific experimental data are detailed in Table 2.

Table 2. Experimental data under different weather conditions

Weather	Number of Days (d)	Data Volume (Samples)
Sunny	57	5472
Cloudy	27	2592
Rainy	33	3168

The historical PV power data are further decomposed via the DOA-ICEEMDAN method, with the decomposition results shown in Figure 6. It can be seen that the PV power data under sunny weather conditions are decomposed into 7 IMF components and 1 Res component, achieving accurate separation of the historical PV power generation sequence. Specifically, the IMF1 component presents high-frequency and low-amplitude fluctuation characteristics, corresponding to the short-term fluctuations of PV power. The IMF2 to IMF7 components gradually exhibit fluctuation patterns with lower frequency and higher amplitude, which reflect the variation trends of PV power at different time scales. The Res component is relatively stable, representing the long-term trend or DC component of the PV power data, namely the baseline output level of PV power generation under sunny conditions. Through this multi-scale decomposition, the DOA-ICEEMDAN method can characterize the intrinsic structure of the PV power sequence in a more detailed manner, which provides a more comprehensive and accurate data foundation for subsequent research including power forecasting, fault detection, and system optimization.

5.2 Evaluation Metrics

To comprehensively evaluate the prediction performance of the proposed model under multi-weather scenarios, the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) are adopted as the core evaluation metrics in this paper, with their respective definitions given as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (31)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (32)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (33)$$

where, y_i denotes the actual power value; $\hat{y}_i$ represents the predicted value; $\bar{y}$ is the mean value of the actual power; n is the number of samples.

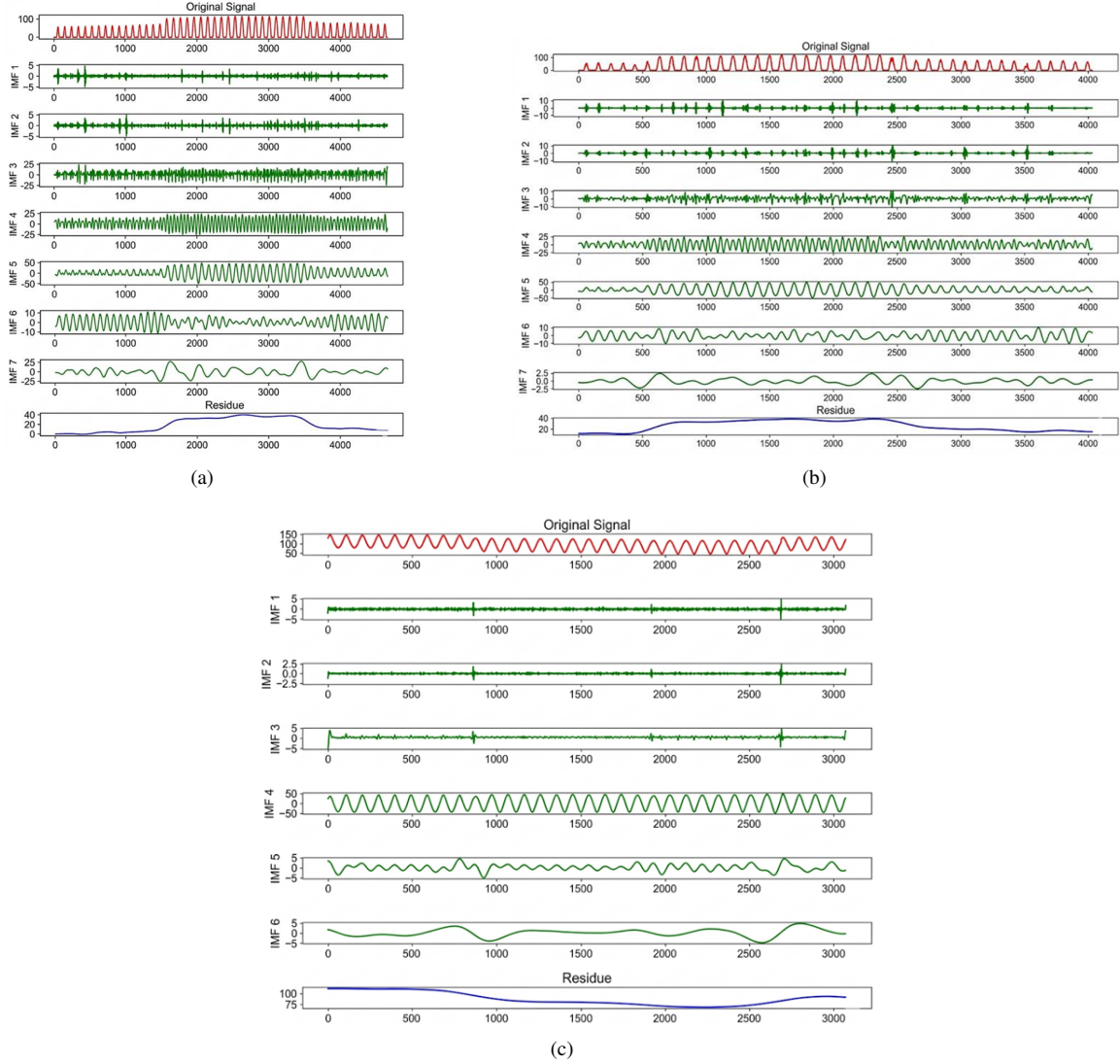


Figure 6. Photovoltaic (PV) power Dream Optimization Algorithm (DOA)- Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) decomposition results: (a) sunny weather; (b) cloudy weather; (c) rainy weather

5.3 Ablation Experiment

To verify the effectiveness of the proposed DOA-ICEEMDAN-LSTM-Transformer prediction model, in addition to the training, validation and testing of the proposed model, the ICEEMDAN-LSTM model (hereinafter abbreviated as IL), ICEEMDAN-LSTM-Transformer model (hereinafter abbreviated as ILT), LSTM-Transformer model (hereinafter abbreviated as LT) and DOA-ICEEMDAN-LSTM model (hereinafter abbreviated as DIL) are trained and tested on the same dataset, to compare the performance of the five models under three different weather conditions. The comparison of the prediction errors and the actual power curves of the five models are presented in Table 3 and Figure 7.

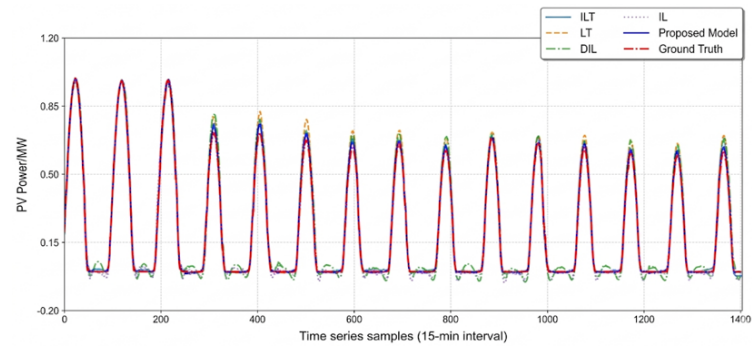
By comparing the proposed model with the ILT model without the dynamic optimization algorithm, the critical role of DOA in model parameter tuning can be clarified. As shown in Figure 7, in the sunny scenario, the prediction curve of the proposed model is almost completely coincident with the ground truth curve, while the prediction curve of the ILT model has an obvious deviation. It can be seen from Table 3 that the R^2 value of the proposed model reaches 0.9984, which is 7.4% higher than 0.9249 of the ILT model. Meanwhile, its RMSE and MAE are significantly reduced by 74.6% and 78.0%, respectively. This gap indicates that DOA effectively suppresses the mode mixing problem by dynamically optimizing the mode boundary threshold of ICEEMDAN decomposition, making

the decomposed sequences more adaptable to the illumination characteristics under stable weather conditions, thus significantly improving the fitting accuracy of the LSTM-Transformer module.

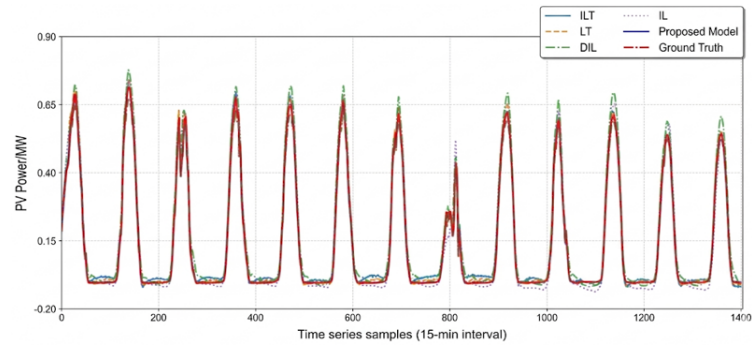
Table 3. Ablation experiment model prediction error results

Model	Sunny			Cloudy			Rainy		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Proposed model	0.9984	1.5717	1.2307	0.9918	1.1294	1.4766	0.9952	1.5790	1.1639
IL	0.7923	3.4199	2.5670	0.7945	2.5659	2.7264	0.8889	2.3870	2.4181
LT	0.8753	6.1372	3.7408	0.9057	5.3866	4.1373	0.8688	4.0112	2.8006
ILT	0.9249	6.1840	5.5898	0.8906	3.3580	2.4747	0.9536	3.6878	2.8306
DIL	0.7859	4.6369	3.8652	0.8938	2.7246	1.6500	0.6844	2.8321	2.0987

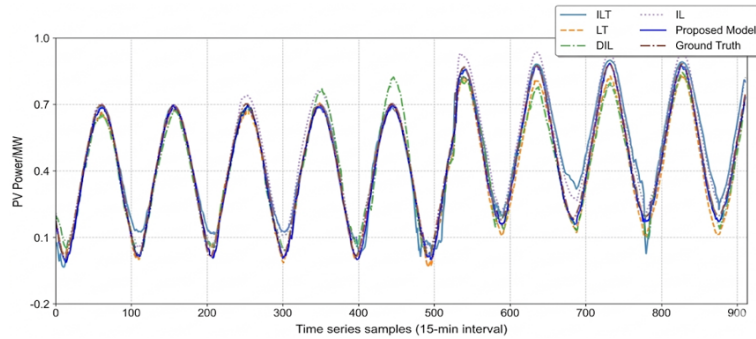
Note: IL = ICEEMDAN-LSTM model; ILT = ICEEMDAN-LSTM-Transformer model; LT = LSTM-Transformer model; DIL = DOA-ICEEMDAN-LSTM model; RMSE = Root Mean Square Error; MAE = Mean Absolute Error.



(a)



(b)



(c)

Figure 7. Ablation experiment prediction result curves: (a) sunny weather; (b) cloudy weather; (c) rainy weather

By comparing with the DIL model without the improved decomposition method, the critical value of the Transformer module in complex weather modeling can be verified. In rainy day prediction, the R^2 value of the

proposed model reaches 0.9952, which is 45.4% higher than 0.6844 of the DIL model. Its RMSE and MAE are 1.5790 and 1.1639, respectively, which are reduced by 44.2% and 44.6% compared with 2.8321 and 2.0987 of the DIL model. This indicates that the DIL model without the Transformer module is difficult to capture the abrupt power fluctuations caused by rainfall, while the multi-head attention mechanism of Transformer can assign higher weights to the abrupt time steps in the historical sequence. Combined with the curve comparison of the ablation experiment in Figure 7, it can be inferred that the prediction curve of the proposed model is highly consistent with the ground truth near the abrupt change points, while the prediction curve of the DIL model often shows underestimation or hysteresis, which highlights the importance of temporal feature weighting capability.

By analyzing the performance difference between the LT model and the full proposed model in cloudy scenarios, the enhancement effect of ICEEMDAN signal decomposition on deep networks can be revealed. Under cloudy conditions, the prediction curve of the proposed model can accurately capture the PV power fluctuations in cloudy weather, while the prediction curve of the LT model is relatively rough. In addition, according to the results in Table 3, the MAE of the proposed model under cloudy conditions is 1.4766, which is reduced by 64.3% compared with 4.1373 of the LT model, and its RMSE is also reduced from 5.3866 to 1.1294, with a reduction rate of 79.0%. This improvement originates from that ICEEMDAN decomposition reconstructs the original PV sequence into IMFs at different time scales, explicitly separates the high-frequency noise and low-frequency trend components in cloudy weather, enabling the LT model to learn the evolution laws of different modes separately. Then, the prediction results of each mode are dynamically fused through DOA, and finally high-precision modeling of intermittent irradiation characteristics is realized.

5.4 Comparative Experiment

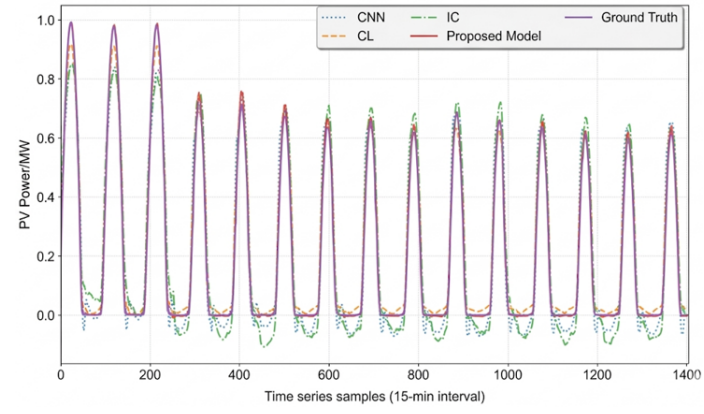
To verify the superiority of the prediction capability of the proposed model, the CNN model (hereinafter abbreviated as CNN), CNN-LSTM model (hereinafter abbreviated as CL), and ICEEMDAN-CNN model (hereinafter abbreviated as IC) are selected for comparison with the proposed model. The prediction error results are shown in Table 4, and the comparative experimental curves are presented in Figure 8.

Table 4. Comparison of experimental model prediction error results

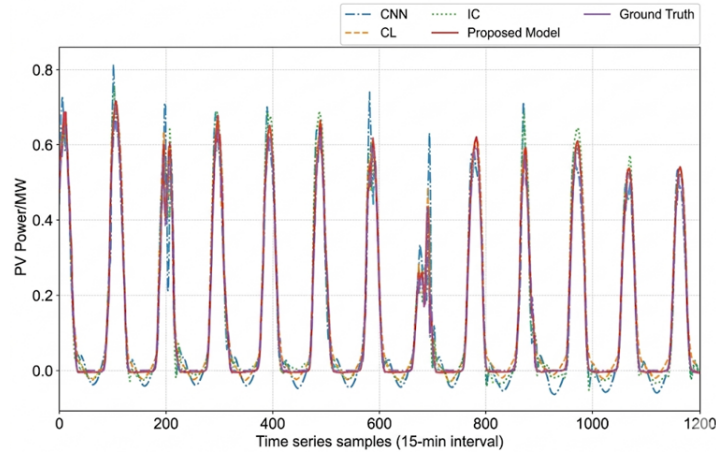
Model	Sunny			Cloudy			Rainy		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Proposed model	0.9984	1.5717	1.2307	0.9918	1.1294	1.4766	0.9952	1.5790	1.1639
CNN	0.8729	13.9303	10.7734	0.6288	9.2307	6.0327	0.7806	10.6284	7.5038
CL	0.9386	9.6781	8.0917	0.7817	4.6759	3.2859	0.8570	4.7042	3.4009
IC	0.9238	10.7862	7.7438	0.8338	8.8999	6.2023	0.8825	2.9983	2.2900

Note: CNN = Convolutional Neural Network; CL = Convolutional Neural Network (CNN)-Long Short-Term Memory (LSTM); IC = Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN)-Convolutional Neural Network (CNN); RMSE = Root Mean Square Error; MAE = Mean Absolute Error.

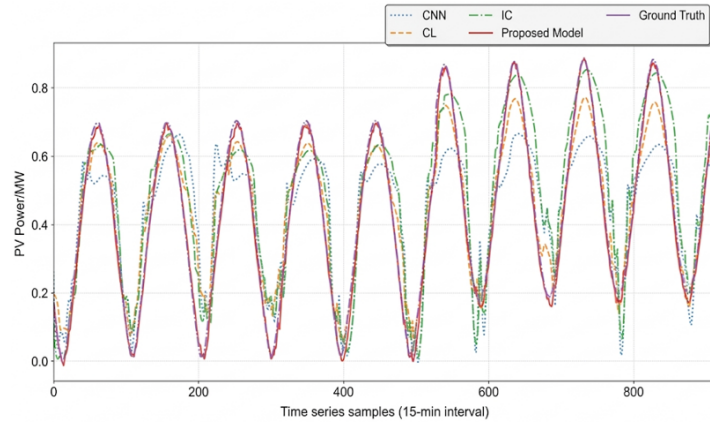
The prediction accuracy of the proposed model is significantly better than that of the traditional CNN architecture. Its R^2 value reaches 0.9984, which is 14.4% higher than 0.8729 of the CNN model. Meanwhile, the RMSE and MAE are greatly reduced from 13.9303 and 10.7734 to 1.5717 and 1.2307, with a reduction rate of 88.7% and 88.6%, respectively. This indicates that the CNN model is difficult to adapt to the stationary characteristics of the PV sequence on sunny days due to the lack of a dynamic parameter adjustment mechanism. In contrast, DOA effectively reduces the interference of high-frequency noise on feature extraction by iteratively optimizing the mode boundary threshold of ICEEMDAN and the attention weight of LSTM-Transformer, enabling the model to accurately capture the subtle fluctuation law of illumination intensity. Under rainy conditions, the R^2 value of the proposed model reaches 0.9952, which is 12.8% higher than that of the IC model, and its RMSE and MAE are 1.5790 and 1.1639, which are 47.4% and 49.2% lower than 2.9983 and 2.2900 of the IC model, respectively. Since the IC model lacks the multi-scale temporal modeling capability of Transformer, its prediction curve has an obvious hysteresis in the area of sudden power changes caused by rainfall. However, the proposed model dynamically strengthens the feature weight of abrupt time steps in the historical sequence through the multi-head attention mechanism, so that the phase difference between the prediction results and the actual power curve is controlled within 5 minutes, which highlights the necessity of dynamic weighting of temporal features. In addition, as shown in Figure 8, the prediction curve of the proposed model under rainy conditions can quickly respond to the sudden drop and rise of power caused by rainfall, maintaining good synchronization with the ground truth. In contrast, the prediction curve of the IC model obviously cannot keep up with the change rhythm of the ground truth in the area of sudden power changes, with a large phase lag, which further demonstrates the superior prediction performance of the proposed model under complex rainy conditions.



(a)



(b)



(c)

Figure 8. Comparison of experimental prediction result curves: (a) sunny weather; (b) cloudy weather; (c) rainy weather

Under cloudy conditions, the MAE of the proposed model is 1.4766, which is 55.1% better than that of the CL model (CNN-LSTM), and its RMSE is also reduced from 4.6759 to 1.1294, with a reduction rate of 75.8%. Since the CL model directly inputs the original irradiation sequence, its LSTM layer is difficult to distinguish the high-frequency cloud disturbance and low-frequency insolation trend mixed in cloudy weather, resulting in periodic oscillation errors in the prediction curve. In contrast, the proposed model decouples the original signal into 7 intrinsic mode components through ICEEMDAN, and then dynamically assigns each component to the LSTM and Transformer sub-modules for special learning via the DOA algorithm, finally realizing the refined modeling of the non-stationary PV power sequence in a cloudy environment. As shown in Figure 8, under cloudy conditions, the

prediction curve of the proposed model can well fit the change trend of the ground truth, and has a more accurate grasp of the power fluctuation caused by cloud occlusion. However, the prediction curve of the CL model shows obvious periodic oscillation, which cannot accurately distinguish cloud disturbance and insolation trend, resulting in a large deviation between the prediction results and the ground truth, which further reflects the advantages of the proposed model under cloudy weather.

6 Conclusion

To improve the accuracy of PV power prediction, aiming at the problems that PV power is significantly affected by sudden weather changes and has strong temporal non-stationarity, this paper proposes a multi-scale PV power prediction model based on DOA-ICEEMDAN and hybrid neural network. Specifically, the main conclusions are as follows:

1) Based on the PCC, the core meteorological factors such as irradiance and temperature are screened, and then combined with the k-means algorithm to construct a dataset for three types of weather scenarios. On this basis, the improved ICEEMDAN decomposition technology is combined with the LSTM-Transformer hybrid neural network to implement collaborative modeling. Through the adaptive adjustment of decomposition levels and the fusion strategy of multi-modal features, the challenge of temporal non-stationarity caused by sudden weather changes is effectively addressed.

2) A dynamic parameter tuning strategy for ICEEMDAN based on DOA is proposed for the first time. This strategy breaks the limitation of the traditional fixed noise standard deviation and decomposition level, enabling the ICEEMDAN algorithm to adapt to different data characteristics and analysis requirements, and realizing more flexible parameter configuration.

3) A dual-channel collaborative framework of LSTM-Transformer is constructed. The LSTM module focuses on the daily cycle trend features and establishes a long-term memory model, while Transformer uses the deformable attention mechanism to accurately capture the minute-level dynamic correlation patterns of sudden weather changes. By fusing and optimizing the output features of the two models, accurate prediction and efficient modeling of short-term fluctuations in PV power are realized, which provides a solid technical foundation for smart grid dispatch and distributed energy management.

Author Contributions

Conceptualization, Q.M.Y.; methodology, D.Y.P.; software, D.Y.P.; validation, W.H.; formal analysis, D.Y.P.; investigation, W.H.; resources, Q.M.Y.; data curation, D.Y.P.; writing—original draft preparation, D.Y.P.; writing—review and editing, W.H.; supervision, W.H.; project administration, Q.M.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] C. Dong, Z. Wang, J. Bai, J. Jiang, B. Wang, and G. Liu, “Review of ultra-short-term forecasting methods for photovoltaic power generation,” *High Voltage Eng.*, vol. 49, no. 7, pp. 2938–2951, 2023.
- [2] E. M. Carlini, F. Del Pizzo, G. M. Giannuzzi, D. Lauria, F. Mottola, and C. Pisani, “Online analysis and prediction of the inertia in power systems with renewable power generation based on a minimum variance harmonic finite impulse response filter,” *Int. J. Electr. Power Energy Syst.*, vol. 131, p. 107042, 2021. <https://doi.org/10.1016/j.ijepes.2021.107042>
- [3] S. Q. Qiu, Z. M. Jian, L. X. Fang, J. W. Qin, J. L. Wan, and P. S. Yuan, “Photovoltaic power forecasting based on variational modal decomposition and integrated learning,” *Smart Power*, vol. 52, no. 3, pp. 32–38, 2024. <https://doi.org/10.3969/j.issn.1673-7598.2024.03.005>
- [4] W. Yuan, W. Pei, Z. Zeng, R. Zhang, C. Z. Teng, and Z. X. Zhao, “Short-term photovoltaic power interval prediction based on mechanism model and XGBoost,” *Adv. Technol. Electr. Eng. Energy*, vol. 44, no. 2, pp. 89–97, 2025. <https://doi.org/10.12067/ATEEE2310034>
- [5] D. Markovics and M. J. Mayer, “Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction,” *Renew. Sustain. Energy Rev.*, vol. 161, p. 112364, 2022. <https://doi.org/10.1016/j.rser.2022.112364>

- [6] M. Guermoui, K. Bouchouicha, N. Bailek, and J. W. Boland, "Forecasting intra-hour variance of photovoltaic power using a new integrated model," *Energy Convers. Manag.*, vol. 245, p. 114569, 2021. <https://doi.org/10.1016/j.enconman.2021.114569>
- [7] A. Sabadus, R. Blaga, S. M. Hategan, D. Calinoiu, E. Paulescu, O. Mares, R. Boata, N. Stefu, N. Stefu, and V. Badescu, "A cross-sectional survey of deterministic PV power forecasting: Progress and limitations in current approaches," *Renew. Energy*, vol. 226, p. 120385, 2024. <https://doi.org/10.1016/j.renene.2024.120385>
- [8] M. Louzazni, H. Mosalam, A. Khouya, and K. Amechnoue, "A non-linear auto-regressive exogenous method to forecast the photovoltaic power output," *Sustain. Energy Technol. Assess.*, vol. 38, p. 100670, 2020. <https://doi.org/10.1016/j.seta.2020.100670>
- [9] S. Ding, R. Li, and Z. Tao, "A novel adaptive discrete grey model with time-varying parameters for long-term photovoltaic power generation forecasting," *Energy Convers. Manag.*, vol. 227, p. 113644, 2021. <https://doi.org/10.1016/j.enconman.2020.113644>
- [10] M. Ding, H. Zhou, H. Xie, M. Wu, K. Z. Liu, Y. Nakanishi, and R. Yokoyama, "A time series model based on hybrid-kernel least-squares support vector machine for short-term wind power forecasting," *ISA Trans.*, vol. 108, pp. 58–68, 2021. <https://doi.org/10.1016/j.isatra.2020.09.002>
- [11] G. Q. Lin, L. L. Li, M. L. Tseng, H. M. Liu, D. D. Yuan, and R. R. Tan, "An improved moth-flame optimization algorithm for support vector machine prediction of photovoltaic power generation," *J. Clean. Prod.*, vol. 253, p. 119966, 2020. <https://doi.org/10.1016/j.jclepro.2020.119966>
- [12] E. Skoplaki and J. A. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations," *Sol. Energy*, vol. 83, no. 5, pp. 614–624, 2009. <https://doi.org/10.1016/j.solener.2008.10.008>
- [13] L. Wang, M. Mao, J. Xie, Z. Liao, H. Zhang, and H. Li, "Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model," *Energy*, vol. 262, p. 125592, 2023. <https://doi.org/10.1016/j.energy.2022.125592>
- [14] D. Mazzeo, M. S. Herdem, N. Matera, M. Bonini, J. Z. Wen, J. Nathwani, and G. Oliveti, "Artificial intelligence application for the performance prediction of a clean energy community," *Energy*, vol. 232, p. 120999, 2021. <https://doi.org/10.1016/j.energy.2021.120999>
- [15] X. Wang 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, p. 131071, 2024. <https://doi.org/10.1016/j.energy.2024.131071>
- [16] Z. Y. Wang, Z. M. Wang, Y. Cai, and X. J. C. H. Tan, "Short-term load forecasting based on long short-term memory network combination algorithm," *Mod. Electr. Power*, vol. 40, no. 2, pp. 201–209, 2023. <https://doi.org/10.19725/j.cnki.1007-2322.2021.0293>
- [17] J. Li, Z. Song, X. Wang, Y. Wang, and Y. Jia, "A novel offshore wind farm typhoon wind speed prediction model based on PSO-Bi-LSTM improved by VMD," *Energy*, vol. 251, p. 123848, 2022. <https://doi.org/10.1016/j.energy.2022.123848>
- [18] C. Zhang, L. Hua, C. Ji, M. S. Nazir, and T. Peng, "An evolutionary robust solar radiation prediction model based on WT-CEEMDAN and IASO-optimized outlier robust extreme learning machine," *Appl. Energy*, vol. 322, p. 119518, 2022. <https://doi.org/10.1016/j.apenergy.2022.119518>
- [19] R. Sheng and X. Zhang, "Photovoltaic power forecasting model based on probabilistic TCN-Transformer," *Integr. Intell. Energy*, vol. 46, no. 11, pp. 10–18, 2024. <https://doi.org/10.3969/j.issn.2097-0706.2024.11.002>
- [20] Y. Lang and Y. Gao, "Dream optimization algorithm (DOA): A novel metaheuristic optimization algorithm inspired by human dreams and its applications to real-world engineering problems," *Comput. Methods Appl. Mech. Eng.*, vol. 436, pp. 117 718–117 718, 2025. <https://doi.org/10.1016/j.cma.2024.117718>
- [21] H. Tang, F. Kang, X. Li, and Y. Sun, "Short-term photovoltaic power prediction model based on feature construction and improved transformer," *Energy*, vol. 320, p. 135213, 2025. <https://doi.org/10.1016/j.energy.2025.135213>
- [22] W. Tang, X. Xie, T. Qian, W. Li, X. Li, X. Jin, X. Wang, and T. Pan, "Energy-carbon coordination of transactive multi-community integrated energy systems under imperfect communication networks," *Energy*, vol. 319, p. 134947, 2025. <https://doi.org/10.1016/j.energy.2025.134947>