PIER C

(Received 18 August 2025, Accepted 26 September 2025, Scheduled 6 October 2025)

Photovoltaic Power Prediction Model Based on K-Shape-NGO-CNN-BiLSTM with Secondary Decomposition

Zhongan Yu, Faneng Wu*, Long Chen, Siqi Zhu, and Junjie Zhang

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

ABSTRACT: With the development of photovoltaic industry, accurate power prediction is critical to grid stability. To address photovoltaic power's high sensitivity to meteorological conditions, nonlinearity, and non-stationarity, this paper develops a prediction model that integrates multi-scale features and intelligent optimization. First, correlation coefficients are used to screen key weather factors, and K-shape clustering is applied to classify operational scenarios into sunny, cloudy, and rainy types. For the power data of each scenario, multi-scale features are extracted via Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), sample entropy secondary clustering, and Variational Mode Decomposition (VMD)-based deep decomposition. After fusing these features with weather factors, the integrated data is input into a Convolutional Neural Network-Bidirectional Long Short-Term Memory Network (CNN-BiLSTM), with hyperparameters optimized using Northern Goshawk Optimization (NGO) algorithm. Verification with actual data sets indicates that this model outperforms traditional counterparts. Specifically, compared with the traditional BiLSTM model, its Mean Absolute Error (MAE) is reduced by 70.8%, 20.7%, and 47.0% under sunny, cloudy, and rainy scenarios, respectively — providing effective support for efficient dispatching and stable operation of photovoltaic power grids.

1. INTRODUCTION

With the global energy structure transitioning towards clean energy, photovoltaic power generation, leveraging its advantages of renewability and non-pollution has been continuously increasing its proportion in the energy system [1, 2]. However, the output of photovoltaic power is susceptible to meteorological factors such as solar irradiance, temperature, and cloud cover, exhibiting significant nonlinearity, non-stationarity, and randomness. This not only poses challenges to the safe and stable operation of power systems but also restricts the efficient absorption of photovoltaic energy [3]. Therefore, constructing high-precision photovoltaic power prediction models holds important practical significance for optimizing power dispatching, reducing the curtailment rate of photovoltaic power, and improving the economy of power grids [4, 5].

In recent years, scholars have carried out extensive research in the field of photovoltaic power prediction. From traditional statistical methods to machine learning algorithms and then to deep learning models, the prediction accuracy has been continuously improved [5–8]. Statistical models use statistical principles to analyze historical data and build prediction models. For example, the autoregressive moving average model is relatively simple to calculate. However, in the face of highly nonlinear, complex, and changeable photovoltaic power data, the prediction accuracy is difficult to meet the actual needs [9, 10]. Prediction methods based on machine learning and deep learning have become research hotspots. In the field of machine learning, [11] studies the photovoltaic power prediction method combining machine learning methods with weather models. In the field of deep learning, [12] uses a model combining convolutional neural network (CNN) and Informer to predict photovoltaic power. Ref. [13] combines bidirectional long short-term memory (BiLSTM) with transformer to form a new model, so as to improve the accuracy of photovoltaic power prediction. Ref. [14] points out that its high accuracy may stem from the non-autoregressive strategy rather than the ability to extract temporal relationships. Ref. [15] proposes DS former, which mines key features of multivariate long sequences through dual sampling. TFEformer in [16] integrates multiple attention mechanisms to optimize photovoltaic power prediction. GinAR+ in [17] addresses the challenge of predicting multivariate sequences with missing values.

To further improve the prediction accuracy, decomposing and clustering input data, as well as optimizing hyperparameters, are also key steps. Ref. [18] has constructed a photovoltaic power prediction model integrating VMD, Informer, and DCC. This model embeds a VMD layer in the encoder part of Informer, and effectively weakens the fluctuation characteristics of the feature sequence by decomposing it. Ref. [19] uses the improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) algorithm to carry out the initial decomposition of the original photovoltaic power data, and then further splits each intrinsic mode function (IMF) into high-frequency, medium-frequency, and low-frequency components. Ref. [20] applies K-means++ clustering method to divide the data set into sunny, cloudy, and rainy days, so as to improve the similarity of training samples. Ref. [21] uses fuzzy c-means (FCM) clustering algorithm to cluster multiple photovoltaic stations into different groups, and then carries out

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

mainstream clustering prediction work. Refs. [22] and [23] respectively adopt improved snake optimization algorithm and Improved Particle Swarm Optimization-Least Squares Support Vector Machine (IPSO-LSSVM) optimization algorithm to optimize hyperparameters of the proposed model.

Despite significant progress in research, existing models still have three shortcomings: (1) The division of weather scenarios is rough. Most models use simple threshold methods to distinguish sunny, cloudy, and rainy weather, failing to capture subtle differences under the same type of weather; (2) The depth of feature extraction is insufficient, and a single decomposition method is difficult to fully characterize multi-time scale characteristics.

To address the above issues, this paper proposes a photovoltaic power prediction model integrating multi-scale feature extraction and intelligent optimization. The model first screens key weather factors through correlation analysis, and then uses K-shape clustering algorithm to divide the photovoltaic power time series into three typical scenarios: sunny, cloudy, and rainy days, so as to realize the hierarchical processing of data. For each type of scenario data, CEEMDAN method is used for initial decomposition, and after secondary clustering combined with sample entropy, the VMD method is used to deeply extract multi-scale features. These features are fused with key weather factors and input into the CNN-BiLSTM model, and NGO algorithm is introduced to optimize hyperparameters to improve prediction performance. It aims to provide a high-precision and high-reliability solution for photovoltaic power prediction under complex meteorological conditions and provide technical support for the sustainable development of the photovoltaic industry.

2. RESEARCH MODEL

2.1. Correlation Analysis

In photovoltaic power prediction, Pearson Correlation Coefficient is a fundamental tool for analyzing the linear relationship between influencing factors and photovoltaic output. Its core function is to quickly identify variables that have a significant linear impact on photovoltaic output, providing a basis for feature selection and input design in prediction models.

The correlation coefficient is a statistical indicator that measures the degree of linear correlation between two continuous variables, with a value range of $[-1, 1]$. The closer the absolute value is to 1, the stronger the linear correlation is. A positive value indicates a positive correlation; a negative value indicates a negative correlation; and a value close to 0 indicates a weak or no linear correlation [13]. The calculation formula for Pearson correlation coefficient r is:

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

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

2.2. K-Shape Clustering Algorithm

K-shape clustering is a clustering algorithm specifically designed for time series data, whose core goal is to group time series with similar shapes, without focusing on differences in their numerical scales or offsets [24]. Compared with traditional clustering algorithms (such as K-means), K-shape is more suitable for handling the dynamic characteristics of time series, such as trends, periodicity, and pattern similarity. The specific steps are as follows:

Step 1: Initialization:

Randomly select K time series as the initial cluster centers.

Step 2: Calculate similarity:

For each time series, calculate its shape distance from each cluster center. For two time series X and Y, first perform standardization on them to eliminate the influence of mean and variance:

$$\begin{cases} \hat{x}_i = \frac{x_i - \overline{X}}{\sigma_X} \\ \hat{y}_i = \frac{y_i - \overline{Y}}{\sigma_Y} \end{cases} \tag{2}$$

where $\overline{X}$ and $\overline{Y}$ are the means of the sequences, respectively, and σ_X and σ_Y are standard deviations of the sequences, respectively. Subsequently, calculate the maximum cross-correlation coefficient of the two standardized sequences:

$$\max\_\text{corr}(X, Y) = \max_{\tau} \left(\frac{1}{n} \sum_{i=1}^{n-\tau} \hat{x}_i \cdot \hat{y}_{i+\tau} \right) \tag{3}$$

where τ is the time lag (used to capture the similarity of temporal offsets between sequences, such as one sequence appearing later than another but with the same shape), and its value range is $[-(n-1), n-1]$. Finally, the shape distance is:

$$d(X, Y) = 1 - \max\_\text{corr}(X, Y) \tag{4}$$

Step 3: Assign clusters:

Assign each time series to the category of the cluster center with the closest distance.

Step 4: Update centers:

For each cluster, recalculate its shape center (i.e., the "average shape" of all time series in the cluster).

Step 5: Repeat iteration:

Repeat Steps 2–4 until the cluster centers no longer change significantly, or the maximum number of iterations is reached.

2.3. CEEMDAN Algorithm

CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) is an improved Empirical Mode Decomposition (EMD) method, mainly used for the adaptive decomposition of nonlinear and non-stationary time series

data [25]. It decomposes the original signal into a series of physically meaningful Intrinsic Mode Functions (IMFs) and a residual component. Its core advantages lie in solving the problems of mode mixing and noise residue existing in traditional EMD and EEMD (Ensemble Empirical Mode Decomposition), thus improving the stability and accuracy of decomposition. The decomposition process of CEEMDAN can be summarized as the following steps:

Step 1: Let the original signal be $x(t)$. On this basis, add a white noise signal $\omega_n(t)$ to obtain the sequence to be decomposed $y_n(t)$.

Step 2: Perform EMD decomposition on $y_n(t)$, take the mean of the decomposition results, and obtain the first-order intrinsic mode function $\mathrm{IMF}_1(t)$ and the residual term.

$$R_1(t) = x(t) - \mathrm{IMF}_1(t) \tag{5}$$

Step 3: Add white noise to the first-order residual term to obtain

$$R_1(t) + \varepsilon_1 E_1[\omega_n(t)] \tag{6}$$

Perform EMD decomposition on it to obtain the second-order intrinsic mode function $\mathrm{IMF}_2(t)$ and its residual term $R_2(t)$;

$$\mathrm{IMF}_2(t) = \frac{1}{k} E_1\{R_1(t) + \varepsilon_1 E_1[\omega_n(t)]\} \tag{7}$$

$$R_2(t) = R_1(t) - \mathrm{IMF}_1(t) \tag{8}$$

Step 4: Add white noise to the $(i-1)$-th residual term, and the i-th IMF component and residual term are obtained as follows:

$$\mathrm{IMF}_i(t) = \frac{1}{k} \sum_{n=1}^{k} E_1\{R_{i-1}(t) + \varepsilon_{i-1} E_{i-1}[\omega_n(t)]\} \tag{9}$$

$$R_i(t) = R_{i-1}(t) - \mathrm{IMF}_i(t) \tag{10}$$

where $E_{i-1}(*)$ is the $(i-1)$-th IMF component after decomposition; $R_{i-1}(t)$ is the $(i-1)$-th residual term; ε_{i-1} is the white noise weight coefficient; $\omega_n(t)$ is the white noise generated in the n-th processing.

Step 5: Repeat Step 4 until the residual term becomes a monotonic function, and the sequence cannot be further decomposed. Finally, all IMF components and the residual term are obtained.

2.4. VMD Algorithm

Variational Mode Decomposition (VMD) is an adaptive signal decomposition method based on the variational principle [26]. Its core lies in decomposing the original signal into multiple modal components with clear frequency centers and limited bandwidths by iteratively solving the variational model.

The mathematical expression for decomposing photovoltaic time series signals is as follows:

$$\begin{cases} \min \left\{ \sum_{k=1}^{K} \| \partial_t \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t} \right) * \mathbf{u}_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \|_2^2 \right\} \\ \text{s.t.} \sum_{k=1}^{K} \mathbf{u}_k = \mathbf{f}(t) \end{cases} \tag{11}$$

where K is the total number of components; $\boldsymbol{u}_k(t)$ is the k-th modal component; ω_k is the central frequency of the k-th mode; ∂_t is the partial derivative; $\delta(t)$ is the Dirac delta function; $*$ is the convolution operator; $\| \cdot \|_2^2$ is the square of the L_2-norm; and $f(t)$ is the original signal.

Using Lagrange multiplier λ and quadratic penalty term α, the above problem is transformed into an unconstrained problem:

$$\begin{aligned} L(\{\mathbf{u}_k\}, \{\boldsymbol{\omega}_k\}, \boldsymbol{\lambda}) = \alpha \sum_{k=1}^{K} \| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mathbf{u}_k(t) \right] \\ \mathrm{e}^{-\mathrm{j}\boldsymbol{\omega}_k t} \|_2^2 + \| \mathbf{f}(t) - \sum_{k=1}^{K} \mathbf{u}_k(t) \|_2^2 \\ + \left\langle \boldsymbol{\lambda}(t), \mathbf{f}(t) - \sum_{k=1}^{K} \mathbf{u}_k(t) \right\rangle \end{aligned} \tag{12}$$

where $\langle \bullet, \bullet \rangle$ denotes the inner product operation.

2.5. Northern Goshawk Optimization

Northern Goshawk Optimization (NGO) is a meta-heuristic optimization algorithm inspired by the predatory behavior of the northern goshawk. Proposed by Dehghani et al. in 2021 [27], it solves various optimization problems by simulating the goshawk's behaviors such as prey identification, attack, chase, and escape. The main introduction is as follows:

(1) Prey identification and attack (exploration phase): The core task of this phase is to screen out target prey and launch an attack quickly. At this stage, the algorithm will conduct a comprehensive global exploration of the entire search range to determine the most suitable attack area, and the corresponding mathematical model is as follows:

$$\begin{aligned} P_i = X_k, i = 1, 2, \ldots, N, \\ k = 1, 2, \ldots, i-1, i+1, \ldots, N \end{aligned} \tag{13}$$

$$x_{i,j}^{\mathrm{new},P_1} = \begin{cases} x_{i,j} + r(p_{i,j} - I x_{i,j}), & F_{p_i} < F_i \\ x_{i,j} + r(x_{i,j} - p_{i,j}), & F_{p_i} \geq F_i \end{cases} \tag{14}$$

In the formula, P_i represents the position of the prey targeted by the i-th northern goshawk; F_{pi} is the objective function value corresponding to this prey position; k is a random natural number within the range of $[1, N]$; $x_{i,j}^{\mathrm{new};p_1}$ denotes the new position of the i-th northern goshawk in the j-th dimension; r takes values within the interval $[0, 1]$; I is a value randomly chosen as 1 or 2. The parameters r and I serve to introduce random behaviors during the search and position update process; F_i^{new,p_1} is the updated objective function value obtained in the exploration phase.

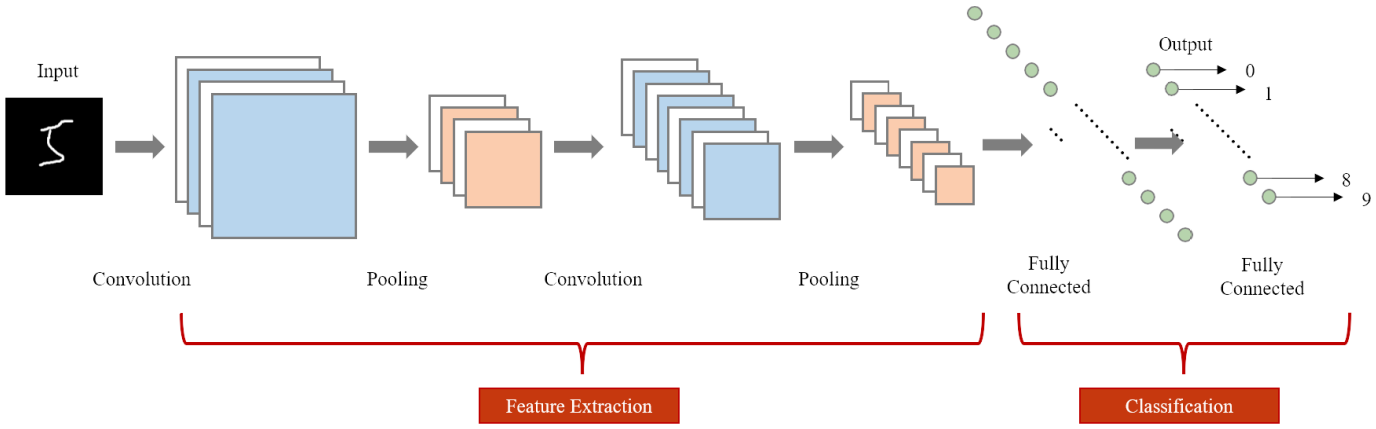

FIGURE 1. Structure diagram of CNN network.

(2) Chase and escape (exploitation phase): This phase mainly simulates the interactive process of the northern goshawk chasing its prey and the prey attempting to escape. Simulating this phase can enhance the algorithm's search efficiency in local spaces. Assuming that the current hunting activity occurs within an area centered at the attack point with a radius of R, the mathematical model before the start of the chase behavior can be expressed as:

$$x_{i,j}^{\text{new},P_2} = x_{i,j} + R(2r-1)x_{i,j} \tag{15}$$

$$R = 0.02\left(1-\frac{t}{T}\right) \tag{16}$$

$$X_i = \begin{cases} X_i^{\text{new},P_2}, & F_i^{\text{new},P_2} < F_i \\ X_i, & F_i^{\text{new},P_2} \geq F_i \end{cases} \tag{17}$$

In the formula, t is the current number of iterations; T is the maximum number of iterations; X_i^{new,P_2} denotes the new position of the i-th northern goshawk in the second phase; and F_i^{new,P_2} is its objective function value.

2.6. Convolutional Neural Network

Convolutional Neural Networks (CNNs) are an important branch of feedforward neural networks and are widely used in the field of deep learning [28]. They draw on the hierarchical information processing mechanism of the biological visual system to excavate the information features hidden in complex data. The main components of CNN include convolutional layers, pooling layers, and fully connected layers, which work together to construct a progressive feature learning mode of "local perception — feature extraction — global mapping".

The convolutional layer uses convolution kernels to perform sliding convolution operations on local regions of the input data, thereby extracting local features of the data in the spatial or temporal dimension. The specific expression for the discrete convolution operation performed by the input data on the convolution kernel is as follows:

$$C(t) = \sum_{i=0}^{k-1}\sum_{j=1}^{D} W(i,j)\cdot X(t+i,j) + b \tag{18}$$

In the formula, X represents the input data, D the feature dimension, k the size of the convolution window, b the bias term, $C(t)$ the output feature value at time step t, and W the convolution kernel. After the data undergoes parallel computation through multi-channel convolution kernels, the CNN can capture fluctuation patterns of the data at different frequencies.

In a CNN network, pooling layer is a key component for feature dimensionality reduction. It employs strategies such as max pooling and average pooling to compress the parameters output by the network, thereby significantly reducing the number and dimensions of parameters and effectively lowering the complexity of computations.

The fully connected layer integrates features through full connections between neurons and ultimately outputs the target results. Considering the prominent advantages of CNNs in processing massive and high-dimensional data, in the task of photovoltaic power prediction, when dealing with multi-source data sets composed of historical meteorological data, power data, etc., CNNs can efficiently extract features from them by virtue of their own structure and use these features as input for the subsequent BiLSTM model. Figure 1 presents the structural composition of the CNN model.

2.7. Bidirectional Long Short-Term Memory

BiLSTM (Bidirectional Long Short-Term Memory) is an extended form of the Long Short-Term Memory network (LSTM) [29]. Its core lies in the integration of forward LSTM and backward LSTM, and it is mainly used to process sequence data. This model can simultaneously utilize the past information (forward) and subsequent information (backward) in the sequence, thereby more comprehensively capturing the internal dependencies of the sequence. Figure 2 presents a schematic diagram of the LSTM structure.

LSTM achieves the storage of long-term time-series information by adding an input gate, a forget gate, and an output gate, and can excellently capture the inherent laws of sequence information. Its formula expression is as follows:

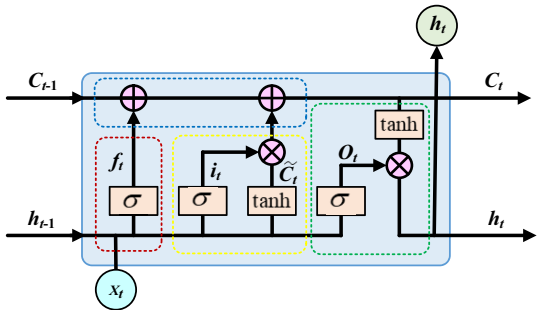

FIGURE 2. LSTM structure.

1. Forget gate: Determines which information in the cell state to discard.

$$f_t = \delta(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{19}$$

In the formula, f_t denotes the forget gate, and δ is the sigmoid activation function, which can map the input to the interval [0, 1]. As a result, part of the input becomes 0 after passing through the activation function, thereby achieving the forgetting effect. W_f represents the weight matrix of the forget gate, h_{t-1} the output at the previous time step, x_t the input at the current time step, and b_f the bias of the forget gate in the current hidden layer.

2. Input gate: Determines which new information is stored in the cell state.

$$\begin{cases} i_t = \delta(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \widetilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \end{cases} \tag{20}$$

In the formula, i_t represents the input gate, which mainly determines the information to be adopted and updated, b_i the parameter of the input gate, tanh an activation function that can limit the output to the range $[-1, 1]$, W_i the weight matrix connecting the input and the neurons in the hidden layer, W_c the weight matrix from the neurons in the hidden layer to the output gate, and b_c the bias of the memory unit in the current hidden layer.

3. Cell state update: Update the long-term memory by combining the results of the forget gate and input gate.

$$C_t = C_{t-1} f_t + \widetilde{C}_t i_t \tag{21}$$

In the formula, C_t and C_{t-1} are the cell state vectors at the current time step and previous time step, respectively.

4. Output gate: Determines the output of the current hidden state (short-term memory).

$$\begin{cases} O_t = \delta(W_o \cdot [h_{t-1}, x_t] + b_o]) \\ h_t = O_t \tanh(C_t) \end{cases} \tag{22}$$

In the formula, O_t represents the output gate, W_o the weight matrix connecting the input and the neurons in the hidden layer, b_o the bias of the output gate in the current hidden layer, and h_t the hidden state at time t.

Since LSTM model can only extract forward time-related features, BiLSTM model has emerged. This model makes up for the defect that the unidirectional LSTM cannot obtain bidirectional feature information during training and has higher accuracy and better performance. To enable the output layer to capture the time-related information of the past and future, BiLSTM is composed of LSTM modules in both forward and backward directions. The outputs of the forward and backward LSTMs are jointly transmitted to the output layer, which not only fills the gap in time-series information but also more comprehensively considers the sequence variation rules of the past and future. Its calculation process is the same as that of LSTM, and Figure 3 shows the structure diagram of BiLSTM.

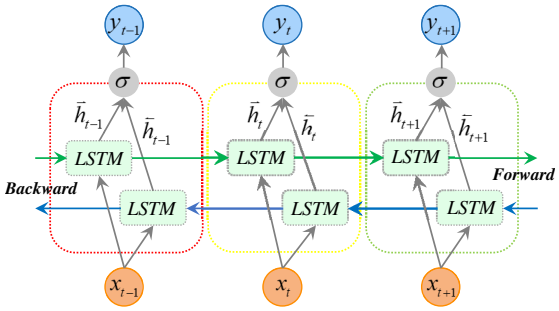

FIGURE 3. BiLSTM structure.

2.8. CNN-BiLSTM Model

CNN-BiLSTM model first uses CNN to perform convolution operations on historical photovoltaic power decomposition data and related data for feature extraction, and then uses BiLSTM neural network to conduct in-depth mining of the data set, capturing the correlation information between data, thereby realizing the prediction of photovoltaic power. The structure of CNN-BiLSTM is shown in Figure 4, where the CNN consists of a convolution layer, a Rectified Linear Unit (ReLU) activation layer, and a max pooling layer; after the Flatten layer completes feature dimensionality reduction, it inputs the processing results of CNN into the BiLSTM network. BiLSTM uses its hidden layer and ReLU activation layer to conduct in-depth

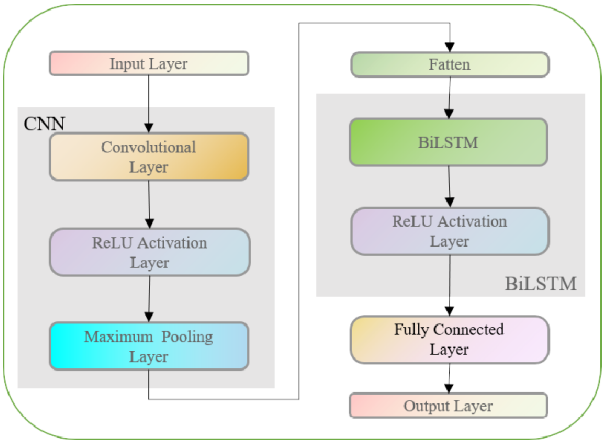

FIGURE 4. CNN-BiLSTM structure.

mining of the data, and then performs full connection operations to output the final prediction results.

3. THE MODEL IN THIS PAPER

Regarding the various weather parameters included in the photovoltaic data set, this study first conducts a systematic correlation analysis. By calculating the correlation coefficients between weather-related features and setting a reasonable threshold, the key weather-influencing factors that exhibit a strong correlation with photovoltaic power are accurately screened out — this process provides high-quality input features for subsequent modeling.

On this basis, K-shape clustering algorithm is used for pattern recognition and clustering of preprocessed photovoltaic power time series data. According to weather characteristics, the time series data are clearly classified into power sequences corresponding to three typical scenarios: sunny days, cloudy days, and rainy days.

For each type of scenario-specific power data, CEEMDAN method is first used for preliminary decomposition to obtain a series of intrinsic mode components of different scales. Then, the sample entropy of each component is calculated to quantify the complexity and irregularity of each component. Based on sample entropy characteristics of each component, secondary clustering is performed to further divide the decomposed components into three types of signals: high-frequency, medium-frequency, and low-frequency — this step realizes the refined extraction of multi-scale features of the power sequence.

Subsequently, following the secondary clustering, VMD method is further applied to the above-classified high-frequency signals for in-depth decomposition — aiming to further explore the dynamic characteristics and potential laws contained in the decomposed signals.

Finally, the multi-scale feature signals obtained through CEEMDAN-VMD secondary decomposition are fused with the key weather factor data screened out earlier, and the fused data is sent as input variables to the CNN-BiLSTM model for training. During model training, NGO algorithm is introduced to conduct global optimization of its hyperparameters, and the optimal combination of hyperparameters is found through iteration — this optimization process constructs a photovoltaic power prediction model with better prediction performance than traditional models. Figure 5 illustrates the comprehensive architecture of the proposed prediction system.

4. CASE ANALYSIS

4.1. Data

The comprehensive data set used in this paper is obtained from Hebei Province, China, which contains detailed atmospheric measurement data and photovoltaic system performance data, specifically involving 7 dimensions of information, namely radiation intensity, air temperature, humidity, wind speed, wind direction, air pressure, and photovoltaic power. Since the photovoltaic power is almost zero at night, all nighttime data are excluded in this study, and the outliers and missing data points in the data are processed, resulting in 6250 valid samples. To screen the characteristics of meteorological data, the study uses Pearson correlation analysis to extract factors that have a strong correlation with photovoltaic power, and the results of the correlation analysis are shown in Figure 6.

Among the factors affecting photovoltaic power output, total radiation intensity, air temperature, wind speed, and wind direction show a positive correlation with photovoltaic power, while air pressure and humidity exhibit a negative correlation. In terms of the degree of association, total radiation intensity has the strongest correlation with photovoltaic power, followed by humidity, and then air temperature. Based on this analysis conclusion, total radiation intensity, humidity, and air temperature are identified as the core factors influencing photovoltaic power output. Therefore, these key variables are incorporated into the prediction model as basic input features.

K-shape clustering algorithm is applied to classify historical photovoltaic data, and the visualization of clustering results is shown in Figure 7. Statistical analysis of meteorological conditions reveals that within the time frame covered by the study, there are 56 sunny days, 27 cloudy days, and 42 rainy days. Specifically, on sunny days, the photovoltaic power generation shows a relatively gentle parabolic fluctuation with changes in solar radiation intensity. On cloudy days, due to the increased thickness of cloud layers and changes in total cloud cover, atmospheric refraction effect is enhanced, and solar radiation is correspondingly reduced, resulting in irregular fluctuations in photovoltaic power. On rainy days, the high relative humidity and increased water vapor hinder the effective reflected radiation from the ground, thereby causing a decrease in photovoltaic power.

4.2. Secondary Decomposition of Signals

For the power data of sunny, cloudy, and rainy days, CEEMDAN method is first used for preliminary decomposition to obtain a series of intrinsic mode functions of different scales. The CEEMDAN results are shown in Figure 8.

Subsequently, the sample entropy value of each component is calculated to quantify the complexity and irregularity of each component. Based on sample entropy features, secondary clustering is conducted to further classify the decomposed components into three types of signals: high-frequency, medium-frequency, and low-frequency. The high, medium, and low-frequency signals are displayed in Figure 9.

After that, the VMD method is continuously applied to the above-classified signals for in-depth decomposition, so as to further explore the dynamic characteristics and potential laws contained in the signals. The finally decomposed signals are shown in Figure 10.

4.3. Model Evaluation Indicators

To effectively evaluate the prediction accuracy and performance of the model, evaluation indicators are usually used for quantitative analysis. In this paper, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square

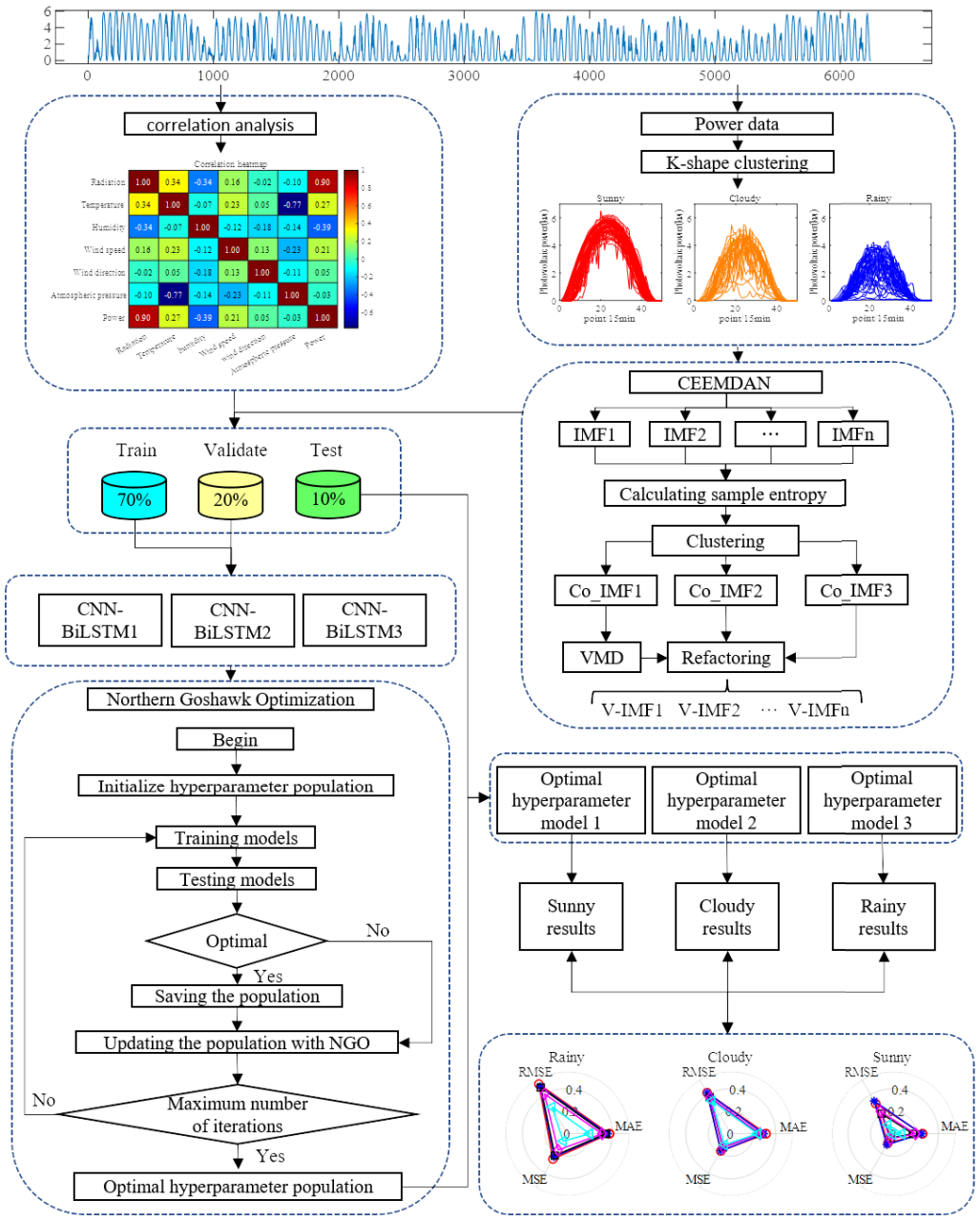

FIGURE 5. Comprehensive framework of the prediction model.

Error (RMSE) are adopted as the evaluation indicators of the model.

(1) Mean Absolute Error

MAE represents the average of the absolute errors between predicted values and actual values, and its calculation formula is:

$$\mathrm{MAE} = \frac{1}{M}\sum_{j=1}^{M}\left|y_{\mathrm{ture},j} - y_{\mathrm{fore},j}\right| \tag{23}$$

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

(2) Mean Squared Error

MSE is the mean of the squares of prediction errors, and its calculation formula is:

$$\mathrm{MSE} = \frac{1}{M}\sum_{j=1}^{M}\left(y_{\mathrm{ture},j} - y_{\mathrm{fore},j}\right)^2 \tag{24}$$

(3) Root Mean Square Error

RMSE is the square root of the mean of the squares of prediction errors, and its calculation formula is:

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(y_{\mathrm{ture},j} - y_{\mathrm{fore},j}\right)^2} \tag{25}$$

TABLE 1. Main parameters of each model.

Model	Weather	Learning rate	Number of neurons	Regularization parameter
BiLSTM, CNN-BiLSTM	Sunny Cloudy Rainy	0.0001	50	1.00e-5
NGO-CNN-BiLSTM	Sunny	0.001381901	21	1.39 e-4
	Cloudy	0.002954543	23	4.21e-4
	Rainy	0.008157245	28	4.57 e-4

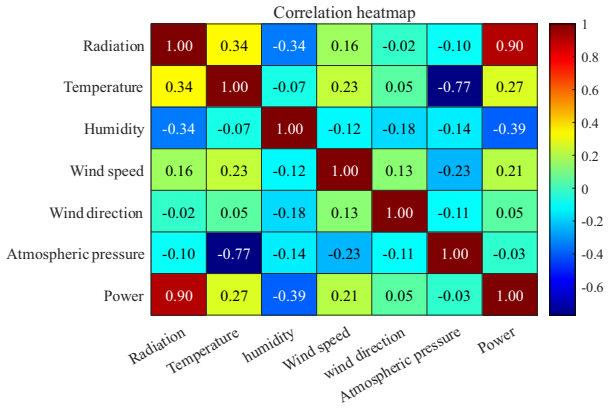

FIGURE 6. Results of correlation analysis.

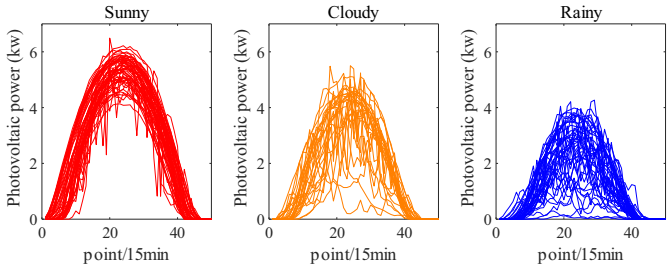

FIGURE 7. K-shape clustering results.

4.4. Comparison of Prediction Results

To verify the effectiveness of the model proposed in this paper, four power prediction models, namely BiLSTM, CNN-BiLSTM, NGO-CNN-BiLSTM, and Transformer, are specifically selected for comparative experiments. Transformer adopts default parameters, while the specific configuration parameters of the other models are detailed in Table 1, and their prediction performance is presented in Table 2.

To further demonstrate the effectiveness and accuracy of the proposed model under different weather conditions, Figures 10 to 14 present photovoltaic power curves and evaluation index values of the five models in various weather scenarios.

In terms of basic prediction capability, BiLSTM, as a pure time-series model, can capture long-term temporal patterns such as diurnal cycles and seasonal trends of photovoltaic power. However, it has limited ability to capture short-term power fluctuations (e.g., sudden rises and falls caused by cloud cover), leading to larger prediction errors, especially in meteorologically unstable scenarios like cloudy or rainy days. This is because it solely relies on the gating mechanism of LSTM units to handle temporal correlations, lacking targeted extraction of local abrupt features.

TABLE 2. Results of different models.

Weather	Model	MAE	RMSE	MSE
Sunny	BiLSTM	0.2585	0.3094	0.0957
	CNN-BiLSTM	0.2490	0.3365	0.1132
	NGO-CNN-BiLSTM	0.1762	0.2058	0.0424
	Transformer	0.0924	0.1132	0.0128
	Proposed Model	**0.0755**	**0.0922**	**0.0085**
Cloudy	BiLSTM	0.3129	0.4208	0.1771
	CNN-BiLSTM	0.2922	0.4189	0.1755
	NGO-CNN-BiLSTM	0.2496	0.3745	0.1402
	Transformer	0.2549	0.3546	0.1257
	Proposed Model	**0.2387**	**0.3357**	**0.1126**
Rainy	BiLSTM	0.3728	0.5084	0.2585
	CNN-BiLSTM	0.3566	0.4850	0.2353
	NGO-CNN-BiLSTM	0.3389	0.4728	0.2235
	Transformer	0.2342	0.3072	0.0944
	Proposed Model	**0.1975**	**0.2739**	**0.0750**

CNN-BiLSTM compensates for the shortcomings of BiLSTM by introducing a CNN module. The convolutional layers can effectively extract local features of photovoltaic power (such as power jumps caused by sudden changes in radiation intensity and short-term fluctuation patterns), which are then processed by BiLSTM to capture temporal dependencies. As a result, its overall prediction accuracy is superior to that of BiLSTM. In scenarios with relatively smooth power curves (e.g., sunny days), the gap between the two models is relatively small; however, in complex scenarios such as cloudy-to-sunny transitions or short-term showers, CNN-BiLSTM shows more obvious advantages, better fitting nonlinear fluctuations of power.

NGO-CNN-BiLSTM builds on CNN-BiLSTM by optimizing hyperparameters through NGO algorithm, addressing the issue of "insufficient parameter adaptability" that may arise from manual parameter tuning. In photovoltaic prediction, this enables the model to more accurately match power characteristics under different meteorological conditions. For example, it can automatically adjust to a larger convolution stride to capture global trends for smooth curves on sunny days and optimize a small convolution kernel to capture detailed changes for

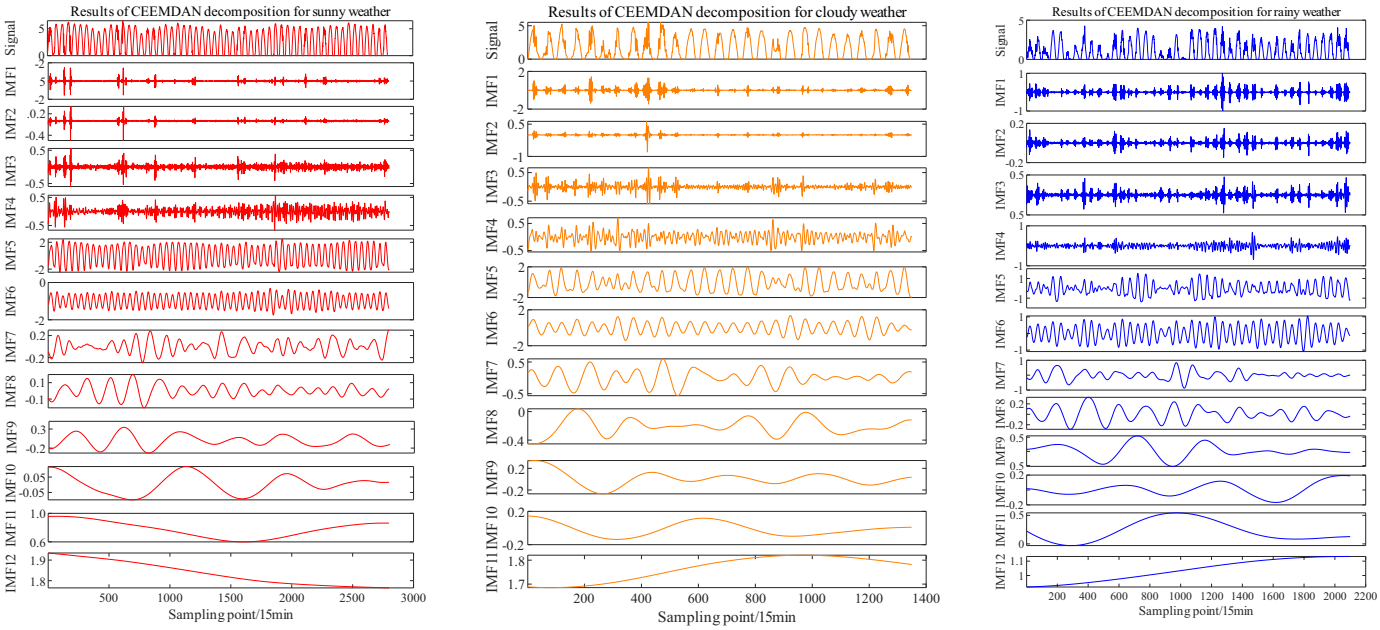

FIGURE 8. CEEMDAN results.

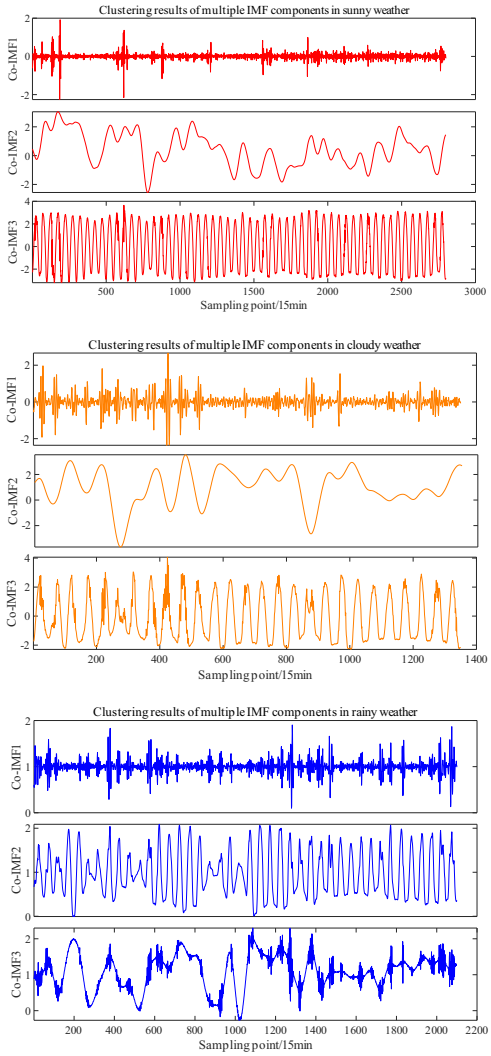

FIGURE 9. High-frequency, medium-frequency, and low-frequency signals.

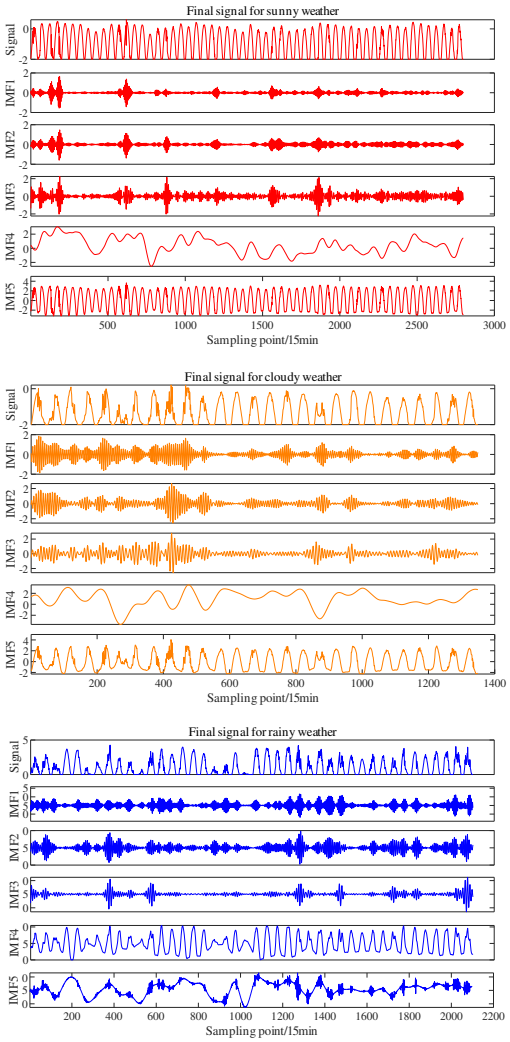

FIGURE 10. Final signal.

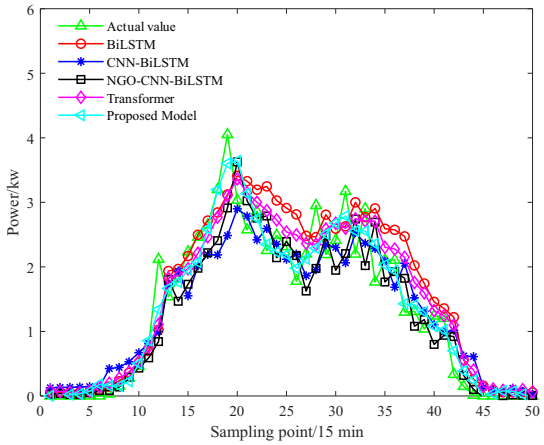

FIGURE 11. Results of rainy day forecast.

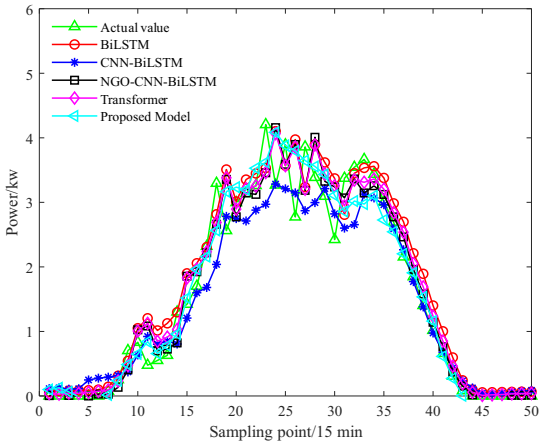

FIGURE 12. Results of cloudy day forecast.

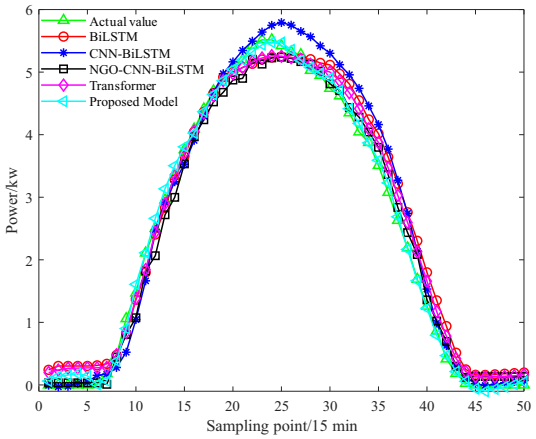

FIGURE 13. Results of sunny day forecast.

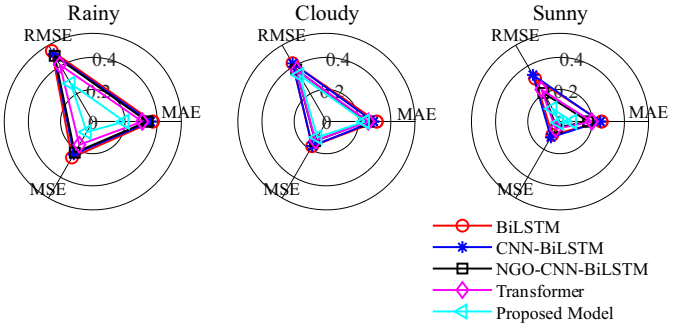

FIGURE 14. Errors in the three types of weather forecasts.

random fluctuations on rainy days. Thus, its prediction error is further reduced compared to CNN-BiLSTM, and it performs more stably, especially on data sets with mixed scenarios.

The core improvement of C-V-CNN-BiLSTM lies in the data preprocessing stage: through the secondary decomposition of CEEMDAN and VMD, the original photovoltaic power sequence is decomposed into high-frequency (short-term noise, such as instantaneous cloud cover), medium-frequency (intra-day fluctuations, such as the morning power rise phase), and low-frequency (long-term trends, such as seasonal changes) signals. This decomposition allows the model to learn features of each frequency band in a targeted manner, reducing interference from the mutual mixing of different frequency components during model learning. In practical prediction, its adaptability to complex meteorological scenarios is significantly enhanced.

The proposed model in this paper integrates the above advantages: secondary decomposition achieves refined decomposition of power signals; the NGO algorithm optimizes the model's adaptability to features of different frequency bands; and the combination of CNN and BiLSTM balances local feature extraction and temporal correlation modeling. In photovoltaic prediction, this model performs optimally across all scenarios — it can accurately capture the parabolic trend on sunny days, effectively fit random fluctuations on cloudy days, and achieve smaller prediction errors for power troughs on rainy days.

Relying on the self-attention mechanism to capture long-term temporal dependencies, the Transformer exhibits outstanding performance in handling long-range correlations such as cross-day cycles. However, its multi-head attention structure leads to a large number of parameters and low computational efficiency. Additionally, it has weak capability in capturing short-term sudden change features of photovoltaic power (e.g., abrupt power drops caused by cloud shading). In complex meteorological scenarios, it is prone to significant errors due to the dilution of local features by global information.

The method proposed in this study specifically addresses the aforementioned issues through a cascaded mechanism of "secondary decomposition (CEEMDAN-VMD) + NGO optimization + CNN-BiLSTM": The decomposition module first separates the high-, medium-, and low-frequency components of the power sequence, enabling the model to focus on learning features of each frequency band. The CNN-BiLSTM combination balances the extraction of local fluctuations and the modeling of temporal correlations, adapting to the nonlinear characteristics of photovoltaic power.

The hyperparameters optimized by the NGO algorithm further balance the model's accuracy and efficiency.

In summary, from single time-series modeling to the integration of "decomposition-optimization-hybrid network", the models' adaptability to the nonlinear and non-stationary

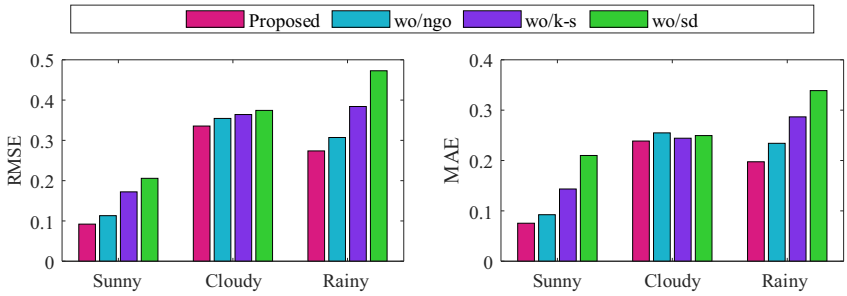

FIGURE 15. Results of ablation experiments.

TABLE 3. Optimal hyperparameter.

Weather		Learning rate	Number of neurons	Regularization parameter
Sunny	IMF1	0.00568849	14	6.10e-5
	IMF2	0.002861757	15	1.36e-4
	IMF3	0.006139866	14	1.97e-5
	IMF4	0.00383468	20	1.21e-4
	IMF5	0.008901037	27	3.13e-4
Cloudy	IMF1	0.00919515	23	1.02e-6
	IMF2	0.01	19	2.52e-4
	IMF3	0.002160373	18	3.93e-4
	IMF4	0.005628152	34	6.52e-4
	IMF5	0.008567446	42	1.00e-3
Rainy	IMF1	0.004613326	18	4.56e-5
	IMF2	0.004902249	25	1.01e-6
	IMF3	0.001190368	42	6.91e-5
	IMF4	0.000783476	40	1.39e-4
	IMF5	0.004089647	23	4.03 e-4

characteristics of photovoltaic power has gradually improved. Among them, the proposed model in this paper exhibits the most significant advantages in prediction accuracy and stability under complex meteorological conditions.

4.5. Ablation Experiments

This study incorporates four key components: secondary decomposition, K-shape clustering, NGO (Northern Goshawk Optimization), and CNN-BiLSTM. To verify that these components contribute to improving the model accuracy, ablation experiments were conducted from the following three aspects:

(1) wo/ngo: The NGO component is removed.

(2) wo/k-s: The K-shape clustering component is removed.

(3) wo/sd: The secondary decomposition component is removed.

Figure 15 presents the results of the ablation experiments. The experiments demonstrate that K-shape scenario clustering, CEEMDAN-VMD secondary decomposition, and NGO hyperparameter optimization are all key modules for improving the model's prediction accuracy, and there is a synergistic effect among the three. Specifically, K-shape clustering provides a clear scenario foundation for feature extraction; secondary decomposition supplies high-quality multi-scale features for model input; and NGO optimization ensures that the model is fully adapted to the features of each scenario. These results further verify the rationality of the model architecture design proposed in this study.

4.6. Analysis of Model Hyperparameters and Efficiency

The model in this paper mainly considers the optimal configuration of three types of key hyperparameters: learning rate, number of neurons, and regularization parameters.

A too small learning rate will lead to slow model convergence, failing to capture the dynamic changes of photovoltaic power in a timely manner, while a too large one will cause oscillations in the training process, making it difficult to achieve stable convergence. An insufficient number of neurons will result in the model's inability to fully learn the temporal characteristics of power, easily leading to underfitting; an excessive number, on the other hand, will increase redundant computations and may cause overfitting. A too small regularization parameter makes the model prone to overfitting in complex weather

scenarios, while a too large one will weaken the model's fitting ability.

Northern Goshawk Optimization (NGO) algorithm, by simulating "exploration-exploitation" mechanism, can efficiently locate the optimal configuration of the three within a reasonable parameter search space, avoiding problems such as unbalanced learning rates, inappropriate number of neurons, and deviations in regularization parameters. The results of the optimal hyperparameter configuration are shown in Table 3.

When training on a data set with only 4,375 samples, this study needs to perform secondary decomposition on the photovoltaic power signals and train the decomposed components separately — resulting in a training workload that is 5 times that of other comparative models. Consequently, the total single-round training time increases to 420.3 seconds. However, thanks to the efficient parameter search capability of NGO optimization, the total convergence time is prevented from increasing by 5 times. By contrast, CNN-BiLSTM model only requires 132 seconds for training, and standalone BiLSTM model takes merely 88 seconds. Although the proposed model in this study demands more training time, it demonstrates significantly superior performance in terms of prediction accuracy.

5. CONCLUSION

The characteristic of photovoltaic power generation lies in its inherent volatility and uncertainty, which pose significant challenges to accurate forecasting. To tackle these challenges, this study proposes a novel prediction model based on the K-shape-NGO-CNN-BiLSTM framework with secondary decomposition. The main objective is to utilize advanced technologies to handle the nonlinearity and time dependence of photovoltaic data, so as to improve the accuracy and reliability of photovoltaic power output prediction under different weather conditions.

1. K-shape clustering accurately classifies typical scenarios such as sunny days, cloudy days, and rainy days, effectively stripping off the cross-interference of photovoltaic power characteristics under different weather conditions. It provides a clear data foundation for subsequent scenario-specific modeling and significantly enhances the model's adaptability to complex meteorological conditions.

2. The secondary decomposition combining CEEMDAN and VMD can decompose the photovoltaic power sequence into signals of different frequencies, realizing the refined analysis of multi-scale features such as instantaneous fluctuations and intra-day trends. This reduces the interference of signal aliasing on model learning and especially improves the prediction accuracy in high-noise scenarios.

3. The optimization of CNN-BiLSTM hyperparameters by the NGO algorithm can adaptively match the characteristics of different scenarios. Combining the local feature extraction capability of CNN with the time-series modeling capability of BiLSTM forms an efficient feature learning mechanism, which takes both prediction accuracy and model stability into account in photovoltaic power prediction.

ACKNOWLEDGEMENT

This work was supported by the Postgraduate Innovation Special Fund of Jiangxi Province (No. YC2023-S617).

REFERENCES

[1] Yi, T., L. Tong, M. Qiu, and J. Liu, "Analysis of driving factors of photovoltaic power generation efficiency: A case study in China," *Energies*, Vol. 12, No. 3, 355, 2019.

[2] Xu, X., C. Guan, and J. Jin, "Valuing the carbon assets of distributed photovoltaic generation in China," *Energy Policy*, Vol. 121, 374–382, 2018.

[3] Başaran, K., F. Bozyiğit, P. Siano, P. Y. Taşer, and D. Kılınç, "Systematic literature review of photovoltaic output power forecasting," *IET Renewable Power Generation*, Vol. 14, No. 19, 3961–3973, 2020.

[4] Sobri, S., S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Conversion and Management*, Vol. 156, 459–497, 2018.

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

[6] Tina, G. M., C. Ventura, S. Ferlito, and S. D. Vito, "A state-of-art-review on machine-learning based methods for PV," *Applied Sciences*, Vol. 11, No. 16, 7550, 2021.

[7] Yu, J., X. Li, L. Yang, L. Li, Z. Huang, K. Shen, X. Yang, X. Yang, Z. Xu, D. Zhang, and S. Du, "Deep learning models for PV power forecasting: Review," *Energies*, Vol. 17, No. 16, 3973, 2024.

[8] Das, U. K., K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. V. Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, Vol. 81, 912–928, 2018.

[9] Massidda, L. and M. Marrocu, "Use of multilinear adaptive regression splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany," *Solar Energy*, Vol. 146, 141–149, 2017.

[10] Li, Y., Y. He, Y. Su, and L. Shu, "Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines," *Applied Energy*, Vol. 180, 392–401, 2016.

[11] Buonanno, A., G. Caputo, I. Balog, S. Fabozzi, G. Adinolfi, F. Pascarella, G. Leanza, G. Graditi, and M. Valenti, "Machine learning and weather model combination for PV production forecasting," *Energies*, Vol. 17, No. 9, 2203, 2024.

[12] Wu, Z., F. Pan, D. Li, H. He, T. Zhang, and S. Yang, "Prediction of photovoltaic power by the informer model based on convolutional neural network," *Sustainability*, Vol. 14, No. 20, 13022, 2022.

[13] 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.

[14] Zeng, A., M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, No. 9, 11 121–11 128, Washington, USA, Jun. 2023.

[15] Yu, C., F. Wang, Z. Shao, T. Sun, L. Wu, and Y. Xu, "Dsformer: A double sampling transformer for multivariate time series long-term prediction," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 3062–3072, Birmingham, United Kingdom, Oct. 2023.

[16] Yu, C., J. Qiao, C. Chen, C. Yu, and X. Mi, "TFEformer: A new temporal frequency ensemble transformer for day-ahead photovoltaic power prediction," *Journal of Cleaner Production*, Vol. 448, 141690, 2024.

[17] Yu, C., F. Wang, Z. Shao, T. Qian, Z. Zhang, W. Wei, Z. An, Q. Wang, and Y. Xu, "GinAR+: A robust end-to-end framework for multivariate time series forecasting with missing values," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 37, No. 8, 4635–4648, 2025.

[18] Wu, Y., X. Pan, and J. Yang, "VMD-Informer-DCC for photovoltaic power prediction," *IEICE Transactions on Communications*, Vol. E107-B, No. 7, 487–494, 2024.

[19] Xu, W., Z. Wang, W. Wang, J. Zhao, M. Wang, and Q. Wang, "Short-term photovoltaic output prediction based on decomposition and reconstruction and XGBoost under two base learners," *Energies*, Vol. 17, No. 4, 906, 2024.

[20] Li, M., W. Wang, Y. He, and Q. Wang, "Deep learning model for short-term photovoltaic power forecasting based on variational mode decomposition and similar day clustering," *Computers and Electrical Engineering*, Vol. 115, 109116, 2024.

[21] Peng, D., Y. Liu, D. Wang, L. Luo, H. Zhao, and B. Qu, "Short-term PV-wind forecasting of large-scale regional site clusters based on FCM clustering and hybrid Inception-ResNet embedded with Informer," *Energy Conversion and Management*, Vol. 320, 118992, 2024.

[22] Wang, Y., Y. Yao, Q. Zou, K. Zhao, and Y. Hao, "Forecasting a short-term photovoltaic power model based on improved snake optimization, convolutional neural network, and bidirectional long short-term memory network," *Sensors*, Vol. 24, No. 12, 3897, 2024.

[23] Jiang, J., S. Hu, L. Xu, and T. Wang, "Short-term PV power prediction based on VMD-CNN-IPSO-LSSVM hybrid model," *International Journal of Low-Carbon Technologies*, Vol. 19, 1160–1167, 2024.

[24] Paparrizos, J. and L. Gravano, "K-shape: Efficient and accurate clustering of time series," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1855–1870, Melbourne Victoria, Australia, May 2015.

[25] Torres, M. E., M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4144–4147, Prague, Czech Republic, May 2011.

[26] Dragomiretskiy, K. and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, Vol. 62, No. 3, 531–544, 2014.

[27] Dehghani, M., Š. Hubálovský, and P. Trojovský, "Northern goshawk optimization: A new swarm-based algorithm for solving optimization problems," *IEEE Access*, Vol. 9, 162 059–162 080, 2021.

[28] 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.

[29] Yang, Z., L. Li, Z. Rao, W. Meng, and S. Wan, "A short-term PV resource assessment method with parallel DenseNet201 and BiLSTM under multiple data features," *Energy Reports*, Vol. 11, 2841–2852, 2024.