energies

MDPI

Article

A Short-Term Photovoltaic Power Forecasting Method Based on Multiscale Decomposition and Adaptive Optimization

Zhengzhong Gao [1], Baokang Peng [1,*] and Fengqi Zang [2]

[1] College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China; skdgzz@163.com
[2] Jining Coal Mine Safety Production Testing and Monitoring Center, Jining 272000, China; 18300328169@163.com
* Correspondence: skdpbk@sdust.edu.cn

Abstract

Existing studies still have limitations in characterizing the complex temporal patterns of photovoltaic time series. In particular, current forecasting models often struggle to capture local abrupt fluctuations and nonlinear relationships among variables, while hyperparameter optimization remains challenging. To overcome these limitations, this paper proposes a CNN-iTransformer short-term photovoltaic power forecasting method based on multiscale decomposition and adaptive optimization. First, the original photovoltaic power time series is decomposed into trend, seasonal, and residual components using seasonal-trend decomposition based on Loess (STL). The residual component is then further decomposed through variational mode decomposition (VMD) to fully extract the latent multiscale temporal information embedded in the sequence. Second, a convolutional neural network (CNN) is employed to extract local fluctuation features, while the iTransformer is utilized to model the nonlinear relationships and long-term temporal dependencies between multiple meteorological variables and photovoltaic power. Finally, the Phototropic Growth Algorithm (PGA) is introduced to optimize key hyperparameters of the forecasting model, thereby improving its generalization capability and forecasting accuracy under multi-seasonal scenarios. Simulation experiments are conducted using real-world data from a photovoltaic power station in Alice Springs, Australia. The experimental results show that, under the same PGA optimization conditions, the proposed model reduces the root mean square error (RMSE) and mean absolute error (MAE) by approximately 19.12% and 17.43%, respectively, compared with the baseline iTransformer. Meanwhile, its performance is better than that of several mainstream deep learning models, which verifies the generalization ability and forecasting accuracy of the proposed method under complex seasonal variations.

Keywords: short-term photovoltaic power forecasting; multiscale decomposition; convolutional neural network; iTransformer; phototropic growth algorithm; time-series data processing

Check for updates

Academic Editor: Michael Negnevitsky

Received: 7 August 2026
Revised: 25 August 2026
Accepted: 27 August 2026
Published: 29 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Traditional fossil energy sources, such as coal, oil, and natural gas, are not only limited in reserves but also release large amounts of greenhouse gases during use, thereby aggravating global climate change [1]. Therefore, clean and renewable energy sources, such as wind energy, solar energy, and geothermal energy, have attracted widespread

attention [2]. Among them, solar energy has become one of the most promising renewable energy sources because it is clean, environmentally friendly, abundant, and widely distributed [3,4]. Photovoltaic (PV) power generation technology can efficiently convert solar energy into electricity to meet the increasing energy demand. With technological advances and cost reductions, PV power generation has been rapidly deployed worldwide.

However, unlike conventional power generation, the output power of PV systems highly depends on meteorological factors, such as solar irradiance, ambient temperature, and cloud cover. As a result, PV power series exhibit significant randomness, intermittency, and non-stationarity [5]. Large fluctuations in PV power over a short period of time significantly increase the difficulty of power balance and frequency regulation in the power grid, raise system operating costs, and may even threaten the safe and stable operation of the power grid. Therefore, accurate PV power forecasting is of great importance for improving grid dispatch flexibility, optimizing energy storage allocation, and supporting electricity market trading [6].

According to the forecasting horizon, PV power forecasting is generally classified into ultra-short-term (0–4 h), short-term (1–72 h), and medium- to long-term (days to years) forecasting [7–9]. Among them, short-term forecasting plays a key role in day-ahead dispatch, energy storage management, and electricity market trading and is one of the core technologies for achieving the efficient grid integration of renewable energy.

Existing short-term PV power forecasting methods can be mainly divided into three categories: statistical methods, machine learning methods, and deep learning methods. A summary of these methods is presented in Table 1.

Table 1. Comparison and analysis of existing photovoltaic power forecasting methods.

Method	Representative References	Advantages	Limitations
Statistical methods	[10–12]	Simple modeling and high computational efficiency	Limited ability to describe the nonlinear and non-stationary characteristics of PV power, resulting in poor adaptability to complex scenarios
Machine learning methods	[13–16]	Strong nonlinear fitting capability	Limited ability to model long-term dependencies
RNN-based methods	[17–20]	Capable of learning temporal dependencies and achieving good forecasting accuracy	Long-term dependency modeling is limited by the recurrent architecture
Transformer-based methods	[21–28]	Capable of learning global relationships among multiple variables and modeling long-term temporal dependencies	Limited ability to capture local continuous variations and short-term fluctuations, and the model performance is sensitive to hyperparameters
Conventional decomposition methods	[29–33]	Reduce the non-stationarity of the series and improve the quality of model inputs	Mainly rely on empirical or frequency-based decomposition and cannot explicitly separate trend, periodic, and random components

Statistical methods were widely used in the early stage. They are mainly used to build mathematical models based on historical data for forecasting. Representative methods include linear regression [10], the autoregressive integrated moving average (ARIMA) model [11], and grey theory [12]. These methods have the advantages of simple modeling

and low computational cost. They are suitable for PV power stations with sufficient data and regular patterns. However, their ability to model nonlinear, non-stationary, and high-dimensional time series is limited. Therefore, they cannot effectively capture the complex dynamic characteristics of PV power series.

To improve nonlinear modeling capability, machine learning methods have been widely introduced into the field of PV forecasting. Representative models include random forest (RF) [13], support vector machine (SVM) [14], extreme learning machine (ELM) [15], and extreme gradient boosting (XGBoost) [16]. These methods perform better than traditional statistical models in modeling nonlinear relationships. However, they have limited ability to model long-term temporal dependencies, and their generalization ability is insufficient under sudden weather changes or multi-time-scale scenarios.

In recent years, with the development of artificial intelligence, deep learning methods have gradually become a research hotspot in PV power forecasting because of their strong feature extraction and sequence modeling capabilities. Existing deep learning models are mainly based on two architectures: recurrent neural networks (RNNs) and Transformers.

RNN-based models, such as long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional long short-term memory (BiLSTM), have long been the mainstream option because of their ability to model temporal dependencies. For example, Ref. [17] proposed a CNN-BiLSTM model that combines convolutional layers for local feature extraction with BiLSTM for bidirectional temporal dependency modeling, thereby improving forecasting accuracy. Ref. [18] adopted a CNN-GRU architecture by combining spatial feature extraction and temporal modeling. The CNN-LSTM models proposed in Refs. [19,20] also achieved good performance. However, these models are still limited by their recurrent structures when processing long sequences. They suffer from the gradient vanishing and exploding problems and cannot fully model long-term dependencies.

To overcome the limitations of RNNs, the Transformer has gradually become an important framework for PV power forecasting because of its self-attention mechanism, which shows advantages in long-sequence modeling. Various improved models, such as Informer, FEDformer, PatchTST, and TimeXer, have significantly enhanced the modeling capability for long time series. For example, Ref. [21] combined the ProbSparse attention mechanism of Informer to further improve the model's adaptability to non-stationary sequences. Ref. [22] proposed the RAD-FEDformer model based on FEDformer. A multilayer perceptron was introduced to enhance deep feature extraction, and a relevance augmentation module (RAM) was designed to incorporate the geographical information of PV power stations into the decoder, thereby strengthening spatial correlation modeling and improving the forecasting accuracy of multi-site PV power. Ref. [23] proposed a parallel TCN-PatchTST forecasting framework, in which TCN and PatchTST were used as two parallel branches to learn the local dynamic features and global temporal dependencies of PV power series, respectively. The forecasting performance was further improved through dual-branch feature fusion. Ref. [24] proposed a cross-attention forecasting model based on TimeXer by integrating exogenous variables. The cross-attention mechanism was used to enable information interaction between PV power and meteorological variables, thereby further enhancing the model's ability to model multivariate time-series features.

Among the improved Transformer models, iTransformer [25] has demonstrated excellent performance in multivariate time-series forecasting because of its unique variable embedding mechanism, and has been successfully applied in many time-series forecasting tasks. Ref. [26] proposed the MKAN-iTransformer model. It enhances the learning ability of complex dynamic features through multiscale temporal modeling and a variable-level attention mechanism, thereby further improving ultra-short-term PV power forecasting performance. Ref. [27] proposed the FE-iTransformer hybrid model for traffic flow forecast-

ing. A feature enhancement module was used to extract and enhance the spatiotemporal features of traffic flow, and iTransformer was employed to learn long-term dependencies, thereby achieving high-accuracy traffic flow forecasting. Ref. [28] combined time-frequency collaborative analysis with iTransformer to fully exploit the multiscale characteristics of stock price series, thereby improving the forecasting performance of complex financial time series.

However, although these models achieve excellent forecasting accuracy, their performance highly depends on the selection of hyperparameters and is sensitive to hyperparameter settings. Therefore, introducing intelligent optimization algorithms has become an important approach to improving the training efficiency and performance stability of Transformer-based models. Automated hyperparameter tuning or adaptive optimization strategies can reduce the cost of manual intervention and identify better model configurations in a larger search space.

At the data processing level of PV power forecasting, most studies focus on data decomposition. The principle of forecasting is to understand the inherent periodicity and fluctuations in the data. PV power data exhibit significant periodic fluctuations and complex non-stationarity. Therefore, effective decomposition methods are required to accurately capture these key characteristics. Ref. [29] proposed an innovative PV power forecasting method by combining empirical mode decomposition (EMD) with a backpropagation neural network (BPNN). Ref. [30] applied improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) to decompose PV power and meteorological variables, thereby improving the stationarity and noise resistance of the input data. The decomposed data were then predicted using a TCN-BiLSTM-MHA model. Ref. [31] used VMD to preprocess the PV power series. By decomposing the series into different frequency components, the complexity and non-stationarity of the series were reduced, thereby improving the model's ability to learn the variation characteristics of PV power.

The above methods all adopt a single decomposition strategy. They reduce the complexity of the series and alleviate the influence of random fluctuations to a certain extent. However, they mainly model features at a single scale and are still insufficient in terms of exploring the hierarchical relationships among features at different time scales in PV power series.

To further explore the potential features of decomposed sequences and achieve finer multiscale feature extraction, some studies have adopted a secondary decomposition strategy. Refs. [32,33] employed a two-stage ICEEMDAN-VMD framework to further decompose highly complex modes. This strategy further improved the separation of features at different time scales and provided more stationary input sequences for subsequent forecasting models. However, these secondary decomposition methods are still mainly based on empirical decomposition and frequency decomposition. They mainly focus on high-frequency fluctuations and random disturbances, but lack explicit modeling of the long-term trends and periodic patterns in PV power series. As a result, the trend, periodic, and random components remain coupled to a certain extent, leading to insufficient physical interpretability of the decomposition results.

In summary, although existing PV power forecasting methods have achieved good forecasting performance, they still have the following limitations:

(1) Existing data decomposition methods cannot effectively consider different timescale characteristics of PV power series, including long-term trends, periodic patterns, and high-frequency fluctuations. The trend, periodic, and random components are still coupled, resulting in insufficient structural representation capability and physical interpretability of the input data.

(2) Although existing Transformer-based models can learn the global relationships between meteorological variables and PV power through the variable embedding mechanism, they mainly rely on the global self-attention mechanism. Their ability to extract continuous variation features and short-term fluctuation information within local time windows is limited. Therefore, they cannot effectively integrate local dynamic features with multivariate correlation information.

(3) The forecasting performance of Transformer models is highly sensitive to key hyperparameters, such as the learning rate, embedding dimension, and the number of attention heads. Most existing studies rely on empirical settings or traditional optimization methods, making it difficult to obtain the optimal hyperparameter combination. As a result, the forecasting accuracy and generalization ability of the model are affected.

To address the above problems and improve the accuracy and robustness of short-term PV power forecasting, this paper proposes a CNN-iTransformer short-term PV power forecasting method based on multiscale decomposition and adaptive optimization. The main contributions are summarized as follows:

(1) To address the coupling of features at different time scales and the insufficient representation of trend and periodic information in PV power series, a two-stage STL-VMD multiscale decomposition method is proposed. First, STL is used to extract the trend and periodic features of the PV power series. Then, VMD is applied to the residual component for secondary decomposition to capture non-stationary fluctuation features, thereby achieving hierarchical and fine-grained characterization of the complex temporal features of PV power.

(2) To address the insufficient extraction of local dynamic features and the inadequate exploration of multivariate relationships in short-term PV power forecasting, a CNN-iTransformer hybrid forecasting model is proposed. CNN is used to extract local fluctuation features and short-term abrupt change information. iTransformer employs the variable embedding mechanism and the multi-head self-attention mechanism to learn the nonlinear relationships and long-term temporal dependencies between multiple meteorological variables and PV power, thereby effectively combining local feature extraction with global relationship modeling.

(3) To address the dependence of the iTransformer model on empirical hyperparameter settings and its unstable forecasting performance, the PGA is introduced. PGA is used to perform adaptive search and optimization of the key hyperparameters of iTransformer, thereby improving the training stability and forecasting accuracy of the model.

2. Methods

2.1. Overall Forecasting Framework of the Model

As shown in Figure 1, the workflow of photovoltaic power forecasting using the proposed model is described. It mainly includes five stages: data preprocessing, dataset splitting, data decomposition, hyperparameter optimization, and model forecasting and result evaluation. The specific process is as follows.

Step 1: Data preprocessing and feature construction.

First, the original data are processed through outlier treatment, missing value filling, normalization, and resampling. Then, correlation analysis is performed to evaluate the correlations between different meteorological variables and photovoltaic power. Meteorological factors with strong correlations are selected as the input features of the model.

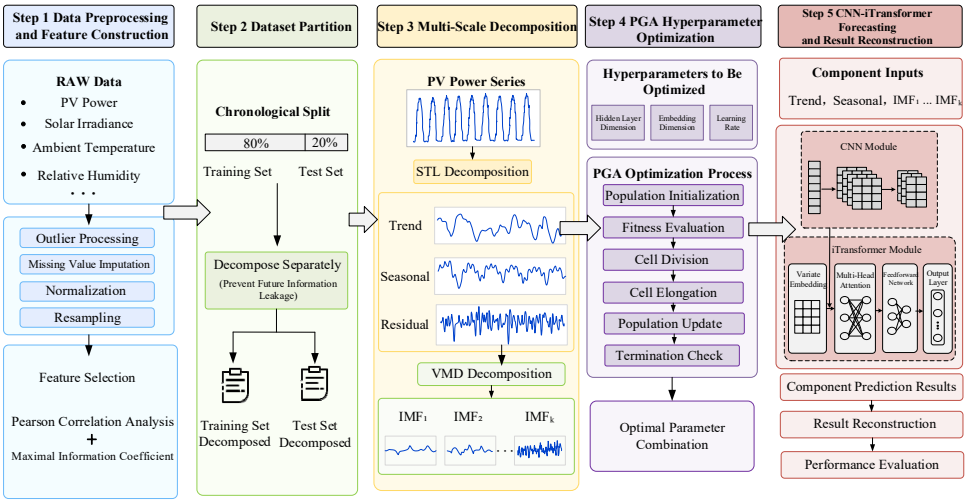


Figure 1. Forecasting framework of the model.

Step 2: Dataset splitting.

The preprocessed data are divided into the training set and the test set in chronological order. The photovoltaic power sequences in the training set and the test set are then decomposed independently.

Step 3: Multiscale decomposition.

First, the STL method is used to decompose the photovoltaic power sequence into the trend component, seasonal component, and residual component. This extracts the long-term trend and periodic variation features of photovoltaic power.

Then, VMD is further applied to the residual component to address the remaining high-frequency fluctuations and random noise. Several residual modal components are obtained. This process captures the non-stationary fluctuations caused by cloud occlusion and sudden weather changes in more detail.

Step 4: PGA hyperparameter optimization.

The hidden layer dimension, embedding dimension, and learning rate of the iTransformer model are set as the three core parameters to be optimized. PGA is used for automatic optimization. The optimal hyperparameter combination is obtained through population initialization, fitness evaluation, cell division, cell elongation, and iterative updating.

Step 5: CNN-iTransformer model forecasting and evaluation.

The trend component, seasonal component, and residual modal components decomposed by VMD are input into the CNN-iTransformer model for forecasting. The CNN module extracts local fluctuations and short-term abrupt variation features. iTransformer models the correlations and long-term temporal dependencies between meteorological variables and photovoltaic power through variable embedding and attention encoding. Finally, the forecasting results of all components are superimposed to obtain the final photovoltaic power forecasting value. RMSE, MAE, and R^2 are used for performance evaluation.

2.2. STL Decomposition

STL decomposition [34] is a time-series decomposition method based on locally weighted regression (Loess). It performs local smoothing on the original sequence. Then, it decomposes the time series into three parts through an iterative process: trend component, seasonal component, and residual component. In photovoltaic time series, the trend component mainly reflects the overall variation in photovoltaic power over time. The seasonal component mainly describes the regular fluctuation characteristics of photovoltaic power within a fixed time period. The residual component represents random fluctuations in photovoltaic sequences after removing trend and periodic components. It usually contains

high-frequency fluctuations caused by cloud occlusion, sudden weather changes, and other factors. STL decomposition can effectively reduce the complexity of photovoltaic sequences. It is suitable for processing photovoltaic power sequences. Its basic formula is as follows:

$$Y_t = T_t + S_t + R_t \tag{1}$$

where Y_t represents the original photovoltaic power time series; T_t represents the trend component; S_t represents the seasonal component; and R_t represents the residual component.

The original photovoltaic time series is decomposed using STL. The specific steps are shown in Figure 2. In the figure, $T_t^{(m+1)}$ and $S_t^{(m+1)}$ indicate the trend component and the seasonal component after the m-th iteration, respectively. The decomposition process consists of an inner loop and an outer loop. After one inner loop is completed, the system performs convergence judgment. If the convergence condition is satisfied, $T_t^{(m+1)}$ and $S_t^{(m+1)}$ are output, and the residual component $R_t^{(m+1)}$. If the convergence condition is not satisfied, the above calculation process continues. After the inner loop satisfies the convergence condition, the process enters the outer loop. The outer loop mainly adjusts the robust weighting coefficient ρ_t. After the above iterative process, the final STL decomposition results are obtained, including the trend component T_t, the seasonal component S_t, and the residual component R_t.

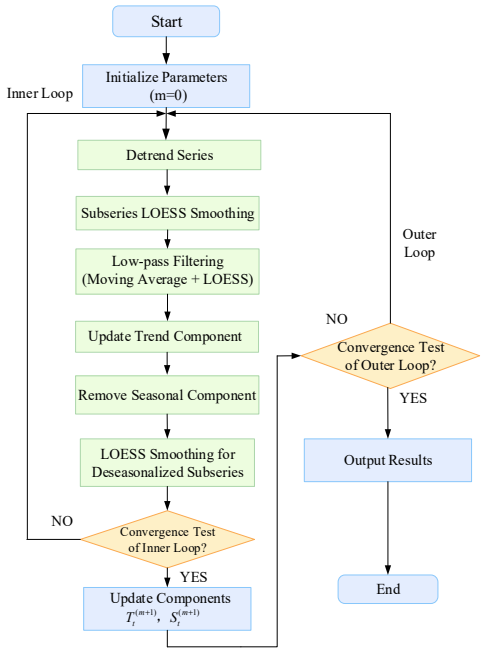


Figure 2. Flowchart of STL decomposition.

2.3. VMD

In this paper, STL decomposition is first used to decompose the original photovoltaic power sequence into trend, seasonal, and residual components. For high-frequency fluctuations in the residual component, VMD [31] is further used for secondary decomposition. Through multiscale modal decomposition of the residual component, random noise inter-

ference can be effectively reduced. This provides more stable data input for subsequent photovoltaic power forecasting. The calculation formula of VMD is as follows:

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \left\{ \sum\limits_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \dfrac{\mathrm{j}}{\pi t} \right) \cdot u_k(t) \right] \cdot \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum\limits_{k=1}^{K} u_k(t) = y(t) \end{cases} \tag{2}$$

where $\delta(t)$ represents the impulse function; $u_k(t)$ represents the modal component function; $\{u_k\}, \{\omega_k\}$ represents the set of all modal components and the set of center frequencies of all modal components, respectively; ω_k represents the center frequency of the k-th modal component; K represents the number of modal decompositions; and $y(t)$ represents the original sequence.

The penalty factor α and the Lagrange operator λ are introduced to process Equation (2). The augmented Lagrangian function is obtained as follows:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t} \right) u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \tag{3}$$

Finally, the alternating direction method of multipliers is adopted. Each modal function u_k and its center frequency ω_k are iteratively updated until K decomposition results are obtained.

2.4. CNN-iTransformer Forecasting Model

Photovoltaic power is jointly affected by multiple meteorological factors, such as solar irradiance, temperature, and cloud cover. It shows obvious nonlinear, random, and local fluctuation characteristics. Traditional Transformer usually treats different time steps as tokens and learns the dependencies between different temporal positions through self-attention along the time dimension. Therefore, it has a strong ability to model long-term dependencies. However, for photovoltaic power forecasting tasks involving multiple meteorological variables, attention calculation only along the time dimension makes it difficult to directly capture the complex cross-variable relationships between different meteorological variables and photovoltaic power.

Different from the traditional Transformer, iTransformer inverts the representation of the input sequence. It independently embeds the sequence of each variable over the entire historical time window as a variable token and performs multi-head self-attention along the variable dimension. Through this variable-level modeling method, the model can directly learn the nonlinear relationships between meteorological variables, such as irradiance, temperature and humidity, and photovoltaic power. At the same time, it preserves the temporal variation information of each variable within the historical window. Since photovoltaic power variation is jointly driven by multiple meteorological factors, the variable-level attention mechanism of iTransformer is more suitable for exploring the coupling relationships between multiple meteorological variables and photovoltaic power than the traditional Transformer, which mainly focuses on dependencies along the time dimension. Therefore, iTransformer is more suitable for the multivariate photovoltaic power forecasting task in this study. However, in short-term photovoltaic power forecasting, iTransformer mainly relies on the global attention mechanism for feature modeling. Its ability to capture continuous variation features within local time windows is relatively limited. When photovoltaic power fluctuates, global attention may not fully extract the local temporal features between adjacent time steps. This can lead to insufficient representation of detailed local information and thus affect the forecasting accuracy of the model.

To address the above problem, this paper introduces a CNN module into the iTransformer model. The CNN first performs local convolution operations on the high-dimensional features after variable embedding. It fully extracts the continuous variation features and local dynamic patterns between adjacent time steps. This process enhances the local representation capability of the embedded features. Then, the enhanced features are fed into the iTransformer encoder for global dependency modeling. This architecture effectively combines local temporal feature extraction with global variable correlation learning. It provides a more comprehensive characterization of the variation patterns of short-term photovoltaic power series. The overall architecture is shown in Figure 3.

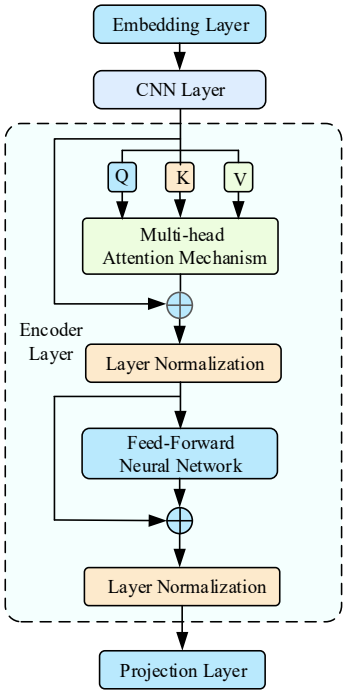


Figure 3. Structure of the CNN-iTransformer model.

(1) Embedding Layer

The input embedding module of iTransformer embeds the entire time series of historical photovoltaic power and each meteorological variable, such as irradiance, temperature, and humidity, as an independent token. This realizes the transformation from time-point correlation to variable correlation. It also preserves the temporal patterns of each variable and provides input features for the subsequent attention mechanism. Therefore, the model can better capture the nonlinear relationships between meteorological variables and photovoltaic power. The calculation formula is as follows:

$$h_n^0 = \text{Embedding}(X) \tag{4}$$

where X represents the time-series input matrix of photovoltaic output and meteorological features; h_n^0 represents the time-series embedded tokens of photovoltaic output and meteorological features; and Embedding represents the embedding operation.

(2) CNN layer

As shown in Figure 4, CNN consists of three parts: the convolutional layer, the pooling layer, and the fully connected layer.

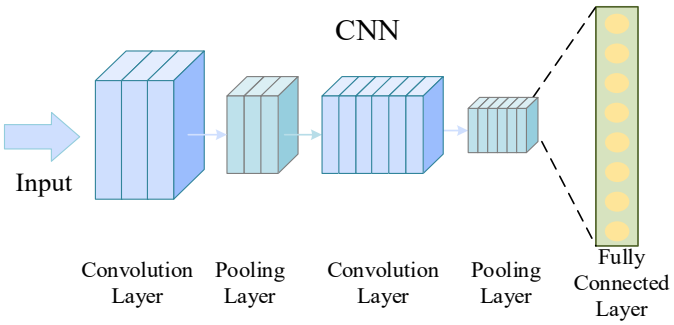


Figure 4. Structure of the CNN model.

The convolutional layer performs convolution operations on the input data through convolution kernels to extract local fluctuation features and potential nonlinear features. It is the core component for feature learning in CNNs. The convolution operation of the convolutional layer is calculated as follows:

$$x_j^l = \sigma\left(\sum_{i \in M_j} x_i^{l-1} \times W_{ij}^l + b_j^l\right) \tag{5}$$

where x_j^l represents the input of the j-th convolution kernel in the l-th layer, and it is also the output of the $(l-1)$-th layer; W_{ij}^l represents the weight matrix of the convolution kernel; and b_j^l represents the bias term.

The pooling layer uses operations such as max pooling to reduce the dimension of the convolved features. It retains key feature information and reduces noise interference. It also lowers the risk of overfitting and improves the generalization ability of the model.

The fully connected layer integrates and maps the features after convolution and pooling. It outputs the photovoltaic power forecasting results. The forward propagation output formula of the fully connected layer is as follows:

$$x_j^{l+1} = \sum_{i=1}^{n} W_{ij}^l x_j^l + b_j^l \tag{6}$$

where x_j^{l+1} represents the value of the j-th output neuron in the $(l+1)$-th layer; W_{ij}^l represent the weight matrix; x_j^l represents the input of the j-th neuron in the l-th layer; and b_j^l represents the bias term.

(3) Encoder Layer

The encoder module consists of two parts: the multi-head self-attention mechanism and the feed-forward neural network. First, temporal feature extraction and interaction modeling are performed on the independently embedded meteorological variables. Then, the attention distribution is calculated. This process dynamically captures the nonlinear relationships among different meteorological features, such as irradiance and temperature. The calculation formula is as follows:

$$Q = W^{\mathrm{Q}} X \tag{7}$$

$$K = W^{\mathrm{K}} X \tag{8}$$

$$V = W^{\mathrm{v}} X \tag{9}$$

$$T_{\mathrm{A}}(Q, K, V) = \mathrm{softmax}(QK^{\mathrm{T}}/\sqrt{d_{\mathrm{k}}})V \tag{10}$$

$$T_{\mathrm{M}}(Q, K, V) = \mathrm{concat}(T_{\mathrm{headl}}, \cdots, T_{\mathrm{head}_i})W^o \tag{11}$$

$$T_{\text{head}_i} = T_{\text{A}}(Q_i, K_i, V_i) \tag{12}$$

where Q represents the query matrix; K represents the key matrix; V represents the value matrix; W^{Q}, W^{K} and W^{v} represent the weight matrices of the query, key, and value, respectively; T_A represents the attention distribution value; T_M represents the multi-head attention distribution value; W^o represents the output weight matrix; T_{headi} represents the output of the i-th attention head; d_k represents the dimension of the key; Q_i, K_i and V_i represent the query matrix, key matrix, and value matrix corresponding to the i-th attention head, respectively; concat represents the concatenation function; and softmax represents the activation function.

Subsequently, the feed-forward neural network performs an independent nonlinear transformation on the feature representation weighted by the attention mechanism. After variable interaction, each meteorological variable token passes through the fully connected network and the nonlinear activation function. This further expands the feature dimensional space and improves the local representation capability of the model. The calculation formula is as follows:

$$T_{\text{FFN}}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{13}$$

$$X' = T_{\text{L}}(X + T_{\text{M(Q,K,V)}}) \tag{14}$$

$$X_{\text{out}} = T_{\text{L}}(X' + T_{\text{FFN}}(X)) \tag{15}$$

where T_{FFN} represents the linear processing of the feed-forward neural network; X represents the input of meteorological features; X' represents the normalized input of meteorological features; T_{L} represents the normalization operation; X_{out} represents the output matrix; b_1 and b_2 represent the bias terms; and W_1 and W_2 represent the weight matrices.

(4) Projection Layer

The projection layer is located at the output end of the encoder. This module maps the high-dimensional features processed by the multi-head attention mechanism and the feed-forward neural network to the target dimension. Through linear transformation, the features extracted by the encoder are converted into the final forecasted photovoltaic power value.

2.5. PGA Optimization Algorithm

The PGA [35] is a swarm intelligence optimization algorithm inspired by the phototropic growth pattern of plants.

2.5.1. Algorithm Flow and Core Mechanism

The overall algorithm construction process includes population initialization, the cell division stage, the cell elongation stage, population updating, and boundary checking.

(1) Population Initialization

First, an initial cell population is randomly generated within the given search space:

$$X_i = L + r(U - L), \quad i = 1, 2, \cdots, N \tag{16}$$

where X_i represents the i-th candidate solution; L and U represent the lower and upper bounds of the search space, respectively; r represents a random vector within $[0, 1]$; and N represents the population size. Then, according to the fitness value, the population is divided into cells in the illuminated region X_L and cells in the shaded region X_S.

(2) Cell Division Stage

The cell division stage simulates the mitotic division process of plant cells. Each parent cell generates new offspring cells through division. Their positions are then updated by the

mutation operator and the auxin redistribution mechanism. For cells in the illuminated region, a new candidate solution is first generated using the mutation operator:

$$X_{Lnew1}(t) = X_{rand} + \alpha\beta r_2 \left| X_{rand} - X_L^i(t) \right| + \alpha\beta r_3 \left| X_{Lbest}^i(t) - X_L^i(t) \right| \quad (17)$$

where X_{rand} represents the random cell; X_{Lbest} represents the current optimal cell in the illuminated region; β represents the random mutation factor; r_2 and r_3 represent random numbers within $[-1, 1]$; and α represents the growth restriction factor. Its formula is as follows:

$$\alpha = e^{-\frac{t}{T}} \quad (18)$$

where T represents the maximum number of iterations and t represents the current number of iterations. As the number of iterations increases, α gradually decreases; the algorithm then shifts from global exploration in the early stage to local exploitation in the later stage. Subsequently, the cell position is further updated through the auxin redistribution mechanism:

$$X_{Lnew2}(t) = X_L^i(t) + \alpha r_4 \left| X_{Lbest}^i(t) - X_L^i(t) \right| \quad (19)$$

This mechanism guides the cells in the illuminated region to move toward the optimal solution. It enhances the local search ability of the algorithm.

For cells in the shaded region, a randomly selected cell from the illuminated region is used to guide them to move toward a better region:

$$X_{Snew}(t) = X_{Lrand} + \alpha\beta r_5 \left| X_{Lbest} - X_S^i(t) \right| \quad (20)$$

where X_{Lrand} represents a randomly selected cell from the illuminated region. This mechanism can improve the global optimization ability of the algorithm and prevent it from falling into a local optimum.

(3) Cell Elongation Stage

The cell elongation stage simulates the bending growth process of plants toward the light source after light stimulation. In this stage, the cell position is further adjusted through the curvature factor and the cell neighborhood mechanism. This enhances the local exploitation ability of the algorithm.

First, the curvature is calculated:

$$curvature = \beta \left(\alpha - \frac{Mean_fitness(X_L)}{Best_fitness(X)} \right) \quad (21)$$

where $Mean_fitness(X_L)$ represents the average fitness of cells in the illuminated region and $Best_fitness(X)$ represents the current global optimal fitness.

Subsequently, the curvature factor is calculated:

$$FOC = r_7 \times curvature \cdot (X_i(t) - X_{Best}(t)) \quad (22)$$

where X_{Best} represents the optimal solution and r_7 represents a random number within $[-1, 1]$. This process enables cells to gradually bend and move toward the optimal solution.

Finally, the cell update formula is obtained as follows:

$$X_{new}^i(t) = X_i(t) + FOC + Cell_vicinity \quad (23)$$

Through the combined effect of the curvature factor and neighborhood information, PGA can maintain population diversity. It can also improve local search accuracy and convergence performance.

(4) Population Updating and Boundary Checking

After each iteration, the fitness values of the updated cell population are recalculated. Individuals with better fitness values are retained for the next generation, and boundary constraints are applied to ensure that all candidate solutions remain within the valid search space. Then, the current global optimal solution is updated. If the maximum number of iterations or the termination condition is reached, the optimal hyperparameter combination is output. Otherwise, the algorithm returns to the cell division stage and continues optimization.

2.5.2. Hyperparameter Optimization Process for the Photovoltaic Power Forecasting Model

In this paper, PGA is used to adaptively optimize the hidden layer dimension, embedding dimension, and learning rate of the iTransformer model. The automatic parameter tuning process based on the PGA is shown in Figure 5. The specific steps are as follows:

1. Set the hyperparameters of the iTransformer model to be optimized and their search ranges.
2. PGA randomly generates multiple groups of hyperparameter combinations. Each group corresponds to a "cell".
3. Each group of hyperparameters is substituted into the overall model. The model is trained on the training set. Then, the forecasting error is calculated.
4. Mean Squared Error (MSE) is used as the fitness evaluation criterion. A smaller error indicates a better hyperparameter combination.
5. PGA divides the hyperparameter combinations into the illuminated region and the shaded region according to their fitness values. Then, the hyperparameter combinations are continuously updated through the cell division and cell elongation mechanisms.
6. Repeat the iterative process until the maximum number of iterations is reached or the fitness no longer improves significantly.
7. Finally, the optimal hyperparameter combination is output.

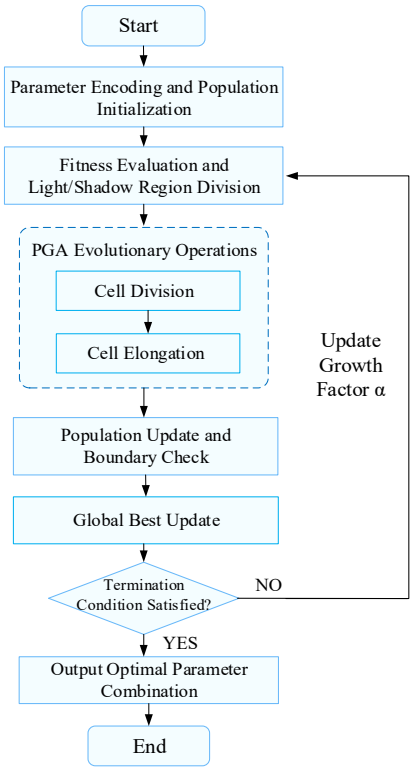


Figure 5. Flowchart of the PGA optimization process.

2.6. Correlation Analysis

Linear and nonlinear relationships usually exist simultaneously between PV power and meteorological variables. The Pearson correlation coefficient (PCC) can accurately measure the linear correlation between variables, while the maximal information coefficient (MIC) can effectively explore complex nonlinear relationships. The two methods are complementary to each other. This paper combines PCC and MIC to perform correlation analysis between meteorological variables and PV power, providing a theoretical basis for subsequent feature selection.

2.6.1. PCC Analysis

PCC is used to quantitatively describe the linear correlation between two variables. Its calculation formula is as follows:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{24}$$

where r_{xy} represents the Pearson correlation coefficient between variable x and variable y; x_i and y_i represent the values of the two variables corresponding to the i-th sample, respectively; $\overline{x}$ and $\overline{y}$ represent the mean values of variable x and variable y, respectively; and n represents the number of samples. The value range of correlation coefficient r_{xy} is $[-1, 1]$. A larger absolute value indicates a stronger linear correlation between two variables.

2.6.2. MIC Analysis

MIC is a statistical method used to measure the correlation between variables. Based on information theory, it can effectively identify both linear and complex nonlinear relationships, including quadratic relationships, periodic relationships, and piecewise functions. Therefore, it has been widely used in nonlinear correlation analysis.

Mutual information (MI) is used to measure the statistical dependence between variables X and Y. It is defined as follows:

$$I(X;Y) = \sum_{x \in X}\sum_{y \in Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{25}$$

where $p(x,y)$ represents the joint probability distribution of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distributions.

MIC partitions the two-dimensional data points (x_i, y_i) into an $m \times n$ grid and calculates the corresponding mutual information $I(X;Y)$ for each grid partition. To reduce the influence of the grid partition on the results, MIC selects the maximum mutual information among all possible grid partitions and then performs normalization. The final formula is given as follows:

$$\mathrm{MIC}(X,Y) = \max_{m,n \le B}\left(\frac{I_m(X;Y)}{\min(\log m, \log n)}\right) \tag{26}$$

where m and n represent the numbers of grid partitions in the horizontal and vertical directions, respectively. $B = n^{0.6}$ is the upper limit of the grid number. The MIC value ranges from 0 to 1. A larger value indicates a stronger correlation between variables, whereas a value closer to 0 indicates a weaker correlation.

3. Experimental Analysis

3.1. Dataset Description

To verify the prediction performance and generalization ability of the proposed short-term photovoltaic power forecasting method under different scenarios, three photovoltaic datasets from different regions of Australia and China are selected for the experiments. These datasets have certain differences in geographical location, climatic conditions, and data time span. Therefore, they can be used to comprehensively evaluate the adaptability of the model under different photovoltaic operating environments. The power series and the division of the training and test sets for each dataset are shown in Figure 6. The key statistical information is listed in Table 2.

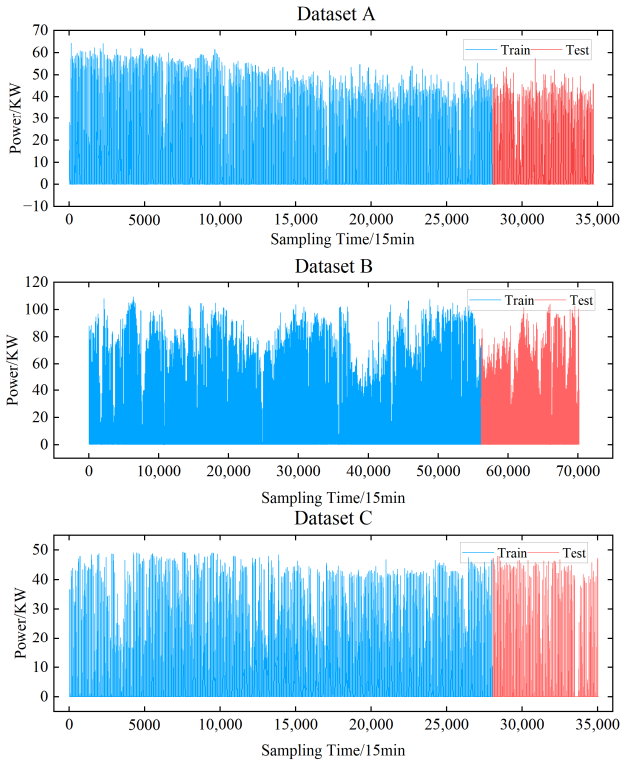


Figure 6. PV power series and training/test set division of different datasets.

Table 2. Statistical information of PV output power for different datasets.

Dataset	Resolution	Samples	Maximum	Mean	Variance
A	15 min	35,040	64.29	15.75	388.36
B	15 min	70,176	109.36	19.57	780.61
C	15 min	35,040	49.31	10.67	222.57

The main experimental data in this study are obtained from the Site1B photovoltaic station of the Desert Knowledge Australia Solar Centre (DKASC) in Alice Springs, Australia (Dataset A). The operating data from 00:00 on 1 January 2017 to 23:45 on 31 December 2017 are selected, with a sampling interval of 15 min. In addition to photovoltaic output power, the dataset contains nine meteorological variables: global horizontal irradiance, diffuse horizontal irradiance, radiation global tilted, radiation diffuse tilted, relative humidity, ambient temperature, rainfall, wind speed, and wind direction. This dataset is used as the main experimental dataset for correlation analysis, different decomposition experiments, optimization algorithm effectiveness analysis, ablation experiments, and comparison experiments among different forecasting models.

Generalization dataset selection was as follows: (1) We selected the public dataset from the China State Grid Renewable Energy Power Forecasting Competition (Dataset B). Data from 00:00 on 1 January 2019 to 23:45 on 31 December 2020 are selected, with a temporal resolution of 15 min. In addition to photovoltaic output power, the dataset contains five meteorological variables: total solar irradiance, direct normal irradiance, global horizontal irradiance, air temperature, and atmosphere. The dataset covers a relatively long time span and can be used to evaluate the forecasting stability of the model under cross-year operating scenarios. (2) We also selected the photovoltaic power station dataset from Xinjiang, China (Dataset C). The full-year operating data from 00:00 on 1 January 2019 to 23:45 on 31 December 2019 are selected, with a sampling interval of 15 min. The dataset contains eight variables: actual power generation, module temperature, temperature, air pressure, humidity, total radiation, direct radiation, and diffuse radiation. These variables can comprehensively characterize the operating status of photovoltaic modules and external meteorological conditions. Xinjiang has abundant solar energy resources. At the same time, it is affected by day–night temperature differences, seasonal variations, and local weather fluctuations. Therefore, the photovoltaic power series shows obvious periodic and non-stationary characteristics. This dataset can be used to evaluate the generalization ability of the proposed method under different climatic conditions and regional characteristics. The above two generalization datasets are only used for comparison experiments between the proposed model and different baseline models.

To improve data quality and reduce the impact of abnormal data on the forecasting results, the three datasets are first preprocessed. The preprocessing mainly includes abnormal data correction and missing value filling to ensure the reliability of subsequent model training and forecasting analysis.

3.2. Simulation Conditions

The experimental computer runs Windows 11 (Microsoft Corporation, Redmond, WA, USA) and is equipped with an NVIDIA GeForce RTX 4060 GPU (NVIDIA Corporation, Santa Clara, CA, USA). PyTorch 2.2.0 is used as the deep learning framework, and Python 3.9.13 is used for model implementation. The initial hyperparameter settings of the model are shown in Table 3. The values in parentheses represent the hyperparameter combinations obtained after optimization by the proposed optimization algorithm. Table 4 presents the main parameter settings of the STL-VMD.

Table 3. Initial hyperparameter settings of the model.

Hyperparameter	Value
Forecasting window	24
Forecasting step size	4
Embedding dimension	64 (84)
Number of CNN convolutional layers	3
CNN convolution kernel size	64
Hidden layer dimension	64 (76)
Encoder–decoder layers	2
Number of attention heads	8
Training–test split ratio	8:2
Dropout rate	0.1
Batch size	64
Training epochs	200
Learning rate	0.0001 (0.00045)
Early stopping strategy ratio	0.15
Optimizer	Adam
Validation method	Rolling-window cross-validation

Table 4. Main parameter settings of STL-VMD.

Method	Parameter	Value
STL	Seasonal period	31 days
VMD	Number of modes K	5
VMD	Penalty factor α	1896

The experiments select FEDformer, NTSformer, TimeXer, PatchTST, Informer, Transformer, DLinear, and iTransformer as baseline forecasting models. CNN-iTransformer and STL-VMD-CNN-iTransformer are also introduced for comparison to further analyze the effects of local feature extraction and multiscale decomposition on forecasting performance. To ensure a fair comparison, all models use the same data split, input window, forecasting horizon, and evaluation metrics for training and testing.

The main dataset, Dataset A, is used for correlation analysis, comparison of different decomposition methods, effectiveness analysis of the optimization algorithm, ablation experiments, and comparison experiments among different forecasting models. Datasets B and C are mainly used for comparison experiments between the proposed model and different baseline models to evaluate the generalization ability of the proposed model under different photovoltaic operating environments and data distributions.

3.3. Model Evaluation Metrics

To evaluate model performance on the test set, three metrics are used: MAE, RMSE, and coefficient of determination (R^2). Their calculation formulas are as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{27}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{28}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2} \tag{29}$$

where y_i and $\hat{y}_i$ represent the actual photovoltaic power value and forecasting value of the i-th sample, respectively; $\overline{y}_i$ represents the mean value of photovoltaic power samples; and N represents the number of samples.

3.4. Correlation Analysis and Variable Selection

To comprehensively evaluate the relationships between meteorological features and PV power, this paper employs the PCC and the MIC to analyze the linear and nonlinear correlations between variables, respectively. The corresponding results are presented in Table 5 and Table 6, respectively.

To comprehensively evaluate the correlation between meteorological variables and PV power, this paper performs a weighted fusion of the PCC and MIC based on their individual calculation results. The fusion results are presented in Table 7, and the calculation formula is given as follows:

$$Score = \alpha|\mathrm{PCC}| + (1-\alpha)\mathrm{MIC} \tag{30}$$

where α represents the weight of the linear correlation. Considering that both strong linear relationships and nonlinear relationships exist between PV power and meteorological

variables, this paper sets α = 0.6 (according to the threshold classification criteria reported in Refs. [36–39]) to comprehensively reflect the overall correlation between the variables and PV power.

Table 5. PCC analysis results for weather features and output power.

	Spring	Summer	Autumn	Winter	Annual
Global Horizontal Irradiance	0.955	0.996	0.932	0.939	0.927
Diffuse Horizontal Irradiance	0.526	0.521	0.599	0.514	0.511
Radiation Global Tilted	0.955	0.987	0.941	0.935	0.956
Radiation Diffuse Tilted	0.812	0.842	0.787	0.756	0.793
Humidity	−0.429	−0.537	−0.476	−0.429	−0.449
Temperature	0.639	0.560	0.602	0.515	0.485
Rainfall	−0.027	0.002	−0.067	−0.082	−0.022
Wind Speed	0.012	0.035	0.021	0.024	0.019
Wind Direction	−0.062	−0.133	−0.113	0.082	−0.083

Table 6. MIC analysis results for weather features and output power.

	Spring	Summer	Autumn	Winter	Annual
Global Horizontal Irradiance	0.994	0.998	0.997	0.996	0.997
Diffuse Horizontal Irradiance	0.996	0.998	0.997	0.995	0.995
Radiation Global Tilted	0.915	0.931	0.955	0.968	0.932
Radiation Diffuse Tilted	0.921	0.929	0.912	0.942	0.912
Humidity	0.424	0.365	0.449	0.325	0.312
Temperature	0.456	0.493	0.465	0.349	0.398
Rainfall	0.128	0.033	0.021	0.139	0.042
Wind Speed	0.238	0.019	0.035	0.018	0.023
Wind Direction	0.099	0.178	0.072	0.162	0.066

Table 7. Comprehensive correlation analysis results.

	Spring	Summer	Autumn	Winter	Annual
Global Horizontal Irradiance	0.971	0.997	0.958	0.962	0.955
Diffuse Horizontal Irradiance	0.715	0.712	0.758	0.706	0.705
Radiation Global Tilted	0.939	0.965	0.947	0.948	0.946
Radiation Diffuse Tilted	0.856	0.877	0.837	0.830	0.841
Humidity	0.426	0.469	0.465	0.388	0.394
Temperature	0.566	0.533	0.547	0.449	0.450
Rainfall	0.067	0.014	0.049	0.105	0.030
Wind Speed	0.023	0.029	0.027	0.022	0.021
Wind Direction	0.077	0.151	0.096	0.114	0.076

As shown in Table 7, global horizontal irradiance, radiation global tilted, radiation diffuse tilted, diffuse horizontal irradiance, and temperature achieve the highest comprehensive correlation scores, indicating that they have strong comprehensive explanatory power for PV power variation. Therefore, these five meteorological features are selected as the input variables of the proposed model.

3.5. Comparison of Different Decomposition Methods

To verify the effectiveness of the proposed STL-VMD in photovoltaic forecasting, STL decomposition, VMD, EMD, and ICEEMDAN decomposition are selected for comparison

experiments. Their performance is evaluated under the same forecasting framework. To ensure the reliability of the comparison among different decomposition methods, each decomposition method is independently run 20 times under the same forecasting framework. The evaluation metrics in the table are the average values of the 20 runs.

As shown in Table 8, the STL-VMD method achieves the best performance in terms of both RMSE and MAE, indicating that the proposed decomposition strategy can more effectively extract the characteristics of the PV power series.

Table 8. Performance comparison of different decomposition methods.

Model	RMSE	MAE	R^2
Without Decomposition	3.124	1.626	0.970
EMD	3.095	1.601	0.970
ICEEMDAN	3.025	1.568	0.971
VMD	2.921	1.529	0.973
STL	2.970	1.545	0.975
STL-VMD	2.711	1.421	0.984

EMD and ICEEMDAN are both empirical mode decomposition methods. They decompose non-stationary series into multiple intrinsic mode functions. However, they are still affected by mode mixing and end effects. As a result, they cannot effectively separate the long-term trend, periodic variation, and random fluctuations in the PV power series.

VMD can effectively suppress mode mixing and has a strong ability to decompose high-frequency fluctuations. Therefore, it achieves a better forecasting performance than EMD and ICEEMDAN. However, VMD mainly performs decomposition based on frequency-domain characteristics and cannot adequately capture the long-term trend and seasonal patterns of the PV power series.

In contrast, STL decomposes the PV power series into trend, seasonal, and residual components. It effectively extracts the long-term trend and seasonal patterns while reducing the non-stationarity of the series. Therefore, it further improves the forecasting performance. However, the residual component obtained by STL still contains high-frequency fluctuations and complex nonlinear information. STL alone cannot fully extract these fine-grained features.

Therefore, the proposed method adopts the STL-VMD multiscale decomposition strategy. This method fully exploits the advantages of STL in extracting trend and seasonal information, combining this model with VMD to further decompose the residual component. As a result, the multiscale characteristics of the PV power series can be represented more comprehensively.

3.6. Multiscale Decomposition

Figure 7 shows the STL decomposition results of the photovoltaic power sequence, including the trend, seasonal, and residual components. The trend component changes smoothly, the seasonal component shows obvious periodicity, and the residual component has strong randomness. As shown in Figure 8, VMD further decomposes the residual component into several modal components with different frequency scales. IMF0 has a large fluctuation amplitude, while the amplitudes of IMF1 to IMF4 gradually decrease. This process reduces noise interference and helps improve model forecasting accuracy.

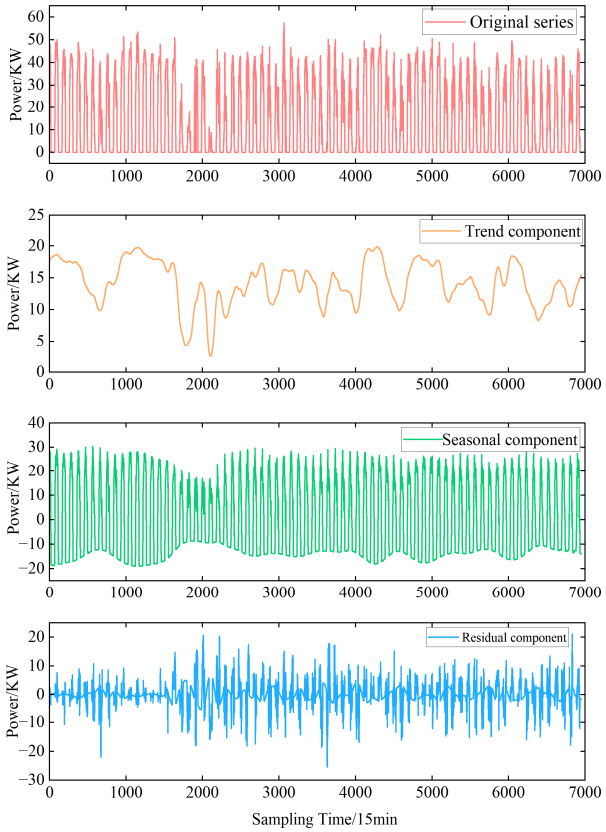


Figure 7. STL decomposition results.

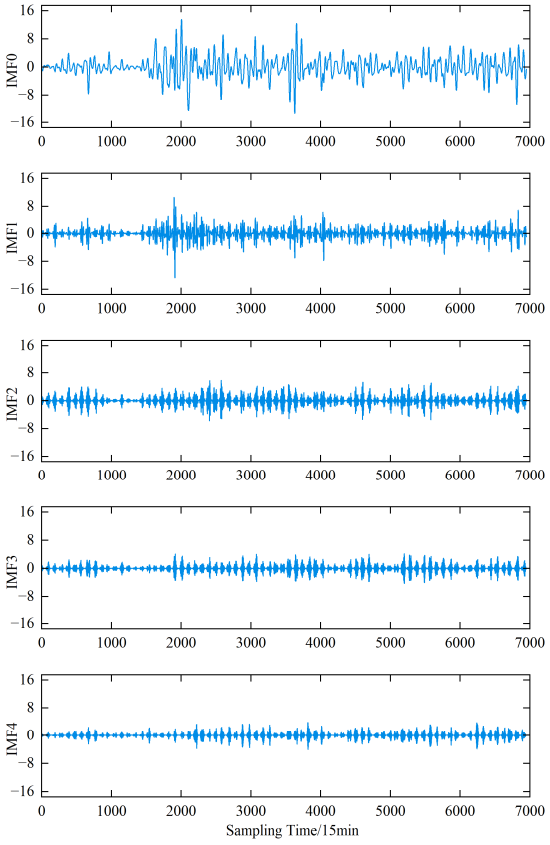


Figure 8. VMD results of the residual component.

In the comparative experiments, VMD secondary decomposition is also performed on the residual component obtained after STL decomposition. Since its decomposition process is consistent with that of the ablation experiment, the seasonal decomposition results are not shown separately.

For possible information leakage in the application of multiscale decomposition to forecasting, this paper provides protection at three levels: data splitting, sliding window construction, and code implementation. First, in the data preprocessing stage, the training set and test set are strictly divided. Decomposition is then performed independently on the two sets. This prevents information from the test set from participating in the training process. Second, during decomposition, a sliding window consistent with the input length of the forecasting model is adopted. Each decomposition is completed based on historical data within the current window only. This prevents information leakage from future time steps. Finally, in the code implementation stage, a masking mechanism is introduced. Unobserved time steps are masked to block potential information transmission paths. Through the above processing, the risk of information leakage during the decomposition process is minimized, thereby ensuring the reliability of the forecasting experiments.

3.7. Effectiveness Analysis of Optimization Algorithms

To verify the effectiveness of the PGA optimization algorithm in model hyperparameter optimization, five intelligent optimization algorithms, including CPO, WOA, SSA, GWO, and PSO, are selected for comparison with PGA. Each optimization algorithm is independently run 20 times, and its convergence performance and stability are analyzed based on the multiple-run results. The initial settings of all optimization algorithms are shown in Table 9.

Table 9. Initial hyperparameter settings of optimization algorithms.

Hyperparameter	Value
Population size	10
Maximum iterations	30
Search dimension	3
Lower bound	[0.0001, 32, 32]
Upper bound	[0.001, 128, 128]
Fitness function	MSE

Its fitness convergence curve is shown in Figure 9; different algorithms show large differences in convergence speed and final fitness value. PGA shows strong search ability at the early iteration stage. Its fitness value decreases rapidly in the first few iterations. It converges to a low level after about the 7th iteration and finally stabilizes around 1.5×10^{-4}. This indicates that PGA can quickly find an optimal solution within a small number of iterations. It can also maintain stable convergence in the subsequent search process. This result suggests that PGA has good global search ability and local exploitation ability.

In contrast, SSA and GWO show small decreases in fitness value at the early stage. Their overall convergence processes are relatively smooth. This indicates that they have limited ability to locate optimal hyperparameter regions in the initial search stage. WOA and PSO can gradually reduce fitness values during iteration. However, they easily enter a plateau stage in later iterations. Their further optimization ability is limited, and their final fitness values are higher than that of PGA. CPO shows a rapid downward trend at the early stage. However, its convergence stability and final optimization accuracy are still weaker than those of PGA.

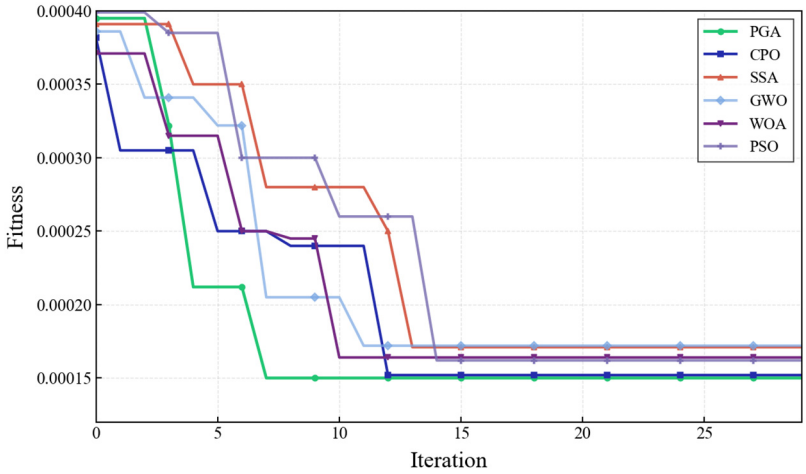


Figure 9. Convergence performance comparison.

Figure 10 is a standard deviation comparison histogram of different optimization algorithms. Standard deviation is used to measure the fluctuation degree of objective function values during multiple independent runs of an optimization algorithm. A smaller standard deviation indicates that algorithm results are more stable under different running conditions. Conversely, a larger standard deviation indicates that the algorithm is more easily affected by random factors and initial parameters, and its convergence process has larger fluctuations.

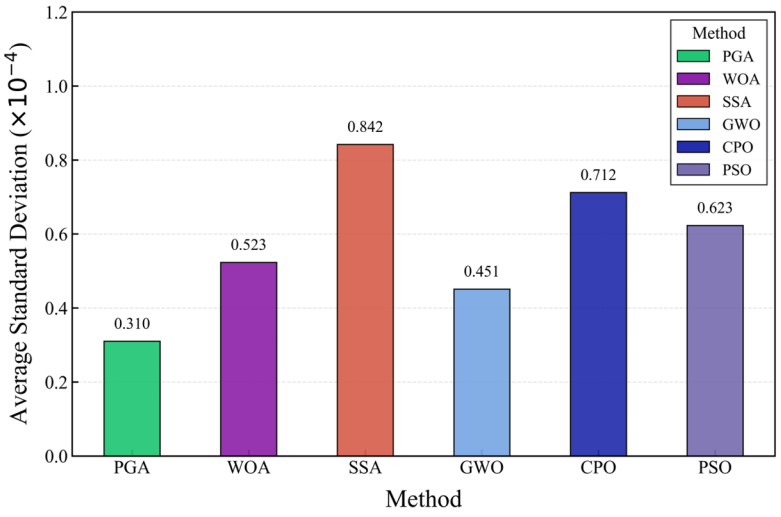


Figure 10. Standard deviation comparison of different optimization algorithms.

As shown in Figure 10, the standard deviation of PGA is 0.31×10^{-4}, which shows a relatively good performance among different optimization algorithms. This indicates that PGA can maintain stable search ability during multiple iterations. It can effectively reduce the influence of random disturbances on optimization results. It also has good global optimization ability and convergence consistency.

Meanwhile, PGA can balance search efficiency and solution stability while ensuring convergence accuracy. It can avoid premature convergence to local optima. Therefore, it improves overall performance in terms of model hyperparameter optimization.

3.8. Ablation Experiment

To verify the accuracy of the proposed model, the full-year photovoltaic power data of 2017 are used for the experiments. iTransformer (M0), CNN-iTransformer (M1), STL-CNN-

iTransformer (M2), VMD-CNN-iTransformer (M3), and STL-VMD-CNN-iTransformer (M4) are selected for the ablation experiments. All models are optimized using the PGA. The experimental results are shown in Table 10. Each ablation model is independently trained 20 times. The RMSE, MAE, and R^2 values in the table are the average results of the 20 runs. The results outside the parentheses are obtained after PGA optimization, while the results inside the parentheses are obtained without PGA optimization. Figure 11 shows comparison results of different models in terms of RMSE, MAE, and R^2. It can directly show differences in forecasting performance before and after PGA optimization.

Table 10. Performance comparison of different models in the ablation experiment.

Number	Model	RMSE	MAE	R^2
M0	iTransformer	3.352 (3.587)	1.721 (1.823)	0.962 (0.955)
M1	CNN-iTransformer	3.124 (3.364)	1.626 (1.722)	0.970 (0.966)
M2	STL-CNN-iTransformer	2.970 (3.125)	1.545 (1.598)	0.975 (0.971)
M3	VMD-CNN-iTransformer	2.921 (3.163)	1.529 (1.634)	0.973 (0.969)
M4	STL-VMD-CNN-iTransformer	2.711 (2.925)	1.421 (1.511)	0.984 (0.981)

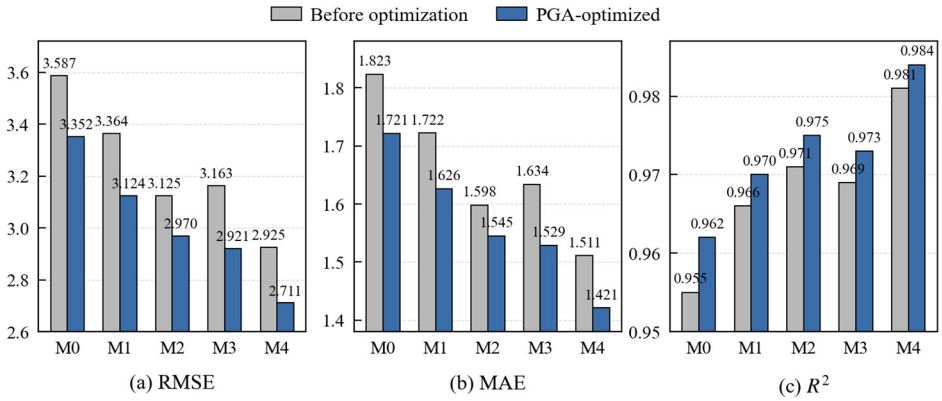


Figure 11. Comparison of forecasting performance of different models before and after PGA optimization.

To separately evaluate the effect of each model on forecasting performance, the experimental results without PGA optimization are first compared, that is, the data in parentheses in Table 10. As shown by data in parentheses in Table 10, RMSE, MAE, and R^2 of M0 are 3.587, 1.823, and 0.955, respectively. Its forecasting error is relatively large. This indicates that M0 has certain limitations in handling local abrupt changes and high-frequency fluctuations in photovoltaic power sequences. After CNN is introduced, RMSE and MAE of M1 decrease to 3.364 and 1.722, respectively. Compared with M0, they decrease by about 6.22% and 5.54%, respectively. This indicates that CNN can effectively extract local fluctuation features. It can also enhance model perception of short-term fluctuations in photovoltaic data. After STL decomposition is added, the RMSE of M2 decreases to 3.125. It is about 7.10% lower than that of M1. These results indicate that STL can separate trend, seasonal, and residual components. It can reduce non-stationarity of photovoltaic sequences. The RMSE of M3 is 3.163, which is about 5.98% lower than that of M1. This indicates that VMD has a certain decomposition ability for high-frequency fluctuations. However, when used alone, it is less sufficient than STL in characterizing trend and periodic features.

In contrast, the RMSE, MAE, and R^2 values of M4 are 2.925, 1.511, and 0.981, respectively. Compared with M0, RMSE and MAE decrease by about 18.46% and 17.11%, respectively, and R^2 increases by about 2.6%. Compared with M2, RMSE and MAE further decrease by about 6.40% and 5.44%, respectively. This indicates that multiscale decomposition based on STL and VMD can extract trend, periodic, and residual high-frequency

features at the same time. Therefore, it improves the model's ability to represent complex photovoltaic sequences.

Further comparisons are conducted under the same PGA optimization conditions. As shown in Table 10, after introducing the PGA optimization algorithm, the RMSE of all models is reduced by about 4.96–7.65%, and the MAE is reduced by about 3.32–6.43%. This indicates that PGA can improve the model parameter configuration through adaptive optimization and thus improve the forecasting performance. Comparing M4 with M2, the RMSE and MAE of M4 are reduced by about 8.72% and 8.03%, respectively. This indicates that, under the same optimization conditions, further applying VMD secondary decomposition to the STL residual component can further enhance the model's ability to represent high-frequency non-stationary fluctuations. After PGA optimization, the RMSE and MAE of M4 are 2.711 and 1.421, respectively. Compared with 3.352 and 1.721 of M0, they are reduced by about 19.12% and 17.43%, respectively. This shows that the overall framework consisting of multiscale decomposition, CNN local feature extraction, and iTransformer variable relationship modeling can effectively improve the forecasting accuracy.

3.9. Generalization Performance Under Different Weather Conditions

To verify the adaptability of the proposed model under different weather conditions, three typical weather scenarios, including sunny, cloudy, and rainy days, are selected for forecasting experiments. The forecasting results are shown in Figure 12. The photovoltaic power variation patterns are clearly different under different weather conditions. The power curve is relatively smooth on sunny days. Under cloudy conditions, the power fluctuates continuously due to cloud changes. On rainy days, the power fluctuates most strongly and shows greater randomness. Therefore, these scenarios can comprehensively evaluate the forecasting performance of the model under different weather conditions. As shown in Figure 12, all models can basically capture the overall variation trend of photovoltaic power, but clear differences appear in rapid ramping, local fluctuations, and abrupt changes.

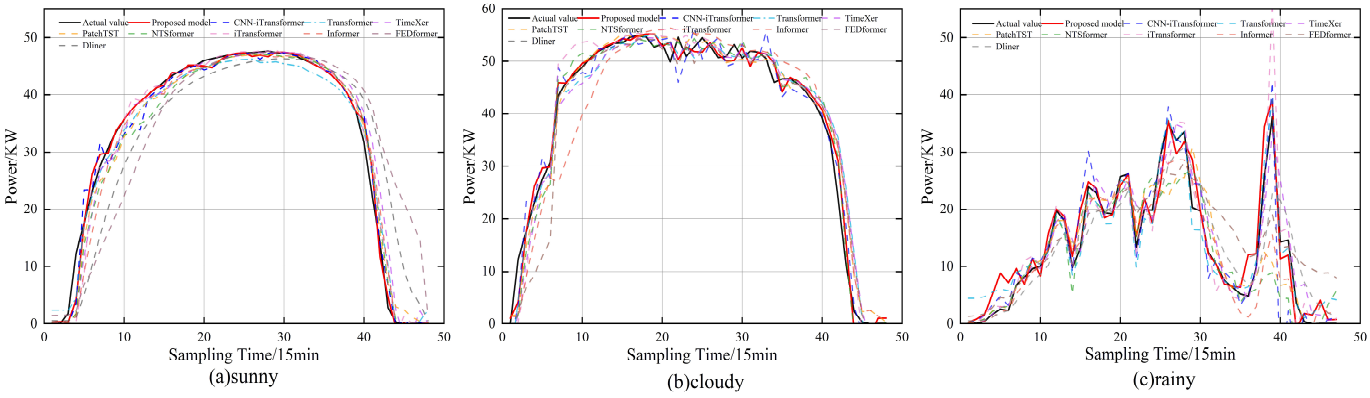


Figure 12. Prediction results of different models under different weather conditions.

Under sunny conditions, the power curve shows a relatively regular single-peak pattern. The differences among the models are mainly concentrated in the rapid ramp-up after sunrise and the rapid decline before sunset. FEDformer and DLinear show relatively obvious deviations in these boundary-change regions. FEDformer focuses more on extracting the main frequency components and global periodic structure, so its response to short-duration rapid changes is relatively smooth. DLinear mainly relies on the linear mapping of trend and periodic components. When the power shows a large nonlinear rate of change within a short period, it may produce a certain lag or amplitude deviation.

Under cloudy conditions, multiple declines and recoveries continuously occur in the medium- and high-power regions. Short-term high-frequency disturbances are superimposed on the stable daily cycle. As shown in the figure, Informer and some models based on global attention show relatively obvious deviations at local sudden drops. This is because the ProbSparse attention of Informer tends to retain dominant relationships with stronger global contributions. However, short-term disturbances caused by cloud cover have a short duration and can be weakened by the relatively stable daily periodic components during global feature aggregation. Therefore, these models can recover the overall trend well, but their responses to continuous rapid declines and recoveries are relatively insufficient.

Under rainy conditions, the overall power level is clearly reduced and is accompanied by multiple irregular peaks and valleys, as well as sudden rises and drops. The randomness and non-stationarity of the series are further increased. In this case, models such as Transformer and FEDformer show relatively obvious amplitude deviations or smoothing at some peaks and valleys. The traditional Transformer mainly relies on global temporal relationships for feature interaction, which makes it difficult to sufficiently constrain short-duration local abrupt changes. Although FEDformer can use frequency-domain information to describe the main periodic structure, the stable periodic characteristics are weakened under rainy conditions, while the proportion of random disturbances increases. This reduces its advantage in modeling the global frequency structure.

In contrast, the proposed model can track the actual power variation well under all three weather scenarios. STL-VMD is used to separate the trend, periodic, and high-frequency fluctuation components. CNN is used to extract local dynamic features, and iTransformer is used to learn the nonlinear relationships between meteorological variables and photovoltaic power. Therefore, the model can effectively handle rapid boundary changes on sunny days, continuous fluctuations under cloudy conditions, and strong random abrupt changes on rainy days. The results show that multiscale decomposition, local dynamic feature extraction, and variable relationship modeling have good complementarity. This improves the adaptability and forecasting stability of the model under different weather conditions.

3.10. Seasonal Comparison Experiments

To further verify the generalization ability of the proposed model under different seasonal conditions, this paper divides the 2017 photovoltaic power data into four seasonal subsets: spring, summer, autumn, and winter. For each seasonal subset, training and test sets are divided at a ratio of 8:2 while maintaining chronological order. To reduce the impact of training fluctuations on the experimental results, each model is independently run 20 times on each seasonal subset. The RMSE, MAE, and R^2 values in the table are the average results of the 20 runs.

As shown in Figure 13, the photovoltaic power series show clear differences across different seasons. All models can generally capture the overall daily variation trend, but they show different degrees of deviation in rapid ramping, peak transitions, and local fluctuation regions. In spring, the power curve is relatively regular and shows a clear ramp-up–high-power–decline process. Therefore, most models can recover the main profile well, and the differences are mainly concentrated in the rapid ramp-up after sunrise and the rapid decline before sunset. Among them, FEDformer shows a relatively obvious response lag during the ramp-up stage. DLinear and some Transformer-based models still show certain deviations near the peak and decline regions, but the overall differences are relatively limited. Autumn is similar to spring. The main trend is relatively stable, and all models show good overall fitting performance. However, amplitude deviations of PatchTST, FEDformer, and other models can still be observed at the local valley in the

early stage and the rapid decline in the later stage. This indicates that when local changes last for a short time, patch-level aggregation or main periodic modeling may weaken transient information. In summer, multiple declines and recoveries continuously occur in the medium- and high-power regions. Local high-frequency disturbances are clearly superimposed on the stable daily cycle. Informer and Transformer show relatively obvious overshooting or insufficient tracking at some peaks and valleys. This is because global attention is more likely to learn the main trend with a longer duration, while short-duration local disturbances can be weakened during feature aggregation. In winter, the overall power level is relatively low and is accompanied by frequent sudden rises, sudden drops, and rapid recoveries. The non-stationarity of the series is further increased. In this case, FEDformer and DLinear show relatively large amplitude deviations or smoothing at some peaks and valleys. This indicates that models mainly relying on frequency structures or linear trend mapping have limited adaptability to strong random fluctuations under a low power baseline.

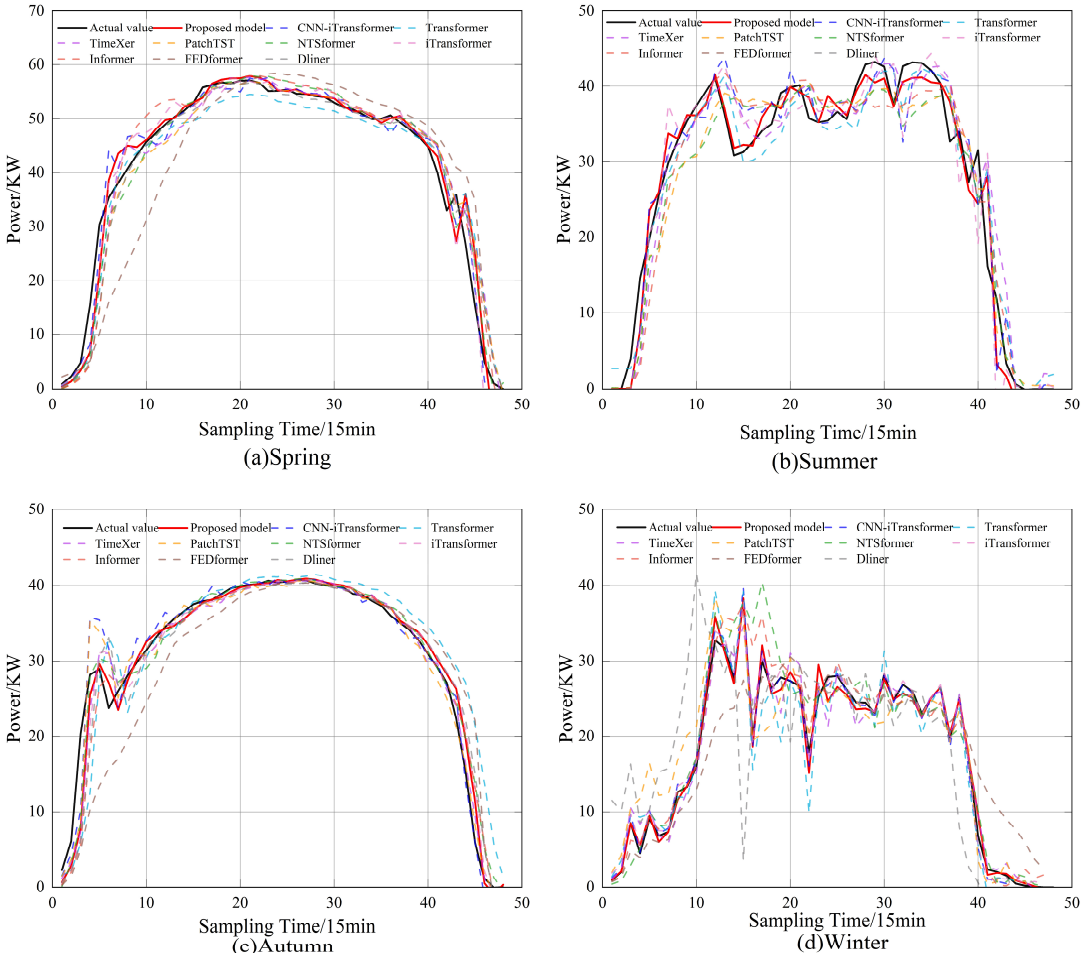


Figure 13. Forecasting results of different models in four seasons.

In contrast, the proposed model maintains good tracking performance in all four seasons. It not only accurately describes the relatively stable overall trends in spring and autumn, but also closely follows the actual power in complex regions, such as continuous fluctuations in summer and rapid changes in winter. This is because STL-VMD first separates the trend, periodic, and residual high-frequency components at different time scales, which reduces the interference between the stable main trend and short-term disturbances. CNN further enhances local continuous changes and short-term abrupt features,

while iTransformer learns the nonlinear relationships between meteorological factors and photovoltaic power from the variable dimension. Therefore, multiscale decomposition, local dynamic feature extraction, and variable relationship modeling complement each other. This enables the model to adapt to both regular variation and strongly non-stationary fluctuation scenarios, showing good seasonal adaptability and forecasting stability.

To further quantify the differences in forecasting performance among different models, Table 11 presents the evaluation results of each model in terms of RMSE, MAE, and R^2. The values in parentheses are the results without PGA optimization, while the values outside the parentheses are the results after PGA optimization. As shown in Table 11, the forecasting errors of most models decrease after PGA optimization, and the goodness of fit is further improved. This indicates that the optimization algorithm can effectively improve the model parameter configuration and enhance the forecasting accuracy and adaptability of the model. The proposed model shows lower forecasting errors than all comparison models. Compared with CNN-iTransformer, the RMSE of the proposed model is reduced by 11.81–14.03% across the four seasons, and the MAE is reduced by 12.54–14.44%. Compared with iTransformer, the RMSE is reduced by 18.35–18.86%, and the MAE is reduced by 17.45–20.57%. Compared with all comparison models in Table 11, the RMSE of the proposed model is reduced by approximately 11.81–62.14%, and the MAE is reduced by approximately 11.41–74.13%.

Table 11. Performance comparison of different models in comparative experiments.

Model	Metrics	Season			
		Spring	Summer	Autumn	Winter
FEDformer	RMSE	7.535 (7.957)	6.082 (6.444)	4.947 (5.228)	5.252 (5.519)
	MAE	4.643 (4.933)	3.871 (4.095)	2.766 (2.931)	3.181 (3.370)
	R^2	0.895 (0.881)	0.877 (0.864)	0.925 (0.907)	0.896 (0.877)
NTSformer	RMSE	5.894 (6.228)	5.103 (5.541)	3.931 (4.215)	3.821 (4.097)
	MAE	2.746 (2.931)	2.942 (3.124)	2.301 (2.449)	2.066 (2.224)
	R^2	0.921 (0.907)	0.901 (0.892)	0.956 (0.939)	0.938 (0.932)
TimeXer	RMSE	3.950 (4.160)	3.870 (4.138)	4.265 (4.512)	3.743 (3.914)
	MAE	1.867 (1.971)	1.805 (1.956)	2.411 (2.562)	1.956 (2.067)
	R^2	0.977 (0.967)	0.976 (0.968)	0.959 (0.942)	0.945 (0.938)
PatchTST	RMSE	3.802 (4.021)	4.528 (4.898)	4.556 (4.856)	5.059 (5.435)
	MAE	1.789 (1.922)	2.570 (2.712)	2.580 (2.759)	2.932 (3.128)
	R^2	0.973 (0.969)	0.936 (0.928)	0.937 (0.921)	0.899 (0.881)
Informer	RMSE	4.006 (4.289)	4.927 (5.235)	3.502 (3.722)	4.099 (4.349)
	MAE	2.369 (2.511)	3.727 (3.987)	1.409 (1.486)	2.364 (2.487)
	R^2	0.965 (0.953)	0.885 (0.871)	0.982 (0.976)	0.942 (0.937)
Transformer	RMSE	5.518 (5.877)	4.955 (5.226)	4.999 (5.412)	3.854 (4.118)
	MAE	3.088 (3.291)	2.749 (2.929)	2.636 (2.841)	1.944 (2.084)
	R^2	0.856 (0.841)	0.924 (0.912)	0.921 (0.901)	0.946 (0.931)
DLinear	RMSE	5.353 (5.674)	5.467 (5.900)	3.413 (3.674)	4.547 (4.795)
	MAE	3.136 (3.321)	2.531 (2.730)	1.340 (1.453)	1.919 (2.011)
	R^2	0.902 (0.893)	0.933 (0.921)	0.987 (0.978)	0.935 (0.928)
iTransformer	RMSE	3.502 (3.738)	3.022 (3.240)	2.354 (2.555)	3.594 (3.819)
	MAE	1.512 (1.605)	1.318 (1.418)	1.096 (1.157)	2.103 (2.254)
	R^2	0.980 (0.973)	0.974 (0.965)	0.984 (0.977)	0.948 (0.941)
CNN-iTransformer	RMSE	3.235 (3.468)	2.859 (3.029)	2.230 (2.383)	3.328 (3.582)
	MAE	1.392 (1.495)	1.244 (1.317)	1.019 (1.072)	1.987 (2.121)
	R^2	0.980 (0.978)	0.980 (0.972)	0.986 (0.978)	0.970 (0.968)
STL-VMD-CNN-iTransformer	RMSE	2.853 (3.058)	2.458 (2.621)	1.922 (2.054)	2.916 (3.132)
	MAE	1.201 (1.292)	1.088 (1.156)	0.872 (0.926)	1.700 (1.821)
	R^2	0.991 (0.983)	0.988 (0.982)	0.992 (0.988)	0.982 (0.978)

In summary, multiscale decomposition, CNN-based local feature extraction, iTransformer-based variable relationship modeling, and PGA-based hyperparameter optimization have good synergistic effects. They can effectively improve the forecasting accuracy of the model under different seasonal scenarios.

3.11. Generalization Experiment Analysis

To further verify the generalization ability of the proposed model under different regions and data distributions, the public dataset from the China State Grid Renewable Energy Power Forecasting Competition (Dataset B) and a photovoltaic power station dataset from Xinjiang, China (Dataset C) are selected for the experiments. The typical-day forecasting results of different models on the two datasets are shown in Figure 14 and Figure 15, respectively. The corresponding evaluation metrics are listed together in Table 12.

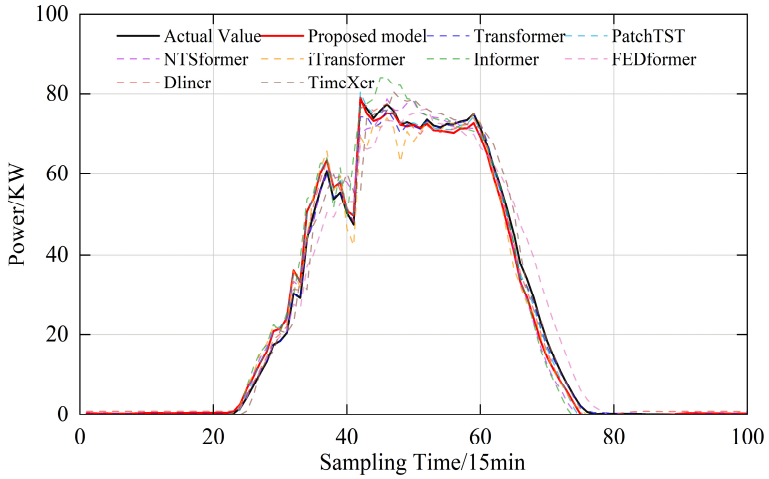


Figure 14. Forecasting results of different models (Dataset B).

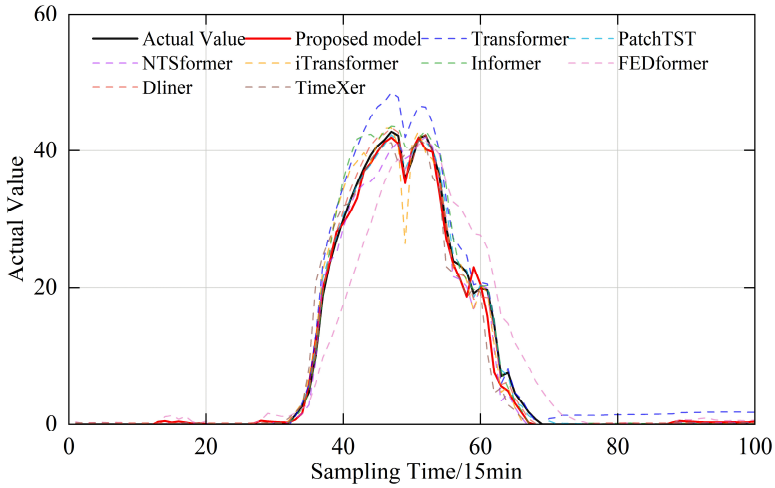


Figure 15. Forecasting results of different models (Dataset C).

Table 12. Performance comparison of different models on generalization datasets.

Model	Dataset B			Dataset C		
	RMSE	MAE	R^2	RMSE	MAE	R^2
FEDformer	7.265	4.412	0.921	5.972	3.349	0.839
TimeXer	5.161	2.712	0.960	2.607	1.168	0.969
Informer	5.153	2.761	0.960	2.912	1.381	0.962

Table 12. *Cont.*

Model	Dataset B			Dataset C		
	RMSE	MAE	R^2	RMSE	MAE	R^2
NTSformer	4.930	2.610	0.963	2.622	1.180	0.969
Transformer	3.761	1.834	0.979	3.207	2.056	0.954
DLinear	3.606	2.015	0.981	2.214	1.013	0.977
PatchTST	3.587	2.193	0.981	2.255	0.980	0.977
iTransformer	3.544	1.728	0.981	2.187	0.899	0.978
Proposed Model	3.364	1.618	0.983	1.797	0.883	0.985

For Dataset B, the power curve shows an obvious intraday variation pattern, but local fluctuations still appear during the rapid ramp-up stage and near the peak. As shown in Figure 14, FEDformer shows a relatively obvious lag and amplitude deviation in the rapid ramp-up and local variation regions, with an RMSE of 7.265. Informer can recover the overall trend relatively well, but certain deviations still appear in local rapid-change regions, with an RMSE of 5.153. In contrast, iTransformer uses variable-level attention to enhance the modeling of relationships among multiple meteorological variables, and its forecasting error is clearly reduced. The proposed model further achieves an RMSE of 3.364, an MAE of 1.618, and an R^2 of 0.983. All three metrics are the best. This indicates that multiscale decomposition and local dynamic modeling can further improve the forecasting accuracy near the peak and in rapid-change regions of this dataset.

Dataset C shows more concentrated local dynamic characteristics. As shown in Figure 15, the power enters the medium- and high-power range after a rapid ramp-up. Continuous peak variations, local sudden drops, and recoveries appear around the 45th–60th sampling points. At this time, the relative performance of different models changes clearly. For example, the RMSE values of DLinear and PatchTST decrease to 2.214 and 2.255, respectively. This indicates that trend mapping and local patch modeling can achieve good fitting performance when the main trend is relatively clear. However, Transformer still shows obvious overshooting and response deviations in the peak and continuous fluctuation regions, with an RMSE of 3.207. This shows that relying only on global temporal relationships is insufficient to fully capture continuous local changes within a short period. The proposed model shows higher consistency with the actual curve at these peaks, valleys, and rapid-change positions. Its RMSE, MAE, and R^2 are 1.797, 0.883, and 0.985, respectively, which are all better than those of the other comparison models.

Considering the two datasets together, the relative advantages of different baseline models change with the power variation patterns and data distributions, while the proposed model maintains relatively stable forecasting performance in both scenarios. This is because STL-VMD can separate the trend, periodic, and high-frequency fluctuation components and reduce the coupling among information at different time scales. CNN further enhances the extraction of rapid ramping, local peaks and valleys, and short-term abrupt-change features. iTransformer uses variable-level attention to learn the nonlinear relationships between meteorological variables and photovoltaic power. Therefore, the proposed model can adapt to local fluctuations under relatively regular trends in Dataset B and can also effectively handle continuous peaks, valleys, and rapid changes in Dataset C. This further verifies its cross-dataset generalization ability.

4. Conclusions

To address the problems in existing studies, this paper proposes a CNN-iTransformer method for short-term photovoltaic power forecasting based on multiscale decomposition and adaptive optimization. These problems include the difficulty in accurately characteriz-

ing complex temporal features of photovoltaic time series, the limited ability of forecasting models to capture local abrupt changes and nonlinear relationships among variables, and the difficulty in model hyperparameter optimization.

(1) Multiscale decomposition is performed on photovoltaic time series. This reduces non-stationarity of photovoltaic power sequences and refines high-frequency fluctuation features in residual components. A CNN-iTransformer hybrid forecasting model is constructed to improve the ability to capture local abrupt variation information in photovoltaic sequence data and nonlinear relationships among variables. Simulation results show that the proposed model achieves RMSE = 2.711, MAE = 1.421, and R^2 = 0.984 in the ablation experiment. It outperforms iTransformer, CNN-iTransformer, and other ablation models. This indicates that multiscale decomposition and the CNN-iTransformer structure have good synergistic forecasting ability.

(2) PGA is introduced to optimize model hyperparameters. According to ablation and comparative experiment results, forecasting errors of all models decrease to different degrees after PGA optimization. Specifically, RMSE decreases by about 4.37–7.90%, and MAE decreases by about 3.32–7.78%. This indicates that PGA can effectively improve model parameter configuration. It also improves model forecasting accuracy and adaptability.

Future research can further introduce multi-source information, such as cloud images and numerical weather prediction data. This can enhance model perception of photovoltaic power changes under abrupt weather conditions. Meanwhile, model structure and parameter search efficiency can be further optimized to reduce training cost. This will make the model more suitable for online forecasting and dispatching applications in practical photovoltaic power stations.

Author Contributions: Conceptualization, Z.G., B.P. and F.Z.; methodology, Z.G. and B.P.; software, Z.G. and B.P.; investigation, Z.G. and F.Z.; formal analysis, Z.G.; visualization, Z.G.; validation, F.Z.; writing—original draft preparation, Z.G. and B.P.; writing—review and editing, F.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 62273215.

Data Availability Statement: The data presented in this study were derived from publicly available resources. The 2017 Alice Springs photovoltaic power and meteorological data were obtained from the Desert Knowledge Australia Solar Centre (DKASC) and are openly available at https://dkasolarcentre.com.au/download?location=alice-springs (accessed on 8 January 2026).

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

Abbreviations

Abbreviation	Full Name
PV	Photovoltaic
ARIMA	Autoregressive Integrated Moving Average
RF	Random Forest
SVM	Support Vector Machine
ELM	Extreme Learning Machine
XGBoost	Extreme Gradient Boosting
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
BiLSTM	Bidirectional Long Short-Term Memory
GRU	Gated Recurrent Unit
CNN	Convolutional Neural Network
TCN	Temporal Convolutional Network
BPNN	Backpropagation Neural Network

STL	Seasonal-Trend Decomposition Using Loess
VMD	Variational Mode Decomposition
EMD	Empirical Mode Decomposition
ICEEMDAN	Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
PGA	Phototropic Growth Algorithm
CPO	Crested Porcupine Optimizer
WOA	Whale Optimization Algorithm
SSA	Sparrow Search Algorithm
GWO	Grey Wolf Optimizer
PSO	Particle Swarm Optimization
PCC	Pearson Correlation Coefficient
MIC	Maximal Information Coefficient
MI	Mutual Information
MSE	Mean Squared Error
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination
DKASC	Desert Knowledge Australia Solar Centre
GPU	Graphics Processing Unit

References

1. Shang, Y.; Sang, S.; Tiwari, A.K.; Khan, S.; Zhao, X. Impacts of renewable energy on climate risk: A global perspective for energy transition in a climate adaptation framework. *Appl. Energy* **2024**, *362*, 122994. [CrossRef]
2. Gielen, D.; Boshell, F.; Saygin, D.; Bazilian, M.D.; Wagner, N.; Gorini, R. The role of renewable energy in the global energy transformation. *Energy Strategy Rev.* **2019**, *24*, 38–50. [CrossRef]
3. Tian, J.; Ooka, R.; Lee, D. Multi-scale solar radiation and photovoltaic power forecasting with machine learning algorithms in urban environment: A state-of-the-art review. *J. Clean. Prod.* **2023**, *426*, 139040. [CrossRef]
4. Sahu, B.K. A study on global solar PV energy developments and policies with special focus on the top ten solar PV power producing countries. *Renew. Sustain. Energy Rev.* **2015**, *43*, 621–634. [CrossRef]
5. Li, G.; Wang, H.; Zhang, S.; Xin, J.; Liu, H. Recurrent neural networks based photovoltaic power forecasting approach. *Energies* **2019**, *12*, 2538. [CrossRef]
6. Zhang, T.; Zhao, W.; He, Q.; Xu, J. Optimization of microgrid dispatching by integrating photovoltaic power generation forecast. *Sustainability* **2025**, *17*, 648. [CrossRef]
7. Gong, J.; Qu, Z.; Zhu, Z.; Xu, H. Parallel TimesNet-BiLSTM model for ultra-short-term photovoltaic power forecasting using STL decomposition and auto-tuning. *Energy* **2025**, *320*, 135286. [CrossRef]
8. Liu, W.; Mao, Z. Short-term photovoltaic power forecasting with feature extraction and attention mechanisms. *Renew. Energy* **2024**, *226*, 120437. [CrossRef]
9. Liu, Z.; Guo, J.; Wang, X.; Wang, Y.; Li, W.; Wang, X.; Fan, Y.; Wang, W. Prediction of long-term photovoltaic power generation in the context of climate change. *Renew. Energy* **2024**, *235*, 121263. [CrossRef]
10. Hassan, M.A.; Bailek, N.; Bouchouicha, K.; Nwokolo, S.C. Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks. *Renew. Energy* **2021**, *171*, 191–209. [CrossRef]
11. Pierre, A.A.; Akim, S.A.; Semenyo, A.K.; Babiga, B. Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches. *Energies* **2023**, *16*, 4739. [CrossRef]
12. Ding, S.; Li, R.; Tao, Z. A Novel Adaptive Discrete Grey Model with Time-Varying Parameters for Long-Term Photovoltaic Power Generation Forecasting. *Energy Convers. Manag.* **2021**, *227*, 113644. [CrossRef]
13. Lateko, A.A.H.; Yang, H.-T.; Huang, C.-M. Short-Term PV Power Forecasting Using a Regression-Based Ensemble Method. *Energies* **2022**, *15*, 4171. [CrossRef]
14. Abuella, M.; Chowdhury, B. Solar Power Forecasting Using Support Vector Regression. *arXiv* **2017**, arXiv:1703.09851. [CrossRef]
15. Tang, P.; Chen, D.; Hou, Y. Entropy Method Combined with Extreme Learning Machine Method for the Short-Term Photovoltaic Power Generation Forecasting. *Chaos Solitons Fractals* **2016**, *89*, 243–248. [CrossRef]
16. Zhou, B.; Chen, X.; Li, G.; Gu, P.; Huang, J.; Yang, B. XGBoost–SFS and Double Nested Stacking Ensemble Model for Photovoltaic Power Forecasting under Variable Weather Conditions. *Sustainability* **2023**, *15*, 13146. [CrossRef]

17. Nadour, M.; Rabehi, A.; Hadroug, N.; Guermoui, M.; Tibermacine, I.E.; Alanazi, A.K.; Habib, M.; Rabehi, A. Deep Hybrid CNN–BiLSTM Model for Accurate Solar Photovoltaic Power Forecasting: A Comparative Study with Classical and Neural Models. *Energy Rep.* **2026**, *15*, 109119. [CrossRef]
18. Oumiguil, L.; Nejmi, A. Short-Term PV Plant Power Forecasting Model Using CNN-GRU with XGBoost Algorithm. *Sci. Afr.* **2026**, *31*, e03184. [CrossRef]
19. Lei, M.; Chen, T.; Xiao, Y.; Yang, C.; Sa-Ngiamvibool, W.; Wattana, S.; Wattana, B. Short-Term Photovoltaic Power Forecasting Based on CNN–LSTM–Attention Mechanism. *Energies* **2026**, *19*, 1747. [CrossRef]
20. Ye, H.; Xu, T. Research on the Distributed Photovoltaic Power Prediction Method Based on CNN–LSTM. *Int. J. Low-Carbon Technol.* **2026**, *21*, ctaf131. [CrossRef]
21. Zhang, J.; Xie, L.; Feng, D.; Fang, S.; Zhang, K. CEEMDAN-Informer: A Hybrid Approach for Enhanced Short-Term PV Power Forecasting. *Comput. Electr. Eng.* **2026**, *134*, 111132. [CrossRef]
22. Wen, Y.; Pan, S.; Li, X.; Li, Z.; Wen, W. Improving Multi-Site Photovoltaic Forecasting with Relevance Amplification: DeepFEDformer-Based Approach. *Energy* **2024**, *299*, 131479. [CrossRef]
23. Li, N.; Xu, Z.; Yan, X.; Zhong, J.; Miao, Q. A Parallel Hybrid TCN-PatchTST Approach to Improved Solar Photovoltaic Power Forecasting. In *Proceedings of the 2025 4th Asian Conference on Frontiers of Power and Energy (ACFPE)*; IEEE: New York, NY, USA, 2025; pp. 12–18. [CrossRef]
24. Nguyen, T.-T.; Nguyen, N.-T.; Pham, M.-H.; Vu, M.-P. Integrating Exogenous Variables into Cross-Attention Transformers for Robust 15-Minute-Ahead Solar Power Prediction. *IEEE Access* **2026**, *14*, 67930–67947. [CrossRef]
25. Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *arXiv* **2023**, arXiv:2310.06625. [CrossRef]
26. Liu, L.; Liu, M.; Han, Z.; Zhao, H. Interpretable Ultra-Short-Term Photovoltaic Power Forecasting with Multi-Scale Temporal Modeling and Variable-Wise Attention. *Sci. Rep.* **2026**, *16*, 10336. [CrossRef]
27. Duan, Y.; Zhang, Y.; Wang, X.; Xue, Y.; Wang, Z.; Wu, D. Feature-Enhanced iTransformer: A Two-Stage Framework for High-Accuracy Long-Horizon Traffic Flow Forecasting. *PLoS ONE* **2026**, *21*, e0340389. [CrossRef]
28. Zou, Z.; Zhou, X.-X.; Liu, S.-J.; Hsu, C.-Y. FT-iTransformer: A Stock Price Prediction Model Based on Time–Frequency Domain Collaborative Analysis. *Technologies* **2026**, *14*, 61. [CrossRef]
29. Yadav, H.K.; Pal, Y.; Tripathi, M.M. Short-Term PV Power Forecasting Using Empirical Mode Decomposition in Integration with Back-Propagation Neural Network. *J. Inf. Optim. Sci.* **2020**, *41*, 25–37. [CrossRef]
30. Li, Y.; Zhai, S.; Yi, G.; Pang, S.; Luo, X. Short-Term Photovoltaic Power Forecasting Based on ICEEMDAN-TCN-BiLSTM-MHA. *Symmetry* **2025**, *17*, 1599. [CrossRef]
31. Jia, P.; Zhang, H.; Liu, X.; Gong, X. Short-Term Photovoltaic Power Forecasting Based on VMD and ISSA-GRU. *IEEE Access* **2021**, *9*, 105939–105950. [CrossRef]
32. Wang, S.; Yan, S.; Li, H.; Zhang, T.; Jiang, W.; Yang, B.; Li, Q.; Li, M.; Zhang, N.; Wang, J. Short-Term Prediction of Photovoltaic Power Based on Quadratic Decomposition and Residual Correction. *Electr. Power Syst. Res.* **2024**, *236*, 110968. [CrossRef]
33. Sibtain, M.; Li, X.; Saleem, S.; Qurat-Ul-Ain; Asad, M.S.; Tahir, T.; Apaydin, H. A Multistage Hybrid Model ICEEMDAN-SE-VMD-RDPG for a Multivariate Solar Irradiance Forecasting. *IEEE Access* **2021**, *9*, 37334–37363. [CrossRef]
34. Xu, Z.; Li, S.; Guo, Y.; Qiao, R. STL-Twinsformer-TTAO: A Short-Term Photovoltaic Power Forecasting Model Based on Seasonal-Trend Interactive Learning. *Renew. Energy* **2026**, *263*, 125462. [CrossRef]
35. Bohat, V.K.; Hashim, F.A.; Batra, H.; Elaziz, M.A. Phototropic Growth Algorithm: A Novel Metaheuristic Inspired from Phototropic Growth of Plants. *Knowl.-Based Syst.* **2025**, *322*, 113548. [CrossRef]
36. Abbas, M.; Che, Y.; Khan, I.U. A Novel Stacked Ensemble Framework with the Kolmogorov-Arnold Network for Short-Term Electric Load Forecasting. *Energy* **2025**, *332*, 137216. [CrossRef]
37. Zhao, X.; Duan, P.; Cao, X.; Xue, Q.; Zhao, B.; Hu, J.; Zhang, C.; Yuan, X. A Probabilistic Load Forecasting Method for Multi-Energy Loads Based on Inflection Point Optimization and Integrated Feature Screening. *Energy* **2025**, *327*, 136391. [CrossRef]
38. Yan, X.; Ta, H.; Lu, J. Dynamic and Static Feature Fusion for Short-Term Photovoltaic Power Forecasting. *Sol. Energy* **2026**, *303*, 114108. [CrossRef]
39. Li, X.; He, T.; Zhu, W.; Xu, Q. HE-DTimeXer: A Short-Term Electricity Price Prediction Model with Exogenous Variables. *Electr. Power Syst. Res.* **2026**, *253*, 112540. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.