

Article

Research on Integrating Physical Constraints with HO-Transformer-KAN for Short-Term Photovoltaic Power Forecasting

Shiyan Gao, Xu Wang *, Ying Zhan, Xiaoxiao Wei, Ye Xu and Wei Li

MOE Key Laboratory of Regional Energy and Environmental Systems Optimization, College of Environmental-Science and Engineering, North China Electric Power University, Beijing 102206, China

* Correspondence: wangxu@ncepu.edu.cn

Abstract

To address the issues of limited interpretability and low predictive accuracy in traditional photovoltaic forecasting models, this paper proposes a hybrid forecasting model named HO-Transformer-KAN-PINN. First, Maximal Information Coefficient (MIC) is used to select the key meteorological features: irradiance and temperature. Then, the grey relational analysis combined with cosine similarity is applied to identify similar days. The prediction framework is then constructed. The Transformer-KAN model provides high predictive accuracy and strong interpretability, while embedding physics-informed neural network (PINN) constraints enforces compliance with the underlying physical laws, yielding the Transformer-KAN-PINN framework. Simultaneously, the Hippopotamus Optimization (HO) algorithm is used to optimize the model hyperparameters. Finally, the photovoltaic power combination prediction model of HO-Transformer-KAN-PINN is constructed. This model has achieved excellent results in short-term photovoltaic power forecasting in Yunnan, Gansu, and Australia. Taking winter in Yunnan Province as an example, the forecasting results of this model yield an MAE of 0.3204 MW, an RMSE of 0.4197 MW, a MAPE of 4.9561%, and an R^2 of 0.9986. Therefore, the hybrid forecasting model proposed in this paper demonstrates a certain degree of advancement and effectiveness. Therefore, it provides reliable technical support for accurate prediction of photovoltaic output.

Keywords: photovoltaic power forecasting; comprehensive similarity; physical constraints; Transformer-KAN; Hippopotamus Optimization algorithm



Academic Editor: Carlo Renno

Received: 9 May 2026

Revised: 25 May 2026

Accepted: 8 June 2026

Published: 29 June 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Traditional energy sources occupy a dominant position in China's energy system and contribute substantially to its economy and social welfare. However, the gradual depletion of fossil fuels is reducing reserves of some conventional energy resources, exacerbating supply–demand imbalances. China's primary energy consumption will continue to grow in the future. It is expected to reach 6.3×10^9 tce (tons of standard coal equivalent) by 2035 and 6.9×10^9 tce by 2060 [1], which conflicts with finite fossil reserves and threatens sustainable development. In this context, constructing a new energy system and vigorously developing renewables are essential responses. Solar energy is widely distributed, clean, low-carbon, and flexible; photovoltaic (PV) generation is therefore poised to expand rapidly and to partially replace traditional power generation. According to data from the National Energy Administration, as of February 2025, the installed capacity of solar

power generation reached 930 GW (9.3×10^8 kW), representing a year-on-year increase of 42.9% [2]. Nevertheless, PV output is highly sensitive to meteorological conditions and exhibits strong intermittency and volatility, which pose operational challenges. Therefore, we must accurately predict photovoltaic output to master the future power generation situation in advance, and then realize scientific dispatching to reduce energy waste and enhance the stability of the grid. This is especially important under the assessment requirements of the “Two Detailed Rules”. In China’s power market operation and grid-connected renewable energy management, the “Two Detailed Rules” generally refer to the detailed rules for power grid operation management and auxiliary service management. These rules include assessment and compensation mechanisms for grid-connected power plants. For PV power stations, inaccurate power forecasting may lead to deviations between scheduled and actual generation, thereby increasing assessment penalties and reducing economic benefits. The decline in prediction accuracy will lead to high assessment costs, which will affect the economic benefits of photovoltaic power plants. Consequently, methods such as similar-day selection, intelligent forecasting algorithms, and hyperparameter optimization have become central research priorities domestically and internationally.

1.1. Similar Day Selection

In photovoltaic power generation prediction, the selection of similar days is an important preprocessing step. The photovoltaic output is closely related to meteorological characteristics. By selecting similar days, it is possible to identify a sample set that is highly consistent with the prediction day, while effectively reducing the sample size, providing an effective data basis for photovoltaic output prediction, and thereby improving the effectiveness of photovoltaic power prediction. Li et al. [3] used the CRITIC weight method to calculate the impact weight of each meteorological element, and then determined the similar day by calculating the weighted Euclidean distance time by time. Yang et al. [4] selected two-dimensional data of irradiance and temperature as similarity variables and, based on two-dimensional Euclidean distance, utilized numerical weather predictions alongside historical daily measured meteorological data to select similar days. Zhang et al. [5] adopted a comprehensive grey relational theory for similar-day selection, considering the degree of correlation among values under various factors, thereby simplifying the required historical data volume and significantly reducing the impact of the randomness of weather factors. Yuan et al. [6] performed meteorological feature similarity assessment through self-supervised learning, followed by dynamic time warping (DTW) for similar-day selection, and then conducted photovoltaic power forecasting. Although the aforementioned methods can effectively select similar days to a certain extent, they calculate similarity from only one perspective, namely distance similarity or shape similarity, making it difficult to capture the complex fluctuation characteristics of photovoltaic power. To address these limitations, this paper combines grey relational analysis (GRA) with the cosine similarity method. This combination avoids the limitations of a single indicator and allows their complementary advantages to be integrated.

1.2. Prediction Model Construction

Based on the screening of similar days, constructing the prediction model with strong generalization ability can effectively improve the accuracy. Physical forecasting methods perform power prediction based on physical relationships using meteorological characteristics and photovoltaic equipment parameters [7]. Statistical methods make predictions based on historical statistical data, leveraging the relationship between a large amount of historical data and photovoltaic power. Statistical methods are widely applied in the field of short-term photovoltaic power prediction [8]. Statistical models include autoregressive moving

average (ARMA), autoregressive integrated moving average (ARIMA), and generalized autoregressive conditional heteroskedasticity (GARCH) [9]. With the vigorous development of artificial intelligence algorithms, scholars both domestically and internationally generally adopt intelligent algorithms as the core framework for building prediction models. Compared with traditional prediction methods, intelligent algorithms have stronger fitting capability and higher data-processing efficiency, thereby achieving better forecasting performance. Bai et al. [10] combined ICEEMDAN with TCN-AM-BiGRU and applied the model to validate it under sunny, cloudy, and rainy weather conditions. Han et al. [11] proposed a combined prediction model of CEEMDAN-VMD-BiGRU, which is input into the BiGRU model for prediction after secondary decomposition. Zhou et al. [12] proposed a novel hybrid forecasting framework, the Attention-DCC-BiLSTM-AR model, based on attention mechanisms and a parallel forecasting architecture, where the parallel framework can model both linear and nonlinear characteristics simultaneously. Zhou et al. [13] constructed a combined forecasting model of CEEMDAN-AM-TCN-BiLSTM for ultra-short-term photovoltaic output prediction and utilized the RIME algorithm for hyperparameter optimization. Liang et al. [14] constructed a hybrid forecasting model combining CEEMDAN, PE, and BiLSTM for short-term photovoltaic power forecasting, and validated the accuracy of the proposed hybrid model. Although the above-mentioned models can improve the accuracy of photovoltaic power forecasting to a certain extent, issues such as overfitting and the need for improved model interpretability remain, which further affect the performance of photovoltaic power prediction. Therefore, this paper adopts the Transformer-KAN model, which is based on the attention mechanism and is adept at capturing long-range dependencies. In the Transformer, the Multi-Layer Perceptron (MLP) layer is replaced by the Kolmogorov–Arnold Network (KAN). Compared with the MLP layer, KAN can approximate functions with fewer parameters [15], thereby improving the model structure and efficiency. The Transformer-KAN model can improve the performance of solving complex tasks and improve the reasoning ability and generalization ability [16].

In the photovoltaic output prediction model, the Transformer-KAN model exhibits excellent performance in terms of prediction accuracy and generalization ability. However, as a black-box model, it still lacks explicit constraints from objective physical laws. Embedding physical mechanisms into the model can help ensure consistency with objective physical laws, compensate for the limitations of purely data-driven models, and further strengthen model generalization ability. Yuan et al. [17] combined deep learning and added physical constraints to the loss function to build a prediction model containing physical model constraints, and verified the accuracy of the model. Fu et al. [18] proposed a prediction model embedded with physical information for predicting fatigue damage in boom systems and conducted verification. Gao et al. [19] introduced a deep learning framework combined with physical constraints for short-term forecasting and enhanced it using signal decomposition.

Currently, the application of physics-informed machine learning in photovoltaic forecasting remains relatively limited. Based on this, this paper introduces physical law-based constraints to regulate model behaviour. This transforms the prediction model from a black-box model into a grey-box model, further enhancing its interpretability. Specifically, the monotonicity constraints can be expressed as objective physical relationships such as “the stronger the solar irradiance, the higher the photovoltaic output” and “the higher the temperature, the lower the photovoltaic output,” with penalties imposed for violations of these physical laws.

By constructing the aforementioned prediction model, although the Transformer model possesses excellent sequence processing capabilities, it may lead to overfitting when the training samples are limited and lack explicit physical interpretability. The application

of the KAN, which replaces the MLP layers, further enhances the interpretability of the model. However, it still relies on sample quality and parameter settings. The incorporation of PINN strengthens the model's interpretability, but still requires appropriate constraint weights. Otherwise, it may further impair the model's fitting capability. Therefore, this paper integrates the above components into a unified framework.

1.3. Hyperparameter Optimization

Hyperparameter optimization is a crucial step in improving the accuracy of the forecasting model. Optimizing and determining the optimal hyperparameter combination for a prediction model can effectively alleviate problems such as poor generalization ability and unsatisfactory fitting performance, thereby better meeting practical application requirements and further improving forecasting accuracy. Ma et al. [20] employed the Gravitational Search Algorithm (GSA) for parameter optimization, thereby enhancing the effectiveness of hyperparameter setting. Wang et al. [21] adopted the Multi-Strategy Improved Sparrow Search Algorithm (MSISSA) to optimize parameters such as the optimal number of hidden-layer nodes, training epochs, and learning rate of the Long Short-Term Memory (LSTM) network, and validated the prediction accuracy of the model. Zhou et al. [22] constructed a CNN-LSTM-attention prediction model to enhance photovoltaic power forecasting, employed Bayesian optimization for hyperparameter tuning, and verified the advantages of their optimized model. Quan et al. [23] adopted a hybrid SATCN-BiLSTM forecasting model and combined it with the Dung Beetle Optimization (DBO) algorithm. Yang et al. [24] focused on short-term photovoltaic output forecasting, employing SGMD for power signal decomposition and using the Grey Wolf Optimizer to optimize the hyperparameters of the BiLSTM model, with case studies validating the applicability of the model. Although the aforementioned optimization algorithms have demonstrated satisfactory performance to a certain extent, they still suffer from certain limitations, such as a tendency to become trapped in local optima. Their global search capability needs further enhancement, and the convergence speed and convergence accuracy still require improvement. In view of the aforementioned issues with these optimization algorithms, this paper adopts the Hippopotamus Optimization Algorithm (HO) to optimize the forecasting model. The HO algorithm simulates the behaviour pattern of hippopotamuses and relies on its excellent global search ability and efficient performance. It can effectively solve the limitations of traditional optimization algorithms.

In summary, this paper employs the Maximal Information Coefficient (MIC) method to select characteristic factors for photovoltaic output, and combines GRA with cosine similarity to form a combined algorithm for similar-day selection. Based on this, the sequence is input into the Transformer-KAN-PINN prediction model, and the Hippopotamus Optimization Algorithm (HO) is employed for hyperparameter optimization. A hybrid forecasting model, denoted as MIC-GRA-Cosine Similarity-HO-Transformer-KAN-PINN, is thereby constructed. The model is validated using data from a photovoltaic power station in Yunnan Province, demonstrating strong forecasting performance.

1.4. Main Contribution

In photovoltaic forecasting research, selecting similar days based solely on distance or shape similarity may lead to suboptimal results due to the limited consideration of single factors. Additionally, physical mechanisms are rarely applied in photovoltaic forecasting, despite their ability to ensure compliance with objective physical laws. On this basis, the selection of model hyperparameters is crucial while also ensuring the rationality of the parameter selection approach. Therefore, this paper addresses the aforementioned issues and achieves the following contributions:

- (i) Develop a physically constrained Transformer-KAN forecasting framework in which data-driven temporal feature extraction and monotonic physical constraints are jointly incorporated, thereby improving both forecasting accuracy and physical consistency.
- (ii) Propose a comprehensive similar-day selection strategy by integrating GRA and cosine similarity, enabling the model to capture both amplitude-related similarity and shape-related similarity of meteorological sequences.
- (iii) Introduce HO-based hyperparameter optimization to jointly optimize the learning rate, hidden dimension, number of attention heads, and physical-constraint weight, reducing the uncertainty caused by manual parameter tuning.

2. Method Overview

2.1. Technical Framework

In this paper, the main prediction process is as follows, the overall technical flowchart is shown in Figure 1.

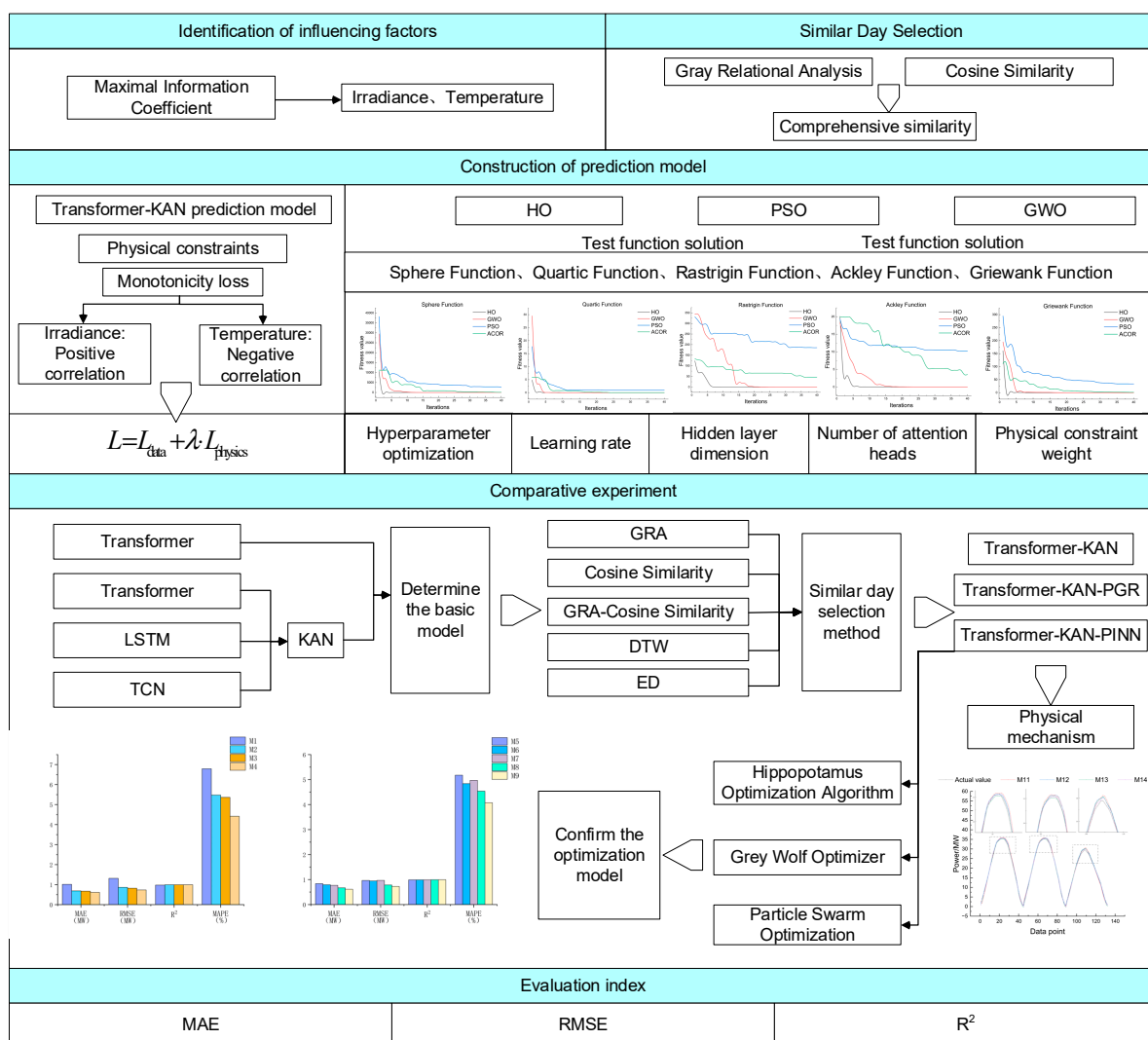


Figure 1. Overall technical flowchart.

- (1) The MIC method is used to analyze the correlation between five meteorological factors—global irradiance, wind speed, wind direction, temperature, and humidity—and short-term photovoltaic output. Two factors with the greatest impact on photovoltaic output, namely global irradiance and temperature, are extracted.

- (2) A similar-day selection method combining grey relational analysis (GRA) and cosine similarity is proposed.
- (3) Physical constraints are embedded into the basic prediction model to enhance its interpretability and forecasting performance.
- (4) The HO algorithm is used for hyperparameter optimization to determine the optimal parameter combination.

2.2. Identification of Key Meteorological Factors

The MIC method is employed to identify important meteorological features, thereby avoiding noise interference caused by redundant features and reducing input dimensionality. The MIC is a method used to measure the correlation between two variables. It can effectively explore both linear and nonlinear relationships between variables [25]. Its value range lies between 0 and 1, where a higher value indicates a stronger correlation:

$$MIC(x, y) = \max_{xy < B(n)} \frac{I(x, y)}{\log_2(\min(x, y))} \quad (1)$$

$$I(x, y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (2)$$

where $I(x, y)$ is the mutual information value between variables x and y ; $p(x, y)$ is the joint probability of variables x and y ; $p(x)$ and $p(y)$ are the marginal probabilities; and $B(n)$ is the upper bound of the grid partition, generally, $B(n) = n^{0.6}$.

2.3. Similar-Day Selection Based on Comprehensive Similarity Combining GRA and Cosine Similarity

This article adopts a comprehensive similarity approach to enhance the selection of similar days, thus avoiding the limitations of a single similarity measure. The specific steps are as follows:

- (1) Calculate the Grey Relational Degree

Calculate the absolute difference between the two sequences, and then compute the correlation coefficient. The calculation formula is as follows:

$$\Delta_{0i}(k) = |X'_0(k) - X'_i(k)| \quad (i = 1, 2, \dots, m; k = 1, 2, \dots, n) \quad (3)$$

$$\varphi_i = \frac{\min_i \Delta_i + \xi \cdot \max_i \Delta_i}{\Delta_i + \xi \cdot \max_i \Delta_i} \quad (4)$$

where $X'_0(k)$, $X'_i(k)$ represents the normalized data sequence and reference sequence; φ_i denotes the grey correlation coefficient; ξ is the distinguishing coefficient.

- (2) Calculate Cosine Similarity

Cosine similarity measures the similarity between two vectors based on the cosine of the angle between them. The closer the value is to 1, the higher the similarity between the two vectors. The formula is as follows:

$$\cos(x, y) = \frac{x \cdot y}{|x| \cdot |y|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (5)$$

where $\cos(x, y)$ denotes the cosine angle between two vectors.

(3) Calculation of Comprehensive Similarity

$$S_i = \alpha \cdot \varphi_i + (1 - \alpha) \cdot \gamma_i \quad (6)$$

where S_i denotes the comprehensive similarity; α denotes the weight coefficient, 0.5; φ_i denotes the grey relational degree; γ_i denotes the cosine similarity.

2.4. Construction of a Hybrid Forecasting Model Based on HO-Transformer-KAN-PINN

In this paper, the Transformer-KAN-PINN combination prediction model is constructed to improve the prediction accuracy. The advantages of this model are as follows: (1) The Transformer model can effectively capture the sequence relationship and has high processing efficiency; (2) the Transformer-Kan model uses a Kan network to replace the MLP layer in the Transformer. The interpretability of the model is further enhanced to achieve accurate prediction of photovoltaic sequences. (3) Embedding PINN, incorporating monotonicity relationships into the model, and ensuring it adheres to objective physical laws.

For the above model, it is still necessary to adjust parameters to meet the complex changes in the prediction scenario. Manual parameter tuning is highly subjective and inefficient, with poor efficiency and stability. Therefore, this paper adopts a novel intelligent optimization algorithm, the HO algorithm, for hyperparameter optimization. This optimization algorithm possesses excellent global search capability and solution efficiency. It optimizes parameters in the Transformer-KAN-PINN model, such as the number of attention heads, physical constraint weight, learning rate, and hidden layer dimension. The HO determines the optimal combination of hyperparameters, thereby enhancing the prediction accuracy of the hybrid forecasting model.

2.4.1. Construction of Forecasting Model Based on Transformer-KAN

The Transformer-KAN model uses a Kolmogorov–Arnold network layer to replace the MLP layer. The expression ability and flexibility of the model are further improved, so as to improve the accuracy of output prediction. The model structure is illustrated in Figure 2.

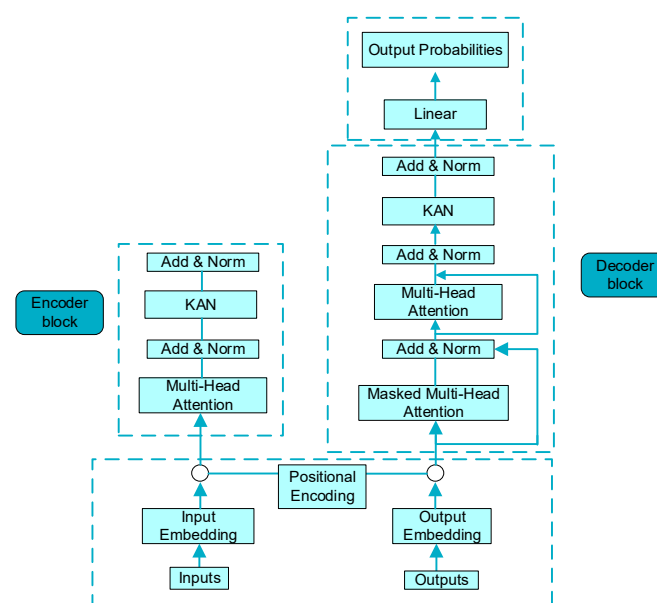


Figure 2. Structure diagram of the Transformer-KAN model.

(1) Transformer Model

The self-attention mechanism introduced in the Transformer can efficiently process sequential data, enabling the model to attend to all elements in the sequence. This helps capture the overall characteristics and establish data relationships [26]. The Transformer consists of an encoder and a decoder. The encoder primarily consists of a multi-head self-attention module and a feed-forward network, incorporating residual connections and layer normalization [27]. The decoder is used to generate the sequence. The model structure is illustrated in Figure 3.

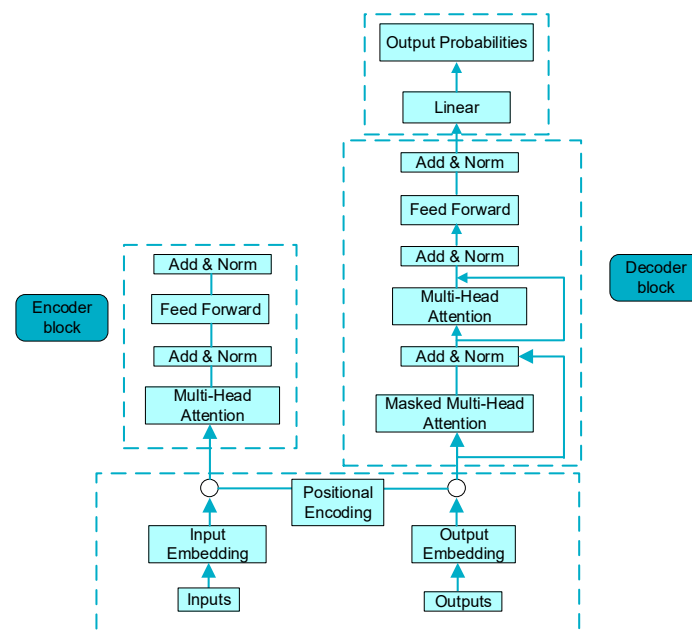


Figure 3. Structure diagram of the Transformer model.

The calculation steps are as follows: the input matrix X obtains the initialization weight matrix through matrix transformation, thereby generating the query vector sequence Q , the key vector sequence K , and the value vector sequence V :

$$Q = W_q X \quad (7)$$

$$K = W_k X \quad (8)$$

$$V = W_v X \quad (9)$$

where W_Q , W_K , and W_V represent the weight matrices.

The self-attention mechanism is a core component of the Transformer model. It takes into account all other elements in the sequence, capturing global characteristics of the sequence. The multi-head attention mechanism is as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (10)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1 \cdots \text{head}_n)W^O \quad (11)$$

$$\text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right), i = 1, 2, \cdots, n \quad (12)$$

where W^O , W^Q , W^K , and W^V denote the learned weight matrices; *Attention* denotes the attention computation function; *MultiHead* denotes the multi-head attention mechanism; d_k

denotes the dimension of the key matrix; *Concat* denotes the concatenation function; *Head_i* denotes the *i*-th sub-attention head; and *n* denotes the number of multi-attention heads.

(2) Kolmogorov–Arnold Network

The Kolmogorov–Arnold Network (KAN) is derived from the Kolmogorov–Arnold Representation Theorem. Under certain conditions, any continuous function of multiple dimensions can be decomposed into a combination of multiple one-dimensional functions [28]. The KAN is capable of learning complex relationships and nonlinear patterns in input data, enabling more accurate prediction results. The KAN can efficiently fit complex functions [29]. Its model structure is illustrated in Figure 4.

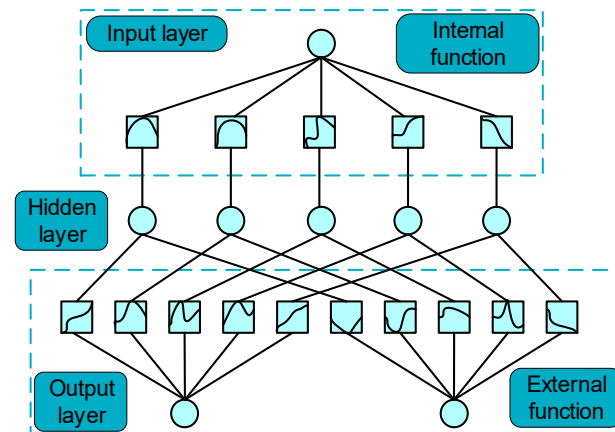


Figure 4. KAN structure diagram.

Its expression is:

$$f(x) = f(x_1, \dots, x_n) = \sum_{p=1}^{2n+1} \varphi_p \left(\sum_{q=1}^n \phi_{p,q}(x_q) \right) \quad (13)$$

where inner univariate function $\phi_{p,q} : [0, 1] \rightarrow \mathbb{R}$, and outer function $\varphi_p : \mathbb{R} \rightarrow \mathbb{R}$.

Residual Activation Function:

$$\phi(x) = w_b b(x) + w_s \text{spline}(x) \quad (14)$$

$$b(x) = \text{silu}(x) = x / (1 + e^{-x}) \quad (15)$$

where w_b , w_s , and c_n are learnable parameters; $\text{silu}(x)$ is the activation function; $\text{spline}(x)$ is a combination of univariate functions.

The output matrix of the *l*-th layer can be expressed as:

$$\mathbf{x}^{(l+1)} = \begin{bmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,2,2}(\cdot) & \cdots & \phi_{l,2,n_l}(\cdot) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \cdots & \phi_{l,n_{l+1},n_l}(\cdot) \end{bmatrix} \mathbf{x}^{(l)} \quad (16)$$

This paper adopts the Transformer-KAN model. The model architecture is as follows: the embedding dimension is 64, the dropout rate is 0.2, the Adam optimizer is used, the learning rate is set to 0.0005, the batch size is 64, B-spline activation is used inside the KAN layer, and the number of training epochs is 80.

2.4.2. Physics-Informed Neural Networks (PINN)

Physics-Informed Neural Networks (PINN) can combine neural networks with physical knowledge to enhance the generalization ability of models [30]. PINN consists of an input layer, hidden layers, and an output layer. Incorporating physical constraints into the loss function enables data fitting and complies with objective physical constraints. That is, the constraints are: “the stronger the solar irradiance, the higher the photovoltaic output” and “the higher the temperature, the lower the photovoltaic output”, and penalties are imposed for violations. This achieves the integration of data and physical constraints. Its structure diagram is illustrated in Figure 5.

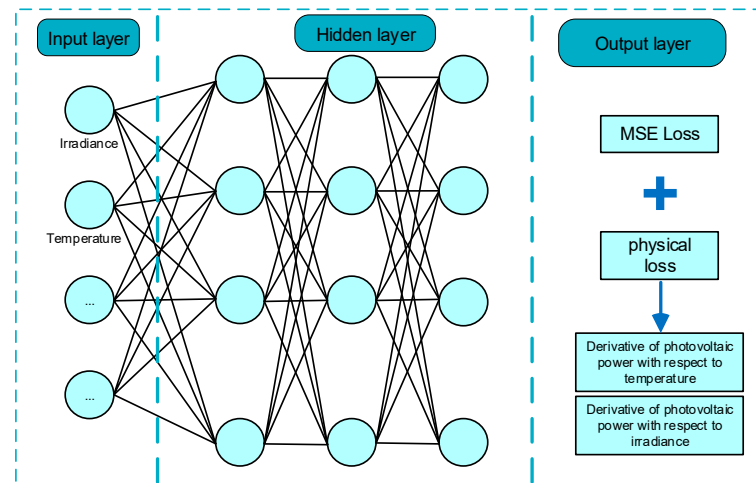


Figure 5. Structure diagram of physical information neural network.

The loss function is composed of the MSE loss and the physical constraint loss. The total loss function is:

$$L = L_{\text{data}} + \lambda \cdot L_{\text{physics}} \quad (17)$$

where L_{data} is the MSE loss; L_{physics} is the physical constraint loss; λ is the coefficient; the final model employs the HO algorithm for hyperparameter optimization.

The MSE Loss Function is as follows:

$$L_{\text{data}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (18)$$

The Physics Loss Function is as follows:

$$L_{\text{physics}} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \max \left(0, -m_j \cdot \frac{\partial \hat{y}_i}{\partial x_{i,j}} \right) \quad (19)$$

where M is the number of features involved in the physical constraints; $\frac{\partial \hat{y}_i}{\partial x_{i,j}}$ denotes the partial derivative of the output of the i -th sample with respect to the j -th input feature; m_j indicates the direction of monotonicity, where $m_j = 1$ represents a positive correlation, and $m_j = -1$ represents a negative correlation; $\max(0, -)$ denotes the ReLU function, which penalizes violations of the constraints.

2.4.3. Hippopotamus Optimization Algorithm (HO)

The HO algorithm was proposed by Mohammad Hussein Amiri et al. in February 2024. This algorithm is inspired by the behavioural patterns of hippopotamuses. Its three-

phase model encompasses hippopotamus position updates, defence strategies, and evasion methods against predators. The algorithm demonstrates strong global and local search capabilities, enabling efficient identification of optimal solutions. This paper employs the HO algorithm to optimize the hyperparameters in the forecasting model, including the learning rate, number of attention heads, hidden layer dimension, and physical constraint weight. These hyperparameters are treated as search variables for the HO. This is performed through location updates, seeking the optimal parameter combination. The principle of the HO algorithm is as follows:

(1) Population Initialization

Hippopotamus individuals are represented by vectors, as shown in the following formula:

$$X_i : x_{ij} = lb_j + r \cdot (ub_j - lb_j), (i = 1, 2, \dots, N; j = 1, 2, \dots, m) \quad (20)$$

where X_i represents the position information of a hippopotamus; x_{ij} denotes the position information of the i -th hippopotamus under the j -th decision variable; r is a random number in the range $[0, 1]$; ub_j and lb_j represent the lower bound and upper bound of the j -th decision variable, respectively; N denotes the number of hippopotamuses in the population; m denotes the number of decision variables.

(2) Phase 1

Hippopotamus individuals move closer to one another, while the dominant hippopotamus protects the population and its territory.

Position information of male hippopotamuses in the population:

$$x_{M,ij} = x_{ij} + y_1 \cdot (D_{hippo} - I_1 x_{ij}) \quad (21)$$

where $x_{M,ij}$ represents the position information of the male hippopotamus; D_{hippo} denotes the position information of the dominant hippopotamus; y_1 is a random number between 0 and 1; I_1 equals 1 or 2.

Position information of female or immature hippopotamuses in the population:

$$x_{ij}^{FBhippo} = \begin{cases} x_{ij} + z_1 \cdot (D_{hippo} - I_2 M_j), & H > 0.6 \\ x_j + z_2 \cdot (M_j - D_{hippo}), & H \leq 0.6 \text{ and } \theta_1 > 0.5 \\ lb_j + \theta_2 \cdot (ub_j - lb_j), & H \leq 0.6 \text{ and } \theta_1 \leq 0.5 \end{cases} \quad (22)$$

where H denotes the selection probability; M_j represents the mean of the randomly selected hippopotamuses in the population; I_2 means 1 or 2; z_1 and z_2 represent random vectors or random numbers; θ_1 and θ_2 indicates a random number belonging to 0–1.

(3) Phase 2

Hippopotamuses defend themselves against predators, primarily by emitting loud calls to deter predators from approaching. The formula for the predator's position is as follows:

$$\text{Predator}_j = lb_j + \delta \cdot (ub_j - lb_j), j = 1, 2, \dots, m. \quad (23)$$

$$L = |\text{Predator}_j - x_{ij}| \quad (24)$$

where δ represents a random vector between 0 and 1; L denotes the distance between the hippopotamus and the predator.

When a predator approaches a hippopotamus, the hippopotamus will defend itself against the predator. The specific expression is as follows:

$$x_{\text{HippoR},ij} = \begin{cases} L_j \cdot \text{Predator}_j + \frac{f}{c-d \times \cos(2\pi g)} \cdot \frac{1}{D_j}, & F_{\text{predator},j} < F_i \\ L_j \cdot \text{Predator}_j + \frac{f}{c-d \times \cos(2\pi g)} \cdot \frac{1}{2 \times D_j + \omega_j}, & F_{\text{predator},j} \geq F_i \end{cases} \quad (25)$$

where $x_{\text{HippoR},ij}$ is the position of the hippopotamus when facing the predator; L_j denotes the j -th element of the Levy random vector; f is a random number, in 2–4; c is a random number, in 1–1.5; d is a random number, in 2–3; g is a random number, in -1 – 1 ; ω_j denotes a random vector, dimension $1 \times m$.

(4) Phase 3

When hippopotamuses are unable to defend against predators, they will leave the current area. The position of the hippopotamuses is as follows:

$$x_{\text{Hippo},ij} = x_{ij} + \eta \cdot [lb_j + \xi \cdot (ub_j - lb_j)], (i = 1, 2, \dots, N, j = 1, 2, \dots, m) \quad (26)$$

where $x_{\text{Hippo},ij}$ indicates the nearest safe location found by the hippopotamuses; η represents a random number between 0 and 1; ξ denotes a random vector or random number.

2.5. Evaluation Metric System

This paper adopts the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE). Mean Absolute Error (MAE) is the average of the “absolute difference between the actual value and the predicted value” of the sample; the Root Mean Squared Error (RMSE) indicates the absolute difference between the actual values and the predicted values; the Coefficient of Determination (R^2) characterizes the goodness of fit of the model. The mathematical formulas are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (27)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (28)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (29)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (30)$$

where n is the sample size of the forecast; y_i is the actual value of photovoltaic power; $\hat{y}_i$ is the predicted value of photovoltaic power.

For the MAE, the RMSE, and the MAPE, the smaller the value, the better the forecasting performance. $R^2 \in [0, 1]$; the closer the value is to 1, the better the forecasting performance.

3. Experimental Analysis

3.1. Introduction to the Dataset

This paper takes a photovoltaic power plant in Yunnan Province as the research subject and validates the aforementioned models, including the identification of key meteorological factors, the selection of similar days based on comprehensive similarity, the construction of a Transformer-KAN prediction model incorporating physical constraints, and hyperparam-

eter optimization using the HO algorithm. This paper conducts research on the full-year data of 2023. The dataset contains a total of 96 data points per day at a time interval of 15 min. This paper removes the relevant nighttime data and studies the photovoltaic data during the time period from 8:15 to 19:00, with 44 data points per day, including relevant parameters such as irradiance, wind speed, wind direction, temperature, humidity, and photovoltaic power. The specific situation is shown in Table 1. Figure 6 illustrates the average values of irradiance, temperature, and photovoltaic power for each month within the dataset.

Table 1. Dataset situation.

Variable	Unit	Mean	Std.	Min	Max
Irradiance	W/m ²	604.51	240.84	87	1053
Temperature	°C	23.99	3.72	10.6	31.4
Humidity	%	36.06	14.15	12.6	80.5
Wind speed	m/s	3.89	1.57	0.1	6.9
Wind direction	°	238.36	27.08	83	280
PV power	MW	20.88	8.31	3.01	36.21

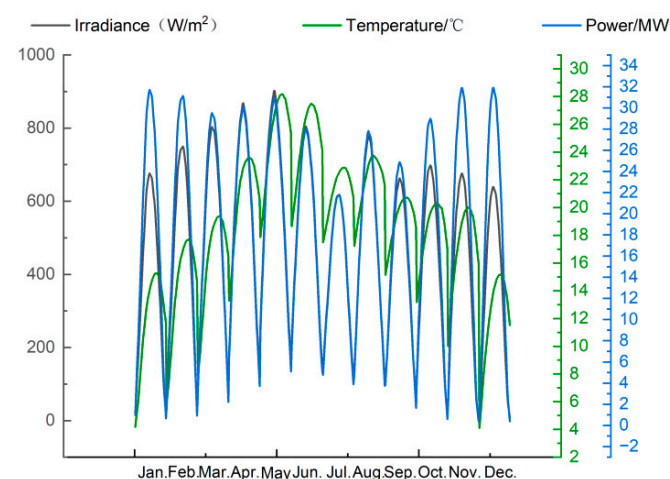


Figure 6. Average values of irradiance, temperature, and photovoltaic power for each month.

3.2. Key Meteorological Factor Identification and Similar-Day Selection Based on Comprehensive Similarity

This paper employs the MIC method to conduct a correlation analysis of various meteorological factors. The MIC values are calculated separately for the four seasons in the dataset, and then averaged to obtain the final MIC value. The MIC value results are shown in Table 2 and Figure 7. In this paper, the top two features with relatively high MIC values are selected to prevent problems such as excessive feature dimensionality and computational burden. Consequently, irradiance and temperature are ultimately chosen as the primary influencing factors for photovoltaic power forecasting.

As shown in Table 3, when the meteorological features of irradiance and temperature were selected, a good result was achieved, with an MAPE of 3.7528%. Adding the humidity feature on this basis, the MAPE decreased by 0.2557%, which was not a significant improvement. Subsequently, more meteorological features were added, but the prediction accuracy decreased. Therefore, through the above research, given the relative complexity of the model, considering the time required for feature improvement and the prediction

accuracy comprehensively, irradiance and temperature were ultimately regarded as the important influencing factors.

Table 2. MIC values of various meteorological factors and photovoltaic power generation output.

(1) Spring	
Meteorological Factors	MIC
Irradiance	0.6557
Temperature	0.2485
Humidity	0.1740
Wind direction	0.1461
Wind speed	0.1527
(2) Summer	
Meteorological Factors	MIC
Irradiance	0.6811
Temperature	0.2331
Humidity	0.2108
Wind direction	0.2077
Wind speed	0.1357
(3) Autumn	
Meteorological Factors	MIC
Irradiance	0.6147
Temperature	0.2071
Humidity	0.1936
Wind direction	0.1940
Wind speed	0.1418
(4) Winter	
Meteorological Factors	MIC
Irradiance	0.6351
Temperature	0.2383
Humidity	0.1837
Wind direction	0.1478
Wind speed	0.1639
(5) Average value calculation	
Meteorological Factors	MIC
Irradiance	0.6467
Temperature	0.2318
Humidity	0.1905
Wind direction	0.1739
Wind speed	0.1485

Table 3. Prediction performance of meteorological feature combinations.

Meteorological Features	MAPE
Irradiance and Temperature	3.7528%
Irradiance, Temperature and Humidity	3.4971%
Irradiance, Temperature, Humidity and Wind direction	4.6118%
Irradiance, Temperature, Humidity, Wind direction and Wind speed	5.0743%

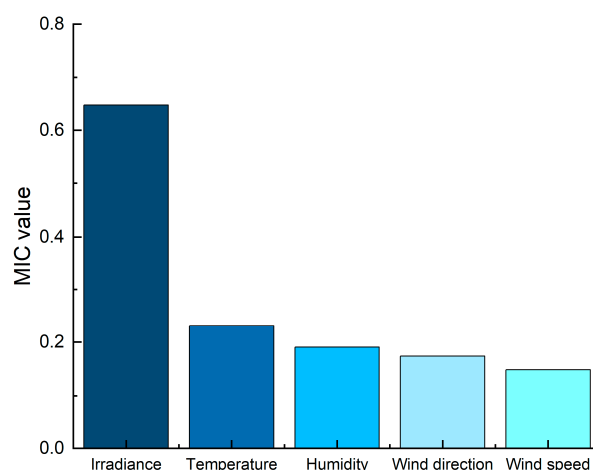


Figure 7. MIC values of various meteorological factors and photovoltaic power.

For the selection of similar days, three days are chosen from each of the four seasons as the forecast days. Specifically, 13–15 April are selected as the forecast days for spring; 15–17 August for summer; 16–18 November for autumn; and 16–18 January for winter. The comprehensive similarity method combining GRA and cosine similarity is employed to select similar days. Based on the calculated similarity and a comprehensive consideration of prediction efficiency, it is ultimately determined that 7 days are selected as the days to be predicted for each similar day. As shown in Table 4, the selected similar-day sets are presented.

Due to the strong correlation between photovoltaic power and irradiance, taking autumn as an example, Figure 8 shows the comparison of photovoltaic power and irradiance between the day to be predicted and the similar days obtained using the comprehensive similarity selection method, and they are highly consistent, proving the effectiveness of this method.

Table 4. Results of selecting similar days based on grey correlation degree—cosine similarity for comprehensive similarity.

Season	Forecast Day	Similar Day
Spring	13 April 2023	27 March 2023, 30 March 2023, 31 March 2023, 9 April 2023, 12 April 2023, 19 April 2023, 24 April 2023
	14 April 2023	10 April 2023, 16 April 2023, 22 April 2023, 25 April 2023, 29 April 2023, 16 September 2023, 17 September 2023
	15 April 2023	23 March 2023, 9 April 2023, 12 April 2023, 19 April 2023, 24 April 2023, 19 July 2023, 13 August 2023
Summer	15 August 2023	8 March 2023, 20 March 2023, 7 April 2023, 8 April 2023, 2 August 2023, 8 August 2023, 26 August 2023
	16 August 2023	18 April 2023, 20 April 2023, 6 May 2023, 12 May 2023, 6 June 2023, 18 July 2023, 25 August 2023
	17 August 2023	18 April 2023, 20 April 2023, 12 May 2023, 20 May 2023, 23 May 2023, 18 July 2023, 25 August 2023
Autumn	16 November 2023	9 November 2023, 15 November 2023, 19 November 2023, 21 November 2023, 22 November 2023, 23 November 2023, 29 November 2023
	17 November 2023	9 November 2023, 15 November 2023, 21 November 2023, 22 November 2023, 23 November 2023, 28 November 2023, 29 November 2023
	18 November 2023	9 November 2023, 12 November 2023, 15 November 2023, 19 November 2023, 21 November 2023, 26 November 2023, 2 December 2023

Table 4. Cont.

Season	Forecast Day	Similar Day
Winter	16 January 2023	12 January 2023, 13 January 2023, 19 January 2023, 22 January 2023, 23 January 2023, 24 January 2023, 27 January 2023
	17 January 2023	12 January 2023, 13 January 2023, 19 January 2023, 22 January 2023, 23 January 2023, 24 January 2023, 26 January 2023
	18 January 2023	3 January 2023, 4 January 2023, 9 January 2023, 15 January 2023, 21 January 2023, 25 February 2023, 9 September 2023

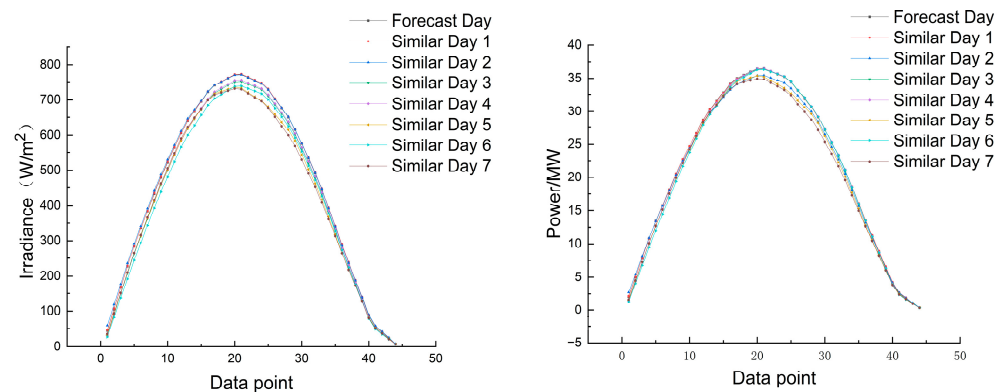


Figure 8. Comparison of irradiance and power between similar days and the day to be predicted.

3.3. Optimization Algorithm Optimization Test

The selection of an optimization algorithm is of critical importance in hyperparameter optimization. This paper conducts optimization tests on relevant optimization algorithms, with the set of test functions covering both unimodal and multimodal problem scenarios. The specific functions and their expressions are shown in Table 5. To compare the HO with the Particle Swarm Optimization (PSO) algorithm, the Grey Wolf Optimizer (GWO), and the Ant Colony Optimization for continuous domains (ACOR), this paper selects the functions in Table 5 as the test functions for evaluating the optimization algorithms.

Table 5. Optimization algorithm optimization test function.

Function Name	Function Expression	Search Range	Optimum
Sphere Function	$f(x) = \sum_{i=1}^n x_i^2$	$[-100, 100]$	0
Quartic Function	$f(x) = \sum_{i=1}^n ix_i^4 + \text{random}[0, 1)$	$[-1.28, 1.28]$	0
Rastrigin Function	$f(x) = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10]$	$[-5.12, 5.12]$	0
Ackley Function	$f(x) = -20 \exp\left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e$	$[-32, 32]$	0
Griewank Function	$f(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	$[-600, 600]$	0

As shown in Figure 9, the convergence behaviour of each optimization algorithm on different test functions is presented. All optimization algorithms are iterated for 1000 iterations, with only the first 40 iterations displayed in the figure.

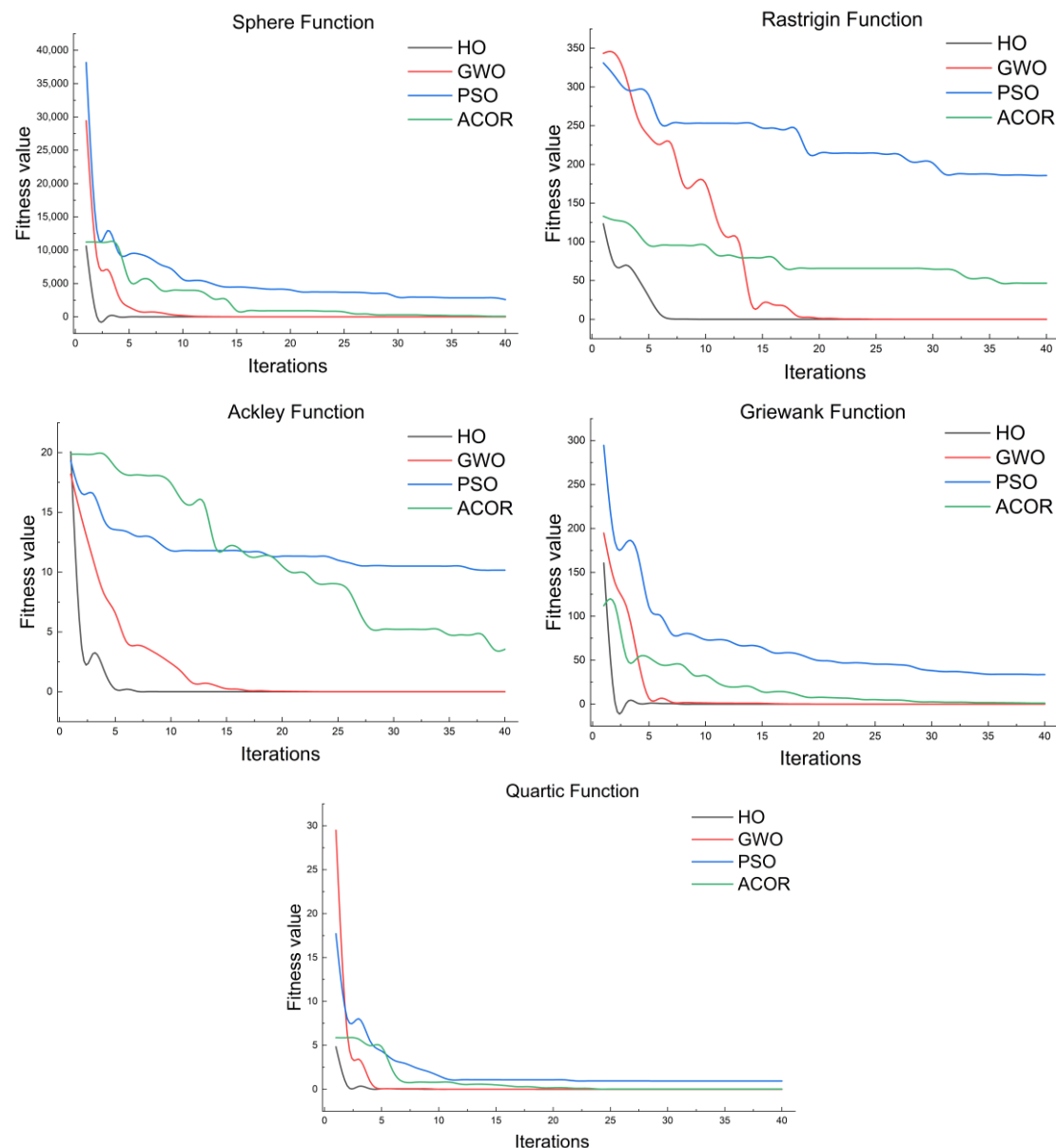


Figure 9. Convergence curve of the test function.

The HO algorithm has demonstrated excellent performance in both convergence accuracy and speed through testing. Taking the Ackley function as an example, the HO algorithm achieves a fitness value below 1 at the 6th iteration, reduces the fitness value to below 0.1 at the 8th iteration, and converges to 0 at the 29th iteration, exhibiting rapid convergence. Meanwhile, the HO algorithm achieves a lower fitness value even in the initial iteration. Taking the quartic function as an example, in the first iteration, its fitness value is 4.84, which is lower than those of the other optimization algorithms. By observing the curves in the figure, we note that during the first five iterations, the convergence effect of the HO algorithm is evident, and its performance is significantly higher than that of other models with a small number of iterations. Among other test functions, the HO algorithm shows lower initial fitness values and faster convergence during the optimization process. Therefore, the HO algorithm possesses superior optimization performance and convergence capability.

4. Case Study

To validate the photovoltaic power forecasting capability of the HO-Transformer-KAN-PINN model proposed in this paper, similar days are selected, and datasets are formed for the four seasons to conduct case analysis. This paper constructs and compares the following models: M1 (Transformer), M2 (LSTM-KAN), M3 (TCN-KAN), M4 (Transformer-KAN), M5 (Cosine Similarity-Transformer-KAN), M6 (DTW-Transformer-KAN), M7 (ED-Transformer-KAN), M8 (GRA-Transformer-KAN), M9 (Comprehensive Similarity-Transformer-KAN), M10 (Comprehensive Similarity-Transformer-KAN-PGR), M11 (Comprehensive Similarity-Transformer-KAN-PINN), M12 (Comprehensive Similarity-PSO-Transformer-KAN-PINN), M13 (Comprehensive Similarity-GWO-Transformer-KAN-PINN), and M14 (Comprehensive Similarity-HO-Transformer-KAN-PINN), among other single and hybrid models for photovoltaic power forecasting. During the prediction process, the computer hardware configuration adopted is as follows: the processor is an Intel Core i7, the installed RAM is 16 GB, and the graphics card is an NVIDIA GeForce GTX 1650. As shown above, through continuous optimization of each model, the performance of photovoltaic power forecasting is progressively improved.

4.1. Comparison of Power Forecasting Results of Baseline Models

To verify the prediction effect of the basic model, the prediction models M1, M2, M3, and M4 are compared. Taking spring as an example, Figure 10 presents the relevant evaluation metrics of each forecasting model.

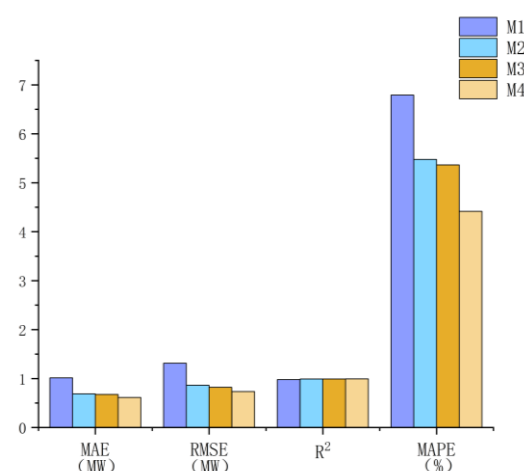


Figure 10. Prediction accuracy of each model.

As shown in Figure 10, the Transformer-KAN model achieves the best performance. Compared with the TCN-KAN and LSTM-KAN models, its MAE is reduced by 0.0623 MW and 0.0754 MW, respectively; its RMSE is reduced by 0.0882 MW and 0.1297 MW, respectively; its MAPE is reduced by 0.9471% and 1.0613, respectively; and its R^2 increases to 0.9931, demonstrating that this model exhibits superior forecasting performance. The substitution of KAN has significantly improved the prediction performance. MAE and RMSE decrease by 0.4001 MW and 0.5794 MW, respectively; MAPE decreases by 2.3754%, and R^2 increases by 0.0154, proving that KAN replacement played an important role. To sum up, the Transformer-KAN model has better fitting ability and prediction performance, which lays the foundation for subsequent model optimization.

4.2. Comparison of Power Forecasting Results Based on Similar-Day Selection Using Comprehensive Similarity

To verify the impact of similar-day selection methods, this paper employs the GRA, cosine similarity, Euclidean distance, DTW, and their combined similarity for selecting

similar days, and conducts a comparison. Taking summer as an example, the evaluation metrics of each forecasting model are detailed in Table 6 and Figure 11.

Table 6. Prediction accuracy of each model.

Forecasting Model	MAE (MW)	RMSE (MW)	R ²	MAPE (%)
M5	0.8372	0.9634	0.9872	5.1763
M6	0.7925	0.9427	0.9877	4.8402
M7	0.7693	0.9662	0.9871	4.9692
M8	0.6692	0.7841	0.9915	4.5381
M9	0.6135	0.7284	0.9927	4.0741

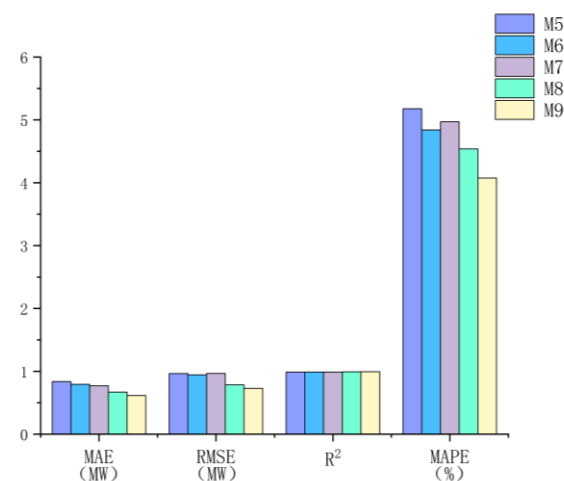


Figure 11. Prediction accuracy of each model.

As shown in the table, the use of comprehensive similarity yields better results. Compared with Euclidean distance, when the comprehensive similarity method is adopted for selection and prediction, the MAE of the model is reduced by 0.1558 MW, the RMSE is reduced by 0.2378 MW, the MAPE is reduced by 0.8951%, and the R² is increased by 0.0056. Compared with DTW, when the comprehensive similarity method is adopted, the MAE of the model is reduced by 0.1790 MW, the RMSE is reduced by 0.2143 MW, the MAPE is reduced by 0.7661%, and the R² is increased by 0.0050. Using the comprehensive similarity method for selection and making predictions, the MAE of the model decreased by 0.0557 and 0.2237, respectively, the RMSE decreased by 0.0557 and 0.2350, respectively, the MAPE decreased by 0.4640% and 1.1022%, respectively, and the R² increased by 0.0012 and 0.0055, respectively. It can be seen that the comprehensive similarity method provides a better dataset for subsequent prediction.

4.3. Comparison of Power Forecasting Results with Incorporation of Physical Mechanisms

To verify the influence of physical mechanisms, this paper adds physical constraints to the Transformer-KAN model. The forecasting results of each model are presented in Table 7. Taking autumn as an example, by adding physical constraints, the MAE and RMSE of the model decrease, while R² increases. Compared with the Transformer-KAN model, the Transformer-KAN-PINN model reduces MAE and RMSE by 0.0474 MW and 0.0176 MW, respectively, reduces MAPE by 3.8853%, and increases R² by 0.0015. Meanwhile, for further validation, this paper combines physical regularization techniques with the prediction model. Compared with the Transformer-KAN-PGR model, the Transformer-KAN-PINN model achieves a lower MAE and MAPE by 0.0434 MW and 1.7406%, respectively, thus further validating the superiority of PINN. Therefore, by imposing constraints to ensure

compliance with objective physical laws, the forecasting performance of the model is further improved.

Table 7. Prediction accuracy of each model.

Forecasting Model	MAE (MW)	RMSE (MW)	R ²	MAPE (%)
M9	0.7206	0.8186	0.9951	12.7610
M10	0.5812	0.6927	0.9965	10.1594
M11	0.5378	0.6749	0.9967	8.4188

4.4. Comparison of Power Forecasting Results Using Optimization Algorithms

To further verify the impact of hyperparameter optimization and optimization algorithms, we compared the performance of the Transformer-KAN-PINN prediction model combined with three different optimization algorithms: the HO, GWO, and PSO. Taking winter as an example, Table 8 shows the prediction accuracy of each model, and Figure 12 shows the comparison between the predicted and actual values of each model.

Table 8. Prediction accuracy of each model.

Forecasting Model	MAE (MW)	RMSE (MW)	R ²	MAPE (%)
M11	0.5764	0.7257	0.9957	7.6900
M12	0.5142	0.6020	0.9970	5.6269
M13	0.3996	0.4924	0.9980	7.3123
M14	0.3204	0.4197	0.9986	4.9561

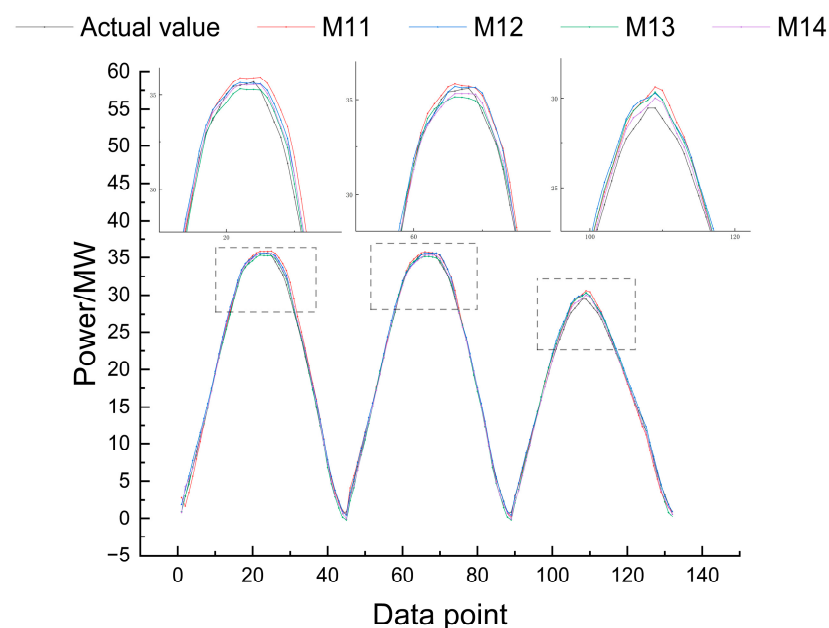


Figure 12. Comparison of predicted values and actual values by various prediction models.

As shown in Table 8 and Figure 12, after hyperparameter optimization, the forecasting performance of each optimized model is improved. Among them, the model optimized with the HO algorithm achieves the best forecasting performance. Compared with the models combined with the other two optimization algorithms, its MAE is reduced by 0.0792 MW and 0.1938 MW, respectively; its RMSE is reduced by 0.0727 MW and 0.1823 MW, respectively; its MAPE is reduced by 2.3562% and 0.6708%, respectively; and its R² is increased by 0.0006 and 0.0016, respectively. To further verify that the differences among M12, M13, and M14 are statistically significant, this paper employs the Wilcoxon test. The

resulting p -value is less than 0.001, which is below the commonly used significance level of 0.05, thereby demonstrating that the improvement in model prediction performance is non-random.

As shown in the figure, the HO-Transformer-KAN-PINN model exhibits an excellent fitting performance. At the same time, in the peak power range, as shown in the enlarged view in Figure 12, the prediction effect of the HO-Transformer KAN-PINN model is better than that of other models. The fitting degree of the first and second forecast days is good. For the third forecast day, the predicted value of each prediction model is slightly higher than the actual value, but the curve obtained using the HO algorithm is closer to the actual value. In conclusion, the model has high prediction accuracy, which proves the superiority of the HO algorithm.

5. Discussion

To further validate the applicability of the model proposed in this paper, another dataset is used for verification. This paper selects a dataset from a photovoltaic power station in Gansu Province and conducts research on the time period of 8:15–19:00. The data interval is still 15 min, and there are still 44 data points per day. Figure 13 shows the prediction results.

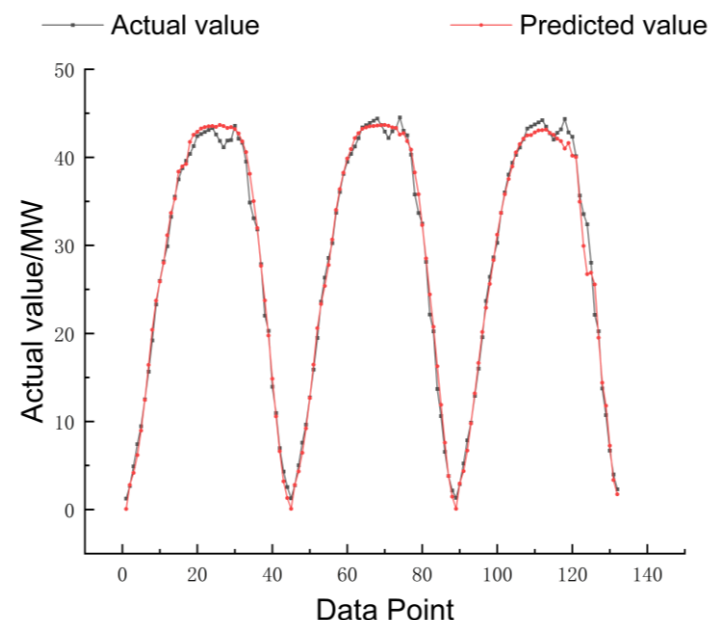


Figure 13. Comparison of predicted values and actual values by prediction models.

The MAE, RMSE, R^2 , and MAPE of the HO-Transformer-KAN-PINN model reached 0.8518 MW, 1.1939 MW, 0.9933, and 6.7957%, respectively, showing a good prediction effect. To further intuitively show the prediction, as shown in Figure 13, the curve fitting effect is good. Meanwhile, validation was conducted on the Australian photovoltaic dataset, where the model achieved MAE, RMSE, R^2 , and MAPE of 0.0634 kW, 0.0824 kW, 0.9908, and 7.0859%, respectively. By predicting different datasets, the excellent prediction ability of the model is further verified.

6. Conclusions

This paper conducts research on photovoltaic power forecasting. First, the Maximal Information Coefficient (MIC) method is employed to identify key factors among various meteorological factors. Subsequently, a comprehensive similar-day selection method combining grey relational analysis (GRA) and cosine similarity is adopted to select similar

days. After the selection of similar days, a forecasting model framework is constructed to perform photovoltaic power forecasting. Finally, to further improve the accuracy of the forecasting model, the Hippopotamus Optimization Algorithm (HO) is used to optimize the hyperparameters within the forecasting model. The following conclusions are drawn:

(1) By employing the Maximal Information Coefficient (MIC) method to identify key meteorological factors, the meteorological features with high correlation to photovoltaic power—namely irradiance and temperature—are selected. The MIC values are 0.6467 and 0.2318, respectively. This accurately identifies the important meteorological characteristics, laying the preliminary foundation for subsequent power forecasting.

(2) The comprehensive similar-day selection method combining grey relational analysis (GRA) and cosine similarity is adopted to select highly consistent similar days for the day to be forecasted, thereby establishing a data foundation for the subsequent power forecasting.

(3) In the forecasting model, the Transformer-KAN architecture is adopted, where the KAN replaces the MLP layer in the traditional Transformer to enhance the interpretability of the model. The replacement with the KAN network leads to a significant improvement in prediction performance. The MAE, RMSE, and MAPE decrease by 0.4001 MW, 0.5794 MW, and 2.3754%, respectively, and R^2 increases by 0.0154. Meanwhile, physical constraints are embedded into this model by imposing monotonicity constraints on irradiance and temperature, ensuring compliance with objective physical laws. The incorporation of physical mechanisms into the prediction model enables the model to achieve satisfactory performance. The MAE, RMSE, R^2 , and MAPE are 0.6191 MW, 0.7641 MW, 0.9957, and 7.2725%, respectively. Finally, the Hippopotamus Optimization Algorithm is employed for hyperparameter optimization. After hyperparameter optimization, the prediction model achieves MAE, RMSE, R^2 , and MAPE of 0.3204 MW, 0.4197 MW, 0.9986, and 4.9561%, respectively, indicating good prediction performance, thus demonstrating the optimization effectiveness of the HO Algorithm. Experimental results demonstrate that through continuous model optimization, the forecasting performance is significantly improved, proving that the hybrid forecasting model proposed in this paper exhibits excellent stability and prediction accuracy, and possesses a certain degree of advancement and innovation. To further verify the applicability of the proposed prediction model in this paper, in addition to Yunnan, validation is also conducted in Gansu and Australia, and it achieved a relatively good prediction result. The proposed prediction model possesses certain applicability and provides a technical pathway for photovoltaic prediction.

7. Outlook and Improvements

(1) In terms of training time, this paper compares the time required for models such as M4, M11, and M14. The respective times are approximately 1 min, 0.3 h, and 9 h. Therefore, through comparison, it is found that the required time gradually increases, especially the time required for hyperparameter optimization increases significantly. However, through accuracy comparison, it is found that the accuracy also gradually improves, with the highest prediction accuracy achieved after hyperparameter optimization. However, the inference time remains acceptable for short-term PV forecasting because the HO-based hyperparameter optimization is performed offline. Therefore, the required time is acceptable. At the same time, the model achieves good prediction performance, and the prediction efficiency can be improved by appropriately reducing the number of training epochs.

(2) In the selection method of similar days, based on the comparison models presented in this paper, additional methods such as Pearson correlation, Spearman correlation, and clustering-based methods can be further incorporated to conduct comparative studies.

(3) In subsequent research, in order to further validate the model presented in this paper, it will be compared with novel prediction models.

(4) In subsequent research, more datasets can be utilized for verification, thereby further validating the applicability of the model.

Author Contributions: S.G.: Writing—original draft, data curation, formal analysis, resources, investigation, and methodology. X.W. (Xu Wang): Validation, conceptualization, and supervision. Y.Z.: Visualization. X.W. (Xiaoxiao Wei): Visualization. Y.X.: Writing—review and editing. W.L.: Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Fundamental Research Funds for the Central Universities (2025MS050).

Data Availability Statement: The data that has been used is confidential.

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

References

1. Jia, A.L.; Wang, G.T.; Li, Y.L. Natural gas development in China: Present situation and prospect. *Nat. Gas Ind.* **2025**, *45*, 31–42.
2. Ding, Y.T. National cumulative installed power generation capacity exceeds 3.4 billion kilowatts. *People's Daily*, 21 March 2025; p. 2.
3. Li, C.R.; Pan, P.C.; Yang, W.R.; Xu, H.S.; Wei, Y.W. Research on PV system power prediction based on improved similar day and HBA-BiLSTM-KELM neural network. *Acta Energ. Sol. Sin.* **2024**, *45*, 508–516.
4. Yang, X.Y.; Wang, S.C.; Zhang, Y.F.; Peng, Y.; Ma, J.C. Short-term PV power prediction based on Grey-Markov and BP_Adaboost by similar days. *Chin. J. Power Sources* **2023**, *47*, 790–794.
5. Zhang, C.; Lin, G.Q.; Huang, J.; Kuang, Y.; Liu, J.J. Short-term photovoltaic power prediction based on AMBOA-DBN combined with similar days. *Acta Energ. Sol. Sin.* **2023**, *44*, 290–299.
6. Yuan, L.; Wang, X.T.; Sun, Y.; Liu, X.B.; Dong, Z.Y. Multistep photovoltaic power forecasting based on multi-timescale fluctuation aggregation attention mechanism and contrastive learning. *Int. J. Electr. Power Energy Syst.* **2025**, *164*, 110389.
7. Wang, R.; Li, Z.; Lu, J. Short-term photovoltaic power forecasting based on fusion clustering and BKA-VMD-TCN-BiLSTM. *J. Univ. Electron. Sci. Technol. China* **2025**, *54*, 592–603.
8. Jiang, J.D.; Chang, Y.Z.; Xu, C.; Guo, J.Q.; Zhang, Y.C. Short-term Photovoltaic Power Forecasting Based on Improved Dung Beetle Optimizer for Optimizing VMD-BiLSTM. *J. Zhengzhou Univ. Eng. Sci.* **2026**, *47*, 59–66.
9. Tian, Z.; Liang, B. PVMTF: End-to-end long-sequence time-series forecasting frameworks based on patch technique and information fusion coding for mid-term photovoltaic power forecasting. *Appl. Energy* **2025**, *396*, 126263.
10. Bai, L.; Yu, B.; Gao, F.; Gu, J.H.; Xu, J. Short-term photovoltaic power prediction based on ICEEMDAN and TCN-AM-BiGRU. *Electron. Meas. Technol.* **2024**, *47*, 61–69.
11. Han, B.; Li, C.Q.; Liu, W.L.; Liu, S.; Liu, C.L.; Xu, J.H.; Wang, X. Ultra-short-term Photovoltaic Power Forecasting Based on Secondary Decomposition and BiGRU. *J. Chin. Soc. Power Eng.* **2025**, *45*, 62–69+79.
12. Zhou, Z.D.; Dai, Y.M.; Leng, M.M. A photovoltaic power forecasting framework based on Attention mechanism and parallel prediction architecture. *Appl. Energy* **2025**, *391*, 125869. [CrossRef]
13. Zhou, D.X.; Liu, Y.J.; Wang, X.; Wang, F.X.; Jia, Y. Combined ultra-short-term photovoltaic power prediction based on CEEMDAN decomposition and RIME optimized AM-TCN-BiLSTM. *Energy* **2025**, *318*, 134847.
14. Liang, J.W.; Yin, L.Y.; Xin, Y.L.; Li, S.C.; Zhao, Y.Q.; Song, T. Short-term photovoltaic power prediction based on CEEMDAN-PE and BiLSTM neural network. *Electr. Power Syst. Res.* **2025**, *246*, 111706.
15. Li, D.; Wang, L.; Wei, Z.; Zhang, K. Fault Diagnosis of High-Speed Train Bogies Based on WOA-VMD and Transformer-Kan. *Comput. Appl. Softw.* **2025**, 1–13.
16. Zhang, X.B.; Wang, J.; Liu, L.Q.; Cao, J.; Quan, Y.; Xie, X.N.; Xiang, P. Prediction of freeze-thaw damage of asphalt concrete based on distributed fiber optic sensors and KAN-Transformer fusion model. *Opt. Fiber Technol.* **2025**, *94*, 104304.
17. Yuan, Y.; Zhao, R.B.; Xu, H.T.; Zhao, Y.Z.; Sun, Z.Q.; Yang, F.X.; Zhan, H.Y.; Li, L.B.; Lv, S.Y. Application of CNN-LSTM based on physical model constraint and attention mechanism in in-situ combustion production prediction. *Pet. Geol. Oilfield Dev. Daqing* **2025**, *44*, 90–100.
18. Fu, L.; She, L.J.; Yan, D.L.; Zhang, P.; Long, X.Y. Fatigue Damage Prediction Framework of The Boom System Based on Embedded Physical Information and Attention Mechanism BiLSTM Neural Network. *J. Mech. Eng.* **2024**, *60*, 205–215. [CrossRef]

19. Gao, X.F.; Zang, Y.F.; Ma, Q.; Liu, M.M.; Cui, Y.M.; Dang, D.Z. A physics-constrained deep learning framework enhanced with signal decomposition for accurate short-term photovoltaic power generation forecasting. *Energy* **2025**, *326*, 136220.
20. Ma, H.Y.; Cheng, G.X.; Wang, H.Y.; Zhao, J.B. Combinatorial Prediction Method Based on Meteorological Irradiation Data and GSA-optimized VMD-BiLSTM. *Mod. Electr. Power* **2026**, *43*, 213–222.
21. Wang, L.Z.; Li, C.Y.; Li, C.; Liu, J. Short-term photovoltaic power prediction of LSTM network optimized by a multi-strategy improved SSA. *Eng. J. Wuhan Univ.* **2025**, *58*, 1356–1366.
22. Zhou, N.; Shang, B.W.; Xu, M.M.; Peng, L.; Feng, G. Enhancing photovoltaic power prediction using a CNN-LSTM-attention hybrid model with Bayesian hyperparameter optimization. *Glob. Energy Interconnect.* **2024**, *5*, 667–681.
23. Quan, R.; Qiu, Z.Z.; Wan, H.; Yang, Z.Y.; Li, X.R. Dung beetle optimization algorithm-based hybrid deep learning model for ultra-short-term PV power prediction. *iScience* **2024**, *27*, 111126. [[PubMed](#)]
24. Yang, S.M.; Luo, Y.M. Short-term photovoltaic power prediction based on RF-SGMD-GWO-BiLSTM hybrid models. *Energy* **2025**, *316*, 134545. [[CrossRef](#)]
25. Fang, J.; Wang, X.D.; Wang, Y.; Wu, Z.W.; Ding, J. Thickness Prediction of Zinc Layer of Hot-Dip Galvanized Sheet in Strip Based on Mic and Ipso-Rerm. *Hot Work. Technol.* **2024**, *53*, 62–68+82.
26. Jiang, D.L.; Xu, W.; Wu, W.; Chen, T.; Zhou, X.J.; Dong, L. Coupling intelligent fault diagnosis method based on PCA and Transformer model centrifugal pump. *J. Mech. Electr. Eng.* **2025**, 1–10.
27. Zhang, W.Q.; Wang, J.X.; Song, Z.F. Research on network intrusion detection based on fusion Transformer and Inception. *Telecommun. Sci.* **2026**, *42*, 127–144.
28. Yu, Y.J.; Wang, D.Q.; Wei, Y.L.; Zhang, L.J. An improved algorithm for mixed excitation linear prediction speech coding based on Kolmogorov-Arnold network. *J. Appl. Acoust.* **2025**, *44*, 1641–1651.
29. Jiao, X.H.; Yang, L.Q.; Li, Z.H. Lithium battery life prediction based on HHO-LSTM-KAN model. *J. Nanjing Univ. Inf. Sci. Technol.* **2025**, 1–15.
30. Zhao, S.S.; Chen, J.F.; He, Y.G.; Zhang, C.L. Online Diagnosis Method of Open-circuit Fault for Four-level Converter Based on Light weight CNN and PINN. *High Volt. Eng.* **2026**, *52*, 1233–1243. [[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.