

## Article

# Research on the Harmonic Prediction Method of a PV Plant Based on an Improved Kernel Extreme Learning Machine Model

Zhenghan Liu <sup>1</sup>, Quanzheng Li <sup>2</sup>, Donglai Wang <sup>1,\*</sup>, Guifan Zhang <sup>1</sup>, Wei Wang <sup>1</sup>, Yan Zhao <sup>1</sup> and Rui Guo <sup>1</sup>

<sup>1</sup> Key Laboratory of Regional Multi-Energy System Integration and Control of Liaoning Province, Shenyang Institute of Engineering, Shenyang 110136, China

<sup>2</sup> State Grid Tieling Power Supply Company, State Grid Liaoning Electric Power Co., Ltd., Tieling 112000, China

\* Correspondence: wangdl@sie.edu.cn

**Abstract:** The harmonics of photovoltaic power plants are affected by various factors including temperature, weather, and light amplitude. Traditional power harmonic prediction methods have weak non-linear mapping and poor generalization capability to unknown time series data. In this paper, a Kernel Extreme Learning Machine (KELM) model power harmonic prediction method based on Gray Relational Analysis (GRA) with Variational Mode Decomposition (VMD) coupled with Harris Hawk Optimization (HHO) is proposed. First, the GRA method is used to construct the similar day set in one screening, followed by further using *K*-means clustering to construct the final similar day set. Then, the VMD method is adopted to decompose the harmonic data of the similar day set, and each decomposition subsequence is input to the HHO-optimized KELM neural network for prediction, respectively. Finally, the prediction results of each subseries are superimposed and numerical evaluation indexes are introduced, and the proposed method is validated by applying the above method in simulation. The results show that the error of the prediction model is reduced by at least 39% compared with the conventional prediction method, so it can satisfy the function of harmonic content prediction of a photovoltaic power plant.

**Keywords:** gray relational analysis; Harris hawk optimization; *K*-means clustering; kernel extreme learning machine; variational mode decomposition



**Citation:** Liu, Z.; Li, Q.; Wang, D.; Zhang, G.; Wang, W.; Zhao, Y.; Guo, R. Research on the Harmonic Prediction Method of a PV Plant Based on an Improved Kernel Extreme Learning Machine Model. *Electronics* **2024**, *13*, 32. <https://doi.org/10.3390/electronics13010032>

Academic Editor: Rui Castro

Received: 15 October 2023

Revised: 6 December 2023

Accepted: 12 December 2023

Published: 20 December 2023



**Copyright:** © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Harmonic prediction in PV power plants is the key to supporting the safe operation of power systems and the development of dispatching strategies [1]. Since voltage harmonics are susceptible to the disturbance of many factors, they are non-linear and non-stationary while maintaining periodicity. This brings a greater challenge for harmonic prediction and detection. Predicting power harmonics based on the construction of similar daily sets can provide the basis for the basic real-time data required for regional smart grid regulation and control technology [2].

The current forecasting methods are mainly divided into conventional statistical methods and machine learning techniques. The traditional statistical techniques are the time series method [3,4], Kalman filter [5], etc. Machine learning methods include artificial neural networks [6], Long Short-Term Memory (LSTM) neural networks [7,8], Least Squares Support Vector Machines (LSSVM) [9], etc. Reference [10] proposed a short-term load prediction method based on a Kalman filter, but the algorithm is not applicable to non-linear non-stationary time series data processing and the prediction accuracy is poor. Reference [11] adopted the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to extract the time-frequency domain feature information of wind speed sequences, used an echo state network to train the input information, and finally adopted LSSVM to correct the error and improve the prediction accuracy. Reference [12] adopted an improved Particle Swarm Optimization (PSO) algorithm to

optimize the parameters of the Multi-Kernel Extreme Learning Machine (MKELM) for prediction, but the global search capability and convergence speed of the PSO were limited, and the prediction effect was not satisfactory. In summary, the traditional methods based on statistics have poor mapping ability for non-linear and non-smooth data, and the processing effect is not satisfactory. The machine-learning-based methods have good generalization ability for unknown non-linear time series data, but the hyper-parameter search is difficult.

Many scholars have proposed a combined prediction model that uses signal decomposition methods to mine time series data features deeply and have adopted swarm intelligence optimization algorithms to optimize the parameters of the machine learning model to address the problem of the difficulty of a hyper-parameter search for the optimization of single machine learning [13–15]. In reference [16], the wind speed data were decomposed by Wavelet Transform (WT) and fed into the LSTM network to obtain the time-series characteristics of wind speed data. In reference [17], the time series were decomposed into several subseries by Variational Mode Decomposition (VMD), and the new subseries were input to the Gated Recurrent Units (GRU) network by using sample entropy to filter and reconstruct each series, and this achieved better prediction results.

The historical data need to be pre-processed for standardization before input. Reference [18] only screened the input data for external factors, and did not construct similar day sets for historical data, resulting in the presence of data with little correlation with the day to be predicted in the input data, interfering with the prediction results. In reference [19], only time series factors were considered when constructing similar day sets considering holiday forecasts, and key factors such as weather conditions and temperature were not taken into account. Reference [20] used the improved Grey Relational Analysis (GRA) to construct a similar day set, but did not consider the clustering selection of secondary similar days, so the final similar day set obtained was not accurate enough. This resulted in insufficient data cleaning and anomalous data removal from the input data, which led to a decrease in prediction accuracy.

In summary, traditional prediction methods cannot take into account the influence of multiple external factors on harmonic prediction, while the Kernel Extreme Learning Machine (KELM) can train the relationship between harmonic data and external data and deal with nonlinear and non-smooth problems better. GRA can filter historical data from geometric aspect similarity, and the *K*-means clustering method can select historical data from data aspect similarity. The advantages of Harris Hawk Optimization (HHO) are quick convergence and great global search reliability. Therefore, the method of photovoltaic power plant harmonic prediction based on GRA with the VMD-HHO-KELM model is proposed. First, the significant impacting elements are chosen using the Pearson correlation coefficient approach, and then the GRA approach and *K*-means clustering are used to create the final collection of comparable days. Then, the VMD approach is adopted to decompose the harmonic data of the collection of comparable days, and each decomposed subsequence is input to the constructed HHO-KELM neural network for prediction, respectively. Finally, all subseries prediction results are superimposed, numerical evaluation indicators are introduced, and the proposed approach is validated by applying the above method in simulation and comparing with LSTM, GRU, and other multi-models.

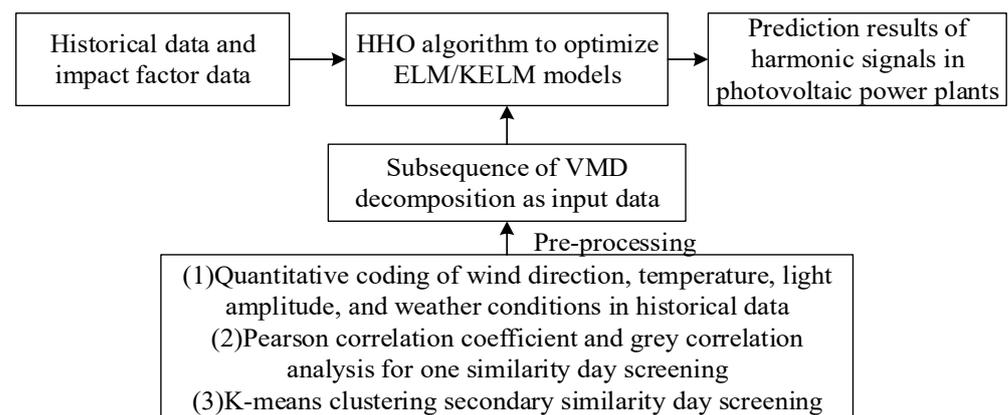
The main contributions of this study are summarized as follows:

- (1) The HHO algorithm has good convergence speed and search efficiency. The HHO algorithm is applied to the KELM neural network, which overcomes the difficulty of selecting parameters, and establishes a good prediction model for the harmonic content of infrastructure construction.
- (2) In response to the problems of unclear spatiotemporal trend flow and difficulty in tracing the source of harmonics in PV power plants, as well as the existing traditional prediction models with delay lag and poor prediction accuracy, this paper constructs a GRA-VMD-HHO-KELM harmonic prediction model, which can eliminate the harm caused by the pollution of each harmonic component and realize the function of accurate prediction of different sub-harmonics and harmonic content.

The structure of the paper is as follows. In Section 2, the problem of harmonic prediction for photovoltaic power plants is described. In Section 3, the theoretical basis is introduced, including the KELM model, the harmonic detection method based on VMD, and the design of the HHO algorithm. The strategy structure of voltage harmonic signal prediction is proposed, and the above methods are combined in Section 4. The experimental analysis is carried out, and the superiority of the proposed method is verified by comparing the prediction results of different methods in Section 5. Conclusions are drawn in Section 6.

## 2. Problem Description of Harmonic Prediction for Photovoltaic Power Plants

Photovoltaic power plant harmonic prediction should not only consider the fluctuation of harmonics with the change in time series factors. It also needs to consider the effects of weather, light amplitude, wind direction, wind power, and temperature on voltage harmonics interference. However, the number of external factors considered is not positively correlated with the prediction accuracy, so it is necessary to screen and identify the external factors with greater harmonic influence [21]. At the same time, since harmonic prediction requires a large amount of data support, there is a high probability that there are historical data with low or no correlation with the date to be predicted, and if these data are directly input into the prediction model, the prediction accuracy will be reduced, so it is necessary to search for historical data with a high correlation with the date to be predicted and train them. Most of the existing reactive power compensation devices are statically installed, and the compensation devices cannot achieve the estimation of harmonic upper limit to complete one installation and commissioning, which brings great challenges to the prediction of harmonic pollution sources, spatial and temporal trend flow analysis, and safety hazard analysis, and the real-time accurate prediction of each harmonic component in PV power plants is of great significance for harmonic pollution management and the dynamic commissioning and tracking management of harmonic suppression devices. The current common harmonic prediction models based on statistical methods still have problems, such as poor prediction accuracy and delay lag. They cannot provide a reasonable basis for the number or capacity of harmonic suppression equipment to be put into operation in the future, so a suitable machine learning model for power harmonic signal prediction needs to be constructed. The power harmonic prediction problem is shown in Figure 1 [22].



**Figure 1.** Power harmonic prediction problem description.

## 3. Theoretical Basis

### 3.1. Kernel Extreme Learning Machine

In this kernel function form of the ELM algorithm, the appropriate kernel function  $K(u,v)$  must be provided and the buried layer feature mapping  $h(x)$  does not need to be known [23]. It is also not necessary to specify the count of nodes of the buried layer  $L$ . The problem of the unsatisfactory generalization ability and stability caused by the randomly given parameters of the buried layer of ELM is effectively improved, and the computational

complexity is largely reduced to avoid the optimization of the total amount of nodes of  $L$ . Since the implicit layer mapping relation function  $h(x)$  is unknown, Huang et al. established a kernel matrix instead of  $HH^T$  by studying the kernel function based on ELM, relying on the Mercer condition [24].

$$\Omega_{i,j} = h(x_i) \cdot h(x_j) = K(x_i, x_j) \quad (1)$$

$$\Omega_{ELM} = HH^T = \begin{bmatrix} K(x_1, x_1) & L & K(x_1, x_N) \\ M & 0 & M \\ K(x_N, x_1) & L & K(x_N, x_N) \end{bmatrix} \quad (2)$$

$$h(x)H^T = \begin{bmatrix} K(x, x_1) \\ M \\ K(x, x_N) \end{bmatrix}^T \quad (3)$$

where  $\Omega_{ELM}$  is the kernel matrix,  $\Omega_{i,j}$  is the elements in matrix.

In the KELM algorithm, the mapping function  $h(x)$  does not need to be known, and there is no requirement to provide the total amount of nodes in  $L$ ; only the corresponding kernel function  $K(u,v)$  is given. The KELM algorithm flows as follows [24].

Input: training samples  $x = \{(x_i, t_i) \mid x_i \in \mathbb{R}^d, t_i \in \mathbb{R}^m, i = 1, \dots, N\}$ , kernel function  $K(u,v)$ . Output:

$$f(x) = \begin{bmatrix} K(x, x_1) \\ M \\ K(x, x_N) \end{bmatrix}^T \left( \frac{I}{\lambda} + \Omega_{ELM} \right)^{-1} T \quad (4)$$

where  $K(x, x_N)$  is the kernel function, and  $T$  and  $I$  are the target vector matrix and diagonal matrix, respectively;  $\lambda$  is the regularization coefficient, and the smaller the  $\lambda$  the stronger the model generalization ability; the larger the  $\lambda$  the higher the model prediction accuracy. The selection of appropriate  $\lambda$  is crucial to the model. The HHO algorithm benefits from having few adjustable parameters, strong full-optimal search capability, and a strong search efficiency, so the algorithm is introduced to optimize the KELM regularization factor  $\lambda$  and the kernel function parameters  $\sigma$ .

### 3.2. Harmonic/Inter-Harmonic Detection Method Based on VMD

VMD is a complete fully non-recursive signal processing method proposed for processing non-smooth and non-linear signals [25]. Since the harmonic voltage signal of PV power plants has a certain periodicity and volatility, it was decided to first apply  $K$ -means clustering to select the group of comparable days, and then to adopt the VMD method to decompose the historical harmonic voltage signal curve into multiple subseries with different frequencies and relative smoothness. Thus, the complexity of the harmonic curve and the non-smoothness and non-linearity of the time series were effectively reduced. The implementation steps are as follows.

The noise is dealt with according to the experimental requirements and the actual situation of the power station in this paper. Since it is not the focus of this paper, the conventional filtering method was applied before the process of decomposing harmonics by VMD method, and higher harmonics exceeding 750 Hz were eliminated. After elimination, the measured harmonics were decomposed into corresponding modes.

For ensuring that each modal function is a component of finite bandwidth with center frequency such that the decomposed modes is minimized, the constrained variational problem is [25]:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \tag{5}$$

$$\text{s.t. } \sum_{k=1}^K u_k = f(t)$$

where  $\{u_k\} = \{u_1, u_2, \dots, u_k\}$  is the  $K$  modal components obtained from the decomposition;  $\{\omega_k\} = \{\omega_1, \omega_2 \dots \omega_k\}$  is the frequency center of each component;  $\delta(t)$  is the impulse function; and  $f(t)$  is the original signal. The main iterative VMD solving process is as follows:

Step 1. Initialize  $\{\hat{u}_k^{(1)}\}, \{\hat{\lambda}^{(1)}\}, \{\hat{\omega}_k^{(1)}\}$  and the maximum number of iterations  $n$ ;

Step 2. Update  $u_k, \omega_k,$  and  $\lambda_k$ .

Step 3. Convergence criterion  $\varepsilon > 0$ . If the iteration stopping condition  $\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \varepsilon$  is not satisfied, then return to step (2); otherwise, terminate the iteration and output the result to obtain each modal component  $u_k$ .

VMD parameter setting: the number of VMD components  $K$  is very important to the decomposition effect. Usually,  $K$  is chosen from 3 to 8 as the number of VMD components; in this paper  $K = 6$ . The VMD penalty factor, initial center frequency, and convergence criterion are set to  $\alpha = 2000, init = 1,$  and  $\varepsilon = 10^{-7},$  respectively.

### 3.3. Design of the Harris Hawk Optimization Algorithm

The algorithm was proposed for multi-dimensional complex problem solving. The algorithm is significantly improved in stability, convergence accuracy and speed, in comparison to certain conventional swarm intelligence optimization methods, and has excellent performance, especially in high-dimensional multi-polarity solution problems [26].

(1) Global search phase:

The Harris hawk’s roundup behavior will be transformed into different behaviors depending on the prey escape energy  $E$ .  $E$  is shown in the following equation:

$$E = 2E_0 \left( 1 - \frac{t}{T} \right) \tag{6}$$

where  $t$  is the current number of iterations,  $E_0$  is a random number in the interval  $(-1, 1),$  and  $E$  represents the current prey’s escape energy.

If the  $|E|$  of the prey is greater than 1, the Harris hawk flock will disperse to fly in a larger range to find the prey, and a random number  $q$  will be randomly generated for the different cases of the Harris hawks finding the prey and not finding prey to obtain the search phase equation:

$$X(t + 1) = \begin{cases} X_{rand}(t) - r_1 |X_{rand}(t) - 2r_2 X(t)| & q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)) & q < 0.5 \end{cases} \tag{7}$$

where  $X(t + 1)$  represents the position vector of the Harris hawk at the next iteration,  $X(t)$  is the current position vector of this hawk,  $X_{rabbit}(t)$  is the prey position vector,  $X_{rand}(t)$  represents the position of random individuals in the hawk population, the upper and lower boundaries of this dimensional variable are represented by  $UB$  and  $LB,$  and  $r_1, r_2, r_3, r_4,$  and  $q$  are random numbers between  $(0, 1).$  The mean position of the eagle  $X_m(t)$  can be calculated by the following equation:

$$X_m(t) = \frac{1}{N} \sum_{i=1}^N X_i(t) \tag{8}$$

where  $N$  represents the number of individuals in the eagle population and  $X_i(t)$  denotes the position of each eagle in iteration  $t$ .

When  $q \geq 0.5$ , the prey is not detected by any of the hawks, and therefore will randomly select individuals in the population to approach the prey location and update their own position. If  $q < 0.5$ , the prey is detected and the Harris hawk targets the prey, circles around it, and updates its position.

(2) Local exploitation phase:

According to the escape behavior of the prey and the Harris hawk's pursuit strategy, the assumption is the chance of the prey escaping before the raid with a successful escape ( $r < 0.5$ ) or unsuccessful escape ( $r > 0.5$ ). When the prey's escape energy  $|E| < 1$ , the prey is physically weak and the hawks will enter the siege phase, in which the hawks have four types of siege depending on whether the escape energy  $E$  is greater than 0.5 and whether the prey escapes the siege [27].

Circling roundup: when  $|E| \geq 0.5$  and  $r \geq 0.5$ , the prey still has enough energy to escape but cannot escape from the encirclement. Harris hawks will circle around the prey and continue to consume the prey's energy:

$$X(t+1) = \Delta X(t) - E|JX_{\text{rabbit}}(t) - X(t)| \quad (9)$$

$$\Delta X(t) = X_{\text{rabbit}}(t) - X(t) \quad (10)$$

where  $X(t+1)$  represents the position of the Harris hawk at the next iteration stage,  $\Delta X(t)$  represents the difference between the prey and the Harris hawk position when the current number of delivery is  $t$ , and  $J$  denotes the random jump strength of the prey.

Strong raid: when  $|E| < 0.5$  and  $r \geq 0.5$ , at this point the Harris hawks consider the prey to be physically exhausted and will make a final raid on the prey at:

$$X(t+1) = X_{\text{rabbit}}(t) - E|\Delta X(t)| \quad (11)$$

Hovering roundup and progressive dive attack: when  $|E| \geq 0.5$  and  $r < 0.5$ , the prey is energetic and still has a chance to escape, and the HHO algorithm introduces the concept of Levy flight (LF) to model the disorienting behavior of the prey with variable routes during the escape phase. The hawks will evaluate their next behavior according to the following equation:

$$Y = X_{\text{rabbit}}(t) - E|JX_{\text{rabbit}} - X(t)| \quad (12)$$

The obtained  $Y$  is then compared with the current position fitness to detect whether the hunt is successful. Assuming that the hunt is unsuccessful, the hawk flock will start to make irregular swoops and perform raids. The flock takes a raid and updates its position based on LF, as follows:

$$Z = Y + S \times LF(D) \quad (13)$$

where  $S$  is a  $D$ -dimensional random vector.  $D$  is the problem dimension, and the final Harris hawk location update final decision for this phase is

$$X(t+1) = \begin{cases} Y & \text{if } F(Y) < F(X(t)) \\ Z & \text{if } F(Z) < F(X(t)) \end{cases} \quad (14)$$

Strong raid with progressive dive attack: when  $|E| < 0.5$  and  $r < 0.5$ . The prey is low on energy but has a chance to escape. The movement of the hawks is similar to a progressive dive attack in a circling roundup, but the difference is that the hawks will try

to reduce the average distance to the prey. The following rules are therefore implemented under strong siege conditions:

$$X(t + 1) = \begin{cases} Y & \text{if } F(Y) < F(X(t)) \\ Z & \text{if } F(Z) < F(X(t)) \end{cases} \tag{15}$$

$$Y = X_{\text{rabbit}}(t) - E|JX_{\text{rabbit}} - X(t)| \tag{16}$$

$$Z = Y + S \times LF(D) \tag{17}$$

#### 4. Strategy Structure for Voltage Harmonic Signal Prediction

##### 4.1. HHO-KELM Power Harmonic Prediction Model

##### 4.1.1. HHO Algorithm Mathematical Model for KELM Parameter Optimization

The Mean Square Error (MSE) between the expected and observed values of KELM is minimized as the objective function.

$$\min \text{MSE} = \frac{1}{N} \sum_{i=1}^N (\bar{f} - y_i)^2 \tag{18}$$

where  $N$  is the sampling point,  $\bar{f}$  is the predicted value of the  $i$ th sampling point, and  $y_i$  is the actual value of the  $i$ th sampling point.

To increase the precision of the algorithm and identify the best parameters of ELM and KELM and to reduce the algorithm's search time, upper and lower bound constraints need to be set on the parameters to be optimized. The RBF kernel function is selected as the KELM kernel function:

$$\begin{cases} \lambda \in [0.01, 50] \\ \sigma \in [0.01, 50] \end{cases} \tag{19}$$

$$\begin{cases} \omega_i \in [-1, 1] \\ b_i \in [0, 1] \end{cases} \tag{20}$$

$$K(x, z) = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right) \tag{21}$$

where  $\lambda$  is the regularization coefficient of the KELM,  $\sigma$  is the parameter of the RBF kernel function in the KELM,  $\omega_i$  is the weight in the ELM, and  $b_i$  is the bias in the ELM. In the HHO algorithm, the relationship between the hunting behavior of Harris hawks and the optimization problem is shown in Table 1, and the process of finding the optimal parameters of ELM and KELM is as follows.

**Table 1.** Correspondence between hunting behavior and parametric optimization problems.

Hawk Hunting Behavior	Specific Optimization Issues
Prey populations	Two-dimensional vector consisting of ELM parameters $\omega_i, b_i$ and KELM parameters $\lambda, \sigma$
Search and hunt for prey process	ELM, KELM prediction process
Current optimal individual position	Adaptation values in the problem
Final optimal individual position	The optimal parameters $\omega_i, b_i$ and $\lambda, \sigma$ for ELM, KELM

The position vector of each eagle in the HHO algorithm corresponds to the total number of hyper-parameters  $\omega_i, b_i, \lambda, \sigma$  in the ELM and KELM of this problem; the total number of eagles in the HHO algorithm corresponds to the total number of hyper-parameters optimized in this problem.

In this paper, the predicted MSE is used as the fitness function, and the maximum number of iterations is set to  $T$ . The value of the fitness function will correspond to the feasible solution as the ELM, KELM hyper-parameters  $\omega_i, b_i, \lambda, \sigma$ , and the historical data in the set of similar days after VMD decomposition are input and trained to obtain the MSE between the expected and observed values of the day to be predicted.

According to the new position of the adjusted population, the fitness value of each individual is traversed in turn to find the best fitness value and compare it with the fitness value of the prey. When it is better than the fitness value of the prey, it is recorded as the MSE between the latest expected and observed value, and the corresponding ELM and KELM hyper-parameters  $\omega_i, b_i, \lambda, \sigma$  are updated and stored to obtain the optimal hyper-parameters corresponding to each implicit layer neuron of the ELM, as shown in Table 2.

**Table 2.** Optimal parameters corresponding to each hidden layer neuron (excerpt of 5 partial neurons).

Number of Neurons in the Hidden Layer	Implicit Layer Bias Value $b_i$	Input Layer Weights $\omega_i$		
1	0.0356	0.1578	−0.1621	0.2194
		0.1600	−0.0451	0.1466
		−0.1910	−0.0565	0.0392
2	0.0241	0.0041	−0.0989	−0.0846
		−0.1218	−0.0181	−0.0818
		−0.0857	−0.0211	−0.2255
3	0.0108	0.0117	0.2694	−0.3432
		−0.2681	0.1456	0.0033
		0.0754	−0.1140	−0.0785
4	0.0318	−0.0326	0.2581	0.0737
		−0.2329	−0.2367	0.1381
		0.2356	−0.1474	−0.0550
5	0.0184	0.0825	0.2727	0.1794
		−0.0774	0.1574	0.0228
		−0.0396	0.1301	0.2202

#### 4.1.2. HHO-ELM/KELM Harmonic Prediction Model for PV Plants

This paper set the number of ELM hidden layer neuron nodes to 20, the number of hidden layers to 15, the range of weights  $\omega_i$  to  $[-1, 1]$ , the range of bias  $b_i$  to  $[0, 1]$ , the range of KELM regularization coefficients  $\lambda$  kernel function parameters  $\sigma$  to  $[0.01, 50]$ , the number of populations of HHO algorithm to 25, and the maximum number of iterations to 100, and the basic steps of HHO-ELM/KELM modeling are as follows:

- (1) Import the PV plant voltage data samples  $N = \{(x_i, t_i) \mid x_i \in R^n, t_i \in R^m, i = 1, 2, \dots, N\}$ ; the excitation function is  $g(x)$ , the number of implied layer nodes is  $L$ , and the samples are divided into training samples and test samples according to the predefined ratio.
- (2) Initialize the number of neurons in the hidden layer of the network  $L$ , the maximum number of iterations  $T$ , the number of populations  $N$ , and the ELM and KELM hyper-parameters  $\omega_i, b_i, \lambda, \sigma$ . The MSE of the ELM and KELM training models are used as the HHO fitness function and a new set of initial populations  $\theta$  is randomly generated.
- (3) For each population  $\theta$ , the implied layer output matrix  $H$  and the output weights  $\beta$  are calculated and the MSE is solved.
- (4) Calculate the individual fitness of the population in turn, and update and save the current optimal individual position and fitness.
- (5) Update and assess prey escape energy  $E$ , escape probability  $r$ , and escape jump strength  $J$ ; complete the update of all individual locations in the population.
- (6) Loop steps (4) and (5) are repeated until the target task is completed, the maximum number of iterations is reached, and the optimal hyper-parameters  $\omega_i, b_i, \lambda, \sigma$  are obtained.

- (7) After finding the optimal  $\theta$ , the output weight  $\beta$  is calculated by training the mathematical model of the network  $H^T H\beta = H^T T$ .
- (8) Import PV plant voltage data test samples and outputting predicted voltage waveforms and errors, etc.

The flowchart of the Harris Hawk Optimization Kernel Extreme Learning Machine algorithm (HHO-KELM) is shown in Figure 2.

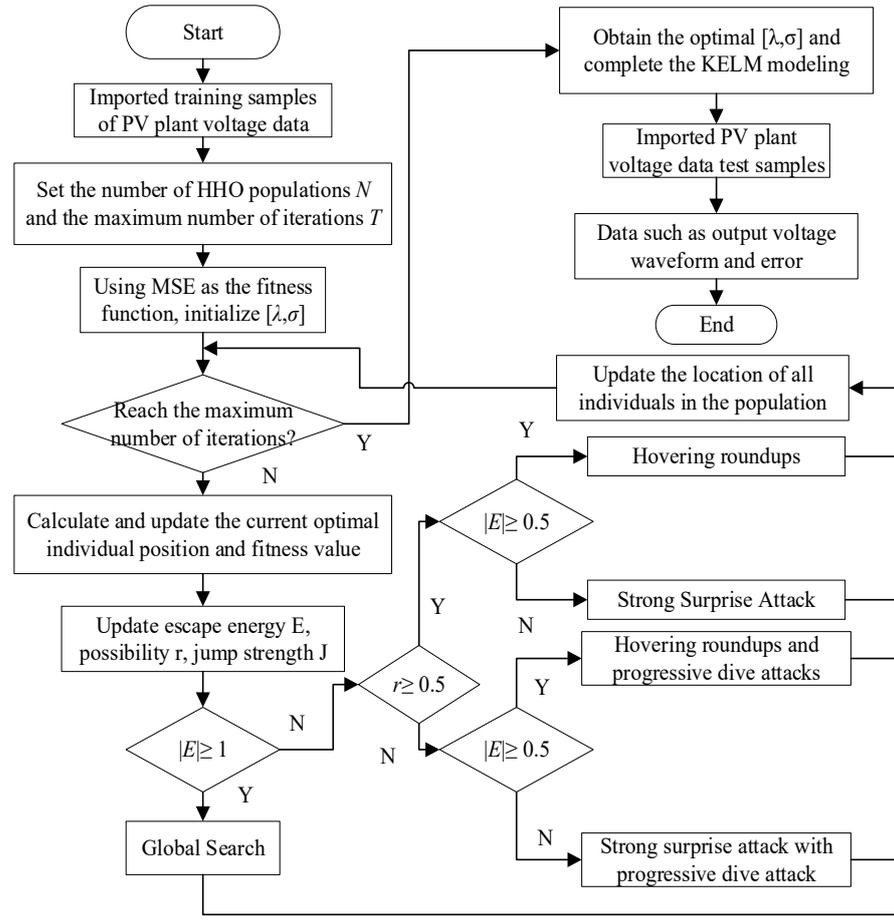


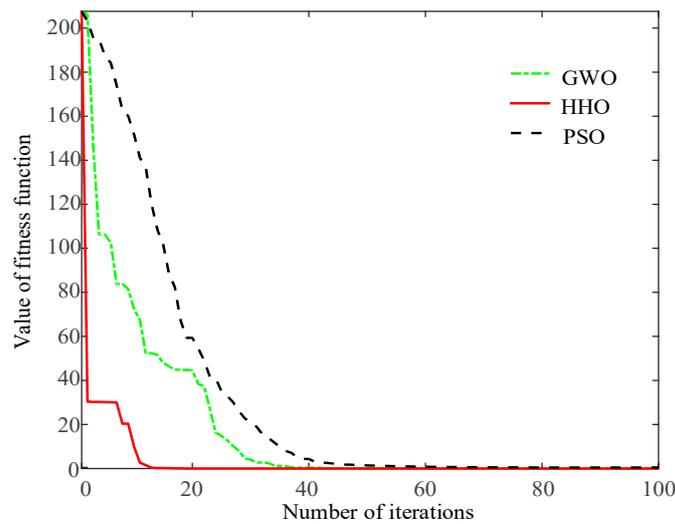
Figure 2. Flow chart of the HHO-KELM algorithm model.

The iteration curves of HHO, grey wolf optimization (GWO), and PSO are shown in Figure 3. The HHO algorithm converges about 15 times and the fitness function finally converges to 0. The GWO algorithm converges about 30 times, and the value of the fitness function finally converges to 15, and the PSO algorithm converges about 50 times and the value of the fitness function eventually converges to 10. The number of convergences of the HHO algorithm and the ultimate convergence values of GWO and PSO are better than those of GWO and PSO.

To further compare the performance of several algorithms, the optimal solutions of several algorithms after 20 runs were extracted, respectively, and relative percentage growth rate  $RPI$  was introduced to assess the effectiveness of several algorithms, including the following:

$$RPI(f) = \frac{(f - f^*)}{f^*} \times 100\% \tag{22}$$

where  $f$  is the smallest MSE for a single run of each method between the expected and observed values of ELM/KELM, and  $f^*$  is the smallest MSE among all minimum MSE.



**Figure 3.** Comparison of HHO, GWO, and PSO convergence iteration curves.

From Table 3, it can be seen that the optimal parameters obtained by the HHO algorithm seeking are more significant for the improvement of the prediction accuracy of ELM, KELM compared to the GWO and PSO algorithms. To more intuitively reflect the performance gap between the algorithms of GWO, PSO, and HHO, a multivariate analysis of variance (MANOVA) was conducted on the above three, in which the algorithms of GWO, HHO, and PSO were defined as factors, and the results at the 95% confidence level are shown in Table 4.

**Table 3.** RPI values of HHO, GWO, and PSO.

Number of Runs	HHO	GWO	PSO	Number of Runs	HHO	GWO	PSO
1	1.762	12.392	25.321	11	0.432	13.569	27.675
2	0.259	9.375	24.287	12	2.187	11.257	21.913
3	3.748	11.382	23.876	13	1.457	9.292	22.481
4	1.321	12.824	27.830	14	1.820	8.295	23.207
5	0.654	10.410	26.495	15	0.351	12.309	28.239
6	4.825	14.732	22.159	16	2.287	13.514	27.865
7	3.335	12.843	22.47	17	0.681	10.543	23.062
8	0.481	12.435	25.481	18	0.731	10.105	23.645
9	1.182	9.365	25.688	19	0.002	12.899	22.733
10	2.231	11.11	24.631	20	1.762	12.562	24.972

**Table 4.** MANOVA of HHO, GWO, and PSO at the 95% confidence level.

Data Type	HHO	GWO	PSO
Average value	1.575	11.561	24.702
Sample variance	1.612	2.956	4.280
Population variance	1.532	2.809	4.066

#### 4.2. Combined VMD-HHO-KELM/ELM Prediction Model Based on GRA

Since the original PV voltage harmonic signal has strong non-linearity and non-smoothness, this paper firstly uses GRA and the K-means algorithm to cluster the original data according to external influencing factors to form a similar day set, and uses the HHO algorithm to search the weights  $\omega_i$ , bias  $b_i$ , kernel function parameters  $\lambda$ , and regularization coefficients  $\sigma$  of ELM/KELM. Then, the PV voltage harmonic curve of each class of data VMD decomposition is performed, and then the ELM/KELM neural network optimized by HHO is used for prediction. Finally, in order to assess the effectiveness of the prediction

model, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) are selected. The prediction process strategy is shown in Figure 4, and the steps are as follows:

- (1) A Pearson correlation coefficient approach is adopted to pinpoint important influencing elements. Secondly, GRA is used to construct a similar day set, and K-means clustering is used to screen similar days again to form the final similar day set.
- (2) The historical PV voltage harmonic signal data in each class of data are decomposed by VMD.
- (3) The HHO algorithm is adopted to search for  $\omega_i, b_i, \lambda, \sigma$  of the optimized ELM/KELM neural network.
- (4) Each subsequence of the VMD decomposition is used as the input of the HHO-optimized ELM/KELM neural network for prediction, respectively. The prediction results of each subsequence are superimposed and the accuracy of the model prediction is verified.

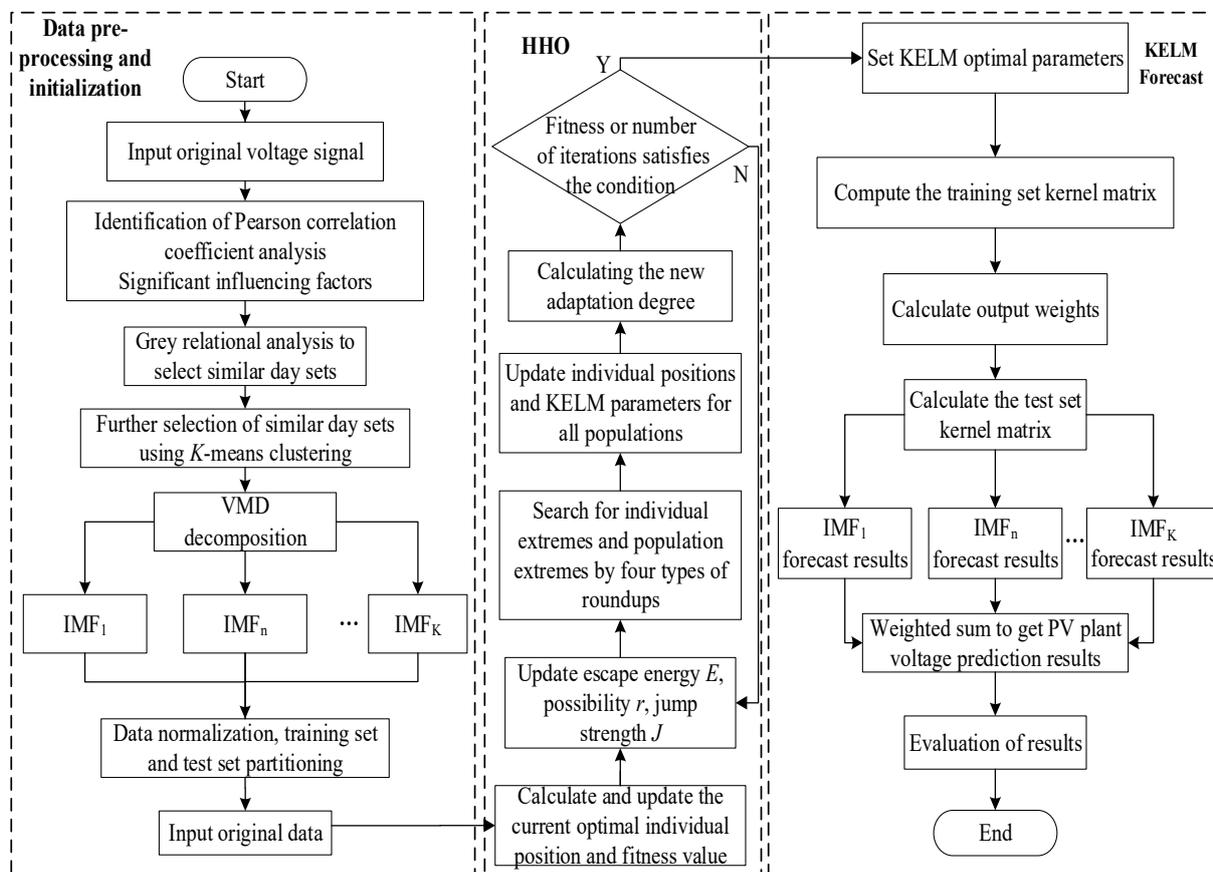


Figure 4. Strategy diagram of the harmonic signal prediction method for photovoltaic power plants.

### 5. Example Analysis

#### 5.1. Dataset Description, Normalization Process, and Performance Evaluation Index

In order to ensure the experimental effect, the hardware in this paper was configured with Intel(R) Core(TM) i5-9300F CPU @ 2.40 GHz, 2400 Mhz and NVIDIA GeForce GTX 1660 Ti. The study used a one-year data set of actual PV plant voltage harmonics signals from Shenyang, China. The data set contains voltage harmonics data, wind direction, wind, temperature, light amplitude, humidity, weather, and data on whether or not it is a business day or a holiday. The influencing factor data sample interval is one day, the multiple sampling intervals of the harmonics data is 10 min, and the single sampling is 30 s. The sampling frequency is 10 kHz, and each point is sampled 5 times.

In order to eliminate the problem of errors arising from inconsistencies in the unit magnitude of the voltage harmonic data and each influence factor, the data are normalized before the voltage data are input into the prediction model:

$$X'_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{23}$$

where  $X'_i$  is the normalized value of historical data;  $X_i$  is the historical original data; and  $X_{\max}$  and  $X_{\min}$  are the maximum and minimum values of voltage harmonics and each influencing factor in historical original data, respectively.

The study uses R-squared ( $R^2$ ), MAE, MAPE, and RMSE as evaluation indexes of the prediction model. The specific indexes are as follows:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{24}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\bar{f} - y_i| \tag{25}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \bar{f}}{y_i} \right| \tag{26}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\bar{f} - y_i)^2}{N}} \tag{27}$$

where  $\bar{f}$  is the mean of the predicted data,  $y_i$  is each actual value, and  $N$  is the number of predicted samples.

### 5.2. Identification of Important Influencing Factors

Influencing factors such as wind direction, weather conditions, and holidays are generally expressed in textual form and cannot be better correlated with voltage harmonics for analysis. Therefore, they should be quantified into numerical indexes, and Monday to Friday set as workdays with a value of 1, and Saturday and Sunday as non-workdays with a value of 0. The influencing factors, such as whether it is a holiday, weather conditions, or wind direction, can be quantified similarly, and the quantified results are shown in Table 5.

**Table 5.** Weather conditions, wind direction, date type quantitative code.

Influencing Factors	Coding	Influencing Factors	Coding
Holiday	1 if the day is a holiday and 0 if otherwise	North wind	1.75
Workday	1	Northeast wind	1.875
Weekend	0	Heavy rain or blizzard	−4
East wind	1	Heavy rain or snow	−3
Southeast wind	1.125	Medium rain or snow	−2
South wind	1.25	Light rain or snow	−1
Southwest wind	1.375	Cloudy day	0
West wind	1.5	Partly cloudy	1
Northwest wind	1.625	Sunny	2

Set each influencing factor data as matrix  $X = [X_1, X_2, X_3, X_4, X_5, X_6]$ , where  $X_1, X_2, X_3, X_4, X_5, X_6$  are wind direction, light amplitude, temperature, whether it is a business day or a holiday, and weather conditions, respectively. The voltage harmonic data are set

as  $Y$ . The Pearson correlation coefficient analysis is performed on the voltage harmonic data and each influencing factor to evaluate how closely they are related, and the value lies within the interval  $[-1, 1]$  and is usually denoted by  $\rho$  [28]. The standard definition of correlation by the Pearson correlation coefficient analysis is shown in Table 6. The Pearson correlation coefficients between the obtained voltage harmonic data and each influencing factor are shown in Table 7.

**Table 6.** Correlation degree and correlation coefficient table.

Relevance	Correlation Coefficient Interval
Very weak or no correlation	$0.0 \leq  \rho  < 0.2$
Weak correlation	$0.2 \leq  \rho  < 0.4$
Moderately relevant	$0.4 \leq  \rho  < 0.6$
Strong correlation	$0.6 \leq  \rho  < 0.8$
Very strong correlation	$0.8 \leq  \rho  < 1.0$

**Table 7.** Pearson correlation coefficient of influencing factors.

Influencing Factors	Coefficients	Influencing Factors	Coefficients
Wind direction	0.125	Whether it is a holiday	0.685
Temperature	0.795	Whether it is a workday	-0.634
Light amplitude	0.764	Weather conditions	0.035

As can be seen from Table 7, among the above influencing factors, the Pearson correlation coefficient of temperature has the largest absolute value of 0.795, representing the largest influence of temperature on harmonic data prediction, and the Pearson correlation coefficient of the weather condition has the smallest value of 0.035, representing the smallest influence of weather conditions on harmonic data prediction. Compared with other influencing factors, wind direction correlation has the least influence, and as shown in the table, weather condition and wind direction are extremely weak correlation factors and are not taken into consideration. The four external influencing factors of temperature, light magnitude, whether the day is a holiday, and whether the day is a workday are used as important bases for selecting similar days.

### 5.3. Similar Day Selection

Using 31 August 2021 as the date to be predicted, data on the influencing factors of historical date and the predicted date were extracted and subjected to gray correlation analysis [29]. The process of similarity day selection is as shown in Figure 5. From the above, it is known that the Pearson correlation coefficient values of temperature and light amplitude are the largest, so the weight taken by the similar days in the selection is the largest. Reference [30] set the gray correlation threshold to 0.7 and obtained a more satisfactory prediction effect. Therefore, in this paper, the threshold value is set to 0.7, and the set of similar days is constructed by screening the historical days greater than 0.7, with a total of 233 historical days.

K-means clustering is utilized to further generate a similar dayset in order to further decrease the previous data in the set with poor correlation with the forecast day [30]. The number of categories of clustering centers directly affects the clustering effect, and the silhouette coefficient (SIL) is an important basis for assessing the number of clustering centers. The closer the contour coefficient is to 1, the better the clustering effect, and for the contrary, the worse the effect. The correspondence between SIL and the number of clustering centers is shown in Figure 6. From Figure 6, it can be seen that the SIL factor is the highest when the number of cluster centers is 3. Therefore, according to the change trend in the figure, the number of clustering centers is determined as 3.

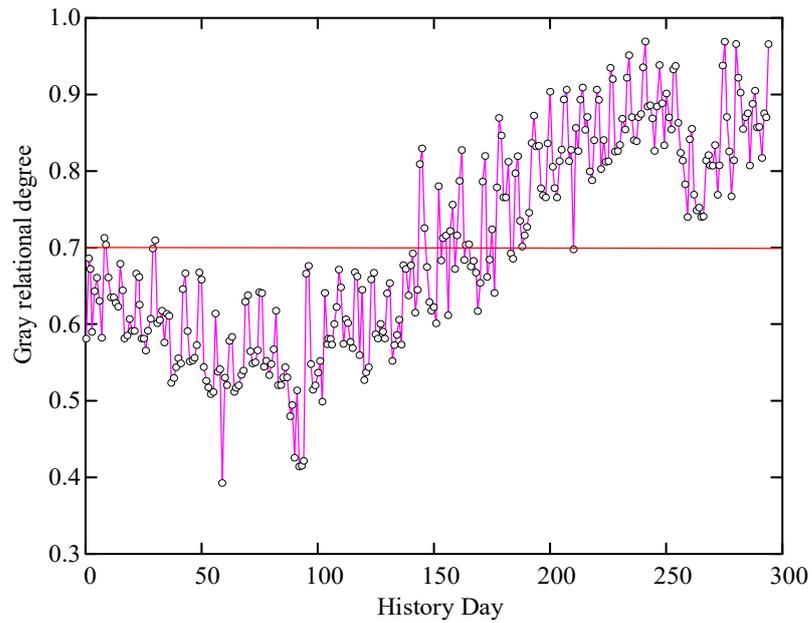


Figure 5. The grey correlation between the historical days and the days to be predicted.

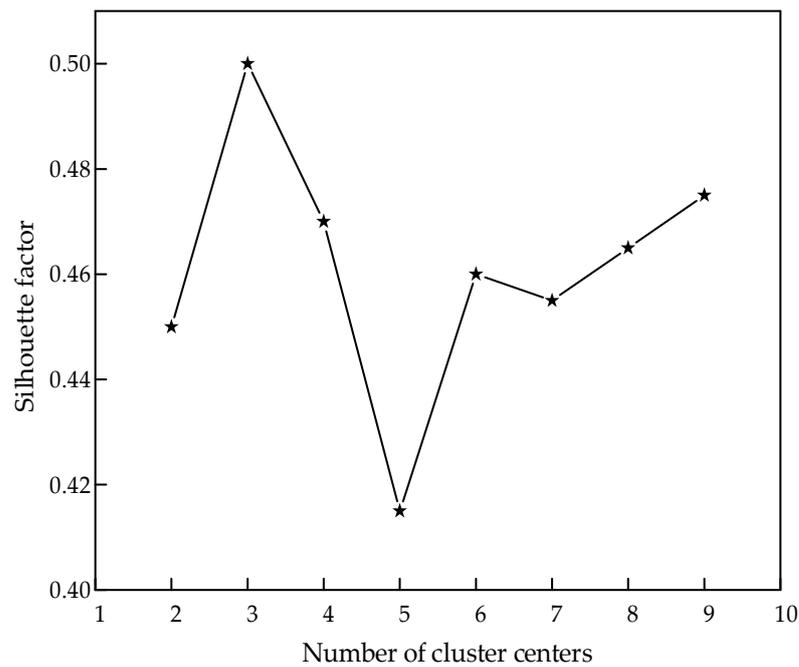


Figure 6. The correspondence between SIL and the number of clustering centers.

The clustering effect of each historical day within the group of comparable days is shown in Figure 7, with 233 data points, and some data points are not shown because of blocking each other. The Euclidean distances between the influences of the day to be predicted and each clustering center are shown in Table 8. From Table 8, it can be seen that the Euclidean distance between the to-be-predicted day and the center of the second category of clusters is smaller, and this category contains 85 historical days' data. Therefore, this category of historical days can be used as the final set of similar days.

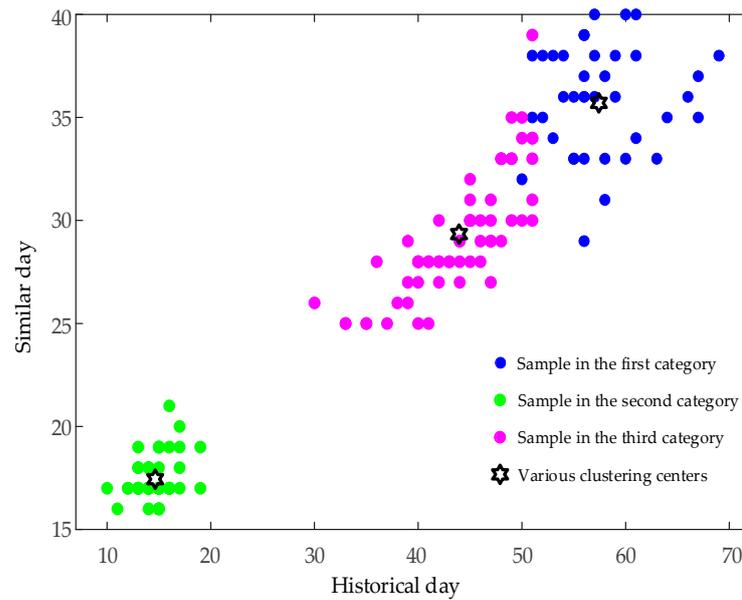


Figure 7. The clustering effect of each historical day in the rough set of similar days.

Table 8. Euclidean distance between the date to be predicted and the center of the cluster.

Category	Category 1	Category 2	Category 3
Euclidean distance	25.053	0.641	12.761

5.4. Correlation Analysis of Each Machine Learning Model

In order to select the best prediction model to compare with the combined VMD-HHO-KELM model based on GRA proposed above, experiments were designed to compare and analyze the prediction error distribution generated by each single machine learning model for harmonic prediction, respectively, taking Pearson correlation coefficient as the correlation index, and the error correlation analysis of each algorithm is shown in Figure 8.

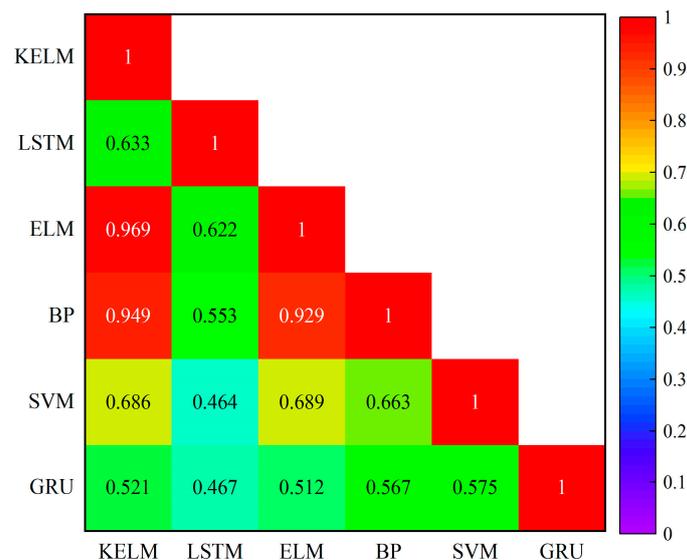


Figure 8. Correlation analysis of forecasting error for each model.

From Figure 8, it can be seen that the error correlation of each algorithm is generally high, which is due to the unavoidable error of the original data itself caused by the model

training. Among these, the error correlation of BP, ELM, and KELM algorithms is the highest, which is due to the fact that the three algorithms are generally improved based on feed forward neural networks trained by gradient descent, although the network structure and training mechanism are somewhat different. Among LSTM, SVM, and GRU, the difference between the SVM and LSTM, GRU training mechanism is larger, so the error correlation is lower. Although GRU is a variant of LSTM, and both belong to the same RNN improvement, compared with LSTM, GRU combines the forgetting gate and input gate into a single update gate, and also mixes the cell state and hidden state, which simplifies the model structure, so the error correlation of both is also lower. Therefore, several algorithms of BP, ELM, KELM, LSTM, and GRU are selected as the comparison algorithms for the combined model.

### 5.5. Analysis of Prediction Results

To evaluate the viability of the models chosen for this paper, the original data of 915 data, with a similar day set to 85 historical day data totaling 1000 data, were decomposed by VMD, with the obtained subsequence as the input data. The VMD decomposition effect is shown in Figure 9. Overall, 80% of the historical day before is selected as the training set and 20% of the historical day data as the test set, and several methods can be programmed and analyzed by examples in this paper. (1) Method 1: using the similar day set as the input data, optimizing the ELM hyper-parameters using the HHO algorithm, and constructing the optimized ELM model, which will be defined as HHO-ELM. (2) Method 2: using the similar day set as the input data, optimizing the KELM hyper-parameters using the HHO algorithm, and constructing the optimized KELM model, which will be defined as HHO-KELM. (3) Method 3: the subseries of the similar day set after VMD decomposition is used as the input data, the KELM hyper-parameters are optimized using the HHO algorithm, and the optimized KELM model is constructed, and the method is defined as VMD-HHO-KELM. (4) Method in this paper: the similar day set is constructed using the GRA method, and then the similar days are screened again using K-means clustering to form final similar day collection. The predicted values of the test set of several methods are compared with the actual values, as shown in Figure 10. The errors of predicted and actual values of the test sets of several methods are shown in Figure 11. The numerical evaluation indexes of each of the several models are shown in Table 8.

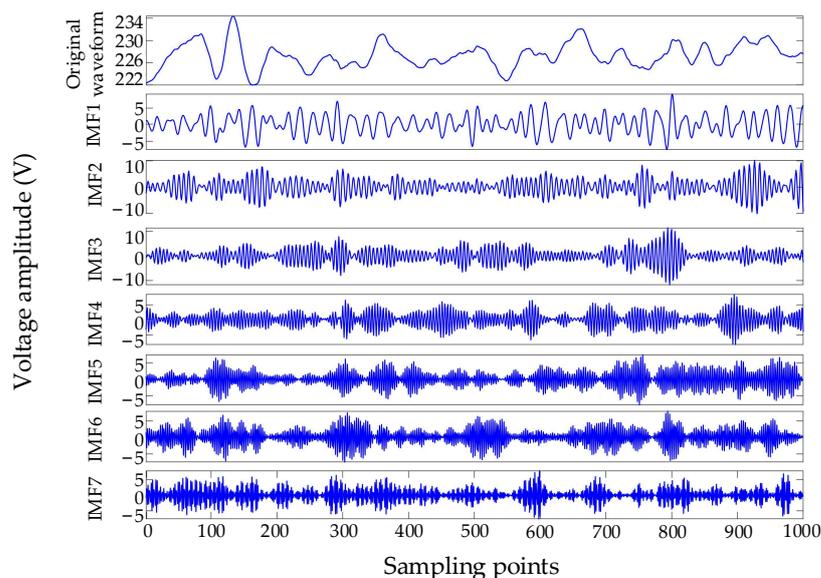
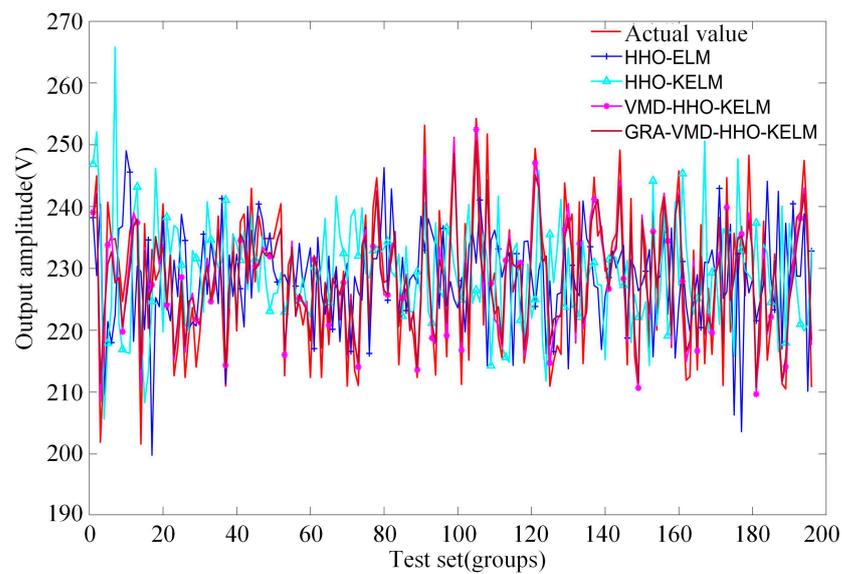
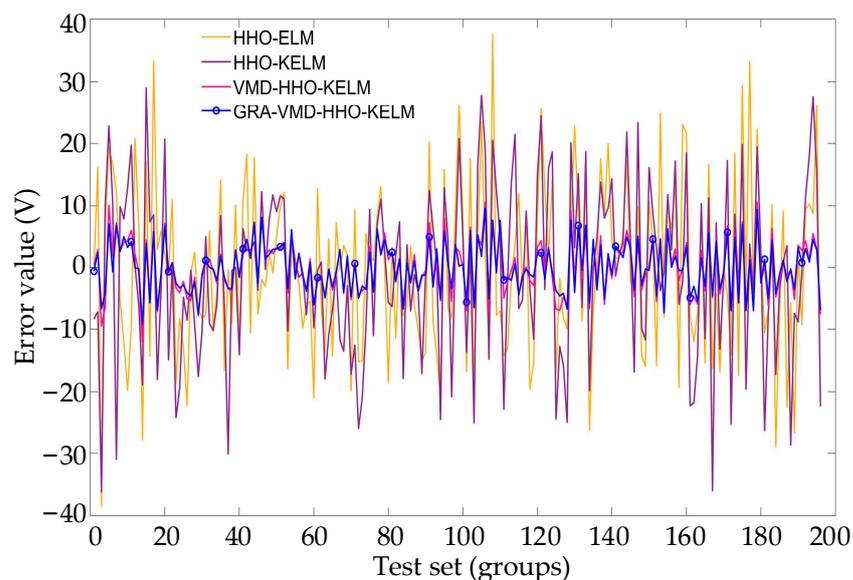


Figure 9. Similar day set VMD decomposition effect.



**Figure 10.** Comparison of predicted and actual values of each model test set.



**Figure 11.** Comparison of error values for each model test set.

By analyzing Figures 11 and 12, and Table 9, it can be seen that the HHO-KELM prediction model reduces 2.436 V, 1.9136 V, 0.8446%, and improves 0.11477 in  $RMSE$ ,  $MAE$ ,  $MAPE$ , and  $R^2$ , respectively, compared with the HHO-ELM prediction model, while the GRA-VMD-HHO-KELM model proposed in this paper does not improve significantly in terms of  $RMSE$ ,  $MAE$ ,  $MAPE$ , and  $R^2$ . The GRA-VMD-HHO-KELM model is not significant compared with VMD-HHO-KELM, but in terms of  $RMSE$ , the value is nearly one-third of HHO-ELM model, which is an order of magnitude improvement, and nearly twice the value of HHO-KELM; in terms of  $MAE$ , the value is nearly one-fourth of the HHO-ELM model, which is an order of magnitude improvement compared with the HHO-ELM model. In terms of  $MAPE$ , the value is one quarter of that of the HHO-ELM model, which is an order of magnitude higher and nearly two times smaller than that of the HHO-KELM; in terms of the  $R^2$ , the GRA-VMD-HHO-KELM model proposed in this paper is significantly more accurate than other models in terms of harmonic signal prediction. Compared with the HHO-ELM model, it improves by 0.234, and compared with the HHO-KELM, it improves by 0.119; in general, the numerical indexes of the model in this paper are

improved substantially compared with other indexes and, in the vertical comparison, the three combined algorithm models are better than the two combined models, which proves that the GRA-VMD algorithm weakens the non-smoothness of time series and reduces the complexity of various component complexity.

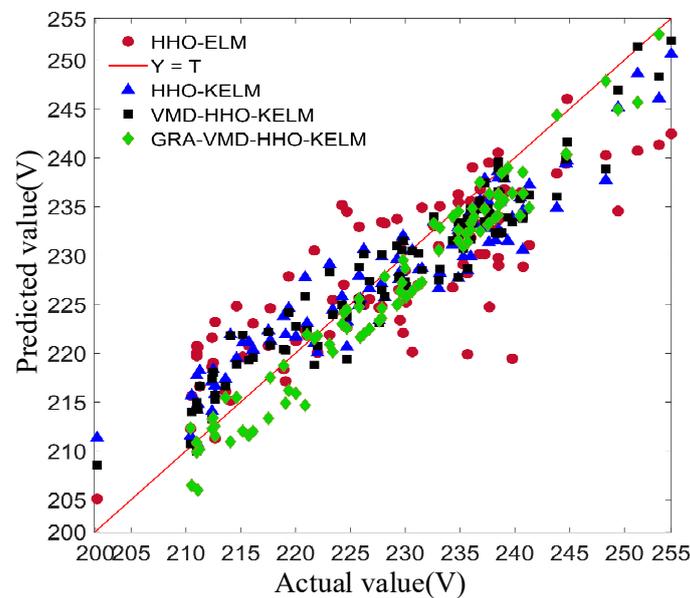


Figure 12. Scatter plot of predicted and actual values of the four models.

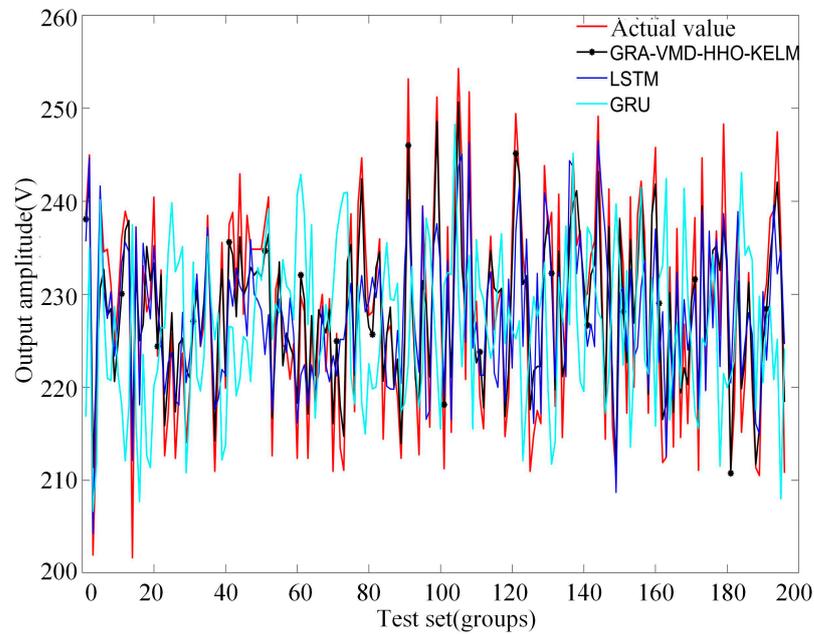
Table 9. Comparison table of evaluation indexes of each model.

Predictive Models	RMSE/V	MAE/V	MAPE/%	R <sup>2</sup>
HHO-ELM	13.8179	11.3103	4.9388%	0.73826
HHO-KELM	11.3819	9.3967	4.0942%	0.85303
VMD-HHO-KELM	4.4411	3.7312	1.6284%	0.94124
GRA-VMD-HHO-KELM	4.0149	3.3658	1.4701%	0.97283

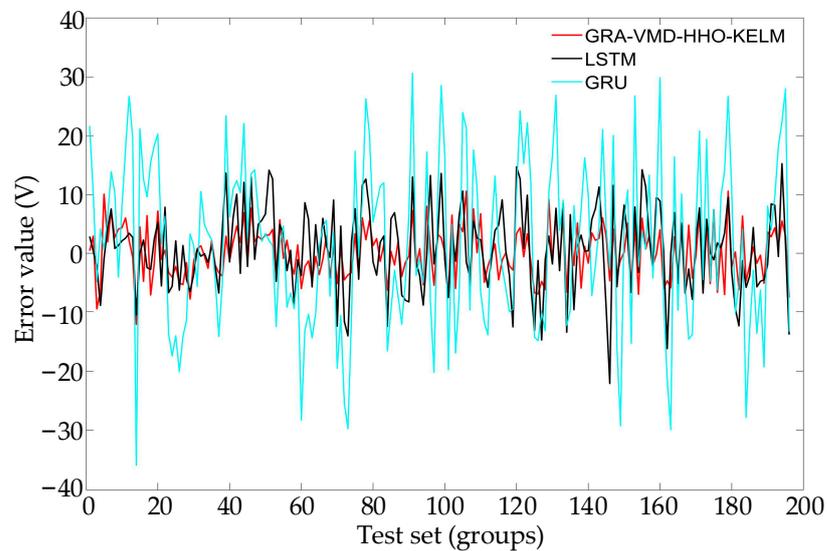
Meanwhile, in order to see the degrees of fit of the four methods more directly, the scatter plot between the expected and observed values of the four models is given, as shown in Figure 12. The diagonal line of  $Y = T$  in the figure indicates that the predicted value is equal to the actual value, and the closer the predicted value is to the actual value, the better the prediction effect, and the closer the points in the figure are to the diagonal line. In order to avoid the limitation of the neural network type for harmonic signal prediction, the proposed algorithm in this paper is compared with the improved Recurrent Neural Network (RNN), LSTM, and GRU mentioned in Section 5.4, and several model test set prediction effect comparison graphs, error value comparison graphs, and numerical evaluation index comparison tables are shown in Figures 13 and 14, as shown in Table 9.

Referring to Figure 14 and Table 10, the prediction model GRA-VMD-HHO-KELM proposed in this paper has the lowest values of RMSE, MAE, and MAPE and the highest value of R<sup>2</sup>. In terms of RMSE, the prediction model in this paper is reduced by 2.9865 V compared with LSTM and 9.9781 V compared with GRU. In terms of MAE, the prediction model in this paper reduces by 2.1851 V compared to LSTM and 8.2852 V compared to GRU, with the error decreasing nearly twice and the fitting accuracy improving by an order of magnitude; in terms of MAPE, the model in this paper is only half of LSTM and decreases nearly four times compared to GRU. In terms of R<sup>2</sup>, the model in this paper improves by 0.10603 compared to LSTM and 0.29013 compared to GRU, and the curve trend is more closely matched to the actual value. The error evaluation index shows that the GRA-VMD-HHO-KELM prediction model is established to have a good prediction

ability for the voltage harmonic signal of PV power plants, and the prediction errors are all relatively small, and can meet the function of the PV power plant harmonic content prediction. It is further verified that the accuracy of the GRA-VMD-HHO-KELM prediction model is better than other types of neural network models in harmonic signal prediction.



**Figure 13.** Comparison of predicted and actual values of LSTM, GRU—the model in this paper test set.



**Figure 14.** Comparison of the error values of this model, LSTM and GRU test set.

**Table 10.** Comparison table of models and LSTM and GRU evaluation metrics in this paper.

Predictive Models	RMSE/V	MAE/V	MAPE/%	R <sup>2</sup>
LSTM	7.0014	5.5509	2.4276%	0.8668
GRU	13.993	11.651	5.1099%	0.6827
GRA-VMD-HHO-KELM	4.0149	3.3658	1.4701%	0.97283

## 6. Conclusions

This paper adopts the GRA and *K*-means clustering method to screen and reconstruct the historical day data twice, and after eliminating the historical day data that are not highly correlated or irrelevant to the day to be predicted, the study constructs the similar day set, and then adopts VMD decomposition for the similar day set, and uses the obtained IMF components as input data, and adopts the HHO algorithm to search and solve for ELM/KELM hyper-parameters. The GRA-VMD-HHO-ELM/KELM power harmonic prediction model is constructed, and finally the power harmonic test set curve of the day to be predicted is obtained, and relevant numerical evaluation indexes are introduced to assess the effectiveness of the method proposed in this paper.

- (1) Similar day set construction: the Pearson correlation coefficient is used to solve the harmonic signal and each external factor Pearson coefficient value is used to screen the external factors that have greater influence on the harmonic value. The set of similar days is initially established, and then the *K*-means clustering method is used to classify the set of similar days, and the class including the center of the minimum Euclidean distance from the influencing factor of the day to be predicted is determined as the final set of similar days.
- (2) Improvement of ELM and KELM algorithms: The HHO algorithm is introduced for simulation comparison with other population intelligence optimization algorithms. The HHO algorithm converges about 15 times, and the fitness function finally converges to 0. Other algorithms require more than 30 times the convergence, and the convergence value is higher than 10. The results of the data comparison verify the excellent convergence speed and search efficiency of the HHO algorithm.
- (3) Construction of the GRA-VMD-HHO-KELM harmonic prediction model: In response to the problems of unclear spatiotemporal trend flow and difficulty in tracing the sources of harmonics in PV power plants, as well as the existing traditional prediction models with delay lag and poor prediction accuracy, this paper constructs a GRA-VMD-HHO-KELM harmonic prediction model, which can eliminate the harm caused by the pollution of each harmonic component and realize the function of the accurate prediction of different sub-harmonics. This model can eliminate the harm caused by the pollution of each harmonic component and realize the function of accurate prediction of different harmonic contents. The results show that the error of the prediction model is reduced by at least 39% compared with the conventional prediction method, so it can satisfy the function of harmonic content prediction of photovoltaic power plants.

Note that the proposed method can provide reliable data support and a theoretical basis for the subsequent harmonic detection, treatment and dynamic operation and the switching of reactive power compensation equipment. The application effect will be verified in future work.

**Author Contributions:** Conceptualization, Z.L.; methodology, G.Z. and R.G.; software, W.W.; validation, W.W.; formal analysis, Q.L.; investigation, D.W.; resources, Y.Z.; data curation, G.Z.; writing—original draft preparation, Z.L.; writing—review and editing, Q.L.; visualization, D.W.; supervision, Q.L.; project administration, Z.L.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Applied Foundational Research Plan Project of Liaoning Province (2022JH2/101300218).

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** Author Quanzheng Li was employed by the company State Grid Tieling Power Supply Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Nomenclature

$\Omega_{\text{ELM}}$	Kernel matrix.
$K(x, x_N)$	Kernel function.
$h(x)$	Implicit layer mapping relation function.
$\lambda$	Regularization coefficient.
$\sigma$	Kernel function parameters.
$\{u_k\}$	$K$ modal components obtained from the decomposition.
$\{\omega_k\}$	Frequency center of each component.
$\delta(t)$	Impulse function.
$f(t)$	Original signal.
$t$	Current number of iterations.
$E_0$	Random number in the interval $(-1, 1)$ .
$E$	Escape energy of Current prey.
$X(t + 1)$	Position vector of the Harris hawk at the next iteration.
$X(t)$	Current position vector of this hawk.
$X_{\text{rabbit}}(t)$	Prey position vector.
$X_{\text{rand}}(t)$	Position of random individuals in the hawk population.
$r_1, r_2, r_3, r_4, q$	Random numbers between $(0, 1)$ .
$N$	Number of individuals in the eagle population.
$X_i(t)$	Position of each eagle in iteration $t$ .
$\Delta X(t)$	Difference between the prey and the Harris hawk position when the current number of delivery is $t$ .
$J$	Random jump strength of the prey.
$\bar{f}$	Predicted value of the $i$ th sampling point.
$y_i$	Actual value of the $i$ th sampling point.
$\omega_i$	Weight in the ELM.
$b_i$	Bias in the ELM.
$f$	Minimum MSE for a single run of each method between the expected and observed values of KELM.
$f^*$	Minimum MSE among all minimum MSEs.
$X'_i$	Normalized value of historical data.
$X_i$	Historical original data.
$X_{\text{max}}, X_{\text{min}}$	Maximum and minimum values of voltage harmonics.

## References

- Liao, N.H.; Hu, Z.H.; Ma, Y.Y.; Lu, W.Y. Review of the short-term load forecasting methods of electric power system. *Power Syst. Prot. Control* **2011**, *39*, 147–152.
- Hao, E.S.; Han, Y.; Lin, X.Y.; Liu, E.P.; Yang, P.; Amr, S.Z. Harmonic characteristics and control strategies of grid-connected photovoltaic inverters under weak grid conditions. *Electr. Power Energy Syst.* **2022**, *142*, 108280–108300.
- Borges, C.E.; Peña, Y.K.; Fenandez, I. Evaluating combined load forecasting in large power system and smart grids. *IEEE Trans. Ind. Inform.* **2013**, *9*, 1570–1577. [[CrossRef](#)]
- Espinoza, M.; Joye, C.; Belmans, R.; Moor, B.D. Short-term power load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series. *IEEE Trans. Power Syst.* **2005**, *20*, 1622–1630. [[CrossRef](#)]
- Sharma, S.; Majumdar, A.; Elvira, V.; Chouzenoux, É. Blind kalman filtering for short-term power load forecasting. *IEEE Trans. Power Syst.* **2020**, *35*, 4916–4919. [[CrossRef](#)]
- Deng, Z.F.; Wang, B.B.; Xu, Y.L.; Xu, T.T.; Liu, C.X.; Zhu, Z.L. Multi-scale convolutional neural network with time-cognition for multi-step short-term power load forecasting. *IEEE Access* **2019**, *7*, 88058–88071. [[CrossRef](#)]
- Rafi, S.H.; Masood, N.A.; Deebea, S.R.; Hossain, E. A short-term power load forecasting method using integrated CNN and LSTM network. *IEEE Access* **2021**, *51*, 32436–32448. [[CrossRef](#)]
- Zhang, M.Y.; Han, Y.; Zalhaf, A.S.; Wang, C.Y.; Yang, P.; Wang, C.L.; Zhou, S.Y.; Xiong, T.L. Accurate ultra-short-term load forecasting based on load characteristic decomposition and convolutional neural network with bidirectional long short-term memory model. *Sustain. Energy Grids Netw.* **2023**, *35*, 101129. [[CrossRef](#)]
- Hu, L.J.; Guo, Z.Z.; Wang, Z.S. Short-term power load forecasting based on ISSA-LSSVM model. *Sci. Technol. Eng.* **2021**, *21*, 9916–9922.
- Mai, H.K.; Xiao, J.H.; Wu, X.C.; Chen, C. Research on ARIMA model parallelization in load prediction based on R language. *Power Syst. Technol.* **2015**, *39*, 3216–3220.

11. Han, H.Z.; Tang, Z.H. Wind speed prediction method based on CEEMDAN and echo state network. *Power Syst. Prot. Control* **2020**, *48*, 90–96.
12. Wang, Y.C.; Dou, Y.K.; Meng, R.Q. Multicore neural network short-term load forecasting model based on fuzzy C-mean clustering-variational model decomposition and chaotic swarm intelligence optimization. *High Volt. Eng.* **2022**, *48*, 1308–1319.
13. Chen, K.J.; Chen, K.L.; Wang, Q.; He, Z.Y.; He, J.L. Short-term load forecasting with deep residual networks. *IEEE Trans. Smart Grid* **2019**, *10*, 3943–3952. [[CrossRef](#)]
14. Farsi, B.; Amayri, M.; Bouguila, N.; Eicker, U. On short-term power load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach. *IEEE Access* **2021**, *9*, 31191–31212. [[CrossRef](#)]
15. Pham, M.H.; Nguyen, M.N.; Wu, Y.K. A novel short-term power load forecasting method by combining the deep learning with singular spectrum analysis. *IEEE Access* **2021**, *9*, 73736–73746. [[CrossRef](#)]
16. Tang, Z.H.; Zhao, G.N.; Cao, S.X. Very short-term wind direction prediction via self-tuning wavelet long-short term memory neural network. *Proc. CSEE* **2019**, *39*, 4459–4467.
17. Xiang, L.; Li, J.X.; Wang, P.H. Wind speed multistep interval forecasting based on VMD-FIG and parameter-optimized GRU. *Acta Energetica Solaris Sin.* **2021**, *42*, 237–242.
18. Zhu, L.J.; Xun, Z.H.; Wang, Y.X.; Cui, Q.; Chen, W.Y.; Lou, J.C. Short-term power load forecasting based on CNN-BiLSTM. *Power Syst. Technol.* **2021**, *45*, 4532–4539.
19. Zhao, P.; Dai, Y.M. Power load forecasting of SVM based on real-time price and weighted grey relational projection algorithm. *Power Syst. Technol.* **2020**, *44*, 1325–1332.
20. Wu, Y.; Lei, J.W.; Bao, L.S. Short-term load forecasting based on improved grey relational analysis and neural network optimized by bat algorithm. *Autom. Electr. Power Syst.* **2018**, *42*, 67–72.
21. Hu, Z.; Han, Y.; Zalhaf, A.S.; Zhou, S.Y.; Zhao, E.S.; Yang, P. Harmonic sources modeling and characterization in modern power systems: A comprehensive overview. *Electr. Power Syst. Res.* **2023**, *218*, 109234–109259. [[CrossRef](#)]
22. Sun, S.Z.; Liu, Z.W.; Zhang, H.; Yu, J.T.; He, Z.Y. Decoupling of FBG flow and temperature composite sensing based on HHO-KELM. *Opt. Precis. Eng.* **2022**, *30*, 1290–1300. [[CrossRef](#)]
23. Huang, G.B. An insight into extreme learning machines: Random neurons, random features and kernels. *Cogn. Comput.* **2014**, *6*, 376–390. [[CrossRef](#)]
24. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
25. Dragomiretskiy, K.; ZOSSO, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
26. Heidari, A.A.; Mirjalili, S.; Faris, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [[CrossRef](#)]
27. Kong, L.; Nian, H. Fault detection and location method for mesh-type DC microgrid using pearson correlation coefficient. *IEEE Trans. Power Deliv.* **2021**, *36*, 1428–1439. [[CrossRef](#)]
28. Kong, X.Y.; Li, C.; Zheng, F. Improved deep belief network for short-term power load forecasting considering demand-side management. *IEEE Trans. Power Syst.* **2020**, *35*, 1531–1538. [[CrossRef](#)]
29. Huang, D.M.; Zhuang, X.K.; Hu, A.D.; Sun, J.Z.; Shi, S.; Sun, Y.; Tang, Z. Short-term load forecasting based on similar-day selection with GRA-K-Means. *Electr. Power Constr.* **2021**, *42*, 110–117.
30. Xi, Y.W.; Wu, J.Y.; Shi, C.; Zhu, X.W.; Cai, R. A refined load forecasting based on historical data and real-time influencing factors. *Power Syst. Prot. Control* **2019**, *47*, 80–87.

**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.