

Article

Short-Term Electrical Load Forecasting Using an Enhanced Extreme Learning Machine Based on the Improved Dwarf Mongoose Optimization Algorithm

Haocheng Wang , Yu Zhang * and Lixin Mu

College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China; wanghaocheng@nefu.edu.cn (H.W.); mulixin@nefu.edu.cn (L.M.)

* Correspondence: zhangyu0902@163.com; Tel.: +86-13804614102

Abstract: Accurate short-term electrical load forecasting is crucial for the stable operation of power systems. Given the nonlinear, periodic, and rapidly changing characteristics of short-term power load forecasts, this paper introduces a novel forecasting method employing an Extreme Learning Machine (ELM) enhanced by an improved Dwarf Mongoose Optimization Algorithm (Local escape Dwarf Mongoose Optimization Algorithm, LDMOA). This method addresses the significant prediction errors of conventional ELM models and enhances prediction accuracy. The enhancements to the Dwarf Mongoose Optimization Algorithm include three key modifications: initially, a dynamic backward learning strategy is integrated at the early stages of the algorithm to augment its global search capabilities. Subsequently, a cosine algorithm is employed to locate new food sources, thereby expanding the search scope and avoiding local optima. Lastly, a “madness factor” is added when identifying new sleeping burrows to further widen the search area and effectively circumvent local optima. Comparative analyses using benchmark functions demonstrate the improved algorithm’s superior convergence and stability. In this study, the LDMOA algorithm optimizes the weights and thresholds of the ELM to establish the LDMOA-ELM prediction model. Experimental forecasts utilizing data from China’s 2016 “The Electrician Mathematical Contest in Modeling” demonstrate that the LDMOA-ELM model significantly outperforms the original ELM model in terms of prediction error and accuracy.

Keywords: electrical load forecasting; machine learning; extreme learning machine; dynamic backward learning; madness factor operator



Citation: Wang, H.; Zhang, Y.; Mu, L. Short-Term Electrical Load Forecasting Using an Enhanced Extreme Learning Machine Based on the Improved Dwarf Mongoose Optimization Algorithm. *Symmetry* **2024**, *16*, 628. <https://doi.org/10.3390/sym16050628>

Academic Editor: Peng-Yeng Yin

Received: 15 April 2024

Revised: 11 May 2024

Accepted: 15 May 2024

Published: 18 May 2024



Copyright: © 2024 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

Accurate electric load forecasting is crucial for the planning and reliable economic operation of power systems. It not only ensures the normal electricity usage of consumers but also reduces costs and guarantees the safety of power systems [1]. However, challenges in predicting electric load demand have increased sharply due to factors such as global climate change, energy supply constraints, increasing numbers of electricity users, and the integration of new energy device loads into the grid [2].

In the realm of electric load forecasting, researchers have employed time series regression models [3] and fuzzy linear regression models [4] to predict load, focusing on the temporal characteristics of load data and providing strong interpretability of the models. However, these methods have limitations in forecasting non-linear load data. In recent years, with the development of intelligent optimization algorithms, an increasing number of researchers have started incorporating these algorithms into electric load forecasting. Han M. C. enhanced the capability and prediction accuracy of capturing characteristics in load data by optimizing LSTM hyperparameters through a sparrow optimization algorithm that integrates Cauchy mutation and inverse learning strategies [5]. Zhang Z. C. quantified the behavior of dragonflies in the dragonfly algorithm to boost the search capability, and

utilized an adaptive noise complete empirical mode decomposition method for preprocessing raw data, thereby improving the prediction accuracy of SVR in load forecasting [6]. Ge Q. B. employed K-means clustering to categorize data and then used a combined predictive algorithm of reinforcement learning and particle swarm optimization along with the least squares support vector machine to predict different types of data [7]. Fan G. F. developed a new model combining the random forest model and mean-generating function model, significantly enhancing the prediction accuracy of peaks and troughs in highly volatile data [8]. Additionally, Xu R. [9] noted that extreme learning machines offer faster learning speeds and less human intervention, and are easier to implement. Deng B. [10] argued that compared to support vector machines, extreme learning machines have milder optimization constraints and quicker learning speeds. Some researchers have also applied optimized ELMs to electric load forecasting. For instance, Wang Tong utilized an improved artificial hummingbird algorithm for optimizing parameters in Extreme Learning Machines (ELM), significantly enhancing prediction accuracy [11]. Long Gan and others have used an improved multiverse algorithm to optimize the input layer weights and thresholds of ELMs, thereby improving their prediction accuracy [12]. Wang Z-X. employed an adaptive evolutionary ELM for data prediction, integrating a chaos-adapted whale optimization algorithm based on a firefly perturbation strategy and a chaotic sparrow search algorithm, which exhibited outstanding performance [13]. Additionally, Zhang S. proposed an ELM model under a moth flame optimization algorithm based on Tsne dimensionality reduction and visualization analysis, which achieved higher prediction accuracy than the original ELM model [14].

Accurate electric load forecasting can impact related decisions in power systems, such as generation control, economic dispatch, and maintenance scheduling. Therefore, to achieve high-precision short-term electric load forecasting, this paper proposes an ELM prediction model based on the improved Dwarf Mongoose Optimization Algorithm. Applied to short-term electric load forecasting, experimental results demonstrate that this model achieves higher accuracy compared to other ELM models.

2. Extreme Learning Machine

The Extreme Learning Machine (ELM) is a type of Single-hidden Layer Feedforward Neural Network (SLFN) algorithm, introduced by Professor Guang-Bin Huang and others based on the theory of the Moore–Penrose pseudoinverse [15]. This algorithm was developed to address several issues inherent in SLFNs, such as slow learning rates, long iteration times, and the traditional need to preset learning rates and step sizes. Unlike conventional neural network learning algorithms, the ELM requires only the appropriate setting of hidden layer node numbers. It autonomously generates all necessary parameters for the hidden layer and determines the final output layer weights through the least squares method. Due to its superior learning and nonlinear approximation capabilities compared to traditional machine learning algorithms, researchers have applied the ELM across a broad range of fields, including fault diagnosis [16], load forecasting [17], and feature recognition [18].

The ELM algorithm operates with a single hidden layer, where each layer from input to output comprises independent neurons, all interconnected in a fully connected manner. The network structure of the ELM is illustrated in Figure 1.

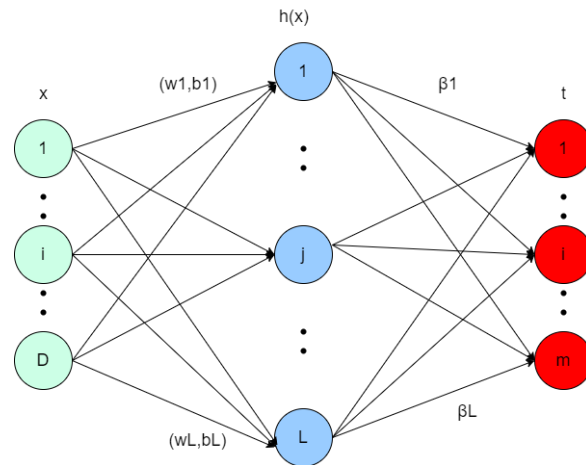


Figure 1. Extreme learning machine network structure.

Assuming there are N arbitrary samples (x_i, t_i) , $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$. This can be represented by a single hidden layer neural network with L hidden nodes, as illustrated in Figure 1 and described by Equation (1).

$$o_j = \sum_{i=1}^L \beta_i g(w_i \cdot x_i + b_i) \quad j = 1, 2, \dots, N \quad (1)$$

Here, $g(x)$ is the activation function, $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$ are the input weights, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T \in R^m$ are the output weights, and b_i is the bias of the i th hidden layer unit, with $w_i \cdot x_i$ being the dot product between them.

Extreme Learning Machines, as a type of single hidden layer neural network, have an output error that asymptotically approaches zero, as shown in Equation (2).

$$\sum_{j=1}^L ||o_j - t_i|| = 0 \quad (2)$$

Equation (2) can be represented using matrices, as shown in Equation (3):

$$H\beta = T \quad (3)$$

Here, H denotes the hidden layer output matrix, β represents the output weights, and T is the target output. The matrix H can be expressed by Equation (4).

$$H = \begin{bmatrix} h_1(x_1) & \dots & h_L(x_1) \\ \vdots & & \vdots \\ h_1(x_D) & \dots & h_L(x_D) \end{bmatrix}, \quad T = [t_1, \dots, t_D] \quad (4)$$

According to Equation (3), Equation (5) can be derived.

$$\beta = H^T T \quad (5)$$

In Equation (5), H^T represents the Moore–Penrose pseudoinverse of the matrix H .

3. Dwarf Mongoose Algorithm

The Dwarf Mongoose Algorithm is an intelligent optimization algorithm inspired by the social behavior of dwarf mongoose groups. This algorithm consists of three parts: the Alpha Group, the Scout Group, and the Babysitter Group. The Alpha Group produces a female leader to guide the group in foraging. The Scout Group is responsible for finding

new locations for sleeping mounds, while the Babysitter Group influences the performance of the algorithm through its numbers.

3.1. Alpha Group

The population is initialized as shown in Equation (6),

$$x_{i,j} = \text{unifrnd}(LB, UB, Dim) \quad (6)$$

where $x_{i,j}$ represents the initial position, LB and UB denote the lower and upper bounds of the solution space, Dim represents the dimension of decision variables, and unifrnd is a uniformly distributed random number.

Furthermore, a female leader emerges within the dwarf mongoose population, as depicted in Equation (7).

$$\alpha = \frac{fit_i}{\sum_{i=1}^N fit_i} \quad (7)$$

Here, fit_i represents the fitness value of the i th individual, and N is the number of individuals in the population. The number of individuals in the Alpha Group is the total population N minus the number of individuals in the Babysitter Group, i.e., $n = N - bs$.

The female leader in the Alpha Group guides the other members to the food source location via calls, as shown in Equation (8):

$$X_{new} = X_i + \text{peep} \times \text{phi} \times (X_i - X_k) \quad (8)$$

Here, X_{new} is the new position of the dwarf mongoose, peep is the calling coefficient, set at $\text{peep} = 2$ in this study, and phi is a uniformly distributed random number within $[0, 1]$. X_i is the current position of the female leader, and X_k is the position of another random individual in the Alpha Group distinct from the leader. Subsequently, the new position X_{new} undergoes a fitness evaluation to obtain fit_{i+1} , and the value of the sleeping mound is determined according to Equation (9).

$$sm_i = \frac{fit_{i+1} - fit_i}{\max\{|fit_{i+1}, fit_i|\}} \quad (9)$$

Here, sm_i represents the value of the sleeping mound, and the average value of the sleeping mound can be calculated according to Equation (10).

$$\varphi = \frac{\sum_{i=1}^N sm_i}{n} \quad (10)$$

3.2. Scout Group

The primary responsibility of the Scout Group is to locate new positions for sleeping mounds, as described by the movement formula in Equation (11).

$$X_{i+1} = \begin{cases} x_i - CF \times \text{phi} \times r \times \left[x_i - \vec{M} \right], & \text{if } \varphi_{i+1} > \varphi_i \\ x_i + CF \times \text{phi} \times r \times \left[x_i - \vec{M} \right], & \text{else} \end{cases} \quad (11)$$

Here, r is a random number within $[0, 1]$, CF is a mobility parameter for the dwarf mongoose population, and $\vec{M}$ is the direction vector determining the mongoose's movement direction. The formulas for calculating CF and $\vec{M}$ are given in Equations (12) and (13), respectively.

$$CF = \left(1 - \frac{Iter}{MaxIt}\right)^{(2 \frac{Iter}{MaxIt})} \quad (12)$$

$$\vec{M} = \sum_{i=1}^N \frac{x_i \times sm_i}{x_i} \quad (13)$$

Here, $Iter$ is the current iteration number, and $MaxIt$ is the maximum number of iterations. The movement parameter CF linearly decreases as the number of iterations increases.

3.3. Babysitter Group

When the timing parameter is greater than or equal to the exchange parameter, i.e., $C \geq L$, the Babysitter Group assumes that the Alpha Group's foraging capability is weak. At this point, the Babysitter Group will swap roles with the Alpha Group, and the dwarf mongoose community will begin searching for a new sleeping mound.

4. Enhancements to the Dwarf Mongoose Algorithm

To address the Dwarf Mongoose Algorithm's tendency to fall into local optima and its weak global search performance, this paper proposes the incorporation of a reverse learning strategy to enhance the algorithm's exploratory capability, thereby improving its global search performance. Additionally, the inclusion of a craziness operator factor and the sine-cosine algorithm expands the local search range of the algorithm, helping to circumvent issues of local optima.

4.1. Dynamic Reverse Learning Strategy

The reverse learning strategy [19] is a common perturbation tactic that expands the algorithm's exploration range to find better solutions. In the reversed learning strategy, the new position generated is symmetrical to the original position at the point $\frac{X_i + x_i}{2}$. This measure enhances the exploratory nature of the algorithm, allowing it to search in the opposite direction for improved population individual positions. Moreover, the new position is compared with the original in terms of fitness, and the individual with the optimal fitness is selected as the population individual position. The reverse learning strategy is depicted in Equation (14).

$$X_i = LB + UB - x_i \quad (14)$$

In the formula, X_i represents the population individual after reverse learning, with LB and UB denoting the lower and upper limits of the solution space, and x_i indicating the original position of the individual within the population. To enhance the search for optimal solutions within the solution space, a random factor is included in the reverse learning strategy, further diversifying the population within the solution space. The dynamic reverse learning strategy introduced in the early stages of the algorithm iteration is shown in Equation (15).

$$X_i = r \times (LB + UB) - x_i \quad (15)$$

Here, r is a random number within (0, 1).

4.2. Sine-Cosine Algorithm

During the search process, the sine-cosine algorithm [20] conducts searches in the form of sine and cosine waveforms. This method enhances the search capabilities of the algorithm, enabling it to avoid becoming trapped in local optima. Additionally, the search process of the algorithm exhibits point symmetry characteristics typical of sine and cosine functions. While the Alpha Group, led by the female leader, is searching for new food sources, this process can easily become trapped in local optima. To avoid such outcomes, this study incorporates the sine-cosine algorithm to expand the search range of the algorithm. The formula for searching new food sources, updated with the sine-cosine algorithm, is shown in Equation (16).

$$X_{new} = \begin{cases} X_i + phi \times r_1 \times sin(r_2) \times (X_i - X_k), & \text{if } r < 0.5 \\ X_i + phi \times r_1 \times cos(r_2) \times (X_i - X_k), & \text{else} \end{cases} \quad (16)$$

In Equation (16), r and r_1 are random numbers within $(0, 1)$, r_2 is a random number within $(0, 2\pi)$, X_i is the position of the female leader, and X_k is the position of another individual distinct from the leader.

4.3. Craziness Factor

In the later stages of the algorithm, when dwarf mongoose individuals seek a sleeping mound, the group tends to converge on this mound, which could lead to local optima. This paper introduces a craziness operator factor, which perturbs the position of the optimal individual to prevent the algorithm from becoming trapped in local optima in its later iterations. The position of the optimal individual after incorporating the craziness operator factor is illustrated in Equation (17).

$$X_i = x_i \times (1 + Pc \times xcraze \times sign) \quad (17)$$

In Equation (17), x_i represents the original optimal individual position, and Pc and $xcraze$ are disturbance factors within the craziness operator factor, with $xcraze$ set at 0.0001. Pc and $xcraze$ are described by Equations (18) and (19), respectively.

$$Pc = \begin{cases} 1, & c < P_r \\ 0, & \text{else} \end{cases} \quad (18)$$

$$sign = \begin{cases} -1, & c > 0.5 \\ 1, & \text{else} \end{cases} \quad (19)$$

In the craziness factor, the $sign$ is determined as either 1 or -1 based on the magnitude of c , exhibiting a kind of symmetry in its values. This method of value assignment can perturb the algorithm, expanding its search range and helping to avoid local optima. In Equation (19), c is a random number within $(0, 1)$, and P_r is the preset craziness probability, set at 0.4 in this study.

4.4. LDMOA Steps and Process

In the enhanced DMOA, the reverse learning strategy is utilized to expand the algorithm's exploratory capacity and search range, thereby enhancing its global search capabilities. Simultaneously, the introduction of the sine–cosine algorithm and the craziness operator factor enhance the local search capability of the algorithm, effectively avoiding situations of local optima. Figure 2 shows the workflow diagram of the LDMOA, and below is the operational process of the LDMOA.

Step 1: Set the initial parameters of the algorithm, such as population parameters, dimensions of the solution space and its limits, and maximum iteration parameters, and utilize the dynamic reverse learning strategy to expand the search range.

Step 2: Select the female leader according to Equation (7) and set the related coefficients. During the process of searching for new food sources by the dwarf mongoose group, incorporate the sine–cosine algorithm to further expand the search for new food source positions.

Step 3: The sleeping mound position is influenced by the optimal position; introduce the craziness operator factor to perturb it, and then determine the sleeping mound position and calculate its average value.

Step 4: Assess whether $C \geq L$; when this condition is met, swap the Alpha Group and Babysitter Group, and proceed with the formula to search for new sleeping mounds and forage.

Step 5: Determine whether the algorithm has reached the maximum iteration count; if not, repeat the above steps, and otherwise output the optimal results.

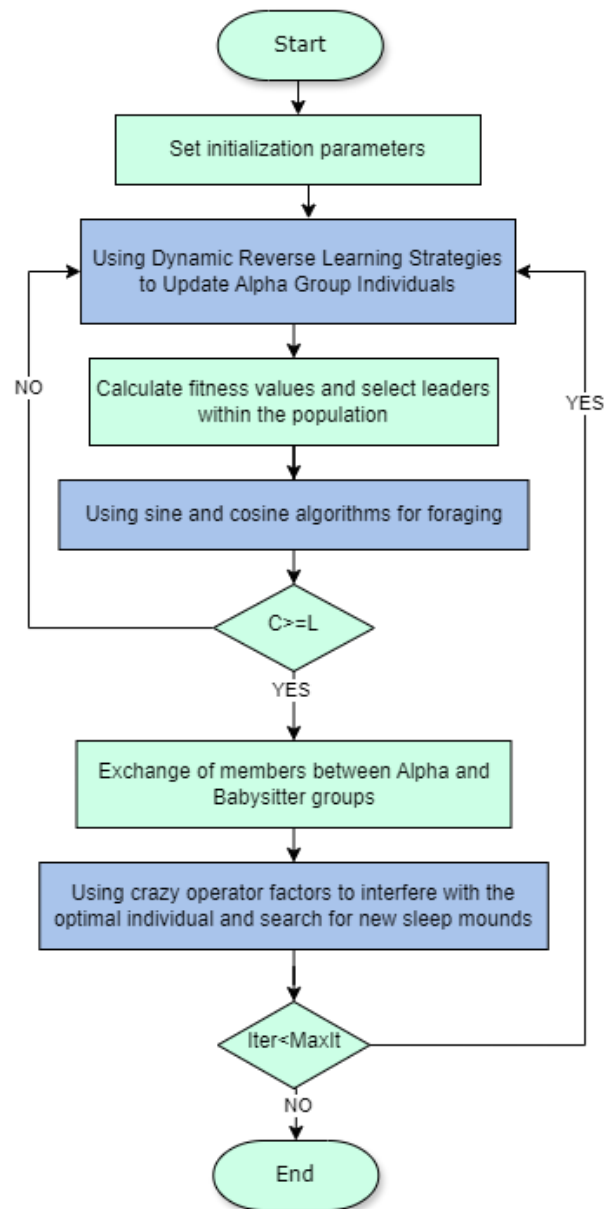


Figure 2. LDMOA Flowchart.

4.5. Benchmark Function Testing

To ensure that the enhanced strategy provides positive improvements over the original Dwarf Mongoose Optimization Algorithm (DMOA), we conducted benchmark function tests comparing the modified algorithm with the original. The selected benchmark functions are shown in Table 1.

Functions f_1 to f_2 are unimodal functions, which test the algorithm's convergence capability. Functions f_3 to f_6 are multimodal functions, evaluating the algorithm's ability to escape local optima. Functions f_7 to f_{11} are hybrid functions, and f_{12} to f_{15} are composite functions, with both sets testing the optimization performance in complex scenarios.

The enhanced and original algorithms were tested using the benchmark functions listed in the table. To ensure the accuracy of the benchmark tests, the algorithms were configured with parameters as shown in Table 2, including population initialization size ($nPop$) and solution space dimensions (Dim). The results are presented in Section 4.6 and Table 3.

Table 1. Benchmark function.

Function	Function Name	Optimal Value
f_1	Shifted and Rotated Bent Cigar Function	100
f_2	Shifted and Rotated Zakharov Function	300
f_3	Shifted and Rotated Rosenbrock's Function	400
f_4	Shifted and Rotated Lunacek Bi_Rastrigin Function	700
f_5	Shifted and Rotated Non-Continuous Rastrigin's Function	800
f_6	Shifted and Rotated Levy Function	900
f_7	Hybrid Function 2 (N = 3)	1200
f_8	Hybrid Function 3 (N = 3)	1300
f_9	Hybrid Function 4 (N = 4)	1400
f_{10}	Hybrid Function 6 (N = 4)	1600
f_{11}	Hybrid Function 6 (N = 5)	1900
f_{12}	Composition Function 1 (N = 3)	2100
f_{13}	Composition Function 3 (N = 4)	2300
f_{14}	Composition Function 5 (N = 5)	2500
f_{15}	Composition Function 9 (N = 3)	2900

Table 2. Parameter settings.

Algorithm	nPop	Dim.	Number of Runs	Number of Iterations
DMOA	50	30	30	500
LDMOA	50	30	30	500

Table 3. Function test result.

Function	Algorithm	Average Value	Standard Deviation
f_1	DMOA	2.2380×10^8	1.5191×10^8
	LDMOA	4.5751×10^6	4.7176×10^6
f_2	DMOA	3.8921×10^5	1.4859×10^5
	LDMOA	1.9088×10^5	3.0369×10^4
f_3	DMOA	6.4582×10^2	59.71001
	LDMOA	4.7798×10^2	25.6402
f_4	DMOA	9.8498×10^2	16.6132
	LDMOA	8.7634×10^2	16.2196
f_5	DMOA	1.050×10^3	15.3858
	LDMOA	9.4086×10^2	11.9651
f_6	DMOA	4.8974×10^3	1.1978×10^3
	LDMOA	2.5027×10^3	5.5686×10^2
f_7	DMOA	3.7911×10^8	1.5928×10^8
	LDMOA	2.1834×10^7	6.8054×10^6
f_8	DMOA	7.4329×10^6	5.6933×10^6
	LDMOA	4.0444×10^5	3.4691×10^5
f_9	DMOA	2.7983×10^5	1.1979×10^5
	LDMOA	2.8142×10^4	1.4802×10^4
f_{10}	DMOA	3.9829×10^3	2.3781×10^2
	LDMOA	2.9550×10^3	1.8911×10^2
f_{11}	DMOA	8.9163×10^4	8.4653×10^4
	LDMOA	2.5640×10^4	1.8000×10^4
f_{12}	DMOA	2.5562×10^3	18.1339
	LDMOA	2.4442×10^3	12.3400
f_{13}	DMOA	2.9142×10^3	17.9570
	LDMOA	2.8032×10^3	17.5984
f_{14}	DMOA	2.9628×10^3	20.6205
	LDMOA	2.8955×10^3	5.5951
f_{15}	DMOA	5.0032×10^3	1.9213×10^2
	LDMOA	3.7782×10^3	1.5826×10^2

4.6. Results of Benchmark Function Testing

In the benchmark function test results, the mean and standard deviation reflect the convergence performance and stability of the algorithms, respectively, with the best values highlighted in bold. The enhanced algorithm shows superior convergence performance and stability compared to the original algorithm. As demonstrated in Table 3, the enhanced algorithm outperforms the original in unimodal functions f_1 to f_2 , multimodal functions f_3 to f_6 , hybrid functions f_7 to f_{11} , and composite functions f_{12} to f_{15} . Based on the analyses in Section 4.5, the enhanced algorithm surpasses the original in convergence, avoiding local optima, and handling complex optimization problems. Figure 3 includes graphical representations of some function iterations: (a) unimodal function, (b) multimodal function, (c) hybrid function, and (d) composite function, all showing improved iterative convergence results for the enhanced algorithm.

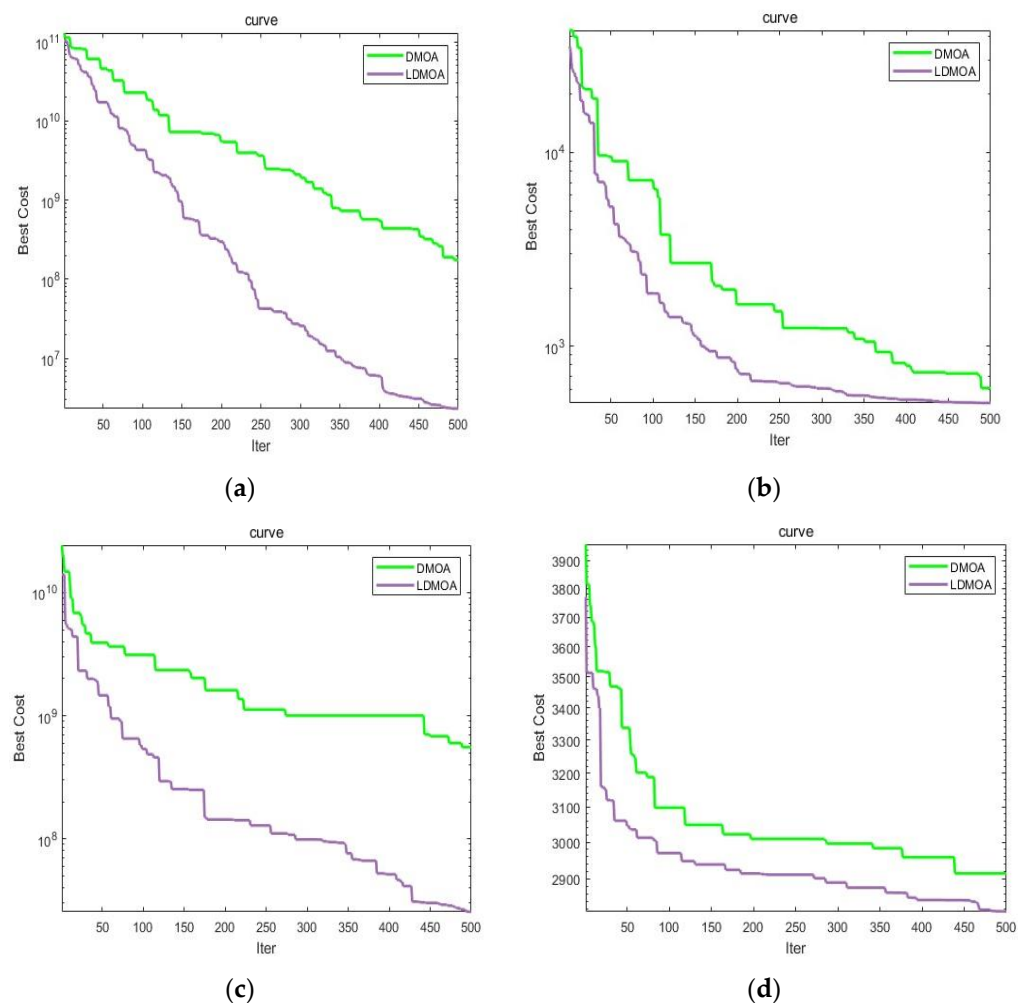


Figure 3. Images of selected test functions: (a) f_1 , (b) f_3 , (c) f_7 , (d) f_{13} .

5. Establishing the LDMOA-ELM Model

In the Extreme Learning Machine (ELM) algorithm, the weights w and biases b significantly influence prediction outcomes. Given that these parameters are randomly generated in ELM, their randomness can significantly affect the model's prediction accuracy. This study uses the LDMOA algorithm to optimize the weights w and biases b in the ELM algorithm, leading to notable improvements in prediction error reduction and accuracy enhancement. The LDMOA-ELM model development process is outlined as follows:

Step 1: Import original load data, normalize it, and split it into training and test sets.

Step 2: Initialize DMOA parameters and use Equations (15)–(17) to optimize the initial population, the Alpha Group's foraging process, and the process of finding new sleeping mounds, respectively, resulting in the LDMOA.

Step 3: Compute the fitness function, using the MAPE of the ELM training set as the fitness measure.

Step 4: Exit the loop if the maximum number of iterations is reached or accuracy requirements are met; otherwise, repeat Steps 2 and 3.

Step 5: Use the optimized parameters as the input weights and biases for the ELM model, and then perform numerical predictions and output the model evaluation metrics.

The steps for establishing the LDMOA-ELM model are depicted in Figure 4.

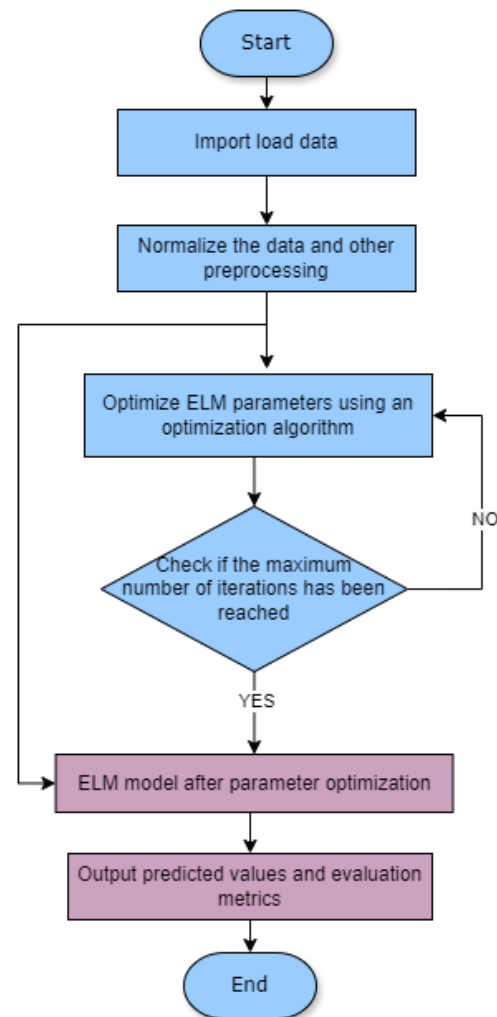


Figure 4. Establishment of the LDMOA-ELM model.

6. Simulation Experiment

In conducting electric load forecasting experiments for comparison, to ensure the efficacy of the improved algorithm, this study pits the proposed LDMOA-ELM algorithm against both the original ELM algorithm and the ELM algorithm optimized by the original Dwarf Mongoose Algorithm. The parameter settings for both the original and the enhanced Dwarf Mongoose Algorithms are detailed in Table 4, where $nPop$ is the population initialization size, Dim is the dimension of the solution space, LB is the lower bound, and UB is the upper bound of the solution space.

Table 4. Parameter settings.

Algorithm	nPop	Dim.	LB	UB	Number of Iterations
DMOA	50	30	−2	2	500
LDMOA	50	30	−2	2	500

6.1. Evaluation Metrics

The electric load forecasting evaluation standards include *MAE* (Mean Absolute Error), *RMSE* (Root Mean Square Error), *MSE* (Mean Square Error), and R^2 (R-Squared, the coefficient of determination) [21]. The formulas for these metrics are shown in Equations (20), (21), (22), and (23), respectively.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (20)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (21)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (22)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (23)$$

Here, $\hat{y}_i$ represents the hourly forecasted electric load; $\bar{y}_i$ is the average of the hourly electric load data; y_i is the actual hourly electric load data; and n is the number of data points.

In these metrics, *MSE* and *RMSE* values are within $[0, +\infty)$, where a value closer to 0 indicates perfect model prediction and, conversely, a higher value indicates greater prediction error. The *MAE* follows the same range and interpretation. An R^2 value closer to 1 indicates a better fit, whereas a lower value indicates a poorer fit.

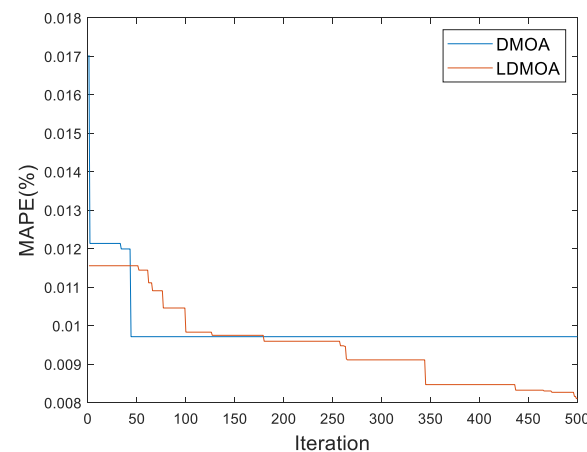
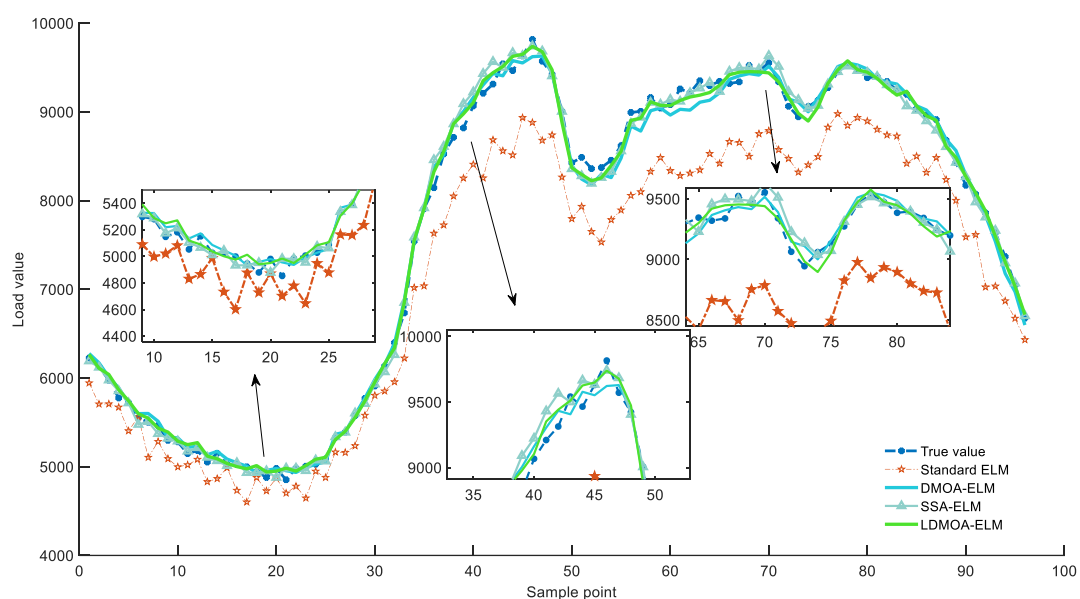
6.2. Forecasting Results Comparison

The paper utilizes the standard dataset provided by the 2016 “The Electrician Mathematical Contest in Modeling” in China, with sampling every 15 min, resulting in 96 samples per day and a total of 35,040 samples. To ensure consistency in experimental results, the number of hidden nodes in the prediction model is uniformly set to 85, with the sample configuration using data from the previous seven days to predict the eighth day, set across 100 sample groups. In this paper, we optimized the ELM model for multi-step-ahead forecasting by dividing 100 sample sets into 99 training sets and one test set. The statistical data of the three methods’ predictions are shown in Table 5. Section 6.1 demonstrates that a higher R^2 value indicates better prediction results, and smaller values for other metrics indicate better performance. Here, the LDMOA-ELM model’s *MAE*, *MAPE*, *MSE*, and *RMSE* values were 61.62, 0.0080845, 5953.4, and 77.158, respectively, all of which were significantly reduced compared to the ELM model. Integrating these five evaluation metrics, it is evident that the LDMOA-ELM used in this study exhibits superior performance across all indicators, with a prediction accuracy of 99.80%, which is an improvement of 15% over the ELM model. This demonstrates that the LDMOA-ELM model achieves lower prediction errors and higher prediction accuracy. The results indicate that the improved predictive model enhances the accuracy of forecasts for the experimental data used in this study.

Table 5. Evaluation metrics.

	MAE	MAPE	MSE	RMSE	R^2
ELM	527.35	0.069746	3.5048×10^5	592.02	0.86538
SSA-ELM	76.713	0.0097348	10,163	100.81	0.99659
DMOA-ELM	72.311	0.009243	9627.1	98.118	0.99689
LDMOA-ELM	61.62	0.0080845	5953.4	77.158	0.99802

As illustrated in Figure 5, LDMOA is more adept at escaping local optima and finding optimal values compared to the original Dwarf Mongoose Algorithm. Figure 6 shows the prediction results graph, indicating that the LDMOA-ELM model's prediction curve best fits the actual value curve. Figure 7 displays the relative prediction errors, revealing that the relative error between the predicted values of the LDMOA-ELM model and the actual values is significantly lower than that of the ELM model. Additionally, by integrating the statistical data from Table 5, it is evident that the LDMOA-ELM model exhibits lower prediction errors and improved accuracy compared to the original ELM model. Particularly in terms of relative prediction errors, as shown in Figure 7, there is a significant difference between the two, with the LDMOA-ELM model outperforming the original ELM model.

**Figure 5.** Iterative optimization.**Figure 6.** Prediction results.

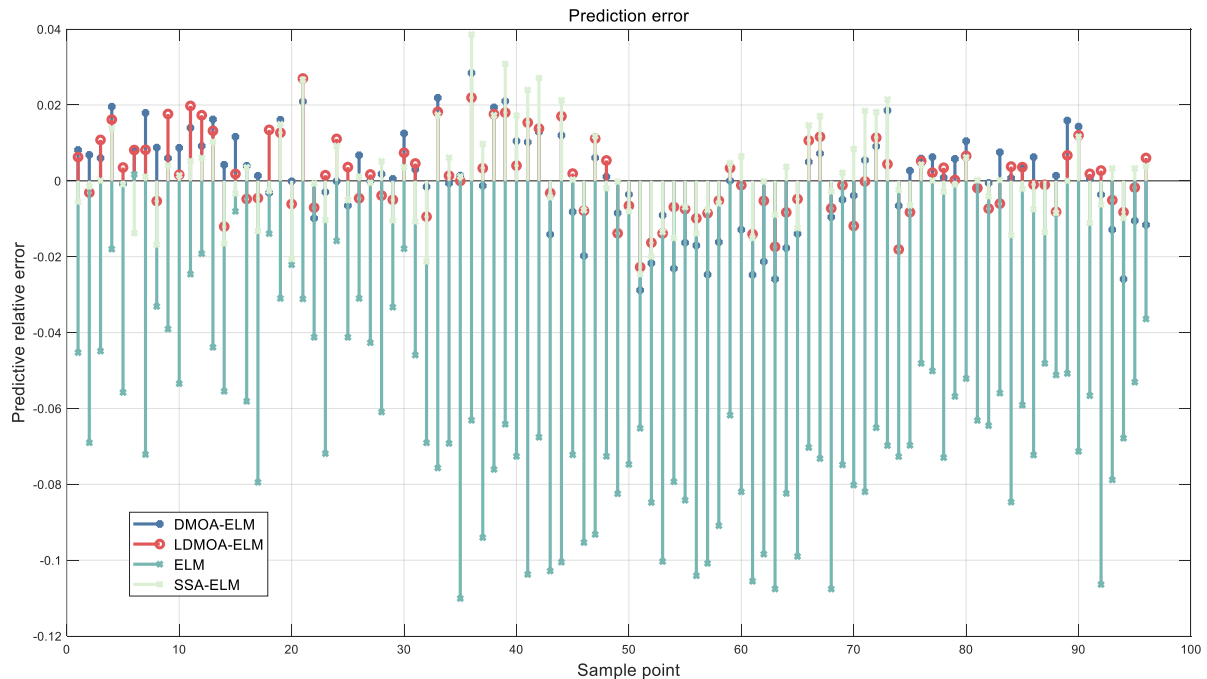


Figure 7. Relative error between predicted and actual load.

Figure 8, the prediction evaluation metrics graph, shows that the ELM model results in the highest error values, while the LDMOA-ELM model yields the smallest error values. Integrating data from Table 5 and Figures 5–8, it is evident that the LDMOA-ELM algorithm, in comparison to both the ELM and DMOA-ELM algorithms, achieves the smallest prediction errors and the highest prediction accuracy, with the LDMOA-ELM algorithm achieving a prediction accuracy of 99.80%. Figure 9 presents the statistical graph for the Mean Absolute Percentage Error (MAPE). It is evident that the original ELM model exhibits significantly higher MAPE values compared to the LDMOA-ELM model. This demonstrates that the LDMOA-ELM model achieves lower prediction errors.

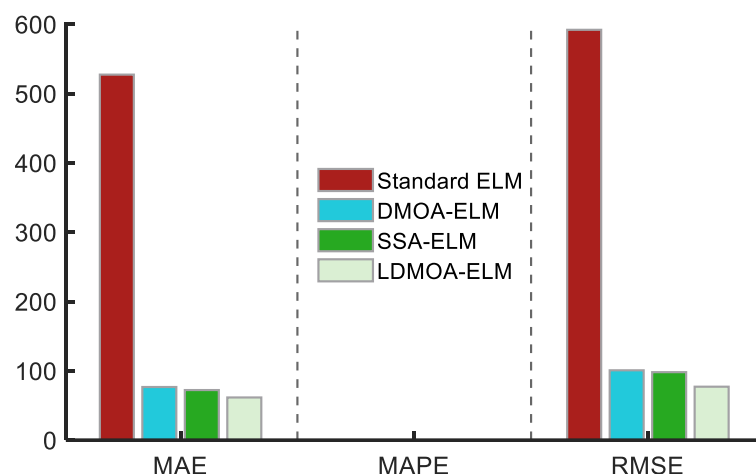


Figure 8. Evaluation metrics.

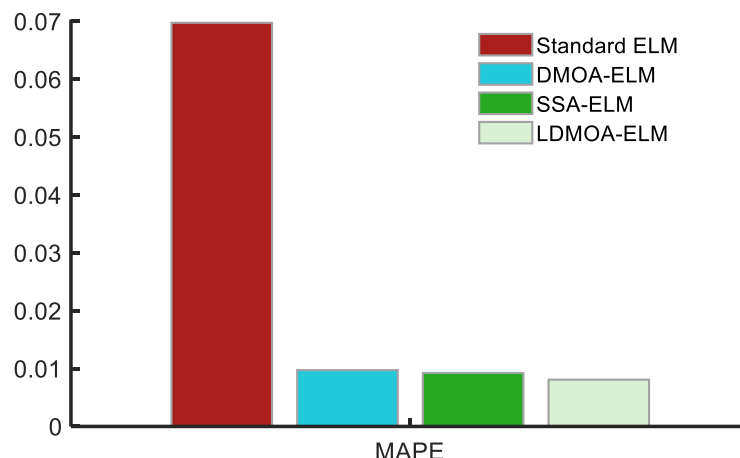


Figure 9. Mean absolute percentage error.

7. Conclusions

In response to the challenges of high randomness and low prediction accuracy in short-term electric load forecasting, this paper introduces a short-term electric load forecasting model that utilizes an enhanced Dwarf Mongoose Algorithm-based ELM. Initially, the Dwarf Mongoose Algorithm was modified by incorporating reverse learning strategies, sine–cosine strategies, and a craziness operator factor, which improved the algorithm’s exploratory capabilities, enhanced its global search ability, and enabled it to escape from local optima more effectively. Subsequently, combining the LDMOA with an Extreme Learning Machine, this model was applied to forecast the relevant experimental data and subjected to experimental analysis. The results demonstrate that, compared to the original ELM and DMOA-ELM models, the LDMOA-ELM model exhibits significantly higher accuracy in predicting short-term electric loads.

The LDMOA-ELM model proposed in this paper exhibits a Mean Absolute Error (MAE) of 61.62, which is significantly lower than that of other models, thereby reducing the prediction error of the Extreme Learning Machine to a certain extent. Although the accuracy of the LDMOA-ELM model surpasses that of the original ELM model, the improvement in predictive accuracy is not markedly evident when compared with other models. Future research should focus on further refining the optimization algorithms and selecting more appropriate predictive data to verify the accuracy and applicability of the predictive model. At the same time, subsequent research projects should consider the processing of raw data and comparisons between different methodologies.

Furthermore, future research in load forecasting could consider incorporating methods such as Variational Mode Decomposition for preprocessing the raw data and compare it with other machine learning prediction methods to highlight the advanced nature of the optimized model. It could also be beneficial to explore the impact of varying the number of nodes in the prediction algorithm on the precision of the forecasts, aiming to achieve better load prediction outcomes.

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

Funding: This research received no external funding.

Data Availability Statement: This article does not contain new data; the results of the experimental data are presented within the article.

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

References

1. Wang, L.; Lin, Y.; Tong, H.; Li, H.; Zhang, T. Short-term load forecasting based on improved Apriori correlation analysis and an MFOLSTM algorithm. *Power Syst. Prot. Control* **2021**, *49*, 74–81.
2. Xie, X.; Zhou, J.; Zhang, Y.; Wang, J.; Su, J. W-BiLSTM Based Ultra-short-term Generation Power Prediction Method of Renewable Energy. *Autom. Electr. Power Syst.* **2021**, *45*, 175–184.
3. Behmiri, N.B.; Fezzi, C.; Ravazzolo, F. Incorporating air temperature into mid-term electricity load forecasting models using time-series regressions and neural networks. *Energy* **2023**, *278*, 127831. [\[CrossRef\]](#)
4. Al-Kandari, A.M.; Soliman, S.A.; El-Hawary, M.E. Fuzzy short-term electric load forecasting. *Int. J. Electr. Power Energy Syst.* **2004**, *26*, 111–122. [\[CrossRef\]](#)
5. Han, M.C.; Zhong, J.W.; Sang, P.; Liao, H.H.; Tan, A.G. A Combined Model Incorporating Improved SSA and LSTM Algorithms for Short-Term Load Forecasting. *Electronics* **2022**, *11*, 1835. [\[CrossRef\]](#)
6. Zhang, Z.; Hong, W.-C. Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. *Nonlinear Dyn.* **2019**, *98*, 1107–1136. [\[CrossRef\]](#)
7. Ge, Q.B.; Guo, C.; Jiang, H.Y.; Lu, Z.Y.; Yao, G.; Zhang, J.M.; Hua, Q. Industrial Power Load Forecasting Method Based on Reinforcement Learning and PSO-LSSVM. *IEEE Trans. Cybern.* **2022**, *52*, 1112–1124. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Fan, G.F.; Zhang, L.Z.; Yu, M.; Hong, W.C.; Dong, S.Q. Applications of random forest in multivariable response surface for short-term load forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *139*, 108073. [\[CrossRef\]](#)
9. Xu, R.; Liang, X.; Qi, J.-S.; Li, Z.-Y.; Zhang, S.-S. Advances and Trends in Extreme Learning Machine. *Jisuanji Xuebao/Chin. J. Comput.* **2019**, *42*, 1640–1670. [\[CrossRef\]](#)
10. Deng, B.; Zhang, X.; Gong, W.; Shang, D. An overview of extreme learning machine. In Proceedings of the 4th International Conference on Control, Robotics and Cybernetics, CRC 2019, Tokyo, Japan, 2–30 September 2019; pp. 189–195.
11. Tong, W. Electric load forecasting based on improved Artificial Hummingbird Algorithm optimized ELM. *Comput. Era* **2023**, *6*, 43–47. [\[CrossRef\]](#)
12. Gan, L.; Mei, H.; Liqian, F.; Chongyin, J.; Yongjun, Z. Short-term power load forecasting based on an improved multi-verse optimizer algorithm optimized extreme learning machine. *Power Syst. Prot. Control* **2022**, *50*, 99–106. [\[CrossRef\]](#)
13. Wang, Z.X.; Ku, Y.Y.; Liu, J. The Power Load Forecasting Model of Combined SaDE-ELM and FA-CAWOA-SVM Based on CSSA. *IEEE Access* **2024**, *12*, 41870–41882. [\[CrossRef\]](#)
14. Zhang, S.; Duan, X.; Zhang, L.; Jiang, A.; Yao, Y.; Liu, Y.; Mu, Y. Tsne Dimension Reduction Visualization Analysis and Moth Flame Optimized ELM Algorithm Applied in Power Load Forecasting. *Proc. Chin. Soc. Electr. Eng.* **2021**, *41*, 3120–3129.
15. Huang, G.-B.; Wang, D.H.; Lan, Y. Extreme learning machines: A survey. *Int. J. Mach. Learn. Cybern.* **2011**, *2*, 107–122. [\[CrossRef\]](#)
16. Wu, Z.; Lu, X. Microgrid Fault Diagnosis Based on Whale Algorithm Optimizing Extreme Learning Machine. *J. Electr. Eng. Technol.* **2024**, *19*, 1827–1836. [\[CrossRef\]](#)
17. Nayak, J.R.; Shaw, B.; Sahu, B.K. A fuzzy adaptive symbiotic organism search based hybrid wavelet transform-extreme learning machine model for load forecasting of power system: A case study. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 10833–10847. [\[CrossRef\]](#)
18. Pan, B.; Hirota, K.; Jia, Z.; Zhao, L.; Jin, X.; Dai, Y. Multimodal emotion recognition based on feature selection and extreme learning machine in video clips. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 1903–1917. [\[CrossRef\]](#)
19. Tizhoosh, H.R. Opposition-Based Learning: A New Scheme for Machine Intelligence. In Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), Vienna, Austria, 28–30 November 2005; pp. 695–701.
20. Mirjalili, S. SCA: A Sine Cosine Algorithm for solving optimization problems. *Knowl.-Based Syst.* **2016**, *96*, 120–133. [\[CrossRef\]](#)
21. Yuan, F.; Che, J. An ensemble multi-step M-RMLSSVR model based on VMD and two-group strategy for day-ahead short-term load forecasting. *Knowl.-Based Syst.* **2022**, *252*, 109440. [\[CrossRef\]](#)

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.