ISSN: 2710-0588

Short-Term Wind Power Prediction based on Adaptive Hybrid Swarm Intelligence Optimization

Qing Xue*

College of Software, Henan Polytechnic University, Jiaozuo, China

*Corresponding Author: xq@home.hpu.edu.cn

Abstract

Wind power prediction is crucial for ensuring the safe and stable operation of high-penetration renewable energy power systems. Addressing the nonlinear and non-stationary characteristics of wind power sequences and the challenges of hyperparameter optimization in deep learning models, this paper proposes a short-term wind power prediction method based on adaptive hybrid swarm intelligence optimization. First, an adaptive elite hybrid algorithm (SAE-PSO-GWO) integrating Grey Wolf Optimizer and Particle Swarm Optimization is constructed, incorporating elite opposition-based learning and parameter adaptive mechanisms to balance global exploration and local exploitation capabilities. Subsequently, a prediction model combining multi-scale CNN and LSTM is designed to extract multi-scale temporal features and model long-term dependencies. Finally, SAE-PSO-GWO is employed to automatically optimize the model hyperparameters. Experimental results demonstrate that the proposed algorithm exhibits superior convergence accuracy and stability on the CEC2010 test functions. On wind power datasets, the optimized model achieves MAE, RMSE, SMAPE, and R^2 values of 57.89, 116.87, 14.17%, and 0.9680, respectively, representing improvements of approximately 14.38%, 19.48%, 4.90%, and 1.83% compared to baseline models. This verifies the effectiveness of the proposed method in enhancing prediction accuracy and robustness.

Keywords

Wind Power Prediction, Hybrid Swarm Intelligence Optimization, Hyperparameter Optimization, Multi-scale Feature Extraction, Deep Learning.

1. Introduction

With the continuous increase in the penetration rate of new energy generation in power systems, wind power, as a clean and renewable[1] form of energy, is playing an increasingly important role in the energy structure transition and the achievement of the "Dual Carbon" goals[2]. However, wind power is influenced by various factors such as meteorological conditions, terrain environment, and atmospheric turbulence, exhibiting significant randomness, nonlinearity, and time-varying characteristics, which pose serious challenges to the safe and stable operation and dispatch decision-making of power systems[3]. Therefore, constructing high-precision and high-robustness wind power prediction models holds significant theoretical importance and engineering value for enhancing wind power integration capacity and reducing system operational risks[4].

To improve the accuracy of wind power prediction, researchers have developed various methods, which can be broadly categorized into four types: physical models, statistical models, artificial intelligence (AI) models, and hybrid methods. Physical models are primarily based on the physical processes of meteorology and wind energy conversion, with common examples including numerical weather prediction (NWP) models[5] and power curve models based on

ISSN: 2710-0588

wind turbines. However, they rely on turbine structural parameters and detailed meteorological field information, resulting in complex modeling and high requirements for prior knowledge. Statistical models, such as autoregressive integrated moving average (ARIMA)[6], exponential smoothing[7], and Kalman filters[8], can capture certain trends in time series to some extent, but often perform inadequately when dealing with strong nonlinearity and long-term dependencies. In recent years, with the rapid development of deep learning techniques, AI-based prediction methods such as artificial neural networks (ANN)[9][10], support vector regression (SVR)[11], random forest (RF)[12], decision trees (DT)[13], k-nearest neighbors (KNN)[14], and extreme gradient boosting (XGBoost)[15] have been increasingly applied to wind power prediction. Among these, convolutional neural networks (CNN)[16] can effectively extract local features, while recurrent neural networks (RNN) and their improved variants such as long short-term memory networks (LSTM)[17] excel in modeling long-term dependencies in time series. Integrating CNN with LSTM to construct hybrid deep learning models has become an important research direction for improving the accuracy of wind power prediction.

While deep learning models possess strong capabilities in nonlinear fitting and feature representation, their predictive performance largely depends on network architecture and hyperparameter configurations, such as the number and size of convolutional kernels, learning rate, hidden layer dimensions, dropout rate, etc. These parameters often exhibit complex coupling and nonlinear interactions, causing the parameter space to present multimodal, high-dimensional, and non-convex characteristics. Traditional parameter tuning methods that rely on manual trial-and-error or exhaustive grid search are not only computationally expensive and inefficient but also prone to falling into local optima, making it difficult to obtain globally optimal parameter combinations. To address this, various intelligent optimization algorithms have been widely introduced to solve the automatic hyperparameter tuning problem of deep learning models, owing to their advantages of being gradient-free and possessing global search capabilities. For instance, the Differential Evolution (DE)[18] algorithm utilizes vector difference perturbation for global exploration, demonstrating good robustness in continuous parameter optimization; the Particle Swarm Optimization (PSO)[19] algorithm simulates the social behavior of bird flocks, converging quickly but being prone to premature convergence; the Whale Optimization Algorithm (WOA)[20] is inspired by bubble-net feeding behavior and exhibits strong global search capability; the Harris Hawks Optimization (HHO)[21] algorithm simulates predation strategies, featuring dynamic exploration and exploitation capabilities; the Sparrow Search Algorithm (SSA)[22] simulates the foraging and anti-predation behaviors of sparrows, showing high activity in the early search phase; the Grey Wolf Optimizer (GWO)[23] algorithm mimics the social hierarchy and hunting mechanism of wolf packs, with a simple structure and strong exploration ability; and the Grasshopper Optimization Algorithm (GOA)[24] simulates the social interactions of grasshopper populations, making it suitable for various optimization scenarios. For example, Tahmasebifar et al.[25] introduced a heuristic method combining Particle Swarm Optimization (PSO) with Extreme Learning Machines for probabilistic wind power forecasting. Zhou et al.[26] proposed an LSTM model using Gaussian mixture clustering, optimized via the Snake Optimization Algorithm for wind power prediction. Wang et al.[27] improved the Hunter–Prey Optimization (HPO) algorithm to optimize the weights and thresholds of ELM networks and validated its effectiveness using real-world data. Devi et al.[28] presented a forecasting model combining LSTM with Enhanced Forget Gate (EFG) networks, employing the Cuckoo Search Optimization (CSO) algorithm for hyperparameter tuning. H et al.[29] first decomposed the original wind power generation signal using Variational Mode Decomposition (VMD) and then fed the decomposed subsequences into an improved LSTM for forecasting. Ding et al.[30] proposed a CEEMD-WOA-KELM framework for wind power forecasting, integrating signal decomposition, meta-heuristic optimization, and machine learning techniques. Shang et al.[31] applied Ensemble Empirical Mode Decomposition (EEMD)

ISSN: 2710-0588

to denoise wind power data and obtained forecasting results using attention mechanisms and Convolutional Neural Networks (CNN). These meta-heuristic algorithms provide diverse approaches for automated hyperparameter tuning.

However, existing intelligent optimization algorithms still face their inherent limitations when dealing with the complex high-dimensional parameter optimization of deep learning models. For example, the Differential Evolution (DE) algorithm may suffer from reduced search efficiency in later stages, Particle Swarm Optimization (PSO) is prone to falling into local optima, the Whale Optimization Algorithm (WOA) sometimes lacks exploitation accuracy, the Harris Hawks Optimization (HHO) and Salp Swarm Algorithm (SSA) have limited parameter adaptive capabilities, the Grey Wolf Optimizer (GWO) exhibits relatively weak local exploitation ability, and the Grasshopper Optimization Algorithm (GOA) faces challenges in convergence stability when handling complex non-convex problems. A single algorithm often struggles to achieve a persistent and adaptive balance between "exploration" and "exploitation," resulting in constrained optimization efficiency and final performance.

To address the aforementioned issues, this paper proposes a wind power prediction model based on adaptive hybrid swarm intelligence optimization. The core of the model lies in introducing an adaptive elite PSO–GWO hybrid optimization algorithm (SAE-PSO-GWO) into the deep learning prediction framework. Through a synergistic mechanism, this algorithm deeply integrates the global exploration capability of GWO, which is based on social hierarchy, with the local exploitation characteristics of PSO, which relies on individual and collective experience. Simultaneously, the algorithm incorporates an elite opposition-based learning strategy to enhance population diversity and adopts an adaptive parameter adjustment mechanism to dynamically balance the demands of different search phases, thereby achieving efficient and precise collaborative optimization of key hyperparameters in the deep learning model. On this basis, this paper constructs a Multiscale-CNN-LSTM wind power prediction model. This model utilizes a multiscale convolutional neural network to extract local and multiscale features from input time-series data in parallel, and then employs a long short-term memory network to deeply capture complex long-term temporal dependencies and dynamic change patterns in the wind power sequence. By applying the aforementioned hybrid optimization algorithm to automatically optimize the key hyperparameters of this prediction model, the accuracy of wind power prediction and the generalization ability of the model are significantly improved. The main contributions of this paper are as follows:

(1) Proposing an adaptive elite PSO–GWO hybrid swarm intelligence optimization algorithm (SAE-PSO-GWO).

(2) Constructing a Multiscale-CNN-LSTM wind power prediction model optimized based on SAE-PSO-GWO.

(3) Validating the effectiveness and robustness of the model through multi-indicator comparative experiments.

2. Theory

2.1. Grey Wolf Optimizer

The Grey Wolf Optimizer (GWO) is a swarm intelligence-based heuristic optimization algorithm inspired by the social hierarchy and cooperative hunting behavior of grey wolf packs in nature. A grey wolf pack is typically divided into four hierarchical levels: α, β, δ, and ω. The α wolf represents the current optimal solution, while the other individuals continuously approach the position of the prey under its guidance, thereby achieving the global optimization process.

In GWO, the position of the prey in the search space is assumed to represent the global optimum, and the positions of individual grey wolves correspond to candidate solutions. The algorithm

ISSN: 2710-0588

conducts the search in the solution space by simulating three phases: "encircling, tracking, and attacking." During the t-th iteration, the position of each grey wolf individual is updated as follows:

$$\mathbf{D} = |\mathbf{C} \cdot \mathbf{X}_p - \mathbf{X}(t)| \tag{1}$$

$$\mathbf{X}(t+1) = \mathbf{X}_p - \mathbf{A} \cdot \mathbf{D} \tag{2}$$

Where $\mathbf{X}_p$ represents the position of the prey (i.e., the optimal solution). $\mathbf{A}$ and $\mathbf{C}$ are control parameters, calculated as follows:

$$\mathbf{A} = 2a \cdot r_1 - a, \quad \mathbf{C} = 2 \cdot r_2 \tag{3}$$

Where r_1 and r_2 are random numbers within the interval [0, 1]. a is the convergence factor, typically implemented using a linearly decreasing strategy:

$$a = 2\left(1 - \frac{t}{T}\right) \tag{4}$$

Where T represents the maximum number of iterations. As the iteration proceeds, the gradual decrease of a enables the algorithm to transition from global search to local search, thereby enhancing solution precision. However, in high-dimensional and complex optimization problems, the traditional GWO is prone to issues such as slower convergence rate in later iterations and limited search accuracy.

2.2. SAE-PSO-GWO Hybrid Swarm Intelligence Optimization Algorithm

2.2.1. GWO-PSO Cooperative Search Mechanism

The GWO algorithm simulates the social hierarchy and hunting behavior of a grey wolf pack to guide the population toward converging to the optimal solution region. Let the top three wolves in the current iteration be denoted as α, β, and δ. The position update rule for the remaining individuals can be expressed as:

$$\mathbf{D}_\alpha = |\mathbf{C}_1 \cdot \mathbf{X}_\alpha - \mathbf{X}| \tag{5}$$

$$\mathbf{D}_\beta = |\mathbf{C}_2 \cdot \mathbf{X}_\beta - \mathbf{X}| \tag{6}$$

$$\mathbf{D}_\delta = |\mathbf{C}_3 \cdot \mathbf{X}_\delta - \mathbf{X}| \tag{7}$$

$$\mathbf{X}(t+1) = \frac{\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3}{3} \tag{8}$$

Where, $\mathbf{X}_1 = \mathbf{X}_\alpha - \mathbf{A}_1 \cdot \mathbf{D}_\alpha$, $\mathbf{X}_2 = \mathbf{X}_\beta - \mathbf{A}_2 \cdot \mathbf{D}_\beta$, $\mathbf{X}_3 = \mathbf{X}_\delta - \mathbf{A}_3 \cdot \mathbf{D}_\delta$, $\mathbf{A}_i = 2a \cdot \mathbf{r}_i - a$, $\mathbf{C}_i = 2 \cdot \mathbf{r}_i$, a decreases linearly with the number of iterations, and $\mathbf{r}_i$ is a random vector within the interval [0,1].

ISSN: 2710-0588

Although GWO exhibits strong exploration capability in the global search phase, its local search accuracy is limited in the later iterations. Therefore, this paper introduces a PSO update mechanism in the neighborhood of elite solutions to locally intensify the search process. The velocity and position update formulas of PSO are:

$$\mathbf{v}_i(t+1) = \omega \mathbf{v}_i(t) + c_1 r_1 (\mathbf{p}_i - \mathbf{x}_i(t)) + c_2 r_2 (\mathbf{g} - \mathbf{x}_i(t)) \tag{9}$$

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \tag{10}$$

Where $\mathbf{p}_i$ denotes the personal best position, $\mathbf{g}$ represents the current global best solution, while ω, c_1, and c_2 refer to the inertia weight and learning factors, respectively.

As illustrated in Figure 1, the search process of SAE-PSO-GWO can be divided into three stages:

(1) GWO performs global exploration across the entire search space;

(2) PSO executes local exploitation around elite solutions;

(3) The elite individual guides the population to accelerate convergence toward the optimal region.

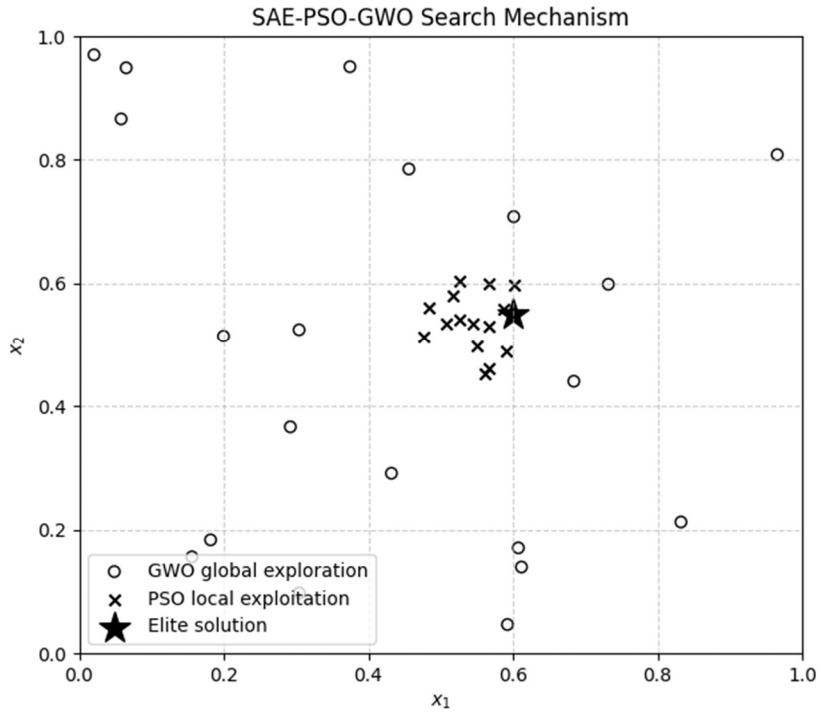


Figure 1. Schematic diagram of the three-stage search process of the SAE-PSO-GWO hybrid algorithm.

2.2.2. Elite Opposition-Based Learning Strategy (EOBL)

To further enhance the algorithm's ability to escape local optima, an elite opposition-based learning strategy is introduced during the iteration process. For an elite solution $\mathbf{X}_\alpha = (x_{\alpha,1}, x_{\alpha,2}, \ldots, x_{\alpha,D})$ in the current iteration, its corresponding opposite solution $\mathbf{X}_{opp}$ is defined as:

ISSN: 2710-0588

$$\mathbf{X}_{opp} = \mathbf{X}_{min} + \mathbf{X}_{max} - \mathbf{X}_{\alpha} \tag{11}$$

Where $\mathbf{X}_{min}$ and $\mathbf{X}_{\max}$ represent the lower bound and upper bound of the search space, respectively.

By comparing the fitness values of $\mathbf{X}_{\alpha}$ and $\mathbf{X}_{opp}$, the superior solution is retained for the next generation, thereby preserving the advantage of elite solutions while significantly enhancing population diversity and global search capability. Figure 2 intuitively illustrates the symmetric distribution relationship between the elite solution and its opposite solution in the search space.

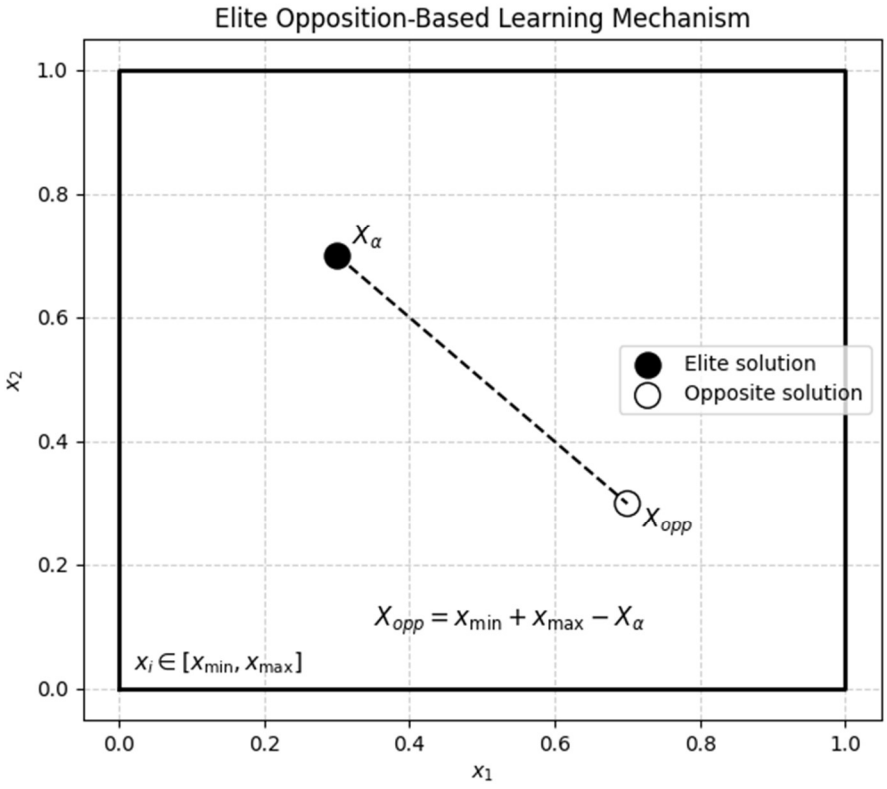


Figure 2. Diagram illustrating the symmetrical distribution relationship between the elite solution and its opposite solution in the search space.

2.2.3. SAE-PSO-GWO Algorithm Workflow

Integrating the aforementioned GWO global search mechanism, PSO local exploitation strategy, and the elite opposition-based learning method, the overall workflow of the proposed hybrid swarm intelligence optimization algorithm, named SAE-PSO-GWO, is as follows.

First, given the boundary conditions of the search space, the grey wolf population positions and PSO velocity vectors are randomly initialized. The fitness values of all individuals are calculated, and the individual with the best current fitness is selected as the initial elite solution. Subsequently, the algorithm enters the iterative optimization phase. In each iteration, the linear decreasing control parameter a is used to adjust the search scope, guiding the population to perform global search in the solution space based on the GWO position update formula, thereby enhancing the algorithm's exploration capability.

When the iteration enters the later stages, the PSO update mechanism is introduced within the neighborhood of the elite solution. It leverages the particle's personal historical best position and global best information to refine the search direction, thereby strengthening local

ISSN: 2710-0588

exploitation and accelerating convergence speed. Building upon this, the elite opposition-based learning operation is performed on the current elite solution. By constructing its symmetrical opposite solution within the search space and comparing the fitness values of both, the superior individual is retained for the next generation. This aims to improve population diversity and reduce the risk of becoming trapped in local optima.

The above process iterates continuously until the maximum iteration count or convergence criteria are met, ultimately outputting the global optimal solution and its corresponding fitness value. By adaptively integrating exploration and exploitation processes, SAE-PSO-GWO achieves efficient solutions for complex high-dimensional optimization problems. Its specific optimization workflow is summarized in Table 1.

Table 1. SAE-PSO-GWO Optimization Algorithm

Algorithm 1：SAE-PSO-GWO Optimization Algorithm
Input:
Objective function $f(\cdot)$
Population size N
Maximum iteration number T
Parameter bounds $[\mathbf{X}_{min}, \mathbf{X}_{max}]$
Output:
Optimal solution $\mathbf{X}_{best}$
Best fitness value $f(\mathbf{X}_{best})$
1: Initialize population X_i (i = 1, 2, ..., N) and velocity V_i
2: Evaluate fitness $f(X_i)$ for all individuals
3: Determine elite solution X_α
4: for t = 1 to T do
5: Update convergence factor a
6: for each individual i do
7: Update position using GWO mechanism
8: if t > 0.6T then
9: Update velocity using PSO rule
10: $X_i \leftarrow X_i + V_i$
11: end if
12: Apply boundary constraints
13: end for
14: Update personal best P_best
15: Update elite solution X_α
16: Generate opposite elite solution X_{opp}
17: if $f(X_{opp}) < f(X_\alpha)$ then
18: $X_\alpha \leftarrow X_{opp}$
19: end if
20: Record best fitness value
21: end for
22: return X_α, $f(X_\alpha)$

ISSN: 2710-0588

2.3. Multi-CNN-LSTM Model

Wind power time series exhibit significant fluctuations across different time scales, making it difficult for a single-scale feature extraction method to fully capture its variation patterns. Therefore, this paper constructs a Multi-CNN-LSTM hybrid forecasting model, which integrates multi-scale convolutional neural networks and long short-term memory networks to deeply mine the complex features of wind power time series.

In the feature extraction stage, a multi-scale convolutional structure is introduced, where convolution kernels of different sizes operate in parallel on the input sequence to simultaneously capture short-term fluctuation features and medium-to-long-term variation trends. The outputs from each convolutional branch are concatenated along the channel dimension to form a richer feature representation.

In the temporal modeling stage, an LSTM network is employed to learn from the fused features. By incorporating a gating mechanism, LSTM effectively mitigates the gradient vanishing problem commonly encountered in traditional RNNs during long-sequence modeling, thereby capturing long-term dependencies in wind power time series. Finally, the model outputs the wind power prediction results through a fully connected layer.

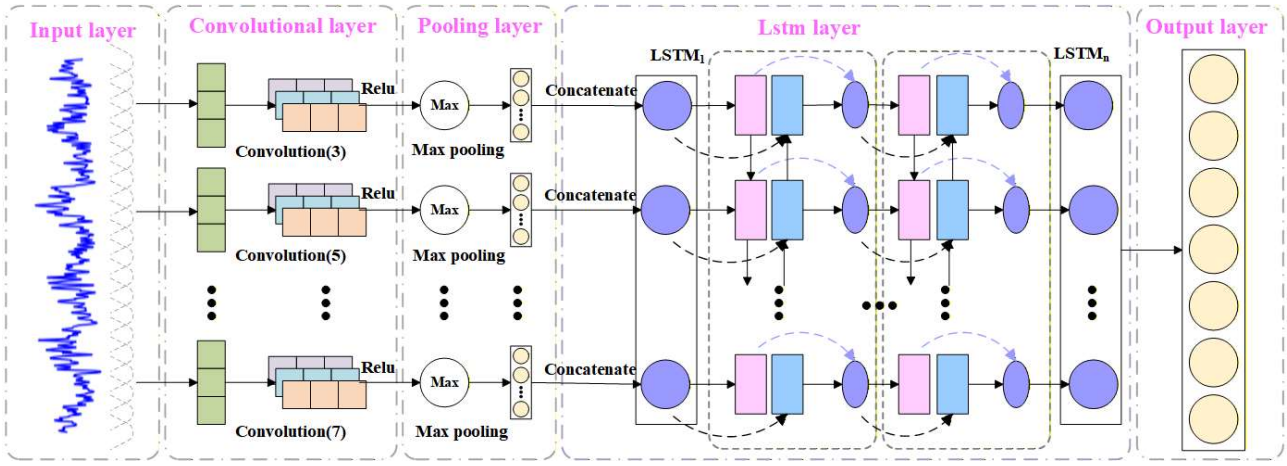


Figure 3. Framework of the Multi-CNN-LSTM model.

2.4. Model Framework Design

The overall framework of the wind power prediction model proposed in this paper is illustrated in the figure below (a schematic diagram of the model structure can be provided here). The model mainly consists of three parts: a data preprocessing module, an SAE-PSO-GWO parameter optimization module, and a Multi-CNN-LSTM prediction module.

First, the raw wind power and meteorological data undergo missing value handling, time series construction, and normalization to ensure the validity and stability of the input data for the model. Subsequently, the SAE-PSO-GWO algorithm is employed to optimize the key hyperparameters of the Multi-CNN-LSTM model. By constructing a fitness function with prediction error as the objective, automated search and optimal configuration of model parameters are achieved.

After obtaining the optimal parameter set, the Multi-CNN-LSTM model is trained based on the training data, and predictions are performed on the test set for performance evaluation. This framework integrates hybrid swarm intelligence optimization algorithms with deep learning models, enabling synergistic optimization of model structure design and parameter tuning, thereby significantly improving the accuracy and robustness of wind power prediction.

ISSN: 2710-0588

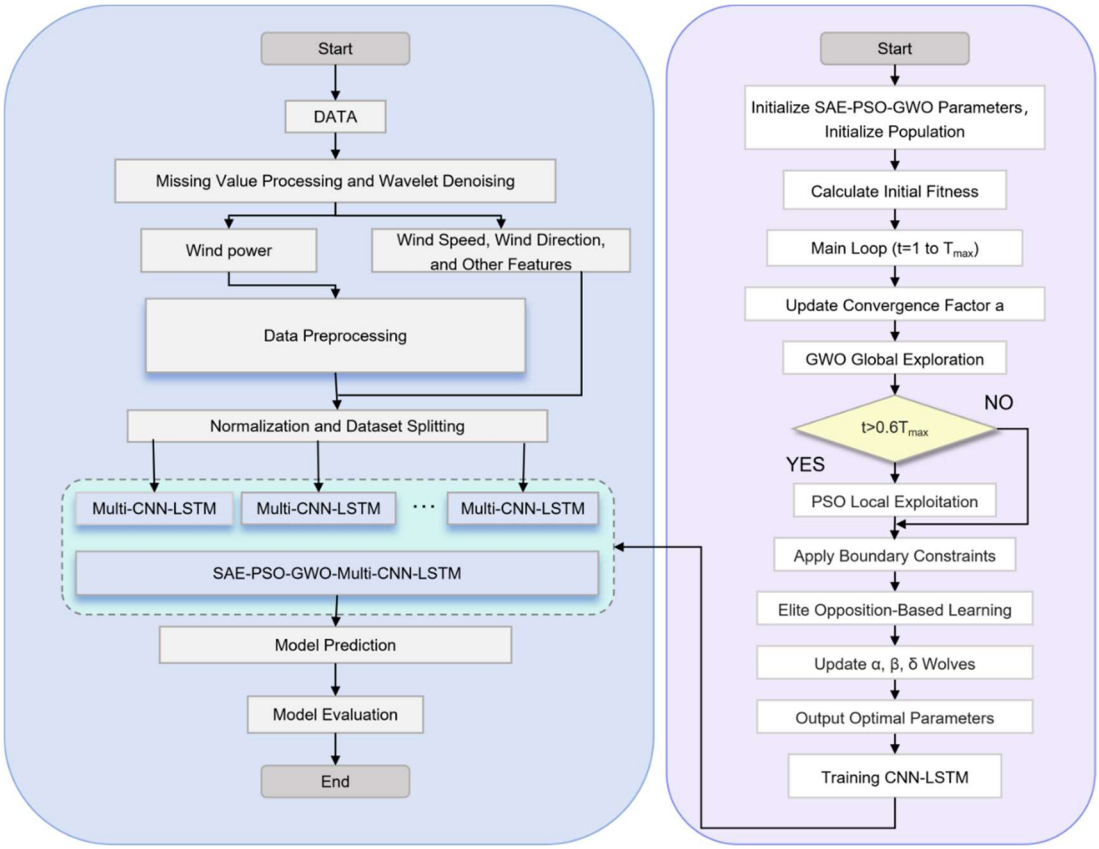


Figure 4. Overall Model Framework Diagram.

3. Optimization Algorithm Test Results and Analysis

3.1. Benchmark Functions

Table 2. Unimodal Test Functions

Test Function	Dimension	Search Range	Minimum
$f_1(x)=\sum_{i=1}^{n}x_i^2$	30	[-100, 100]	0
$f_2(x)=\sum_{i=1}^{n}\lvert x_i\rvert+\prod_{i=1}^{n}\lvert x_i\rvert$	30	[-10, 10]	0
$f_3(x)=\sum_{i=1}^{n}\left(\sum_{j=1}^{i}x_j\right)^2$	30	[-100, 100]	0
$f_4(x)=\max_i\{\lvert x_i\rvert,1\le i\le n\}$	30	[-100, 100]	0
$f_5(x)=\sum_{i=1}^{n-1}\left[100(x_{i+1}-x_i)^2+(x_i-1)^2\right]$	30	[-30, 30]	0
$f_6(x)=\sum_{i=1}^{n}(\lfloor x_i+0.5\rfloor)^2$	30	[-100, 100]	0
$f_7(x)=\sum_{i=1}^{n}ix_i^4+\mathrm{random}[0,1)$	30	[-1.28, 1.28]	0

To evaluate the optimization performance of the SAE-PSO-GWO algorithm, 23 classic test functions from CEC2010 are selected for experiments, as detailed in Table 2 for the seven

ISSN: 2710-0588

unimodal functions (F1–F7), Table 3 for the six multimodal functions (F8–F13), and Table 4 for the ten fixed-dimension multimodal functions (F14–F23). Unimodal functions have only one global optimum and no local optima, primarily used to assess the algorithm's local exploitation ability and convergence speed. Multimodal functions possess multiple local optima and one global optimum, with complex search spaces where the number of local optima increases exponentially with the problem dimension. They are mainly employed to evaluate the algorithm's global exploration capability and its ability to avoid local optima. Fixed-dimension multimodal functions have relatively low dimensions but contain numerous local optima in their search space. This category of functions is primarily used to evaluate the algorithm's performance in low-dimensional yet complex problems. Due to their lower dimensionality, the algorithm's search behavior is more easily visualized, facilitating the analysis of the optimization process. Fixed-dimension multimodal functions effectively test the algorithm's performance in regions dense with local optima. By combining unimodal, multimodal, and fixed-dimension multimodal functions, this study comprehensively evaluates the algorithm's performance across various types of optimization problems.

Table 3. Multimodal Test Functions

Test Function	Dimension	Search Range	Minimum
$f_8(x)=\sum_{i=1}^{n}-x_i\sin(\sqrt{\lvert x_i\rvert})$	30	[-500, 500]	-12569.5
$f_9(x)=\sum_{i=1}^{n}\left[x_i^2-10\cos(2\pi x_i)+10\right]$	30	[-5.12, 5.12]	0
$f_{10}(x)=-20\exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2})$ $-\exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos 2\pi x_i\right)+20+e$	30	[-32, 32]	0
$f_{11}(x)=\frac{1}{4000}\sum_{i=1}^{n}x_i^2-\prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	30	[-600, 600]	0
$f_{12}(x)=\frac{\pi}{n}\{10\sin^2(\pi y_i)$ $+\sum_{i=1}^{n-1}(y_i-1)^2[1+10\sin^2(\pi y_{i+1})]$ $+(y_n-1)^2\}+\sum_{i=1}^{n}u(x_i,10,100,4)$ $y_i=1+\frac{1}{4}(x_i+1),$ $u(x_i,a,k,m)=\begin{cases}k(x_i-a)^m, & x_i>a\\ 0, & -a\le x_i\le a\\ k(-x_i-a)^m, & x_i<-a\end{cases}$	30	[-50, 50]	0
$f_{13}(x)=0.1\{\sin^2(3\pi x_1)$ $+\sum_{i=1}^{n-1}(x_i-1)^2[1+\sin^2(3\pi x_{i+1})]$ $+(x_n-1)^2[1+\sin^2(2\pi x_n)]\}$ $+\sum_{i=1}^{n}u(x_i,5,100,4)$	30	[-50, 50]	30

ISSN: 2710-0588

Table 4. Fixed-Dimension Multimodal Test Functions

Test Function	Dimension	Search Range	Minimum
$f_{14}(x)=\left[\frac{1}{500}+\sum_{j=1}^{25}\frac{1}{j+\sum_{i=1}^{2}(x_i-a_{ij})^6}\right]^{-1}$	2	[-65.536, 65.536]	1
$f_{15}(x)=\sum_{i=1}^{11}\left[a_i-\frac{x_1(b_i^2+b_ix_2)}{b_i^2+b_ix_3+x_4}\right]^2$	4	[-5, 5]	0
$f_{16}(x)=4x_1^2-2.1x_1^4+\frac{1}{3}x_1^6+x_1x_2-4x_2^2+4x_2^4$	2	[-5, 5]	0
$f_{17}(x)=\left(x_2-\frac{5.1}{4\pi^2}x_1^2+\frac{5}{\pi}x_1-6\right)^2$ $+10\left(1-\frac{1}{8\pi}\right)\cos(x_1)+10$	2	[-5, 10]	0.398
$f_{18}(x)=\left[1+(x_1+x_2+1)^2(19-14x_1\right.$ $\left.+3x_1^2-14x_2+6x_1x_2+3x_2^2)\right]$ $\times\left[30+(2x_1-3x_2)^2\right.$ $\left.\times(18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2)\right]$	2	[-2, 2]	3
$f_{19}(x)=-\sum_{i=1}^{4}c_i\exp\left(-\sum_{j=1}^{4}a_{ij}(x_j-p_{ij})^2\right)$	4	[0, 1]	-3.86
$f_{20}(x)=-\sum_{i=1}^{4}c_i\exp\left(-\sum_{j=1}^{6}a_{ij}(x_j-p_{ij})^2\right)$	6	[0, 1]	-3.32
$f_{21}(x)=-\sum_{i=1}^{5}\left[(x-a_i)(x-a_i)^T+c_i\right]^{-1}$	4	[0, 10]	-10
$f_{22}(x)=-\sum_{i=1}^{7}\left[(x-a_i)(x-a_i)^T+c_i\right]^{-1}$	4	[0, 10]	-10
$f_{23}(x)=-\sum_{i=1}^{10}\left[(x-a_i)(x-a_i)^T+c_i\right]^{-1}$	4	[0, 10]	-10

3.2. Parameter Settings

SAE-PSO-GWO was compared with seven typical swarm intelligence optimization algorithms: GWO, DE, PSO, WOA, HHO, SSA, and GOA. The population size for all algorithms was uniformly set to 30, and the maximum number of iterations was set to 500. The specific parameter values for each algorithm are shown in Table 5. To ensure the reliability and stability of the experimental results and avoid potential random errors from single runs, each algorithm was independently executed 50 times. Through statistical analysis, the optimal value, mean value, and standard deviation from each run were extracted as the final performance evaluation metrics. These metrics comprehensively reflect the differences among the algorithms in terms of convergence speed, solution accuracy, and stability during the optimization process. By comparing the performance of SAE-PSO-GWO with that of GWO, DE, PSO, WOA, HHO, SSA, and GOA, the improvements of SAE-PSO-GWO in global search capability, local exploitation ability,

ISSN: 2710-0588

and adaptability can be thoroughly examined, thereby validating its advantages and applicability over traditional swarm intelligence optimization algorithms.

Table 5. Parameter Settings for Different Algorithms

Algorithm	Parameter Settings
DE	F=0.8, CR=0.9
GOA	c_min=4E-05, c_max=1
GWO	a=2.0
HHO	E0=2.0, beta=1.5
PSO	w=0.729, c1=1.494, c2=1494
SSA	PD=0.2, SD=0.1, warn_threshold=0.8
WOA	a=2.0, b=1.0
SAE_PSO_GWO	a_init=2.0, w=0.7, c1=1.5, c2=1.5, pso_start_ratio=0.6

3.3. Analysis of Results

In comprehensive and systematic benchmark tests for function optimization, the SAE_PSO_GWO algorithm demonstrates exceptional overall performance and strong competitiveness. It exhibits a series of significant advantages across unimodal, multimodal, and fixed-dimension multimodal benchmark functions, proving its effectiveness and sophistication as an advanced hybrid optimization algorithm.

3.3.1. Comparative Analysis on Unimodal Benchmark Functions

Unimodal benchmark functions typically possess a single global optimum. Their search space is relatively simple, and they are primarily used to evaluate an optimization algorithm's convergence rate and local search capability. In this section, the performance of the proposed SAE_PSO_GWO algorithm is compared and analyzed against algorithms including GWO, DE, PSO, WOA, HHO, SSA, and GOA on unimodal benchmark functions. Evaluation metrics encompass the *Best*, *Mean*, and *Standard Deviation* values. The experimental results are presented in **Table 6**, with a focus on assessing the algorithms' convergence speed, precision, and stability.

Table 6. Performance comparison of different optimization algorithms on unimodal benchmark functions (F1-F7)

Function	Metric	DE	PSO	WOA	HHO	SSA	GOA	GWO	SAE_PSO_GWO
f_1	Best Value	3.32E+03	1.34E-08	4.03E-112	1.90E+03	4.36E+04	4.12E+04	1.43E-121	2.02E-25
	Mean Value	6.60E+03	3.33E+02	5.93E-95	8.45E+03	6.51E+04	6.18E+04	9.30E-100	6.36E-18
	Std Dev	1.76E+03	1.80E+03	3.19E-94	6.01E+03	7.80E+03	7.41E+03	5.00E-99	2.89E-17
f_2	Best Value	5.00E+01	4.33E-04	3.28E-69	7.32E+01	3.28E+02	6.25E+03	2.24E-73	3.97E-14
	Mean Value	7.91E+01	1.40E+01	4.94E-61	1.16E+02	1.72E+08	2.91E+07	4.28E-63	5.86E-09
	Std Dev	6.54E+01	1.23E+01	1.52E-60	1.71E+01	3.87E+08	4.80E+07	1.97E-62	1.89E-08
f_3	Best Value	7.32E+01	1.91E+00	4.39E-08	6.36E+01	7.86E+01	7.22E+01	2.85E-31	1.37E-10
	Mean Value	8.28E+01	5.65E+00	5.08E+01	7.96E+01	8.59E+01	8.23E+01	2.30E-01	8.25E-09
	Std Dev	4.23E+00	2.42E+00	3.64E+01	6.76E+00	3.00E+00	3.34E+00	1.23E+00	1.22E-08
f_4	Best Value	2.18E+04	4.10E+02	1.32E+04	4.22E+04	6.18E+04	5.64E+04	2.72E+04	2.09E-11
	Mean Value	2.90E+04	1.13E+04	4.22E+04	7.79E+04	1.37E+05	1.11E+05	5.34E+04	5.23E-06
	Std Dev	3.95E+03	7.36E+03	1.30E+04	2.27E+04	5.00E+04	4.02E+04	1.33E+04	1.07E-05
f_5	Best Value	1.62E+06	1.51E+00	2.87E+01	3.15E+05	1.29E+08	9.67E+07	7.00E-02	2.67E+01
	Mean Value	5.16E+06	3.63E+02	2.88E+01	5.17E+06	2.16E+08	1.70E+08	2.27E+01	2.69E+01
	Std Dev	3.16E+06	9.00E+02	2.00E-02	4.23E+06	4.28E+07	3.92E+07	1.05E+01	9.00E-02
f_6	Best Value	3.67E+03	0.00E+00	0.00E+00	5.16E+03	4.35E+04	4.14E+04	0.00E+00	0.00E+00
	Mean Value	6.47E+03	6.70E+02	0.00E+00	1.49E+04	6.50E+04	6.18E+04	0.00E+00	0.00E+00
	Std Dev	1.70E+03	2.49E+03	0.00E+00	6.41E+03	7.80E+03	7.39E+03	0.00E+00	0.00E+00
f_7	Best Value	7.40E-01	1.00E-02	1.51E-05	8.40E-01	5.10E+01	3.57E+01	5.08E-06	8.96E-05
	Mean Value	2.63E+00	1.20E-01	1.00E-03	8.79E+00	6.88E+01	7.75E+01	2.19E-04	2.79E-03
	Std Dev	1.10E+00	4.80E-01	1.32E-03	7.07E+00	9.31E+00	1.74E+01	2.55E-04	3.22E-03

ISSN: 2710-0588

As can be seen from the experimental results in Table 6, on unimodal test functions, SAE_PSO_GWO demonstrates excellent convergence accuracy and strong stability. Its most outstanding performance is observed on function f4, where it achieves an extremely high precision optimal value of 2.09E-11, significantly surpassing various traditional algorithms including DE and PSO. This verifies its powerful capability in precise local optimization. On function f6, the algorithm can stably and consistently converge to the theoretical global optimum value of 0, matching the performance of top metaheuristic algorithms such as WOA and GWO. Moreover, although its optimal value exploration on some functions (e.g., f1, f2) is slightly inferior to a few top-performing algorithms, SAE_PSO_GWO maintains an extremely low standard deviation of average results across all functions. This fully demonstrates its unparalleled stability and robustness in problem-solving, achieving a perfect balance between exploration breadth and exploitation depth on unimodal problems and effectively avoiding severe fluctuations in results.

3.3.2. Analysis of Comparative Results of Algorithms on Multimodal Test Functions

Multimodal test functions typically contain multiple local optima and one global optimum. Their characteristic lies in the complex search space, where the number and distribution of local optima impose higher demands on the global exploration and local exploitation capabilities of algorithms. In this section, the performance of all algorithms on multimodal test functions is compared and analyzed, with a focus on their global exploration ability, local exploitation ability, and robustness.

Table 7. Performance Comparison of Different Optimization Algorithms on Multimodal Test Functions (F8–F13)

Function	Metric	DE	PSO	WOA	HHO	SSA	GOA	GWO	SAE_PSO_GWO
f_8	Best Value	-7.00E+03	-1.01E+04	-1.26E+04	-1.26E+04	-3.99E+03	-4.14E+03	-1.26E+04	-1.26E+04
	Mean Value	-5.65E+03	-8.89E+03	-1.20E+04	-6.61E+03	-2.38E+03	-2.58E+03	-1.25E+04	-8.16E+03
	Std Dev	4.78E+02	6.07E+02	1.52E+03	2.05E+03	5.74E+02	5.22E+02	6.96E+01	1.85E+03
f_9	Best Value	2.55E+02	4.58E+01	0.00E+00	1.61E+02	3.14E+02	3.24E+02	0.00E+00	0.00E+00
	Mean Value	2.87E+02	1.00E+02	0.00E+00	2.39E+02	3.51E+02	3.62E+02	0.00E+00	5.12E-14
	Std Dev	1.88E+01	3.18E+01	0.00E+00	4.30E+01	1.79E+01	1.65E+01	0.00E+00	6.94E-14
f_{10}	Best Value	1.99E+01	7.54E-05	4.44E-16	1.02E+01	1.97E+01	1.97E+01	4.44E-16	4.44E-16
	Mean Value	2.00E+01	2.02E+00	2.69E-15	1.73E+01	2.02E+01	2.01E+01	1.51E-15	7.46E-11
	Std Dev	1.00E-02	3.48E+00	2.67E-15	2.68E+00	1.90E-01	2.10E-01	1.87E-15	2.19E-10
f_{11}	Best Value	3.09E+01	2.79E-08	0.00E+00	1.81E+01	3.98E+02	3.92E+02	0.00E+00	0.00E+00
	Mean Value	6.04E+01	3.03E+00	3.00E-02	7.70E+01	5.95E+02	5.89E+02	0.00E+00	2.59E-17
	Std Dev	1.59E+01	1.62E+01	1.80E-01	5.41E+01	7.12E+01	7.06E+01	0.00E+00	6.83E-17
f_{12}	Best Value	1.79E+05	5.98E-07	0.00E+00	3.07E+01	2.36E+08	1.97E+08	0.00E+00	3.00E-02
	Mean Value	4.27E+06	6.10E-01	5.00E-02	4.00E+06	5.03E+08	3.81E+08	3.00E-02	7.00E-02
	Std Dev	5.24E+06	9.80E-01	4.00E-02	1.09E+07	1.11E+08	9.66E+07	2.00E-02	4.00E-02
f_{13}	Best Value	2.83E+06	5.89E-08	1.20E-01	1.73E+05	5.29E+08	4.60E+08	8.00E-02	5.60E-01
	Mean Value	1.36E+07	1.00E-02	5.90E-01	1.49E+07	9.98E+08	7.92E+08	3.20E-01	1.09E+00
	Std Dev	8.83E+06	2.00E-02	3.60E-01	1.82E+07	1.99E+08	1.72E+08	2.20E-01	3.40E-01

On multimodal test functions, SAE_PSO_GWO demonstrates its exceptional global exploration capability. As shown in Table 7, on functions f9, f10, and f11, the algorithm successfully finds the theoretical global optimum (0 or close to 0), performing on par with or even better than algorithms such as GWO. This clearly indicates its strong potential for accurately locating the global optimum in complex multimodal environments and effectively avoiding getting trapped

ISSN: 2710-0588

in local optima. Especially on the challenging f8 function, its best solution is very close to the theoretical limit, showcasing excellent global guidance capability. Overall, SAE_PSO_GWO maintains good average performance and stability on multimodal problems, proving the effectiveness and reliability of its hybrid strategy in addressing challenges posed by multiple local optima.

3.3.3. Analysis of Comparative Results on Fixed-Dimension Multimodal Test Functions

Fixed-dimension multimodal test functions constitute an important part of optimization algorithm performance evaluation. Their characteristic lies in a fixed search space dimension with multiple local optima, which can effectively test an algorithm's global search capability and local exploitation ability in complex search spaces. A comparative analysis was conducted on the performance of all algorithms on fixed-dimension multimodal test functions, focusing on the algorithms' convergence accuracy, convergence speed, and robustness.

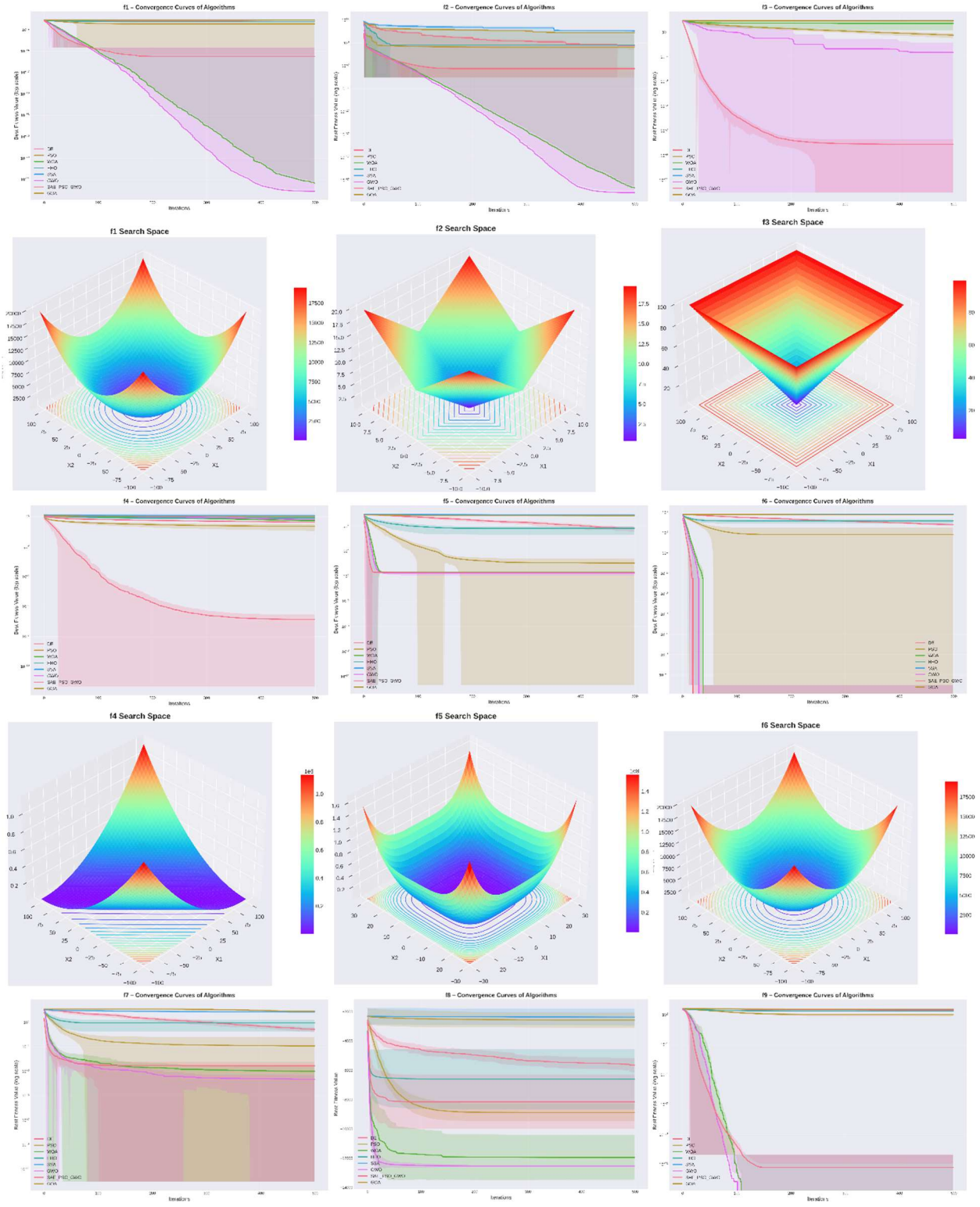
Table 8. Performance Comparison of Different Optimization Algorithms on Fixed-Dimension Multimodal Test Functions (F14–F23)

Function	Metric	DE	PSO	WOA	HHO	SSA	GOA	GWO	SAE_PSO_GWO
f_{14}	Best Value	1.00E+00	1.00E+00	1.00E+00	1.00E+00	3.97E+00	1.00E+00	1.00E+00	1.00E+00
	Mean Value	1.03E+00	1.00E+00	5.50E+00	6.73E+00	2.97E+01	3.33E+00	4.24E+00	6.54E+00
	Std Dev	1.80E-01	0.00E+00	4.41E+00	6.18E+00	5.36E+01	2.52E+00	3.59E+00	4.34E+00
f_{15}	Best Value	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E-02	2.00E-02	0.00E+00	0.00E+00
	Mean Value	0.00E+00	1.00E-02	3.00E-02	6.00E-02	5.00E-02	4.00E-02	3.00E-02	3.00E-02
	Std Dev	0.00E+00	1.00E-02	2.00E-02	3.00E-02	3.00E-02	1.00E-02	2.00E-02	2.00E-02
f_{16}	Best Value	-1.03E+00	-1.03E+00	-1.03E+00	-1.03E+00	-1.03E+00	-1.03E+00	-1.03E+00	-1.03E+00
	Mean Value	-1.03E+00	-1.03E+00	-1.03E+00	-9.50E-01	-1.02E+00	-1.02E+00	-1.03E+00	-1.03E+00
	Std Dev	0.00E+00	0.00E+00	1.01E-08	2.40E-01	1.00E-02	1.00E-02	1.74E-07	1.36E-04
f_{17}	Best Value	4.00E-01	4.00E-01	4.00E-01	4.00E-01	4.00E-01	4.00E-01	4.00E-01	4.00E-01
	Mean Value	4.00E-01	4.00E-01	4.20E-01	4.00E-01	4.40E-01	4.10E-01	4.10E-01	4.40E-01
	Std Dev	0.00E+00	0.00E+00	9.00E-02	0.00E+00	1.10E-01	1.00E-02	5.00E-02	2.00E-01
f_{18}	Best Value	3.00E+00	3.00E+00	3.00E+00	3.00E+00	3.00E+00	3.00E+00	3.00E+00	3.00E+00
	Mean Value	3.00E+00	3.00E+00	9.71E+00	8.40E+00	3.16E+00	3.17E+00	9.70E+00	3.94E+00
	Std Dev	0.00E+00	0.00E+00	1.22E+01	1.08E+01	2.10E-01	1.60E-01	1.22E+01	4.84E+00
f_{19}	Best Value	-4.20E+10	-4.20E+10	-1.42E+04	-4.20E+10	-5.36E+03	-7.45E+03	-9.07E+04	-4.03E+06
	Mean Value	-4.20E+10	-4.20E+10	-2.45E+03	-3.08E+10	-6.70E+02	-6.34E+02	-4.11E+03	-3.04E+05
	Std Dev	0.00E+00	1.75E+00	3.80E+03	1.62E+10	9.22E+02	1.29E+03	1.62E+04	7.55E+05
f_{20}	Best Value	-8.40E+10	-8.40E+10	-9.99E+02	-1.90E+08	-5.72E+02	-1.78E+02	-1.56E+03	-1.34E+04
	Mean Value	-8.40E+10	-7.84E+10	-1.33E+02	-7.00E+06	-1.29E+02	-6.88E+01	-1.68E+02	-1.12E+03
	Std Dev	1.37E+00	2.10E+10	1.98E+02	3.40E+07	1.31E+02	3.94E+01	2.76E+02	2.39E+03
f_{21}	Best Value	-3.73E+00	-3.73E+00	-3.70E+00	-3.73E+00	-2.78E+00	-2.74E+00	-3.69E+00	-3.73E+00
	Mean Value	-3.70E+00	-3.14E+00	-2.84E+00	-2.79E+00	-2.90E-01	-1.48E+00	-2.56E+00	-3.71E+00
	Std Dev	1.30E-01	3.90E-01	5.60E-01	9.00E-01	5.70E-01	7.50E-01	7.20E-01	7.00E-02
f_{22}	Best Value	-3.21E+00	-3.21E+00	-3.20E+00	-3.21E+00	-2.41E+00	-2.52E+00	-3.10E+00	-3.20E+00
	Mean Value	-3.20E+00	-2.98E+00	-2.41E+00	-2.38E+00	-4.40E-01	-1.31E+00	-2.01E+00	-3.18E+00
	Std Dev	0.00E+00	3.70E-01	6.60E-01	9.50E-01	7.00E-01	6.10E-01	6.70E-01	5.00E-02
f_{23}	Best Value	-3.20E+00	-3.20E+00	-2.99E+00	-3.20E+00	-3.05E+00	-2.76E+00	-2.87E+00	-3.18E+00
	Mean Value	-3.06E+00	-2.93E+00	-2.02E+00	-2.10E+00	-5.80E-01	-1.37E+00	-1.91E+00	-2.79E+00
	Std Dev	8.00E-02	3.70E-01	7.60E-01	9.90E-01	8.20E-01	6.60E-01	6.80E-01	5.40E-01

On more complex fixed-dimensional multimodal test functions, SAE_PSO_GWO once again demonstrated its broad applicability and strong optimization capability. As shown in Table 8, for most functions (e.g., f16, f17, f21, f22, f23), the obtained optimal solutions are very close to or even reach the known theoretical optimum values, and the algorithm ranks first alongside other algorithms such as DE, PSO, and GWO in multiple functions. This fully illustrates that the algorithm can still maintain high search efficiency and solution quality when dealing with fixed-dimensional problems characterized by complex nonlinear structures and multimodal properties. Although there is room for improvement in certain functions (e.g., f19), overall, the

ISSN: 2710-0588

high consistency and robustness exhibited by SAE_PSO_GWO across these functions highlight the superiority of its algorithmic structure design, enabling it to adapt to various complex optimization scenarios of different types.



ISSN: 2710-0588

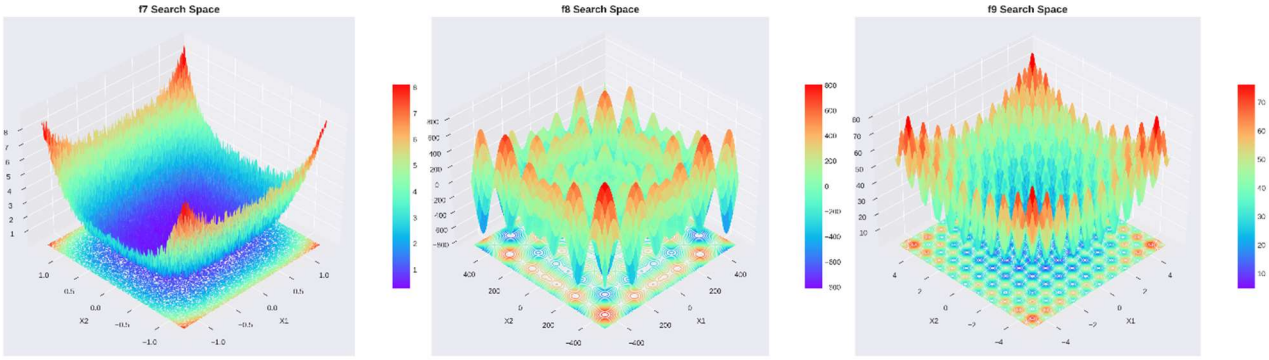
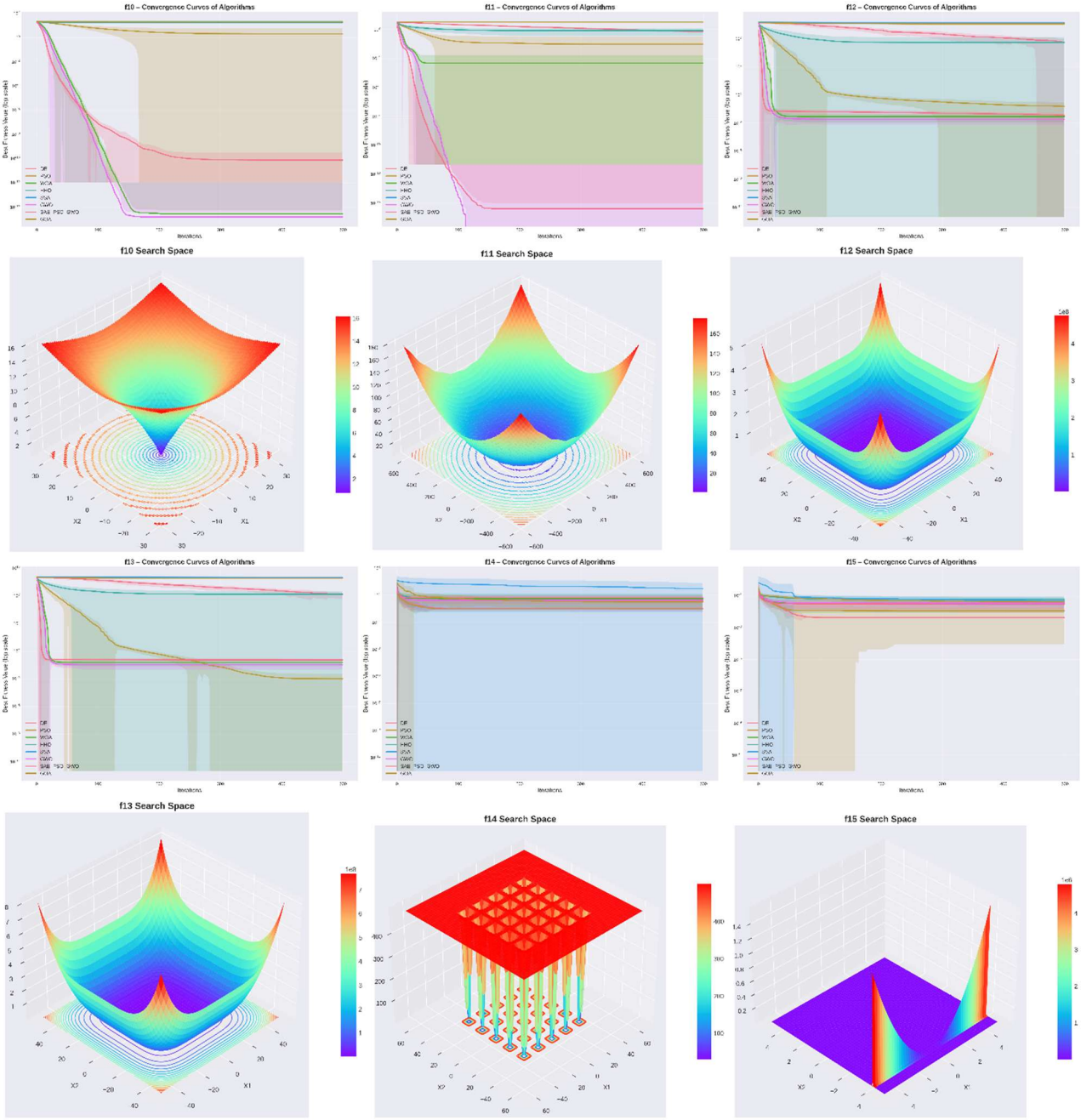


Figure 5. Search Space Diagrams and Convergence Curves of Different Algorithms on Test Functions (F1–F9)



ISSN: 2710-0588

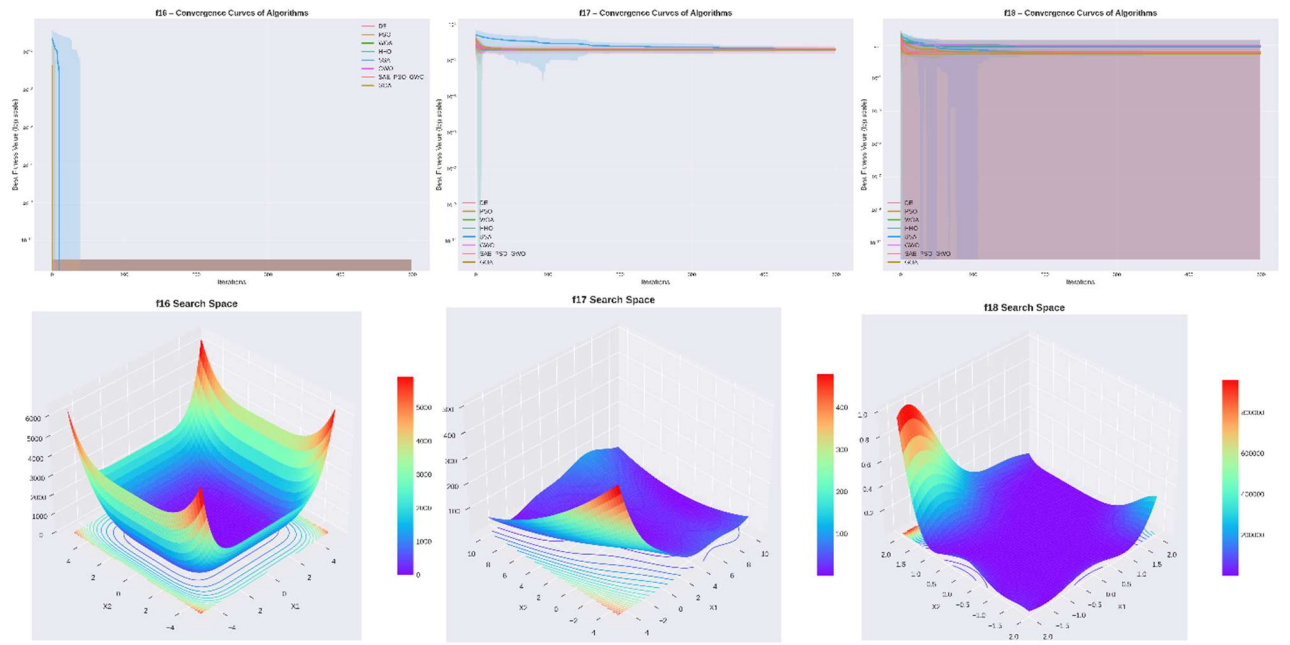


Figure 6. Search Space Diagrams and Convergence Curves of Different Algorithms on Test Functions (F10–F18)

To visually evaluate the search characteristics of the SAE_PSO_GWO algorithm, Figures 5 and 6 present the search space diagrams and convergence curves of different algorithms on the test functions. As illustrated, on various benchmark functions, the optimal solutions obtained by SAE_PSO_GWO exhibit highly dense distributions around the theoretical optimum, and its iterative trajectories demonstrate distinct spatial focusing characteristics. This indicates that the algorithm possesses strong directed search capability and optimization accuracy.

From the comparative convergence curves, it can be further observed that SAE_PSO_GWO exhibits fast and stable convergence behavior across different types of complex test functions. Its convergence curve consistently remains below those of other compared algorithms and approaches the theoretical optimum earlier, suggesting that the algorithm effectively balances global exploration and local exploitation during iteration, thereby avoiding premature convergence to local optima. Considering its performance on multiple complex functions, SAE_PSO_GWO demonstrates superior convergence efficiency and stability, further validating its effectiveness and robustness in handling high-dimensional, multimodal, and other complex optimization problems.

4. Experiments

4.1. Dataset

The dataset[32] used in this study comprises four years of hourly wind power generation data and weather conditions from Romania (longitude 26.1°, latitude 44.417°). This dataset covers wind power generation data in Romania from January 1, 2019, to December 31, 2023, with an hourly temporal resolution, totaling 43,829 data points. Table 1 presents the relevant statistical information. All data are partitioned into an 80% training set and a 20% test set.

ISSN: 2710-0588

Table 9. Statistical Properties of the Dataset

	Mean(MW)	STD(MW)	Min(MW)	25%	50%	75%	Max(MW)
Wind speed	1.20	0.89	0	1	1	2	7
Visibility	9632.12	2072.01	0	10000	10000	10000	65000
Temperature	13.52	9.54	-11.7	5.6	13	20.7	40.8
Humidity	7.01	7.11	-17.4	1.7	7.2	12.8	25.4
Air pressure	1016.66	12.34	0	1011.8	1016.1	1021.6	1044.7
Precipitation1	0.09	0.92	0	0	0	0	32
Precipitation2	0.00	0.06	0	0	0	0	7
Snowfall	0.04	0.63	0	0	0	0	23
Wind	781.31	656.37	0	241	588.4	1184	2623

4.2. Model Training Configuration

To verify the effectiveness of the proposed SAE-PSO-GWO hybrid optimization algorithm, this section applies it to the hyperparameter optimization task of a time-series forecasting model and compares it with the classical GWO algorithm under the same experimental settings. The relevant algorithmic parameters are configured as shown in Table 10.

Table 10. Optimization Algorithm Parameters

Parameter Name	Symbol	GWO	SAE-PSO-GWO	Description
Population Size	N	6	8	Number of candidate solutions per generation
Maximum Iterations	$T_{\max}$	10	15	Termination condition for optimization
Search Dimension	D	6	6	Corresponds to six network hyperparameters
Initial Convergence Factor	a	2	Adaptive	Balances exploration and exploitation
Convergence Factor Update	a_t	Linear Decrease	Linear Decrease	$a_t = 2\left(1-\frac{t}{T_{\max}}\right)$
Leader Wolf	Alpha	1	1 (Elite)	Current best solution
Inertia Weight	w	—	0.7	Parameter in the PSO phase
Learning Factors	c_1, c_2	—	1.5, 1.5	Individual/Social learning factors

During the hyperparameter optimization process, this study selected six key hyperparameters that most significantly impact model predictive performance for tuning. The rationale for their selection is as follows: the number of LSTM units directly determines the model's capacity for modeling temporal dependencies; the learning rate controls the step size of gradient descent and is a core parameter affecting training stability and convergence speed; the batch size influences the variance of gradient estimates and the model's generalization ability; the Dropout rate is employed to prevent overfitting in complex networks; the number of convolutional kernels and the scaling factor jointly determine the width and receptive field size for multi-scale feature extraction, which is crucial for capturing the multi-scale fluctuations of wind speed. These hyperparameters are coupled with each other, collectively forming a high-dimensional, non-convex optimization space. The symbols, types, and reasonable search ranges determined through preliminary experiments for each hyperparameter are detailed in Table 11. All optimization procedures were conducted within these defined boundaries to ensure process rationality and result comparability. The aforementioned parameter configuration aims to balance the algorithm's exploration and exploitation capabilities, ensuring an effective

ISSN: 2710-0588

search for high-quality hyperparameter combinations within a limited number of iterations, thereby enhancing model prediction accuracy and generalization performance.

Table 11. Hyperparameter Optimization Ranges

Hyperparameter Name	Symbol	Type	Lower Bound	Upper Bound	Description
LSTM Units	n_lstm	Integer	64	256	Capacity for temporal feature modeling
Learning Rate (log10)	log10(lr)	Continuous	-4	-2	Effective range: 1e-4 to 1e-2
Batch Size	B	Integer	8	64	Mini-batch gradient descent
Dropout Rate	p	Continuous	0.1	0.5	Prevents overfitting
Number of Conv. Kernels	F	Integer	16	64	Number of channels for multi-scale features
Conv. Kernel Scaling Factor	s	Continuous	0.5	1.5	Controls convolutional receptive field size

4.3. Experimental Results

4.3.1. Evaluation Metrics

To comprehensively evaluate the predictive performance of each model, four widely adopted regression evaluation metrics were employed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Symmetric Mean Absolute Percentage Error (SMAPE), and the coefficient of determination (R^2). The formulas for each metric are as follows:

(1) Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{12}$$

Where y_i represents the true value, $\hat{y}_i$ denotes the predicted value, and N is the number of samples. The MAE reflects the average absolute prediction error and is less sensitive to outliers.

(2) Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{13}$$

RMSE (Root Mean Square Error) amplifies the influence of larger errors through the squared term, making it an important metric for evaluating prediction stability.

(3) SMAPE (Symmetric Mean Absolute Percentage Error):

$$SMAPE = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \tag{14}$$

SMAPE measures prediction error in percentage terms, making it suitable for comparing data with different scales.

(4) Coefficient of Determination (R^2):

ISSN: 2710-0588

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{15}$$

where $\bar{y}$ is the mean of the actual values. R^2 represents the model's ability to explain the variance of the target variable. A value closer to 1 indicates a better fit of the model.

4.3.2. Experimental Results Analysis

To systematically evaluate the effectiveness of the SAE-PSO-GWO algorithm in hyperparameter optimization, this section applies it to the tuning task of a time-series prediction model. A comparative analysis is conducted against the non-optimized Multi-CNN-LSTM model (Baseline) and the model optimized by the classic GWO algorithm. Figure 5 visually illustrates the performance comparison trends of the three models across four key evaluation metrics (MAE, RMSE, SMAPE, and R^2), while Table 4-2 presents the specific numerical results for each metric.

Table 12. Evaluation Metrics for Different Models

Model	MAE	RMSE	SMAPE (%)	R^2
Multi-CNN-LSTM(Baseline)	67.61	145.14	14.90	0.9506
GWO-Multi-CNN-LSTM(GWO)	65.03	128.87	15.16	0.9611
SAE-PSO-GWO-Multi-CNN-LSTM(SAE-PSO-GWO)	57.89	116.87	14.17	0.9680

Based on the analysis of the chart, it can be concluded that the model optimized by the SAE-PSO-GWO algorithm achieves the best performance across all evaluation metrics, demonstrating significant improvement:

(1) In terms of prediction error, the MAE of the SAE-PSO-GWO model is 57.89, which is approximately 14.38% and 10.98% lower than that of the Baseline and GWO models, respectively. Its RMSE is 116.87, representing reductions of approximately 19.48% and 9.31% compared to the Baseline and GWO models, respectively. The simultaneous optimization of both metrics indicates that the model not only provides more accurate predictions on average but also exhibits stronger robustness to outliers and greater overall output stability.

(2) Regarding percentage error, the SMAPE of the SAE-PSO-GWO model is 14.17%, lower than those of the Baseline (14.90%) and GWO (15.16%) models. This indicates a smaller deviation between the predicted and actual values, further confirming the model's better adaptability to data of varying scales.

(3) In terms of model goodness-of-fit, the R^2 of the SAE-PSO-GWO model reaches 0.9680, significantly higher than those of the Baseline (0.9506) and GWO (0.9611) models. This suggests that the model possesses the strongest capability to explain data variance and can more effectively capture the underlying patterns and trends in time-series data.

In summary, the SAE-PSO-GWO algorithm achieves a better balance between global exploration and local exploitation during hyperparameter optimization by integrating the swarm learning mechanism of PSO and the hierarchical search strategy of GWO, thereby searching for significantly better configurations for the prediction model. Experimental results consistently demonstrate that the algorithm can effectively improve the comprehensive prediction performance of the model, exhibiting high application value and promotion potential.

ISSN: 2710-0588

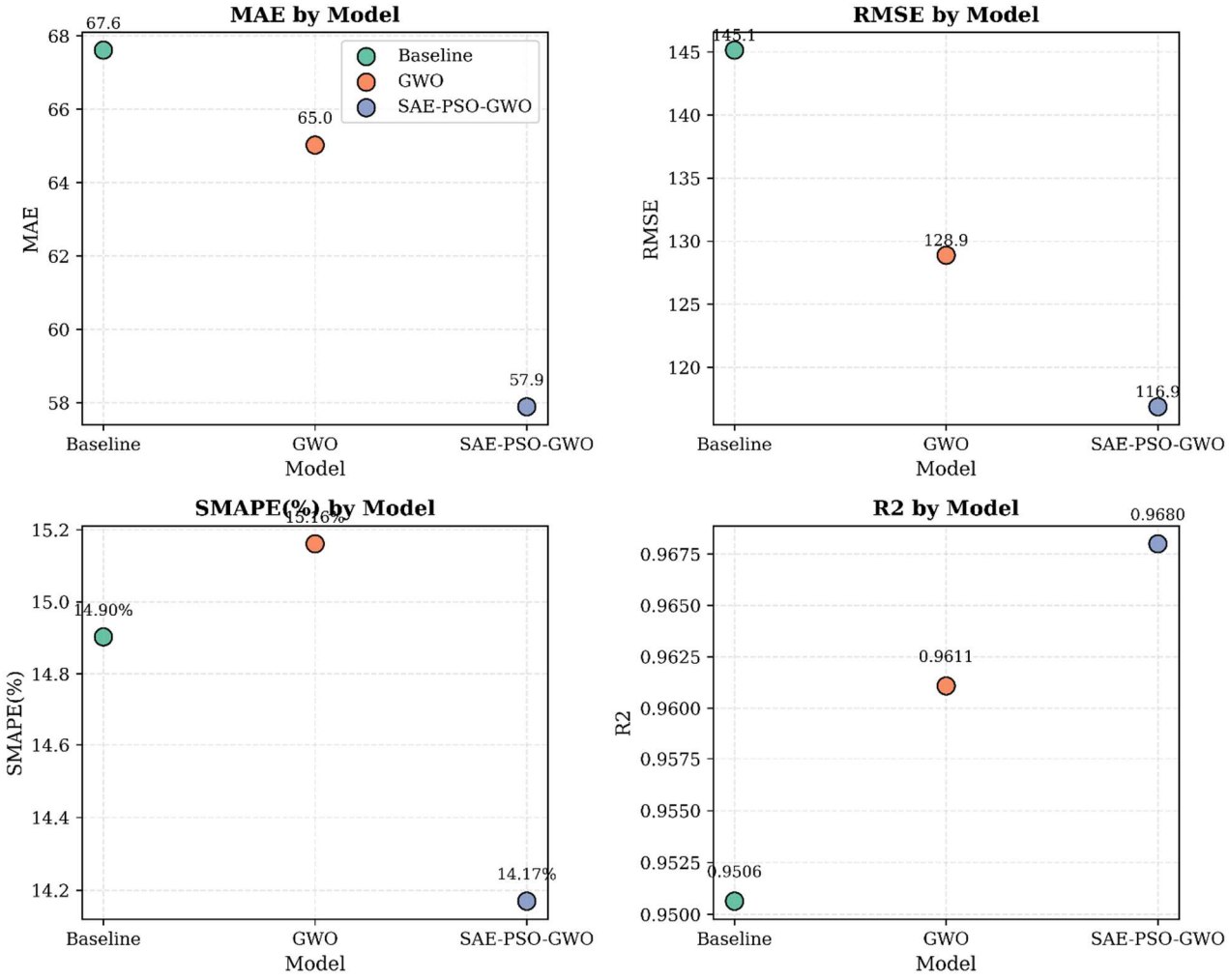


Figure 7. Evaluation Metrics of Different Models

5. Summary and Outlook

This paper addresses the difficulty of hyperparameter optimization for deep learning models in wind power prediction and proposes a short-term wind power prediction framework based on SAE-PSO-GWO hybrid swarm intelligence optimization. By deeply integrating the global exploration mechanism of the GWO algorithm and the local exploitation capability of the PSO algorithm, and introducing an elite opposition-based learning strategy to enhance population diversity, the proposed approach effectively overcomes the shortcomings of traditional optimization algorithms, such as premature convergence and insufficient accuracy, in complex high-dimensional parameter spaces. On this basis, the constructed Multi-CNN-LSTM prediction model utilizes a multi-scale convolutional structure to extract local and medium-to-long-term fluctuation features of the input sequence in parallel and models long-term temporal dependencies through an LSTM network, significantly enhancing the model's ability to capture the multi-scale nonlinear dynamics of wind power. Systematic experiments show that the proposed method exhibits superior optimization performance and prediction accuracy on both standard test functions and actual wind power data, providing a reliable technical approach for high-precision wind power prediction.

Looking ahead, this research can be further deepened in the following aspects: First, at the algorithmic level, the combination of SAE-PSO-GWO with deep learning surrogate models and parallelization strategies can be explored to improve its optimization efficiency and adaptability in higher-dimensional and dynamic environments. Second, at the model level,

ISSN: 2710-0588

attention mechanisms, Transformer architectures, or graph neural networks can be introduced to better integrate spatiotemporal features and multi-source meteorological information. Third, at the application level, efforts can be made to promote the engineering deployment of this framework in practical systems such as collaborative prediction for wind farm clusters and day-ahead and real-time rolling prediction, while also studying online learning mechanisms to enable autonomous updating and continuous optimization of the model under changing operational environments.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This is the place to fill in information about funds, sponsors, etc. that need to be thanked.
National Natural Science Foundation of China (No. 52404163),
Natural Science Foundation of Henan Polytechnic University (No. B2021-31),
Fundamental Research Funds for the Universities of Henan Province (No.NSFRF220415, NSFRF230503),
Henan Natural Science Foundation (No. 222300420168),
Science and Technology Plan Project of Henan Province (No. 232102211007, 252102221003),
Henan University of Technology Outstanding Youth Science Fund (J2025-3).

References

[1] Y. Liu, Y. Wang, Q. Wang, K. Zhang, W. Qiang, Q.H. Wen. “Recent advances in data-driven prediction for wind power”, Frontiers in Energy Research, Vol. 11, Article 1204343, 2023. https://doi.org/10.3389/fenrg.2023.1204343

[2] S.-X. Lv, L. Wang. “Multivariate wind speed forecasting based on multi-objective feature selection approach and hybrid deep learning model”, Energy, Vol. 263, Article 126100, 2023.

[3] Y. Chen, C. Lin, Y. Zhang, J. Liu, D. Yu. “Day-ahead load forecast based on conv2d-gru_sc aimed to adapt to steep changes in load”, Energy, Vol. 302, Article 131814, 2024. https://doi.org/10.1016/j.energy.2024.131814

[4] H. Wang, S. Han, Y. Liu, J. Yan, L. Li. “Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system”, Applied Energy, Vol. 237, pp. 1-10, 2019. https://doi.org/10.1016/j.apenergy.2018.12.076

[5] J. Bessac, E. Constantinescu, M. Anitescu. “Stochastic simulation of predictive space–time scenarios of wind speed using observations and physical model outputs”, The Annals of Applied Statistics, Vol. 12, pp. 432–458, 2018.

[6] P. Gomes, R. Castro. “Wind speed and wind power forecasting using statistical models: autoregressive moving average (ARMA) and artificial neural networks (ANN)”, International Journal of Sustainable Energy Development, Vol. 1, 2012.

[7] S. Nouhitehrani, E. Caro, J. Juan. “Computation of prediction intervals of wind energy based on the ewma and BOA techniques”, Sustainable Energy Technologies and Assessments, Vol. 66, Article 103806, 2024.

[8] F. Cassola, M. Burlando. “Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output”, Applied Energy, Vol. 99, pp. 154-166, 2012.

[9] K. Methaprayoon, C. Yingvivatanapong, W.J. Lee, J.R. Liao. “An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty”, IEEE Transactions on Industry Applications, Vol. 43, 2007.

ISSN: 2710-0588

[10] Y. Kassa, J.H. Zhang, D.H. Zheng, D. Wei. "A GA BP hybrid algorithm based ANN model for wind power prediction", Proceedings of the 2016 4th IEEE International Conference on Smart Energy Grid Engineering (SEGE), 2016.

[11] Y. Zhang, H. Sun, Y. Guo. "Wind power prediction based on PSO-svr and grey combination model", IEEE Access, Vol. 7, pp. 136254-136267, 2019.

[12] S.S. Soman, H. Zareipour, O. Malik, P. Mandal. "A review of wind power and wind speed forecasting methods with different time horizons", Proceedings of the North American Power Symposium 2010 (NAPS), 2010.

[13] P. Nikolaidis. "Wind power forecasting in distribution networks using non-parametric models and regression trees", Discover Energy, Vol. 2, No. 1, p. 6, 2022.

[14] R. Mahaseth, N. Kumar, A. Aggarwal, A. Tayal, A. Kumar, R. Gupta. "Short term wind power forecasting using k-nearest neighbour (KNN)", Journal of Information and Optimization Sciences, Vol. 43, No. 1, pp. 251-259, 2022.

[15] X. Xiong, X. Guo, P. Zeng, R. Zou, X. Wang. "A short-term wind power forecast method via xgboost hyper-parameters optimization", Frontiers in Energy Research, Vol. 10, Article 905155, 2022.

[16] Y. Wang, R. Zou, F. Liu, L. Zhang, Q. Liu. "A review of wind speed and wind power forecasting with deep neural networks", Applied Energy, Vol. 304, Article 117766, 2021.

[17] K. Wang, J. Zhang, X. Li, Y. Zhang. "Long-term power load forecasting using LSTM-informer with ensemble learning", Electronics, Vol. 12, No. 10, pp. 1–19, 2023.

[18] O. Bilal et al. "Differential evolution optimization based ensemble framework for accurate cervical cancer diagnosis", Applied Soft Computing, Vol. 167, Article 112366, 2024.

[19] J. Kennedy, R.C. Eberhart, Y. Shi. "The particle swarm", Swarm Intelligence, pp. 287-325, 2001.

[20] S. Mirjalili, A. Lewis. "The whale optimization algorithm", Advances in Engineering Software, Vol. 95, pp. 51-67, 2016.

[21] A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen. "Harris hawks optimization: algorithm and applications", Future Generation Computer Systems, Vol. 97, pp. 849–872, 2019. https://doi.org/10.1016/j.future.2019.02.028

[22] S. Mirjalili, A.H. Gandomi, S.Z. Mirjalili, S. Saremi, H. Faris, S.M. Mirjalili. "Salp swarm algorithm: A bio-inspired optimizer for engineering design problems", Advances in Engineering Software, Vol. 114, pp. 163-191, 2017.

[23] S. Mirjalili, S.M. Mirjalili, A. Lewis. "Grey wolf optimizer", Advances in Engineering Software, Vol. 69, pp. 46-61, 2014.

[24] C. Dey, R. Bose, K.K. Ghosh, S. Malakar, R. Sarkar. "LAGOA: Learning automata based grasshopper optimization algorithm for feature selection in disease datasets", Journal of Ambient Intelligence and Humanized Computing, Vol. 13, pp. 3175–3194, 2022.

[25] R. Tahmasebifar, M.P. Moghaddam, M.K. Sheikh-El-Eslami, R. Kheirollahi. "A new hybrid model for point and probabilistic forecasting of wind power", Energy, Vol. 211, Article 119016, 2020.

[26] Y. Zhou, R. Huang, Q. Lin, Q. Chai, W. Wang. "Probabilistic optimization based adaptive neural network for short-term wind power forecasting with climate uncertainty", International Journal of Electrical Power and Energy Systems, Vol. 157, Article 109897, 2024.

[27] X. Wang, J. Li, L. Shao, H. Liu, L. Ren, L. Zhu. "Short-term wind power prediction by an extreme learning machine based on an improved Hunter–Prey optimization algorithm", Sustainability, Vol. 15, No. 2, Article 991, 2023. https://doi.org/10.3390/su15020991

[28] A.S. Devi, G. Maragatham, K. Boopathi, A.G. Rangaraj. "Hourly day-ahead wind power forecasting with the EEMD-CSO-LSTM-EFG deep learning technique", Soft Computing, Vol. 24, No. 16, pp. 12391–12411, 2020. https://doi.org/10.1007/s00500-020-04680-7

[29] H. Li, H. Jing, R. Zhang, Z. Gao. "Wind power forecast based on improved Long Short Term Memory Network", Energy, Vol. 189, Article 116300, 2019. https://doi.org/10.1016/j.energy.2019.116300

ISSN: 2710-0588

[30] Y. Ding, Z. Chen, H. Zhang, X. Wang, Y. Guo. "A short-term wind power prediction model based on CEEMD and WOA-KELM", Renewable Energy, Vol. 189, pp. 188–198, 2022. https://doi.org/10.1016/j.renene.2022.02.108

[31] Z. Shang, Y. Chen, Y. Chen, Z. Guo, Y. Yang. "Decomposition-based wind speed forecasting model using causal convolutional network and attention mechanism", Expert Systems with Applications, Vol. 223, Article 119878, 2023. https://doi.org/10.1016/j.eswa.2023.119878

[32] Stefan Comanita,Hourly Electricity Consumption and Production.Hourly Electricity Consumption and Production by Type in Romania for 5.5 years,https://www.kaggle.com/datasets/stefancomanita/hourly-electricity-consumption-and-production.