

Article

Short-Term Wind Power Prediction Based on Improved SAO-Optimized LSTM

Zuoquan Liu ^{1,*}, Xinyu Liu ^{2,*} and Haocheng Zhang ^{2,*}

¹ Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China

² School of Electrical Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450045, China

* Correspondence: zuoquan.liu@connect.polyu.hk (Z.L.); lxy@ncwu.edu.cn (X.L.); zhc923532798@163.com (H.Z.)

Abstract

To enhance the accuracy of short-term wind power forecasting, this study proposes a hybrid model combining Northern Goshawk Optimization (NGO)-optimized Variational Mode Decomposition (VMD) and an Improved Snow Ablation Optimizer (ISAO)-optimized Long Short-Term Memory (LSTM) network. Initially, NGO is applied to determine the optimal parameters for VMD, decomposing the original wind power series into multiple frequency-based subsequences. Subsequently, ISAO is employed to fine-tune the hyperparameters of the LSTM, resulting in an ISAO-LSTM prediction model. The final forecast is obtained by reconstructing the subsequences through superposition. Experiments conducted on real data from a wind farm in Ningxia, China demonstrate that the proposed approach significantly outperforms traditional single and combined models, yielding predictions that closely align with actual measurements. This validates the method's effectiveness for short-term wind power prediction and offers valuable data support for optimizing microgrid scheduling and capacity planning in wind-integrated energy systems.

Keywords: short-term wind power prediction; variational modal decomposition; long- and short-term memory networks; northern goshawk optimization algorithm; snow ablation optimizer



Academic Editor: Krzysztof Rogowski

Received: 27 April 2025

Revised: 17 June 2025

Accepted: 1 July 2025

Published: 9 July 2025

Citation: Liu, Z.; Liu, X.; Zhang, H. Short-Term Wind Power Prediction Based on Improved SAO-Optimized LSTM. *Processes* **2025**, *13*, 2192. <https://doi.org/10.3390/pr13072192>

Copyright: © 2025 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

Since the proposal of the “dual-carbon” goal, China has actively advanced the development of renewable energy to address climate change and enhance energy security. Among these sources, wind power plays a crucial role due to its renewable nature, low cost, and zero emissions, leading to its growing presence in the national energy mix [1]. Nevertheless, the inherent variability and unpredictability of wind power often result in generation intermittency, which complicates energy utilization and poses technical challenges for grid integration. Accurate forecasting of wind power can help overcome these issues by supporting effective grid scheduling and ensuring the stable operation of wind-connected systems [2,3].

Short-term wind power forecasting models are generally categorized into three types: physical models, statistical models, and hybrid (combined) models [4]. Physical models [5] rely on meteorological forecasts and turbine characteristics specific to geographic locations. Although useful for long-term predictions, their reliance on extensive and complex weather data limits their accuracy and practicality for short-term forecasting. Statistical models [6], on the other hand, establish data-driven relationships between inputs—such as numerical

weather forecasts and historical generation data—and predicted outputs. These models are simpler to construct and often yield better accuracy and generalization than physical models. However, their performance can vary across different time scales, especially when handling nonlinear patterns, potentially resulting in larger prediction errors. To overcome the limitations of both approaches, hybrid models integrate optimization algorithms with prediction techniques [7], enabling more accurate modeling of wind speed and power fluctuations. By tuning model parameters through optimization, these combined approaches enhance forecasting accuracy and robustness, meaning that they are widely adopted in modern wind power prediction tasks.

At present, conventional single prediction models include the neural network prediction model [8], the integrated learning [9] prediction model and the Random Forest [10] (RF) prediction model. Optimization algorithms mainly include Particle Swarm Optimization [11] (PSO), Genetic Algorithm (GA) and Differential Evolution (DE).

An enhanced PSO-BP model was introduced [12], where Particle Swarm Optimization (PSO) was used to optimize the weights and thresholds of the BP neural network. However, due to the BP network's reliance on gradient descent, it is prone to getting trapped in local optima, indicating the need for more advanced neural network architectures to address this limitation. In [13], a forecasting method utilizing the Spotted Hyena Algorithm (SHA) was proposed to optimize the penalty coefficients and kernel parameters of the Support Vector Machine (SVM), resulting in notable improvements in prediction accuracy and stability. Ref. [14] presented a wind power prediction model combining the Sparrow Search Algorithm (SSA) with a Gated Recurrent Unit (GRU), where SSA is employed to fine-tune model parameters through iterative optimization, enhancing forecasting performance. These hybrid approaches illustrate the effectiveness of integrating optimization algorithms with predictive models to address the uncertainty and volatility inherent in wind power forecasting.

Given the stochastic, volatile, and long-tailed characteristics of wind power sequences, many current hybrid models incorporate signal decomposition techniques to simplify data structure and enhance forecasting accuracy. These methods help reduce the complexity and randomness of the original signal, making them widely adopted in wind power data preprocessing. Common techniques include Fourier Transform [15] (Fast Fourier Transform, FFT), Empirical Mode Decomposition (EMD) [16], and others. For instance, Ref. [17] presents an EMD-based hybrid model that integrates an improved GA-BP algorithm with Adaboost, introducing a novel hidden layer node selection strategy. By applying EMD to obtain decomposed input data, the model improves prediction accuracy by capturing relationships between components and the output. However, EMD, despite its suitability for nonlinear and non-smooth signals, suffers from issues like mode mixing and sensitivity to boundary effects, limiting its robustness. In contrast, Variational Mode Decomposition (VMD) [18] offers a more stable approach by breaking down highly volatile and random sequences into multiple smooth and regular subcomponents, effectively preserving the intrinsic power characteristics of the original data while enhancing consistency.

To address the challenge of improving both accuracy and stability in short-term wind power forecasting, this study proposes a hybrid prediction model that integrates the Northern Goshawk Optimization (NGO) algorithm and an Improved Snow Ablation Optimizer (ISAO) within a VMD-LSTM framework. First, Variational Mode Decomposition (VMD) is applied to break down the original wind power sequence into multiple components. The NGO algorithm is employed to optimize the key VMD parameters $[k, \alpha]$, using minimum alignment entropy as the fitness criterion, thus enhancing the decomposition's effectiveness by reducing noise and ensuring more stable input for the forecasting stage. Next, a separate LSTM model is constructed for each decomposed component, with ISAO used to fine-tune the LSTM hyperparameters, resulting in the ISAO-LSTM forecasting model.

After training with the optimized parameters, the final output is generated through the reconstruction of individual predictions. Experimental results confirm that the proposed approach significantly enhances forecasting accuracy and robustness.

2. Raw Wind Power Decomposition

2.1. VMD Decomposition

VMD [19] is a fully non-recursive signal decomposition method, proposed in 2014 by Konstantin Dragomiretskiy. VMD, as a fully non-recursive model, is essentially the process of building variational problems and then solving them to realize the decomposition of non-smooth sequences. The following are the three specific steps of the model:

(1) Construct the variational

Assuming that the original signal can be decomposed into components, the objective is to minimize the total estimated bandwidth of the IMF [20]. In addition, the constraints need to be satisfied at the same time, expressed as

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ \text{s.t. } \sum_k u_k(t) = f(t) \end{cases} \quad (1)$$

where k is the number of modes to be decomposed; $\{\omega_k\}$ is the center frequency; ∂_t is the partial derivative; $\delta(t)$ is the DeLillek function; and $*$ is the convolution operation.

(2) Constrained variational

In order to obtain the optimal solution, the Lagrange multiplier and the second-order penalty factor α are introduced, and the expression is

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 \\ & \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (2)$$

where λ is the Lagrange multiplier; and α is the second-order penalty factor.

(3) Solving the variational components

The nonlinear normalized model is solved distributively using the alternating direction multiplier method [21], where all components can be obtained by Fourier isometric transformation with the expression

$$\hat{u}_k^{n_1+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\omega_k^{n_1+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)| d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (4)$$

where $\hat{u}_k^{n_1+1}(\omega)$, $\hat{f}(\omega)$, $\hat{u}_k(\omega)$, $\hat{\lambda}(\omega)$ and are the Fourier transforms of the components, respectively.

2.2. The NGO-VMD Model

Considering that the VMD is not adaptive, the number of modes, the penalty factor α , the fidelity coefficient, and other parameters need to be set manually. If the number of modes is selected improperly, overfitting or loss of key information will occur, while the

improper selection of the penalty factor α will lead to distortion and slow convergence, etc. In this study, the VMD is optimized by using NGO.

In this study, we adopt NGO to optimize VMD, using Minimum Permutation Entropy (MPE) as the fitness function to establish the optimization model.

2.2.1. Entropy of a Permutation

Alignment entropy [22] quantifies the level of vector aggregation in high-dimensional space and serves as an indicator of time-series regularity. A lower entropy value corresponds to a more structured and predictable sequence. To accurately determine the optimal parameter combination $[k, \alpha]$, this study adopts minimum alignment entropy as the fitness function. The calculation process is outlined as follows:

$$H_{PE}(m) = - \sum_{q=1}^S P_q \ln P_q \quad (5)$$

where P_q is the probability of the time series, and q is the position before being aligned.

2.2.2. Northern Goshawk Optimization Algorithm

The Northern Goshawk Optimization Algorithm [23] is a population-based optimization algorithm proposed by Dehghani et al. in 2021, inspired by the hunting behavior of hawks, which consists of two phases: identifying attacks (global search) and pursuing actions (local search).

(1) Recognize Attack (Exploration)

In this phase, the northern hawk randomly selects prey in the search space and launches an attack. This step aims to perform a global search to identify the optimal region. Equation (6) represents the new state in the j th dimension, and Equation (7) represents the population update formula:

$$X_{i,j}^{\text{new},P1} = \begin{cases} x_{i,j} + r(p_{i,j} - Ix_{i,j}), & F_{P_i} < F \\ x_{i,j} + r(x_{i,j} - p_{i,j}), & F_{P_i} \geq F \end{cases} \quad (6)$$

$$X_i = \begin{cases} X_i^{\text{new},P1}, & F_i^{\text{new},P1} < F_i \\ X_i, & F_i^{\text{new},P1} \geq F_i \end{cases} \quad (7)$$

where F_i is the value of the objective function, $F_i^{\text{new},P1}$ is the value of the objective function of the stage, r and I are the random numbers generated during the iteration process.

(2) Manhunt (development)

This stage enhances the algorithm's ability to exploit the search space by simulating the trailing and chasing process of the northern goshawk. Assuming that the hunting range radius of the northern goshawk is, according to Equation (8), the updating formula for the search radius, Equation (9) is the state updating formula for the first dimension, and Equation (10) is the updating formula for the population membership:

$$R = 0.02(1 - \frac{t}{T}) \quad (8)$$

$$x_{i,j}^{\text{new},P2} = x_{i,j} + R(2r - 1)x_{i,j} \quad (9)$$

$$X_i = \begin{cases} X_i^{\text{new},P2}, & F_i^{\text{new},P2} < F_i \\ X_i, & F_i^{\text{new},P2} \geq F_i \end{cases} \quad (10)$$

where t is the current number of iterations, T is the maximum number of iterations, and $F_i^{\text{new},P2}$ is the value of the objective function of the stage

2.3. NGO-Optimized VMD

After determining the required parameters and of the NGO algorithm, the population is randomly initialized, by obtaining the minimum fitness value of the population and setting the minimum alignment entropy as the fitness function. Subsequently, a global search is performed on the minimum arrangement entropy. The output is the optimal $[k, \alpha]$. This series of steps aims to achieve effective exploration and optimization of the search space to obtain the best results.

The optimization flow of the NGO optimization VMD is shown in Figure 1.

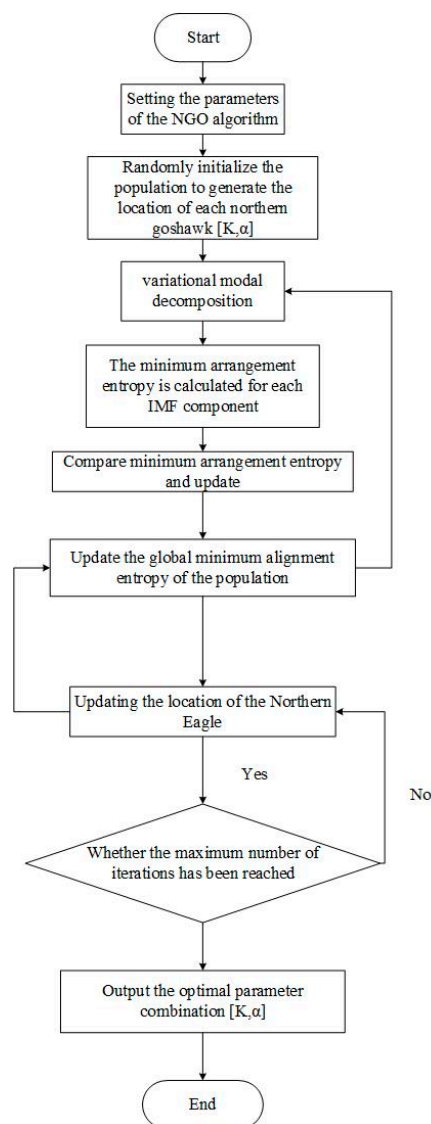


Figure 1. NGO optimization of VMD process.

3. Short-Term Wind Power Forecasting Model Based on NGO-VMD-ISAO-LSTM

Accurate and reliable wind power forecasting is crucial for the efficient scheduling of microgrids, as it directly influences energy management, storage planning, and operational costs. However, due to the nonlinear, non-stationary, and noisy nature of wind power data, traditional approaches—such as physical models, statistical methods, and

standalone artificial intelligence techniques—often struggle with limited generalization and inconsistent performance, making them inadequate for real-time microgrid operations. In contrast, hybrid models that integrate signal decomposition with optimization algorithms address these challenges more effectively. By extracting components at different frequency levels and simultaneously tuning model parameters, these approaches overcome the shortcomings of individual models and significantly enhance forecasting accuracy and stability. Therefore, advancing research on such combined models holds both theoretical and practical value for improving prediction precision, refining microgrid scheduling strategies, and boosting overall system efficiency.

3.1. Snow Ablation Optimizer

The Snow Ablation Optimizer (SAO) [24], introduced by Lingyun Deng et al. in 2023, is a meta-heuristic algorithm inspired by the natural processes of snow sublimation and melting. Its design aims to balance global exploration and local exploitation within the solution space. SAO operates through four key stages: initialization, exploration, exploitation, and a dual-population strategy to enhance search efficiency and diversity.

(1) Initialization phase

In SAO, the iterative process begins with a randomly generated population. Equation (11) describes the entire population, which is usually modeled as a matrix containing an operation vector and rows and columns.

$$Z = L + \theta \times (U - L)$$

$$= \begin{bmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,Dim-1} & z_{1,Dim} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,Dim-1} & z_{2,Dim} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{N-1,1} & z_{N-1,2} & \cdots & z_{N-1,Dim-1} & z_{N-1,Dim} \\ z_{N,1} & z_{N,2} & \cdots & z_{N,Dim-1} & z_{N,Dim} \end{bmatrix}_{N \times Dim} \quad (11)$$

where L and U are the lower and upper bounds of the solution space, respectively, and θ denote the random numbers generated in $[0, 1]$.

(2) Exploration phase

During the transition of snow or meltwater into vapor, the resulting movement is irregular and highly dispersed. This behavior is mimicked in the exploration phase of the algorithm through Brownian motion, which captures the randomness observed in the sublimation process.

In standard Brownian motion, the step size is determined using a probability density function derived from a normal distribution with a mean of zero and a variance of one. Consequently, the displacement at any moment follows this distribution, and is mathematically described as

$$f_{BM}(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \times \exp\left(-\frac{x^2}{2}\right). \quad (12)$$

Brownian motion utilizes dynamics and uniform step sizes, making it an exploration tool for exploring potential regions in space, which can be well modeled in the process of vapor diffusion. The position of the exploration process is calculated as

$$Z_i(t+1) = Elite(t) + BM_i(t) \otimes (\theta_1 \times (G(t) - Z_i(t)) + (1 - \theta_1) \times (\bar{Z}(t) - Z_i(t))) \quad (13)$$

where $Z_i(t)$ is the position of the first particle in the first iteration; $BM_i(t)$ is a vector of random numbers based on Gaussian distribution to represent the Brownian motion; \otimes is the multiplication by rows; θ_1 is the random number generated in $[0, 1]$.

The solution formula for the location of the center of mass of the group is

$$Z_c(t) = \frac{1}{N_1} \sum_{i=1}^{N_1} Z_i(t) \quad (14)$$

$$Elite(t) \in [G(t), Z_{second}(t), Z_{third}(t), Z_c(t)] \quad (15)$$

where $G(t)$ is the current optimal particle; $Elite(t)$ is a random individual among several elite groups in the overall population; $\bar{Z}(t)$ is the center-of-mass position of the individual whose fitness value is ranked in the top 50% of the whole population; $Z_{second}(t)$ and $Z_{third}(t)$, denote the second and third best individuals in the current population, respectively.

In each iteration, a selection is made at random from a set that includes the current best solution, the second and third top-performing individuals, and the center-of-mass position of the leading factor.

(3) Development phase

When snow melts into liquid water, the process is often modeled around the current optimal solution using snowmelt simulation techniques. A widely used approach for capturing this behavior is the classical degree-day method, which effectively represents the dynamics of snowmelt. The position update equation for this stage is

$$Z_i(t+1) = M \times G(t) + BM_i(t) \otimes (\theta_2 \times (G(t) - Z_i(t)) + (1 - \theta_2) \times (\bar{Z}(t) - Z_i(t))) \quad (16)$$

where θ_2 is a random number in $[-1, 1]$; and M is the degree-day snowmelt model.

The general form of the method is

$$M = (0.35 + 0.25 \times \frac{1}{e^{\frac{t}{t_{\max}}} - 1}) \times e^{\frac{-t}{t_{\max}}} \quad (17)$$

where t is the current iteration number; and t_{\max} is the maximum iteration number.

In each iteration, update the expression for the degree-day factor (DDF) as

$$DDF = 0.35 + 0.25 \times \frac{e^{\frac{t}{t_{\max}}} - 1}{e - 1} \quad (18)$$

where DDF is the degree-day factor, which ranges from 0.35 to 0.6.

(4) Dual-population mechanism

The SAO algorithm distinguishes itself from other optimization methods through its dual-population mechanism, effective exploration–exploitation strategy, and adaptable position-update process. These characteristics enhance its ability to maintain a strong balance between global and local searches, improve convergence efficiency, and adapt to complex challenges, particularly in multi-modal and high-dimensional scenarios like short-term wind power forecasting.

$$Z_i(t+1) = \begin{cases} Elite(t) + BM_i(t) \otimes (\theta_1 \times (G(t) - Z_i(t)) \\ \quad + (1 - \theta_1) \times (\bar{Z}(t) - Z_i(t))), i \in index_a \\ M \times G(t) + BM_i(t) \otimes (\theta_2 \times (G(t) - Z_i(t)) \\ \quad + (1 - \theta_2) \times (\bar{Z}(t) - Z_i(t))), i \in index_b \end{cases} \quad (19)$$

3.2. Improved Snow Abatement Optimizer

Although the SAO algorithm demonstrates strong performance in optimization tasks, it still faces challenges such as limited convergence accuracy and a tendency to fall into

local optima. To address these issues, this section introduces an Improved Snow Ablation Optimizer (ISAO), which enhances global search ability and convergence precision through the integration of multiple optimization strategies. These improvements effectively mitigate the original algorithm's shortcomings and offer a more robust and efficient approach for solving complex optimization problems.

(1) Sinusoida Chaos Mapping Initialization

In traditional SAO, random initialization often leads to limited initial population diversity due to the lack of uniform distribution characteristics, which in turn leads to premature convergence problems. In this study, Sinusoida chaotic mapping was introduced to improve the initialization process of SAO. Sinusoida chaotic mapping generates pseudo-random sequences with ergodicity and non-repeatability through nonlinear iterative equations. Its mathematical expression is

$$x_{n+1} = \mu \cdot \sin(\pi x_n) \quad (20)$$

where the system exhibits typical chaotic properties for the control parameter $\mu \in [0, 2.3]$. Compared with other traditional chaotic models, Sinusoida mapping exhibits better traversal uniformity and dynamic range coverage ability in the parameter space.

(2) Levy flight strategy

In traditional SAO, when reaching a certain number of iterations, the exploration phase will be transformed into the development phase, and at this time, the fitness function value is no longer changed. In order to avoid falling into the local optimum, the Levy flight mechanism was introduced to update the exploration and development phases in order to improve the global search capability. The specific implementation is as follows.

The Mantegna algorithm is used to generate the Levy step, and the mathematical expression is

$$s = 0.01 \cdot \frac{u \otimes \sigma}{|v|^{1/\beta}} \quad (21)$$

where u, v is the independent normally distributed random vector; β is the Levy index; and σ is the scaling factor.

The original Brownian motion based on Gaussian distribution was replaced with a Levy flight, and the original exploration phase formula was updated to

$$X_i^{t+1} = X_i^t + r_1 \cdot S \odot (C_{\text{centroid}} - X_i^t) \quad (22)$$

where S is the step vector generated by the Levy flight; $r_1 \in [0, 1]$ is a uniform random number; and C_{centroid} is the center of mass of the elite population.

The long-jump property of the Levy flight was combined with a local dense search to effectively balance global dispersion and local concentration in the exploration phase. By adjusting the scaling factor of s (e.g., 0.01), the step size magnitude can be controlled to avoid excessive deviation from the potential optimal region.

3.3. Long- and Short-Term Memory Networks

Long Short-Term Memory (LSTM) [25] is a variant of Recurrent Neural Networks (RNNs), which enables the management of memory units by introducing special storage units and gate mechanisms to better capture long-term dependencies in sequential data.

The basic structure of LSTM is shown in Figure 2. The LSTM cell contains the states of the memory cells as well as three gate control structures: forgetting gates, input gates, and output gates. The input forgetting gate performs selective forgetting, decides which information gets stored by computation, and acquires new cells by constantly updating the computation.

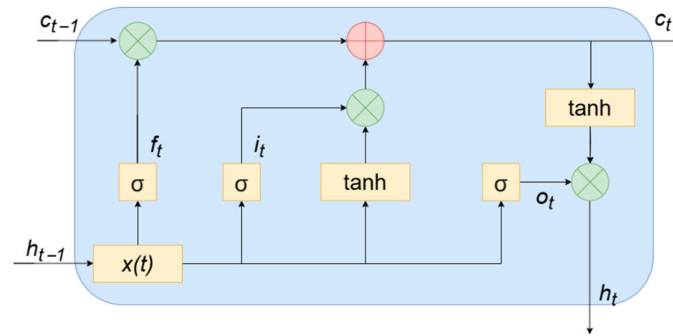


Figure 2. LSTM network architecture.

In Figure 2, h_{t-1} , h_t are the hidden layer vectors at the time $t - 1$ and the time t , respectively; c_{t-1} , c_t are the cell states at the time $t - 1$ and the time t , respectively; x_t is the input at the time t ; σ is the Sigmoid activation function, with the value domain of $[0, 1]$; and \tanh is the hyperbolic tangent activation function, with the value domain of $[-1, 1]$. The formulas for the input gate, forget gate and output gate are shown below.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (23)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (24)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (25)$$

where W_i , U_i is the weight assigned to the input gate; b_i is the bias of the input gate; W_f , U_f is the weight assigned to the forget gate; b_f is the bias of the forget gate; W_o , U_o is the weight assigned to the output gate; and b_o is the bias of the output gate.

The metameric state is updated at time t . The state is computed as

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (26)$$

Candidates for the new cell state information store are

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (27)$$

where W_c , U_c is the weight assigned to the candidate representative; and b_c is the bias of the candidate representative.

The implicit layer state at the time output is computed as

$$h_t = o_t \cdot \tanh(c_t) \quad (28)$$

3.4. ISAO-Optimized LSTM

There are difficulties in the selection of certain hyperparameters in LSTM, and the correct selection of hyperparameters often affects the overall prediction accuracy. Traditional methods usually learn the parameters initially and cross-validate them based on experience. In this study, we used the ISAO algorithm to optimize the parameters of LSTM, adaptively search for appropriate neural network parameters, reduce the difficulty of learning and prediction, and improve the accuracy of prediction.

The ISAO learning parameter optimization process for LSTM is shown in Figure 3.

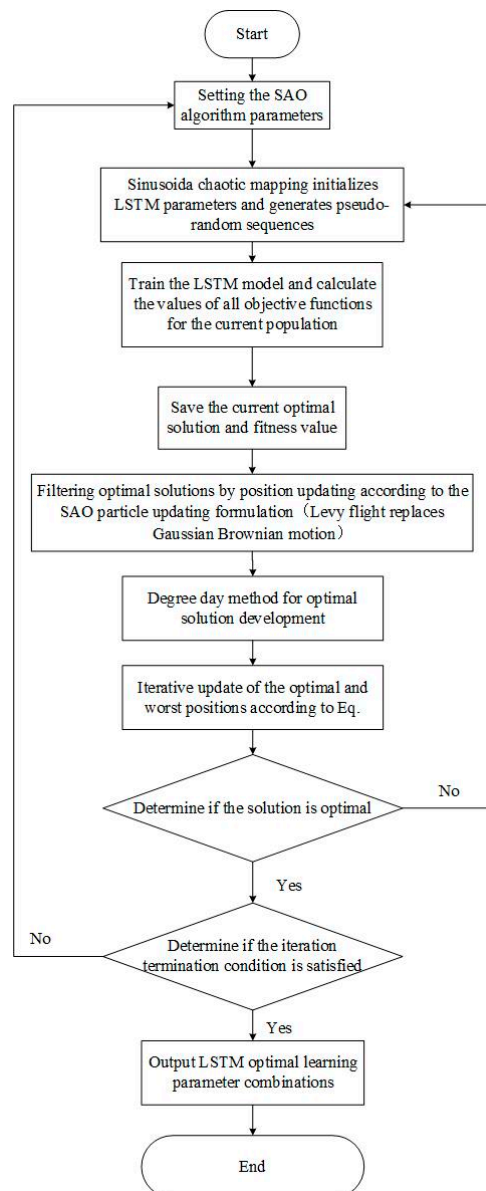


Figure 3. SAO-optimized LSTM process.

3.5. NGO-VMD-ISAO-LSTM Combined Prediction Model Building

This study combines the NGO optimization algorithm, VMD modal decomposition technique, ISAO optimization algorithm and LSTM prediction model to build the combined NGO-VMD-ISAO-LSTM model, and the specific steps are as follows:

- (1) The NGO algorithm optimizes the VMD parameters $[k, \alpha]$, enabling the decomposition of raw wind power data into multiple subsequences using the enhanced VMD method;
- (2) An ISAO-LSTM prediction model is constructed for each IMF component obtained from the decomposition. The ISAO algorithm adaptively tunes the neural network's hyperparameters, thereby enhancing forecasting accuracy;
- (3) The total prediction result is obtained by superposition reconstruction;
- (4) Appropriate indicators are selected to analyze the errors.

The flow chart of NGO-VMD-ISAO-LSTM prediction is shown in Figure 4.

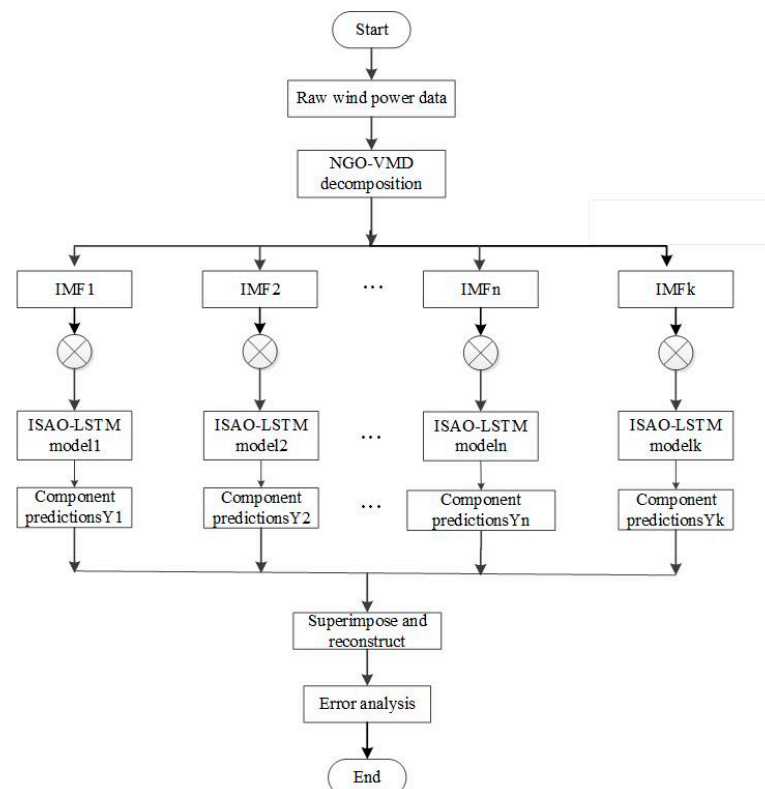


Figure 4. NGO-VMD-SAO-LSTM flowchart.

3.6. Evaluation Indicators

To evaluate the forecasting performance of the proposed model, three key metrics were selected. Their calculation formulas are as follows

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

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (30)$$

$$R^2 = 1 - \frac{\sum_{o=1}^N (\hat{y}_o - y_o)^2}{\sum_{o=1}^N (\bar{y} - y_o)^2} \quad (31)$$

4. Simulation Analysis

The data used in this paper were collected over a 10-day period from 25 September to 5 October 2021 at a wind farm in Ningxia, China. The wind power data were sampled at 15 min intervals, resulting in a total of 1056 sets of data. Wind speed, wind direction, temperature and pressure were selected as input features. The first 80% of the data was selected as the training set and the remaining 20% as the test set, and the prediction time span was 45 h. Because wind power data are often affected by problems such as weather conditions and equipment failures, which can adversely affect the subsequent modeling decision-making process, the collected data were preprocessed accordingly. In this study, a combination of linear interpolation and forward/backward padding was used: for smooth trends with little variation, linear interpolation usually provides smoother and more reasonable padding; for data with large variations or no obvious trends, forward or

backward padding can be considered, especially when missing values are at the beginning or end of the data. These methods can effectively fill in any gaps in the data to ensure data integrity and continuity.

In this study, the hardware equipment used for the wind power prediction experiment included an Nvidia GeForce RTX 4090 model graphics card, an i9-13900k model CPU, etc., and the software platform used was MATLAB2022B version.

4.1. Analysis of Single LSTM Model Prediction Results

To demonstrate the advantages of the LSTM prediction model proposed in this paper over other individual models in short-term wind power forecasting, ablation experiments were conducted using BP neural networks and Convolutional Neural Networks (CNN). The performance metrics for each model are presented in Table 1.

Table 1. Evaluation indicators for each single prediction model.

Model	RMSE	MAE	R^2
BP	5.528	4.279	0.936
CNN	4.452	3.516	0.958
LSTM	4.309	3.418	0.961

As shown in Table 1, the LSTM model achieved lower RMSE and MAE values compared to the BP and CNN models, reflecting a noticeable improvement in prediction accuracy. This suggests that the LSTM model is better suited to handling wind power data with pronounced temporal characteristics. The forecasting results of each model are illustrated in Figure 5.

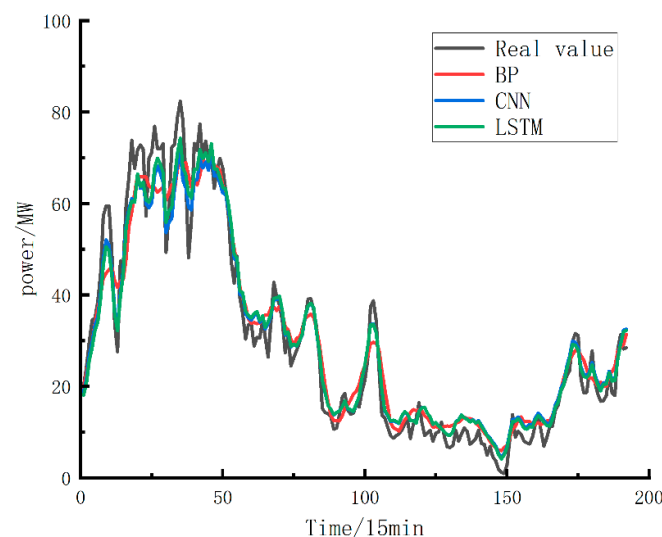


Figure 5. Prediction results of different single prediction models.

Figure 5 illustrates that, despite initial preprocessing, the non-stationary nature of wind power data still causes discrepancies between predicted and actual values at certain points, negatively impacting model fitting. Due to these challenges, relying solely on a single prediction model is insufficient. Therefore, the VMD algorithm was employed to decompose the data and reduce noise. Additionally, to address the difficulty in selecting LSTM hyperparameters and enhance prediction accuracy, the ISAO algorithm was introduced for hyperparameter optimization.

4.2. Analysis of VMD-SAO-LSTM Model Prediction Results

In order to verify the performance of the VMD-SAO-LSTM model, three models, VMD-LSTM, VMD-SAO-LSTM and VMD-ISAO-LSTM, were selected for the comparison of the prediction results. The sampling frequency of the VMD was 1000 Hz, the modal number of the center frequency rule of thumb was six, and the penalization factor α was adjusted to 3000 according to the smoothness of the signals. The convergence tolerance criterion was 10^{-7} with no DC part. The VMD decomposition results are shown in Figure 6.

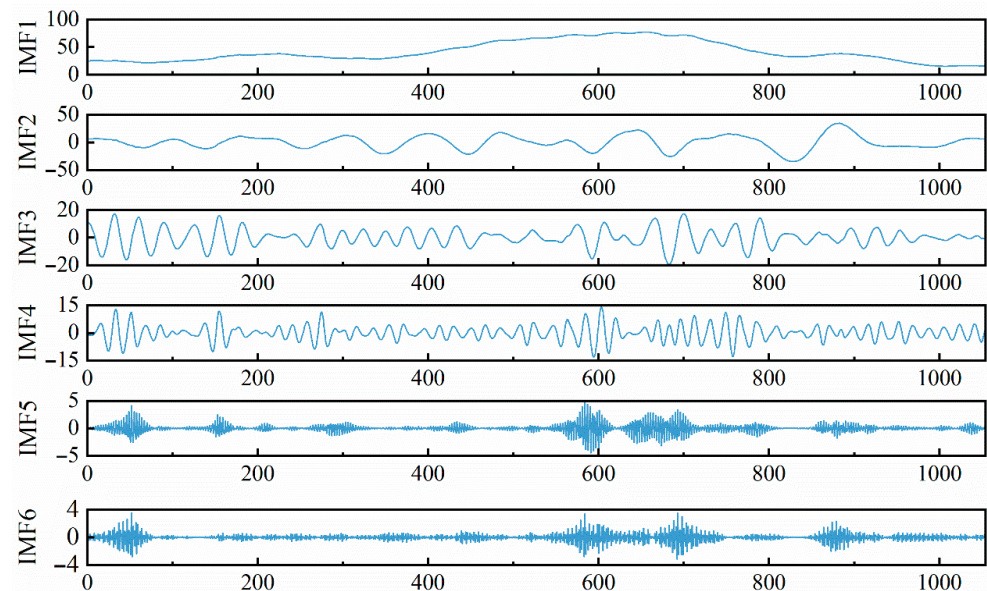


Figure 6. VMD decomposition results.

Both SAO and ISAO use a population size of 10, 30 iterations, and a degree-day factor of 0.35. These algorithms adaptively optimize the LSTM model's initial learning rate, hidden unit count, and L2 regularization parameter. The evaluation metrics for each combined prediction model are presented in Table 2.

Table 2. Evaluation indicators for various combination prediction models.

Model	RMSE	MAE	R^2
VMD-LSTM	3.363	3.150	0.973
VMD-SAO-LSTM	3.281	2.982	0.977
VMD-ISAO-LSTM	2.630	2.378	0.980

Combined with Table 2, it can be seen that after the introduction of VMD decomposition technology and the SAO optimization algorithm, all evaluation indexes were significantly improved. Specifically, the ISAO optimization algorithm improved RMSE and MAE by 5.1% and 6.8%, respectively, compared to the SAO optimization algorithm, which proves the feasibility of applying VMD decomposition technology and the ISAO optimization algorithm in the field of wind power prediction. The prediction results of each combined prediction model are shown in Figure 7.

Figure 7 shows that most predicted points closely follow the actual wind power trends. However, because VMD parameters are manually set and not adaptive, prediction efficiency is limited. To address this, this study introduced the NGO optimization algorithm to adaptively optimize VMD hyperparameters, identifying the best combination of modal number and penalty factor to enhance prediction accuracy.

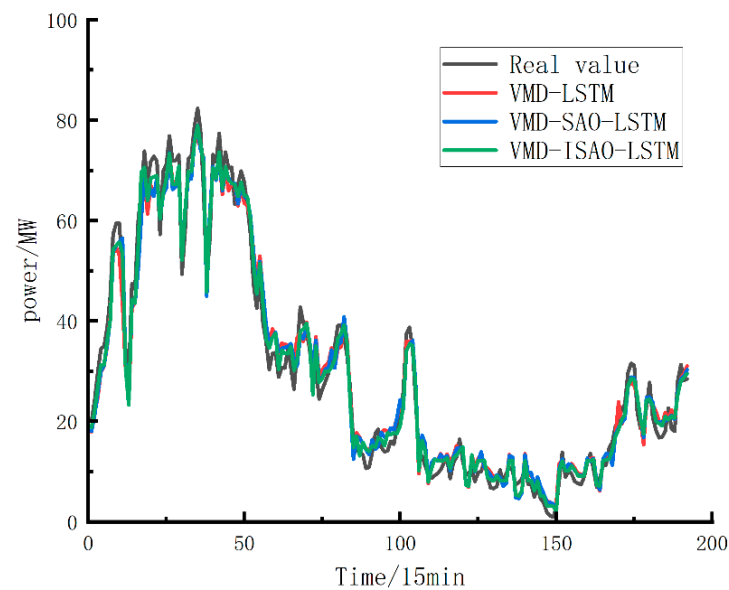


Figure 7. Prediction results of different combination prediction models.

4.3. Analysis of the Prediction Results of the NGO-VMD-SAO-LSTM Model

The VMD parameters, including sampling frequency, were set as described, while the optimal modal number and penalty factor were determined through NGO optimization. The NGO algorithm uses a population size of 10 and runs for a maximum of 30 iterations. Using minimum arrangement entropy as the fitness function, NGO adaptively finds the best combination of modal number and penalty factor—resulting in eight modes and an α value of 2867. The decomposition results of NGO-VMD are shown in Figure 8.

Figure 8 shows that the NGO-VMD model decomposed the original sequence into eight subsequences. IMF1 and IMF2, as the primary modes, exhibit smoother curves; IMF3, IMF4, and IMF5 are roughly symmetric, easing prediction difficulty; while IMF6, IMF7, and IMF8 capture the overall volatility of the wind power series. Compared to the single VMD decomposition shown in Section 3.2, these eight components are more regular and better capture the original series' features. This demonstrates that incorporating NGO optimization improves VMD's ability to preserve data characteristics and reduce modal aliasing, providing a stronger foundation for accurate prediction.

The prediction results based on the NGO-VMD-ISAOLSTM prediction model and the comparison model are shown in Figure 9, and the corresponding evaluation indexes are shown in Table 3.

Compared to the NGO-VMD-SAO-LSTM model, the NGO-VMD-ISAOLSTM achieved improvements of 13.64%, 18.51%, and 0.5% in RMSE, MAE, and R^2 . Among all the models discussed in Sections 3.1 and 3.2, the proposed model exhibited the lowest prediction error and highest accuracy, further demonstrating its effectiveness and reliability for short-term wind power forecasting.

Table 3. Evaluation indicators for NGO-VMD-SAO-LSTM prediction model and comparison model.

Model	RMSE	MAE	R^2
NGO-VMD-SAO-LSTM	2.023	1.864	0.981
NGO-VMD-ISAOLSTM	1.747	1.519	0.986

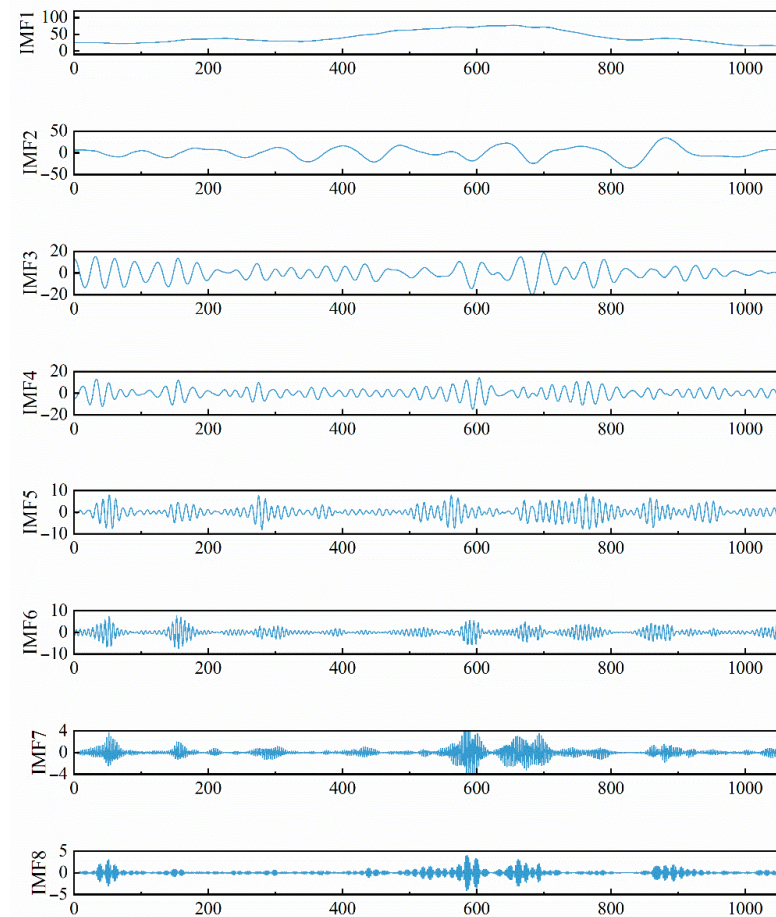


Figure 8. NGO-VMD decomposition results.

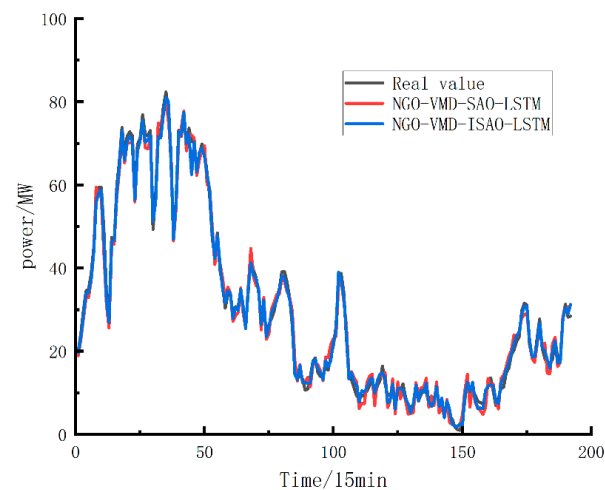


Figure 9. Prediction results of NGO VMD SAO-LSTM prediction model and comparison model.

5. Conclusions

This paper has proposed a combined prediction model integrating NGO-optimized VMD decomposition with ISAO-optimized LSTM, validated using measured wind power data from a Ningxia wind farm. The main conclusions are as follows:

- (1) Decomposition of raw wind power data by the VMD method reduces wind power fluctuation, solves the data feature capture problem, effectively improves the quality of input features in the subsequent prediction model, and thus significantly improves prediction accuracy.

- (2) The introduction of the NGO algorithm optimized the VMD parameters, reduced the influence of noise on signal decomposition, improved the computational efficiency of VMD in processing wind power data, and improved the quality and stability of signal decomposition. At the same time, the NGO algorithm adaptively seeks the optimal combination of modal number and penalty factor $[k, \alpha]$, and the decomposed modal components can accurately capture the periodicity of the data as well as the trend of change, which improves its usability in the subsequent short-term wind power prediction work.
- (3) The ISAO algorithm was applied to optimize LSTM hyperparameters, enhancing both training efficiency and model performance. Compared to single and existing combined models, the proposed approach yielded predictions closer to actual values, making it well-suited to short-term wind power forecasting. Multiple error metrics confirmed its effectiveness in this application.
- (4) The integration of NGO-VMD decomposition and ISAO-LSTM prediction enhances the accuracy and practicality of short-term wind power forecasting, offering valuable theoretical support for optimizing new energy integration into power systems.

Author Contributions: Conceptualization, Z.L.; methodology, Z.L.; software, X.L.; validation, X.L. and H.Z.; data curation, X.L.; writing—original draft preparation, Z.L. and X.L.; writing—review and editing, H.Z.; visualization, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We thank everyone who helped with this paper.

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

References

1. Xue, Y.; Yin, J.; Hou, X. Short-Term Wind Power Prediction Based on Multi-Feature Domain Learning. *Energies* **2024**, *17*, 3313.
2. Nikulins, A.; Sudars, K.; Edelmers, E.; Namatevs, I.; Ozols, K.; Komasilovs, V.; Zacepins, A.; Kviesis, A.; Reinhardt, A. Deep Learning for Wind and Solar Energy Forecasting in Hydrogen Production. *Energies* **2024**, *17*, 1053. [\[CrossRef\]](#)
3. Maduabuchi, C.; Nsude, C.; Eneh, C.; Eke, E.; Okoli, K.; Okpara, E.; Idogho, C.; Waya, B.; Harsito, C. Renewable Energy Potential Estimation Using Climatic-Weather-Forecasting Machine Learning Algorithms. *Energies* **2023**, *16*, 1603.
4. Liu, S.; Zhang, Y.; Du, X.; Xu, T.; Wu, J. Short-Term Power Prediction of Wind Turbine Applying Machine Learning and Digital Filter. *Appl. Sci.* **2023**, *13*, 1751.
5. Huo, X.; Su, H.; Yang, P.; Jia, C.; Liu, Y.; Wang, J.; Zhang, H.; Li, J. Research of Short-Term Wind Power Generation Forecasting Based on mRMR-PSO-LSTM Algorithm. *Electronics* **2024**, *13*, 2469.
6. Yang, M.; Jiang, Y.; Che, J.; Han, Z.; Lv, Q. Short-Term Forecasting of Wind Power Based on Error Traceability and Numerical Weather Prediction Wind Speed Correction. *Electronics* **2024**, *13*, 1559. [\[CrossRef\]](#)
7. Lei, P.; Ma, F.; Zhu, C.; Li, T. LSTM Short-Term Wind Power Prediction Method Based on Data Preprocessing and Variational Modal Decomposition for Soft Sensors. *Sensors* **2024**, *24*, 2521. [\[CrossRef\]](#)
8. Xiang, L.; Liu, J.; Yang, X.; Hu, A.; Su, H. Ultra-short term wind power prediction applying a novel model named SATCN-LSTM. *Energy Convers. Manag.* **2022**, *252*, 115036.
9. Masoumi, M. Machine Learning Solutions for Offshore Wind Farms: A Review of Applications and Impacts. *J. Mar. Sci. Eng.* **2023**, *11*, 1855.
10. Kontopoulou, V.I.; Panagopoulos, A.; Kakkos, I.; Matsopoulos, G. A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet* **2023**, *15*, 255.
11. Liu, X.; Yang, L.; Zhang, Z. Short-Term Multi-Step Ahead Wind Power Predictions Based On A Novel Deep Convolutional Recurrent Network Method. *IEEE Trans. Sustain. Energy* **2021**, *12*, 1820–1833.
12. Zhao, Z.; Yun, S.; Jia, L.; Guo, J.; Meng, Y.; He, N.; Li, X.; Shi, J.; Yang, L. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105982.

13. Ateş, K.T. Estimation of Short-Term Power of Wind Turbines Using Artificial Neural Network (ANN) and Swarm Intelligence. *Sustainability* **2023**, *15*, 13572. [[CrossRef](#)]
14. Wang, C.-H.; Zhao, Q.; Tian, R. Short-Term Wind Power Prediction Based on a Hybrid Markov-Based PSO-BP Neural Network. *Energies* **2023**, *16*, 4282.
15. Wang, Z.; Ying, Y.; Kou, L.; Ke, W.; Wan, J.; Yu, Z.; Liu, H.; Zhang, F. Ultra-Short-Term Offshore Wind Power Prediction Based on PCA-SSA-VMD and BiLSTM. *Sensors* **2024**, *24*, 444.
16. Wang, Z.; Wang, S.; Cheng, Y. Fault Feature Extraction of Parallel-Axis Gearbox Based on IDBO-VMD and t-SNE. *Appl. Sci.* **2023**, *14*, 289.
17. Dou, D.; Jiang, J.; Wang, Y.; Zhang, Y. A rule-based classifier ensemble for fault diagnosis of rotating machinery. *J. Mech. Sci. Technol.* **2018**, *32*, 2509–2515.
18. Chen, J.; Liu, L.; Guo, K.; Liu, S.; He, D. Short-Term Electricity Load Forecasting Based on Improved Data Decomposition and Hybrid Deep-Learning Models. *Appl. Sci.* **2024**, *14*, 5966.
19. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544.
20. Lu, P.; Ye, L.; Zhao, Y.; Dai, B.; Pei, M.; Tang, Y. Review of meta-heuristic algorithms for wind power prediction: Methodologies, applications and challenges. *Appl. Energy* **2021**, *301*, 117446.
21. Liu, X.; Sun, W.; Li, H.; Hussain, Z.; Liu, A. The method of rolling bearing fault diagnosis based on multi-domain supervised learning of convolution neural network. *Energies* **2022**, *15*, 4614. [[CrossRef](#)]
22. Liu, B.; Cai, J.; Peng, Z. Rolling Bearing Fault Diagnosis Method Based on VMD-IMDE-PNN. *Noise Vib. Control* **2022**, *42*, 96–101+133.
23. Dehghani, M.; Hubalovsky, S.; Trojovsky, P. Northern Goshawk Optimization: A New Swarm-Based Algorithm for Solving Optimization Problems. *IEEE Access* **2021**, *9*, 162059–162080.
24. Deng, L.; Liu, S. Snow ablation optimizer: A novel metaheuristic technique for numerical optimization and engineering design. *Expert Syst. Appl.* **2023**, *225*, 120069.
25. Pan, H.; He, X.; Tang, S.; Meng, F. An improved bearing fault diagnosis method using one-dimensional CNN and LSTM. *Stroj. Vestn.-J. Mech. Eng.* **2018**, *64*, 443–452.

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.