modelling

MDPI

Article

A Novel Short-Term Wind Power Forecasting Model Based on Improved Ensemble Learning

He Jiang *, Tianhui Shi, Qingzheng Li and Xinyu Wang

School of Renewable Energy, Shenyang Institute of Engineering, Shenyang 110136, China; 18579026256@163.com (T.S.); umcu1670373@163.com (Q.L.); 15640176099@163.com (X.W.)
* Correspondence: jianghescholar@163.com

Abstract

The development of renewable energy is vital for addressing future climate change and environmental degradation. Nevertheless, the irregular and fluctuating essential features of wind power presents a considerable barrier to grid operational stability. Hence, precise prediction of wind energy output is crucial for improving power system management, boosting the reliability of the supply, and minimizing reserve expenditure. This study presents a predictive model designed for predicting short-term wind speeds using a stacking ensemble approach, which is based on an enhanced Multi-Feature Zebra Optimization Algorithm (IZOA-Stacking). In the data preprocessing phase, to minimize computational costs and prevent overfitting, a module tailored to the various features affecting wind power is developed for the IZOA-Stacking model. Grey relational analysis and Pearson correlation analysis are employed to determine and filter feature correlations. Critically, the preprocessing module demonstrates strong robustness: the One-Class Support Vector Machine (OneSVM) model is applied to identify and replace 100% of anomalous wind speed data, which leads to a substantial and measurable increase in feature correlation and overall model performance. For instance, when retaining wind speed features, the One-Class Support Vector Machine (OneSVM) model is employed to eliminate anomalous wind speed data. During model construction, a stacking ensemble learning strategy integrates multiple prediction models, including Long Short-Term Memory (LSTM) net-works, Extreme Gradient Boosting (XGBoost), ridge regression (RR), and Residual Networks (ResNets). This integration leverages the predictive strengths of each model. Additionally, the improved Zebra Optimization Algorithm (ZOA) optimizes the hyperparameters of each constituent model, further enhancing forecasting accuracy. The findings suggest that the proposed model demonstrates better performance than reference competitor models with regard to predictive accuracy.

Keywords: wind power forecasting model; multi-feature processing and analysis; optimization algorithms; improved ensemble learning

Check for updates

Academic Editor: Wei Gao

Received: 16 March 2026
Revised: 5 May 2026
Accepted: 7 May 2026
Published: 19 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The utilization and expansion of renewable energy resources is recognized as a critical strategy in the fight against future climate change and environmental degradation [1,2]. Across the globe, a wide range of eco-friendly power options exist [3,4]. Among these, wind-based generation, recognized for minimal emissions and ample availability, has drawn heightened interest. In consideration of the inherent variability and intermittent nature of wind energy [5], ensuring the stability of its grid integration poses a series of challenges.

Consequently, the implementation of accurate wind power forecasting is paramount for optimizing energy management within power systems, enhancing power supply reliability, and reducing the costs associated with system reserve capacities [6].

The prediction of wind power is typically categorized into three primary groups: physical modelling techniques, statistical methods, and algorithms based on artificial intelligence [7,8]. Physical modelling approaches are more suitable for medium to long-term forecasting, as they take into account such parameters as wind field topography and atmospheric dynamics. Nevertheless, their application is generally constrained by uncertainties related to model inputs, parameter estimation, and structural configurations [9,10]. Conversely, statistical methods rely on historical data to predict wind power through the development of time series models or regression models [11]. Conventional statistical approaches include methodologies like the Autoregressive Integrated Moving Average (ARIMA) model, in addition to strategies including Support Vector Regression (SVR) [12,13]. These methods offer rapid computational speeds and excel in forecasting linear data patterns but exhibit drawbacks in dealing with nonlinear and sophisticated relationships. In comparison to physical and statistical methods, artificial intelligence approaches such as recurrent neural networks (RNNs) [14] and LSTM [15] leverage big data analytics to demonstrate superior accuracy and broader applicability.

Considering that single predictive models in artificial intelligence methods have limited capability in handling some extremely complex sequences in wind power forecasting, techniques such as data preprocessing, algorithm optimization, and ensemble modelling have gradually been adopted by numerous studies [16–18]. In the preprocessing phase of wind power feature analysis, reference [19] applies the grey relational analysis technique for the purpose of assessing the effects of diverse wind power features on wind power generation, pinpointing the key determinants influencing the output. Reference [20] employs the Pearson correlation coefficient to mine important features within the dataset, effectively eliminating features with minimal impact on wind power output. However, due to the inherent preference biases of different correlation analysis methods' evaluation metrics, their influence on model prediction results necessitates further research.

Regarding wind speed data preprocessing, reference [21] corrects weather forecast wind speeds through curve fitting and performs wind speed screening based on meteorological types, thereby excluding anomalous wind speed values. Nevertheless, curve fitting exhibits a strong dependence on data and is susceptible to the influence of outliers when data quality is poor. As demonstrated in reference [22], the application of a Kalman filter to the smoothing of multi-scale signals in decomposed wind speed time series has been shown to be an effective method of reducing noise and outliers in wind speed data. However, this approach demonstrates poor real-time performance when handling large-scale data and is highly sensitive to model parameter selection, which may result in suboptimal filtering outcomes. Reference [23] constructs a boundary using a OneSVM to describe the feature space of normal data and identifies anomalous data by detecting samples that significantly deviate from this boundary. As an unsupervised learning method, this approach effectively handles high-dimensional and nonlinear data while maintaining robust performance.

In optimizing predictive model parameters, reference [24] proposes the implementation of a Grey Wolf Optimization (GWO) algorithm for the purpose of optimizing the hyperparameters of a Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) network. The GWO algorithm identifies an optimal combination of hyperparameters for the LSTM-CNN model. Reference [25] enhances the parameters of the Support Vector Machine (SVM) by utilizing Particle Swarm Optimization (PSO), aiming to increase the predictive accuracy of wind power forecasting. Nevertheless, both algorithms exhibit a tendency to get stuck in local optima, which undermines their ability to effectively perform

global search, necessitating additional exploration and refinement. More generally, many existing swarm intelligence and metaheuristic optimization algorithms suffer from several common drawbacks in high-dimensional, multimodal hyperparameter spaces: they may prematurely converge due to insufficient population diversity, struggle to balance global exploration and local exploitation, and display high sensitivity to initialization and control parameters. These limitations significantly constrain their ability to achieve stable and reliable global optimization in complex wind power forecasting tasks. The ZOA, introduced by Trojovská E et al. in 2022 [26], is a novel optimization method that mimics zebra behaviour. Compared to other optimization algorithms, it boasts strong optimization capabilities and rapid convergence speed. Owing to its specific search mechanism, the ZOA can better maintain population diversity and more effectively trade off exploration and exploitation, thus reducing the risk of trapping in local optima and improving global search performance compared with traditional algorithms such as PSO and GWO in complex optimization scenarios. Reference [27] implements an augmented iteration of the ZOA for hyperparameters of a LSTM network, demonstrating that the enhanced ZOA achieves faster convergence speeds than other algorithms. Elite Opposition-Based Learning (EOBL) is a method designed to improve the global exploration efficiency and the rate of convergence of optimization algorithms by producing inverse solutions, thus helping to by-pass local optima.

Reference [28] improves the GWO algorithm using EOBL; this leads to an improvement in the predictive performance of the optimized LSTM network when compared to the baseline model. These studies indicate that combining the ZOA with EOBL has the potential to further enhance convergence speed and global optimization capability, making it a promising candidate for hyperparameter optimization in advanced wind power forecasting models.

In studies focusing on single predictive models, reference [29] employs XGBoost in order to predict wind power. XGBoost, representing a highly efficient and powerful distributed gradient boosting algorithm, demonstrates a computation speed ten times faster than traditional gradient boosting algorithms. However, the XGBoost model is sensitive to anomalous wind speed values in wind power features; excessive anomalies in wind speed data may degrade model performance. Reference [30] utilizes a LSTM network for the estimation of wind power. LSTM networks, renowned for their deep learning capabilities in capturing historical information within time sequence data, are widely employed in a variety of sequential prediction tasks. However, within the scope of wind power forecasting, LSTM networks are unable to handle multi-scale temporal features of wind speed fluctuations. Reference [31] selects a Residual Neural Network (ResNet) for wind power prediction. Results indicate that the ResNet exhibits strong feature extraction capabilities in processing two-dimensional features and time series data due to its unique convolutional architecture. However, compared to other neural networks, the ResNet tends to overlook the correlations between local and global features because convolutional layers typically handle only a limited range of time steps. Reference [32] employs a RR model for wind power forecasting. RR is a technique specifically designed to address multicollinearity in data analysis by incorporating an L2 regularization term into the loss function, thereby preventing model weights from becoming excessively large. This effectively controls model complexity and mitigates overfitting. However, the highly nonlinear relationships between various wind power features often lead to a diminished predictive capacity of the RR model. These differences suggest that the above models are not redundant, but rather complementary. Specifically, LSTM is suitable for learning long-term temporal dependence and sequential evolution patterns in wind power series; ResNet is effective at extracting local, hierarchical, and fluctuation-related features from input data through convolutional residual learning; RR provides a stable and interpretable

linear baseline that is useful for capturing approximately linear relationships and mitigating multicollinearity among meteorological variables; and XGBoost is powerful in modelling nonlinear feature interactions and complex decision boundaries while maintaining high computational efficiency. Therefore, combining these highly heterogeneous base learners can enable the forecasting framework to simultaneously capture temporal dependence, local structural patterns, linear trends, and nonlinear interactions, thereby improving both prediction accuracy and model robustness.

In the domain of hybrid predictive model ensembles, reference [33] introduces a combined approach utilizing a ResNet alongside LSTM, achieving an average accuracy improvement of 10.97% compared to single ResNet and LSTM predictive models. Nonetheless, its accuracy remains inferior to other ensemble models, indicating the need for more comprehensive ensemble strategies. Reference [34] employs a stacking framework for multi-model fusion in load forecasting. The prediction results demonstrate that the stacking framework outperforms both single models and ensemble models without a stacking framework in terms of accuracy. However, its predictive accuracy is still lower compared to multi-model fusion frameworks optimized with algorithmic enhancements. Therefore, it is necessary to design stacking frameworks that incorporate strongly heterogeneous base learners and are equipped with more powerful optimization algorithms, so as to fully exploit model complementarity and further improve forecasting performance.

Taking into consideration the merits and demerits of the research methods previously discussed, this research puts forward a short-term wind power forecasting model which incorporates an improved multi-feature ZOA within a stacking ensemble structure. The model encompasses multi-feature data processing, anomaly detection and filtering, and an optimized parameter algorithm within a stacked deep learning architecture. The research objectives include:

1. Feature Analysis and Selection: To mitigate the escalation of computational costs and the risk of overfitting, grey relational analysis and Pearson correlation coefficient analysis are employed to evaluate each feature influencing wind power output. Considering that different correlation analysis methods may introduce biases in evaluation metrics affecting prediction accuracy, the scores from the two analysis methods for each feature are averaged. Subsequently, the degree of association for each feature is determined, and feature selection is conducted accordingly. If wind speed features are retained, the OneSVM model is utilized to screen and eliminate anomalous wind speed data. Compared to existing data preprocessing methods, this approach not only quantitatively assesses various wind power features but also effectively handles further wind speed anomalies.
2. Ensemble Model Construction: Given that single models and certain composite models still exhibit relatively poor prediction accuracy, it is necessary to consider more comprehensive ensemble models and combination strategies. Employing the stacking ensemble learning strategy, multiple forecasting models, including LSTM, XGBoost, RR, and ResNet are integrated to forecast wind power generation. These base learners are intentionally selected for their heterogeneity and complementarity: LSTM captures long-term temporal dependencies, ResNet extracts local and hierarchical fluctuation features, RR models linear relationships under multicollinearity, and XGBoost learns nonlinear interactions among multiple features. By combining these models within a stacking framework, the proposed method aims to exploit their complementary strengths and improve forecasting accuracy and robustness.
3. Hyperparameter Optimization: Due to the superior optimization capabilities and faster convergence speed of the ZOA compared to other optimization algorithms, it is adopted for optimizing the hyperparameters of the presented model. In this study, an

improved ZOA incorporating Elite Opposition-Based Learning (EOBL) is employed to further enhance global exploration ability and avoid premature convergence. The enhanced ZOA is a sophisticated machine learning technique that utilizes elite reverse learning to maximize the performance of each model within the stacking ensemble architecture. This optimization aims to determine the ideal set of parameters that most effectively align with the predictive models, thus enhancing the precision of the forecasting approach.

The structure of this document is organized in the subsequent manner: Section 2 describes the development of the preliminary processing element within the IZOA-Stacking wind turbine power forecasting model. Section 3 describes the IZOA optimization algorithm, the stacking composite learning strategy, and the conceptual framework of the IZOA-Stacking model. Section 4 introduces the model evaluation metrics and provides a relative analysis of the forecasting results from multiple models, thereby validating the feasibility of the proposed model. Finally, Section 5 concludes this study.

2. Data Preprocessing and Filtering

The initial wind speed data collected from wind farms contain some abnormal values with short-term drastic fluctuations, which mask the true features of wind speed data. Moreover, the different characteristics of wind power generation in each period have varying impacts on power generation, which results in challenges when directly applying the raw data to forecasting wind power, making it hard to attain optimal prediction accuracy. The wind power prediction model, referred to as Improved Zebra Optimization Algorithm-Stacking (IZOA-Stacking), is outlined in this study. Its integration of a preprocessing module for initial data handling and preparation is intended to enhance the precision of wind power forecasts and reduce its inherent unpredictability. In addition, it is intended to evaluate the impact of various wind power features. The development procedure for the preprocessing module of the IZOA-Stacking wind power forecasting model is outlined as follows:

1. Analysis and Processing of Different Characteristics of Wind Power

A correlation analysis is conducted on the various characteristics that influence wind power to determine the impact of each characteristic on wind power. After determining the correlation of each characteristic, they are screened, retaining some characteristics that are highly correlated with wind power. The aforementioned characteristics, which have been retained, are then subjected to a process of normalization and weighting, with a view to utilizing them as input features for the forecasting model.

2. Wind Speed Anomaly Data Processing

If the wind speed characteristic is successfully retained after the first screening step, the second step is carried out. This involves screening the wind speed anomaly data and removing abnormal data to obtain data that can reflect the true properties of wind speed. Then, the missing wind speed data is filled with normal data to obtain regular and continuous wind speed data.

2.1. Analysis and Processing of Different Characteristics of Wind Power

Grey relational analysis (GRA) evaluates the closeness of relationships based on the geometric similarity between sequences; i.e., the proximity of sequence curves determines the degree of correlation. This analysis method effectively overcomes the limitations that traditional mathematical statistics analysis may encounter in system evaluation. It is insensitive to sample size and regularity, has a simple calculation process, and can reduce inconsistencies between quantitative calculation and qualitative evaluation.

Spearman's Rank Correlation Coefficient Analysis, due to its intuitiveness and high reliability, is widely used in data analysis and modelling in various fields. This method is a common means of assessing the linear association between two variables, and it is particularly effective when the variables follow a normal or approximately normal distribution.

Before implementing GRA and Spearman's Rank Correlation Coefficient Analysis, quantitative analysis is needed to preprocess these wind power characteristics with dimensionless processing to allow for more accurate correlation analysis. The main process of correlation analysis for influencing wind power characteristics is as follows:

1. Obtain a 60-day dataset of wind power $P_t = [P_{t-5760}, P_{t-5759}, \ldots, P_{t-1}]^{\mathrm{T}}$ from a specific wind power plant, featuring a 15 min time resolution, sourced from the Xihe Energy Meteorological Big Data Platform. This dataset serves as the reference sequence. For comparison, the other influencing factors are regarded as comparison sequences $X_\delta = \{x_\delta(t-5760), x_\delta(t-5759), \ldots, x_\delta(t-1)\}, \delta \in \{1, 2, \ldots, D\}$ where D represents the number of features.
2. Both the reference sequence and the comparison sequence are normalized using the Z-score standardization method, and the mathematical formula is:

$$\begin{cases} \mu_\delta = \dfrac{1}{N}\sum_{k=1}^{N} x_\delta(k) \\ \mu_0 = \dfrac{1}{N}\sum_{k=1}^{N} p_t(k) \\ \sigma_\delta = \sqrt{\dfrac{1}{N}\sum_{k=1}^{N}(x_\delta(k)-\mu_l)^2} \\ \sigma_0 = \sqrt{\dfrac{1}{N}\sum_{k=1}^{N}(p_t(k)-\mu_0)^2} \\ r_\delta(k) = \dfrac{x_\delta(k)-\mu_\delta}{\sigma_\delta} \\ r_0(k) = \dfrac{p_t(k)-\mu_0}{\sigma_0} \end{cases} \tag{1}$$

In this context, N denotes the complete set of data points. The variables $x_\delta(k)$ and $p_t(k)$ denote the values of the δ-th feature and wind power generation, respectively, at the k-th data point. The mean values of the δ-th feature and wind power generation across all data points are represented by μ_δ and μ_0, respectively, while σ_δ and σ_0 denote their respective standard deviations. Finally, $r_\delta(k)$ and $r_0(k)$ represent the standardized values of the δ-th data point for the δ-th feature and wind power, respectively.

3. Conduct grey relational analysis on each feature that affects wind power. The mathematical formula is:

$$\begin{cases} \varepsilon_\delta^1(k) = \dfrac{\min\limits_{\delta}\min\limits_{k}|r_0(k)-r_\delta(k)| + \phi\max\limits_{\delta}\max\limits_{k}|r_0(k)-r_\delta(k)|}{|r_0(k)-r_\delta(k)| + \phi\max\limits_{\delta}\max\limits_{k}|r_0(k)-r_\delta(k)|} \\ \varepsilon_\delta^1 = \dfrac{1}{N}\sum\limits_{k=1}^{N}\varepsilon_\delta^1(k) \end{cases} \tag{2}$$

where $\varepsilon_\delta^1(k)$ represents the grey relational coefficient of the k-th data point for the δ-th feature, ϕ denotes the distinguishing coefficient within the range of [0, 1], and ε_δ^1 represents the grey relational grade of the δ-th feature.

https://doi.org/10.3390/modelling7030098

4. Pearson correlation coefficient analysis is performed on each feature that influences wind power, and the absolute values are taken. The mathematical formula is as follows:

$$\varepsilon_{\delta}^{2} = \left| \frac{\sum_{k=1}^{N} (r_{\delta}(k) - \bar{r}_{\delta})(r_0(k) - \bar{r}_0)}{\sqrt{\sum_{k=1}^{N} (r_{\delta}(k) - \bar{r}_{\delta})^2 (r_0(k) - \bar{r}_0)^2}} \right| \tag{3}$$

where ε_{δ}^{2} represents the Pearson correlation coefficient of the δ characteristic, and $\bar{r}_{\delta}$ and $\bar{r}_0$ denote the average values of the δ feature and wind power output, in turn.

5. After determining the relevance of each feature, the relevance is first normalized. Then, the correlation coefficients of each feature from the two correlation analysis methods are averaged. Finally, the correlation of each feature is analyzed. The mathematical formula is as follows:

$$\begin{cases} \varphi_{\delta}^{(1,2)} = \dfrac{\varepsilon_{\delta}^{(1,2)}}{\sum_{\delta=1}^{6} \varepsilon_{\delta}^{(1,2)}} \\ \varphi_{\delta} = \dfrac{\varphi_{\delta}^{1} + \varphi_{\delta}^{2}}{2} \end{cases} \tag{4}$$

where $\varphi_{\delta}^{(1,2)}$ represents the normalized value of the grey relational analysis and Pearson interrelation coefficient of the δ-th feature, and φ_{δ} is the mean-processed interrelation coefficient of the δ-th feature. An excessive number of parameters amplifies model complexity, driving up computational costs and making it more prone to overfitting, while an inadequate selection of features may result in the loss of crucial details, thereby affecting prediction performance; two features, exhibiting minimal correlation with wind power, were excluded to derive $X_{\delta'}$ and attain the intended predictive performance of the model, where $\delta' \in \{1, 2, 3, 4\}$.

2.2. Wind Speed Outlier Processing

If the wind speed feature is successfully retained after feature selection, wind speed outlier processing is performed. Outliers typically refer to values that are significantly different from other data points. These data may be caused by equipment errors, input errors, or very rare weather conditions. If these outliers are not processed during model training, they may distort the model's learning process, thereby reducing the model's prediction accuracy under normal conditions. OneSVM is an algorithm that utilizes SVM technology for anomaly detection, specifically designed for training with only normal sample data. The algorithm identifies abnormal samples by recognizing data points that deviate significantly from the decision boundary. As an unsupervised learning algorithm, OneSVM exhibits efficiency and robustness in processing high-dimensional datasets with intricate nonlinear features. The process of employing the OneSVM algorithm to screen wind speed outliers is as follows:

1. Obtain the 60-day historical wind speed data $V_t = [v_{t-5760}, v_{t-5759}, \ldots, v_{t-1}]^{\mathrm{T}}$ with a time scale of 15 min from the XiHe Energy Meteorological Big Data Platform wind farm. Utilize the Radial Basis Function (RBF) kernel in calculating the similarity between each pair of data points. Mapping the dataset into a higher-dimensional feature space allows OneSVM to effectively address nonlinear challenges. The mathematical formula is as follows:

$$K(v_i, v_j) = \exp\left(-\beta \left\| v_i - v_j \right\|^2\right) \tag{5}$$

where v_i and v_j are the wind speeds at the i, j time point, respectively, where $i, j \in [t-5760, t-5759, \ldots, t-1]$ means β represents the width parameter of the RBF kernel; $K(v_i, v_j)$ represents the similarity between v_i and v_j.

2. The OneSVM objective function is employed to calculate and find the Lagrange multiplier α value corresponding to each time point. The formula for calculation is provided as follows:

$$\begin{cases} \max\alpha = \left[\sum_{i=1}^{5760} \alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{5760} \alpha_i \alpha_j K(v_i, v_j)\right] \\ \text{s.t.}\sum_{i=1}^{5760} \alpha_i = 1 \\ 0 \le \alpha_i \le C \end{cases} \tag{6}$$

where α is a column vector of the optimal combination of Lagrange multipliers for each time point; α_i and α_j are the Lagrange multipliers of any two data points in the 60-day historical wind speed data; and, denoted as C, a positive constant is employed to regulate the model's flexibility.

3. The decision function is employed for decision screening. If the decision function at a certain time point is greater than 0, it is a normal data point; if it is less than 0, it is an outlier. The calculation formula is as follows:

$$f(z_i) = \text{sign}\left(\sum_{j=1}^{5760} \alpha_j K(v_i, v_j) - \chi\right) \tag{7}$$

where $f(z_i)$ is the decision function, which serves to assess if the wind speed value at the i data point constitutes an anomaly; the function $\text{sign}(\cdot)$ represents the signum function, which outputs either 1 or -1, based on the polarity of the input value; and χ is the offset of the decision boundary.

After removing the wind speed outlier data points, they are replaced and updated with their adjacent normal values. The update formula is:

$$v_i' = \begin{cases} v_i & v_i \notin V_{\text{problem}} \\ \frac{v_{i-1}+v_{i+1}}{2} & v_i \in V_{\text{problem}}, v_{i-1} \notin V_{\text{problem}}, v_{i+1} \notin V_{\text{problem}} \\ v_{i-1} & v_i \in V_{\text{problem}}, v_{i-1} \notin V_{\text{problem}}, v_{i+1} \in V_{\text{problem}} \\ v_{i+1} & v_i \in V_{\text{problem}}, v_{i-1} \in V_{\text{problem}}, v_{i+1} \notin V_{\text{problem}} \end{cases} \tag{8}$$

where v_i' is the updated and replaced wind speed data at time point i. If v_i is an outlier and $v_{i-1} \in V_{\text{ab}}$, $v_{i+1} \in V_{\text{ab}}$, then the wind speed value v_i' at the i time point after updating selects a farther normal data point; v_{i-1}, v_i, v_{i+1} are the wind speed data before updating at time points $i-1$, i, and $i+1$, respectively; $V_{problem}$ is the set of outlier wind speed data. The updated wind speed model dataset is then denoted as $V_t' = \left[v_{t-672}', v_{t-671}', \ldots, v_{t-1}'\right]^{\text{T}}$.

3. IZOA-Stacking Wind Power Prediction Model

3.1. IZOA Optimization Algorithm

The Improved Zebra Optimization Algorithm (IZOA) is an innovative swarm intelligence-based optimization method. By utilizing the IZOA, the parameters of each base learner in stacking can be optimized to achieve improved predictive performance. Unlike alternative algorithms, the IZOA method demonstrates a faster convergence rate, improved global search efficiency, and greater robustness in terms of adaptability. The IZOA mainly comprises the subsequent stages:

Step 1: Randomly perform initialization and establish a zebra population of zebra individuals, each with dimensional position information.

Step 2: Improve the ZOA through EOBL, introducing a dynamic opposite point:

$$W'_\mu = \left(W_\mu^{1'}, W_\mu^{2'}, \ldots W_\mu^{d'}\right) \tag{9}$$

$$W_\mu^{j\prime} = \nu_1\left(W_{\mu,\min}^j + W_{\mu,\max}^j\right) + (1-\nu_2)\left(W_{\mu,\min}^j + W_{\mu,\max}^j\right) - W_\mu^j \tag{10}$$

where ν_1, ν_2 are random numbers between (0, 0.5), conforming to a uniform distribution; $W_{\mu,\min}^j$ and $W_{\mu,\max}^j$ represent the min and max values, respectively, of the j-th dimensional positional data of the μ-th zebra individual; W_μ^j refers to the updated j-th dimensional position of the μ-th zebra individual after the application of the dynamic reverse point technique; and W_μ^j denotes the j-th dimensional position of the μ-th zebra individual prior to the dynamic reverse point's introduction, $\mu \in [1, a]$, $j \in [1, d]$.

Obtain a reverse zebra population with a zebra individuals, each having d-dimensional position information, through the dynamic reverse point of the elite reverse learning method.

Step 3: Compare the fitness values associated with the positional data of each individual zebra in the initial population and its dynamic reverse zebra population. Select a zebra individuals with better fitness from high low to form the final initialized population. Then, the zebra population W improved by the elite reverse learning method is expressed as:

$$W = \begin{bmatrix} W_1^1 & W_1^2 & \ldots & W_1^d \\ W_2^1 & W_2^2 & \ldots & W_2^d \\ \ldots & \ldots & \ldots & \ldots \\ W_a^1 & W_a^2 & \ldots & W_a^d \end{bmatrix} \tag{11}$$

where d represents the maximum position dimension of each individual zebra, referring to the parameters of each model to be optimized in the stacking model, namely the parameters of XGBoost, LSTM, ResNet, and RR.

The fitness calculation formula is:

$$f_{fit}^1(W_\mu) = \left(1 - \frac{\sum_{t=1}^{\overline{m}} \left|P_t - \overline{P}_t\right|}{\overline{m}}\right) \times 100\% \tag{12}$$

where $f_{fit}^1(W_\mu)$ represents the fitness of the μ-th zebra; the quantity $\overline{m}$ represents the total count of wind power prediction samples; $\overline{P}_t$ is the true power value at time t.

Step 4: If the number of iterations $\tau \leq 0.75 iter_{\max}$, then perform the zebra foraging behaviour, i.e., the first stage, by imitating the behaviour of zebras searching for food to refresh the population composition. Zebras' primary diet includes grasses and sedges; in times of food scarcity, the zebra may also consume other plant material, including shoots, fruit, bark, roots, and leaves. Foraging time for zebras, depending on the quality and availability of plants, can occupy 60 to 80% of their daily time. Plains zebras, in particular, tend to graze the taller, less nutritious grasses first, creating opportunities for other species seeking shorter, more nutrient-rich forage. In the IZOA model, the best-performing members of the population are considered the leading zebras, leading fellow members to the designated area within the search space. Hence, the adjustment of position during the zebra foraging process can be accomplished using the subsequent mathematical model:

$$\begin{cases} \overline{W}_\mu^{j,(\tau+1)} = W_\mu^{j,(\tau)} + r \cdot \left(W_{\mu,\max}^{j,(\tau)} - I \cdot W_\mu^{j,(\tau)}\right) \\ W_\mu^{j,(\tau+1)} = \begin{cases} \overline{W}_\mu^{j,(\tau+1)}, F_\mu^{(\tau)} < F_\mu^{(\tau+1)} \\ W_\mu^{j,(\tau)}, \quad else \end{cases} \end{cases} \tag{13}$$

where $\overline{W}_{\mu}^{j,(\tau+1)}$ denotes the provisional new location of the μ-th zebra in the j-th dimension at the τ-th iteration; r is a random number between 0 and 1; $W_{\mu}^{j,(\tau)}$ is the original position of the μ-th zebra in the j-th dimension at the τ-th iteration; $W_{\mu,\max}^{j,(\tau)}$ is the position of the fittest zebra (i.e., the pioneer zebra) in the j-th dimension at the τ-th iteration; and $W_{\mu}^{j,(\tau+1)}$ is the new position of the μ-th zebra in the j-th dimension at the $(\tau+1)$-th iteration. If the fitness $F_{\mu}^{(\tau+1)}$ of the new position of the μ-th zebra is better than the fitness $F_{\mu}^{(\tau)}$ of the current position, the zebra moves to the new position; otherwise, it remains at the current position. The Root Mean Square Error (RMSE) is used to evaluate the fitness level.

Step 5: When the number of iterations $\tau > 0.75 iter_{\max}$ reaches 75, the zebras engage in defensive behaviour against predators, marking the second stage. At this stage, the escape behaviour of zebras in reaction to predatory threats is employed to alter the locations of zebra individuals within the search domain of the IZOA throughout the specified region. When confronted with lion attacks, zebras employ an erratic zigzag route characterized by unpredictable lateral shifts. Nevertheless, when facing smaller predators such as hyenas and wild dogs, zebras exhibit increased aggression in their behaviour, utilizing collective strength to confuse and intimidate the predators. In the IZOA settings, the following two scenarios are assumed to occur with equal probability:

1. When facing lion attacks, zebras adopt an escape strategy;
2. When facing attacks from other predators, zebras adopt a counterattack strategy. In scenario 1, zebras under attack by lions will quickly escape near their current location. Mathematically, this escape strategy can be expressed by Equation (14) S1 below. In scenario 2, when threatened by predators, the zebra group will approach the individual under threat, trying to establish a defensive arrangement to confuse and ward off the attacker. This collective defence strategy can be mathematically expressed by Equation (14) S2 below. During the update of the zebra's position, if the new position results in an enhanced objective function value, then the adoption of this updated position is recommended. This update mechanism is modelled by the following equation:

$$\begin{cases} \overline{W}_{\mu}^{j,(\tau+1)} = \begin{cases} S1: W_{\mu}^{j,(\tau)} + R\cdot(2r-1)\cdot\left(1-\frac{\tau}{iter_{\max}}\right)\cdot W_{\mu}^{j,(\tau)}, P_s \le 0.5 \\ S2: W_{\mu}^{j,(\tau)} + r\cdot\left(W_{\mu_{at}}^{j,(\tau)} - I\cdot W_{\mu}^{j,(\tau)}\right), else \end{cases} \\ W_{\mu}^{j,(\tau+1)} = \begin{cases} \overline{W}_{\mu}^{j,(\tau+1)}, F_{\mu}^{(\tau)} < F_{\mu}^{(\tau+1)} \\ W_{\mu}^{j,(\tau)}, \quad else \end{cases} \end{cases} \tag{14}$$

where R and P_s are random numbers between 0 and 1; $W_{\mu_{at}}^{j,(\tau)}$ represents the position of the attacked zebra in the j-th dimension at the τ-th iteration; and $\left(1-\frac{\tau}{iter_{\max}}\right)$ simulates the increasing fatigue level of the zebra over time.

Step 6: Update the zebra position according to the fourth and fifth steps. If the iteration exceeds the maximum iteration number, the training ends and outputs the position information of all global optimal zebra dimensions, that is, the parameters of XGBoost, LSTM, ResNet, and RR. Otherwise, jump to the fourth step to continue parameter optimization.

It is important to emphasize that the IZOA optimizes hyperparameters using only the training set (1 April–20 May) with 5-fold cross-validation. The test set (21 May–30 May) remains completely unseen during the entire tuning and training process, and is used only once for the final evaluation of the optimized stacking model.

3.2. Stacking Ensemble Learning Strategy

The stacking ensemble learning strategy, often referred to as stacked generalization, is a method for combining multiple models in a hierarchical manner. Its workflow is summarized as follows:

1. Define the dataset D serving as the input for the stacking ensemble learning approach. With a time scale of 15 min, the data point samples comprise 5760 samples. Each sample consists of a feature vector $X_{\delta'}$ and its corresponding wind power, forming the dataset $D = \left\{\left(x_1^{(t-5760)}, x_2^{(t-5760)}, x_3^{(t-5760)}, x_4^{(t-5760)}, P_{t-5760}\right), \left(x_1^{(t-5759)}, x_2^{(t-5759)}, x_3^{(t-5759)}, x_4^{(t-5759)}, P_{t-5759}\right), \ldots, \left(x_1^{(t-1)}, x_2^{(t-1)}, x_3^{(t-1)}, x_4^{(t-1)}, P_{t-1}\right)\}$.
2. Perform a split on the dataset to obtain training and test datasets.
3. Develop multiple beginner models for the first layer and apply 5-fold validation on the training dataset, constituting a fundamental aspect of the methodology. The procedure involves the division of the training dataset into five segments of uniform size. For each base model, one part is used for validation, while the remaining four parts are used for training. For each beginner learner, record its prediction results on the validation subset and test set, thereby generating a corresponding set of validation subset and test set prediction results for each type of beginner learner.
4. Establish a second layer learner, namely the secondary learner. This layer learner uses all the validation subset prediction results of each beginner learner in the first layer as training data, and the averaged test set prediction results as new test data. The meta-learner is trained on this basis to form the ultimate prediction model.

By employing this approach, the results obtained from each phase of learning serve as the input for the subsequent phase, ultimately improving the model's generalization capabilities.

To mitigate overfitting given the limited 60-day dataset, the LSTM architecture employs dropout (0.3) and recurrent dropout (0.3), together with early stopping (patience = 10). The ResNet model includes dropout layers (0.3) after each residual block and uses batch normalization. These regularization strategies, combined with the stacking ensemble and deterministic cross-validation, ensure stable generalization.

3.3. IZOA-Stacking Ensemble Algorithm Framework Design

In constructing the IZOA-Stacking ensemble learning framework, the foremost task is to select base learners that are both high-quality and diverse. XGBoost trees represent a class of extremely efficient and predictive distributed gradient boosting algorithms, boasting computational speeds up to ten times faster than traditional gradient boosting methods. However, the XGBoost model exhibits sensitivity to outliers in wind speed features within wind power data; a prevalence of such outliers in wind speed measurements can bring about a degradation in the performance of the model. LSTMs leveraging their deep learning capabilities to capture historical information in time series data are extensively utilized in various time series forecasting tasks. However, in wind power forecasting, LSTM models struggle to effectively manage the multi-scale temporal characteristics inherent in wind speed fluctuations. ResNets, with their distinctive convolutional architecture, demonstrate robust feature extraction capabilities when processing spatial data types like images and time series, and thus are regarded as a significant breakthrough in the field of deep learning.

However, compared to other neural network architectures, the ResNet tends to overlook the correlations between local and global aspects in the processing of various wind power features due to convolutional layers typically managing only a limited range of temporal steps. RR is a technique specifically designed to address multicollinearity in data analysis by incorporating an L2 regularization term into the loss function, thereby preventing model weights from becoming excessively large. This effectively controls model complexity and mitigates overfitting. Nevertheless, the typically complex and nonlinear as-

sociations between wind power and its characteristics lead to erratic prediction accuracy in the RR model. Figure 1 depicts the structural diagrams of each model within the ensemble learning framework with stacking strategy.

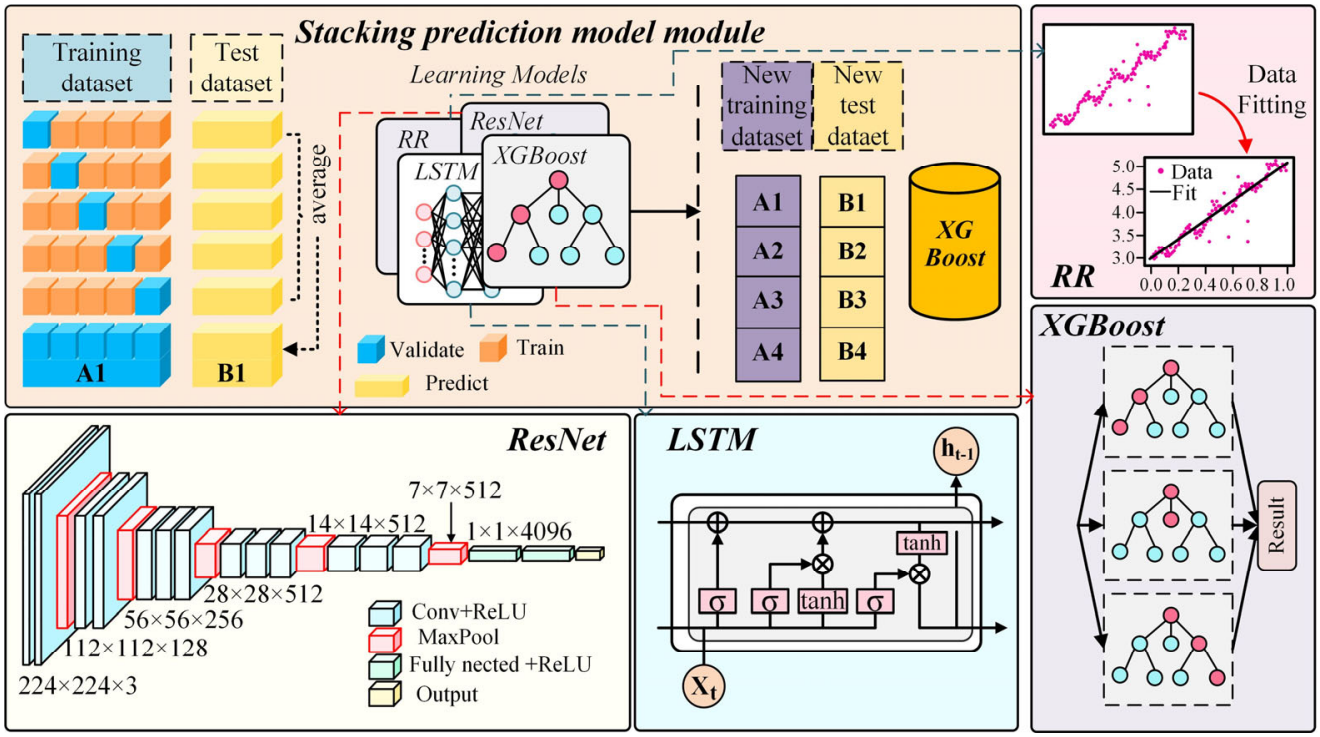


Figure 1. Architectural schematics of constituent models within the stacking ensemble learning framework.

Based on the above considerations, the selection of base learners in this study follows a task-oriented principle, aiming to match the characteristics of wind power data with complementary model families rather than searching for a single universally best algorithm. From a methodological perspective, the proposed framework does not introduce a completely new learning algorithm; instead, it emphasizes a task-oriented design of the ensemble components. The four base learners are chosen to reflect a variety of model families, including tree-based boosting, recurrent neural networks, convolutional residual networks, and linear regularized regression, allowing the stacking ensemble to capture nonlinear interactions, temporal dynamics, and approximate linear trends concurrently. This diversity is important for wind power forecasting, where meteorological variables, turbine characteristics and operational constraints interact at multiple scales. By embedding IZOA-based hyperparameter optimization and the tailored preprocessing module into this heterogeneous ensemble, the proposed framework aims to provide a practically useful and reproducible recipe for short-term wind power prediction rather than a purely algorithmic novelty.

In the IZOA-Stacking ensemble model strategy adopted in this study, the primary learners consist of XGBoost, LSTM, ResNet, and RR, while the secondary learner is chosen as XGBoost. The model's parameters are optimized through the use of the Zebra Optimization Algorithm, and incorporating the 5-fold KFOLD cross-validation approach ensures the generalization ability and the model's prediction accuracy. The presented IZOA-Stacking wind power forecasting model is illustrated in Figure 2.

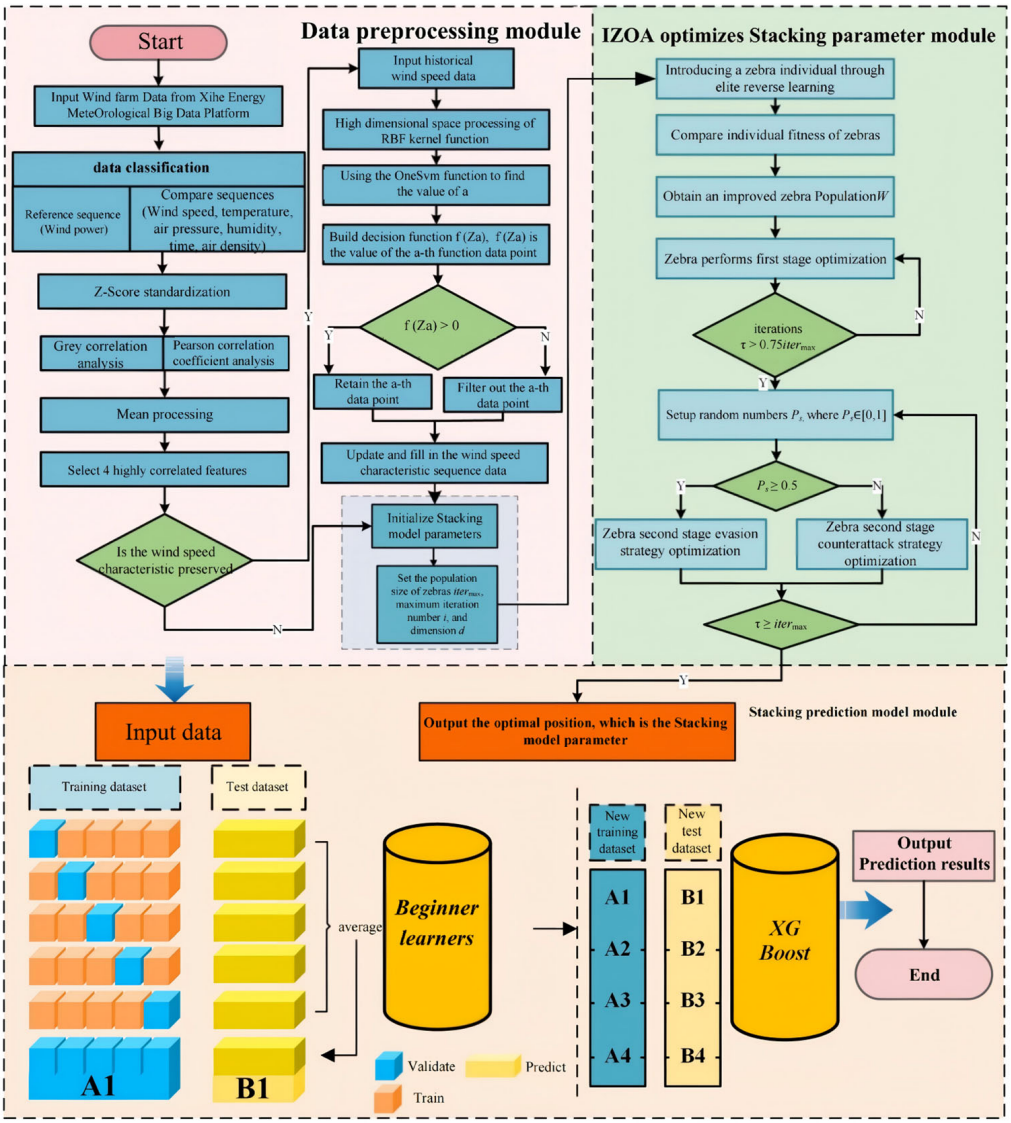


Figure 2. Flowchart of the IZOA-Stacking wind power forecasting model.

4. Case Analysis

This study selected wind farms from the Xihe Energy Meteorological Big Data Platform as case study subjects, utilizing wind power data from 1 April 2024 to 30 May 2024, for analysis. Each wind farm possessed a total capacity of 50 MW, with data recorded every 15 min, resulting in 96 data points per day. This study aimed to conduct short-term predictions of wind power output by leveraging data from the preceding three time intervals to estimate the generation for the upcoming period. Data samples from 1 April 2024 to 20 May 2024 were designated as the training set, while samples from 21 May 2024 to 30 May 2024 were allocated to the test set. The experimental platform utilized in this study was PyCharm 2021.

4.1. Model Evaluation Metrics

Cost of Purchasing Electricity

To assess the predictive accuracy, model fit, and reliability of the proposed wind power prediction model, the following assessment criteria are used: the RMSE is utilized to assess precision, the standard deviation of training error (SD) is utilized as a measure

of robustness, and the coefficient of determination (Rsquared) is adopted to indicate the quality of fit. Their mathematical formulas are:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\varsigma=1}^{960}(p_\varsigma - \hat{p}_\varsigma)}{960}} \tag{15}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{\varsigma=1}^{960}(p_\varsigma - \hat{p}_\varsigma)^2}{\sum_{\varsigma=1}^{960}\left(p_\varsigma - \overline{p}_\varsigma\right)^2} \tag{16}$$

$$\mathrm{SD} = \sqrt{\frac{\sum_{\varsigma=1}^{960}\left(\left(p_\varsigma - \overline{p}_\varsigma\right) - \overline{E}\right)^2}{960}} \tag{17}$$

where p_ς represents the actual wind power value at the ς data point; $\hat{p}_\varsigma$ denotes the predicted wind power value at the ς data point; $\overline{p}_\varsigma$ represents the average actual wind power value at the ς data point; and $\overline{E}$ represents the average training error.

4.2. Analysis of Results of the Preprocessing Module Based on the IZOA-Stacking Wind Power Prediction Model

To address the significant discrepancies of various wind power characteristics, the impact of this factor on wind power output being considerable, the data preprocessing module of the established model was employed to analyze and filter each distinct wind power feature. Both grey relational analysis and Pearson Correlation Coefficient Analysis were conducted on each wind power characteristic. The grey relational degree diagram and Pearson radar chart are displayed in Figures 3 and 4, respectively, with the findings outlined in Table 1. As indicated in Table 1, wind speed characteristics exert the most substantial impact on wind power output, followed by atmospheric pressure, humidity, and temperature.

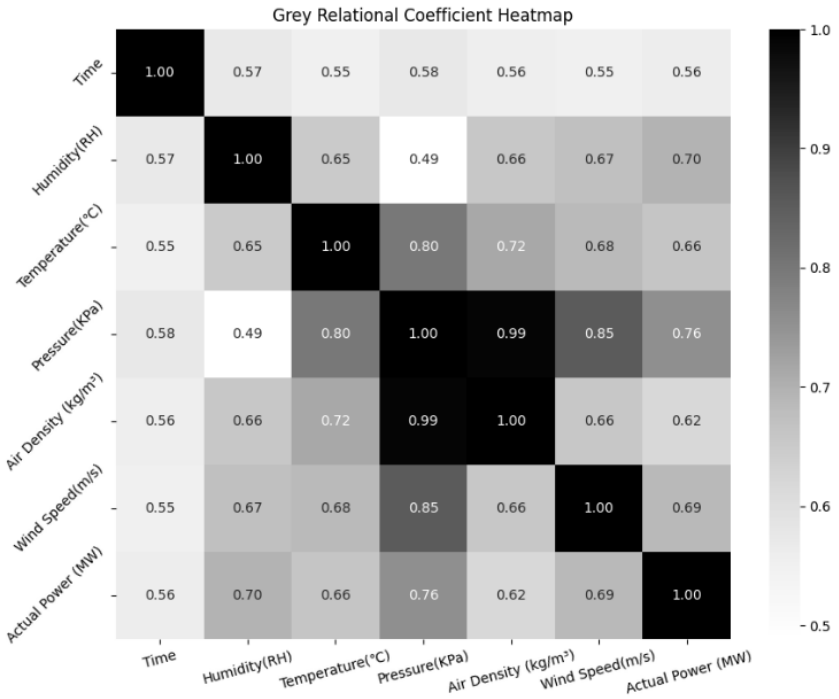


Figure 3. Grey relational degree analysis of various features in wind power forecasting.

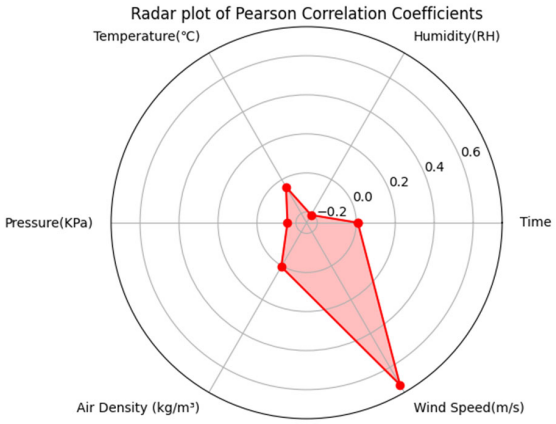


Figure 4. Pearson correlation radar chart of various features in wind power forecasting.

Table 1. Correlation analysis data values of various wind power characteristics.

Correlation Analysis	Wind Speed	Temperature	Air Pressure	Humidity	Time	Air Density
Grey Relational Analysis	0.6919	0.6623	0.7639	0.7012	0.5634	0.6183
Pearson Correlation Coefficient Analysis	0.7016	0.036	0.1563	0.2084	0	0.006
Average	0.6958	0.3492	0.4601	0.4548	0.2817	0.3122

After preserving the wind speed characteristics, anomaly detection and preprocessing are conducted on the wind speed data. To evaluate the impact of handling anomalous wind speed data on the correlation with wind power generation, a correlation analysis is performed across varying degrees of wind speed data processing. Figure 5 presents a scatter plot comparison between wind speed-associated power generation prior to outlier treatment and the predicted power generation corresponding to different levels of wind speed processing.

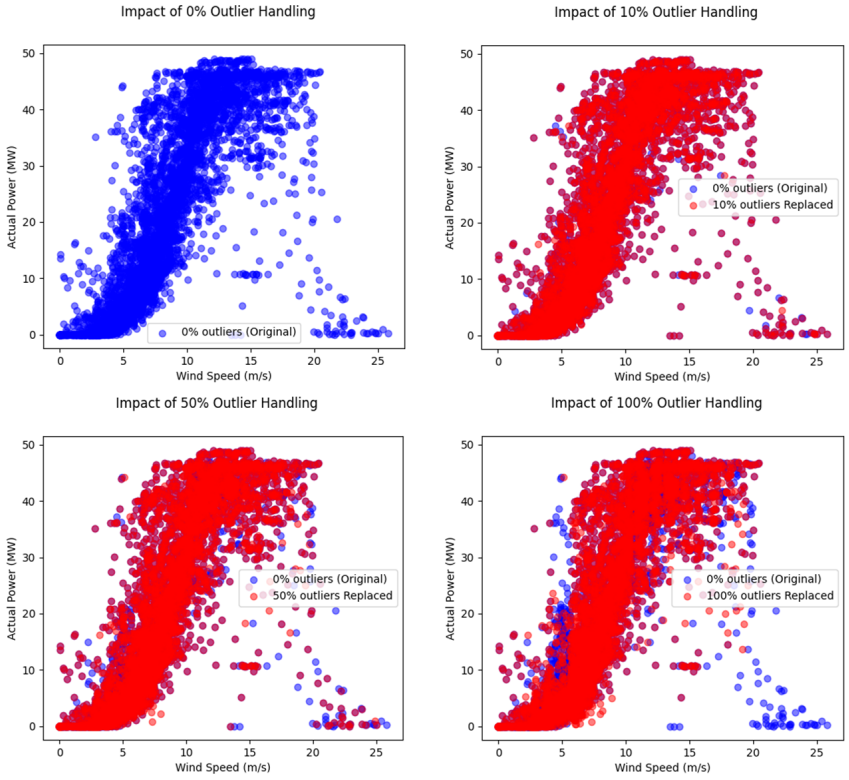


Figure 5. Scatter plot comparison of wind speed data before and after outlier mitigation.

The figure demonstrates that data cleansed using the OneSVM algorithm effectively maintains the intrinsic wind power characteristics of the wind farm, efficiently eliminating anomalous power records and wind power data during wind curtailment periods. Table 2 illustrates the impact on wind power correlation after randomly processing 10%, 50%, and 100% of the wind speed data. As shown in Table 2, upon retaining wind speed characteristics, randomly processing 10%, 50%, and 100% of the wind speed data results in enhanced correlation coefficients compared to unprocessed wind speed data. Notably, randomly processing 100% of the wind speed data exhibits the highest correlation with wind power generation.

Table 2. Association analysis data values of different wind speed processing degrees.

Association Analysis	Randomly Processed 10% Wind Speed Data	Randomly Processed 50% Wind Speed Data	Randomly Processed 100% Wind Speed Data
Grey Relational Analysis	0.7113	0.7518	0.7916
Pearson Correlation Coefficient Analysis	0.7374	0.7638	0.8293
Average	0.7244	0.7578	0.8105

To quantify the contribution of the preprocessing module to the final forecasting performance, an ablation study is conducted. Five configurations are considered: Model 1 uses all raw features without feature selection or wind speed outlier processing; Model 2 performs feature selection only; and Models 3–5 apply both feature selection and wind speed outlier processing, where 10%, 50% and 100% of the wind speed samples are corrected, respectively. All other components of the stacking framework and the IZOA optimization procedure are kept identical so that the performance differences can be attributed to the preprocessing strategy. The parameter search ranges are summarized in Table 3, and the corresponding prediction results of the five configurations are reported in Figures 6 and 7.

Table 3. Stacking model parameter setting search range.

Stacking Model Parameters	Bounds
alpha for RR	(0.1, 10)
estimators for XGBoost	(100, 1000)
LSTM_hidden_size	(10, 100)
num_layers for LSTM	(1, 5)
LSTM _learning_rate	(0.0001, 0.01)
Resnet_channels	(1, 64)
final_n_estimators for stacking XGBoost	(100, 1000)

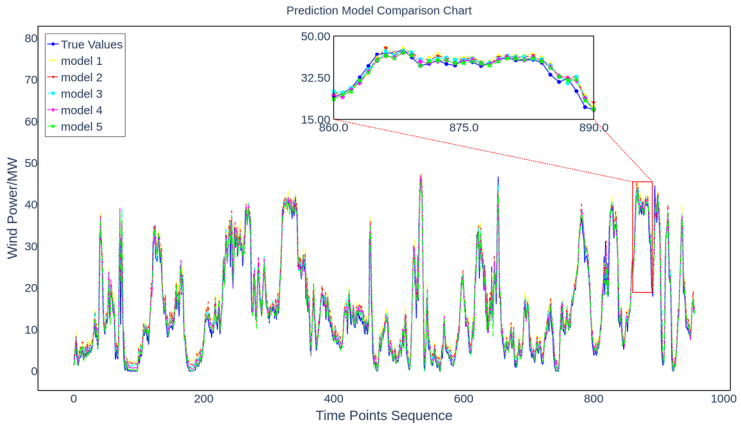


Figure 6. Wind power prediction plots for each model.

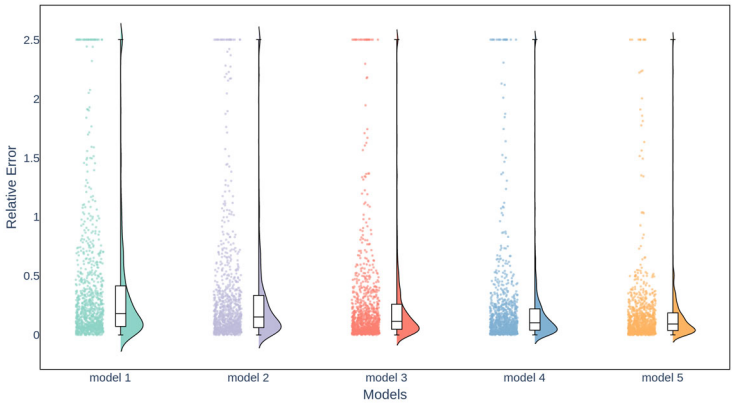


Figure 7. Box plot of wind power prediction error for each model.

The RMSE, R^2, and SD prediction evaluation metrics of each model's prediction results are shown in Table 4. As shown in Table 4, after selecting different wind power features, the RMSE, R^2, and SD of the IZOA-Stacking wind power prediction model decrease by 0.307 MW, increase by 1.27, and decrease by 0.036 MW, respectively. With 10%, 50%, and 100% of wind speed data randomly processed, the RMSE gradually decreases by an average of 0.161 MW, R^2 gradually increases by an average of 0.57, and SD gradually increases by an average of 0.019 MW. These results demonstrate that the IZOA-Stacking wind power prediction model preprocessing module does reduce wind power prediction error and improves prediction accuracy and goodness of fit.

Table 4. Evaluation metrics for each model.

Evaluation Metric	Model 1	Model 2	Model 3	Model 4	Model 5
RMSE/MW	2.917	2.610	2.345	2.208	2.128
R^2/%	93.64	94.91	95.89	96.35	96.62
SD/MW	2.095	2.059	2.076	2.099	2.117

4.3. Power Prediction Result Analysis Based on the IZOA-Stacking Wind Power Prediction Model

The preprocessing module of the IZOA-Stacking wind power forecasting model analyzes and processes the wind farm data to extract relevant features related to wind power, such as processed wind speed data, atmospheric pressure, humidity, and temperature. These features are integrated with the wind power data to create multi-dimensional input data for the IZOA-Stacking wind power forecasting model. To assess the efficiency and precision of the suggested wind power prediction approach, various models including XGBoost, LSTM, ResNet, RR, stacking, and IZOA-Stacking are compared. The results of the forecast are depicted in Figure 8, while Figure 9 presents the boxplots for each forecasting model, where models a-f correspond to XGBoost, LSTM, ResNet, RR, stacking, and IZOA-Stacking, respectively. The findings presented in Figures 8 and 9 demonstrate that the IZOA-Stacking model for wind power forecasting exhibits a closer alignment with actual measurements than alternative forecasting approaches, suggesting superior accuracy in prediction.

Table 5 shows the RMSE, R^2, and SD prediction evaluation metrics for the prediction results of each model. In terms of RMSE and SD metrics, the proposed model in this paper reduces the values by 2.749 MW and 2.12 MW, 3.58 MW and 1.762 MW, 1.905 MW and 1.859 MW, 2.23 MW and 1.641 MW, 2.211 MW and 2.202 MW, respectively, compared to the XGBoost, LSTM, ResNet, RR, and stacking prediction models. In terms of the R^2 metric, the proposed model improves the values by 14.39%, 20.96%, 8.77%, 10.81%, and 10.69%, respectively, compared to the XGBoost, LSTM, ResNet, RR, and stacking

prediction models, demonstrating superior predictive performance and enhanced stability in forecasting results.

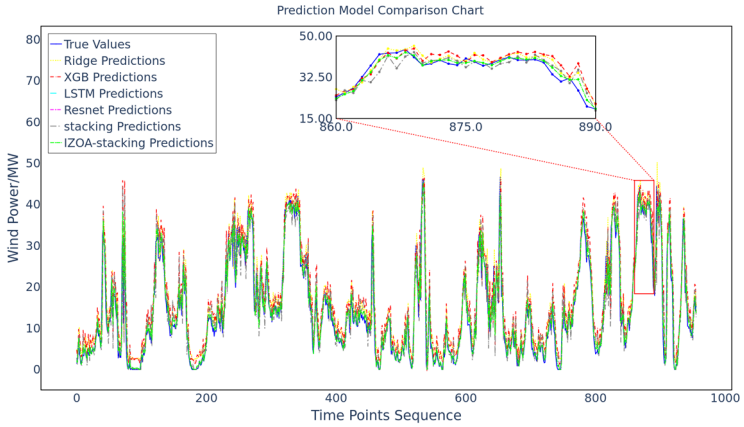


Figure 8. Wind power prediction diagram of each model.

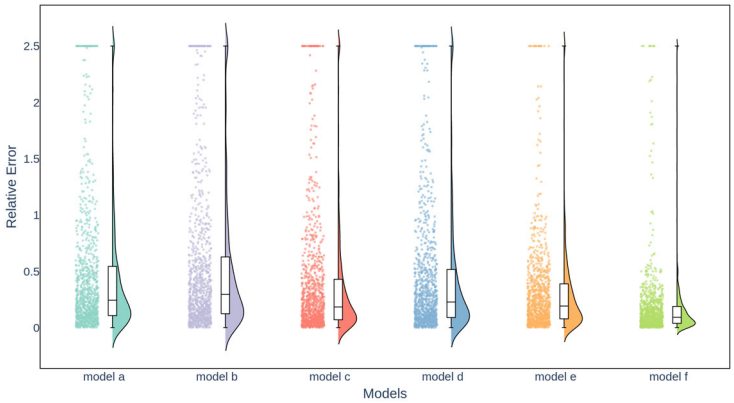


Figure 9. Boxplot of wind power prediction errors for each model.

Table 5. Evaluation metrics for each wind power forecasting model.

Evaluation Metric	XGBoost	LSTM	ResNet	RR	Stacking	IZOA-Stacking
RMSE/MW	4.877	5.708	4.033	4.358	4.339	2.128
R^2/%	82.23	75.66	87.85	85.81	85.93	96.62
SD/MW	4.237	3.879	3.976	3.758	4.319	2.117

4.4. Discussion

1. Limitations of the Data and Seasonal Generalizability

This study uses wind power data only from 1 April 2024 to 30 May 2024, covering a two-month spring period. This limited scope does not capture seasonal variations in wind intensity and turbulence. The model's performance under winter, summer, or autumn conditions remains untested, which restricts generalizability. Future work should extend the dataset to a full year.

2. Comparison with Existing Stacking Ensemble Literature

A similar methodological genre can be found in Yang et al. [35]. Compared to their work, our IZOA-Stacking introduces three distinct contributions: (1) a OneSVM-based outlier replacement module that corrects 100% of anomalous wind speed data, measurably increasing feature correlation; (2) an Improved Zebra Optimization Algorithm (IZOA) for hyperparameter tuning; and (3) a task-oriented selection of four complementary model

families (XGBoost, LSTM, ResNet, ridge regression). Our experiments confirm that the full preprocessing pipeline outperforms both raw-feature baselines and the unoptimized stacking ensemble.

3. Computational Cost and Practical Deployment Trade-offs

Beyond forecasting accuracy, the computational cost of the proposed framework is also considered. In terms of computational cost, the IZOA-Stacking framework is more demanding than individual base models because it requires repeated training of multiple learners during the cross-validation and hyperparameter search stages. In our implementation, the training time is approximately 5 to 10 times higher than that of a single XGBoost model on the same hardware, which constitutes a clear computational cost constraint. However, the forecasting time for new samples remains comparable to that of a single ensemble model, as inference only involves a forward pass through the trained base learners and meta-learner. For practical deployment, this trade-off suggests that the proposed method is suitable for scenarios where models can be updated offline while near-real-time prediction is required online. Nevertheless, this computational cost constraint potentially limits the use of the proposed framework in highly resource-constrained environments that require frequent offline retraining, as the repeated training overhead may become prohibitive. For online real-time forecasting, where future data v_{i+1} is unavailable, any newly arriving data point detected as anomalous by the OneSVM model is replaced by the most recent normal value v_{i-1}. This ensures the model can operate in real-world dispatch scenarios without look-ahead bias.

4. Implications for Practice, Policy, and Society

Implications for practice and economic impact: More accurate 15 min ahead forecasts reduce imbalance penalties, reserve costs, and wind curtailment, directly improving wind farm profitability. Implications for policy and grid management: Reliable forecasting supports high renewable integration, reduces the need for fossil-fuel reserves, and enhances grid stability. Policymakers could incentivize advanced ensemble methods via grid codes. Implications for society: Stable and cost-effective wind power lowers electricity prices, reduces pollution, and accelerates the renewable energy transition.

5. Contributions to Research and Education

This work provides a reproducible case study with fixed random seeds for data splitting and 5-fold cross-validation, ensuring all metrics (RMSE, R^2, SD) can be exactly reproduced. The framework can serve as a benchmark in graduate-level courses on renewable energy forecasting and ensemble learning. Ablation experiments clearly demonstrate the marginal contribution of each component.

5. Conclusions

This paper develops a short-term wind power forecasting framework termed IZOA-Stacking, which combines a dedicated multi-feature preprocessing module, a set of complementary base learners, and an Improved Zebra Optimization Algorithm within a stacking ensemble architecture. Rather than pursuing algorithmic novelty for its own sake, the strength of the proposed model lies in its task-oriented design: explicitly matching complementary model families to the complex, multi-scale characteristics of meteorological and operational variables in wind power forecasting. Rather than proposing an entirely new learning paradigm, the contribution of this work lies in integrating and tailoring established techniques to the specific characteristics of wind power data and in systematically evaluating the effect of each component.

First, the preprocessing module jointly applies grey relational analysis and Pearson correlation analysis to quantify the relevance of meteorological and operational variables

to wind power output. By averaging the normalized relevance scores from the two criteria and discarding weakly correlated variables, the module mitigates the imbalance of feature scales and reduces the risk of overfitting during model training. When wind speed is retained, a One-Class SVM is used to detect and correct anomalous wind speed values, which further improves the consistency between the input features and the actual operating conditions of the wind farm.

Second, an ensemble of XGBoost, LSTM, ResNet and ridge regression is constructed as the first-level learners in the stacking framework. These models are intentionally selected because they exhibit complementary inductive biases: XGBoost handles tabular and nonlinear relationships efficiently, LSTM is suitable for sequential temporal dependencies, ResNet enhances multi-scale feature extraction, and ridge regression provides a simple, low-variance linear baseline. This task-oriented selection ensures that the ensemble can simultaneously capture nonlinear interactions, temporal dynamics, multi-scale patterns, and approximate linear trends—capabilities that are essential for wind power data characterized by fluctuating meteorological variables and operational constraints. Using XGBoost as the meta-learner allows the framework to exploit heterogeneous base-model outputs and alleviate individual model weaknesses.

Third, an Improved Zebra Optimization Algorithm is employed to tune the hyperparameters of all base learners and of the meta-learner. The elite opposition-based strategy in IZOA enhances global exploration and accelerates convergence, reducing the sensitivity of the overall framework to manual parameter selection. Comparative experiments show that, under the same training and test sets, the IZOA-Stacking model achieves lower RMSE and SD and a higher R^2 than the individual base models, the conventional stacking ensemble and other reference approaches. To providing a clearer view of the role of each component, ablation experiments are conducted on the preprocessing and ensemble modules. When only feature selection is applied but wind speed outlier processing is omitted, the predictive performance improves compared with using all raw features, but remains inferior to the full preprocessing pipeline. When different proportions of wind speed data are corrected, the correlation between wind speed and power output, as well as the overall forecasting accuracy, is further enhanced. Similarly, comparing the single models, the unoptimized stacking ensemble and the IZOA-Stacking model, confirms that the combination of stacking and IZOA-based hyperparameter tuning yields the most accurate and stable forecasts.

Despite these advantages, several limitations should be acknowledged. The case study is based on data from a single wind farm within a relatively short time span, which may restrict the generalization of the results to other locations, seasons and turbine configurations. In addition, the multi-stage framework, including feature analysis, anomaly detection and meta-learning, introduces additional training-time computational cost compared with simpler single-model approaches. Although the prediction stage remains efficient once the model is trained, future work will consider more detailed complexity analysis and potential model simplifications for real-time or resource constrained deployments. Moreover, the current experimental design, while including multiple benchmark models, does not yet cover the full range of state-of-the-art optimization algorithms and ensemble strategies.

We acknowledge that the current study is based on a single 50 MW wind farm with only two months of data (April–May 2024), which limits the generalizability of our findings across different seasons, geographies, and turbine configurations. In future research, we plan to extend the proposed methodology to longer forecasting horizons and multiple wind farms, and to incorporate additional baseline models and optimization techniques for a more comprehensive comparison. Another promising direction is to formulate multi-objective optimization schemes that explicitly balance forecast accuracy, model complexity and computational cost. Beyond these extensions, the real-time application of wind power

predictions can be further enhanced by coupling them with intelligent control systems. For instance, exploring the integration of the proposed forecasting model with reinforcement learning controllers in off-grid microgrids could optimize hydrogen storage utilization and minimize unmet electrical loads [36]. Such a combination would enable adaptive decision-making under uncertainty, leveraging forecasted wind power to schedule hydrogen production and storage operations dynamically. Additionally, more advanced learning paradigms, such as meta-reinforcement learning, offer the potential to accelerate the adaptation to new wind farms or rapidly changing weather patterns with minimal retraining [37]. These directions, together with the multi-objective optimization schemes mentioned above, are expected to provide power system operators with more robust, interpretable, and operationally aware tools for planning and real-time control in renewable-dominated power systems.

Author Contributions: Conceptualization, H.J.; data curation, X.W.; formal analysis, H.J.; methodology, T.S.; resources, H.J.; software, T.S. and X.W.; project administration, H.J.; supervision, Q.L.; validation, T.S.; visualization, Q.L.; writing—original draft preparation, H.J.; writing—review and editing, T.S. and Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (62473269), Basic Scientific Research Project of Liaoning Provincial Department of Education (LJ222411632036), Liaoning Province Key Research and Development Project (2024JH2/102500093), and Liaoning Revitalization Talents Program (XLYC2403160).

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

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

References

1. Tsai, W.C.; Hong, C.M.; Tu, C.S.; Lin, W.M.; Chen, C.H. A review of modern wind power generation forecasting technologies. *Sustainability* **2023**, *15*, 10757. [CrossRef]
2. Guo, Z.; Wei, F.; Qi, W.; Han, Q.; Liu, H.; Feng, X.; Zhang, M. A time series prediction model for wind power Based on the Empirical Mode Decomposition-Convolutional Neural Network-Three-Dimensional Gated Neural Network. *Sustainability* **2024**, *16*, 3474. [CrossRef]
3. Wang, X.; Hao, Y.; Yang, W. Novel wind power ensemble forecasting system based on mixed-frequency modeling and interpretable base model selection strategy. *Energy* **2024**, *297*, 131142. [CrossRef]
4. Meng, A.; Zhang, H.; Dai, Z.; Xian, Z.; Xiao, L.; Rong, J.; Li, C.; Zhu, J.; Li, H.; Yin, Y.; et al. An adaptive distribution-matched recurrent network for wind power prediction using time-series distribution period division. *Energy* **2024**, *299*, 131383. [CrossRef]
5. 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]
6. Cao, W.; Wang, G.; Liang, X.; Hu, Z. A STAM-LSTM model for wind power prediction with feature selection. *Energy* **2024**, *296*, 131030. [CrossRef]
7. Lipu, M.S.; Miah, M.S.; Hannan, M.; Hussain, A.; Sarker, M.R.; Ayob, A.; Saad, M.; Mahmud, M.S. Artificial Intelligence Based Hybrid Forecasting Approaches for Wind Power Generation: Progress, Challenges and Prospects. *IEEE Access* **2021**, *9*, 102460–102489. [CrossRef]
8. Sideratos, G.; Hatziargyriou, N.D. An advanced statistical method for wind power forecasting. *IEEE Trans. Power Syst.* **2007**, *22*, 258–265. [CrossRef]
9. Liu, H.; Chen, C. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Appl. Energy* **2019**, *249*, 392–415. [CrossRef]
10. Simankov, V.; Buchatskiy, P.; Teploukhov, S.; Onishchenko, S.; Kazak, A.; Chetyrbok, P. Review of Estimating and Predicting Models of the Wind Energy Amount. *Energies* **2023**, *16*, 5926. [CrossRef]
11. Zhang, J.; Li, H.; Cheng, P.; Yan, J. Interpretable Wind Power Short-Term Power Prediction Model Using Deep Graph Attention Network. *Energies* **2024**, *17*, 384. [CrossRef]

12. Singh, S.; Parmar, K.S.; Makkhan, S.J.S.; Kaur, J.; Peshoria, S.; Kumar, J. Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries. *Chaos Solitons Fractals* **2020**, *139*, 110086. [CrossRef]
13. Chen, K. Forecasting systems reliability based on support vector regression with genetic algorithms. *Reliab. Eng. Syst. Saf.* **2007**, *92*, 423–432. [CrossRef]
14. Srivastava, T.; Vedanshu; Tripathi, M.M. Predictive analysis of RNN, GBM and LSTM network for short-term wind power forecasting. *J. Stat. Manag. Syst.* **2020**, *23*, 1. [CrossRef]
15. Alhussan, A.; El-kenawy, E.M.; Abdelhamid, A.A.; Ibrahim, A.; Eid, M.; Khafaga, D.S. Wind Speed Forecasting Using Optimized Bidirectional LSTM Based on Dipper Throated and Genetic Optimization Algorithms. *Front. Energy Res.* **2023**, *11*, 1172176. [CrossRef]
16. Sun, H.; Cui, Q.; Wen, J.; Kou, L.; Ke, W. Short-term wind power prediction method based on CEEMDAN-GWO-Bi-LSTM. *Energy Rep.* **2024**, *1*, 1487–1502. [CrossRef]
17. Jiang, L.; Wang, Y. A wind power forecasting model based on data decomposition and cross-attention mechanism with cosine similarity. *Electr. Power Syst. Res.* **2024**, *229*, 110156. [CrossRef]
18. Zhang, Y.; Sun, H.; Guo, Y. Wind power prediction based on PSO-SVR and grey combination model. *IEEE Access* **2019**, *7*, 136254–136267. [CrossRef]
19. Zhu, Y. Research on adaptive combined wind speed prediction for each season based on improved gray relational analysis. *Environ. Sci. Pollut. Res.* **2023**, *30*, 12317–12347. [CrossRef] [PubMed]
20. Li, F.; Sun, C.; Han, W.; Yan, T.; Li, G.; Zhao, Z.; Sun, Y. Medium-term load forecasting of power system based on BiLSTM and parallel feature extraction network. *IET Gener. Transm. Distrib.* **2023**, *18*, 190–201. [CrossRef]
21. Yang, M.; Shi, C.; Liu, H. Day-ahead wind power forecasting based on the clustering of equivalent power curves. *Energy* **2021**, *218*, 119515. [CrossRef]
22. Qin, L.; Sun, N.; Dong, H. Adaptive Double Kalman Filter Method for Smoothing Wind Power in Multi-Type Energy Storage System. *Energies* **2023**, *16*, 1856. [CrossRef]
23. Li, N.L.; Zhao, X.; Tseng, M.; Tan, R.R. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *J. Clean. Prod.* **2020**, *242*, 118447. [CrossRef]
24. Rangaswamy, S.; Vineeth, N.; Asha, G. Deep Learning Approaches for Ransomware Detection: Assessing CNN and CNN-LSTM Models using Class Imbalance Methods. In Proceedings of the 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 23–25 October 2024.
25. An, G.; Jiang, Z.; Cao, X.; Liang, Y.; Zhao, Y.; Li, Z.; Dong, W.; Sun, H. Short-Term Wind Power Prediction Based On Particle Swarm Optimization-Extreme Learning Machine Model Combined with Adaboost Algorithm. *IEEE Access* **2021**, *9*, 94040–94052. [CrossRef]
26. Trojovská, E.; Dehghani, M.; Trojovský, P. Zebra Optimization Algorithm: A New Bio-inspired Optimization Algorithm for Solving Optimization Algorithm. *IEEE Access* **2022**, *10*, 49445–49473. [CrossRef]
27. Krithiga, G.; Senthilkumar, S.; Alharbi, M.; Mangaiyarkarasi, S.P. Design of modified long short-term memory-based zebra optimization algorithm for limiting the issue of SHEPWM in multi-level inverter. *Sci. Rep.* **2024**, *14*, 22439. [CrossRef] [PubMed]
28. Zhang, X.; Qin, X.S.; Qin, J.; Wang, B.; Yuan, Y. Short-term wind power prediction with a new PCC-GWO-VMD and BiGRU hybrid model enhanced by attention mechanism. *J. Renew. Sustain. Energy* **2025**, *17*, 3. [CrossRef]
29. Xiong, X.; Guo, X.; Zeng, P.; Zou, R.; Wang, X. A Short-Term Wind Power Forecast Method via XGBoost Hyper-Parameters Optimization. *Front. Energy Res.* **2022**, *10*, 905155. [CrossRef]
30. 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]
31. Han, Y.; Cao, L.; Geng, Z.; Ping, W.; Zuo, X.; Fan, J.; Wan, J.; Lu, G. Novel economy and carbon emissions prediction model of different countries or regions in the world for energy optimization using improved residual neural network. *Sci. Total Environ.* **2022**, *860*, 160410. [CrossRef]
32. Zheng, Y.; Ge, Y.; Muhsen, S.; Wang, S.; Elkamchouchi, D.H.; Ali, E.; Ali, H.E. New ridge regression, artificial neural networks and support vector machine for wind speed prediction. *Adv. Eng. Softw.* **2023**, *179*, 103426. [CrossRef]
33. Jia, C.X.; Zhang, L.; Zhang, C.J.; Li, Y.T. Summary of demand-side response and energy transaction strategy of intelligent building clusters driven by data. *J. Sustain. Energy* **2022**, *1*, 8–17. [CrossRef]
34. Guo, F.; Mo, H.; Wu, J.; Pan, L.; Zhou, H.; Zhang, Z.; Li, L.; Huang, F. A hybrid stacking model for enhanced short-term load forecasting. *Electronics* **2024**, *13*, 2719. [CrossRef]
35. Yang, Y.; Li, Y.; Cheng, L.; Yang, S. Short-Term Wind Power Prediction Based on a Modified Stacking Ensemble Learning Algorithm. *Sustainability* **2024**, *16*, 5960. [CrossRef]

36. Zhang, T.; Giannelos, S.; Pudjianto, D.; Strbac, G. Performance Evaluation of Reinforcement Learning for Hydrogen Integration in Renewable Microgrids. In Proceedings of the 14th International Conference on Renewable Power Generation, Shanghai, China, 24–26 October 2025; Volume 38, pp. 877–883.
37. Peng, S. Overview of Meta-Reinforcement Learning Research. In Proceedings of the 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 18–20 December 2020; pp. 54–57.

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.