Online ISSN: 3104-9338
Print ISSN: 3104-932X

An Ultra-Short-Term Wind Power Prediction Model Based on Crested Porcupine Optimizer and Cross-Domain Attention Mechanism

Jieni Ning[1], Weishang Huang [2*]

[1] South China Normal University

[2] Belarusian National Technical University

Received: October 29, 2025

Accepted: October 29, 2025

Published online: December 1, 2025

To appear in: *International Journal of Advanced AI Applications,* Vol. 1, No. 8 (December 2025)

* Corresponding Author: Wei shang Huang (hws178170365@163.com)

Abstract. Wind power prediction is one of the critical tasks for ensuring power grid stability. Traditional forecasting methods often encounter challenges due to the non-stationarity and complexity of wind power data. To address these issues, this paper proposes a CPO-VMD-WTC-Crossformer model that integrates Crested Porcupine Optimizer (CPO)-optimized Variational Mode Decomposition (VMD) with Wavelet Transform Convolution (WTC) to enhance the accuracy and robustness of wind power forecasting. The CPO algorithm is first applied to optimize VMD's mode number and penalty factor, thereby improving its decomposition performance for wind power signals and providing high-quality input for the forecasting model. WTC utilizes its distinctive multiresolution analysis capability to precisely capture local key features such as subtle wind speed variations, compensating for Crossformer's limitations in local feature processing. Crossformer employs innovative cross-attention and cross-domain feature fusion mechanisms to efficiently analyze the spatiotemporal coupling characteristics of wind power data with low computational cost, enabling accurate forecasting. Experimental results demonstrate that compared to conventional forecasting models, the proposed model achieves significant improvements in forecasting accuracy, contributing to more scientific and stable power system dispatching.

Keywords: *Wind Power Prediction; Crested Porcupine Optimizer; Variational Mode Decomposition; Wavelet Transform Convolution; Crossformer*

1. Introduction

Against the backdrop of continuously growing global energy demand, the reserves of conventional energy resources such as coal, oil, and natural gas are gradually diminishing. The development of renewable energy sources, including solar, wind, and hydropower, has thus become crucial to overcoming resource constraints and ensuring energy supply. As a clean energy source, wind power holds significant potential in reducing greenhouse gas emissions, decreasing energy dependence, and promoting sustainable economic development [1]. Despite the abundance of global wind energy resources and their vast theoretical reserves, their inherent intermittency and volatility pose challenges to power system stability. Unstable wind speeds may lead to fluctuations in power generation, affecting grid load balancing and the continuity of electricity supply. Therefore, accurate wind power forecasting is essential for optimizing grid dispatch, improving operational efficiency, enhancing the competitiveness of wind power in energy markets, and reducing wind curtailment [2].

Based on the time scale, wind power prediction (WPP) can be classified into three main types: ultra-short-term, short-term, and long-term forecasting [3]. Ultra-short-term wind power prediction primarily utilizes advanced machine learning models to deeply analyze historical data characteristics of wind farms, including wind speed, wind direction, and power output, thereby achieving accurate estimation of wind power generation for the next 0–4 hours [4]. Accurate ultra-short-term wind power prediction can significantly enhance the dispatch efficiency and stability of power systems. Grid operators can optimize power dispatch strategies, proactively respond to power fluctuations, and reduce reliance on conventional energy sources. Furthermore, it improves the grid integration capacity of wind power, mitigates wind curtailment, and maximizes the utilization of renewable energy.

Wind power prediction methods can be categorized into three main approaches: physical models, statistical models, and artificial intelligence models. Physical models forecast wind power by analyzing complex meteorological factors, utilizing Numerical Weather Prediction (NWP) data [5] while accounting for geographical features such as terrain and surface roughness. These models generate predictions by solving thermodynamic partial differential equations, making them particularly suitable for newly constructed wind farms or those with limited historical data [6]. Although established physical models including SOWIE [7] and HIRLAM [8] are available, their effectiveness is constrained by the need for highly accurate meteorological data, computational complexity, and limited robustness - particularly in ultra-short-term forecasting. However, they demonstrate relatively better performance in long-term predictions. Statistical methods perform forecasting by identifying statistical patterns from historical wind power data and meteorological information. Statistical methods demonstrate advantages in computational simplicity and reduced data requirements compared to physical models, yet their strong dependence on historical data may compromise prediction accuracy and adaptability under novel meteorological conditions. In contrast, artificial intelligence (AI) approaches exhibit superior performance in processing large-scale datasets and capturing complex nonlinear relationships between wind power and meteorological factors, owing to the rapid development of AI technologies and large-scale models, which has garnered increasing recognition of their significant potential in wind power forecasting [9]. Conventional AI techniques, including Artificial Neural Networks (ANN) [10-11], Support Vector Machines (SVM) [12-13], Random Forest (RF) [14], and XGBoost [15], often rely on manual feature extraction, limiting their ability to model complex nonlinear relationships and high-dimensional data features, consequently constraining prediction accuracy and adaptability. Deep learning methods address these limitations through automated feature learning, enabling effective handling of complex patterns and high-dimensional data, thereby significantly enhancing prediction performance, as exemplified by Convolutional Neural Networks (CNN) [16], Gated Recurrent Units (GRU) [17-18], Generative Adversarial Networks (GAN) [19-20], and Bidirectional Long Short-Term Memory networks (BiLSTM) [21]. Although Recurrent Neural Networks (RNN) excel in processing sequential data, they frequently encounter gradient vanishing and explosion problems during long-sequence learning, adversely affecting model training efficiency and effectiveness. The Transformer model effectively addresses these limitations by incorporating a self-attention mechanism, thereby providing more efficient sequence modeling capabilities. Wang et al. [22] employed Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose wind power time series, utilized sample entropy to distinguish between high-frequency and low-frequency components, and subsequently applied Transformer and BiGRU-Attention models for component-wise prediction before integrating the results. Qu et al. [23] adapted the Transformer architecture from natural language processing (NLP) to wind power forecasting, successfully capturing inter-farm correlations to generate accurate predictions. Wang et al. [24] proposed an asymmetric multi-quantile generative Transformer approach that produces interval predictions through a single forward pass while employing conformal quantile regression to calibrate prediction intervals, yielding narrower yet more precise uncertainty estimates. Although Transformers excel at capturing global features and long-term dependencies in wind power prediction, their capability in modeling local characteristics remains limited when handling complex wind data patterns. The Crossformer model [25] enhances prediction accuracy and performance through its cross-attention modules and cross-domain feature fusion mechanism, which effectively addresses spatiotemporal coupling and adapts to temporal variations, though it may still exhibit prediction biases under conditions with minimal wind speed fluctuations.

With the increasing demands for wind power forecasting accuracy, data preprocessing methods have become crucial for enhancing prediction performance, among which decomposition strategies have gained widespread application as an effective approach. Jiajun et al. [27] proposed a hybrid method combining Wavelet Transform (WT), Deep Belief Network (DBN), Light Gradient Boosting Machine (LGBM), and Random Forest (RF), where wind speed series are first decomposed before employing DBN for high-dimensional feature extraction and tree-based models for prediction. However, WT suffers from difficulties in selecting appropriate wavelet basis functions and scales, along with significant edge effect errors. Gao et al. [28] utilized Empirical Mode Decomposition (EMD) to decompose power load sequences, selecting highly correlated intrinsic mode functions as features to input into Gated Recurrent Unit (GRU) networks together with original sequences. Nevertheless, EMD demonstrates notable limitations including noise sensitivity leading to unstable decomposition results and mode mixing phenomena. Ding et al. [29] applied Complementary Ensemble Empirical Mode Decomposition (CEEMD) to decompose non-stationary wind power time series into relatively stable components, then employed Whale Optimization Algorithm (WOA)-optimized Kernel Extreme Learning Machine (KELM) for individual component prediction before reconstructing final results through superposition. However, CEEMD exhibits high computational complexity, noise sensitivity, and potential reconstruction errors. Compared to WT, EMD and CEEMD, Variational Mode Decomposition (VMD) demonstrates superior mode decomposition capability in wind power forecasting, effectively avoiding mode mixing while maintaining higher decomposition accuracy and computational efficiency without requiring repeated noise addition like CEEMD. VMD shows stronger adaptability in handling complex non-stationary signals, thereby improving both prediction accuracy and stability. In wind power decomposition, VMD parameter configuration proves critical as improper settings (particularly for mode number k and penalty factor ε) may lead to decomposition inaccuracies that compromise prediction performance, thus typically requiring optimization algorithms for parameter tuning to enhance both decomposition quality and forecasting accuracy.

To address these challenges, we propose an innovative hybrid model integrating Crowned Porcupine Optimizer (CPO) [30]-optimized Variational Mode Decomposition (VMD), Wavelet Transform-based Convolution (WTC), and Crossformer architectures. The proposed approach employs the CPO algorithm to automatically optimize VMD's critical parameters (mode number k and penalty factor ε), significantly enhancing the decomposition accuracy of non-stationary wind power signals. The WTC module effectively captures time-frequency localized features through wavelet convolution operations, while the Crossformer architecture provides superior spatiotemporal feature extraction capabilities through its cross-attention mechanisms. This comprehensive framework demonstrates remarkable improvements in both prediction accuracy and stability for wind power forecasting applications. The main innovations of this paper include the following aspects:

(1) The CPO algorithm is a novel optimization method that optimizes VMD's key parameters (mode number k and penalty factor) to improve the accuracy and stability of VMD decomposition. It effectively avoids local optima issues common in traditional optimization algorithms and achieves better global optimization performance for complex wind power forecasting tasks.

(2) Crossformer is a Transformer-based model that utilizes self-attention mechanisms to capture both long-term and short-term dependencies in data, making it suitable for time series modeling. For wind power forecasting, Crossformer can extract useful information from multiple feature dimensions and demonstrates strong predictive performance on complex dataset

(3) WTC wavelet transform is a time-frequency analysis method that extracts signal features across different time and frequency domains. By integrating WTC with Crossformer, the model effectively processes local signal characteristics and improves prediction accuracy for wind power fluctuations.

2. Methods and principles

2.1. Variational Mode Decomposition (VMD)

Variational Mode Decomposition (VMD), proposed by Dragomiretskiy K., is an adaptive signal decomposition method primarily used for analyzing non-stationary signals. This method decomposes complex signals into K amplitude-modulated and frequency-modulated sub-signals, with wide applications in processing historical data sequences such as wind power generation. VMD performs non-recursive decomposition of signals into a specified number of mode components, each possessing physical significance. The approach demonstrates excellent noise suppression capabilities while effectively reducing signal non-stationarity. The VMD decomposition process essentially involves formulating and solving a set of optimization problems, beginning with Wiener filtering for noise reduction and parameter initialization to estimate the center angular frequencies w_k of K modes. Subsequently, the Alternating Direction Method of Multipliers (ADMM) updates each mode's frequency and signal components while modulating them to corresponding frequency bands to minimize the total bandwidth. VMD achieves signal decomposition by formulating and solving a constrained variational problem, as established in Equations (1) and (2).

$$\min_{\{u_k\}\{\omega_t\}}\left\{\sum_k\left\|\vartheta_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\} \tag{1}$$

$$s.t.\ \sum_k u_{k=f} \tag{2}$$

In the equations: $L(\cdot)$ represents the augmented Lagrangian function; ε denotes the penalty factor; $\lambda(t)$ stands for the Lagrangian multiplier; and $\langle\cdot\rangle$ indicates the inner product operator.

VMD (Variational Mode Decomposition) applications, selecting appropriate mode number k and penalty factor ε is crucial. These parameters collectively determine VMD decomposition performance, particularly critical for wind power signal analysis. The k value affects the number of resulting Intrinsic Mode Functions (IMFs) - insufficient k may cause incomplete signal component extraction leading to mode mixing, while excessive k may induce over-decomposition generating redundant features. The ε value controls IMF bandwidth: excessive ε produces overly narrow bandwidth failing to capture signal characteristics, whereas insufficient ε causes bandwidth broadening leading to component mixing. Traditional manual parameter tuning often fails to achieve optimal decomposition, prompting this study to employ intelligent optimization algorithms for automatically determining these key parameters' optimal combination.

2.2 The CPO-Optimized VMD Methodology

The Crested Porcupine Optimizer (CPO) is a novel metaheuristic algorithm that demonstrates superior global search capability, enhanced adaptability, and robust performance compared to conventional optimization approaches such as Simulated Annealing, Genetic Algorithm (GA), and Particle Swarm Optimization (PSO). This method exhibits particular efficacy in addressing complex, multimodal optimization problems with multiple constraints.

The algorithm draws inspiration from the defensive behaviors of crested porcupines, simulating four distinct defense strategies employed when confronting predators: visual intimidation, acoustic intimidation, odor attack, and physical assault. These strategies correspond to exploration and exploitation behaviors within the algorithmic framework.The fundamental principles of the Crested Porcupine Optimizer (CPO) algorithm are presented as follows:

2.2.1 Population initialization

The positions of individuals in the initial population are generated as follows:

$$X_J = L + y \times (T - L) \tag{3}$$

where L and T represent the lower and upper bounds of the search space respectively, and y is a random number uniformly distributed between 0 and 1.

2.2.2 Cyclic Population Reduction Technique

The mathematical model of the cyclic population reduction technique is expressed as:

$$K = K\left(K^{'} - K_{min}()\left(1 - \left(\frac{t\%T_{max}}{Tmax \times T}\right)\right)\right)_{min} \tag{4}$$

where K represents the cyclic count variable, t denotes the current function evaluation count, T_{max} signifies the maximum function evaluation count, and K_{min} corresponds to the minimum individual number in the newly generated population.

2.2.3 Exploration Phase

(a) Primary Defense Strategy

In the first defense strategy, the formula for updating individual positions is:

$$\overrightarrow{X_\iota^{t+1}} = \overrightarrow{X_\iota^{t}} + \tau_1 \times \overrightarrow{\left|2 \times \tau_2 \times \overrightarrow{X^t{}_{CP}} - \overrightarrow{Y_\iota^{t}}\right|} \tag{5}$$

Y_i^t represents the vector generated between the current CP and a randomly selected CP

(b) Second Defense Strategy

In the second defense strategy, the formula for updating individual positions is:

$$X_i^{t+1} = (1 - U_1) \times X_i^t + U_1 \times \left(Y + \tau_3 \times (X_{r1}^t - X_{r2}^t)\right) \tag{6}$$

2.2.4 Exploitation Phase

(a) Third Defense Strategy

In the third defense strategy, the formula for updating individual positions is:

$$\overrightarrow{x_i^{t+1}} = \left(1 - \overrightarrow{U_1}\right) \times \overrightarrow{x_i^t} + \overrightarrow{U_1} \times \left(\overrightarrow{x_{r1}^{'}} + S_i^{'} \times \left(\overrightarrow{x_{r2}^{'}} - \overrightarrow{x_{r3}^{'}}\right) - \tau_3 \times \vec{\delta} \times \gamma_t \times S_i^{'}\right) \tag{7}$$

where δ is the parameter controlling search direction, $S_i^{'}$ denotes the defense factor, and $S_i^{'}$represents the scent diffusion factor.

(b) Fourth Defense Strategy

In the fourth defense strategy, the formula for updating individual positions is:

$$\overrightarrow{x_i^{t+1}} = \overrightarrow{x_{CP}^t} + (\alpha(1-\tau_4)+\tau_4) \times \left(\vec{\delta} \times \overrightarrow{x_{CP}^t} - \overrightarrow{x_i^t}\right) - \tau_5 \times \vec{\delta} \times \gamma_t \times \overrightarrow{F_i^t} \tag{8}$$

Where α represents the convergence speed factor. These formulas collectively reflect the update mechanisms of the CPO algorithm across different phases, which integrate the natural defense behaviors of crested porcupines to achieve optimization objectives. The specific steps for CPO to optimize VMD parameters are shown in Fig. 1.

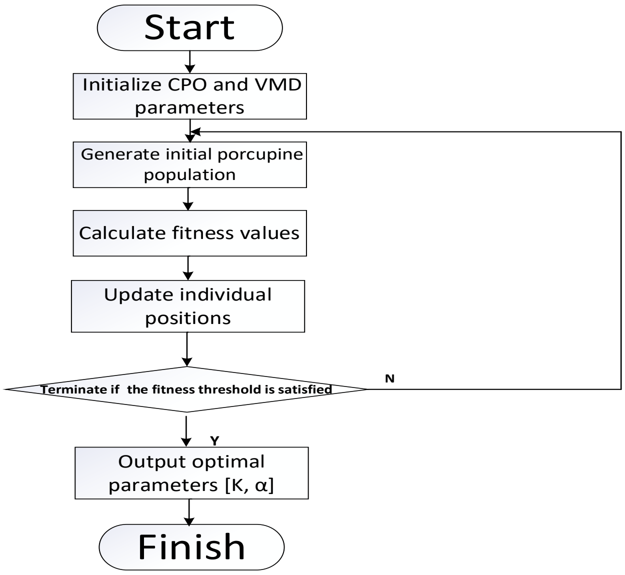

Fig. 1 The process of VMD parameters optimized by CPO.

2.3 Wavelet Transform Convolution

WTC (Wavelet Transform Convolution) is a neural network architecture that integrates wavelet transform and convolutional operations, aiming to enhance the receptive field and feature extraction capabilities of convolutional neural networks (CNNs) while effectively avoiding the issue of excessive parameters. With its advantages of a large receptive field, small parameter size, and precise frequency selectivity, WTC convolution demonstrates superior performance and robustness compared to other convolutions when processing data with complex frequency structures and rich contextual information. The core of WTC convolution lies in decomposing the input data into different frequency components through wavelet transform, performing convolutional operations on these components, and then integrating the results through inverse wavelet transform. This method can expand the receptive field without significantly increasing the number of parameters, thereby capturing broader contextual information.

2.3.1 Wavelet Transform (WT)

First, the input data X is processed through the wavelet transform, which employs a set of filters (including low-pass and high-pass filters) to perform convolutional operations on the data, thereby decomposing it into low-frequency components (e.g X_{LL}) and high-frequency components (e.g X_{LH},X_{HL},X_{HH}).

$$[X_{LL}, X_{LH}, X_{HL}, X_{HH}] = Conv([f_{LL}, f_{LH}, f_{HL}, f_{HH}], X) \tag{9}$$

where f_{LL} is the low-pass filter and f_{LH}, f_{HL}, f_{HH} is the high-pass filter.

2.3.2 Cascaded Wavelet Decomposition

Subsequently, the extracted low-frequency components undergo further wavelet transform to perform cascaded decomposition, enabling a more refined separation of frequency components and thereby enhancing frequency resolution. Through this multi-level decomposition mechanism, the complex frequency structures within the input data can be captured with greater precision.

$$\left[X_{LL}^{(i)}, X_{LH}^{(i)}, X_{HL}^{(i)}, X_{HH}^{(i)}\right] = WT\left(X_{LL}^{(i-1)}\right) \tag{10}$$

where $X_{LL}^{(i)}$ represents the low-frequency component at the i level and $X_{LL}^{(i-1)}$ denotes the low-frequency component at the $i-1$ level.

2.3.3 Convolutional Operation

Depthwise convolutions with small kernels are performed on the frequency maps obtained through wavelet transform. Since the input data has been decomposed into components of different frequencies, the convolutional operations can be conducted at reduced spatial dimensions, which not only expands the receptive field but also enables effective extraction and subsequent integration of features across various frequency components.

$$\left[Y_{(LL)}^{i}, Y_{LH}^{(i)}, Y_{HL}^{(i)}, Y_{HH}^{(i)}\right] = Conv(W^{(i)}, \left[X_{LL}^{(i)}, X_{LH}^{(i)}, X_{HL}^{(i)}, X_{HH}^{(i)}\right]) \tag{11}$$

where $W^{(i)}$ denotes the convolutional kernel at the i level.

2.3.4 Inverse Wavelet Transform (IWT)

The convolved outputs are reconstructed through the inverse wavelet transform to form the final feature map. As the reverse process of wavelet decomposition, the inverse transform employs specific filters to reintegrate frequency components, synthesizing a comprehensive feature map with enriched informational representation.

$$Z^{(i)} = IWT\left(Y_{LL}^{(i)} + Z^{(i+1)},\ Y_{LH}^{(i)},\ Y_{HL}^{(i)},\ Y_{HH}^{(i)}\right) \tag{12}$$

where $Z^{(i)}$ represents the output feature map at the i level and $Z^{(i+1)}$ denotes the output feature map at the $i+1$ th level.

The final output feature map $Z^{(0)}$ is generated by fusing and reconstructing multi-level convolutional results through the inverse wavelet transform.

$$Z^{(0)} = Y_{LL}^{(0)} + Z^{(1)} \tag{13}$$

2.4 Crossformer

In multivariate time series forecasting, model performance is closely related to data feature capture capability. Traditional Transformers process multivariate time series by concatenating data from different dimensions at the same timestep into vectors, which can capture cross-time dependencies but fail to explicitly model cross-dimension dependencies, thereby limiting predictive performance. To overcome this limitation, the 2023 ICLR conference proposed Crossformer, a Transformer-based architecture specifically designed for multivariate time series forecasting. Its core advantage lies in addressing the insufficient cross-dimension dependency capture of traditional models through a unique structure and algorithm that leverages inter-variable dimensional correlations to enhance prediction accuracy and reliability.

The Crossformer model, the multivariate time series data is first embedded into a 2D vector space through the Dimension-wise Segment Embedding layer (DSW), preserving both temporal and dimensional features; then the Time-Series Attention mechanism (TSA) is introduced to capture dependencies across timesteps and dimensions; finally, these components are integrated to construct a Hierarchical Encoder-Decoder architecture (HED) that fuses multi-scale information for prediction, thereby enhancing the modeling capability for complex data patterns.

2.4.1 Constructing the 2D Vector Array

Traditional Transformers concatenate data from different dimensions at the same timestep, whereas Crossformer's Dimension-wise Segment Embedding (DSW) embeds sequential timestep data within each dimension through segment-wise processing, dividing the time series into segments of length τ. These segments undergo linear projection and positional embedding to generate a 2D vector array that preserves temporal characteristics, captures cross-dimensional dependencies, and minimizes information loss.

2.4.2 Two-Stage Attention Layer

The computational process of the Two-Stage Attention (TSA) layer can be simplified as follow:

(a) Cross-Time Stage: Multi-head self-attention is applied to each dimension to capture temporal dependencies, with a computational complexity of $O(DL^2)$, where D denotes the number of dimensions and L represents the number of segments.

(b) Cross-Dimension Stage: To reduce computational complexity, a learnable router is employed to first aggregate and then propagate information, thereby lowering the complexity to $O(DL)$.

(c) The overall computational complexity of the Two-Stage Attention (TSA) layer is ultimately.

$$O(DL^2 + DL) = O(DL^2) \tag{14}$$

2.4.3 Hierarchical Encoder-Decoder (HED)

The Hierarchical Encoder-Decoder (HED) model architecture can be divided into an encoder component and a decoder component.

In the encoder component, dependencies across different scales are first captured through the TSA layer and segment merging operations. Except for the first layer, every two adjacent vectors are merged to obtain coarser-scale representations. Ultimately, the encoder employs the TSA layer to model dependencies among these scales. The computational complexity of the encoder is :

$$O(DT^2/L^2seg) \tag{15}$$

where T denotes the total sequence length and $Lseg$ represents the segment length within each dimension.

The decoder component performs prediction at each layer by utilizing the output of the TSA layer as queries (Q), while the encoded vectors across all dimensions from the encoder serve as keys (K) and values (V), thereby establishing encoder-decoder alignment. These representations are multiplied with a learnable weight matrix W to generate the predicted value for the i-th segment. Finally, the predictions from all layers are aggregated through summation to produce the final output.

3. Composition of the proposed model

The nonlinearity, volatility, and complexity of wind power constitute fundamental challenges that critically impact prediction accuracy, posing significant difficulties for wind power forecasting models. To address these challenges, this paper proposes a hybrid forecasting model based on CPO-VMD-WTC Wavelet Convolution-Crossformer. First, the CPO algorithm is employed to optimize the parameters of Variational Mode Decomposition (VMD), decomposing the wind power signal into multiple stationary modal components, thereby mitigating the interference of volatility and chaotic characteristics on model performance. Subsequently, the WTC wavelet convolution performs multi-scale feature extraction on the decomposed modal components to thoroughly uncover latent patterns and complexity within the wind power signal. Finally, the Crossformer model is adopted for adaptive learning and prediction of each modal component, where Crossformer autonomously adjusts its network architecture and weights according to the characteristics of the modal components, substantially enhancing prediction accuracy. In conclusion, the proposed CPO-VMD-WTC Wavelet Convolution-Crossformer hybrid model effectively addresses the nonlinearity and volatility challenges in wind power forecasting, demonstrating outstanding performance in improving both prediction accuracy and stability. The flow chart of CPO-VMD-WTC-Crossrormer model is shown in Fig. 2.

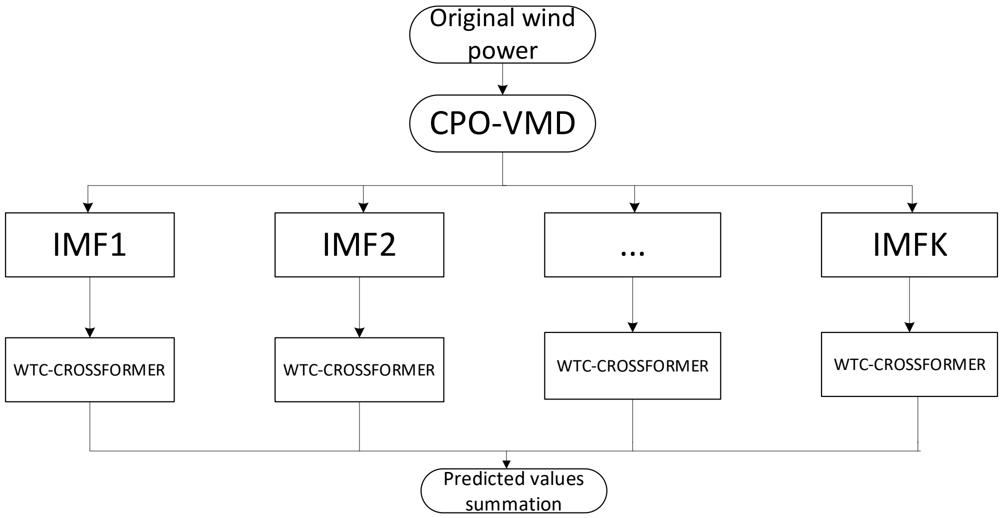

Fig. 2 Flow chart of CPO-VMD-WTC-Crossrormer model.

4. Case Analysis

4.1 Data Preprocessing

The experiment utilizes 35,040 wind power generation data points collected online from an operational wind farm in Hami, Xinjiang, between January 1, 2019 and December 31, 2019. These time-series data were acquired through equidistant sampling at 15-minute intervals, yielding 96 sample points per day. The complete dataset was partitioned proportionally, with 94% allocated for training and 6% reserved for testing. Additionally, the sliding window step size was set to 96. Fig. 3 shows the original wind power.

As evident from the figure above, the data exhibits intense volatility in the short term with pronounced stochastic characteristics. Under such conditions, direct application of the

Crossformer neural network for prediction proves highly challenging, as the disordered fluctuations interfere with the network's ability to capture underlying patterns. Consequently, preprocessing the data using decomposition algorithms becomes essential. This processing stabilizes the data and reveals inherent periodicity, thereby significantly reducing subsequent prediction complexity and laying the foundation for enhanced forecasting accuracy. The experiment comparatively implements VMD, CEEMDAN, EEMD, and EMD algorithms to decompose the signal f The results are shown in Table 1.

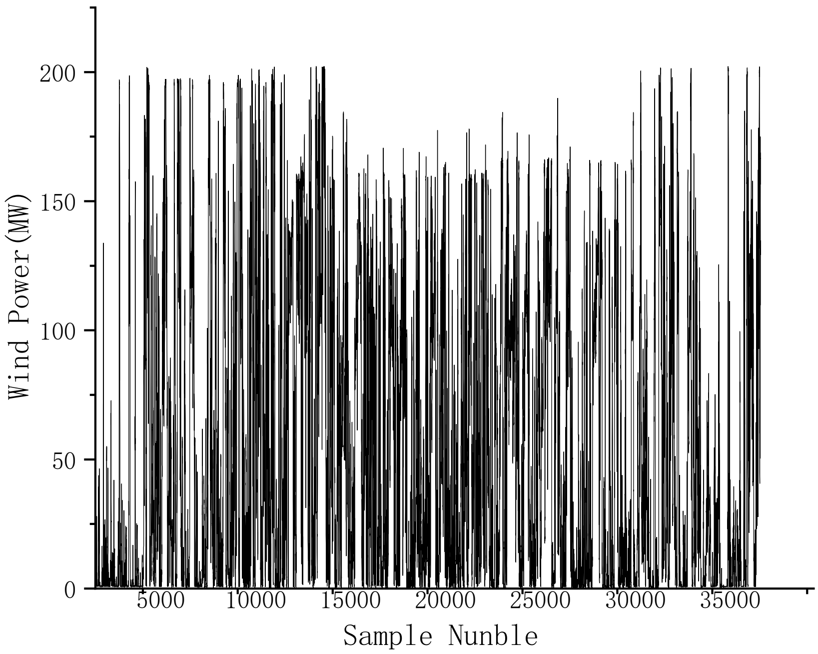

Fig. 3 The original wind power

Table 1. Comparative Results of Decomposition Algorithms

Algorithm	Time/s	MMI	RMSE	SNR	ETCE
VMD	9.8	0.05	4.1	12.6	93.5
EMD	12.3	0.25	8.5	5.2	68.7
EEMD	45.6	0.18	6.2	8.1	75.4
CEEMDAN	38.9	0.12	5.8	9.3	82.1

The experimental results demonstrate that the VMD algorithm outperforms other methods with a superior ETCE comprehensive score of 93.5, exhibiting significant advantages. It achieves exceptional performance in key metrics including decomposition error (RMSE: 4.1MW) and noise resistance (SNR gain: 12.6dB), delivering optimal decomposition and reconstruction of wind power signals. In comparison, CEEMDAN attains a comprehensive score of 82.1, while EEMD's introduction of noise during decomposition increases its error to 6.2MW and reduces its score to 75.4. Although EMD avoids additional noise injection, its decomposition error remains notably high at 8.5MW with a comprehensive score of only 68.7, indicating comparatively weaker overall performance. These findings conclusively validate the superiority of VMD for wind power signal decomposition and robustly justify its selection for related applications.

4.2 CPO-VMD-Based Wind Power Results

Therefore, this study employs the VMD algorithm to decompose wind power data, utilizing the Crested Porcupine Optimizer (CPO) to optimize VMD's penalty coefficient (α) and mode number (K). The optimization parameters are configured with a population size of 10 and maximum iterations of 30, while the search ranges for K and α are set to [5,12] and [500,100000] respectively. The optimal solution yields a minimum fitness value (Ep) of -2.35, corresponding to the best α parameter of 8769 and optimal K value of 5. The time-domain and frequency-domain representations of the CPO-VMD (Crested Porcupine Optimizer-Variational Mode Decomposition) decomposed power signals are shown in Figs. 4,5 The results demonstrate that CPO-VMD accurately decomposes the original power signal into multiple Intrinsic Mode Functions (IMFs): IMF1 and IMF2 represent low-frequency high-amplitude components with smooth trajectories and large magnitudes; IMF3, IMF4, and IMF5 exhibit higher frequencies, lower amplitudes, and frequent fluctuations, precisely extracting key factors influencing power variations and validating the technique's reliability and efficiency in power data processing. The remaining IMF components, despite their diverse frequency characteristics, display discernible patterns that collectively enhance the model's predictive accuracy.

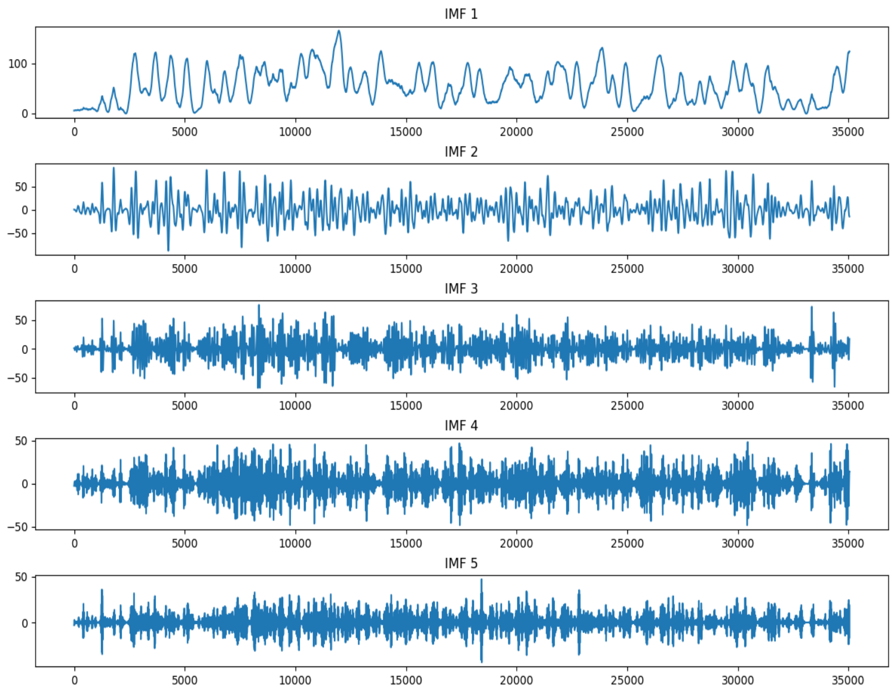

Fig. 4 The time-domain Decomposition results of wind power by CPO-VMD.

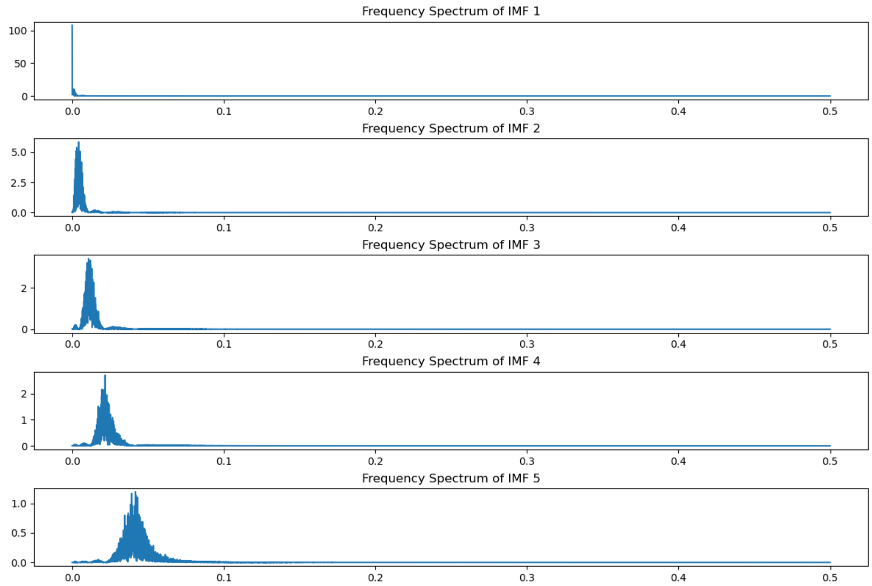

Fig. 5 The frequency-domain Decomposition results of wind power by CPO-VMD.

Fig. 6 shows the comparison results of fitness values when different optimization algorithms are used to optimize VMD parameters. The Firefly Algorithm (FA) achieves its minimum fitness value of -2.29 after 26 iterations, while the Whale Optimization Algorithm (WOA) reaches -2.11 at the 9th iteration, and the Sparrow Search Algorithm (SSA) obtains -2.14 by the 7th iteration. In contrast, the Crested Porcupine Optimizer (CPO) demonstrates superior convergence speed, attaining the globally optimal fitness value of -2.35 as early as the 8th iteration. These results conclusively establish CPO's dominance in this VMD parameter optimization task, evidenced by both its fastest convergence rate and the lowest achieved fitness value.

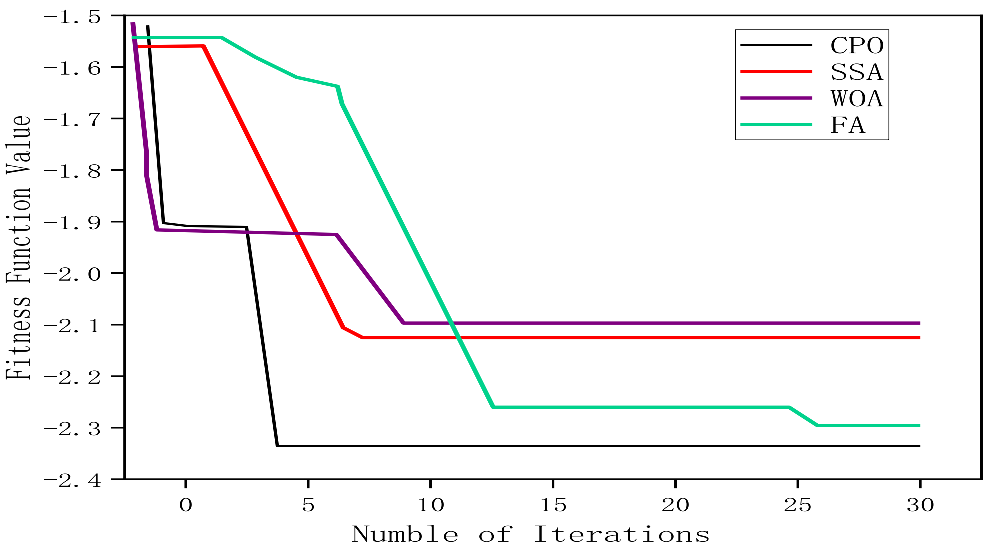

Fig. 6 Variation trend of fitness curve of different optimization algorithms

4.3 Analysis of Predictive Results Across Different Models

In this experiment, the input sequence length of the model is set to 96 with a prediction sequence length of 1, and the training iteration count is fixed at 20. The initial learning rate, which critically influences both convergence speed and prediction accuracy, is determined as 0.0001 through extensive empirical validation. To rigorously verify the model's efficacy, comparative experiments are conducted with Crossformer, NTSformer, PatchTST, TimeNet, and DLinear, alongside ablation studies for the CPO-VMD-WTC-Crossformer framework. All prediction models are objectively evaluated by selecting their parameter configurations through multiple trials to ensure optimal performance. The comparative results are visualized in Figs. 10,11, with corresponding error metrics quantitatively presented in Table 2.

As evidenced by the prediction curves in Fig. 7 and error metrics in Table 2, Crossformer demonstrates superior forecasting performance. Compared to TimeNet, DLinear, PatchTST, and TSMixer models, it achieves reductions of 93.60%, 32.12%, 28.91%, and 10.56% in MSE; 74.38%, 17.20%, 15.23%, and 4.28% in RMSE; and 75.70%, 18.10%, 27.64%, and 17.33% in MAPE respectively. This stems from Crossformer's exceptional cross-scale information fusion capability, which effectively captures wind power characteristics across diverse temporal scales by integrating local details with global trends, thereby accurately resolving complex time-series patterns. Its unique architecture enhances computational efficiency, enabling rapid learning of critical features during large-scale data processing, which significantly outperforms other models. Fig. 8 shows the ablation study results of the CPO-VMD-WTC-Crossformer model, where the prediction curves and Table 2's error metrics confirm that incorporating WTC convolution substantially improves performance over the baseline. Specifically, WTC-Crossformer reduces MSE, RMSE, and MAE by 14.15%, 5.72%, and 6.44% respectively compared to vanilla Crossformer, demonstrating WTC convolution's efficacy in enhancing feature extraction and interpretation to capture pivotal data characteristics more precisely, thereby reducing prediction errors and improving reliability. Furthermore, VMD-WTC-Crossformer achieves additional reductions of 17.28% (MSE), 3.09% (RMSE), and 3.23% (MAE) over WTC-Crossformer, highlighting Variational Mode Decomposition's (VMD) critical role in transforming volatile input data into stabilized modal components for improved feature learning. Notably, CPO-optimized VMD parameters yield higher accuracy than manual VMD configurations, with CPO-VMD-WTC-Crossformer further lowering MSE, RMSE, and MAE by 6.69%, 11.47%, and 5.68% compared to VMD-WTC-Crossformer, proving that automated parameter optimization mitigates under/over-decomposition issues and enhances prediction precision.

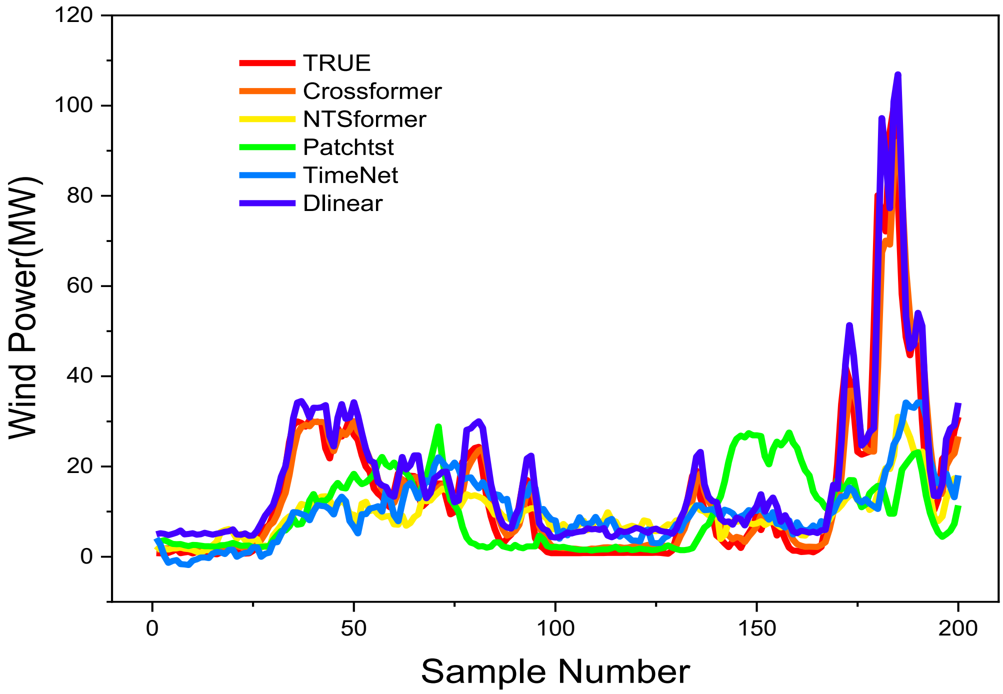

Fig. 7 Comparison of prediction results of single model.

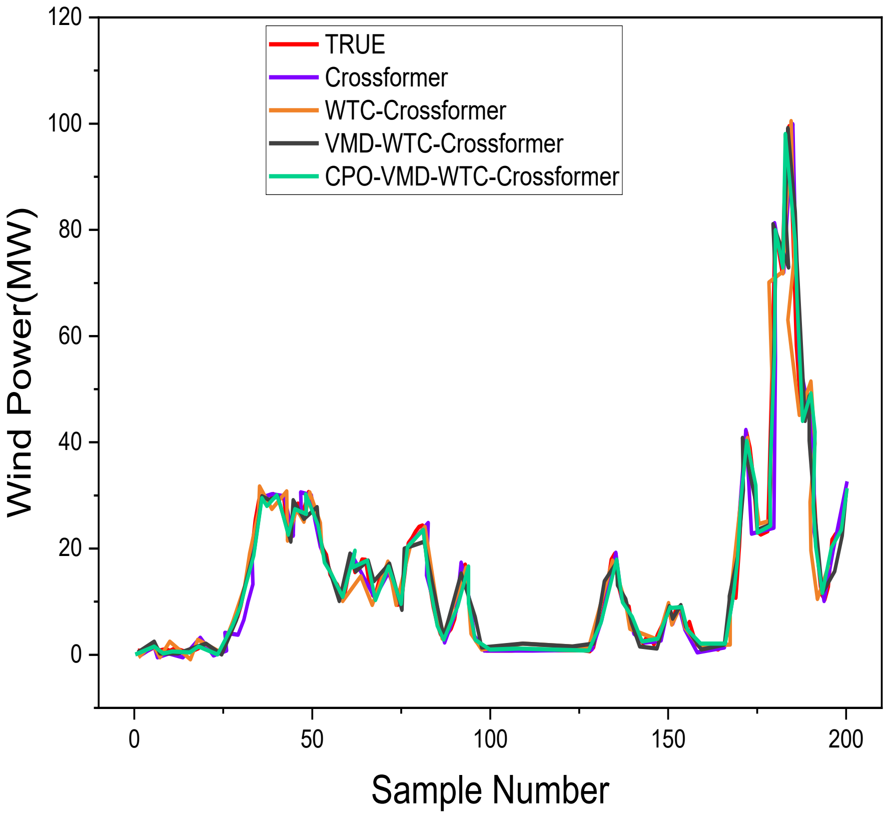

Fig. 8 Comparison of prediction results of ablation experiment

Table 2. Error evaluation of different models.

	MSE	**RMSE**	**MAE**
TimeNet	170.4458	13.1729	19.7925
Dlinear	103.7349	10.1850	5.8733
Patchtst	99.0568	9.9483	6.6480
Tsmixer	78.7246	8.8095	5.5184
Crossformer	70.4156	8.4328	4.8102
WTC-Crossformer	60.4522	7.9506	4.5003
VMD-WTC-Crossformer	50.0048	7.7053	4.3549
CPO-VMD-WTC-Crossformer	46.6587	6.8213	4.1073

5. Conclusion

Accurate wind power forecasting is critical for ensuring grid stability and power supply reliability. This paper proposes the CPO-VMD-WTC-Crossformer model for wind power prediction. Experimental results demonstrate that optimizing VMD parameters with CPO yields superior performance: compared to conventional optimization algorithms, CPO achieves the fastest convergence, reaching the minimum fitness value of -2.35 by the 8th iteration, while effectively escaping local optima and enhancing overall model efficacy. In terms of prediction accuracy, Crossformer outperforms TimeNet, PatchTST, and other benchmarks with MSE, RMSE, and MAE values of 70.4156, 8.4328, and 4.8102 respectively, highlighting its unique advantages in processing wind power time-series data. Furthermore, the integrated CPO-VMD-WTC-Crossformer framework reduces MSE, RMSE, and MAE by 33.74%, 19.09%, and 14.64% compared to standalone Crossformer, proving that the combined preprocessing mechanism significantly improves prediction precision and error reduction. In conclusion, the proposed CPO-VMD-WTC-Crossformer model exhibits exceptional performance in wind power forecasting, combining algorithmic convergence advantages with substantial accuracy improvements, thereby providing a more reliable and efficient solution for the field.

References

[1] Algarni S, Tirth V, Alqahtani T, et al. Contribution of renewable energy sources to the environmental impacts and economic benefits for sustainable development[J]. Sustainable energy technologies and assessments, 2023, 56: 103098.

[2] Wang J, Zhu H, Zhang Y, et al. A novel prediction model for wind power based on improved long short-term memory neural network[J]. Energy, 2023, 265: 126283.

[3] Wei J, Wu X, Yang T, et al. Ultra-short-term forecasting of wind power based on multi-task learning and LSTM[J]. International Journal of Electrical Power & Energy Systems, 2023, 149: 109073.

[4] Wei H, Wang W, Kao X. A novel approach to ultra-short-term wind power prediction based on feature engineering and informer[J]. Energy Reports, 2023, 9: 1236-1250.

[5] Markovics D, Mayer M J. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction[J]. Renewable and Sustainable Energy Reviews, 2022, 161: 112364

[6] Hu S, Xiang Y, Zhang H, et al. Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction[J]. Applied Energy, 2021, 293: 116951

[7] Chen W, Zhou H, Cheng L, et al. Prediction of regional wind power generation using a multi-objective optimized deep learning model with temporal pattern attention[J]. Energy, 2023, 278: 127942.

[8] Tian Z. A state-of-the-art review on wind power deterministic prediction[J]. Wind Engineering, 2021, 45(5): 1374-1392.

[9] Wang Y, Pei L, Li W, et al. Short-term wind power prediction method based on multivariate signal decomposition and RIME optimization algorithm[J]. Expert Systems with Applications, 2025, 259: 125376.

[10] Guo N Z, Shi K Z, Li B, et al. A physics-inspired neural network model for short-term wind power prediction considering wake effects[J]. Energy, 2022, 261: 125208.

[11] Dosdoğru A T, İpek A Ä B. Hybrid boosting algorithms and artificial neural network for wind speed prediction[J]. International Journal of Hydrogen Energy, 2022, 47(3): 1449-1460.

[12] Peng X, Li C, Jia S, et al. A short-term wind power prediction method based on deep learning and multistage ensemble algorithm[J]. Wind energy, 2022, 25(9): 1610-1625.

[13] Alkesaiberi A, Harrou F, Sun Y. Efficient wind power prediction using machine learning methods: A comparative study[J]. Energies, 2022, 15(7): 2327

[14] Lei L, Shao S, Liang L. An evolutionary deep learning model based on EWKM, random forest algorithm, SSA and BiLSTM for building energy consumption prediction[J]. Energy, 2024, 288: 129795.

[15] Guan S, Wang Y, Liu L, et al. Ultra-short-term wind power prediction method combining financial technology feature engineering and XGBoost algorithm[J]. Heliyon, 2023, 9(6)

[16] Garg S, Krishnamurthi R. A CNN encoder decoder LSTM model for sustainable wind power predictive analytics[J]. Sustainable Computing: Informatics and Systems, 2023, 38: 100869.

[17] Farah S, Humaira N, Aneela Z, et al. Short-term multi-hour ahead country-wide wind power prediction for Germany using gated recurrent unit deep learning[J]. Renewable and Sustainable Energy Reviews, 2022, 167: 112700.

[18] Chen H, Wu H, Kan T, et al. Low-carbon economic dispatch of integrated energy system containing electric hydrogen production based on VMD-GRU short-term wind power prediction[J]. International Journal of Electrical Power & Energy Systems, 2023, 154: 109420

[19] Huang L, Li L, Wei X, et al. Short-term prediction of wind power based on BiLSTM–CNN–WGAN-GP[J]. Soft Computing, 2022, 26(20): 10607-10621.

[20] Liu R, Song Y, Yuan C, et al. Gan-based abrupt weather data augmentation for wind turbine power day-ahead predictions[J]. Energies, 2023, 16(21): 7250.

[21] Xiong J, Peng T, Tao Z, et al. A dual-scale deep learning model based on ELM-BiLSTM and improved reptile search algorithm for wind power prediction[J]. Energy, 2023, 266: 126419.

[22] Wang S, Shi J, Yang W, et al. High and low frequency wind power prediction based on Transformer and BiGRU-Attention[J]. Energy, 2024, 288: 129753.

[23] Qu K, Si G, Shan Z, et al. Short-term forecasting for multiple wind farms based on transformer model[J]. Energy Reports, 2022, 8: 483-490.

[24] Wang W, Feng B, Huang G, et al. Conformal asymmetric multi-quantile generative transformer for day-ahead wind power interval prediction[J]. Applied Energy, 2023, 333: 120634

[25] Zhang Y, Yan J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting[C]//The eleventh international conference on learning representations. 2023.

[26] Finder S E, Amoyal R, Treister E, et al. Wavelet convolutions for large receptive fields[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 363-380.

[27] Jiajun H, Chuanjin Y, Yongle L, et al. Ultra-short term wind prediction with wavelet transform, deep belief network and ensemble learning[J]. Energy Conversion and Management, 2020, 205: 112418.

[28] Gao X, Li X, Zhao B, et al. Short-term electricity load forecasting model based on EMD-GRU with feature selection[J]. Energies, 2019, 12(6): 1140.

[29] Ding Y, Chen Z, Zhang H, et al. A short-term wind power prediction model based on CEEMD and WOA-KELM[J]. Renewable Energy, 2022, 189: 188-198.

[30] Abdel-Basset M, Mohamed R, Abouhawwash M. Crested Porcupine Optimizer: A new nature-inspired metaheuristic[J]. Knowledge-Based Systems, 2024, 284: 111257.