



ARTICLE

A Short-Term Wind Power Forecasting Method Based on Adaptive BKA-TCN-BiLSTM Hybrid Model with AP Clustering

Mingxuan Ji¹, Jing Gao^{1,*}, Dantian Zhong¹, Yingqi Xu¹, Shuxiang Yang¹, Zhongxiao Du¹ and Yingming Liu²

¹School of Electrical Engineering, Shenyang Institute of Engineering, Shenyang, 110136, China

²School of Electrical Engineering, Shenyang University of Technology, Shenyang, 110870, China

*Corresponding Author: Jing Gao. Email: gaojing1@sie.edu.cn

Received: 15 October 2025; Accepted: 25 December 2025; Published: 12 July 2026

ABSTRACT: The intermittency of wind power poses severe challenges to the safe and stable operation of power grids, while conventional forecasting models are deficient in prediction accuracy and adaptability to variable weather conditions. To address these issues, this study proposes an adaptive short-term wind power forecasting model integrating affinity propagation (AP) clustering and a black-winged kite algorithm (BKA)-optimized temporal convolutional network-bidirectional long short-term memory (TCN-BiLSTM) hybrid architecture. First, mutual information was employed to screen key meteorological features, and AP clustering categorized historical data into six distinct weather scenarios. A scenario-specific TCN-BiLSTM model was then constructed for each cluster: TCN was utilized to capture multi-scale local temporal features, while BiLSTM modeled global sequence dependencies, with BKA implementing global hyperparameter optimization. Final predictions were generated by invoking the model corresponding to the nearest cluster center. Comparative experiments against baseline models (LSTM, BiLSTM, CNN-LSTM, TCN-BiLSTM) demonstrate that the proposed model achieves remarkable performance gains: normalized root mean square error (nRMSE), normalized mean absolute error (nMAE), and normalized mean square error (nMSE) are reduced by over 1.27%, 1.73%, and 12.09%, respectively, with the coefficient of determination (R^2) reaching 0.9084. This verifies substantial improvements in prediction accuracy and data fitting capability. The scenario-based modeling framework combined with intelligent hyperparameter optimization effectively enhances the model's adaptability to complex weather, confirming its high accuracy and strong generalization for wind power forecasting tasks.

KEYWORDS: TCN-BiLSTM; weather scenarios; BKA optimization algorithm; wind power forecasting

1 Introduction

With the rapid advancement of wind power technology, the capacity of individual wind turbines continues to increase. The number and scale of grid-connected wind farms are also expanding. As a result, the wind power penetration level has significantly increased. However, the inherent randomness and intermittency of wind power have notable impacts on power system stability, power quality, power flow distribution, and dispatch strategies [1]. High-accuracy wind power forecasting is critical for secure and economic grid dispatch, electricity market operation, and optimal control of wind farms [2].

Currently, existing short-term power forecasting mainly relies on models that learn the mapping relationship between meteorological data and power output. In reference [3], the Long Short-Term Memory (LSTM) network was applied to wind power forecasting. However, the prediction accuracy of the single



LSTM model decreases as the length of the time series increases. In reference [4], a Convolutional Neural Network (CNN) was adopted. The CNN extracts complex features from time series through multiple convolution layers. Due to its local connectivity and global parameter sharing properties, the CNN significantly reduces training time and improves forecasting efficiency. In reference [5], heteroscedastic Support Vector Regression (SVR) was used. Unlike conventional SVR models, the heteroscedastic SVR can effectively capture the uncertainty in wind power, leading to improved prediction accuracy. While all above studies train models by learning the meteorological-power mapping, each employs a single architecture for all weather conditions. This lack of adaptability results in poor generalization, especially under complex meteorological scenarios. As a result, significant prediction errors may occur in such cases.

In the deep learning framework for wind power forecasting, hybrid models have become the core technical approach to boost prediction performance. In the deep learning framework for wind power forecasting, hybrid models are the core approach to boost prediction performance. In reference [6], Abdelkader et al. proposed an RNN-LSTM hybrid temporal architecture, which integrates Recurrent Neural Networks (RNN)'s capacity to model local temporal correlations and Long Short-Term Memory Networks (LSTM)'s advantage in capturing long-range temporal dependencies. The model achieved notable accuracy gains in wind power forecasting, verifying the practical validity of hybrid temporal neural networks for non-stationary wind power sequence data. In reference [7], Liu et al. deeply coupled three capabilities—Convolutional Neural Networks (CNN)'s spatial feature extraction, Bidirectional Gated Recurrent Units (BiGRU)'s bidirectional temporal perception, and the Attention mechanism's key feature focusing—to construct the CNN-BiGRU-Attention multi-module hybrid model. They also introduced the BKA algorithm for global adaptive tuning of model hyperparameters, which balances feature extraction efficiency and prediction stability, significantly enhancing short-term wind power forecasting reliability. Nevertheless, the performance merits of the abovementioned hybrid model studies are mostly verified under conventional meteorological conditions. When facing compound complex weather conditions, such models show obvious deficiencies in cross-scenario generalization and anti-interference robustness, which severely restricts their engineering deployment in the complex meteorological scenarios of actual wind farms.

On the other hand, some studies have attempted scenario-based modeling according to different weather scenarios or power output patterns. In reference [8], the k-means clustering algorithm was used to group meteorological variables. A separate neural network was built for each cluster, resulting in improved prediction accuracy. However, the k-means algorithm requires the number of clusters and initial cluster centers to be specified manually. Its clustering performance largely depends on these settings. In reference [9], a method based on weather scenarios was proposed. Different datasets were used to forecast corresponding weather conditions. However, historical meteorological data do not include explicit weather scenario labels. These labels must be manually defined based on prior knowledge. Although the above studies perform scenario-based modeling for different weather conditions and wind power peaks or valleys, the classification methods used are often based on predefined rules or strong assumptions. A data-driven and dynamically adaptive partitioning mechanism is lacking. As a result, the classification outcomes are not objective or precise enough. This limits the ability to fully capture the meteorological-power relationships. Thus, further improvements in forecasting accuracy are constrained [10].

In view of the above limitations, a model architecture that integrates AP clustering with BKA-TCN-BiLSTM is proposed in this study. Experimental results show that the combination of scenario-based modeling and intelligent optimization enhances the model's adaptability to complex weather conditions. This demonstrates that the proposed model achieves high wind power forecasting accuracy and strong generalization capability.

2 Overall Structure of the Prediction Model

The architecture of AP-BKA-TCN-BiLSTM adaptive model can be divided into training module and prediction module, and the overall model construction process is shown in Fig. 1. The following describes the modeling process from two aspects, namely, the steps of model construction and the content of subsequent research:

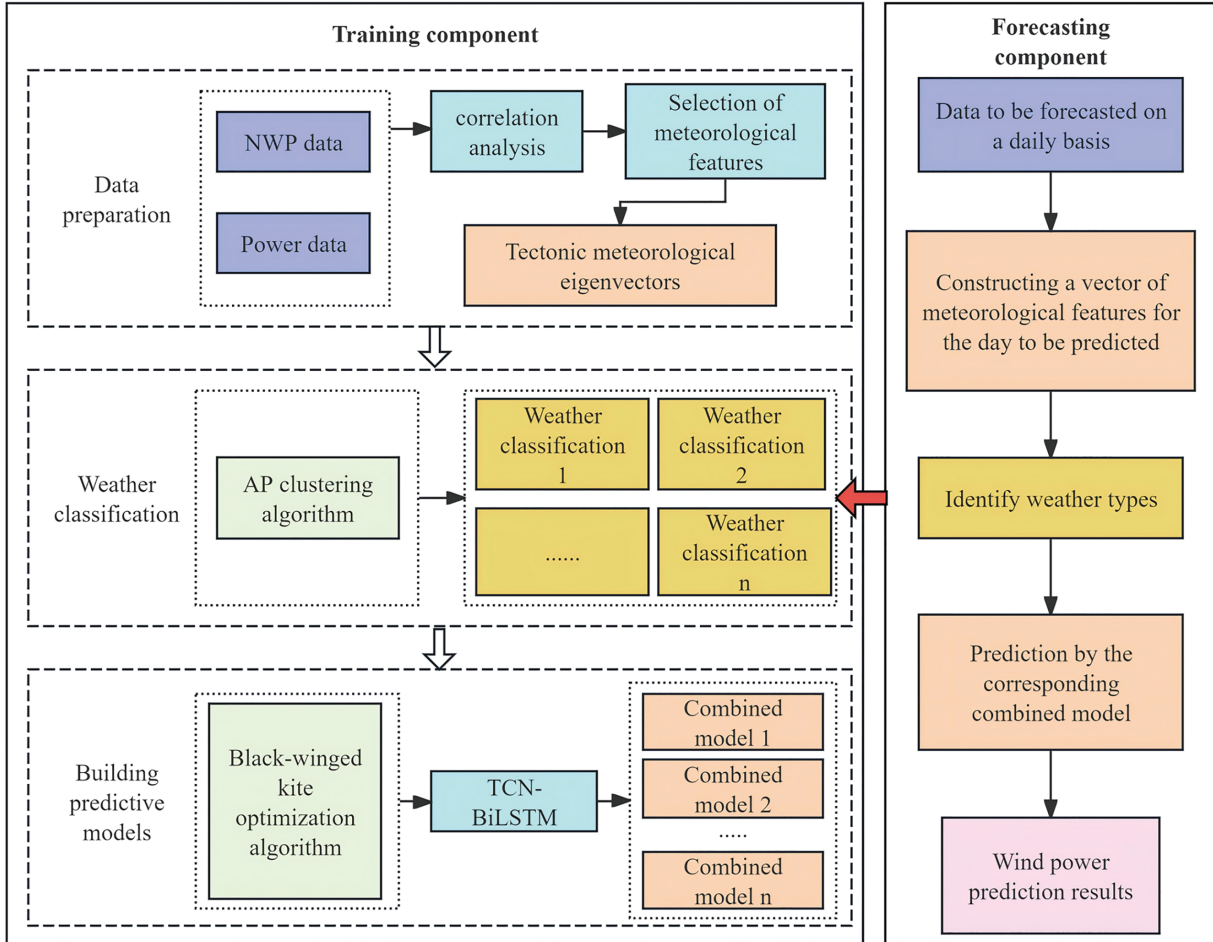


Figure 1: Predictive modeling structure

- (1) In the training phase, the correlation between each meteorological feature and the power sequence is first quantified using mutual information values. Significant features are then selected to conduct scenario clustering analysis. After high-correlation features are obtained, feature reconstruction is conducted to optimize the data structure.
- (2) The AP clustering algorithm is applied to the reconstructed feature dataset to perform unsupervised classification. Several weather scenario types are generated in this way. For each scenario type, a BKA-TCN-BiLSTM hybrid model is constructed and trained separately. The power sequence and feature sequence are merged into a feature matrix, which serves as the input to the TCN layer. Benefiting from the strong nonlinear feature extraction capability of TCN, potential correlations between wind power and multi-dimensional features are captured. The extracted features are then input to the BiLSTM network, enhancing the model's ability to learn complex temporal dependencies. On this basis, the

BKA optimization algorithm is introduced to globally optimize the hyperparameters of the TCN-BiLSTM model. Finally, an optimal forecasting model is established for each weather scenario.

- (3) In the prediction phase, the feature space distance between the sample to be predicted and each cluster center is calculated first. The corresponding weather scenario type is then determined. The pre-trained model for that specific scenario is subsequently called to perform power forecasting. This enables precise and refined prediction output.

3 Theory of Correlation Algorithm

3.1 AP Clustering Algorithm

The AP clustering algorithm is a representative-point-based clustering method [11]. The concept of data point preference is used to measure similarity. Negative Euclidean distances are used as similarity metrics between data points. A matrix is constructed based on self-similarity and mutual similarity. Optimal cluster centers are determined through iterative refinement. Data points are then clustered according to pairwise similarity, ensuring clustering quality and determining the number of clusters.

AP clustering algorithm is based on the data similarity between Numerical Weather Prediction (NWP) data.

$$S(i, k) = -\|x_i - x_k\|^2 \quad (1)$$

where: $S(i, k)$ is the negative Euclidean distance, which represents the mutual correlation between NWPs on each day; x_i is the NWP time series on day i ; x_k is the NWP time series on day k . The AP clustering algorithm is based on the data similarity (similarity) between NWPs.

The two mechanisms of responsiveness (R) and availability (A) are also applied to realize the selection of the representative weather scenario, which in turn realizes the categorization of other weather scenarios under the representative weather scenario.

$$R(i, k) = \begin{cases} S(i, k) - \max_{j \neq k} \{A(i, j) + S(i, j)\}, & i \neq k \\ S(i, k) - \max_{j \neq k} \{S(i, j)\}, & i = k \end{cases} \quad (2)$$

$$A(i, k) = \begin{cases} \min \{0, R(k, k) + \sum_{j \neq i, j \neq k} \max \{0, R(j, k)\}\}, & i \neq k \\ \sum_{j \neq k} \max \{0, R(j, k)\}, & i = k \end{cases} \quad (3)$$

$$A(i, k) = (1 - \lambda) \times A(i, k) + \lambda \times A(i, k)_{\text{old}} \quad (4)$$

where: $R(i, k)$, $R(i, k)_{\text{old}}$ is the degree of attraction of the current, last campaign clustering representative weather scenario k to other weather scenario i ; λ is the damping factor; $A(i, k)$, $A(i, k)_{\text{old}}$ is the degree to which the current, last campaign clustering representative weather scenario k is supported by other weather scenario i as a center of clustering. the larger the $R(i, k) + A(i, k)$, the greater the likelihood of candidate data point k as a clustering center. The transfer process is shown in Fig. 2.

In terms of determining the number of clusters, this study selects the final number of clusters based on the principle of optimal clustering quality. The clustering quality evaluation adopts the internal evaluation index system, which specifically contains the Davies-Bouldin (DB) index and the correction index based on the Euclidean morphological distance, and its mathematical expression is as follows:

(1) DB index

The DB indicator can comprehensively take into account the degree of agglomeration within clusters and the degree of separation between clusters, and its calculation formula is as follows:

$$I_{DB} = \frac{1}{K} \sum_{k=1}^K R_k \quad (5)$$

$$R_k = \max_{k \neq j} \frac{d(X_k) + d(X_j)}{d(C_k, C_j)} \quad (6)$$

where $d(X_k)$ and $d(X_j)$ are the average Euclidean distances between sample points and sample centers in cluster C_k and cluster C_j , respectively, indicating the degree of aggregation between samples within a cluster; $d(C_k, C_j)$ is the average Euclidean distance between sample centers in cluster C_k and cluster C_j , indicating the degree of separation between two different classes of clusters.

(2) Modified MDB (modified DB, MDB) metric based on Euclidean morphological distance

$$I_{MDB} = \frac{1}{K} \sum_{k=1}^K \max_{k \neq j} \frac{d_{emd}(X_k) + d_{emd}(X_j)}{d_{emd}(C_k, C_j)} \quad (7)$$

where $d_{emd}(X_k)$ and $d_{emd}(X_j)$ are the average Euclidean morphological distances between the sample points and the sample centers in clusters C_k and C_j ; and $d_{emd}(C_k, C_j)$ is the average Euclidean morphological distance between the sample centers in clusters C_k and C_j . IDB and IMDB both combine the degree of intraclass clustering and the degree of interclass separation between any two clusters as a ratio, so lower values represent higher clustering quality. When choosing the optimal number of clusters, we generally look for the point of its minimal value.

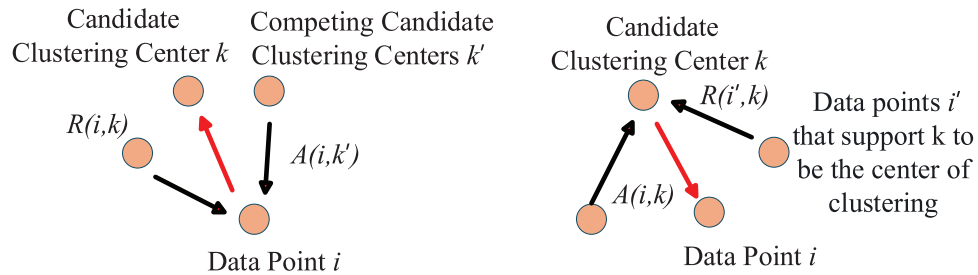


Figure 2: AP clustering algorithm delivery process

Based on the historical sample set, meteorological conditions can be clustered into several typical scenarios, but the prediction needs to determine the weather scenario to which the moment to be predicted belongs, so as to use the power prediction model corresponding to the weather scenario to which the moment belongs to make the prediction. Let AP clustering generate n weather scenarios, the clustering center of each scenario is $c_i = [c_{i1}, c_{i2}, \dots, c_{im}]$ ($i = 1, 2, \dots, n$), m is the number of weather feature dimensions, stored as a matrix $C \in R^{n \times m}$. In this paper, negative Euclidean distance is used to calculate the similarity between the new samples and the clustering centers, and the mathematical expression is as follows.

$$s(x, c_i) = - \sum_{k=1}^m (x_k - c_{ik})^2 \quad (8)$$

where $s(x, c_i)$ is the negative Euclidean distance between the new sample and the clustering center, x is the feature vector of the new sample, c_i is the clustering center vector of the i th scenario, x_k is the k th original eigenvalue of the new sample, and c_{ik} is the k th eigenvalue of the i th clustering center ($k = 1, 2, \dots, m$).

The scene corresponding to the clustering center with the largest similarity is selected and the mathematical expression is as follows:

$$i^* = \arg \max_{i=1, \dots, n} s(x, c_i) \quad (9)$$

where i^* is the index of the determined target scene category.

3.2 Black-Winged Kite Optimization Algorithm

The Black-Winged kite optimization algorithm [12] is a nature-inspired population intelligence optimization algorithm proposed in 2024, which is inspired by the fact that Black-Winged kites exhibit highly adaptive and intelligent behaviors during attacks and migrations. The mathematical model is briefly described as follows.

(1) Population initialization

In BKA, creating a set of random solutions is the first step to initialize the population. The following matrix is used to represent the location of each Black-Winged kite (BK):

$$BK = \begin{bmatrix} BK_{1,1} & BK_{1,2} & \dots & \dots & BK_{1,\text{dim}} \\ BK_{2,1} & BK_{2,2} & \dots & \dots & BK_{2,\text{dim}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ BK_{m,1} & BK_{m,2} & \dots & \dots & BK_{m,\text{dim}} \end{bmatrix} \quad (10)$$

where m represents the number of potential solutions; dim is the size of the given problem dimension; and $BK_{i,j}$ is the j th dimension of the i th Black-Winged kite. Each Black-Winged kite is assigned a position of:

$$X_i = BK_{lb} + \text{rand}(BK_{ub} - BK_{lb}) \quad (11)$$

where i is an integer between 1 and m ; BK_{lb} and BK_{ub} denote the lower and upper bounds of the i th Black-Winged kite in the j th dimension, respectively; and rand is a random value between $[0, 1]$.

(2) Attack behavior

The two attack strategies of Black-Winged kites are hovering in the air waiting for an attack and hovering in the air looking for prey. These strategies include different attack behaviors used for global exploration and search. The following is the mathematical model of Black-Winged kite attack behavior:

$$n = 0.05 \times e^{-2 \times \left(\frac{t}{T}\right)^2} \quad (12)$$

$$y_{t+1}^{i,j} = \begin{cases} y_t^{i,j} + n(1 + \sin(r)) \times y_t^{i,j}, & p < r \\ y_t^{i,j} + n \times (2r - 1) \times y_t^{i,j}, & \text{else} \end{cases} \quad (13)$$

where $y_t^{i,j}$ and $y_{t+1}^{i,j}$ denote the position of the i th Black-Winged kite in the j th dimension in the t th and $t + 1$ -th iteration, respectively; r is a random number between 0 and 1; p is a constant value of 0.9; T is the total number of iterations; t is the number of iterations that have been completed by so far; and n is a balance parameter.

(3) Migration behavior

The migration behavior is led by a leader. If the current population's fitness value is less than that of a random population, the leader abandons its leadership and joins the migrating group. In contrast, if the current population's fitness value is higher, the leader guides the group to the destination. This strategy allows for the dynamic selection of high-quality leaders, ensuring successful migration. The mathematical model of the Black-Winged kite's migration behavior is as follows:

$$m = 2 \times \sin(r + \pi/2) \quad (14)$$

$$y_{t+1}^{i,j} = \begin{cases} y_t^{i,j} + C(0,1) \times (y_t^{i,j} - L_t^j), & F_i < F_{ri} \\ y_t^{i,j} + C(0,1) \times (L_t^j - m \times y_t^{i,j}), & \text{else} \end{cases} \quad (15)$$

where, L_t^j is the leader of the Black-Winged kite in dimension j in the t th iteration so far; $y_t^{i,j}$ and $y_{t+1}^{i,j}$ are the position of the i th Black-Winged kite in dimension j in the t th and $t+1$ st iteration, respectively; F_i is the current position in dimension j obtained by any Black-Winged kite in the t th iteration; F_{ri} is the fitness value of the random position in dimension j obtained by any Black-Winged kite in the t th iteration; $C(0,1)$ is the Cauchy mutation.

3.3 BKA-TCN-BiLSTM Model

The TCN preserves the spatial feature extraction capability of CNN [13,14]. The architecture of the TCN is improved to enhance the processing of sequential information. Gradient vanishing and gradient explosion are effectively avoided by the TCN. Low memory usage, stable gradients, high parallelism, and flexible receptive field are achieved [15,16]. The BiLSTM extends the unidirectional processing mechanism of LSTM [17,18]. A bidirectional structure, consisting of forward and backward layers, is integrated. Bidirectional analysis of temporal data is enabled. Both past and future information in wind power data are considered. The dynamic characteristics of the wind power sequence are captured more comprehensively. Forecasting performance of the model is thereby improved.

The TCN-BiLSTM model architecture comprises an input layer, a TCN layer, a BiLSTM layer, and an output layer [19–21]. The core innovation of the structure lies in the integration of TCN and BiLSTM to form a hierarchical feature processing and dynamic modeling mechanism [22–24]. Deep local features and long-range dependencies are efficiently extracted by the TCN layer through dilated causal convolution and residual connections. Advantages include fast parallel computation, large receptive field, and stable training. These characteristics address the limitations of traditional recurrent neural networks in handling very long sequences. The BiLSTM layer captures complex temporal dynamics and strong nonlinear mapping relationships based on the feature sequences output by the TCN layer. This is achieved through a bidirectional gating mechanism. Temporal fluctuations in wind power can be accurately modeled as a result. A decoupling and deep collaboration between feature extraction and temporal modeling is achieved through the combination of these two components. The TCN serves as a powerful feature extractor, providing high-dimensional and robust feature representations for the BiLSTM. Based on these features, the BiLSTM performs fine-grained temporal learning. The inherent limitations of single models are thereby avoided.

The hierarchical fusion architecture of TCN-BiLSTM is especially suitable for the key needs of wind power forecasting: TCN is good at dealing with the long-period trend and spatial correlation of meteorological factors affecting power, and BiLSTM is good at modeling short-term dramatic fluctuations and dynamic response of power itself. The two complement each other's strengths and work together to tackle the strong non-stationarity, significant randomness, and complex spatiotemporal dependencies of wind power data, providing a solid algorithmic foundation for improving prediction accuracy. The output layer ultimately

maps the dynamic states learned by BiLSTM into power prediction values, and the model structure is shown in Fig. 3.

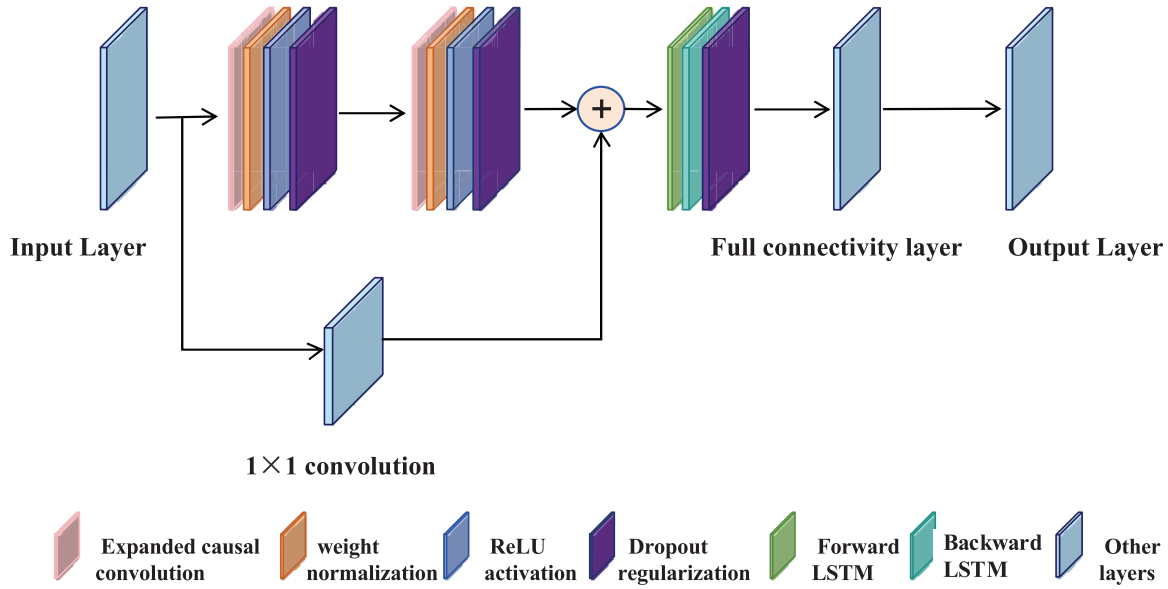


Figure 3: TCN-BiLSTM framework

In this paper, BKA optimization algorithm is used to find the optimization of TCN-BiLSTM parameters. The unique group collaboration and dynamic evolution mechanism of the BKA algorithm is utilized to achieve efficient global exploration of model parameter combinations, in order to solve the problem that traditional heuristic parameter tuning methods are prone to fall into the local optimum in multivariate optimization scenarios, and then to enhance the ability of the TCN-BiLSTM model to capture the complex features of the wind power sequence. The specific steps are as follows, and the flow chart is shown in Fig. 4.

4 Experimental Results and Discussion

4.1 Data Description

The data source is a wind farm in Northeast China with an installed capacity of 49.5 MW, and the actual power and NWP data of the whole year of 2023 and January to April 2024 are used as the data samples for the model, with a time resolution of 15 min and 96 daily sampling points, and the data are divided into training set, validation set and test set according to the ratio of 8:1:1. In order to reduce the arithmetic pressure and improve the efficiency of model training, so when conducting experiments, we firstly use the min-max normalization method to eliminate the data magnitude, normalize the original data, and its mathematical expression is as follows.

$$P' = \frac{P - P_{\min}}{P_{\max} - P_{\min}} \quad (16)$$

where: P' is the normalized data, P is the original wind power data, $P_{\max}$ and $P_{\min}$ are the maximum and minimum values of the wind power series data, respectively.

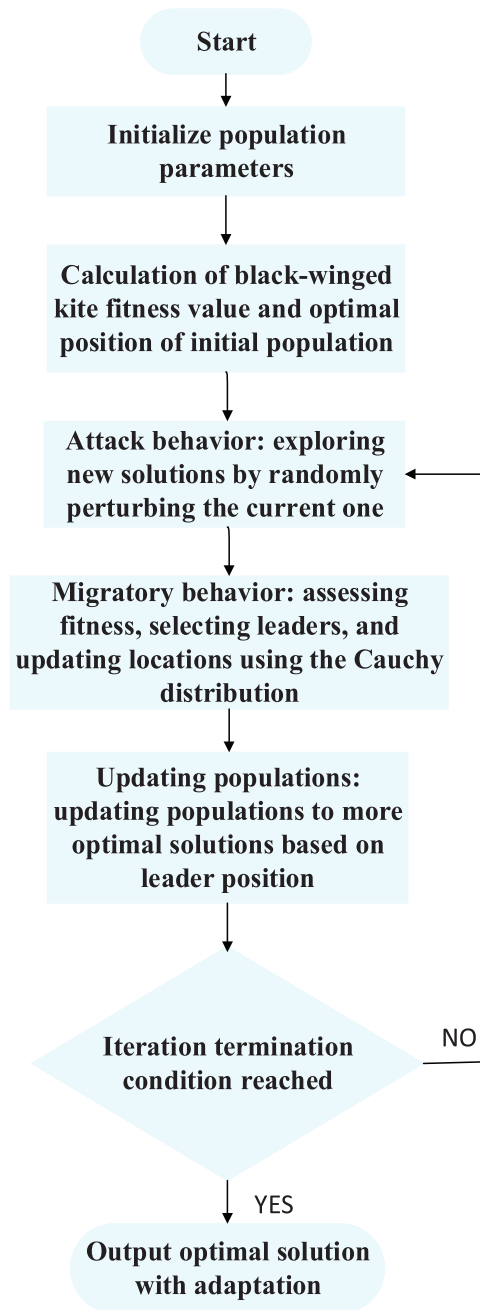


Figure 4: Flow of BKA optimization algorithm

4.2 Evaluation Metrics

In the prediction task, the evaluation index in regression analysis is usually used to measure the degree of deviation of the model prediction results from the actual values. In this study, the normalized root mean square error ($nRMSE$), the normalized mean absolute error ($nMAE$), the normalized mean square error ($nMSE$) and the coefficient of determination (R^2) are selected as the core indexes for model performance evaluation, and its mathematical expression is as follows:

$$nMSE = \frac{1}{n} \sum_{i=1}^n \left(\frac{P(i) - \hat{P}(i)}{C} \right)^2 \quad (17)$$

$$nRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{P(i) - \hat{P}(i)}{C} \right)^2} \quad (18)$$

$$nMAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{P(i) - \hat{P}(i)}{C} \right| \quad (19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{P}(i) - P(i))^2}{\sum_{i=1}^n \left(\hat{P}(i) - \frac{1}{n} \sum_{i=1}^n P(i) \right)^2} \quad (20)$$

where $\hat{P}(i)$ is the predicted value of wind power; $P(i)$ is the real value of wind power; C is the installed capacity of wind farm; n is the number of test samples.

4.3 Parameter Setting

To enhance the predictive performance and learning capability of the models, the key parameters of the TCN and BiLSTM architectures were systematically configured within predefined ranges. The primary parameter settings for the TCN and BiLSTM models are summarized in [Tables 1](#) and [2](#), respectively.

Table 1: Parameters of the TCN model

Model	Num filters	Filter size	Num blocks	Dropout factor	Batch size	Epoch	Optimizer	Activation
TCN	[10, 64]	[1, 15]	[1, 2]	[0.01, 0.5]	50	100	Adam	ReLU

Table 2: Parameters of the BiLSTM model

Model	Hidden layer	LSTM layers	Learning rate	Dropout factor	Batch size	Epoch	Optimizer	Activation
BiLSTM	[16, 64]	2	0.001	[0.01, 0.5]	50	100	Adam	ReLU

4.4 Hyperparameter Sensitivity Analysis of BKA

To systematically evaluate the parameter sensitivity of the BKA optimization algorithm, single-factor control experiments were conducted with fixed model architecture. Three critical parameters—population size, maximum iteration count, and dynamic inertia weight—were analyzed based on predictive performance metrics. Optimal parameter configurations were determined through trade-off analysis between comprehensive performance and computational efficiency.

(1) Population Size Impact

With maximum iterations fixed at 30 and dynamic inertia weight at 0.85, the influence of population size on model performance was evaluated. The results are presented in [Table 3](#).

Table 3: Performance under varying population sizes

Population size	RMSE (MW)	MAE (MW)	Average convergence generation	Time per iteration (s)
5	9.32	7.15	28	1.2
10	8.76	6.83	25	1.6
20	8.65	6.75	23	3.5
30	8.59	6.69	22	5.1

When population size increased from 5 to 10, RMSE decreased by 7.5% ($p < 0.05$), indicating that moderate population expansion enhances search diversity. However, further increases to 20 or 30 yielded marginal improvements (1.7% enhancement) with near-linear growth in computational cost. Considering accuracy-efficiency balance, a population size of 10 was selected.

(2) Maximum Iteration Count Impact

Maintaining population size at 10 and inertia weight at 0.85, varying iteration counts were tested. The results are presented in [Table 4](#).

Table 4: Performance under different maximum iteration limits

Maximum iterations	RMSE (MW)	MAE (MW)	Proportion of early convergence
10	9.67	7.41	80%
20	9.01	7.02	60%
30	8.76	6.83	40%
40	8.73	6.81	35%

RMSE improved by 9.8% when iterations increased from 10 to 30, demonstrating that extended iterations facilitate convergence to optimal solutions. Beyond 30 iterations, however, performance gains became negligible (0.3% improvement) with persistent premature convergence. Consequently, 30 iterations were determined sufficient to ensure convergence while avoiding redundant computations.

(3) Dynamic Inertia Weight Impact

The inertia weight directly affects the search step size and convergence speed. The performance of different weight values was tested under the conditions of a population size of 10 and a maximum number of iterations of 30. The results are presented in [Table 5](#).

Table 5: Performance with different dynamic inertia weight strategies

Dynamic inertia weight	RMSE (MW)	MAE (MW)	Convergence stability (Standard deviation of RMSE)
0.65	9.12	7.21	0.38
0.75	8.91	7.05	0.32
0.85	8.76	6.83	0.27
0.95	8.89	6.98	0.30

When the inertia weight was set to 0.85, the root mean square error (RMSE) reached the minimum value with the optimal stability, indicating that this value achieved a well-balanced trade-off between global exploration and local exploitation. Excessively low weights would lead to slow convergence, while excessively high weights tend to induce oscillation and thus compromise the stability of the algorithm.

The parameters of the BKA algorithm and the optimized hyperparameters of the TCN-BiLSTM model are presented in [Tables 6](#) and [7](#), respectively.

Table 6: Parameters of the BKA algorithm

BKA	Population size	Maximum iterations	Dynamic inertia weight
	10	30	0.9

Table 7: Optimized model hyperparameter values

Model	Num filters	Filter size	Num blocks	Dropout factor
TCN-BiLSTM	45	7	1	0.35814

4.5 Comparison of Algorithmic Optimization Results

In order to evaluate the performance of the proposed BKA algorithm, we compared it with Genetic Algorithm (GA), Sparrow Search Algorithm (SSA), and Particle Swarm Optimization (PSO) in the context of optimizing hyperparameters for the TCN-BiLSTM model. The maximum number of iterations for each optimization algorithm was set to 50, with a population size of 10. [Fig. 5](#) illustrates the fitness changes of these algorithms during the training process. It is evident from the figure that the BKA optimization algorithm exhibits a significantly faster convergence rate and requires fewer iterations compared to the other three optimization algorithms.

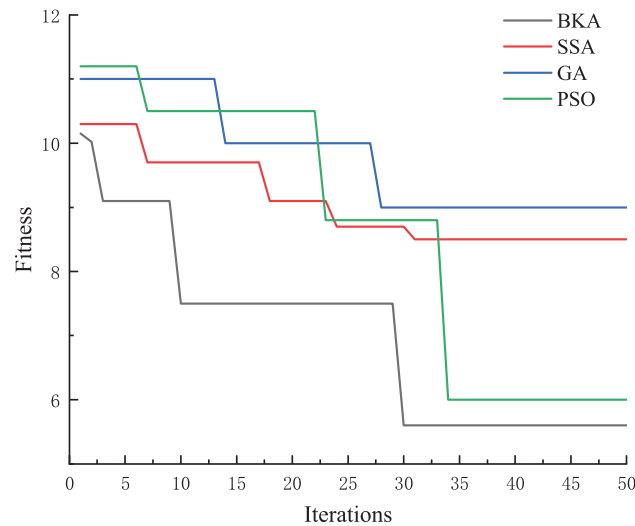


Figure 5: Optimization algorithm optimization comparison

4.6 Correlation Analysis

In this paper, the mutual information value is used to calculate the correlation between each observation feature and power, and the observation features with strong correlation with power are extracted for scenario clustering.

The mutual information of two discrete random variables X and Y is defined as shown in Eq. (21), where $p(x, y)$ is the joint probability distribution function of X and Y , while $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y , respectively. In the case of continuous random variables, $p(x, y)$ is the joint probability density function of X and Y , while $p(x)$ and $p(y)$ are the marginal probability density functions of X and Y , respectively, as shown in Eqs. (21) and (22).

$$MI(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x) p(y)} \right) \quad (21)$$

$$MI(X; Y) = \int_Y \int_X p(x, y) \log \left(\frac{p(x, y)}{p(x) p(y)} \right) dx dy \quad (22)$$

In this case, each variable is continuous, so kernel density estimation is first used for $p(x, y)$, $p(x)$ and $p(y)$, and then their mutual information is calculated using Eq. (21). The mutual information value between each feature is calculated and then the mutual information value between NWP features and wind power can be obtained as shown in Fig. 6. So 70 m wind speed, 50 m wind speed, 70 m wind direction, 50 m wind direction and temperature are selected as model input features.

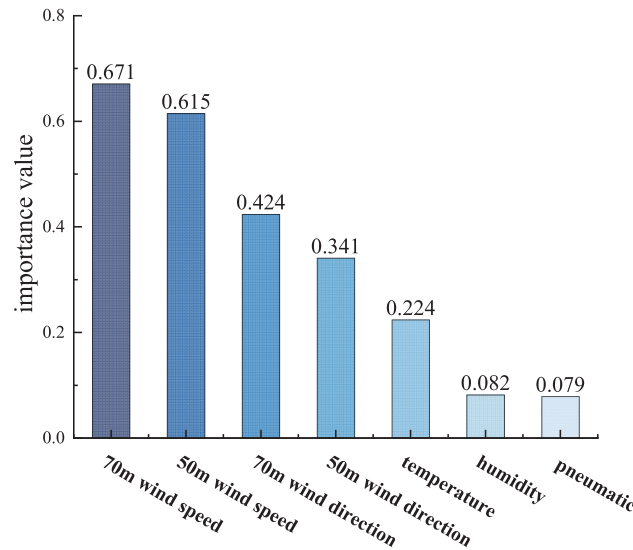


Figure 6: Mutual information values between NWP features and wind power

4.7 Clustering Results

This study adopts daily-scale feature extraction, which primarily considers that the daily operation patterns of wind farms exhibit distinct periodic characteristics, and meteorological conditions exhibit relatively stable patterns at the daily scale. Furthermore, daily-level features align with day-ahead power grid dispatching plans in terms of time scale, which holds practical engineering significance. Thus, this study employs daily-level time granularity for feature extraction and scenario clustering. Specifically, for each day, we extract wind speed at 50 m/70 m, wind direction, and temperature data at 15-min intervals, forming a

5×96 -dimensional temporal feature matrix. To reduce dimensionality while preserving temporal patterns, we calculate the daily mean, standard deviation, maximum value, minimum value, and rate of change for each meteorological variable, resulting in a 25-dimensional feature vector. After clustering, the rolling iteration method is adopted. The samples of each scenario are concatenated in chronological order. When input into the TCN-BiLSTM model, the power and features of the first 16 time steps are used as inputs, while the power of the subsequent 2 time steps is designated as outputs.

The weather scenarios were clustered using AP clustering algorithm with Euclidean distance with morphological features. The weight coefficients α and β for the two similarity measures were computed via the entropy weighting method, yielding values of 0.6832 and 0.5275, respectively. The damping coefficient was set to 0.95, and the optimal number of clusters was determined by tuning the reference parameter p . The clustering results were summarized using the DB metric and the MDB metric. The optimal number of clusters is selected using DB metrics and MDB metrics. Fig. 6 shows the relationship between number of clusters and clustering metrics.

As shown in Fig. 7, when the number of clusters is 6, both DB and MDB metrics have minimum values, so the weather scenarios are categorized into 6 classes. Cluster centers for each weather scenario are listed in Table 8. From the table, weather scenarios 4 and 6 are dominated by moderately warm and humid air masses. weather scenario 4 shows southeast-south gentle breezes, while weather scenario 6 exhibits southwest-west moderate winds. These correspond to transitional weather during spring and autumn. weather scenarios 1 and 5 feature low-temperature wind fields. weather scenario 1 presents northwest-west moderate winds, indicative of stable wind conditions following a cold front passage. weather scenario 5 is characterized by northwest-west strong winds and lower temperatures, corresponding to winter cold wave conditions. weather scenarios 2 and 3 represent low-temperature weak-wind scenarios. weather scenario 2 displays northeast calm weak winds, while weather scenario 3 shows very weak north winds at low temperatures, with the atmosphere tending towards stability. Vertical wind direction consistency is high across all types. Wind shear characteristics (notably significant in weather scenario 3) reflect different meteorological dynamic mechanisms. This provides crucial evidence for scenario-based modeling.

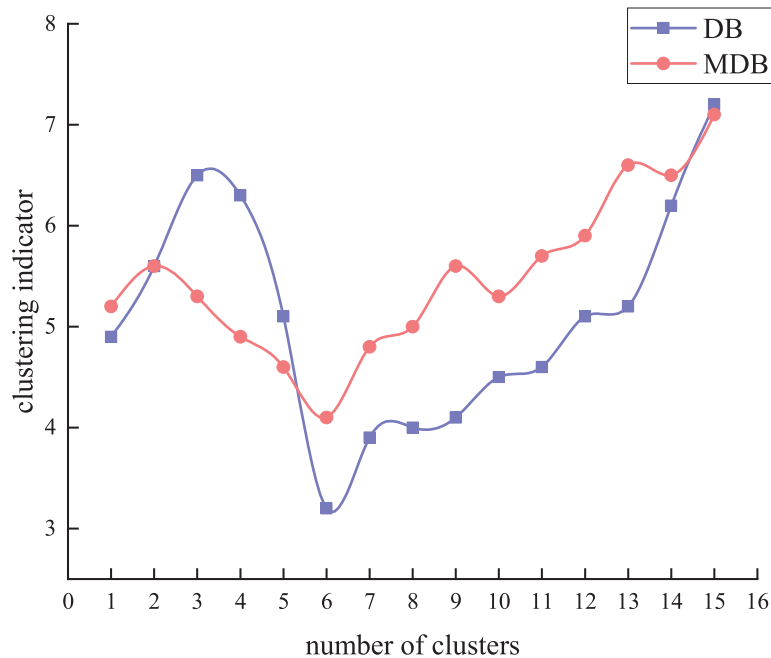
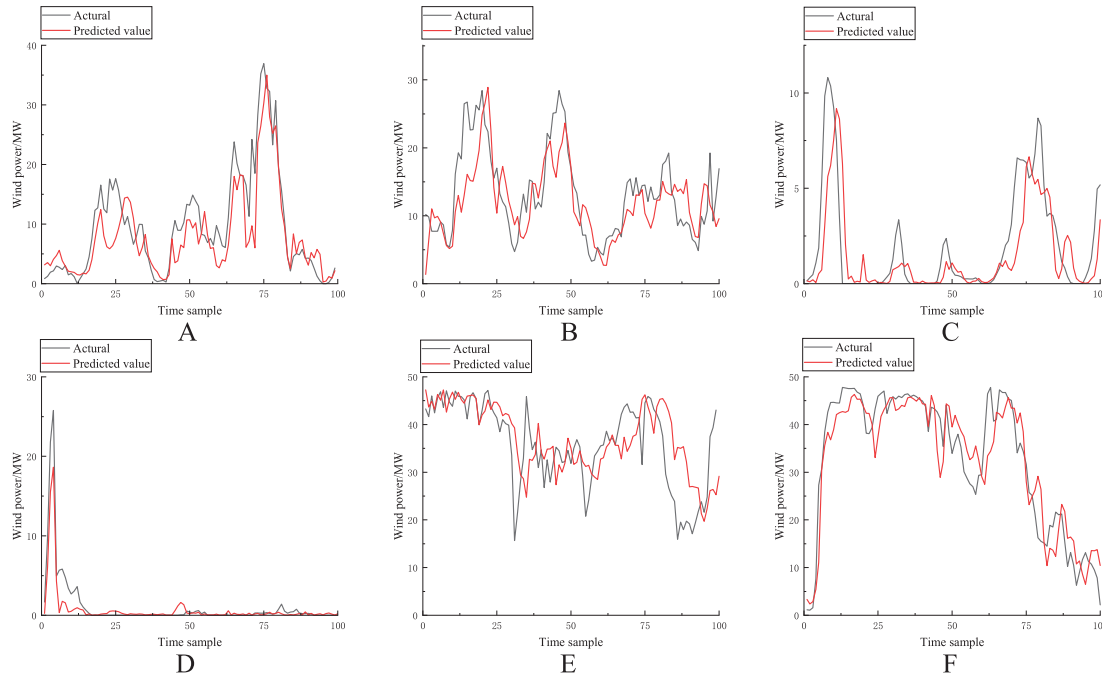


Figure 7: Relationship between number of clusters and evaluation metrics

Table 8: Cluster centers for each weather scenario

	50 m wind direction (°)	70 m wind direction (°)	50 m wind speed (m/s)	70 m wind speed (m/s)	Temp (°C)
Weather scenario 1	278.93	281.85	7.43	7.88	4.99
Weather scenario 2	25.34	28.13	3.02	3.25	4.77
Weather scenario 3	353.11	5.31	4.21	4.36	3.94
Weather scenario 4	156.56	163.4	4.66	4.78	12.86
Weather scenario 5	319.41	322.94	7.88	8.18	−2.15
Weather scenario 6	222.82	227.29	5.03	5.41	10.23

In Fig. 8, six subplots illustrate wind power forecasting performance under different weather scenarios: Subplot A and F correspond to moderately warm and humid airflow scenarios (southeast/southwest winds, moderate-to-low wind speeds). Prediction values closely match actual values, accurately capturing smooth fluctuations and step changes, with good detail matching. Subplots B, E, and C correspond to low-temperature wind fields. In subplot B and E, representing northwest-west/moderate-strong wind scenarios (winter cold fronts/cold waves), predicted curves closely follow the intense power oscillation trends. In subplot C, for cold wave strong wind scenarios, high-power peaks are effectively tracked, demonstrating adaptability to extreme meteorological conditions. In subplot D, concerning a low-temperature weak-wind stable scenario (northeast/very weak north winds), predicted values align with low-power steady states and occasional disturbance trends. While slight deviations exist in peak details, the basic fitting remains reliable. Overall, differentiation in prediction performance across subplots verifies the precise adaptation of scenario-based modeling to various meteorological types. This visually demonstrates the model's superiority in trend capture and detail depiction.

**Figure 8:** Comparison of wind power forecasting results under different weather scenarios

4.8 Analysis of Prediction Results

In order to intuitively and accurately evaluate model performance and prediction accuracy, comparison experiments are conducted using models LSTM, SVM, BiLSTM, CNN-LSTM, TCN-BiLSTM vs. the proposed model. Furthermore, to mitigate model training randomness, enhance robustness and generalization ability, multiple runs are executed, and average values are obtained for each model's evaluation metrics, as shown in Table 9 and Fig. 9. Fig. 10 presents scatter plots of wind power forecasting values vs. actual values for all models.

Table 9: Evaluation index of different model

Model	R ²	nMAE ($\times 10^{-2}$)	nMSE ($\times 10^{-2}$)	nRMSE ($\times 10^{-2}$)	Absolute MAE (MW)	Daily cumulative error (MWh)
SVM	0.8462	9.2425	2.7423	14.7418	2.65	254.40
LSTM	0.8712	8.4958	1.9204	13.8582	2.44	234.24
BiLSTM	0.8842	8.2246	1.5841	12.5863	2.36	226.56
CNN-LSTM	0.8904	8.0215	1.4122	12.2882	2.30	220.80
TCN-BiLSTM	0.8938	7.9151	1.3372	12.0375	2.27	217.92
AP-TCN-BiLSTM	0.9084	7.7816	1.1755	11.8521	2.23	214.08

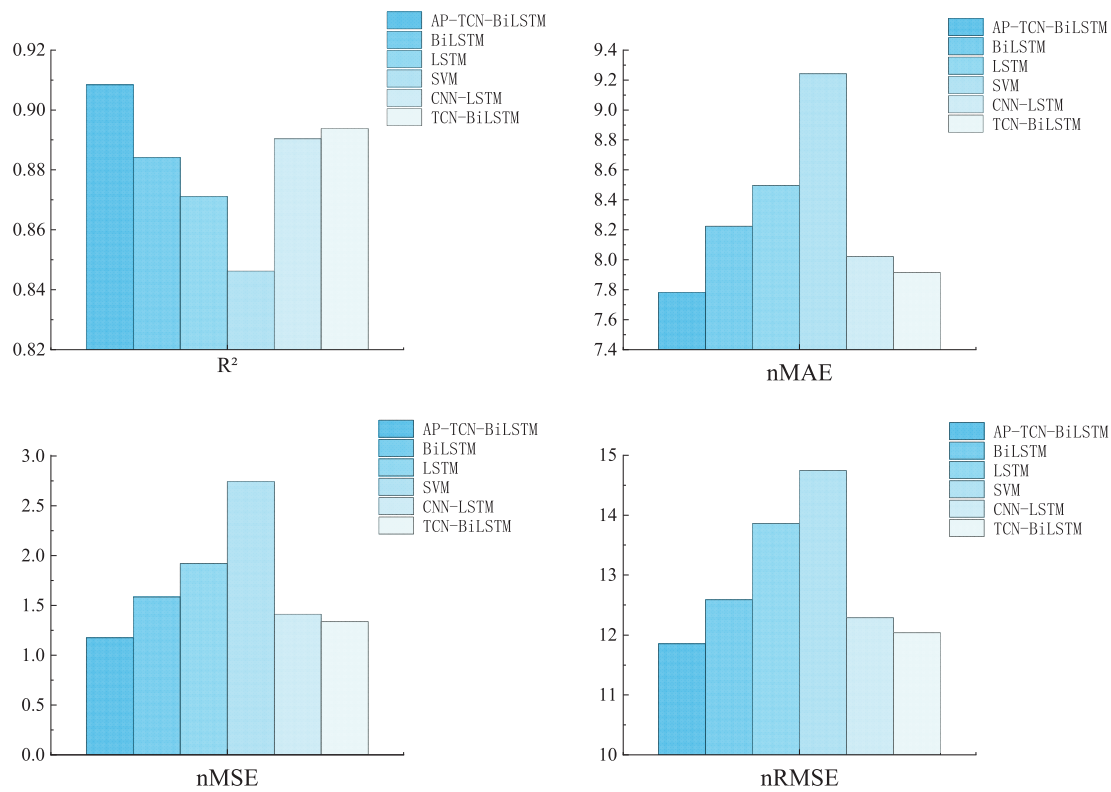


Figure 9: Histogram of the performance evaluation of each model

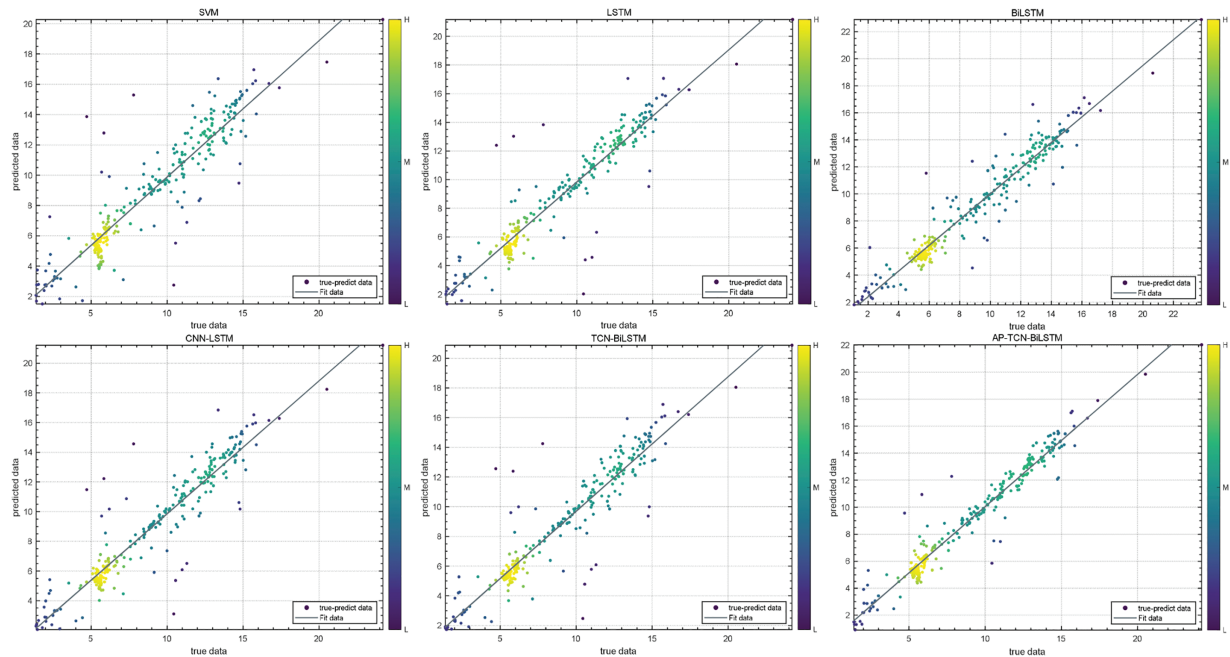


Figure 10: Scatter plots of wind power predicted values vs. actual values for all models

As shown in Table 9 and Fig. 9, the lowest values of nRMSE, nMAE, and nMSE are achieved by the proposed model, while the highest R^2 value is observed. Additionally, the scatter plot distribution is shown to be more convergent. These results indicate superior prediction accuracy performance. Compared to single-model architectures, the advantages of sub-models are fully utilized in the combined approach, with sequence dependency relationships are more effectively modeled. Further improvements in all evaluation metrics are observed after incorporating weather clustering into the combined model. This enhancement demonstrates that fluctuation characteristics of wind power are better learned through meteorological scenario classification.

The proposed model demonstrates superior predictive accuracy, achieving the lowest nRMSE, nMAE, and nMSE values among all benchmark models. It is noteworthy that the installed capacity of this wind farm is 49.5 MW, and the average output power during the test period is 28.7 MW. In terms of specific metrics: For the CNN-LSTM model, the normalized mean absolute error (nMAE) is 8.02% and the normalized root mean squared error (nRMSE) is 12.29%, corresponding to an approximate mean absolute error (MAE) of 2.30 MW. For the TCN-BiLSTM model, the nMAE is 7.92% and the nRMSE is 12.04%, with a corresponding MAE of about 2.27 MW. For the BiLSTM model, the nMAE is 8.22% and the nRMSE is 12.59%, corresponding to an MAE of approximately 2.36 MW. In contrast, the proposed AP-TCN-BiLSTM model in this study achieves a low nMAE of 7.78% and an nRMSE of 11.85%, equivalent to an MAE of 2.23 MW. Compared with the CNN-LSTM model, the proposed model reduces the 15-min prediction error by 0.07 MW; compared with the TCN-BiLSTM model, it reduces the 15-min error by 0.04 MW; and compared with the BiLSTM model, it reduces the 15-min error by 0.13 MW. Based on 96 daily prediction points, the proposed model reduces the daily cumulative error by 6.72 MWh relative to the CNN-LSTM model, 3.84 MWh relative to the TCN-BiLSTM model, and 12.48 MWh relative to the BiLSTM model. Notably, the cumulative error reduction compared with the BiLSTM model accounts for 2.5% of the daily power generation of this medium-sized wind farm, which holds significant value for power grid dispatching. The highest R^2 value is attained,

confirming minimal deviation between predicted and actual values. Fig. 10 visually validates this close agreement via comparative result plots.

As demonstrated in Fig. 11, the closest fit between predicted and actual wind power curves is achieved by the proposed model. This optimal fitting confirms superior performance and the highest prediction accuracy for wind power series data. Consequently, improved power resource allocation is enabled, while operational risks in wind power systems are reduced and grid equipment reliability is enhanced.

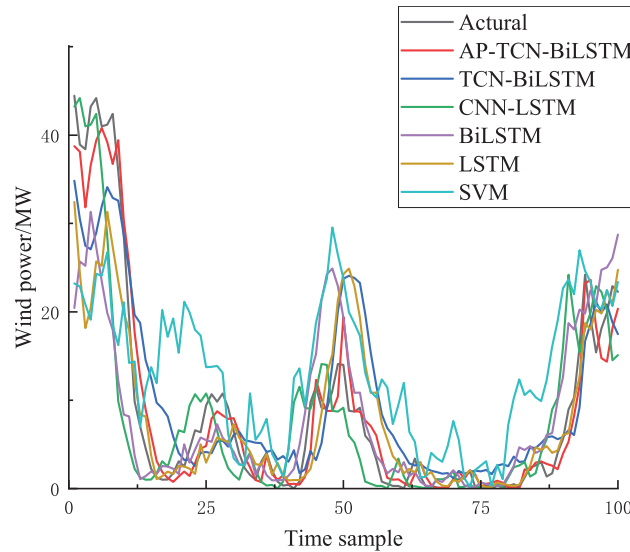


Figure 11: Results of wind forecasting curve

5 Conclusion

To improve wind power forecasting accuracy, weather scenario classification and sequence feature modeling are integrated, and an adaptive forecasting model based on Affinity Propagation (AP) clustering and BKA-TCN-BiLSTM is proposed. Case studies are conducted, and the following conclusions are derived:

- (1) The mutual information method is employed to quantify the nonlinear correlations between meteorological features and power sequences, thereby enabling the selection of highly correlated core features. Historical data are classified into typical weather scenarios via unsupervised AP clustering. This approach overcomes the limited universality of traditional single models under diverse meteorological conditions and also enhances the statistical regularity of feature distributions within each scenario. Consequently, a reliable data foundation is established for customized modeling of sub-scenarios.
- (2) A TCN-BiLSTM fusion architecture is constructed, which captures multiscale local features through dilated causal convolutional layers. Meanwhile, global sequence dependencies are modeled using bidirectional long short-term memory (BiLSTM) networks, effectively improving wind power forecasting accuracy. Unlike traditional fixed parameter settings, the BKA optimization algorithm is introduced to determine the model's hyperparameters. This optimization mechanism significantly enhances generalization capability while maintaining computational efficiency. Experimental results demonstrate the model's superior root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) values in comparison with benchmark models, confirming its effectiveness in short-term wind power forecasting.
- (3) Future research efforts will focus on two interconnected directions: methodological advancements to enhance wind power forecasting performance and the translation of algorithmic innovations

into practical engineering applications; specifically, methodologically, we will prioritize improving the dimensionality and accuracy of weather scenario classification by incorporating additional meteorological factors, furthermore, signal decomposition methods will be employed to address the complexity and nonlinear characteristics of operational data, thereby facilitating more in-depth extraction of latent patterns in wind power time series, concurrently, explorations of more efficient optimization algorithms and model fusion strategies will be pursued to continuously adapt to the evolving demands of wind power forecasting, and additionally, we will develop a lightweight API system to enable seamless integration with existing power grid dispatching platforms—this initiative aims to advance the transition of the proposed techniques from algorithmic innovation to tangible engineering practice, thus enhancing the practical utility of the developed forecasting framework in real-world power system operations.

Acknowledgement: Authors thank those who contributed to write this article and give some valuable comments.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Mingxuan Ji and Jing Gao; methodology, Mingxuan Ji; software, Mingxuan Ji; validation, Mingxuan Ji, Jing Gao and Dantian Zhong; formal analysis, Mingxuan Ji; investigation, Mingxuan Ji; data curation, Mingxuan Ji, Yingqi Xu, Zhongxiao Du and Shuxiang Yang; writing—original draft preparation, Mingxuan Ji; writing—review and editing, Jing Gao and Mingxuan Ji; visualization, Mingxuan Ji and Yingming Liu; supervision, Jing Gao and Dantian Zhong; project administration, Jing Gao and Dantian Zhong. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Ethics Approval: Not applicable, for studies not involving humans or animals.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Luo J, Liu B, Guo X, Yang L, Liu C, Xu Y, et al. Oscillation stability induced by wind power integration: incidents, mechanism, countermeasures and future challenges. *Renew Sustain Energy Rev.* 2025;217(5):115692. doi:10.1016/j.rser.2025.115692.
2. Guo X, Zeng P, Xiong X, Wang G, Cui Y. Short-term wind power forecasting methods based on machine learning: a review and case study. *Energy Rep.* 2025;14(4):3753–82. doi:10.1016/j.egyr.2025.10.040.
3. Guo L, Xu C, Yu T, Wumaier T, Han X. Ultra-short-term wind power forecasting based on long short-term memory network with modified honey badger algorithm. *Energy Rep.* 2024;12(6):3548–65. doi:10.1016/j.egyr.2024.09.021.
4. Liu S, Li Q, Lu B, He J. Prediction of offshore wind turbine wake and output power using large eddy simulation and convolutional neural network. *Energy Convers Manag.* 2025;324:119326. doi:10.1016/j.enconman.2024.119326.
5. Li LL, Cen ZY, Tseng ML, Shen Q, Ali MH. Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic-support vector regression machine. *J Clean Prod.* 2021;279(3):123739. doi:10.1016/j.jclepro.2020.123739.
6. AbdElkader AG, ZainEldin H, Saafan MM. Optimizing wind power forecasting with RNN-LSTM models through grid search cross-validation. *Sustain Comput Inform Syst.* 2025;45(15):101054. doi:10.1016/j.suscom.2024.101054.
7. Liu H, Li P, Wang Q. Short-term prediction of wind power based on BKA-CNN-BiGRU-attention. *Electr Power Syst Res.* 2026;253(6):112543. doi:10.1016/j.epr.2025.112543.
8. Chen T, Qian Z, Jing B, Han M. Research on short-term wind power forecasting model based on K-means++ and ELM. *Electr Meas Instrum.* 2024;61(6):45–50. doi:10.19753/j.issn1001-1390.2024.06.006.

9. Wang B, Feng S, Liu C. Study on weather typing based wind power prediction. *Power Syst Technol.* 2014;38(1):93–8. doi:10.13335/j.1000-3673.pst.2014.01.015.
10. Yang J, Wang X, Luo X, Jiang C. Intelligent combination prediction of wind power based on numerical weather prediction and fuzzy clustering. *Acta Energiæ Solaris Sin.* 2017;38(3):669–75. doi:10.19912/j.0254-0096.2017.03.014.
11. Gao J, Ji M, Wang H, Du Z. Research on the IL-bagging-DHKELM short-term wind power prediction algorithm based on error AP clustering analysis. *Comput Mater Contin.* 2024;79(3):5017–30. doi:10.32604/cmc.2024.050158.
12. Wang J, Wang WC, Hu XX, Qiu L, Zang HF. Black-winged kite algorithm: a nature-inspired meta-heuristic for solving benchmark functions and engineering problems. *Artif Intell Rev.* 2024;57(4):98. doi:10.1007/s10462-024-10723-4.
13. Sun Y, Yang J, Zhang X, Hou K, Hu J, Yao G. An ultra-short-term wind power forecasting model based on EMD-EncoderForest-TCN. *IEEE Access.* 2024;12:60058–69. doi:10.1109/ACCESS.2024.3373798.
14. Liu M, Qin H, Cao R, Deng S. Short-term load forecasting based on improved TCN and DenseNet. *IEEE Access.* 2022;10:115945–57. doi:10.1109/ACCESS.2022.3218374.
15. Sheng Y, Wang H, Yan J, Liu Y, Han S. Short-term wind power prediction method based on deep clustering-improved temporal convolutional network. *Energy Rep.* 2023;9(2):2118–29. doi:10.1016/j.egyr.2023.01.015.
16. Shao Z, Han J, Zhao W, Zhou K, Yang S. Hybrid model for short-term wind power forecasting based on singular spectrum analysis and a temporal convolutional attention network with an adaptive receptive field. *Energy Convers Manag.* 2022;269(3):116138. doi:10.1016/j.enconman.2022.116138.
17. Joseph LP, Deo RC, Prasad R, Salcedo-Sanz S, Raj N, Soar J. Near real-time wind speed forecast model with bidirectional LSTM networks. *Renew Energy.* 2023;204(7):39–58. doi:10.1016/j.renene.2022.12.123.
18. Peng S, Zhu J, Wu T, Yuan C, Cang J, Zhang K, et al. Prediction of wind and PV power by fusing the multi-stage feature extraction and a PSO-BiLSTM model. *Energy.* 2024;298(8):131345. doi:10.1016/j.energy.2024.131345.
19. Zhou D, Liu Y, Wang X, Wang F, Jia Y. Combined ultra-short-term photovoltaic power prediction based on CEEMDAN decomposition and RIME optimized AM-TCN-BiLSTM. *Energy.* 2025;318(1):134847. doi:10.1016/j.energy.2025.134847.
20. Xie C, Yang X, Chen T, Fang Q, Wang J, Shen Y. Short-term wind power prediction framework using numerical weather predictions and residual convolutional long short-term memory attention network. *Eng Appl Artif Intell.* 2024;133:108543. doi:10.1016/j.engappai.2024.108543.
21. Limouni T, Yaagoubi R, Bouziane K, Guissi K, Baali EH. Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model. *Renew Energy.* 2023;205:1010–24. doi:10.1016/j.renene.2023.01.118.
22. Liu S, Xu T, Du X, Zhang Y, Wu J. A hybrid deep learning model based on parallel architecture TCN-LSTM with Savitzky-Golay filter for wind power prediction. *Energy Convers Manag.* 2024;302(3):118122. doi:10.1016/j.enconman.2024.118122.
23. Xiong J, Peng T, Tao Z, Zhang C, Song S, Nazir MS. A dual-scale deep learning model based on ELM-BiLSTM and improved reptile search algorithm for wind power prediction. *Energy.* 2023;266(2):126419. doi:10.1016/j.energy.2022.126419.
24. Özüpak Y. Real-time detection of photovoltaic module faults using a hybrid machine learning model. *Sol Energy.* 2025;302(9):114014. doi:10.1016/j.solener.2025.114014.