Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A wind power forecasting method based on CEEMDAN- LightGBM-Transformer

Jingsong ZHANG
Xichang University
Yukai Zhou
Xichang University
Xiaorui CHU
xcc20230015@xcc.edu.cn

Xichang University
Yuanyuan Yang
Xichang University
Yangdi Xu
Xichang University

Research Article

Keywords: Power prediction, Mode decomposition, Gradi entenhancement, Self attention mechanism

Posted Date: September 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10563026/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A wind power forecasting method based on CEEMDAN- LightGBM-Transformer

Jingsong ZHANG [1], Yukai ZHOU [1], Xiaorui CHU [1,*],Yuanyuan YANG [1], Yangdi XU [1]

[1]School of Mechanical and Electrical Engineering,Xichang University, Xichang Sichuan 61500,China; 187815567881@163.com
*Correspondence: xcc20230015@xcc.edu.cn;

Abstract

Addressing the challenge of accurately predicting wind power output due to its intermittency, volatility, and randomness, this paper proposes a CEEMDAN-LightGBM -Transformer model. Initially, CEEMDAN decomposes the data into multiple IMFs and employs the "drop_k" strategy to select IMF components with prominent feature information and low redundancy. Subsequently, LightGBM is utilized to construct nonlinear features from the data, which are then combined with strongly correlated linear features to rank feature importance. Finally, the multi-dimensional features are fused using a Transformer with an Encoder-Decoder architecture to generate a low-error power prediction sequence. Experimental results demonstrate that, compared to models such as Random Forest and GRU, the proposed model achieves a 15% reduction in both mRMSE and mMAE, with an R^2 value exceeding 98%. Furthermore, CEEMDAN, LightGBM, and Transformer exhibit synergistic effects in enhancing the model's performance. This model provides effective support for grid dispatching, power consumption management, and equipment maintenance decision-making.

Keywords: Power prediction; Mode decomposition; Gradi entenhancement; Self attention mechanism

1. Introduction

In recent years, the installed capacity of clean energy sources, such as wind power, has been increasing annually. By the end of 2025, the total installed capacity of wind power in China is projected to exceed 60 GW. However, wind power exhibits characteristics of intermittency, volatility, and randomness. If effective prediction of wind power generation cannot be achieved[1], issues related to its grid connection scheduling, power consumption, and equipment maintenance will significantly undermine the stable and economic operation of the power system. This represents one of the critical challenges that urgently need to be addressed in the current multi-energy complementary power system[2][3].Typical power prediction methods include Auto-Regressive Integrated Moving Average(ARIMA)[4][5], Support Vector Machine (SVM)[6][7], Multiple Linear Regression(MLR)[8], and Random Forest[9][10]. Nevertheless, these methods are susceptible to multicollinearity interference and are unable to effectively handle the nonlinear and highly volatile nature of wind power.

With the rapid application and advancement of deep learning technology in the power sector, power prediction methods based on deep learning have also yielded promising results. Zhang Y[11] integrated the chaos method with the Cauchy distribution to optimize the

parameters of the seagull algorithm and LSTM model, subsequently designing an interval prediction model combined with kernel density estimation based on data augmentation theory. Mo S[12] constructed the Powerformer model on the foundation of Transformer, enhancing the model's capacity for temporal feature extraction while reducing its complexity. Karijadi I[13] combined CEEMDAN and EWT to denoise high-frequency data sequences, and then employed LSTM to predict and output the sub-sequences. Li Q[14] utilized sample entropy to classify the modal information decomposed by VMD, and constructed a dual-channel architecture incorporating TCN and Informer models to extract feature information from modal groups of varying complexity. Yang T[15] integrated CNN, BiLSTM, and a multi-head attention mechanism to model the modal information decomposed by MVMD, thereby constructing a hybrid predictive deep neural network. Li N[16] employed EMD to decompose wind power sequences, introduced a convolutional attention mechanism to replace the dot-product attention mechanism in the Transformer model, and constructed the CCTransformer model to reduce prediction errors. Xiao Y[17] improved the artificial lemming algorithm, combined it with TCN and BiLSTM methods, and addressed the issues of premature convergence and local optima in handling high-dimensional problems. Hussain RMS[18] proposed a modular spatiotemporal wind energy prediction framework that considered the spatiotemporal dependencies between wind turbines and combined dynamic time warping with Transformer to reduce model prediction errors. Jiang W[19] designed a spatiotemporal multimodal dynamic graph network for wind power prediction based on physical information, integrating meteorological data with multimodal methods to overcome the spatiotemporal dependencies between different meteorological data patterns during power prediction. Wang D[20] introduced adaptive weighting parameters into the LSTM model to fuse historical and predicted wind speeds, enabling autonomous calibration of predicted wind speeds and enhancing the model's ability to reconcile discrepancies between predicted and actual wind speeds.

The innovative aspects of this study are as follows:

(1)CEEMDAN is employed for wind power time series decomposition. To overcome the subjective segmentation drawback of traditional modal decomposition, a noise indicator is introduced to quantitatively differentiate the predictive performance of each intrinsic mode function (IMF) component. Furthermore, tailored to the temporal evolution characteristics of wind power sequences, a dropout filtering strategy that discards the first several high-frequency IMFs is developed. By merely removing the dominant high-frequency component, the proposed approach achieves a favorable trade-off between noise reduction and dynamic feature retention, mitigates modal aliasing, and enables effective multi-scale signal decoupling, thereby providing refined and reliable feature inputs for LightGBM and Transformer models.

(2)Developing a multidimensional feature selection framework by integrating Pearson correlation analysis with the LightGBM algorithm. Wind speed is screened and retained as the dominant linear predictive feature, whereas nonlinear meteorological variables including atmospheric pressure and humidity are reserved based on their contribution rankings of LightGBM feature gain. On this basis, a hybrid prediction model that integrates univariate linear correlations and multi-factor nonlinear coupling relationships is established, which

realizes the complementary fusion of diverse feature information and further improves the rationality and comprehensiveness of model input features.

(3) Adopting an encoder-decoder causal architecture integrated with masked multi-head attention to effectively eliminate temporal data leakage. Cross-attention is introduced to implement cross-dimensional coupling between historical feature information and prediction sequences, while the multi-head parallel mechanism is utilized to extract multi-scale temporal dependencies synchronously. Furthermore, the incorporation of sinusoidal positional encoding enables the model to capture inherent periodic characteristics of wind power sequences, which facilitates the accurate fitting of complex wind power operating conditions, including power fluctuation and peak variation patterns.

2.CEEMDAN-LightGBM-Transformer Power Forecast Model

2.1. CEEMDAN Data Mode Dcomposition

Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [20] is an improved variant derived from empirical mode decomposition (EMD) [21]. The EMD algorithm first identifies local extrema of the raw signal to construct upper and lower envelopes, and subsequently computes their mean curves. By subtracting these mean components from the original signal, one intrinsic mode function (IMF) can be extracted. This sifting procedure is executed iteratively until the residual signal turns monotonic, where no additional IMFs can be separated out.Nevertheless, conventional EMD suffers from severe mode mixing in its decomposition results. Individual IMFs frequently contain components with overlapping frequency bands, making it impossible to distinguish modal information effectively. To address this defect, CEEMDAN injects adaptive white noise into the IMFs yielded by EMD. Leveraging the stochastic property of white noise, this approach substantially mitigates or even eliminates mode mixing, generates more reliable IMFs, and thus enhances the precision and stability of signal decomposition. The core implementation steps of the CEEMDAN algorithm are elaborated as follows:

Let $E_j(\cdot)$ table denote the jth IMF component obtained after performing EMD decomposition on the sequence, let ε_m Table denote the noise coefficient added to the input sequence by CEEMDAN in the mth stage, and let $C_m(t)$ denote the mth IMF component generated by CEEMDAN.

(1) add N instances of Gaussian white noise to the original signal $y(t)$ to construct a total of N preprocessed sequences $y_n(t)$ 。

$$y_n(t) = y(t) + \varepsilon_0 \delta_n(t) \tag{1}$$

In Equation (1), ε_0 is the weight coefficient of the Gaussian white noise, and $\delta_n(t)$ is the Gaussian white noise at the nth processing step, where $n=1,2,\cdots,N$。

(2) Perform EMD decomposition on all preprocessed sequences $y_n(t)$ to obtain the first IMF component $c_1^n(t)$; take its mean as the first IMF component $c_1(t)$ obtained from the CEEMDAN decomposition; and simultaneously obtain the first residual sequence $r_1(t)$, as shown below.

$$c_1(t)=\frac{1}{N}\sum_{n=1}^{N}c_1^n(t),\quad r_1(t)=y(t)-c_1(t) \tag{2}$$

Similarly, by adding Gaussian white noise to the residual sequence $r_1(t)$ to construct N new sequences $r_1(t)+\varepsilon_1 E_1(\delta_n(t))$, performing EMD decomposition on these N sequences, and then calculating their mean to obtain the second IMF component $c_2(t)$, and taking the difference to obtain $r_2(t)$, as shown below.

$$c_2(t)=\frac{1}{N}\sum_{n=1}^{N}E\left(r_1(t)+\varepsilon_1 E_1(\delta_n(t))\right),\quad r_2(t)=r_1(t)-c_2(t) \tag{3}$$

By analogy, the residuals for the first stage are given by the following equation.

$$r_m(t)=r_{m-1}(t)-c_m(t) \tag{4}$$

(4) Performing N EMD decompositions on $r_m(t)+\varepsilon_m E_m(\delta_n(t))$ yields the (m+1)th IMF sequence resulting from the CEEMDAN decomposition, as shown below.

$$c_{m+1}(t)=\frac{1}{N}\sum_{n=1}^{N}E_1\left(r_m(t)+\varepsilon_m E_m(\delta_n(t))\right) \tag{5}$$

(5) Repeat the above steps until the decomposition stops; the final residual sequence is given by the following equation.

$$R(t)=y(t)-\sum_{m=1}^{M}c_m(t) \tag{6}$$

The expression for the signal sequence $y(t)$ after CEEMDAN decomposition is shown below.

$$y(t) = R(t) + \sum_{m=1}^{M} c_m(t) \tag{7}$$

2.2. LightGBM Multidimensional feature reconstruction and selection

Light gradient boosting machine(LightGBM)[22] constitutes an optimized evolution of the Gradient Boosting Decision Tree (GBDT) [23]. Fundamentally, GBDT assembles multiple weak decision trees into a robust strong learner. The training pipeline is initialized using statistical baselines such as the mean value of sample labels. Decision trees are trained sequentially in an iterative manner; the predictive outputs of newly trained trees are incorporated into the pre-established model under a predefined learning rate to iteratively rectify fitting residuals, before the final prediction is yielded via weighted aggregation of all iteratively generated decision trees.

Nevertheless, the sequential training paradigm inherent to GBDT severely restricts parallel computing capability. The algorithm necessitates repeated full traversal of input datasets, thereby incurring inferior computational efficiency. Additionally, GBDT exhibits pronounced sensitivity to outlier samples, is susceptible to disturbance from data noise, and delivers degraded predictive performance when confronted with high-dimensional sparse feature inputs.

Retaining the core gradient boosting and residual fitting framework of GBDT, LightGBM integrates a series of optimized modules including histogram-based discretization, depth-limited leaf-wise tree growth, Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). Such multi-pronged optimizations substantially alleviate the critical drawbacks of GBDT, namely sluggish training speed, excessive memory overhead and poor scalability on large-volume datasets, and further diminish the model’s prediction error. The primary implementation procedures of the LightGBM algorithm are detailed as follows:

（1）Histogram Data Preprocessing

The histogram algorithm discretizes continuous floating-point feature variables into k discrete integer intervals and establishes a histogram with k bins. During data traversal, the discretized indices are adopted to accumulate statistical information within the histogram. Subsequently, optimal feature segmentation thresholds are determined by traversing all discrete intervals. Since the algorithm only stores discretized feature information instead of original continuous values, it effectively reduces the memory occupation of the model and improves computational efficiency.

（2）Optimization Strategies

During the node-splitting procedure of decision trees, the leaf-wise growth strategy only segments the leaf node that yields the maximum split gain at each iteration, which inevitably introduces the risk of model overfitting. To mitigate this issue, the depth-restricted leaf-wise growth mechanism imposes an upper bound on the maximum tree depth. This improvement suppresses overfitting risks and lowers prediction residuals without compromising computational efficiency. Gradient-based One-Side Sampling (GOSS) reserves samples with large gradient magnitudes to preserve critical gradient information and alleviate computational overhead. Exclusive Feature Bundling (EFB) aggregates mutually exclusive

sparse features extracted from sample datasets into integrated composite features, which drastically cuts down feature dimensionality and the overall parameter scale of the model.

（3）Decision Output

A set of decision trees is constructed, where each tree incorporates depth-constrained leaf-wise splitting for parameter calculation. The predicted outputs from all individual trees are aggregated to generate the final model result. In this work, the maximum tree depth is empirically set to 8.The LightGBM algorithm structure is shown in Figure 1.

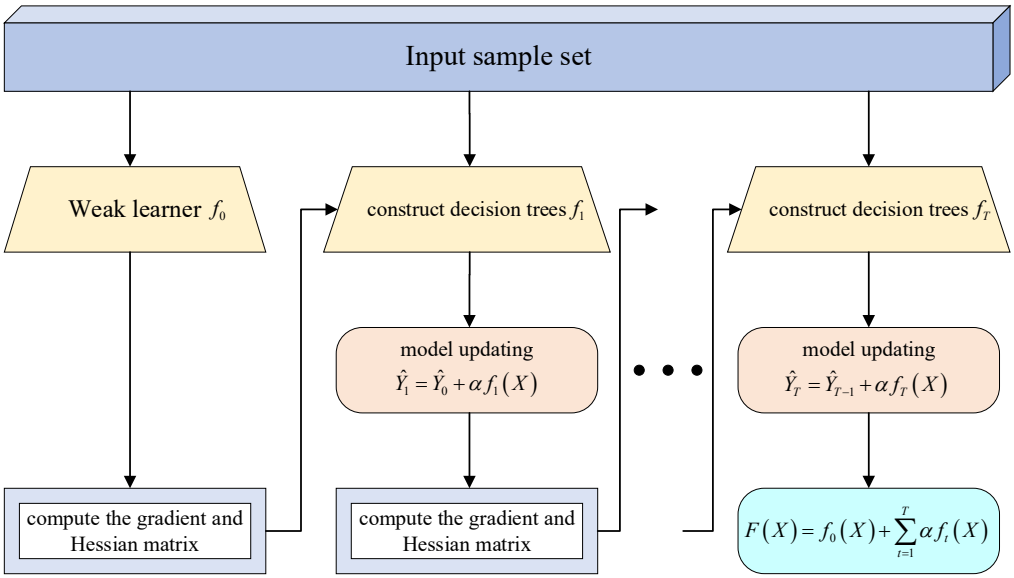


Figure 1. LightGBM algorithm structure

In this study, systematic feature engineering is implemented on the acquired sample datasets. Primitive meteorological variables extracted from raw anemometer tower data, including multi-height temperature, humidity, wind speed and wind direction, are adopted as baseline input features. Furthermore, eight additional feature categories are artificially constructed, covering temporal periodic features, wind speed standard deviation, and approximated wind energy density, yielding a total of 28-dimensional feature inputs. Subsequent feature significance screening is conducted to refine the feature set, and 20 high-contribution features are ultimately retained, dominated by wind direction consistency, 70 m wind direction, and 70 m wind speed from the anemometer tower. The screened feature data is converted into tensor format and fed into a customized DataLoader configured with fixed sliding windows, batch size settings, and standardized sampling strategies. Following feature and label extraction, the processed dataset is imported into the LightGBM model for iterative training and preliminary wind power prediction.

2.3. Transformer Power Sequece Forecasting

The Transformer architecture [24] mainly comprises an input layer, an Encoder‑Decoder module [25], and an output layer. For sequence prediction tasks, the core advantage of the Transformer lies in the capability of the self-attention mechanism to capture the inherent continuity, periodicity, and long-range dependency characteristics of time-series data. In this work, positional encoding (PE) is firstly imposed on the multi-dimensional feature matrix derived from the LightGBM module. Specifically, the fixed sinusoidal encoding strategy, as formulated in Equations (8) and (9), is adopted to inject positional

embedding vectors into the feature matrix, which explicitly characterizes the temporal order of the sequential samples.

$$\mathrm{PE}_{(\mathrm{pos},2\mathrm{i})} = \sin\left(\mathrm{pos}/10000^{2\mathrm{i}/\mathrm{d}_{\mathrm{model}}}\right) \quad (8)$$

$$\mathrm{PE}_{(\mathrm{pos},2\mathrm{i}+1)} = \cos\left(\mathrm{pos}/10000^{2\mathrm{i}/\mathrm{d}_{\mathrm{model}}}\right) \quad (9)$$

In Equations (8) and (9), pos denotes the time step positions within the historical sliding window, i is the feature dimension index, and $\mathrm{d}_{\mathrm{model}}$ is the dimension of the model's hidden layer. This encoding method enables the model to accurately perceive the relative and absolute positional relationships of wind power across different time steps.

Most existing studies adopt an encoder-only architecture for sequence encoding and decoding tasks. Such a structure implements bidirectional encoding over historical time series and acquires global information of the entire historical window in a one-shot manner, rather than generating predictions through autoregressive stepwise inference. By directly mapping historical inputs to all future time steps in a single forward pass, this paradigm ignores temporal causal constraints and consequently suffers from implicit future information leakage. To address this limitation, this paper employs a canonical encoder–decoder framework. The encoder module is responsible for modeling historical sequence data, wherein the self-attention mechanism [26] captures sophisticated internal dependencies within time-series samples, and the feed-forward network (FFN) [27] excavates high-order latent temporal features, yielding comprehensive semantic representations containing bidirectional sequential information. The decoder integrates the refined latent features output by the encoder with future prediction placeholders. With residual connection and layer normalization operations, the model ultimately generates causally constrained and recursively reasoned prediction results that conform to real temporal evolution rules.

Compared with the conventional encoder-only architecture, the proposed encoder–decoder framework strictly obeys temporal causal principles, effectively eliminating latent information leakage. Furthermore, this structure supports accurate multi-step prediction for both short-term and long-term time-series scenarios. The overall architecture of the transformer network adopted in this study is visualized in Figure 2.

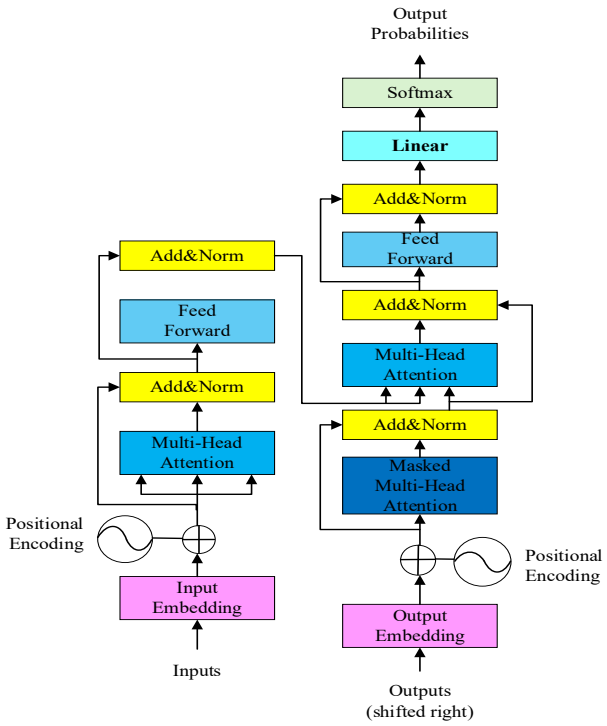


Figure 2. Transformer predicting network structures

Considering the inherent seasonal periodicity and time-lagged correlation characteristics of wind power time-series data, this study adopts a multi-head self-attention (MHA) mechanism [28] to replace the conventional single self-attention structure. This modification enables the model to capture richer and more comprehensive modal feature representations, and the specific implementation formulation of the MHA mechanism is defined in Equation (10).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (10)$$

In Equation (10), d_k represents the dimension of the key vector, and $\sqrt{d_k}$ acts as a scaling factor to prevent vanishing gradients. It is capable of simultaneously capturing long-range temporal dependencies, adaptively focusing on key time periods, and extracting multi-scale periodic and trend features from the data. Its structure is shown in Figure 3.

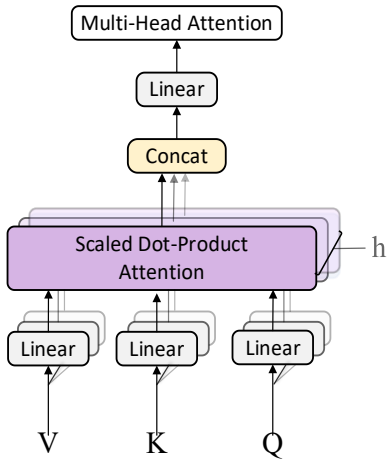


Figure 3. Multi-head attention echanism architecture

Furthermore, the vanilla self-attention module inside the decoder is substituted with masked multi-head self-attention and cross-attention modules [29]. Masked multi-head self-attention introduces causal masks to shield future temporal signals, restricting the model to exclusively exploit historical observations prior to and including the current time step, which rigorously maintains the causal constraint of time-series forecasting. As for cross-attention, it regards the target sequence as queries and the source sequence as key-value pairs to realize cross-sequence information matching and latent feature fusion, so as to precisely excavate the inherent correlation dependencies between historical and future series segments.

3. Result

3.1. Test Environment

The platform parameters used in this experiment are as follows: The CPU is intel(R) Core(TM)i5-14600KF, the RAM is 32GB, The operating system is Windows10, the GPU isNVIDIA GeForce RTX 2080 16GB; the selected software isPycharm 2026、Anaconda.

3.2. Experimental Data

The experimental data in this article comes from the power generation data of a wind farm in China from January 1 to December 31, 2019. The installed capacity is 200MW, and the sequence sampling interval is 15 minutes. The dataset is used as the training set for the first 80% of the time series, the validation set for the middle 10%, and the test set for the last 10%. The sample order is not disrupted throughout the entire process.The format of the dataset is shown in Table 1.

Table 1. Data set format

Time	Wind speed (m/s) 10m…70m	Hub Height Wind Speed (m/s)	Wind direction (o) 10m…70m	Rim Height Wind Direction (o)	Temper-ature (c^o)	Air pressure (hPa)	Humidity (%)	Power (MW)
2019/01/01 -0:00	2.8…4.5	4.5	214.5…248.1	248.1	13.2	874.7	56.9	3.5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2019/12/31 -23:45	11.4…16.5	16.5	71.7…75.8	75.8	13.1	864.8	48.9	199.3

3.2.1.Pearson correlation coeficient analysis

In this paper, the raw data were analyzed using Pearson’s correlation coefficient [30]; the correlations between the various input factors and actual power generation are shown in Figure 4.

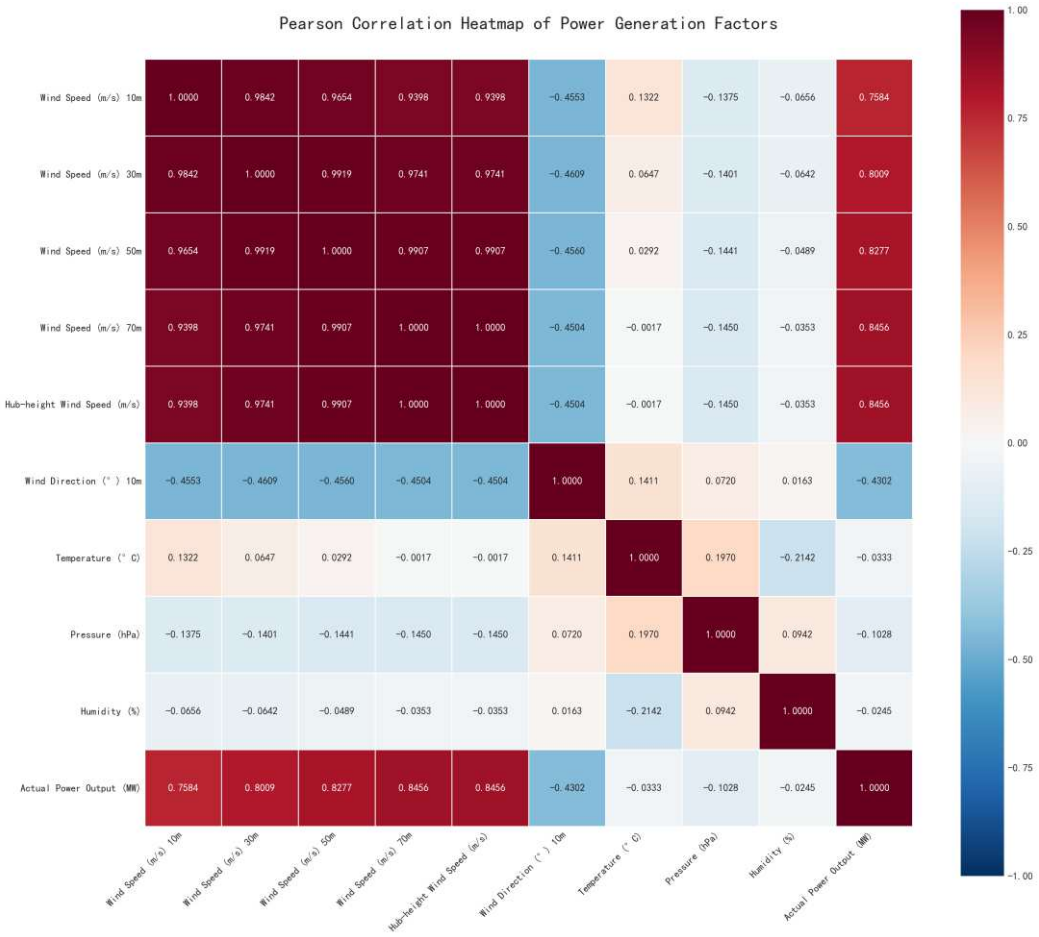


Figure 4. Data pearson correlatin coefficient heatmap figure

As illustrated in Figure 7, wind speed at the hub height and wind speeds measured at heights ranging from 10 m to 70 m on the anemometer tower exhibit strong positive correlations with actual wind power output. Among these variables, wind speed at the hub height and the 70 m wind speed yield the highest correlation coefficients exceeding 0.84 with power generation. Such correlation results can serve as a reference for screening IMFs derived from CEEMDAN decomposition, and the strongly linearly correlated features are subsequently fed into the LightGBM model as core inputs.

3.2.1.Mode decompositio

In this section, the cleaned and filtered wind power time series data undergoes CEEMDAN decomposition, resulting in 8-dimensional IMFs, as shown in Figure 5.

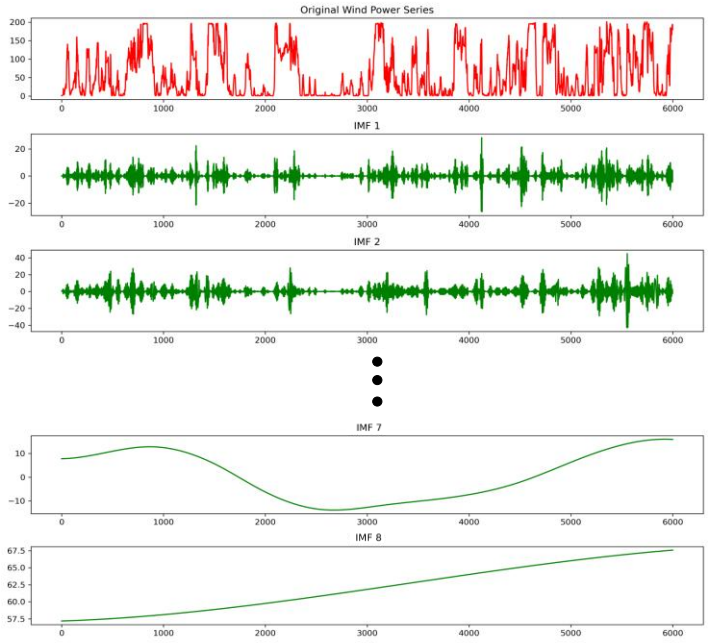


Figure 5. Decomposition results of CEEMDAN

Wind power time series inherently manifest non-stationarity, multi-scale oscillation features and prominent stochastic noise. The high-frequency intrinsic mode functions (IMFs), dominated by IMF1, are dominated by turbine mechanical vibration signals, instantaneous turbulent interference and random stochastic noise, which can be categorized as irregular white-noise components. These components carry no valid information associated with wind power output tendencies or meteorological evolutionary rules, and thereby possess negligible predictive capacity. Instead, such high-frequency noise terms introduce unordered temporal perturbations, exacerbating the model fitting difficulty and magnifying overall forecasting residuals.

Medium-frequency IMFs (IMF2-IMF6) correspond to short-timescale meteorological variations including wind gusts and minor temperature fluctuations, which constitute the primary source of dynamic wind power oscillation and abrupt short-term power mutations. These components embed the most vital temporal dynamic signatures required for short-term forecasting, and govern the model’s capability to identify power peak values and reproduce rapid power ramp-up and ramp-down operating conditions.

Low-frequency IMFs (IMF7-IMF8) depict diurnal and seasonal periodic power output tendencies as well as steady-state baselines of average wind speed, which reflect the long-term stable operational patterns of wind farms. Such low-frequency components dominate the long-range forecasting precision and mitigate bias errors arising from distorted overall trend fitting. The noise indicator values of all decomposed IMFs are visualized in Figure 6.

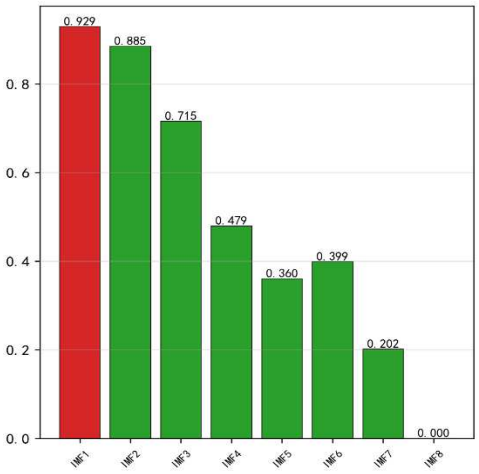


Figure 6. Noise index of each IMF

As can be seen from Figure6, IMF1 has the highest noise index. This paper continues to systematically evaluate the impact of CEEMDAN's "drop_k" strategy (discarding the top k IMFs) on the denoising effect of wind power signals through ablation experiments.

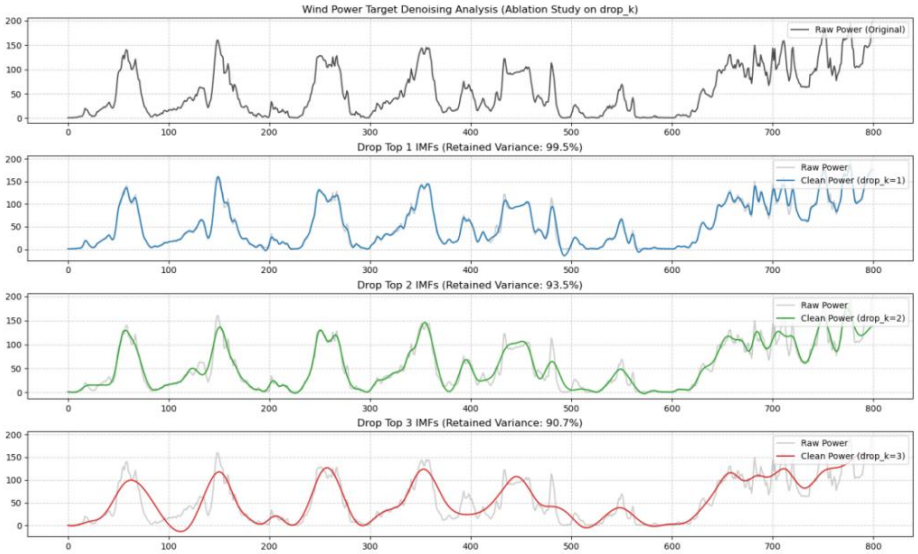


Figure 7. Data denoising results

As illustrated in Figure 7, eliminating merely the first-order IMF (i.e., setting the truncation parameter drop_k = 1) enables effective removal of high-frequency stochastic noise while preserving 99.5% of the original signal variance. The reconstructed signal achieves prominent smoothing effects without erasing critical transient pulse characteristics embedded in the raw sequence.

Although discarding additionalIMFs($drop_k$ = 2or3)further improves signal smoothness, such operation incurs irreversible loss of valuable medium-frequency components and introduces severe signal distortion; the retained variance declines to 93.5% and 90.7% correspondingly. More importantly, the dynamic response signatures with core predictive significance for wind power forecasting are substantially attenuated. Accordingly, within the wind power denoising framework proposed in this work, truncating only the first IMF strikes an optimal trade-off between noise suppression and signal fidelity, rendering this denoising strategy well-suited for short-term forecasting tasks sensitive to subtle power fluctuations.

3.3. Experimental results

The evaluation metrics selected for this experiment include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of Determination (R-Square, R^2), as well as the mean MAE (mMAE), mean RMSE (mRMSE), and mean R^2 (mR^2). The training parameters for the CEEMDAN-LightGBM-Transformer prediction model are set as follows: LightGBM learning_rate=0.03, max_depth=8; Transformer lr=0.0005, with the Adam optimizer selected for both.

3.3.1. Comparative experiment

This paper compares three forecasting models:Random Forest (RF), Gate Recurrent Unit (GRU), and Autoregressive Integrated Moving Average (ARIMA) with CEEMDAN-LightGBM-Transformer (CLT). The performance metrics and parameters for each model are shown in Table 2.

Table 2. Index Parameters of Each Model

	RMSE			MAE			R^2		
Model	**Training set**	**Test set**	**mRMSE**	**Training set**	**Test set**	**mMAE**	**Training set**	**Test set**	**mR^2**
CLT	3.176	6.245	4.711	2.110	3.676	2.893	0.997	0.985	0.991
RF	7.646	17.419	12.533	4.278	10.934	7.606	0.987	0.930	0.959
GRU	27.242	23.824	25.533	19.337	16.150	17.744	0.828	0.868	0.851
XGBoost	26.004	25.953	25.979	18.240	20.282	19.261	0.818	0.826	0.822

As summarized in Table 2, the proposed CLT model yields optimal performance in terms of RMSE, MAE and determination coefficient R^2 across both training and test datasets. In comparison with the XGBoost baseline, the mRMSE and mMAE of the CLT model decline by over 81%, while the mR^2 rises by more than 20% and attains 0.991. The proposed model achieves remarkable superiority over all comparative alternatives in both forecasting error metrics and generalization capability. The fitted power prediction curves of all evaluated models are visualized in Figure 8 for intuitive comparison.

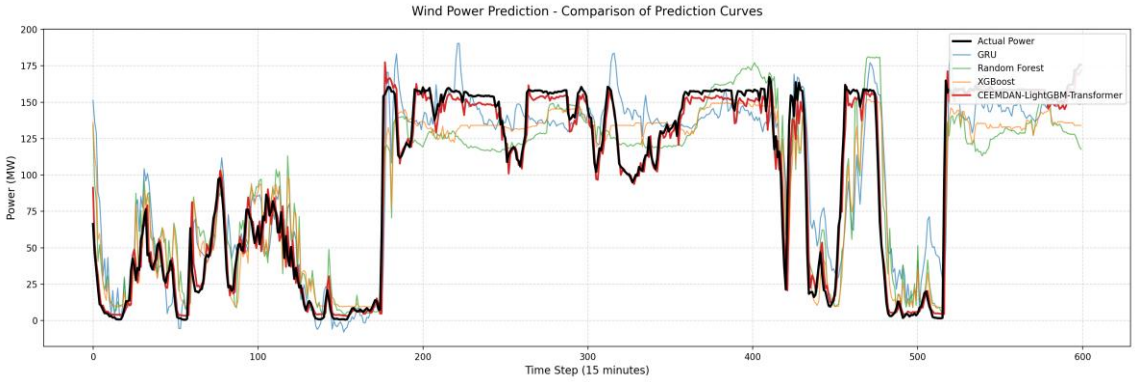


Figure 8. Comparison of prediction performance across models

As visualized in Figure 8, comparative baseline models including GRU, XGBoost and random forest manifest obvious prediction lag and fitting bias under critical operational scenarios characterized by drastic power fluctuations and near-peak output states, which reveals their insufficient capacity to track transient power variations. In sharp contrast, the prediction curve derived from the proposed model achieves a high consistency with the real wind power sequence. The presented framework accurately captures both the overall variational tendency of power output and the amplitude of power peaks, and sustains favorable fitting precision under conventional steady operating conditions as well as extreme volatile scenarios. Furthermore, the scatter points of predicted versus ground-truth values distribute tightly along the ideal perfect-fit diagonal line, which verifies the outstanding predictive superiority of the proposed method for complex wind power time-series forecasting tasks.

3.3.2. Ablation Experiment

To quantitatively explore the performance contribution of each core module, ablation experiments are implemented under unified experimental settings in this study. Several variant models, namely CEEMDAN-LightGBM (CL), LightGBM-Transformer (LT), and CEEMDAN-Transformer (CT), are constructed by removing different functional modules of the proposed framework. Subsequent comparative analysis is conducted to evaluate the performance discrepancies among these variant structures. The quantitative experimental results of all ablation models are presented in Table 3.

Table 3. Model Metrics and Parameters

	RMSE			**MAE**			R^2		
Model	**Training set**	**Test set**	**mRMSE**	**Training set**	**Test set**	**mMAE**	**Training set**	**Test set**	**m R^2**
CLT	3.176	6.245	4.711	2.110	3.676	2.893	0.997	0.985	0.991
CL	4.244	6.864	5.554	2.849	4.285	3.567	0.994	0.982	0.988
LT	5.376	6.644	6.010	3.455	4.041	3.748	0.991	0.982	0.987
CT	5.892	7.180	6.536	3.615	4.363	3.989	0.989	0.979	0.984
Transformer	6.158	7.717	6.938	3.891	4.685	4.288	0.989	0.977	0.983

As summarized in Table 3, the proposed CLT framework attains optimal performance regarding RMSE, MAE and coefficient of determination R^2 on both training and test datasets. Relative to the vanilla Transformer baseline, the mRMSE and mMAE of CLT drop by over 32%, whereas the mR^2 rises by more than 0.8% and peaks at 0.991. The proposed architecture delivers remarkable advantages over all alternative variants in both forecasting error metrics and generalization capacity.

Comparative results between LT and CT reveal that LightGBM-driven feature enhancement contributes more prominent performance gains for denoised non-stationary time-series data than CEEMDAN signal decomposition. Meanwhile, the contrast between CL and CT indicates that after CEEMDAN decomposition, LightGBM outperforms Transformer in residual fitting and subsequence modeling tasks. By synergistically combining the

three-stage pipeline: signal decomposition denoising, tree-model-based rough prediction, and sequential model fine-tuning, the CLT model drastically suppresses forecasting residuals and substantially strengthens overall model robustness.

To further verify the superiority and advancement of the improved Transformer architecture proposed in this work, several state-of-the-art time-series forecasting baselines including PatchTST, Informer and TimesNet are adopted for comparative evaluation. The quantitative experimental metrics of all competing models are tabulated in Table 4.

Table 4. Index Parameters of Each Mode

	RMSE			MAE			R^2		
Model	**Training set**	**Test set**	**mRMSE**	**Training set**	**Test set**	**mMAE**	**Training set**	**Test set**	**m R^2**
Transformer	6.158	7.717	6.938	3.891	4.685	4.288	0.989	0.977	0.983
Informer	11.419	10.879	11.149	8.152	7.545	7.849	0.970	0.971	0.971
TimesNet	31.214	26.966	29.090	21.475	19.091	20.283	0.761	0.804	0.783
PatchTST	42.252	39.937	41.095	29.380	26.087	27.734	0.741	0.690	0.716

As summarized in Table 4, the improved Transformer architecture proposed in this work achieves superior performance against Informer, TimesNet and PatchTST across all evaluation metrics, including RMSE, MAE, R2, mRMSE, mMAE and mR2, with forecasting residuals reduced by a minimum of 37%. These quantitative outcomes demonstrate that, for wind power forecasting scenarios featured with intensive high-frequency random fluctuations, the proposed modified Transformer can sufficiently excavate latent multi-dimensional input features and capture the coupled evolutionary characteristics of time-series power output.

4. Discussion

This study proposes a novel integrated wind power forecasting framework based on the CEEMDAN-LightGBM-Transformer hybrid architecture. Firstly, the raw time-series data are decomposed via the CEEMDAN algorithm, and a drop_k truncation strategy is adopted to screen valid IMF components. This procedure retains high-fidelity signal characteristics while eliminating high-frequency stochastic noise, yielding optimized input samples for subsequent modeling.

Secondly, the LightGBM module conducts nonlinear feature evaluation and quantitative feature importance ranking, whereby a 20-dimensional high-correlation feature subset closely associated with wind power output is precisely selected.

Finally, the embedded Encoder-Decoder Transformer architecture performs fixed sinusoidal positional encoding on the refined multi-dimensional features. The self-attention mechanism captures intrinsic long-range temporal dependencies within historical sequences, and the FFN further excavates deep high-order temporal features to generate accurate forecasting outputs.

Quantitative experimental results demonstrate that the proposed hybrid model reduces both mRMSE and mMAE by 15% compared with conventional baseline models, including random forest and GRU, while achieving R^2 exceeding 0.98. Ablation experiments further validate the individual and collaborative contributions of each embedded module to the overall predictive performance. This predictive model offers robust support for the efficient exploitation of wind power resources and the scheduling of power systems.

References

1. Tuncar, E.A.; Sağlam, Ş.; Oral, B. A review of short-term wind power generation forecasting methods in recent technological trends[J]. Energy Reports, 2024, 12: 197-209.
2. Wang, G.; Zhang, Z.; Lin, J. Multi-energy complementary power systems based on solar energy: A review[J]. Renewable and Sustainable Energy Reviews, 2024, 199: 114464.
3. Zhao. M.; Wang, Y.; Wang, X.;Chang, J.;Chen,Y.Zhou,Y.;Guo,A. Flexibility evaluation of wind-PV-hydro multi-energy complementary base considering the compensation ability of cascade hydropower stations[J]. Applied Energy, 2022, 315: 119024.
4. Alsamamra, H.R.; Salah, S.; Shoqeir, J.H. Performance analysis of ARIMA Model for wind speed forecasting in Jerusalem, Palestine[J]. Energy Exploration & Exploitation, 2024, 42(5): 1727-1746.
5. Shrivastava, R.; Iyer, I.S.; Batra, A. Short term wind prediction with an ARIMA model[J]. Wind Energy and Engineering Research, 2025: 100020.
6. Abedinia, O.; Ghasemi-Marzbali, A.; Shafiei; M.; Sobhani,B.;Gharehpetian,B.G. A multi-level model for hybrid short term wind forecasting based on SVM, wavelet transform and feature selection[C]//2022 IEEE international conference on environment and electrical engineering and 2022 IEEE industrial and commercial power systems europe (EEEIC/I&CPS europe). IEEE, 2022: 1-6.
7. Dong, X.; Deng, S.; Wang, D. A short-term power load forecasting method based on k-means and SVM[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(11): 5253-5267.
8. Sias, Q.A.; Gantassi, R.; Choi, Y.; Bae, J.H. Recurrence Multilinear Regression Technique for Improving Accuracy of Energy Prediction in Power Systems. Energies 2024, 17, 5186.
9. Mustaffa, Z.; Sulaiman, M.H. Random forest based wind power prediction method for sustainable energy system[J]. Cleaner Energy Systems, 2025: 100210.
10. Jamii, J.; Trabelsi, M.; Mansouri, M.; Kouadri, A.; Mimouni,F.M.; Nounou, M. Medium-term wind power forecasting using reduced principal component analysis based random forest model[J]. Wind Engineering, 2024, 48(4): 597-616.
11. Zhang, Y.; Kong, X.; Wang, J.; Wang, H.;Cheng, X. Wind power forecasting system with data enhancement and algorithm improvement[J]. Renewable and Sustainable Energy Reviews, 2024, 196: 114349.
12. Mo, S.; Wang, H.; Li, B.; Xue, Z.; Fan, S.; Liu, X. Powerformer: A temporal-based transformer model for wind power forecasting[J]. Energy Reports, 2024, 11: 736-744.
13. Karijadi, I.; Chou, S.Y.; Dewabharata, A. Wind power forecasting based on hybrid CEEMDAN-EWT deep learning method[J]. Renewable Energy, 2023, 218: 119357.
14. Li, Q.; Ren, X.; Zhang, F.; Guo, L.; Hao, B. A novel ultra-short-term wind power forecasting method based on TCN and Informer models[J]. Computers and Electrical Engineering, 2024, 120: 109632.
15. Yang, T.; Yang, Z.; Li, F.; Wang, H. A short-term wind power forecasting method based on multivariate signal decomposition and variable selection[J]. Applied Energy, 2024, 360: 122759.
16. Li, N.; Dong, J.; Liu, L.; Li, H.; Yan, J. A novel EMD and causal convolutional network integrated with Transformer for ultra short-term wind power forecasting[J]. International Journal of Electrical Power & Energy Systems, 2023, 154: 109470.
17. Xiao, Y.; Yin, Y.; Zhong, R.; Slowik, A.; Chen, H. Q-learning driven artificial lemming algorithm with elite pool and dynamic vertical crossover strategy:a case study in wind power forecasting[J].Expert Systems With Applications,2026,320:132220.

18. Hussain, R.M.S.; Baig, A.M.M; Yousuf, U.M .A modular based framework for multi-step spatio-temporal wind power forecasting [J]. Energy Strategy Reviews,2026,65:102203.
19. Jiang, W.;Wang, J.; He, X.; He, M. PSTG-Net: A physics-informed spatiotemporal multimodal dynamic graph network for wind power forecasting[J].Expert Systems With Applications, 2026,320:132224.
20. Su, Y.; Wang, Z.; Dong, Z.; Hua, X.; Ye, T.; Song, Z.; Shao, Y. Frequency-aware ultra-short-term wind power forecasting using CEEMDAN – VMD – SE and Transformer – GRU networks[J].Energy,2025,338138715-138715.
21. Ning, L.; Jie, D.; Lingyue, L.; Li, H.; Yan, J. A novel EMD and causal convolutional network integrated with Transformer for ultra short-term wind power forecasting[J].International Journal of Electrical Power and Energy Systems,2023,154.
22. Liao, S.; Tian, X.; Liu, B.; Liu, T.; Su, H.; Zhou, B. Short-Term Wind Power Prediction Based on LightGBM and Meteorological Reanalysis[J].Energies, 2022,15, 6287.
23. Kopyt, M.; Piotrowski, P.; Baczyński, D. Short-Term Energy Generation Forecasts at a Wind Farm—A Multi-Variant Comparison of the Effectiveness and Performance of Various Gradient-Boosted Decision Tree Models.Energies,2024,17, 6194.
24. Sun, S.; Liu, Y.; Li, Q.; Wang, T.; Chu, F. Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks[J].Energy Conversion and Management, 2023, 283(c):116916.
25. Wang, Q.; Wang, Y.; Zhang, K.; Liu, Y.; Qiang, W.; Han Wen, Q. Artificial Intelligent Power Forecasting for Wind Farm Based on Multi-Source Data Fusion[J].Processes, 2023,11, 1429.
26. Sun, G.; Zhao, Y.; Li, J.; Cui, F.; Qi, X. DG-LSTM-SA model: A deep gated LSTM network with self-attention mechanism for power generation and load forecasting[J]. PloS one, 2026, 21(6): e0350071.
27. Jing, S.; Xi, X.; Su, D.; Han, Z.; Wang, D. Spatio-Temporal Photovoltaic Power Prediction with Fourier Graph Neural Network[J]. Electronics, 2024, 13, 4988.
28. Zi, X.; Liu, F.; Liu, M.; Wang, Y. Transformer with Adaptive Sparse Self-Attention for Short-Term Photovoltaic Power Generation Forecasting[J]. Electronics, 2025, 14(20): 3981.
29. Pan, C.; Liu, Y.; Oh, Y.; Lim, C. Short-Term Photovoltaic Power Forecasting Using PV Data and Sky Images in an Auto Cross Modal Correlation Attention Multimodal Framework[J]. Energies, 2024,17, 6378.
30. Xu, P.; Zhang, S.; Luo, P.; Wang, X.; Yan, Y.; Qi, D. Data-feature hybrid optimization for ultra-short-term wind power forecasting[J]. International Journal of Electrical Power & Energy Systems, 2026, 176: 111742.

Author Contributions: J.Z. conceived the algorithmic model of this paper, wrote part of it, and conducted comparison experiments with representative algorithms and performed data analysis. Y.Z. conducted the ablation experiments and analyzed the data. X.C. determined the research direction and wrote some of the content. Y.Y. wrote some chapters and made the final revisions. Y.X. created the figures and performed the paper search. All authors have read and agreed to the published version of the manuscript

Funding: This research was funded by Liangshan Prefecture Science and Technology Plan Project(2024YYYJ0029);Extracurricular Scientific Research Programs for Undergraduates at Xichang University(2026KW028).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article. Our code link is ghrepo clone zjs210/wind-power-forecasting

Coflicts of Interest: The authors declare no conflict of interest.