EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

A wind power prediction 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, Electrical and Aerospace Engineering, Xichang University, Xichang Sichuan, China

Abstract

In the development of green energy, wind power, as the core clean energy, has the inherent intermittency, volatility and randomness of its output, which is very easy to cause the power imbalance of the energy system and the surge of grid accommodation pressure, seriously restricting the safe and stable operation of the energy system. This study proposes the CEEMDAN-LightGBM-Transformer model. First, Complete ensemble empirical mode decomposition with adaptive noise(CEEMDAN) decomposes the data into multiple intrinsic mode function(IMF), and uses the "drop_k" strategy to select IMF components with prominent feature information and low redundancy, effectively stripping the noise and invalid fluctuations in wind power output data, and consolidating the foundation for accurate analysis of energy data. Then, Light gradient boosting machine(LightGBM) is used to construct nonlinear features, which are combined with strong correlation linear features to rank the importance of features, and accurately explore the correlation between wind power output and multidimensional factors. Finally, the multi-dimensional features are fused using a Transformer with an Encoder-Decoder architecture to generate a low-error power prediction sequence. The experimental results show that compared with multiple classical baseline models such as random forest, the mean value of Root Mean Squared Error (mRMSE) and mean value of Mean Absolute Error(mMAE) of this model are reduced by 61%, and the R^2 value is more than 98%, which has significant advantages in prediction performance. In addition, CEEMDAN, LightGBM and Transformer have shown good synergy in enhancing the performance of the model, which can effectively cope with the complex dynamic operating characteristics of wind power systems. The accurate prediction model can effectively stabilize the fluctuation risk of energy system caused by wind power integration, and provide reliable technical support for the grid dispatching of new power system.

Keywords: Power prediction, Mode decomposition, Gradient boosting, Self attention mechanism

Received on 27 August 2026, accepted on 12 November 2026, published on 16 November 2026

Copyright © 2026 Jingsong Zhang *et al.*, licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ew.14761

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 600 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 seriously affect the stability and economy of 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

*Corresponding author. Email: xcc20230015@xcc.edu.cn

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 et al [11] integrated the chaos method with the cauchy distribution to find the optimal parameters of seagull algorithm and long short-term memory(LSTM) model, subsequently designing an interval prediction model combined with kernel density estimation based on data augmentation theory. Mo et al[12] constructed the Powerformer model on the foundation of Transformer, enhancing the model's capacity for temporal feature extraction while reducing its complexity. Karijadi et al[13] combined CEEMDAN and empirical wavelet transform(EWT) to denoise high-frequency data sequences, and then employed LSTM to predict and output the sub-sequences. Li et al[14] utilized sample entropy to classify the modal information decomposed by variational mode decomposition (VMD), and constructed a dual-channel architecture incorporating temporal convolutional network(TCN) and Informer models to extract feature information from modal groups of varying complexity. Woldegiyorgis et al [15] used artificial neural network (ANN) and Weibull probability distribution function (PDF) to study the wind power. The artificial neural network is closer to the actual wind speed, and achieves better results. Li et al[16] used empirical mode decomposition(EMD) to process data, and introduces convolution mechanism into the transformer to reduce the prediction error. Qazi et al[17] proposed a gated recurrent neural network structure based on Tree-structured Parzen Estimator (TPE), which has fast convergence speed and low loss in the training of prediction model. Hussain et al [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. Ahsan et al [19] converted SCADA variable data into multi-channel gray-scale images and optimized the LeNet-5-LSTM hybrid neural network to achieve active and reactive power prediction. Venkateswaran et al [[20] used CEEMDAN decomposes the original wind power series into multiple IMF components. The multi-head self-attention mechanism of Transformer captures long-range global temporal dependencies, while LSTM supplements local temporal features, thereby reducing model errors.

The innovative aspects of this study are as follows:

(i) CEEMDAN is employed for wind power data decomposition. To overcome the subjective segmentation drawback of traditional modal decomposition, a noise indicator is introduced to quantitatively differentiate the predictive performance of each 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 approach achieve optimal choice 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.

(ii) 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.

(iii) 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, sinusoidal position coding enables the model to obtain periodicity of data series, 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 Decomposition

CEEMDAN [21] is an improved variant derived from EMD [22]. 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 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. The CEEMDAN algorithm flowchart is shown in Figure 1.

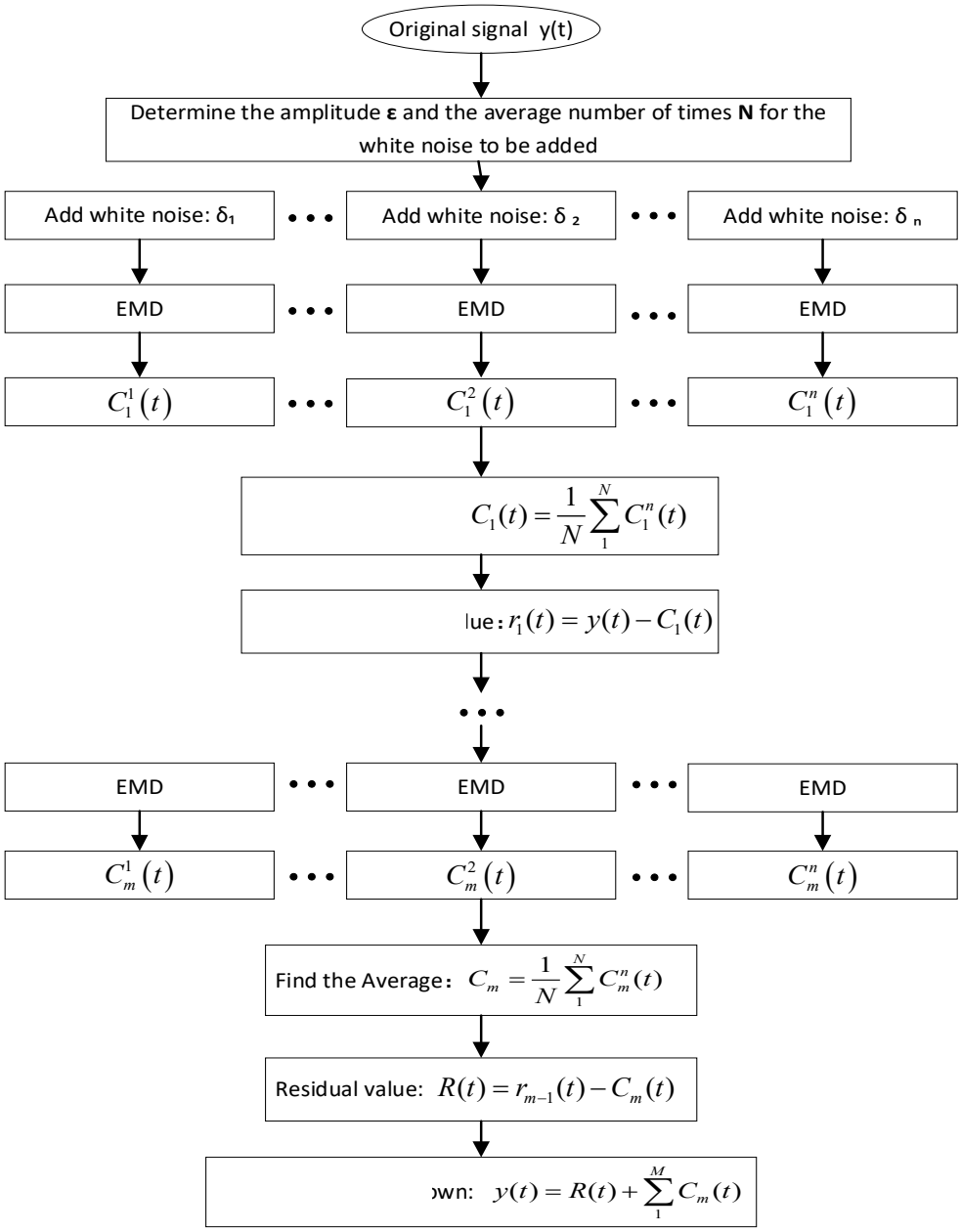


Figure 1. CEEMDAN algorithm structure

In Figure 1, y(t) is the original signal,$\varepsilon$is the amplitude of Gaussian white noise, N is the number of additions, $\delta_n(t)$ is the Gaussian white noise after the nth processing, $C_1^n(t)$ is the IMF components, $C_1(t)$ is the average of IMF components, $r_1(t)$ is the residual sequences, $C_m(t)$ is the final mean value of IMF components, $R(t)$ is the final residual sequence.

2.2. LightGBM Multidimensional feature reconstruction and selection

LightGBM [23] constitutes an optimized evolution of the Gradient Boosting Decision Tree (GBDT) [24]. 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 main implementation procedures of the LightGBM algorithm are detailed as follows:

(i) 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.

(ii) 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.

(iii) 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 2.

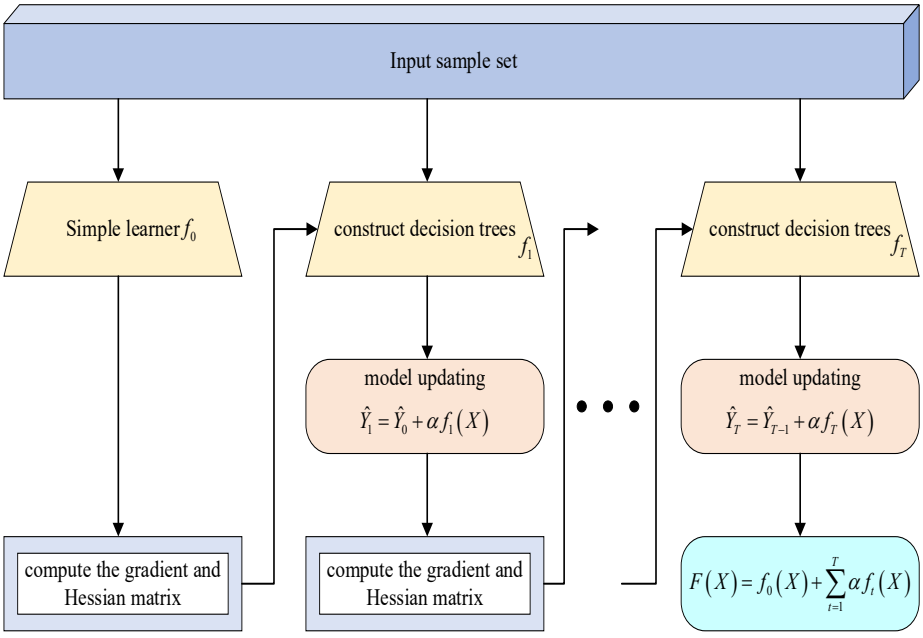


Figure 2. LightGBM algorithm structure

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, 70m wind direction, and 70m 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 for iterative training and preliminary wind power prediction.

2.3. Transformer Power Sequece Forecasting

The transformer [25] architecture mainly comprises an input layer, an Encoder–Decoder module [26], 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 (1) and (2), is adopted to inject positional embedding vectors into the feature matrix, which explicitly characterizes the temporal order of the sequential samples.

$$PE_{(pos,2i)} = \sin\left(pos/10000^{2i/d_{model}}\right) \quad (1)$$

$$PE_{(pos,2i+1)} = \cos\left(pos/10000^{2i/d_{model}}\right) \quad (2)$$

In Equations, pos denotes the time step positions within the historical sliding window, i is the feature dimension index, d_{model} is the dimension of the model's hidden layer.

Future-information risk in time-series forecasting originates jointly from input construction, prediction targets and training protocols, and cannot be simply attributed to encoder-only architectures. In many one-shot multi-step wind-power prediction implementations, encoder-only Transformer takes a concatenated window containing historical observations and future prediction placeholders as input. Even with chronological dataset splitting, bidirectional self-attention allows each token to access all positions within the input window. This is not explicit leakage of ground-truth future values, but implicit causal violation in attention computation, which may produce over-optimistic offline metrics and deteriorate online inference performance.

To address this causal violation, this study adopts a masked Encoder-Decoder framework. The encoder models fully-observed historical time-series and outputs high-dimensional semantic representations via self-attention [27] and feed-forward network (FFN) [28]. The decoder leverages masked multi-head self-attention to block future-step tokens and strictly enforces temporal causality. Meanwhile, cross-attention realizes cross-dimensional coupling between encoder historical representations and autoregressive decoding outputs. Residual connections and layer normalization are applied to generate prediction results complying with real-world temporal evolution. This autoregressive scheme better matches the rolling forecasting requirement of practical wind-farm dispatching, compared with the one-shot inference of encoder-only structures. Its structure is shown in Figure 3.

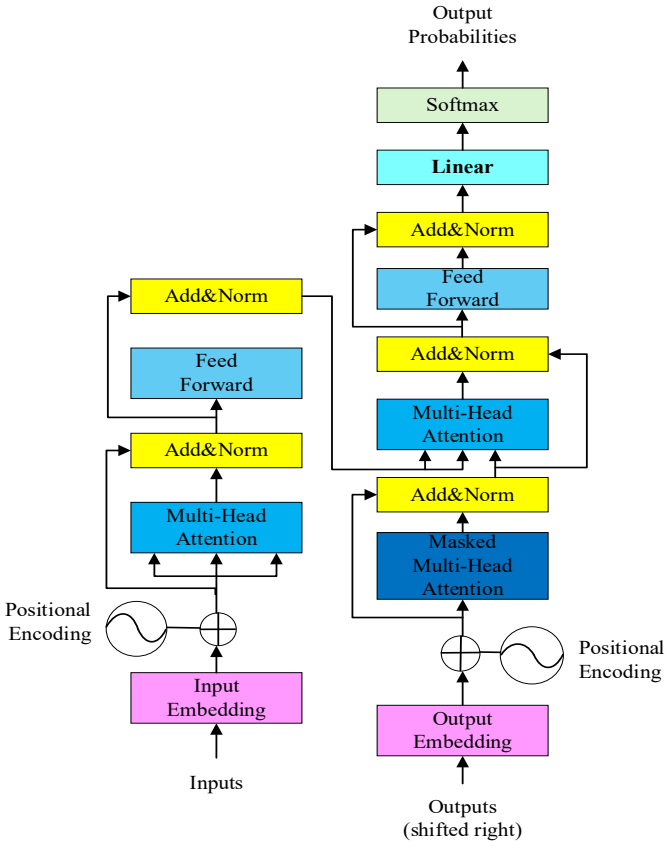


Figure 3. Transformer predicting network structure

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 [29] to replace the conventional single self-attention structure. This modification enables the model to get more comprehensive modal feature

representations, and the specific implementation formulation of the MHA mechanism is defined in Equation (3).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

In Equation (3), 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 4.

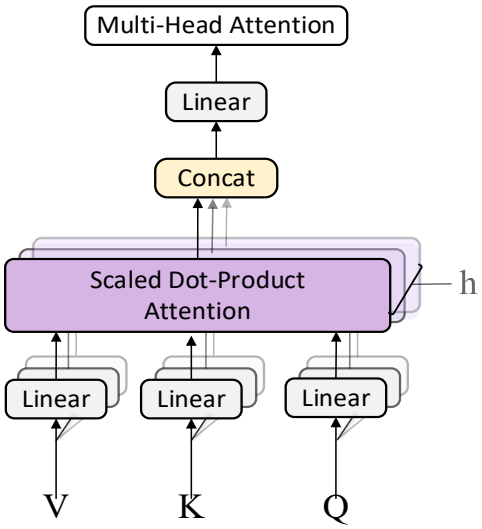


Figure 4. Multi-head attention mechanism architecture

Furthermore, the vanilla self-attention module inside the decoder is substituted with masked multi-head self-attention and cross-attention modules [30]. 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 is NVIDIA GeForce RTX 2080 16GB; the selected software is Pycharm 2026、Anaconda.

3.2. Test Data

The experimental data from 200MW wind farm in China from January 1 to December 31, 2019. Sampling SCADA/EMS for data statistics, and the sequence sampling interval is 15 minutes. After the data set is cleaned and filtered to eliminate outliers, it is divided according to time series, with the first 80% as the training set, the middle 10% as verification set, and the last 10% as test set. The sequence of samples remained unchanged throughout the process. Data set format is shown in the 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)	Temperature (oC)	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 coefficient analysis

In this paper, the raw data were analyzed using Pearson's correlation coefficient [31]; the correlations between the various input factors and actual power generation are shown in Figure 5.

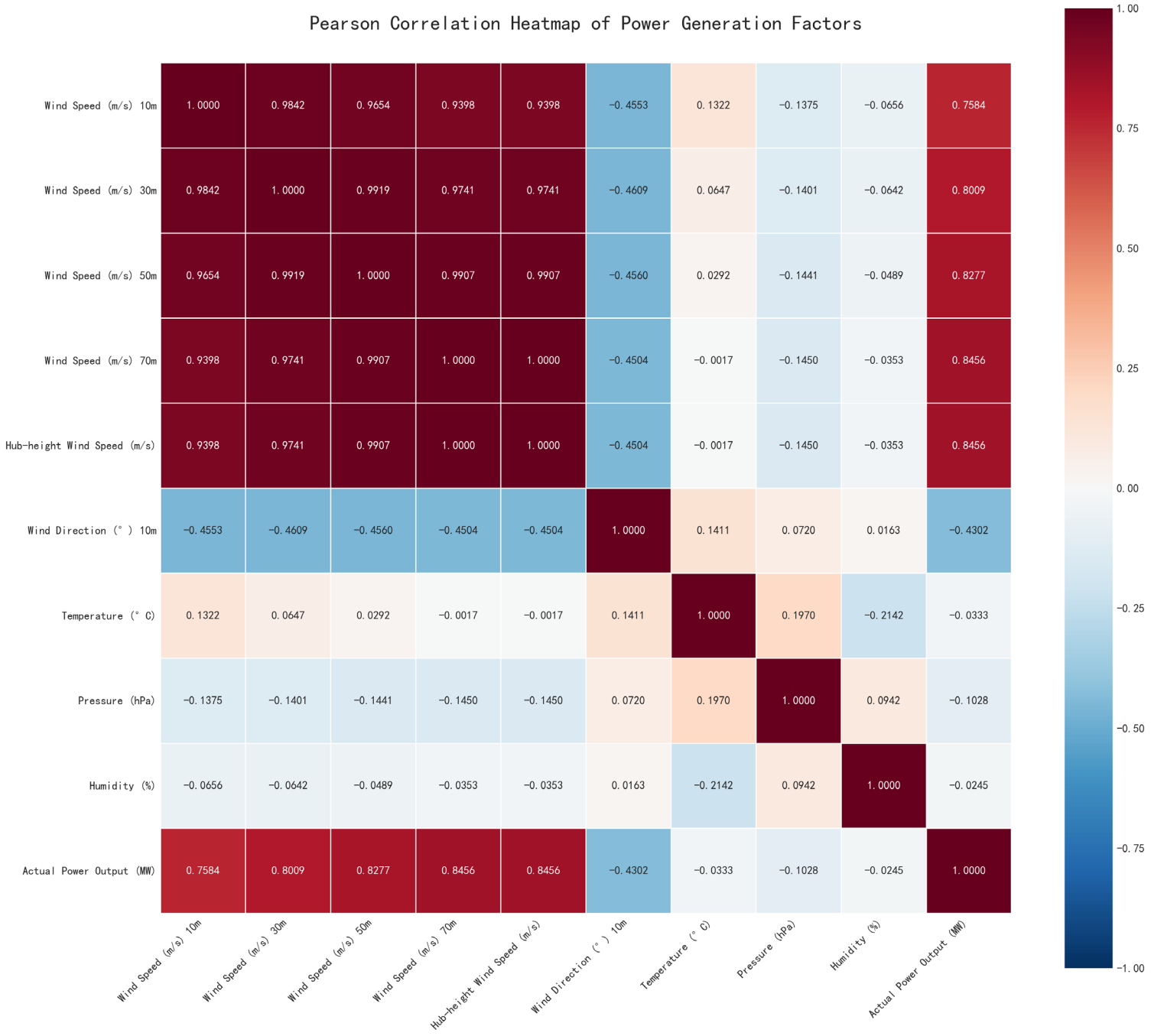


Figure 5. Data pearson correlatin coefficient heatmap figure

As illustrated in Figure 5, 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.

3.2.2. Mode decomposition

In this section, the cleaned and filtered wind power time series data undergoes CEEMDAN decomposition, resulting in 8-dimensional IMFs, as shown in Figure 6.

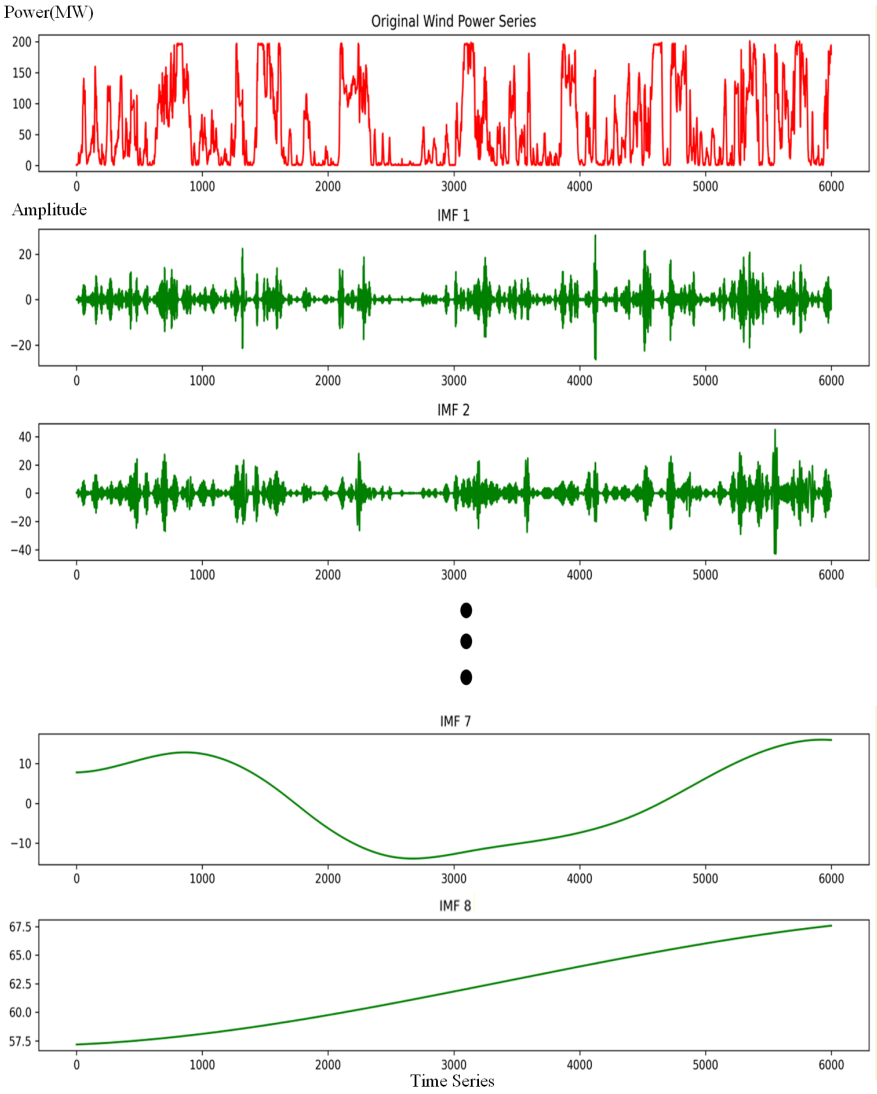


Figure 6. Decomposition results of CEEMDAN

Wind power data has 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 7.

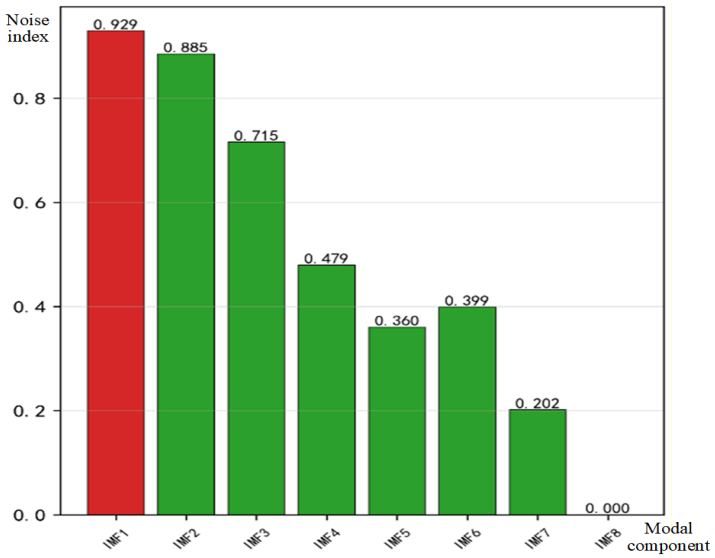


Figure 7. Noise index of each IMF

As can be seen from Figure7, 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.

Unlike IMF screening methods such as sample entropy and mutual information criterion, which require complex iterative calculations for each component, this paper adopts the noise index as the quantitative judgment criterion. It features low computational overhead and is more suitable for short-term wind power prediction scenarios sensitive to power fluctuations.

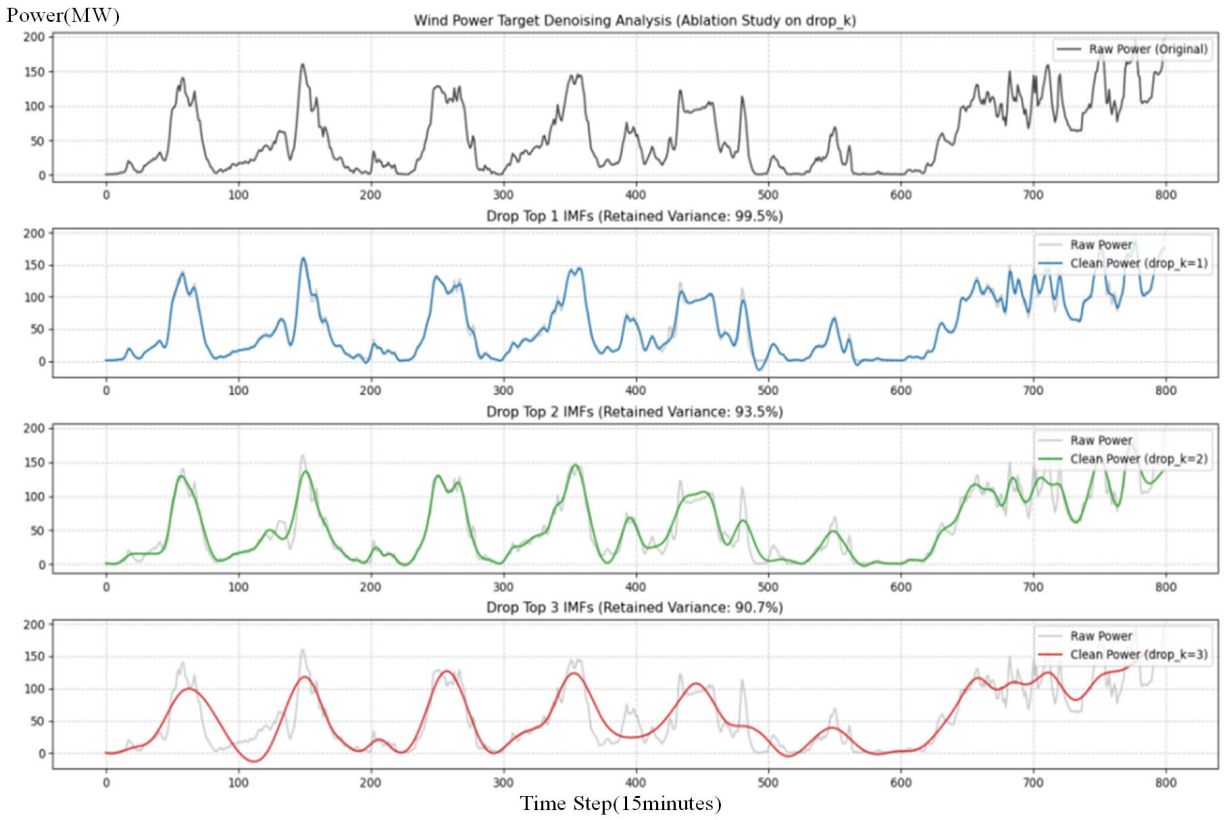


Figure 8. Data denoising results

As illustrated in Figure 8, 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 additional IMFs($\mathrm{drop_k} = 2\mathrm{or}3$) 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 MAE, RMSE, Coefficient of Determination (R-Square, R^2), mMAE, mRMSE, and mean of 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 eXtreme Gradient Boosting (XGBoost) 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

Model	RMSE		mRMSE	MAE		mMAE	R^2		mR^2
	Training set	Test set		Training set	Test set		Training set	Test set	
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 parameters are optimal in terms of RMSE, MAE and determination coefficient R2 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 mR2 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 shown in Figure 9 for intuitive comparison.

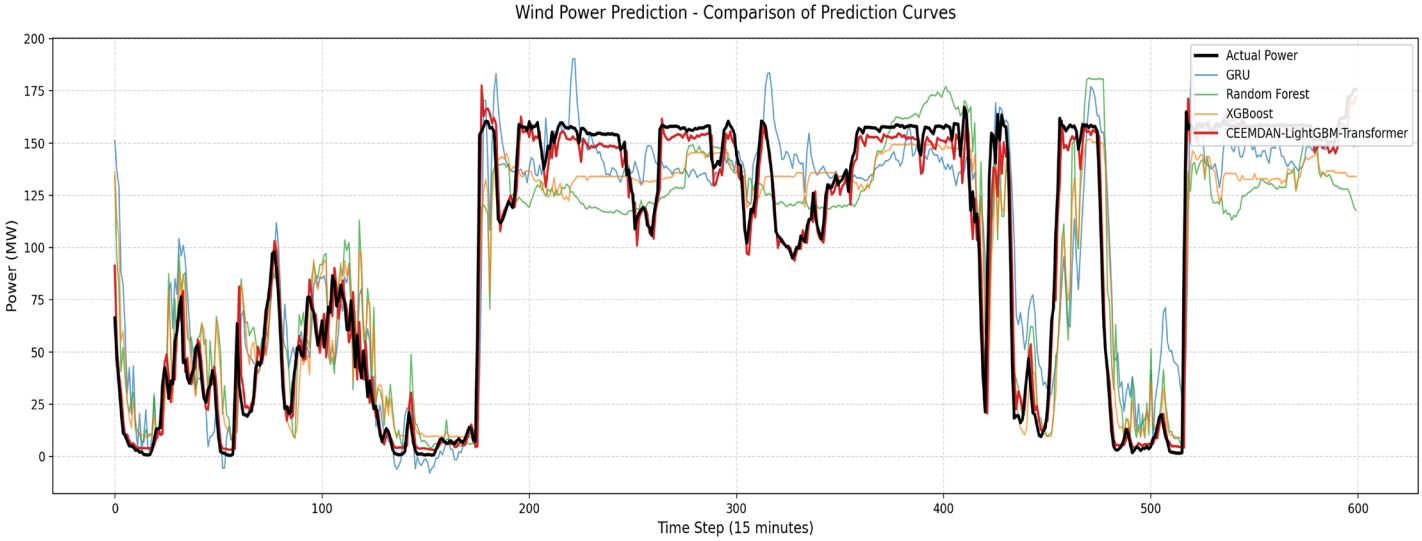


Figure 9. Comparison of prediction performance across models

As visualized in Figure 9, 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 prediction tasks.

3.3.2. Ablation Experiment

To quantitatively explore the performance 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), CEEMDAN-Transformer (CT), Transformer(T), 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 experimental results are shown in Table 3.

Table 3. Model Metrics and Parameters

Model	RMSE		mRMSE	MAE		mMAE	R^2		mR^2
	Training set	Test set		Training set	Test set		Training set	Test set	
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 R2 on both training and test datasets. Relative to the vanilla Transformer baseline, the mRMSE and mMAE of CLT drop by over 32%, whereas the mR2 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 the raw wind power time series in this dataset exhibits strong weather-driven nonlinear coupling. LightGBM-based feature reconstruction and filtering effectively extracts latent nonlinear correlations between meteorological variables and wind power output, yielding notable performance gains. CEEMDAN removes partial high-frequency random noise while fully preserving the intermediate-frequency oscillatory components that dominate prediction accuracy, 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. The line diagram of each model residual box is shown in Figure 10.

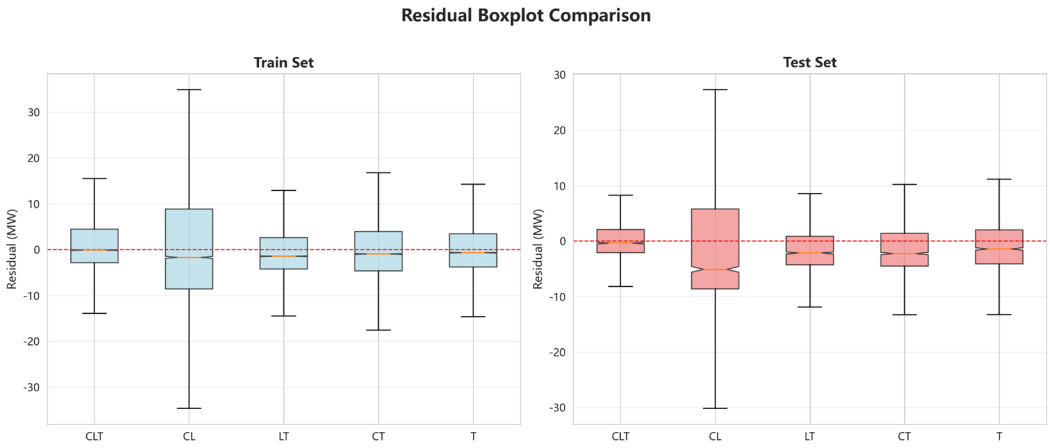


Figure 10. Residual box diagram of each model

The residual boxplots show that the CLT model achieves the optimal generalization performance, with residuals tightly distributed around zero and limited extreme errors on both training and test sets. The CL model has the most scattered residuals and unstable prediction. The LT, CT and T models perform moderately. The CLT hybrid model can effectively reduce wind power prediction errors and suppress extreme prediction deviations.

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 first 80% of the data set is still used as the training set, the middle 10% is the verification set, and the last 10% is the test set.lr=0.0005, Optimizer selects Adam. The quantitative experimental metrics of all competing models are tabulated in Table 4.

Table 4. Index Parameters of Each Mode

Model	RMSE		mRMSE	MAE		mMAE	R^2		mR^2
	Training set	Test set		Training set	Test set		Training set	Test set	
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. Conclusion

Firstly, CEEMDAN decomposed wind power time series data, and the optimal IMF component is selected through the "drop_k" strategy, which can effectively filter out the high-frequency noise while retaining the core effective signal, taking into account the noise reduction effect and signal integrity. Then,LightGBM reconstructed and adaptively filter the input data to obtain the core features and effectively reduce the feature redundancy. Then, Transformer extracted the deep time series characteristics through FFN to complete the power prediction. The experimental results show that, compared with other models such as random forest, the mRMSE and mMAE decreased by more than 61%, and the R2 can reach more than 99%; Ablation experiments further confirmed the effectiveness of CEENDAN noise reduction, lightgbm feature selection and transformer feature extraction, and verified the superiority of multi module fusion strategy.

Although this study has achieved excellent prediction results, there are still two core limitations. First, the application scenarios of the model are limited. Only single wind farm data is used for verification, and the multi region and multi model wind power scenario tests are not carried out. The generalization and migration ability of the model needs to be improved. The second is the lack of real-time and lightweight of the model. The integrated model in this paper uses a multi module series structure, which has larger parameter volume, more complex reasoning process, and higher calculation time than a single prediction model.It's difficult to adapt to real-time prediction of wind power, power grid real-time scheduling and other application scenarios with high response speed requirements.

In view of the above shortcomings, the following training tests will be carried out by using the data of multi region and multi model wind farms in combination with migration learning to optimize the model generalization and migration performance. The lightweight upgrading of the model is realized by means of structure simplification, parameter adaptive optimization and parallel operation transformation, which improves the real-time prediction model performance and adapts to the dynamic dispatching requirements of the power grid. Build an integrated intelligent forecasting and dispatching system based on the load characteristics of the power grid, and further promote the efficient, intelligent and large-scale application for wind power.

Funding

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

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] Chaudhary, K.A.;Guha, D.; Negi, R. Dynamic event-triggered recurrent neural network framework for synchronized frequency control of delayed power system[J].Journal of the Franklin Institute,2026,363(14): 108961-108961.

[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] Benbouhenni, H.; Colak, I.; Bizon, N. Backstepping control of multilevel modified SVM inverter in variable speed DFIG-based dual-rotor wind power system[J].Transactions of the Institute of Measurement and Control,2025,47(10): 1917-1930.

[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] Woldegiyorgis, A. T.;Chaka, D.M.;Asmare, E.;Asmare, E.; Admassu, C.;Anose, A.F. Assessment of artificial neural network and Weibull distribution methods for wind power density prediction in Ethiopia[J].Environmental Research Communications,2026,8(8):085006-085006.
[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] Qazi, A. M.;Chiang, D. H.Gated recurrent neural network enhanced wind power prediction accuracy using TPE Bayesian optimization[J].Results in Engineering,2026,PP: 32112136-112136.
[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] Ahsan, M.; Dao, B.P.Multi-Channel SCADA-Based Image-Driven Power Prediction for Wind Turbines Using Optimized LeNet-5-LSTM Hybrid Neural Architecture[J]. Energies,2026,19(5):1169-1169.
[20] Venkateswaran, R.; Venkateswaran, D.; Choi,W.;Jeong, J.H.Hybrid CEEMDAN – Transformer – LSTM Framework for enhanced multi-horizon wind speed forecasting[J].Journal of Wind Engineering and Industrial Aerodynamics, 2026, 272:106408-106408.
[21] 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.
[22] Lu, X.; Shao, J.; Lai, K. K.; Shi,Y.;.Chen, Q.; Cui, H. Bitcoin price forecasting with a deep hybrid EMD-CNN-GRU model[J].Soft Computing, 2026, 30(1):561-580.
[23] Randrianirina, S. F. M. J.; Rakotomalala, F. F. L.; Andriamahitasoa, A. B.;Bernard, A.A.;Zely, A.R.;Liva , G.R.Optimizing Wind Power Forecasting Using Machine Learning: A Comparative Study with Emphasis on LightGBM for Predictive Maintenance[J].International Journal of Advanced Engineering Research and Science, 2025,12(6):69-80.
[24] 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.
[25] Oriol, T.; Cifuentes, J.;Marulanda, G. Enhancing Wind Power Forecasting in the Spanish Market Through Transformer Neural Networks and Temporal Optimization[J]. Sustainability,2025,17(19):8655-8655.
[26] 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.
[27] 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.
[28] 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.
[29] 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.
[30] 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.
[31] Taheri, N.;Tucci, M. Enhancing Regional Wind Power Forecasting through Advanced Machine-Learning and Feature-Selection Techniques[J].Energies,2024,17(21): 5431-5431.