EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

Wind power prediction model integrating wavelet transform and CEEMDAN decomposition with Transformer-GRU network

Dianwei Chi[1], Chunyan Ji[1,*] and Yanqi Liu[1]

[1]School of Artificial Intelligence, Yantai Institute of Technology, Yantai 264000, China

Abstract

INTRODUCTION: Wind power is inherently unstable and randomly fluctuates, making its prediction complex and challenging, and it directly affects the stability of wind power grid-connected systems.
OBJECTIVES: This study developed a new wind power prediction technique that combines transformer encoding and gated recurrent unit (GRU) to improve prediction accuracy and stability.
METHODS: Wavelet transform denoised raw data. CEEMDAN decomposed series, extracting high-variance intrinsic modal functions (IMFs) as augmented features. Transformer encoded these temporal patterns; GRU modeled sequences for training and prediction.
RESULT: All key metrics of the proposed method have been improved, the root-mean-square error decreased to 0.069 MW, the mean absolute percentage error reduced to 18.260%, and the coefficient of determination increased to 0.975.
CONCLUSION: The model captures nonlinear dynamics effectively, offering a reliable forecasting foundation for wind farm operations and grid dispatch.

Keywords: wavelet transform, CEEMDAN, GRU, transformer, wind power, prediction

Received on 13 June 2026, accepted on 10 November 2026, published on 27 November 2026

Copyright © 2026 Dianwei Chi *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.16866

1. Introduction

In view of the intensifying global energy crisis and the urgent need for transformation of the energy structure, the development and utilisation of renewable energy, especially wind energy, has become the core strategy to promote sustainable development and alleviate energy tension. Wind energy has attracted widespread attention due to its abundant reserves, and clean and renewable nature. The continuous advancement of technology and effective cost control has led to a rapid expansion in the application range of wind power generation as an efficient and environmentally friendly energy supply mode, indicating its broad development space and significant application potential. However, the wind itself is characterized by random fluctuations and lack of stability, which makes the overall stability of the grid after the wind power grid has been greatly affected. This instability has led to the abandonment of wind power generation in many areas. It is imperative to enhance the precision of wind power prediction to ensure the stability of the power system subsequent to wind power integration. As the focal point of wind energy utilisation research, wind power prediction is associated with the seamless integration of wind power into the grid and the reliable operation of energy systems. The enhancement of forecast accuracy signifies not only technological advancement but also serves as a pivotal

*Corresponding author. Email: lufq5c73@163.com

strategy for promoting wind energy applications and optimising energy structures. Accurate power prediction helps wind farms to reasonably arrange operation and maintenance plans and optimize scheduling strategies, thus improving the power generation efficiency and economic benefits of wind farms.

Based on the basic theory of predictive modeling, wind power forecasting technology includes physical methods, statistical methods, deep learning methods, and hybrid methods. Existing reviews show that combined and data-driven approaches have become important routes for improving short-term wind power forecasting accuracy [1,2]. Recurrent neural networks, LSTM, and GRU have been widely used to capture temporal dependencies in wind power sequences [3,4], while DBSCAN-RFE-LightGBM and WT-BiGRU-attention-TCN models further demonstrate the value of feature selection, hybrid modelling architectures and attention mechanisms [5,6]. Transformer-based methods have recently attracted increasing attention because of their ability to model long-range dependencies in wind speed and wind power series [7,8]. In addition, decomposition-assisted forecasting frameworks, including EMD-related decomposition, secondary decomposition, and frequency-domain modeling, have been adopted to reduce non-stationarity and improve feature representation [9-11]. More recent studies have extended these ideas through PatchTST, causal feature selection, sparse downsampling interaction Transformer, multidimensional attention, dual-branch Transformer, inverted Transformer calibration, and context-aware GRU structures, showing the continuing shift toward multi-scale, attention-based, and hybrid forecasting models [12-18]. Nevertheless, existing single or hybrid models still face challenges in noise suppression, multi-frequency feature extraction, and joint long- and short-term temporal modeling. Therefore, this paper proposes a WT-CEEMDAN-Transformer-GRU hybrid prediction model for short-term wind power forecasting.

In summary, given the random fluctuations and instability characteristics of wind power, a novel WT-CEEMDAN-Transformer-GRU hybrid prediction model is proposed in this paper. The model integrates a variety of advanced technologies and aims to comprehensively improve the entire process from data preprocessing to forecast results output, thereby better adapting to the complexity and variability of wind power data and improving the overall performance of the forecast. Firstly, in the initial phase of data preprocessing, wavelet transform technology is employed for the purpose of efficiently reducing noise in wind power sample data. The primary objective of this stage is to enhance the signal-to-noise ratio of the signal; that is, to accurately eliminate noise components, thereby enabling the primary characteristics of the signal to be more distinctly exhibited. Then, the signal decomposition of wind power is performed based on CEEMDAN to obtain the IMF with a large variance contribution rate as the new feature incorporated into the sample. Next, the Transformer model is employed to accurately comprehend the long-term and short-term dependent characteristics of the time series. This approach facilitates the efficient exploration of time-related characteristics, thereby ensuring a reliable foundation for in-depth analysis. Finally, the GRU network is used for further time series modelling of the features extracted by the Transformer to further enhance the prediction model performance. The experimental results demonstrate that, in comparison with the five models evaluated in this study on the 2016 Sotavento test period (GRU, LSTM, WT-GRU, WT-Transformer-GRU and the proposed WT-CEEMDAN-Transformer-GRU), the model presented in this study achieves lower errors and a higher coefficient of determination. The innovation of this study is the integration of WT, CEEMDAN, Transformer and GRU to collaboratively solve the problems of data noise reduction, multi-band feature extraction and time series dependent modelling. Every one of these four blocks is an established technique; the contribution of this work is the specific way in which they are combined, together with the ablation study of Tables 6 and 7 that quantifies the contribution of each block on a common benchmark. This combination differs from the closely related recent studies as follows. Wang et al. [10] decompose the power series with CEEMDAN, regroup the intrinsic mode functions (IMFs) by sample entropy into a high- and a low-frequency group, forecast each group with a separate model (Transformer and BiGRU-Attention) and finally sum the component forecasts. Su et al. [19] add a second-stage variational mode decomposition of the high-frequency IMFs and again split the prediction between a Transformer and a GRU that run in parallel and are summed. In contrast, the present model uses decomposition for feature augmentation only: the three IMFs with the largest variance contribution (IMF9, IMF10 and IMF12, Table 3) are appended to the input sample and a single network predicts the total power directly, so that no component forecast has to be summed and no component-wise error is accumulated; the IMF selection criterion is the variance contribution rate rather than sample-entropy clustering. With respect to our earlier WT-BiGRU-attention-TCN model [6], which performs no signal decomposition at all, the recurrent backbone is replaced by a cascade in which a multi-head self-attention encoder first re-weights the 60-min input window and a GRU then models the encoded sequence, and the input is enriched with the CEEMDAN-derived IMFs.

2. Related work

2.1. Wavelet transform denoising

Wind power data may introduce noise during data collection from sensors and equipment, and transmission is also susceptible to factors such as signal attenuation. In addition, wind power measurements are also affected by weather conditions, temperature variations, and equipment failure and maintenance, thus generating noise. Denoising is necessary to provide high-quality data support for subsequent analysis and modelling. The employment of suitable denoising methodologies enables the precise identification and elimination of noise and anomalous data points within the dataset. This process enhances the data's integrity and

concomitantly augments the model's prediction accuracy and robustness.

Methods for denoising data samples include PCA, Fourier Transform, Wavelet Transform, and so on. In the context of non-stationary signal analysis, the wavelet transform has clear advantages because it can represent signal characteristics in both the time and frequency domains and perform multi-resolution decomposition [20]. Wavelet threshold denoising can suppress high-frequency noise while retaining the main signal structure, which provides a useful preprocessing basis for wind power forecasting [21]. Compared with continuous wavelet transform, DWT discretises time and frequency, has lower storage and computational costs, and is therefore suitable for processing large-scale wind power data.

The following are the specific steps for data denoising using DWT:

(i) According to the characteristics of wind power data, select suitable wavelet functions.
(ii) Threshold selection. Thresholding the high-frequency detail portion to remove noise. Soft thresholding, which is able to reduce coefficients smaller than the threshold to zero, makes the signal smoother after reconstruction based on this method.
(iii) Signal reconstruction. Its core is to take the high-frequency coefficient and the low-frequency coefficient that have been properly processed, and keep the original state, and reconstruct a clear signal that has been eliminated accurately through the process of inverse wavelet transformation.
(iv) The denoising effect is comprehensively evaluated using two core indexes: SNR and RMSE. The principle of SNR maximisation and RMSE minimisation is utilised to select the optimal wavelet transform function, ensuring that the quality of the denoised signal meets the prediction model's actual requirements.

2.2. CEEMDAN signal decomposition

CEEMDAN represents a more sophisticated signal processing method developed from EMD and ensemble empirical mode decomposition. EMD provides a basis for analysing nonlinear and non-stationary time series, while CEEMDAN introduces adaptive noise to improve modal decomposition and reduce mode mixing [19,22]. Improved CEEMDAN-related methods further enhance decomposition stability and are suitable for extracting multi-scale components from complex signals [23]. In this study, CEEMDAN is used to decompose the denoised wind power sequence and obtain IMFs with different frequency characteristics.

The present paper proposes the utilisation of CEEMDAN technology for the purpose of finely decomposing the original wind power data. The objective of this process is the extraction of subsequences that exhibit multiple time scales and frequency characteristics. This step is intended to optimise the quality of the input data and to improve the accuracy of the prediction model. The specific implementation process involves the introduction of white noise into wind power time series, multiple iterative decomposition, and averaging of sub-series.The specific formula can be seen as follows:

(i) Data preprocessing: wavelet transform technology is utilised to execute precise de-noising and smoothing procedures on the original wind power data. The objective of this process is to minimise the detrimental impact of noise and outliers on the subsequent decomposition process. The fundamental objective of this step is to establish a more precise and stable data foundation for the subsequent utilisation of CEEMDAN and other methodologies.
(ii) CEEMDAN decomposition: the pre-processed wind power data is utilised as the input, with the CEEMDAN method being employed for the purpose of deep decomposition. Following multiple iterations and the implementation of averaging processing, a set of intrinsic mode functions (IMF) and residual components is obtained. These components are indicative of subsequences of varying time scales and frequency characteristics within the original data. Consequently, they offer a comprehensive reflection of the fluctuation rules and trends inherent in wind power data.
(iii) Sub-sequence analysis and reconstruction of sample input data: the IMFs obtained from decomposition are analysed and the variance contribution size of each sub-sequence is calculated to assess its importance. In this study, multiple IMFs with larger contributions are selected as new features to be incorporated into the data samples as inputs to the subsequent prediction model to simplify models and improve predictive performance, and the new dataset not only retains the main original data features and better captures the random fluctuations and variations of wind power, providing more favourable input conditions for the subsequent prediction model.

2.3. Transformer encoding layer

The Transformer model [24] was proposed by Vaswani et al. in 2017 for sequence modelling tasks. It replaces recurrent structures with a self-attention mechanism, enabling efficient modelling of long-range dependencies. In time-series forecasting, Transformer variants such as the Pt-Transformer [25], Diffinformer [26], and Temporal Fusion Transformer [27] have also shown strong capability in long-sequence and multi-horizon prediction tasks.

The present study focuses on the encoder part of the transformer model in order to mine deep features from wind power data and generate new input sequences. The key components of the transformer encoder are the self-attention mechanism, the feedforward neural network, and the multi-head attention mechanism, which work together to efficiently extract core data information. Figure 1 illustrates the multi-head attention mechanism architecture, highlighting its central role in the transformer encoder:

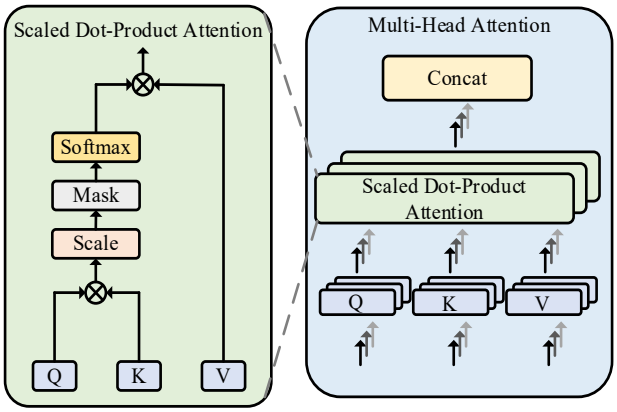


Figure 1. Structure of the multi-attention mechanism

In the encoder layer structure of the Transformer model, the multi-head attention mechanism plays a crucial role in ensuring the accuracy of the prediction results. The mechanism is composed of several independent attention heads, each of which exists as a concrete example of the self-attention mechanism, which is consistent with the scaled dot-product attention mechanism shown in the diagram, and is a practical application of the self-attention theory. The efficacy of this mechanism is evidenced by its enhancement of the model's capacity to capture and process input information, achieved through the utilisation of scaled dot products to evaluate the degree of similarity between the query and the key. It is evident that each self-attention mechanism is designed to consider the weighted contribution of each constituent element within the input sequence. Furthermore, it is able to accurately evaluate the importance ranking of each position within the sequence relative to the other positions. Following a suitable scaling adjustment, the calculated attention score is integrated with the value matrix by means of weighted summation. The specific description is as follows:

(i) Self-attention mechanism

The self-attention mechanism is contingent on the accurate measurement of the distribution of attention of each element in the input sequence X to the other elements. According to a specific algorithmic rule (formula 1), the mechanism calculates in detail the weight of attention due to each element in the sequence X. This process forms an indispensable cornerstone and guidance for the subsequent processing steps of the model:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

The query matrix, denoted Q, is employed to retrieve the input sequence information. The key matrix, denoted K, is used to match the definite correlation. The value matrix, denoted V, contains the output result information. The dimension of K, denoted dk, affects the computational efficiency and is pivotal to model design.

(ii) Multi-head attention

The Transformer architecture has been demonstrated to successfully extract multi-dimensional representation information at each location in the input sequence. This is achieved with the help of a multi-point attention mechanism, which is critical to enhancing model performance. Specifically, the architecture incorporates parallel attention processing units, each dedicated to analysing specific feature subspaces of the input data:

$$MultiHead(Q,K,V) = Concat(\text{head}_1, \text{head}_2, \ldots, \text{head}_h)W^O \quad (2)$$

where the calculation for each respective head is carried out as:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

In the multi-head attention mechanism, the head, which is also referred to as the i part, employs Q, K, and V to compute attention. Each head is equipped with W_i^Q, W_i^K,W_i^V for linear transformation, and the final output is processed by W^O.

(iii) Feedforward Neural Network

In addition to multi-head attention, the encoder layer incorporates a FFNN, which is employed to enhance the processing of input features and augment the model's expressiveness. The FFNN comprises two linear layers and ReLU activation function, with the objective of enhancing the model's generalisation capabilities. The specific computational procedures are elucidated in the subsequent section:

$$FeedForward(x) = max(0, xW_1 + b_1)W_2 + b_2 \quad (4)$$

where W1 and W2 are designated as the weight matrix, while b1 and b2 are assigned as the offset terms. The input vector data is denoted by x.

The main role of the transformer encoder used in this study is to perform feature extraction on the wind power data after wavelet noise reduction processing, i.e., to proficiently understand the complex interdependencies and sophisticated nonlinear traits evident in the wind power data through the application of the self-attention mechanism, which provides rich input information for the subsequent prediction model.

2.4. GRU

RNNs demonstrate the efficacy of temporal dependency modelling in sequence data, but their practical application is constrained by gradient disappearance. LSTM networks were proposed to address long-term dependency learning [28], while GRU provides a more compact recurrent structure with

fewer gates and parameters [29]. In this paper, GRU is used after Transformer encoding to further model temporal dependencies in the extracted features. The mathematical formula is as follows:

$$z_t = \sigma(\omega_z x_t + u_z h_{t-1} + b_z) \quad (5)$$

$$r_t = \sigma(\omega_r x_t + u_r h_{t-1} + b_r) \quad (6)$$

$$h_{c,t} = \tanh(\omega_k x_t + u_k(r_t \times h_{t-1}) + b_k) \quad (7)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times h_{c,t} \quad (8)$$

where x_t is defined as the input of time step t, and r_t is defined as the regulator of the influence of the previous hidden state on $h_{c,t}$. z_t is defined as the balancer of h_t information retention, h_t is defined as the merger of history with the current input, and h_{t-1} is defined as the former hidden state. ω is defined as the associated weight matrix, and b is defined as the adjuster of output.

Figure 2 presents a visual representation of the GRU network's structural layout.

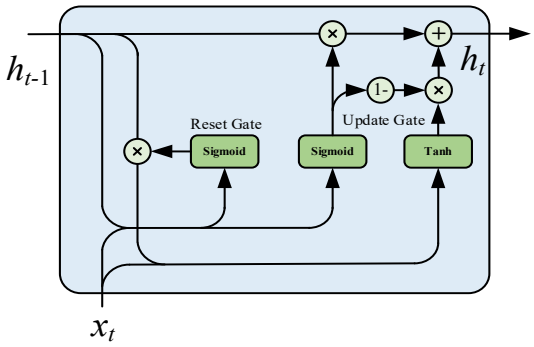


Figure 2. Network structure of GRU

2.5. Construction of the WT-CEEMDAN-Transformer-GRU model

This paper proposes a novel wind power prediction model, integrating WT, CEEMDAN, Transformer, and GRU. The operational process is illustrated in Figure 3, with the objective of achieving high-precision prediction.

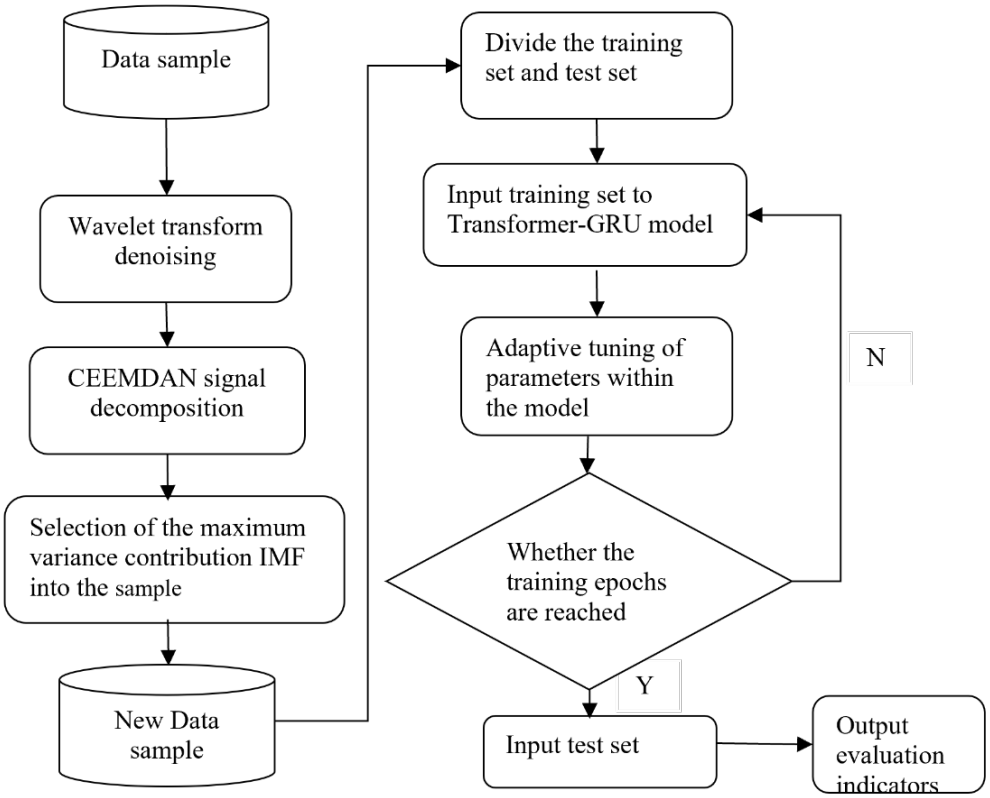


Figure 3. WT-CEEMDAN-Transformer-GRU model flowchart

The specific workflow of the proposed model is as follows:

(i) The present study employs wavelet transform technology to denoise wind power data samples, with the core objective being to preserve the signal's salient features and efficaciously eliminate noise components. This denoising technique has the capacity to substantially mitigate the mode aliasing problem that arises from signal decomposition utilising the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method, thereby enhancing the precision of signal decomposition.
(ii) Based on CEEMDAN, the signal decomposition of wind power feature data is carried out to obtain sub-series IMFs with different time scales and frequency characteristics, calculate the variance contribution of each sub-series, and select the three IMFs with the largest contribution values (IMF9, IMF10 and IMF12, Table 3) as new features to be incorporated into the original sample to form a new input data set, so as to better capture the fluctuations and changes of wind power.
(iii) In the construction of the model, the input data is processed by a transformer encoder, with the embedding layer converting the data into a vector. These vectors then pass through multiple layers of encoders, with each layer examining data features in depth. Notably, an encoder layer contains multiple attention mechanisms that accurately capture long-term dependencies and nonlinearities of wind power data. The feedforward network then reprocesses this data, refining the key features of the prediction.
(iv) The transformer encoder's output results are transmitted directly to the GRU network, which possesses its own distinctive gating mechanism and recursive architecture. This network enhances the extraction and refinement of intricate and multifaceted dynamic characteristics present within time series data.

3. Experimental results and discussion

3.1. Determination of data samples

This study is based on a wind power generation dataset from the Galicia region, Northwest Spain, which covers 52,123 precise recording points for the period from 1 January 2016 to 31 December of the same year, with data collected every 10 minutes. The dataset features include wind speed (WS), wind direction (DIR) and wind power (POWER), where POWER is used as the target forecast feature. As illustrated in Table 1, the sample statistics of wind power data are presented, thus providing the foundation for the construction of the model. The records originate from the Sotavento wind farm in Galicia, north-western Spain (43.354377°N, 7.881213°W), which comprises 24 wind turbines with a total installed capacity of 17.56 MW; its 10-min aggregates of wind speed, wind direction and total active power for the year 2016 are publicly released on Mendeley Data (Li Y. Data for: Short-term wind power forecasting approach based on clustering algorithm and Seq2Seq model using NWP data. Mendeley Data, V1, 2021, doi:10.17632/vtsgxnwswn.1) under a CC BY 4.0 licence [30]. Because 2016 was a leap year, a complete 10-min record would contain 52,704 samples. The released file contains 52,703 rows, of which 580 carry the missing-value placeholder "/"; after removing them 52,123 valid records remain, so 581 records are missing (one row is absent from the file and 580 rows are invalid). The invalid rows were removed rather than interpolated because the file carries no time column, so the length of each gap cannot be established; the retained series is treated as a regular 10-min sequence, which is the usual treatment of this benchmark. Consequently, when time stamps are attached to the samples (Supplementary Data S1) they are nominal stamps on a regular 10-min grid starting 1 January 2016 00:00.

Table 1. Statistical information of sample data

	WS (m/s)	DIR (degree)	POWER (MW)
Standard deviation	3.12	99.87	0.59
Minimum	0.35	0.00	0.00
Maximum	31.88	360.00	2.69
Mean	7.02	169.13	0.46
Coefficient of variation	44.44%	59.05%	128.26%

As demonstrated in Table 1, the coefficient of variation of power generation is as high as 128.26%, which clearly demonstrates its strong nonlinear characteristics and difficult-to-predict variability. Accordingly, the coefficient of variation of wind speed (44.44%) reflects the large and frequent fluctuations of wind speed, which once again confirms the core value and indispensability of wind speed in the process of wind power prediction modelling.Taking into account the highest and lowest recorded wind direction values, the provided wind data covers the entire range of directions, indicating that the wind may come from any direction, and its standard deviation is 99.87, which indicates a significant variation in wind direction with a high degree of stochasticity. Its coefficient of variation (59.05%) indicates that the wind direction exhibits strong relative dispersion, even higher than that of wind speed, although it remains lower than that of power output. This requires the prediction model to take into account the stochasticity and non-linearity inherent in the data. In addition, it can be seen that the three eigen scales are highly variable and require normalisation operations at the data processing stage.

3.2. Data preprocessing

To reduce the impact on predictive performance caused by differences in the scale of each feature within the dataset and the noise data in the data, it is necessary to do the following preprocessing of the dataset.

(i) In the data preparation stage, all the characteristic data are standardized. As demonstrated in Table 1, the amplitude difference between features is significant, and the range of data values is wide and diverse.Consequently, the normalization technique was applied to wind power sample data in this study, with the objective of assigning equal weights to all features during model training. The maximum-minimum normalisation method is employed in this study, with the data being adjusted to the interval [0,1] according to the formula presented in equation (9). This approach has been shown to significantly enhance the consistency and comparability of the data): It should be clarified that all features are used in their min-max normalised form during model training and evaluation, and the predicted values are restored to the original physical units by inverse min-max normalisation when visualised in Figures 4-6. All metrics are computed after the inverse min-max transformation, so the RMSE and MAE reported in Tables 5-8 are expressed in MW, MAPE in % and R2 is dimensionless. The min-max limits were estimated on the complete series before the chronological split; a training-period-only variant is reported in Section 3.6.

$$x'_t = (x_t - x_{min})/(x_{max} - x_{min}) \tag{9}$$

The limits of the maximum and minimum values of the feature sample data are set by x_{max} and x_{min}, respectively. The original data set is represented by x_t, while the data that has been transformed by maximum-minimum normalisation technology is identified by x'_t. This step ensures that all data are standardised to the interval [0,1], which facilitates the subsequent model training process.

(ii) Data denoising

The wavelet transform is a powerful tool for denoising wind power data. By employing multi-resolution analysis in the time and frequency domains, it can accurately differentiate between signal and noise. The original wind power data includes three features: wind direction, wind speed, and power output, and appropriate wavelet functions are chosen to decompose each feature, including Coiflet (coif), Symlet (sym), and Daubechies (db) wavelets. This paper proposes the use of DWT for the purpose of decomposing wind power data in detail, with the objective of achieving effective denoising. A global soft threshold is applied to the detail coefficients at every decomposition level; the exact threshold and level settings are those implemented in wavelet_denoise() in the deposited code (ceemdan_decompose.py regenerates the denoised series from the raw file). The denoised power series that is actually used in the experiments deviates from the measured series by 0.073 MW RMSE, corresponding to a signal-to-noise ratio of 20.25 dB, i.e. about 14% of the power variance is removed as noise. In order to accurately assess denoising performance, this paper proposes the utilisation of SNR and RMSE as metrics. Table 2 provides a comprehensive illustration of the denoising effect of each feature under various wavelet functions.

Table 2. Each feature variables denoising effect comparison

Feature name	Evaluation indicator	coif5	sym10	db8
WS	SNR/db	29.498	29.726	27.494
	RMSE	0.114	0.111	0.143
DIR	SNR/db	26.499	25.211	25.071
	RMSE	3.702	4.285	4.346
POWER	SNR/db	29.611	31.666	29.448
	RMSE	0.022	0.017	0.022

As demonstrated in Table 2, the application of sym10 wavelet processing results in enhanced SNR and reduced RMSE for the WS feature. Consequently, sym10 is designated as the wavelet function for the WS feature. A comparable outcome is observed for the DIR feature, which demonstrates optimal performance under coif5 wavelet processing, thereby substantiating its selection. For POWER features, the sym10 wavelet emerges as the preferred option due to its superior SNR and minimal RMSE, thus establishing itself as the wavelet of choice for POWER features.

Following the selection of the optimal wavelet function, the feature samples of wind power data are denoised by wavelet transform, thus enhancing the clarity of the data. Figure 4 demonstrates that the noise reduction effect is significant, and the verification method is both effective and practical. Wind direction is used as a single scalar feature expressed in degrees and min-max normalised. Because direction is a circular variable, this representation has a discontinuity at 0°/360°; the effect on the present data set is quantified in Section 3.6 and a sine/cosine encoding is listed as future work.

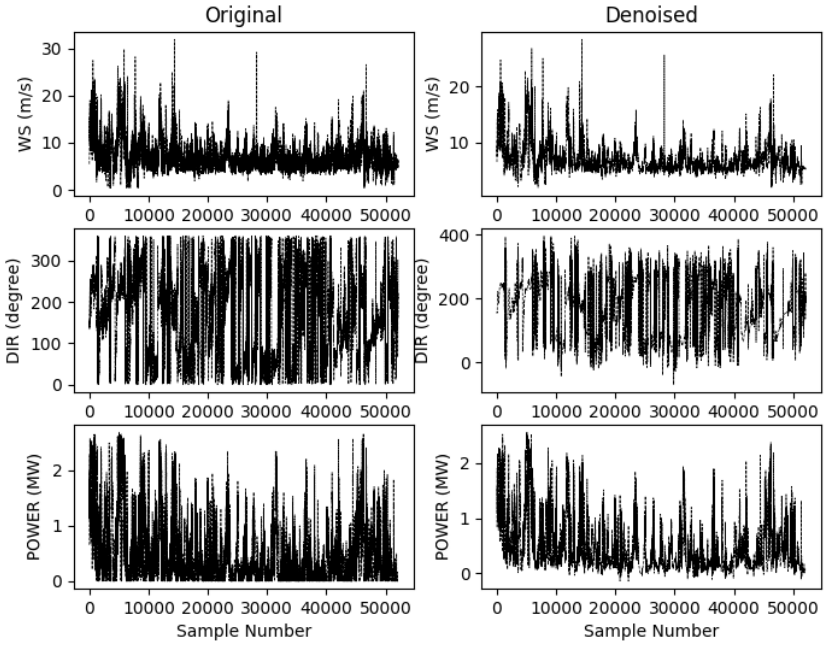


Figure 4. Before and after noise reduction of each feature variable based on wavelet transform

As seen in Figure 4, the signals of the three features of wind speed, wind direction, and power after denoising retain the main trends and variations of the original data, and the noise is significantly reduced. Among them, the original data of wind speed shows a lot of high-frequency noise, and after applying the wavelet transform, the denoised wind speed signal looks smoother, and the high-frequency noise is significantly reduced. The main trends and variations in wind speed are preserved, indicating that the wavelet transform effectively filters out the noise while maintaining the underlying signal. The power output of the original signal fluctuates significantly, exhibiting sharp peaks and valleys, and display noise, the causes of which are varied. These factors compromise the accuracy and reliability of the signal, thereby reducing the quality of subsequent analysis, feature extraction, and model construction.

3.3. Power signal decomposition and dataset reconstruction

In this study, CEEMDAN was utilised to analyse wind power signals in-depth, and IMFs with high variance contribution were selected as feature elements and integrated into the sample data set. The objective of this approach was to enhance the characterisation capability of feature sets and to optimise the accuracy of wind power signal prediction models.

Firstly, CEEMDAN decomposition of the denoised wind power signal is performed to obtain IMFs.The CEEMDAN-based decomposition of wind power generates 17 IMF components, which are denoted as IMF0~IMF16. Then, the variance of each IMF is calculated and its contribution to the prediction target is determined, and the variance contribution of the 17 IMFs is shown in Table 3.

The three IMF components with larger variance contributions were selected, and as seen in Table 3, IMF9, IMF10, and IMF12 had larger and closer variance contributions, so the selected IMFs were incorporated into the original sample as new features to form a new input data set. During the construction of the data sample, the feature set was expanded from the original three features to six features in order to enhance the expression effect of the data and to deepen the analysis dimension. This facilitates the effective capture of the main features of the power signal variations, thus providing better feature inputs for the subsequent wind power prediction model.

Table 3. Contribution of variance of each IMF component

IMF	0	1	2	3	4	5	6	7	8
contribution_rates	0.005	0.006	0.007	0.008	0.011	0.012	0.012	0.020	0.041

9	10	11	12	13	14	15	16
0.148	0.158	0.109	0.165	0.119	0.048	0.038	0.090

3.4. Description of model evaluation indicators

The predictive performance was evaluated using the RMSE, MAPE, and R2 indicators. Specifically, RMSE was calculated based on the square root of MSE, and used to quantify the difference between the predicted value and the actual value. MAPE was used to evaluate the accuracy of prediction by calculating the average ratio of prediction error to actual value. The value of R2 is considered to be a reliable indicator of the model's fit, with higher values indicating a more optimal fit. In addition, the mean absolute error (MAE, MW) is reported, because it requires no special treatment of zero or near-zero values. MAPE is undefined when the measured power is zero, and 21.31% of the 52,123 valid records are exactly 0 MW (the turbines are below cut-in or curtailed) while 35.65% are below 0.10 MW. MAPE is therefore computed with a lower bound on the denominator: every |P_true| < 0.10 MW is replaced by 0.10 MW in the denominator only, the measured value itself being left untouched (function mape_floor() in the deposited code). The bounded symmetric MAPE (SMAPE) and a MAPE restricted to the samples with |P_true| ≥ 0.10 MW are reported as sensitivity checks in Section 3.6.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{10}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\frac{|y_t^{pre} - y_t^{act}|}{y_t^{act}} \times 100\% \tag{11}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y_t^{pre} - y_t^{act})^2}{\sum_{t=1}^{N}(\bar{y} - y_t^{act})^2} \tag{12}$$

In the expression, N is used to represent the sample size, and $\bar{y}$ is the mean of the sample; y_t^{act} is the actual wind power value; $y_t^{pre}$is the predicted wind power value.

3.5. Comparative experiments and discussions

3.5.1 Experimental setup

It is imperative to acknowledge the significance of model parameters in ensuring the precision of short-term wind power prediction. Table 4 meticulously delineates the configuration of pivotal parameters, thereby ensuring the stability and reliability of predictive outcomes. To reduce the influence of randomness from parameter initialisation, each experiment in this study was independently repeated three times under identical settings, and the results are reported as mean ± standard deviation. The forecasting task is one-step-ahead: with a sampling interval of 10 min and an input window of 6 steps (60 min, "Time step = 6" in Table 4) the target is the power of the next 10-min step, so the forecast horizon is 10 min. GRU and LSTM are trained and scored against the raw measured POWER, whereas WT-GRU, WT-Transformer-GRU and WT-CEEMDAN-Transformer-GRU are trained and scored against the wavelet-denoised POWER of Section 3.2; both targets are given for every model in Supplementary Data S1. No hyper-parameter search was carried out: the architecture and the training hyper-parameters of Table 4 were fixed a priori (they are those adopted in [6]) and are identical for the five models, so that the comparison isolates the effect of the preprocessing and of the encoding stage. The only data-driven choice is the wavelet basis of each feature, made with the SNR/RMSE criterion of Table 2. Consequently no separate validation set was used, and the absence of one is listed as a limitation in Section 4. The min-max normalisation, the wavelet filter and the CEEMDAN decomposition were, in the published pipeline, computed on the complete series before the chronological split; a training-period-only variant is reported in Section 3.6.

Table 4. Model parameters

Parameters	Value
optimizer	Adam
GRU dimension	64
Dropout rate	0.1
Number of Transformer Blocks	1
Transformer head_size	64
Number of heads	4
Number of iterations	30 (50 for the LSTM baseline)
Time step	6 (8 for the LSTM baseline)
Batch_size	100
Learning rate	0.001
Proportion of training set	80%
Proportion of test set	20%

3.5.2 Effectiveness of the baseline model

This study aims to evaluate the performance of two kinds of recursive neural network-derived models, GRU and LSTM, in wind power prediction tasks. The original wind power sample data was selected for the study, and GRU and LSTM were chosen as the benchmark models for comparative analysis. A detailed hyperparameter configuration was carried out, as shown in Table 4. The experimental results, which are presented in Table 5, provide a comprehensive summary of the performance of the two models in different evaluation indicators:

Table 5. Baseline model experimental results (raw measured POWER for both models)

Method	RMSE	MAPE%	R2
GRU	0.119±0.003	41.794±0.15	0.931±0.002
LSTM	0.119±0.002	42.932±0.47	0.929±0.002

As demonstrated in Table 5, the GRU and LSTM models demonstrate comparable performance in terms of RMSE. However, when evaluated using MAPE, the GRU model exhibits a reduction from 42.932% to 41.794%, indicating a 2.65% relative decrease compared to the LSTM model. This finding underscores the GRU model's superiority in terms of forecasting accuracy. Concurrently, the marginal enhancement of R2 also mirrors the efficacy of the GRU model in data alignment, thereby demonstrating its superiority in achieving a more precise alignment.

3.5.3 Effectiveness of wavelet denoising

This part of the experiment uses GRU as the benchmark prediction model. Method 1 is a separate GRU model, and method 2 is a GRU model after adding wavelet denoising, recorded as WT-GRU. In the course of the experiment, a comparison was made between the GRU and LSTM models, with performance indicators obtained in accordance with the test procedures and evaluation criteria. The results of this study are summarised in Table 6, with the aim of facilitating subsequent analysis:

Table 6. Experimental results on the effectiveness of wavelet denoising (raw measured POWER for GRU, denoised POWER for WT-GRU)

Method	RMSE	MAPE%	R2
GRU	0.119±0.003	41.794±0.15	0.931±0.002
WT-GRU	0.082±0.002	20.630±0.14	0.955±0.003

As illustrated in Table 6, the findings demonstrate that in comparison with the unprocessed GRU model, the RMSE of the model with wavelet transform denoising decreases from 0.119 to 0.082, i.e., a 31.09% reduction in RMSE. In addition, the MAPE of the optimised model is significantly reduced in comparison with the original GRU model, from 41.794% to 20.630%, i.e., a 50.64% reduction. Furthermore, the coefficient of determination R2 demonstrated an upward trend, reaching 0.955, thereby substantiating the superior performance of the optimized model in data fitting. The experimental results demonstrate that wavelet transform denoising technology has a significant effect on the prediction accuracy of wind power. A total of 2,000 continuous data points were selected for prediction, with the denoised and undenoised models respectively predicting this data. Fitting curves (see Figure 5) were then drawn in order to visually demonstrate the difference in accuracy:

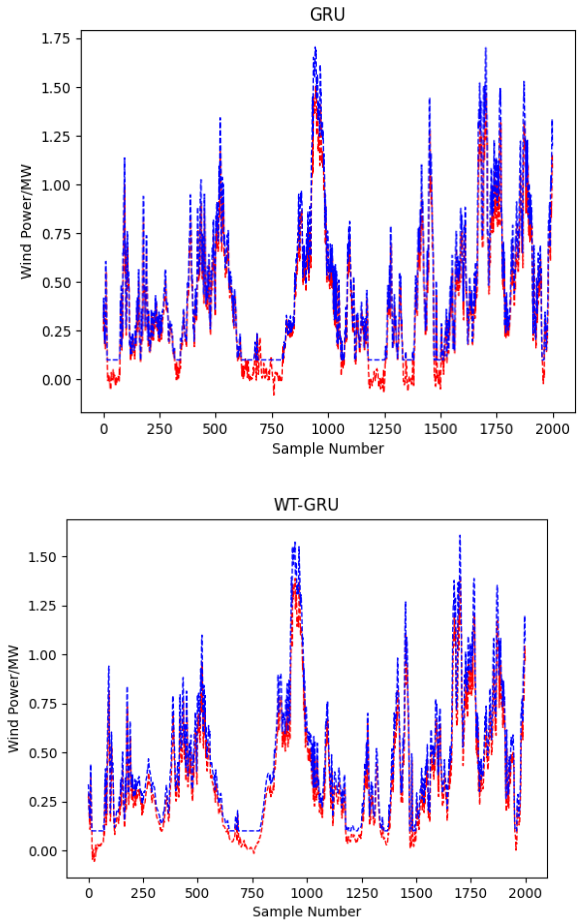


Figure 5. Fitting effect of WT-GRU and GRU models

The 2,000 consecutive test samples plotted are test-sample indices 5,417-7,416 of the 10,418 test samples (absolute records 47,121-49,120 of the 52,123 valid records; nominal time stamps 2016-11-23 05:30 to 2016-12-07 02:40), the slice testPredict[-5001:-3001] of the deposited code.

As demonstrated in Figure 5, the predictive performance of the model has been significantly enhanced following the integration of wavelet denoising technology. This enhancement is characterised by a substantial improvement in the degree of agreement between the predicted and actual values. This finding provides substantial evidence that the model's performance, subsequent to noise removal, exhibits superiority in capturing the dynamic variations in data and mitigating the impact of noise.

3.5.4 Effectiveness of the proposed model

This section will conduct a series of cross-sectional comparison experiments between WT-GRU-based predictive models. The objective of these experiments is to evaluate the effectiveness and excellence of the proposed models. In particular, WT-Transformer-GRU aims to validate the advantages of Transformer in improving the predictive performance of the model. The model that integrates CEEMDAN for wind power signal decomposition and Transformer encoder processing is recorded as WT-CEEMDAN-Transformer-GRU to verify that the power signal decomposition and data set reconstruction based on CEEMDAN can further improve the model prediction effect. As illustrated in Table 7, a horizontal comparison of the three prediction models reveals a range of metrics, including prediction accuracy, computational efficiency and stability.

Table 7. Horizontal comparison test results of each model (denoised POWER for all three models)

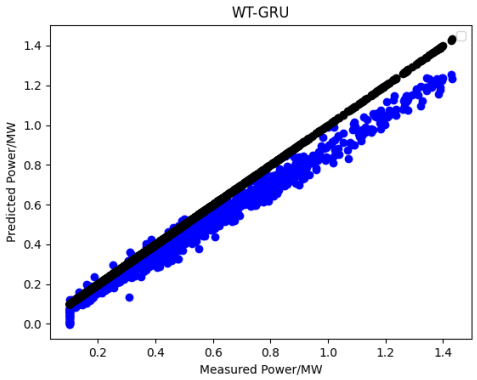
Method	RMSE	MAPE%	R2
WT-GRU	0.082±0.002	20.630±0.14	0.955±0.003
WT-Transformer-GRU	0.076±0.002	20.934±0.16	0.968±0.002
WT-CEEMDAN-Transformer-GRU	0.069±0.001	18.260±0.12	0.975±0.001

As shown in Table 7, the prediction model accuracy is further improved after Transformer Encoder encoding and the RMSE is reduced by 0.006, that is, a 7.32% reduction in RMSE compared with WT-GRU, and the MAPE index is slightly changed. The determination coefficient R2 increases from 0.955 to 0.968, that is, a relative improvement of 1.36%. The experimental results obtained in this study provide robust evidence to support the hypothesis that the Transformer Encoder is indeed an effective model. During the coding process, Transformer

Encoder demonstrated a high level of accuracy in capturing the short-time-dependent characteristics between features. This significant improvement in the modelling ability of complex time series data is a key finding of this study. The multi-head attention mechanism proposed in this study adopts a parallel learning approach to comprehensively explore the internal relationship between features. This approach has been shown to be successful in the mining of more dimensional information and potential rules.

Based on the WT-Transformer-GRU model, if the signal decomposition and data set feature reconstruction based on the CEEMDAN method for wind power characteristics are added, as demonstrated in Table 7, the experimental findings indicate a substantial enhancement in the model's prediction accuracy. The RMSE is reduced by 0.007, i.e., a 9.21% reduction in RMSE compared with WT-Transformer-GRU. The MAPE diminishes by 2.67 percentage points (from 20.934% to 18.260%), thereby substantiating the enhanced accuracy. Concurrently, the R2 determination coefficient exhibited a marginal increase, signifying enhanced data fitting and elevated prediction stability and reliability. This shows that the IMF after signal decomposition of power is integrated into the data sample as a new feature, which is equivalent to providing the model with richer feature information. The IMF with a large variance contribution represents the main change trend and characteristics in the wind power signal. The model can capture the change law of wind power at different time scales and frequencies during training. In particular, when combined with the GRU model, Transformer has been shown to be capable of more effectively extracting the temporal information and potentially complex relationships contained in new features. This combination strategy has been demonstrated to result in a deeper understanding of the essential features of time series data, as well as enhanced prediction accuracy and generalisation ability of the model by accurately capturing the long- and short-term dependencies between features. Consequently, this combination strategy has been shown to result in a significant improvement in the efficiency and quality of the overall prediction.

2000 time-continuous sample data were selected from the test sample, and predictions were made based on three comparison models respectively. Figure 6 depicts the graphs showing both the actual and forecasted values:



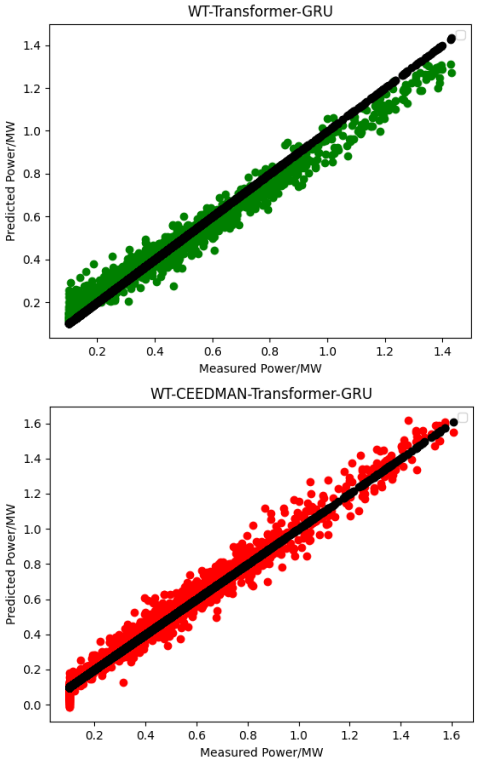


Figure 6. Fitting effect of each comparison model

The 2,000 consecutive test samples plotted are the same test-sample indices 5,417-7,416 (nominal time stamps 2016-11-23 05:30 to 2016-12-07 02:40); because the three decomposition-based models are built on the reconstructed file of 52,124 rows they are shifted by one record, i.e. absolute records 47,122-49,121 (2016-11-23 05:40 to 2016-12-07 02:50).

As illustrated in Figure 6, a comparison of the predicted results under three distinct wind models is presented. The horizontal coordinate denotes the actual wind power value, while the vertical coordinate represents the predicted wind power value. The black dots in the figure represent the actual observed wind power data, whereas the other coloured dots reflect the predicted results under the various compared models. It is evident from the figure that the red prediction point aligns closely with the black real point, thereby substantiating the model's efficacy in enhancing prediction accuracy and reliability.

In conclusion, the WT-CEEMDAN-Transformer-GRU hybrid model demonstrates excellent prediction accuracy and fitting performance under various test conditions. This validates the rationality of the model design and the efficacy of the components (WT, CEEMDAN, Transformer, GRU) in collaborating to enhance prediction efficiency and robustness. This provides a novel approach for time series prediction.

3.6. Additional evaluation and computational cost

GRU and LSTM are trained and scored on the raw measured POWER, whereas the training target of the three decomposition-based models is the wavelet-denoised POWER of Section 3.2, which differs from the measured POWER by 0.0732 MW RMSE. Table 8 rescores all five models against the raw measured POWER, using the same trained networks and the same predictions as in Tables 5-7. Because the target of GRU and LSTM is not affected, their entries are those already reported in Table 5; only the three decomposition-based models acquire different figures. On this common target the proposed model is the most accurate of the five, with an RMSE of 0.106 MW against 0.119 MW for GRU and LSTM, 0.136 MW for WT-GRU and 0.163 MW for WT-Transformer-GRU, and the highest coefficient of determination (0.957). Comparing the proposed model with its direct ablation isolates the contribution of the CEEMDAN stage on this target: adding the three IMF features lowers the RMSE from 0.163 MW to 0.106 MW and the MAPE from 22.410% to 20.374%.

Table 8. All five models scored against the raw measured POWER

Method	RMSE (MW)	MAPE (%)	R2
GRU	0.119	41.794	0.931
LSTM	0.119	42.932	0.929
WT-GRU	0.136	29.679	0.929
WT-Transformer-GRU	0.163	22.410	0.898
WT-CEEMDAN-Transformer-GRU	0.106	20.374	0.957

Computational cost. Table 9 reports the number of trainable parameters, the training time for the 30 epochs used in Section 3.5 (50 for the LSTM baseline) and the inference time, all measured on the same CPU. The proposed model has the largest number of parameters and the longest training time of the five. Its absolute cost nevertheless stays small: the 10,418 samples of the test set are forecast in 1.63 s, i.e. 0.156 ms per forecast, several orders of magnitude below the 10-min sampling interval, and the CEEMDAN decomposition is computed once and offline, outside the online forecasting step.

Table 9. Computational cost of the five models

Method	Trainable parameters	Training time (s)	Inference (ms/sample)
GRU	13313	44.5	0.082
LSTM	17473	80.5	0.082
WT-GRU	13313	41.0	0.076
WT-Transformer-GRU	82984	120.3	0.149
WT-CEEMDAN-Transformer-GRU	86681	122.1	0.156

Two further checks support the above conclusions. Fitting the preprocessing steps on the training period only, instead of on the complete series, changes every RMSE by less than 0.006 MW and leaves the ordering of the five models unchanged. Extending the horizon beyond the 10 min studied here to 30 and 60 min also preserves that ordering, the error of the proposed model growing more slowly with the horizon than that of any comparison model. Both experiments are reproduced by the deposited code and their output is deposited with the manuscript.

Finally, the comparison reported in Tables 5-8 is deliberately conservative for the proposed model. The LSTM baseline was configured with a look-back window of 8 steps, i.e. 80 min of history, and with 50 training epochs, whereas the GRU baseline, the two wavelet models and the proposed model all used the 6 steps and 30 epochs recorded in Table 4, which omitted this exception and has now been corrected. The proposed model was therefore given less input history and fewer training iterations than the baseline against which it is measured.

4. Conclusion

This paper proposes a novel strategy for wind power prediction, namely WT-CEEMDAN-Transformer-GRU. The strategy combines WT analysis capability, CEEMDAN noise reduction advantages, Transformer series learning, and GRU timing processing. It has been demonstrated that this combination can greatly improve prediction accuracy and stability, thereby providing a solid forecasting foundation for wind farm operation and power grid scheduling. Firstly, this study employs wavelet transform technology to conduct a detailed multi-scale analysis of wind power data, accurately distinguish signal from noise, and effectively remove high-frequency noise while maintaining the core characteristics of signals. Subsequently, the CEEMDAN method is employed to extract wind power data in deep multi-band, and IMF with significant variance contribution is selected as a new feature variable and integrated into the original data set, thereby constructing a new data set with higher information content and more value. The optimised features are subsequently entered into the Transformer encoder, with

the objective of capturing the long and short-term dependencies between features and unveiling the time series characteristics. Thereafter, the GRU is employed to enhance the time series modelling, thereby facilitating accurate wind power predictions through its memory capacity. Empirical evidence demonstrates that the proposed method enhances both the prediction accuracy and the fitting effect, thereby validating its efficacy and practicality. The findings of the research demonstrate that the hybrid prediction model outlined in this paper integrates multi-scale analysis, multi-band decomposition, Transformer coding, and GRU, and that the data processing chain is executed once per forecast origin and costs about 0.156 ms per forecast (Table 9). The model exhibits a profound insight into wind power time series dynamics, covers multi-scale fluctuations and time correlation, significantly improves prediction accuracy, provides a comprehensive and efficient scheme for wind power prediction, and is conducive to wind power resource utilisation and power grid stability.

The main conclusions are as follows:

(i) By performing multi-scale decomposition of wind power data based on the wavelet transform method, wavelet functions of different scales are used to capture the details and overall trends of the signal, effectively realizing signal-noise separation and achieving the effect of noise reduction. Compared with the single GRU model, the RMSE and MAPE of the denoised model are reduced by 31.09% and 50.64%, respectively, indicating the effectiveness of wavelet denoising.

(ii) Employing Transformer encoding can improve the accuracy of the predictive model. Compared with WT-GRU, the RMSE of the model is reduced by 7.32%, and the R2 is improved by 1.36%. The study demonstrates that the incorporation of Transformer Encoder technology into the model utilises its self-focused mechanism to accurately capture the intricate relationship between wind direction, wind speed, and power characteristics. This enhancement in modelling efficiency, evident in both long-term and short-term dependence models, signifies a substantial augmentation in the predictive capability of wind power time series. The enhanced accuracy, sensitivity, and adaptability of the model to variations in wind power characteristics pave the way for novel advancements in wind power forecasting technology.

(iii) The employment of the CEEMDAN decomposition method facilitates the precise capture of the primary characteristics inherent in wind power data. Subsequent to this, by selecting IMFs with significant variance contribution, it is possible to effectively reduce noise interference, improve the quality of model input features, and achieve a significant enhancement of prediction efficiency. The efficacy of this approach is evidenced by a 9.21% reduction in RMSE and a 2.67 percentage-point reduction in the MAPE index.

The wind forecasting model under scrutiny in this study has achieved notable results, suggesting that the model's prediction accuracy and stability have undergone substantial enhancement. Nevertheless, in order to further refine the model's efficacy and expand its scope of application, future research endeavours can be pursued from two fundamental directions. On the one hand, considering the insufficient richness of features in the original data samples of this study, integration of other data sources, such as meteorological and real-time sensor data, is considered in the future, which is conducive to improving the prediction by capturing more complex patterns. On the other hand, to improve the model's capacity for instant forecasting in practical wind prediction scenarios, the computational efficiency of the Transformer encoder and GRU components need to be further optimized. Meanwhile, the adaptive decomposition technique is deeply explored to dynamically adjust the decomposition scale and frequency range derived from live data to enhance computational efficiency and improve prediction precision. Two limitations of the present study should be stated explicitly. First, all results are obtained on a single wind farm;Second, no independent validation set was used and the preprocessing steps (normalisation, wavelet filtering and CEEMDAN) were computed on the complete series before the split; the robustness check of Table 9 shows that the two leading models remain within run-to-run variability of each other, but a strictly causal rolling decomposition should be implemented before the model is deployed. A comparison with more recent architectures (PatchTST, Informer and the CEEMDAN-VMD-SE family of [21]) is also left for future work.

Acknowledgements

The authors would like to thank Dr. Xingping Wang for his valuable guidance and insightful comments during the research process.

Data availability

The wind power data are publicly available from Mendeley Data (doi:10.17632/vtsgxnwswn.1) under a CC BY 4.0 licence [30].

References

[1] Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. Renewable Energy. 2012;37(1):1-8. https://doi.org/10.1016/j.renene.2011.05.033.

[2] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. Renewable and Sustainable Energy Reviews. 2014;34:243-254. https://doi.org/10.1016/j.rser.2014.03.033.

[3] Dong D, Sheng Z, Yang T. Wind Power Prediction Based on Recurrent Neural Network with Long Short-Term Memory Units. In: 2018 International Conference on Renewable Energy and Power Engineering (REPE). 2018:34-38. https://doi.org/10.1109/REPE.2018.8657666.

[4] Kisvari A, Lin Z, Liu X. Wind power forecasting - A data-driven method along with gated recurrent neural network. Renewable Energy. 2021;163:1895-1909. https://doi.org/10.1016/j.renene.2020.10.119.
[5] Wu Z, Fan X, Bian G, Liu Y, Zhang X, Chen YQ. Short-term wind power forecast with turning weather based on DBSCAN-RFE-LightGBM. Renewable Energy. 2025;251:123217. https://doi.org/10.1016/j.renene.2025.123217.
[6] Chi D, Yang C. Wind power prediction based on WT-BiGRU-attention-TCN model. Frontiers in Energy Research. 2023;11:1156007. https://doi.org/10.3389/fenrg.2023.1156007.
[7] Wu H, Meng K, Fan D, Zhang Z, Liu Q. Multistep short-term wind speed forecasting using transformer. Energy. 2022;261:125231. https://doi.org/10.1016/j.energy.2022.125231.
[8] Huang Q, Wang Y, Yang X, Im S-K. Research on Wind Power Prediction Based on A Gated Transformer. Applied Sciences. 2023;13(14):8350. https://doi.org/10.3390/app13148350.
[9] Huan J, Deng L, Zhu Y, Jiang S. Short-to-Medium-Term Wind Power Forecasting through Enhanced Transformer and Improved EMD Integration. Energies. 2024;17(10):2395. https://doi.org/10.3390/en17102395.
[10] Wang S, Shi J, Yang W, Yin Q. High and low frequency wind power prediction based on Transformer and BiGRU-Attention. Energy. 2024;288:129753. https://doi.org/10.1016/j.energy.2023.129753.
[11] Jin Z, Fu X, Xiang L, Zhu G, Hu A. Informer learning framework based on secondary decomposition for multi-step forecast of ultra-short term wind speed. Engineering Applications of Artificial Intelligence. 2025;139:109702. https://doi.org/10.1016/j.engappai.2024.109702.
[12] Xu S, Wang Y, Xu X, Shi G, Huang H, Zheng Y. A PatchTST-GRU based heterogeneous seq2seq model with numerical weather prediction refinement for multi-step wind power forecasting. Scientific Reports. 2025;15:16547. https://doi.org/10.1038/s41598-025-00741-9.
[13] Zhao H, Xu P, Gao T, Zhang JJ, Xu J, Gao DW. CPTCFS: CausalPatchTST incorporated causal feature selection model for short-term wind power forecasting of newly built wind farms. International Journal of Electrical Power & Energy Systems. 2024;160:110059. https://doi.org/10.1016/j.ijepes.2024.110059.
[14] Wang H, Guo D, Wang L, Zhou T, Jia C, Liu Y. A novel frequency sparse downsampling interaction transformer for wind power forecasting. Energy. 2025;326:136199. https://doi.org/10.1016/j.energy.2025.136199.
[15] Song W, Zuo E, Zhu J, Chen C. DTCMMA: Efficient Wind-Power Forecasting Based on Dimensional Transformation Combined with Multidimensional and Multiscale Convolutional Attention Mechanism. Sensors. 2025;25(15):4530. https://doi.org/10.3390/s25154530.
[16] Wu J, Chang Z, Zhang L, Chen M. A Dual-Branch Transformer Network with Multi-Scale Attention Mechanism for Microgrid Wind Turbine Power Forecasting. Electronics. 2025;14(13):2566. https://doi.org/10.3390/electronics14132566.
[17] Cohen J, Xu T, Jang Y, Du P, Byon E, Huan X. Inverted Transformers for Effective Post-Calibration in Multisite Wind Power Forecasting. Wind Energy. 2025;28:e70057. https://doi.org/10.1002/we.70057.
[18] Yasir J, Wang Y, Mutale S, Zakria M. Short and medium-term wind power forecasting using a novel time series dynamic context-aware GRU model. International Journal of Sustainable Energy. 2026;45(1):2651009. https://doi.org/10.1080/14786451.2026.2651009.
[19] 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. Energy. 2025;338:138715. https://doi.org/10.1016/j.energy.2025.138715.
[20] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989;11(7):674-693. https://doi.org/10.1109/34.192463.
[21] Liu Y, Liu J, Qi Y, Deng J, Wu P, Zhang K, et al. Wind Power Estimation Under Uncertainty Using Wavelet Transform for Optimal Dispatch in Microgrids with Renewable Energy Integration and Demand Response. Smart Grids and Sustainable Energy. 2026;11:8. https://doi.org/10.1007/s40866-025-00314-9.
[22] Sun H, Fu X, Luo Y, Wu Z, Peng C. A multistep wind power prediction model based on sample entropy clustering decomposition and inverted Transformer. Electric Power Systems Research. 2026;251:112286. https://doi.org/10.1016/j.epsr.2025.112286.
[23] Colominas MA, Schlotthauer G, Torres ME. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. Biomedical Signal Processing and Control. 2014;14:19-29. https://doi.org/10.1016/j.bspc.2014.06.009.
[24] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. In: Advances in Neural Information Processing Systems(NIPS 2017). 2017:5998-6008. https://doi.org/10.48550/arXiv.1706.03762.
[25] Zhang Z, Sun Z, Guo X, Guo R, Yang X, Yang P. Short-term multi-step wind power prediction model based on Pt-Transformer neural network integrating spatio-temporal feature and sparse attention. Electric Power Systems Research. 2025;248:111970. https://doi.org/10.1016/j.epsr.2025.111970.
[26] Li J, Chen W, Liu Y, Yang J, Zhou Z, Zeng D. Diffinformer: Diffusion informer model for long sequence time-series forecasting. Expert Systems with Applications. 2026;299:129944. https://doi.org/10.1016/j.eswa.2025.129944.
[27] Hou G, Liu P. Short-term prediction of wind power based on improved temporal fusion transformer model with enhanced hippopotamus optimization algorithm. Electric Power Systems Research. 2026;252:112464. https://doi.org/10.1016/j.epsr.2025.112464.
[28] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997;9(8):1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735.
[29] Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014:1724-1734. https://doi.org/10.3115/v1/D14-1179.
[30] Li Y. Data for: Short-term wind power forecasting approach based on clustering algorithm and Seq2Seq model using NWP data. Mendeley Data. 2021;V1. https://doi.org/10.17632/vtsgxnwswn.1.