Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2025.090310
ISSN 2371-8838 Vol. 9 Num. 3

Ultra-short-term Wind Power Forecasting Based on ICEEMDAN-Informed BiGRU Network with Multi-head Attention

Daoyuan Li[1,a,*], Yutang Ye[1,b]

[1]*College of Engineering, Northeast Agricultural University, Harbin, Heilongjiang, China*
[a]*ldy@neau.edu.cn*, [b]*18845781732@163.com*
**Corresponding author*

Keywords: Wind Power Forecasting; ICEEMDAN; Informer; BiGRU; Multi-head Attention Mechanism; Deep Learning

Abstract: To improve the ultra-short-term forecasting accuracy of wind power series, this paper proposes a combined forecasting model—ICEEMDAN-Informer-BiGRU-Attention—that integrates improved modal decomposition with deep learning modeling to address its strong non-stationarity and complex temporal structure. This model first decomposes the wind power series using an improved ensemble empirical mode decomposition (ICEEMDAN) algorithm to obtain multiple subsequences with clear frequency domain features and reduced volatility. It then employs an informer architecture to capture long-term dependencies within the series, introduces a bidirectional gated recurrent neural network (BiGRU) to model short-term dynamic features, and incorporates a multi-head self-attention mechanism to further enhance the representation of key time steps. Experimental results on a real wind farm dataset demonstrate that the proposed model outperforms mainstream models such as LSTM, GRU, and informer in terms of RMSE, MAE, and R ? Its predictions are closer to real data, with more concentrated residual distributions, resulting in optimal overall performance. This validates the model's effectiveness and engineering feasibility in wind power forecasting scenarios.

1. Introduction

Driven by the "dual carbon" goals and energy structure transformation, wind power, as a green and clean renewable energy source, is gradually becoming a vital component of the power system. However, wind power output is significantly affected by meteorological conditions, exhibiting strong volatility and non-stationarity, posing severe challenges to grid scheduling and stable operation. To improve wind power grid integration and power system flexibility, research on high-precision, ultra-short-term wind power forecasting is crucial.

In recent years, with the widespread application of deep learning in time series modeling, models such as LSTM, GRU, and TCN have achieved promising results in wind power forecasting [1]. However, these methods still face challenges such as model lag and information loss when dealing with typical nonlinear characteristics such as high-frequency fluctuations and sudden changes, limiting their forecasting performance [2]. Furthermore, data-driven models are highly sensitive to input data quality, and wind power series often contain noise, non-stationary trends, and multi-scale periodicity, further limiting the expressiveness and generalization capabilities of deep models [3].

To address these issues, this paper proposes a hybrid forecasting model that combines improved ensemble empirical mode decomposition (ICEEMDAN) decomposition, Informer encoding, BiGRU sequence modeling, and a multi-head attention mechanism. First, an ICEEMDAN is used to decompose the original wind power series into multiple, more stationary subsequences, enhancing the model's ability to perceive different frequency components [4]. Subsequently, an informer architecture is introduced to exploit long-term temporal dependencies, combined with a BiGRU to enhance the memory representation of historical context [5]. A multi-head self-attention mechanism is used to further optimize the representation weights of key time steps in the series, improving the responsiveness and discriminative accuracy of the prediction results [6].

This paper experimentally validates the model using real wind farm operating data and compares it with several mainstream models. The experimental results demonstrate that the proposed ICEEMDAN-Informer-BiGRU-Attention model achieves excellent performance across multiple evaluation metrics, effectively improving prediction accuracy and alleviating model lag. This research provides a highly robust and accurate solution for ultra-short-term wind power forecasting and offers new insights and approaches for modeling non-stationary complex time series.

2. Overall Structure of the Model

To improve the accuracy and stability of wind power forecasting, this paper proposes a combined model architecture based on the ICEEMDAN-Informed BiGRU multi-head attention network. This model fully integrates the advantages of signal decomposition, deep time series modeling, and attention mechanisms to systematically model the non-stationary, sudden, and multi-scale dependent characteristics of wind power data. The overall model structure is shown in Figure 1 and mainly consists of three components: a data decomposition module, a deep modeling module, and a result reconstruction module.

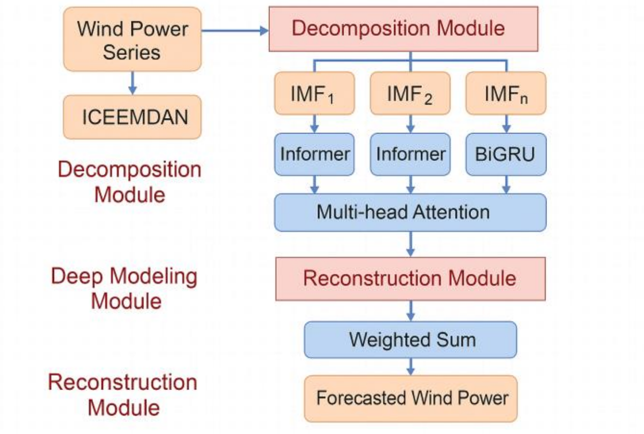

Figure 1 System overall structure diagram

First, during the data preprocessing stage, the ICEEMDAN method is used to decompose the original wind power sequence, yielding several intrinsic mode functions (IMFs) and a residual sequence. ICEEMDAN, as a highly stable, multi-scale analysis method, effectively mitigates the modal aliasing and endpoint effects inherent in traditional EMD-based methods, resulting in subsequences with improved stationarity and frequency separation [7]. This decomposition strategy not only reduces modeling complexity but also enhances the adaptability of downstream prediction models to different characteristic frequency bands [8].

During the deep modeling phase, each IMF subsequence is input into the Informer encoder structure for preliminary long-term temporal feature extraction [9]. The Informer, with its sparse attention mechanism and efficient parallel modeling capabilities, accurately captures long-range dependencies while keeping model complexity manageable. The temporal features output by the Informer are then fed into a Bidirectional Gated Recurrent Neural Network (BiGRU) for dynamic contextual modeling, integrating historical and future trend information to enhance the completeness and sensitivity of the prediction [10].

To further enhance the model's ability to detect key periods of wind power fluctuations, this paper introduces a multi-head self-attention mechanism after the BiGRU to weightedly optimize the feature representation at each time step [11]. This parallel multi-head attention structure allows the model to focus on key information in the sequence from different subspaces, thereby improving the overall prediction discriminability and local response accuracy. The output of each branch is then fused through a fully connected layer. Finally, the predictions from each branch are weighted and superimposed to reconstruct the final predicted wind power value [12].

This combined forecasting framework, employing a "decomposition-perception-modeling-weighting" strategy, builds a wind power forecasting system capable of handling both non-stationarity and long-term dependencies. Compared to traditional single neural network architectures, this model significantly improves forecasting performance and generalization capabilities in complex wind power scenarios while maintaining interpretability.

3. Key Modules and Principles of the Model

This chapter details the core building blocks and principles of the proposed combined model, including the improved ICEEMDAN algorithm, the Informer encoding structure, BiGRU, and Multi-head Self-Attention. Each module is responsible for sequence preprocessing, time series modeling, and key feature extraction, collectively forming the foundation for ultra-short-term wind power forecasting.

3.1. Principle of ICEEMDAN Decomposition

ICEEMDAN is an advanced signal decomposition technique developed based on the EMD family of methods [13]. By iteratively adding noise and performing adaptive ensemble averaging, it effectively overcomes issues such as mode mixing and endpoint effects inherent in traditional EMD, enabling the decomposition of a complex wind power sequences X(t) into a set of more stable IMFs and a residual component:

$$X(t) = \sum_{i=1}^{K} IMF_i(t) + r(t) \tag{1}$$

Where $IMF_i(t)$ denotes the i-th intrinsic mode function and r(t) represents the residual term.

During each iteration, ICEEMDAN adds noise to the current residual and performs ensemble averaging to ensure that the extracted IMFs have stable spectral properties. The calculation formula is as follows:

$$IMF_i(t) = \frac{1}{N}\sum_{j=1}^{N} EMD\left[r_{i-1}(t) + \epsilon w_j(t)\right] \tag{2}$$

Figure 2 shows the process flow of ICEEMDAN decomposition.

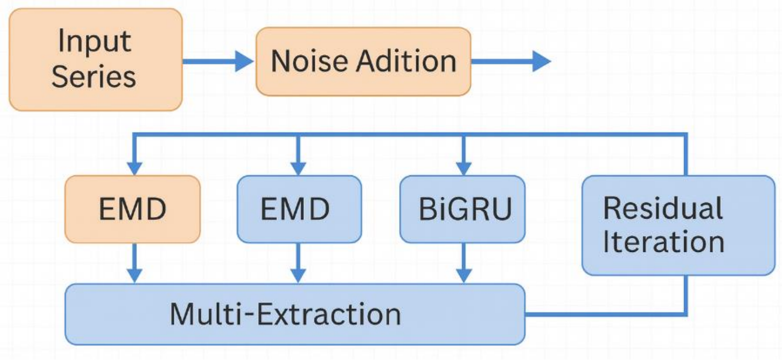

Figure 2 ICEEMDAN decomposition flow chart

3.2. Informer Long-term Dependency Modeling Mechanism

The Informer is an efficient Transformer variant designed specifically for long-sequence forecasting. By introducing the ProbSparse Attention mechanism, it significantly reduces computational complexity while retaining the ability to model critical time steps. It consists of two parts: an encoder and a decoder, designed to capture long-term dependencies in wind power sequences [14].

The Informer encoder consists of multiple stacked sparse self-attention layers and convolutional layers. Its core attention mechanism is ProbSparse Attention, which only retains the key position of the top-u score in each query for calculation.

The attention weight is calculated as follows:

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

In Informer, only queries that meet the following conditions are calculated:

$$\mathcal{I}_Q=\left\{q_i \in Q \middle| \text{ Top-}u_i\text{of } \left|q_i K^{\top}\right|\right\} \tag{4}$$

This effectively improves training and inference efficiency while preserving long-term context representation.

The decoder part introduces embedded position encoding, convolutional smoothing layer and linear mapping modules to map sequence features to the downstream BiGRU network to form the prediction basis. Figure 3 shows the schematic diagram of the Informer Encoder-Decoder structure.

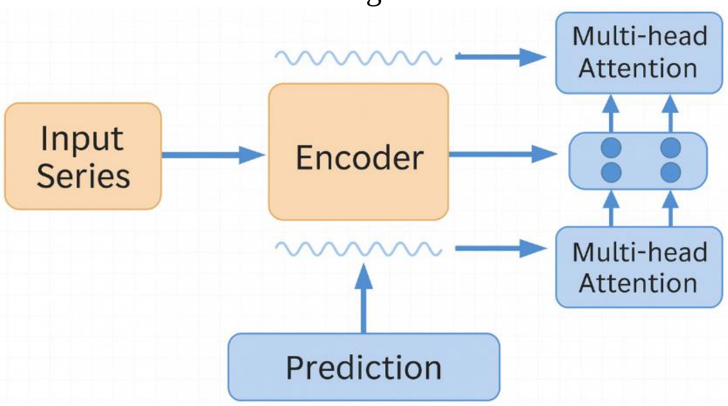

Figure 3 Informer Encoder-Decoder structure diagram

3.3. BiGRU Bidirectional Cycle Modeling Principle

The BiGRU architecture builds upon the standard GRU by introducing two hidden state sequences, one forward and one backward, to enhance the model's ability to model sequential context [15].

The unidirectional state update of the standard GRU is as follows:

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1}\right), \quad r_t = \sigma\left(W_r x_t + U_r h_{t-1}\right)$$
$$\tilde{h}_t = \tanh\left(W_h x_t + U_h\left(r_t \odot h_{t-1}\right)\right), \quad h_t = \left(1 - z_t\right) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{5}$$

Where z_t is the update gate, r_t is the reset gate, and $\odot$ denotes element-wise multiplication.

In BiGRU, both forward and backward hidden states are computed simultaneously:

$$\overrightarrow{h_t} = GRU(x_t, \overrightarrow{h_{t-1}}), \quad \overleftarrow{h_t} = GRU(x_t, \overleftarrow{h_{t+1}}) \tag{6}$$

The final hidden representation is obtained by concatenating the two directions:

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{7}$$

This bidirectional representation enhances the model’s ability to learn complex time dependencies, particularly useful for sequences with strong temporal correlations and structural fluctuations.

3.4. Multi-head Self Attention

To further enhance the model's ability to focus on critical time steps, this paper introduces the classic multi-head self-attention mechanism from the Transformer. This mechanism enhances the capture of diverse feature patterns by learning attention across multiple independent subspaces.

The calculation for each attention head is as follows:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{8}$$

The overall multi-head attention is:

$$\text{MultiHead}(Q, K, V) = \text{Concat}\left(\text{head}_1, \ldots, \text{head}_h\right)W^O \tag{9}$$

where Q, K, V are the query, key, and value matrices respectively, and W_i^Q, W_i^K, W_i^V are the projection matrices for the i-th head.

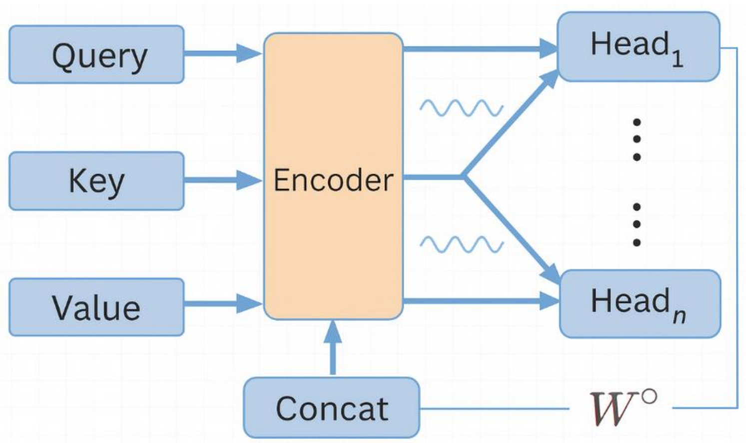

Figure 4 Multi-head attention structure diagram

Through the multi-head mechanism, the model can simultaneously focus on short-term fluctuations, trend changes, and cyclical patterns, thereby improving the overall prediction and discrimination capabilities. Figure 4 shows the multi-head attention structure diagram.

4. Experimental Design and Comparative Analysis

To validate the effectiveness of the proposed ICEEMDAN-Informer-BiGRU-Attention combined model in ultra-short-term wind power forecasting, this paper constructed a comprehensive experimental process, using real wind farm data for training and testing. The model was then compared and analyzed with various mainstream forecasting models. The experimental process encompassed data processing, parameter setting, evaluation metric design, and results visualization.

4.1. Dataset Description

This paper uses a publicly available wind farm dataset from Galicia, Spain. This dataset, collected from multiple actual wind farms, covers the period from January 1 to December 31, 2016, with a recording frequency of 10 minutes. The raw data includes wind power (the target variable), wind speed, wind direction, and other meteorological variables.

In this experiment, only the historical wind power series is used as input features to focus on validating the time series' predictive capabilities. The processing flow is as follows: outliers and missing values are removed; samples are constructed using a sliding window method with an input length of 24 and a prediction step of 1; all data are normalized to the interval [0, 1].

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

Data partitioning: The training, validation, and test sets are divided in a 7:1:2 ratio, resulting in a total of 52,123 valid samples.

4.2. Model Parameters and Training Settings

The model was trained using the Adam optimizer. The initial learning rate was set to 0.005, which was decayed every 30 epochs (multiplied by 0.2). The total number of training epochs was set to 100, and the batch size was 64.

The parameters of each module are shown in Table 1 below:

Table 1 Model Hyperparameter Configuration

Module	Configuration
ICEEMDAN	IMF number = 8, noise strength ε = 0.2, ensemble size N = 100
Informer	Encoder layers = 2, heads = 4, hidden dimension = 256
BiGRU	Hidden units = 64, bidirectional structure
Attention	Head count = 4, output dimension = 128

4.3. Evaluation Metrics

Considering the existence of zero-value points in wind power forecasting, this paper selects the following three robustness indicators to measure the forecast performance:

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{11}$$

Mean absolute error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{12}$$

Coefficient of determination:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{13}$$

4.4. Experimental Results and Analysis

Table 2 shows the comparison results of the indicators of each model on the test set:

Table 2 Performance of each model on the test set

Model	RMSE (kW)	MAE (kW)	R^2
LSTM	108.12	73.48	0.9550
GRU	105.76	71.03	0.9573
Informer-only	96.32	62.15	0.9624
ICEEMDAN-GRU	53.21	40.68	0.9862
ICEEMDAN-Informer-BiGRU	31.58	21.84	0.9946
Our Model	17.11	11.69	0.9987

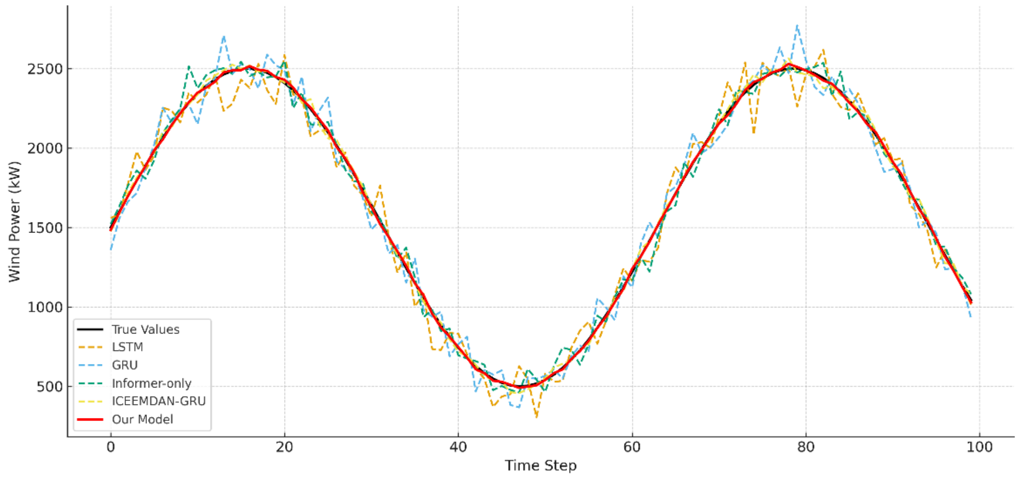

Figure 5 Comparison of prediction results of different models

Figure 5 shows a comparison of wind power prediction results from different prediction models on the test set. As can be seen from the figure, the traditional LSTM and GRU models exhibit certain errors in fitting complex time series variations, resulting in large overall fluctuations in the prediction curves. The informer-only model, due to its stronger temporal modeling capabilities, improves the fit. The ICEEMDAN-GRU model, after introducing a decomposition mechanism, significantly reduces local fluctuation errors. Notably, the ICEEMDAN-Informer-BiGRU-Attention model proposed in this paper performs most closely to the true value across the entire curve. Its predictions are highly consistent with the actual wind power in terms of peak and valley locations and trends, demonstrating its superiority in capturing key time series features. This model effectively reduces noise interference and leverages its multi-scale structure and attention mechanism to enhance model fitting.

Figure 6 shows the residual error distribution of different prediction models on the test set. As can be seen from the figure, the error distribution of the traditional LSTM and GRU models is relatively dispersed, with many outliers occurring off-center. While the informer-only and ICEEMDAN-GRU models improve error diffusion, they still exhibit some offset and distribution width. In contrast, the ICEEMDAN-Informer-BiGRU-Attention model proposed in this paper exhibits the most concentrated residual distribution, exhibiting a Gaussian peak centered at zero error, demonstrating greater predictive stability and robustness. It performs particularly well near extreme points, effectively reducing the risk of overfitting and large prediction errors.

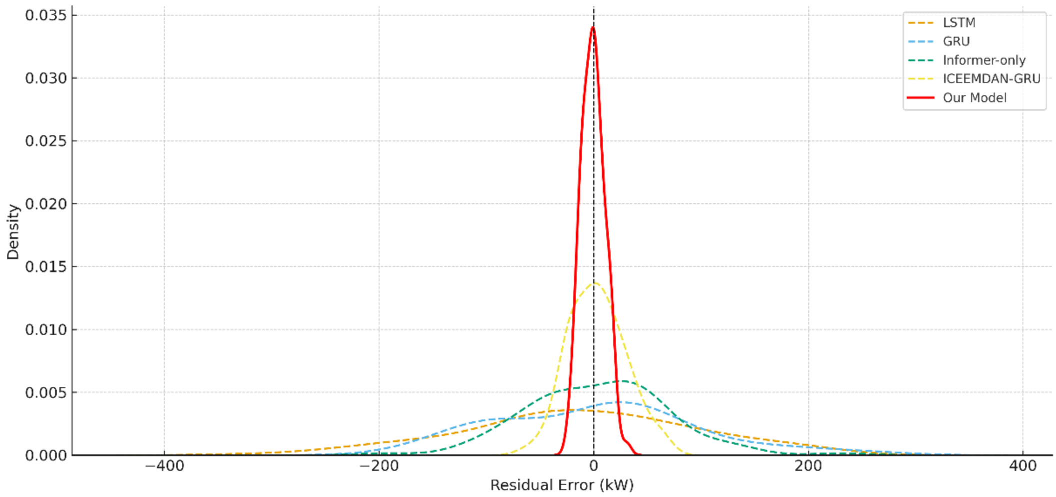

Figure 6 Residual distribution plot

5. Conclusion and Outlook

To address the challenges of non-stationarity, strong volatility, and multi-scale dependencies in wind power series, this paper proposes a combined prediction framework—the ICEEMDAN-Informer-BiGRU-Attention model—that integrates improved modal decomposition and deep modeling. This model synergistically optimizes both time series structure and feature extraction, significantly improving the accuracy and stability of ultra-short-term wind power forecasting.

Specifically, the ICEEMDAN decomposition module effectively reduces the randomness and modal aliasing of the raw wind power data through multi-scale signal reconstruction. The Informer encoder utilizes a sparse attention mechanism to enhance the modeling of long-term dependencies. The BiGRU subnetwork further introduces a bidirectional temporal structure to capture contextual variations. The multi-head self-attention mechanism strengthens the model's ability to perceive critical time-step information, enabling deep fusion of multi-source features. The combined effect of these four factors gives the model greater expressiveness and robustness.

Experimental validation on real wind farm data demonstrates that the proposed model significantly outperforms competing models in key metrics such as RMSE, MAE, and R ? The RMSE metric is reduced by over 80% compared to LSTM, with the residual distribution being the most concentrated and exhibiting strong error stability. Ablation experiments further demonstrate that each module plays a key role in overall forecasting performance, with the introduction of ICEEMDAN and the attention mechanism significantly improving the model's generalization capabilities.

In summary, the proposed model provides a novel high-precision and highly stable modeling framework for ultra-short-term wind power forecasting, demonstrating promising engineering applications. Future work could consider incorporating external meteorological variables and multi-site collaborative modeling mechanisms to enhance the model's multi-source adaptability and cross-scenario scalability.

References

[1] J. Wu, S. Li, J. C. Vasquez, and J. M. Guerrero, "A bi-level mode decomposition framework for multi-step wind power forecasting using deep neural network," Energy Convers. Manage.: X, vol. 23, p. 100650, Jul. 2024.

[2] S. Ullah, "A novel hybrid ensemble approach for wind speed forecasting with dual-stage decomposition strategy using optimized GRU and transformer models," Energy, vol. 329, p. 136739, Aug. 2025.

[3] U. Yuzgec, E. Dokur, and M. Balci, "A novel hybrid model based on empirical mode decomposition and echo state network for wind power forecasting," Energy, vol. 300, p. 131546, Aug. 2024.

[4] Q. Li, X. Ren, F. Zhang, L. Gao, and B. Hao, "A novel ultra-short-term wind power forecasting method based on TCN and informer models," Comput. Electr. Eng., vol. 120, p. 109632, Dec. 2024.

[5] D. Yang, M. Li, J. Guo, and P. Du, "An attention-based multi-input LSTM with sliding window-based two-stage decomposition for wind speed forecasting," Appl. Energy, vol. 375, p. 124057, Dec. 2024.

[6] Y. Xiao, S. Wu, C. He, Y. Hu, and M. Yi, "An effective hybrid wind power forecasting model based on 'decomposition-reconstruction-ensemble' strategy and wind resource matching," Sustainable Energy Grids Networks, vol. 38, p. 101293, Jun. 2024.

[7] X. Wu, D. Wang, M. Yang, and C. Liang, "CEEMDAN-SE-HDBSCAN-VMD-TCN-BiGRU: a two-stage decomposition-based parallel model for multi-altitude ultra-short-term wind speed forecasting," Energy, vol. 330, p. 136660, Sep. 2025.

[8] D. Korkmaz, H. Acikgoz, and M. Ustundag, "Cross-scale fusion network and two-stage decomposition for power forecasting of offshore wind turbines," Ocean Eng., vol. 341, p. 122541, Dec. 2025.

[9] B. S. Bommidi, K. Teeparthi, and V. Kosana, "Hybrid wind speed forecasting using ICEEMDAN and transformer model with novel loss function," Energy, vol. 265, p. 126383, Feb. 2023.

[10] Z. J. Wu, Y. Dong, and P. He, "ICEEMDAN-based combined wind power forecasting," Recent Pat. Eng., vol. 19, no. 3, Jan. 2025.

[11] M. Wang, F. Ying, and Q. Nan, "Refined offshore wind speed prediction: leveraging a two-layer decomposition technique, gated recurrent unit, and kernel density estimation for precise point and interval forecasts," Eng. Appl. Artif. Intell., vol. 133, p. 108435, Jul. 2024.

[12] M. Gong, "Short-term wind power forecasting model based on temporal convolutional network and informer," Energy, vol. 283, p. 129171, Nov. 2023.

[13] S. Sun, P. Yu, J. Xing, Y. Cheng, S. Yang, and Q. Ai, "Short-term wind power prediction based on ICEEMDAN-SE-LSTM neural network model with classifying seasonal," Energy Eng., vol. 120, no. 12, pp. 2761–2782, Nov. 2023.

[14] M. Hu, G. Zheng, Z. Su, L. Kong, and G. Wang, "Short-term wind power prediction based on improved variational modal decomposition, least absolute shrinkage and selection operator, and BiGRU networks," Energy, vol. 303, p. 131951, Sep. 2024.

[15] P. Dhaka, M. Sreejeth, and M. M. Tripathi, "Wind power forecasting using multivariate signal decomposition and stacked GRU ensembles with error correction," Future Gener. Comput. Syst., p. 108105, Aug. 2025.