



ARTICLE

A CEEMDAN-CNN-BiLSTM-Attention Framework with Bayesian Optimization for Wind Power Forecasting at Multiple Temporal Resolutions

Md Shafiul Alam*

Department of Electrical Engineering, College of Engineering, King Faisal University, Al-Ahsa, Saudi Arabia

*Corresponding Author: Md Shafiul Alam. Email: shafiul@kfu.edu.sa

Received: 13 July 2026; Accepted: 15 September 2026; Published: 21 October 2026

ABSTRACT: In modern power systems with increasing penetration of renewable energy sources, wind power forecasting at multiple temporal resolutions is of crucial importance for reliable and economic operation. But, the strong nonlinearity, intermittency and nonstationarity features of wind power, cause serious problems for traditional forecasting methods. In this paper, a new hybrid forecasting framework is proposed for accurate wind power forecasting at multiple temporal resolutions. The proposed framework combines Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), Attention Mechanism, and Bayesian Optimization. In the proposed framework, CEEMDAN firstly decomposes the original wind power series into multiple intrinsic mode functions to reduce the complexity of the signal. Then each decomposed component is modeled by a CNN-BiLSTM-Attention network independently and Bayesian Optimization is applied to automatically select the optimal network hyperparameters, which eliminates the necessity of manual tuning. Finally, the predicted intrinsic components are reconstructed to get the predicted wind power. The proposed method is evaluated on two real-world wind power datasets with significantly different temporal resolutions and statistical properties. Extensive comparison of deep learning models such as LSTM, BiLSTM, GRU and CNN-LSTM to four popular models is performed. The experimental results demonstrate that the proposed framework achieves superior overall forecasting performance across both datasets, particularly in terms of MAE, RMSE, and R^2 . The proposed model achieves the MAE of 89.56 MW, RMSE of 115.36 MW, and R^2 of 0.9796 on Dataset 1 and the MAE of 230.19 MW, RMSE of 299.59 MW, and R^2 of 0.9957 on Dataset 2, outperforming all the competing methods in these matrices. Additional evaluation using sMAPE, nMAE, and nRMSE further confirms the robustness of the proposed framework, particularly for Dataset 1 where conventional MAPE is sensitive to near-zero wind-power observations. Moreover, the statistical significance tests (Friedman and Nemenyi tests) justify the effectiveness of the proposed framework against conventional deep learning approaches. The results show that the combination of signal decomposition, deep feature extraction, temporal dependency learning, attention-based feature weighting and automated hyperparameter optimization provides a robust and highly accurate solution for wind power forecasting at multiple temporal resolutions.

KEYWORDS: Wind power forecasting; convolutional neural network; CEEMDAN; attention mechanism; bidirectional long short-term memory; deep learning; Bayesian optimization; renewable energy

1 Introduction

Wind energy has become one of the most important renewable resources for decreasing dependence on fossil-fuel based generation and supporting low-carbon power systems. Yet, the increasing penetration of wind power brings significant operational issues due to the intrinsic intermittency, nonlinearity and

high weather dependence of wind generation. Unlike conventional generators, wind turbines cannot be dispatched independently of the atmospheric conditions, and their output may be highly variable over short time intervals [1]. Such variations affect the unit commitment, economic dispatch, reserve scheduling, frequency regulation, participation in the electricity market and secure operation of transmission and distribution networks. Thus, accurate wind power forecasting on multi-horizon is of great importance to improve the reliability of the grid and to reduce the operational risk, as well as to allow large-scale integration of renewable energy. Recent studies have highlighted the importance of accurate wind power forecasting in reducing scheduling uncertainty, improving supply-demand balancing, and enhancing the efficiency of wind power trading [2,3]. Some recent studies also show that large-scale wind integration needs accurate and stable forecasting to reduce operational risks and improve grid scheduling efficiency [4,5].

Wind power forecasting methods can be classified broadly into physical, statistical, machine learning and deep learning methods [6]. Physical models are based on numerical weather prediction, terrain data, turbine data and the equations for the atmospheric boundary-layer. While physically meaningful, these approaches require detailed meteorological and site-specific information and may become computationally expensive for short-term operational applications. Statistical methods, for example autoregressive models and Kalman-filter-based techniques [7], are more straightforward and computationally efficient, but they often cannot account for the nonlinear and nonstationary behavior of wind power time series. Machine learning techniques such as artificial neural networks, support vector machines, random forests and extreme learning machines have improved the forecasting capability by learning nonlinear relationships directly from the historical data. However, traditional machine learning models are often highly dependent on manual feature engineering, and may fail to capture long-range temporal dependencies in sequential wind data [8]. The reference [9] introduces a multi-view spatiotemporal graph neural network encoder and a frozen pre-trained large language model (LLM) for modeling spatial, temporal, meteorological, and terrain dependencies for multi-step wind power forecasting. It also proposes a frequency domain interpolation method to reconstruct missing or anomalous sensor data which improves the forecasting accuracy and reliability. To address the non-independent and identically distributed data problem and improve personalized wind power forecasting in multiple wind farms, an adaptive parameter updater that combines federated learning and meta-learning-based fine-tuning is proposed in [10]. This method jointly optimizes global and local models with the preservation of data privacy, which achieves better accuracy especially for data-scarce wind farms. A stacked bidirectional LSTM model with SHAP-based explainable AI is proposed in [11] for accurate and interpretable wind power forecasting with consideration of multiple environmental variables. The reference [12] proposes a heterogeneous PatchTST-GRU sequence-to-sequence model integrated with numerical weather prediction refinement for multi-step wind power forecasting, where PatchTST extracts temporal features and GRU generates future prediction sequences. In [13], a hybrid model of SMA-Autoformer-BiLSTM-LightGBM is proposed for multi-step wind power forecasting, which uses Autoformer to capture the long-term temporal trends, BiLSTM to model the short-term fluctuations, LightGBM to correct the residual error, and the slime mold algorithm (SMA) to adaptively optimize the hyperparameters.

Recently, deep learning has been demonstrated to be a dominant approach for forecasting wind speed and wind power by virtue of its capability to automatically learn complex temporal and nonlinear representations. Long short-term memory (LSTM) networks and gated recurrent units (GRUs) are particularly useful for time-series forecasting due to their ability to alleviate the vanishing-gradient problem and capture temporal information across multiple time steps [14]. In [15], a hybrid LSTM-Single Candidate optimizer model was developed for short-term wind power forecasting and the optimization method was used to optimize the LSTM configuration. The proposed method achieved better forecasting accuracy compared to

the conventional models used in the study. Another study [16] proposed a wavelet transform convolution-Transformer model for wind power forecasting under extreme operating conditions. The method integrated wavelet based feature extraction and Transformer learning, which improved forecasting performance under highly variable wind power conditions. Bidirectional LSTM (BiLSTM) further enhances the temporal learning by extracting information from two directions, forward and backward [17]. Convolutional neural networks (CNNs) are widely used for image processing, but they are also useful in one-dimensional time-series forecasting as they can extract local patterns and short-term fluctuations from sequential data. This leads to the increasing use of hybrid CNN-LSTM and CNN-BiLSTM models for jointly learning local features and temporal dependencies [18–20]. For example, recent research has shown that CNN-LSTM and other similar hybrid models can improve the prediction of wind speed or wind power by learning spatial, local and sequential features simultaneously.

However, direct application of deep learning models to raw wind power data has been challenging despite these advancements. The wind power signals are nonlinear, nonstationary, noisy and multi-scale in general. A single forecasting model trained on the original signal may not be able to distinguish high frequency fluctuations, medium term variations, and low frequency trends [21,22]. This limitation has led to the development of decomposition based hybrid forecasting models. Signal decomposition methods decompose the original wind power series into several intrinsic mode functions or subseries with different frequency characteristics. Then each of the components is predicted separately. The final prediction is obtained by reconstructing the predicted components. Empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), variational mode decomposition (VMD), empirical wavelet transform (EWT), and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) have been widely used for this purpose [23–25].

Among these methods, CEEMDAN has attracted much attention because it is able to improve the decomposition completeness and reduce the mode-mixing problems in comparison with the conventional EMD and EEMD. CEEMDAN introduces adaptive noise to improve the stability of the decomposition process, decomposing the original signal into a large number of intrinsic mode functions and a residual component [26]. This is useful for wind power forecasting where the signal typically has irregular high frequency fluctuations and nonstationary patterns. In recent years, CEEMDAN has been successfully integrated with LSTM, BiLSTM, grey wolf optimization, improved whale optimization, empirical wavelet transform, and other learning frameworks [27–29]. For instance, Karijadi et al. introduced a CEEMDAN-EWT-LSTM-based approach for the ultra-short-term wind power prediction, and validated on the wind power datasets from France and Turkey [25]. The original wind power data was decomposed by CEEMDAN and denoising was performed by using EWT before LSTM-based forecasting. Similarly, Sun et al. proposed a CEEMDAN-GWO-BiLSTM model, in which CEEMDAN is used to decompose the wind power signal, BiLSTM is used to forecast each component, and GWO is used to optimize the model parameters before the final signal reconstruction [30].

Another significant development is the application of attention mechanisms in wind forecasting. In standard recurrent networks temporal information is compressed into hidden states, but not all time steps or extracted features are equally important for the final prediction. Attention mechanisms overcome this limitation by adaptively weighting important temporal features, which allows the model to concentrate on the most relevant information [31]. Tian et al. proposed a deep learning wind power forecasting system with a dual-stage self-attention mechanism and noted that previous studies focused on preprocessing and parameter optimization, but the mining of hidden-layer features was limited [32]. They used attention to weigh the importance of input and hidden features, which improved the accuracy and stability of the forecasts in their study. This aligns with the emerging consensus that attention-based models are particularly valuable

for volatile renewable energy forecasting, where the forecasting model needs to find the most informative parts of noisy time-series data [33,34].

Hyperparameter optimization is another critical issue in wind forecasting based on deep learning. The performance of CNN, LSTM, BiLSTM and attention-based models depends heavily on lookback window length, number of filters, kernel size, recurrent units, dropout rate, dense-layer size, batch size and learning rate. These parameters were set by trial-and-error in many previous studies, which can lead to sub-optimal performance and poor reproducibility. The model tuning has been increasingly automated using metaheuristic and Bayesian optimization methods. Deep learning models are combined with grey wolf optimization, whale optimization, coati optimization, and other smart optimizers for renewable forecasting [29]. More specifically, optimized hybrid frameworks such as ICEEMDAN-KPCA-IWOA-BiLSTM have tackled the simultaneous problems of noise suppression, feature construction, and unstable hyperparameter selection [29]. However, many meta-heuristic algorithms need to be carefully parameterized and may entail high computational cost. Bayesian optimization, for example an Optuna-based search, is a more systematic and sample-efficient way to select deep learning hyperparameters, by probing promising regions of the parameter space sequentially.

Existing studies have made great progress, but some gaps still exist in the research. Firstly, many decomposition-based models use CEEMDAN or VMD but use relatively simple forecasting networks such as LSTM or GRU which cannot fully leverage both local and bidirectional temporal patterns. Second, some wind forecasting systems with attention mechanisms have been proposed, but they are often not combined with decomposition-based component-wise learning. Third, it is common to utilize different metaheuristic algorithms to optimize model hyperparameters, but there are few studies integrating CEEMDAN decomposition, CNN based local feature extraction, BiLSTM based temporal learning, attention pooling, and Bayesian hyperparameter optimization into a single forecasting framework. Fourth, many of the studies improve accuracy but do not provide much discussion on the reconstruction consistency, physical output constraints, and component-wise prediction behavior. These limitations indicate the necessity of a more integrated hybrid model which can handle signal nonstationarity, extract multi-level temporal features, tune critical hyperparameters automatically and reconstruct the final wind power forecast in a physically meaningful way.

In this paper, we propose a Bayesian-optimized CEEMDAN-CNN-BiLSTM-Attention framework for wind power forecasting at multiple temporal resolutions to fill these gaps. In the proposed method, the original wind power series is normalized, decomposed by CEEMDAN into some intrinsic mode functions. Each component is transformed into supervised learning samples by a sliding-window approach. Then, a hybrid CNN-BiLSTM-Attention network is trained separately for each decomposed component. The CNN layer is used to extract the local fluctuation patterns, the stacked BiLSTM layers are used to capture the bidirectional temporal dependencies, and the attention pooling layer is used to assign higher weights to the more informative hidden representations. We tune the key hyperparameters such as lookback window, convolution filters, kernel size, BiLSTM units, dense units, dropout rate, learning rate and batch size by using Bayesian optimization. Finally, the predicted components are reconstructed to get the final wind power forecast and the results are inversed normalized and evaluated using MAE, RMSE, MAPE and R^2 . Thus, the proposed framework combines signal decomposition, automatic hyper-parameter optimization, hybrid deep feature extraction, attention-based temporal weighting and component-wise reconstruction into a unified forecasting architecture. The novelty of this research does not lie in the individual development of CEEMDAN, CNN, BiLSTM, attention, or Bayesian optimization but in their systematic integration into a unified forecasting architecture. CEEMDAN decomposes the nonlinear wind-power signal into several components with different frequency characteristics, CNN extracts the local features of the decomposed

signals, BiLSTM captures the temporal dependencies, and the attention mechanism adaptively emphasizes the more informative temporal features. Bayesian optimization also provides data-driven selection of the principal model hyperparameters. The resulting framework, thus, combines the decomposition, feature extraction, temporal modeling, adaptive feature weighting and hyperparameter optimization into a coordinated wind-power forecasting pipeline.

2 Proposed Methodology

2.1 Overview of the Proposed Forecasting Framework

This study proposes a hybrid CEEMDAN–CNN–BiLSTM–Attention framework with Bayesian optimisation for wind power forecasting at multiple temporal resolutions as shown in Fig. 1. A separate leakage-controlled evaluation, in which chronological partitioning precedes normalization and CEEMDAN decomposition, is presented in Section 4.2. The proposed architecture is designed to address three major challenges commonly encountered in wind power prediction: the nonlinear and nonstationary characteristics of wind power signals, the complex temporal dependency among consecutive observations, and the sensitivity of deep learning models to hyperparameter selection. In the proposed framework, the original wind power time series is first normalised and then decomposed into multiple intrinsic components using CEEMDAN. Instead of directly forecasting the original wind power signal, each decomposed CEEMDAN component is independently transformed into supervised learning samples using a sliding-window strategy. A separate CNN–BiLSTM–Attention model is then trained for each decomposed component. The CNN layer extracts local temporal features, the BiLSTM layers capture forward and backward temporal dependencies, and the attention pooling layer assigns adaptive weights to the most informative hidden representations. Finally, the predicted decomposed components are reconstructed through summation to obtain the final normalised wind power forecast, which is subsequently transformed back to the original scale. The individual modules of the proposed framework are selected to perform complementary functions within the forecasting pipeline. CNN is employed to extract local fluctuation patterns from the decomposed component sequences, while BiLSTM subsequently captures temporal dependencies within the historical input window. Compared with TCN or Transformer architectures, this combination provides a relatively compact structure suitable for the short-lookback, univariate component sequences considered in this study. The attention mechanism is applied after BiLSTM to assign greater importance to the most informative temporal features rather than treating all hidden states equally. Thus, the CNN, BiLSTM, and attention modules respectively provide local feature extraction, temporal dependency learning, and adaptive temporal feature weighting within the CEEMDAN-based forecasting framework.

The main methodological novelty of the proposed framework lies in the integration of CEEMDAN-based signal decomposition, component-wise deep learning, attention-based temporal feature weighting, and Bayesian hyperparameter optimisation within a unified forecasting pipeline. CEEMDAN is adopted to reduce the complexity of the original wind power signal by separating it into simpler oscillatory components, thereby improving the learnability of the forecasting model. The CNN module extracts local fluctuation patterns from each decomposed component, while the BiLSTM structure enhances temporal learning by processing the sequence in both forward and backward directions. The attention mechanism further improves the representation capability of the model by allowing it to focus on the most relevant temporal features [35]. Bayesian optimisation is employed to automatically search for an effective set of hyperparameters, reducing manual trial-and-error tuning and improving model reproducibility.

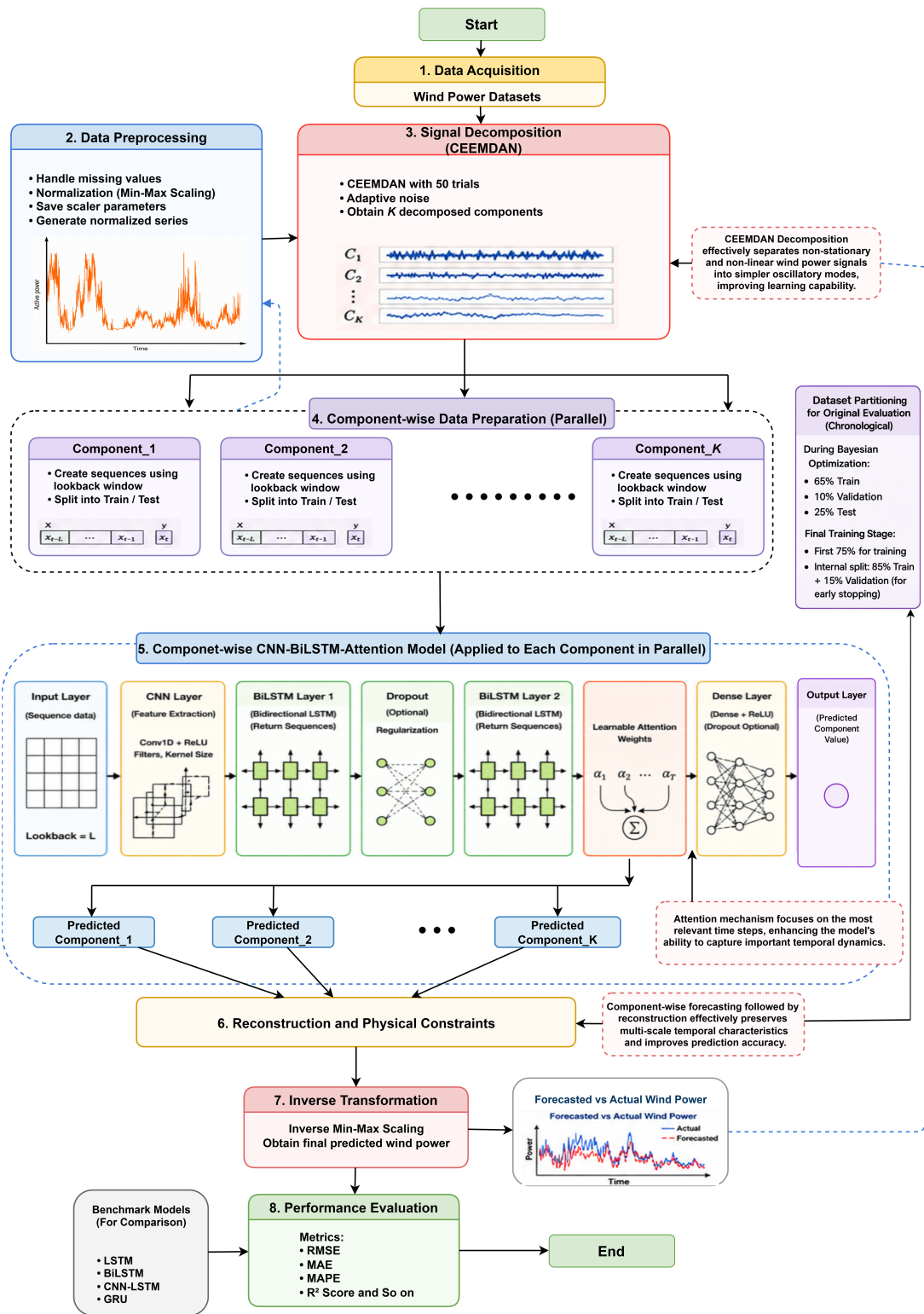


Figure 1: Overall workflow of the proposed CEEMDAN-CNN-BiLSTM-Attention framework under the original evaluation protocol for wind energy prediction. A separate leakage-controlled evaluation applies chronological partitioning before normalization and CEEMDAN decomposition.

2.2 Data Preprocessing and Normalisation

Let the original wind power time series be represented as

$$P = \{P_1, P_2, \dots, P_T\}, \quad (1)$$

where P_t denotes the observed wind power at time index t , and T is the total number of samples. Since deep learning models are sensitive to the numerical scale of input data, the raw wind power series is normalised using min-max scaling as

$$\tilde{P}_t = \frac{P_t - P_{\min}}{P_{\max} - P_{\min}}, \quad (2)$$

where $P_{\min}$ and $P_{\max}$ are the minimum and maximum values of the original wind power series, respectively. The normalised wind power series $\tilde{P}_t$ is bounded within the interval $[0, 1]$, which improves numerical stability during model training and prevents large-magnitude values from dominating the optimisation process.

During the modeling process, the samples are processed in the chronological order and no random shuffling is done. In the original evaluation protocol, the normalization and CEEMDAN decomposition are applied to the entire time series before the chronological partitioning. This protocol can therefore include future information through the preprocessing and decomposition steps. To explicitly evaluate this effect, a separate leakage-controlled evaluation is conducted in [Section 4.2](#) where the data are split chronologically before normalization and CEEMDAN decomposition. During Bayesian hyperparameter optimisation, the first 65% of the normalised series is used for model training, while the following 10% is used for validation. The validation period is used to evaluate the reconstructed wind power prediction produced by summing the forecasts of all decomposed components. After identifying the best hyperparameter configuration, the final models are trained using the first 75% of the dataset, and the remaining 25% is reserved for testing. Within the final training stage, an internal validation split is also used for early stopping to reduce overfitting.

2.3 CEEMDAN-Based Signal Decomposition

Wind power signals are highly nonlinear and nonstationary because of stochastic wind behaviour, turbulence, seasonal variation, and short-term atmospheric fluctuations. Directly training a deep learning model on the original wind power series may force the model to learn multiple frequency scales simultaneously, including high-frequency fluctuations, medium-frequency variations, and low-frequency trends. This can increase model complexity and reduce forecasting accuracy. To address this issue, CEEMDAN is applied to decompose the normalised wind power signal into several simpler components.

The CEEMDAN decomposition of the normalised wind power signal can be generally expressed as

$$\tilde{P}_t = \sum_{k=1}^K C_k(t), \quad (3)$$

where $C_k(t)$ denotes the k -th CEEMDAN-decomposed component, with the final component corresponding to the residual (trend), and K is the total number of components returned by the decomposition process. In this study, all CEEMDAN-generated components are treated consistently as learnable subseries. Therefore, instead of explicitly separating the residual term from the IMFs in the forecasting stage, each returned component is forecasted using an individual deep learning model, and the final wind power prediction is obtained by summing all predicted components.

This decomposition strategy transforms the original complex forecasting problem into a set of simpler component-wise forecasting tasks. High-frequency components generally represent rapid fluctuations,

while low-frequency components capture smoother and more slowly varying patterns. By modelling each component individually, the proposed framework enables the CNN–BiLSTM–Attention network to learn component-specific temporal behaviour more effectively than a single model trained directly on the original signal. The CEEMDAN method was performed with the PyEMD package with 50 ensemble trials, and the random noise generation was initialized with a fixed seed of 123 for reproducibility purposes. The amplitude of the noise and the stopping criteria for the decomposition were not manually altered and were kept at the default values of the PyEMD implementation.

2.4 Sliding-Window Supervised Learning Formulation

For each decomposed CEEMDAN component $C_k(t)$, a sliding-window approach is used to construct supervised learning samples. Given a lookback window length L , the input sequence and target output for the k -th component are formulated as

$$\mathbf{X}_{k,t} = [C_k(t-L), C_k(t-L+1), \dots, C_k(t-1)], \quad (4)$$

and

$$y_{k,t} = C_k(t). \quad (5)$$

Thus, the model learns to predict the next value of each decomposed component using its previous L observations. The input tensor for each component is reshaped into the form

$$\mathbf{X}_{k,t} \in \mathbb{R}^{L \times 1}, \quad (6)$$

which is compatible with the one-dimensional convolutional and recurrent layers used in the proposed model.

The lookback window length L plays an important role in forecasting performance. A very short lookback window may fail to capture sufficient temporal information, whereas a very long lookback window may introduce redundant information and increase computational complexity. Therefore, in the proposed framework, L is not selected manually but is included as one of the hyperparameters optimised through Bayesian optimisation.

2.5 CNN–BiLSTM–Attention Forecasting Model

Each decomposed CEEMDAN component is forecasted using a dedicated CNN–BiLSTM–Attention model. The model receives the input sequence $\mathbf{X}_{k,t}$ and produces the predicted component value $\hat{C}_k(t)$. The model architecture consists of a one-dimensional convolutional layer, two stacked BiLSTM layers, an attention pooling layer, a dense hidden layer, and an output layer.

First, the input sequence is passed through a one-dimensional convolutional layer to extract local temporal patterns. The convolution operation can be written as

$$z_j = f \left(\sum_{m=0}^{q-1} w_m x_{j+m} + b \right), \quad (7)$$

where q is the convolution kernel size, w_m is the convolution weight, b is the bias term, and $f(\cdot)$ denotes the rectified linear unit (ReLU) activation function. The CNN layer helps identify short-term variations and local fluctuation patterns within each decomposed wind power component.

The output features from the CNN layer are then passed to stacked BiLSTM layers. A standard LSTM cell uses gating mechanisms to control information flow. The input gate, forget gate, output gate, candidate cell state, cell state, and hidden state are defined as

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \quad (8)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \quad (9)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \quad (10)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c), \quad (11)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad (12)$$

and

$$h_t = o_t \odot \tanh(c_t), \quad (13)$$

where i_t , f_t , and o_t are the input, forget, and output gates, respectively; $\tilde{c}_t$ is the candidate memory state; c_t is the cell state; h_t is the hidden state; $\sigma(\cdot)$ is the sigmoid activation function; and $\odot$ denotes element-wise multiplication.

In the BiLSTM structure, the sequence is processed in both forward and backward directions. The bidirectional hidden representation is given by

$$h_t^{bi} = [\vec{h}_t; \overleftarrow{h}_t], \quad (14)$$

where $\vec{h}_t$ and $\overleftarrow{h}_t$ represent the forward and backward hidden states, respectively. This bidirectional structure enables the model to learn richer temporal representations from both directions within the input sequence. In the proposed model, two BiLSTM layers are stacked to improve the temporal feature extraction capability. Dropout is applied after the first BiLSTM layer and after the dense layer when the selected dropout rate is greater than zero.

After the BiLSTM layers, an attention pooling mechanism is used to assign adaptive importance to the hidden states. For the hidden representation h_t^{bi} , the attention score is calculated as

$$e_t = \mathbf{u}^T \tanh(\mathbf{W} h_t^{bi} + \mathbf{b}), \quad (15)$$

where $\mathbf{W} \in \mathbb{R}^{d \times d}$ is the trainable weight matrix, $\mathbf{b} \in \mathbb{R}^d$ is the bias vector, $\mathbf{u} \in \mathbb{R}^d$ is the trainable context vector, and d denotes the hidden feature dimension of the BiLSTM output. The attention weights are computed using the softmax function:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{j=1}^L \exp(e_j)}. \quad (16)$$

The final attention-based context vector is then obtained as

$$\mathbf{c} = \sum_{t=1}^L \alpha_t h_t^{bi}. \quad (17)$$

This mechanism allows the proposed model to focus more strongly on the most informative time steps instead of treating all hidden states equally. This is especially useful for wind power forecasting because not all previous observations contribute equally to the next-step prediction.

The context vector is passed through a dense layer with ReLU activation, followed by an output neuron that generates the predicted value of the corresponding decomposed component:

$$\widehat{C}_k(t) = F_{\theta_k}(\mathbf{X}_{k,t}), \quad (18)$$

where $F_{\theta_k}(\cdot)$ represents the trained CNN–BiLSTM–Attention model for the k -th CEEMDAN component, and θ_k denotes its learnable parameters.

2.6 Training Strategy and Loss Function

For each CEEMDAN component, a separate CNN–BiLSTM–Attention model is trained. The proposed implementation uses the Adam optimiser, which adaptively updates the model parameters based on first- and second-order moment estimates of the gradients. The trainable parameters of the k -th component model are updated by minimising the Huber loss function. The Huber loss is less sensitive to large errors than the mean squared error while still maintaining quadratic behaviour for small errors. It is defined as

$$\mathcal{L}_\delta(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2, & \text{if } |y - \hat{y}| \leq \delta, \\ \delta|y - \hat{y}| - \frac{1}{2}\delta^2, & \text{otherwise.} \end{cases} \quad (19)$$

In this formulation, y is the actual component value, $\hat{y}$ is the predicted component value, and δ is the threshold parameter controlling the transition between quadratic and linear loss behaviour.

In the implementation we set the Huber threshold to $\delta = 1.0$. The learning rate of the Adam optimizer was tuned using Bayesian optimization, while all other parameters of the optimizer were set to the default values in TensorFlow/Keras. For Bayesian optimization, each candidate model was trained for up to 8 epochs, with early stopping based on the training loss, a patience of 2 epochs, and restoring the best weights. During final training the initial 75% of the chronological data was used as the training region; within this region the first 85% was used for fitting a model, and the remaining 15% was used as a chronological internal validation set. Last training was done for max 25 epochs with early stopping watching validation loss, patience of 5 epochs and restoring the best weights. The threshold of minimum improvement (`min_delta`) was not changed explicitly and therefore followed the default setting of TensorFlow/Keras.

2.7 Bayesian Hyperparameter Optimisation

The forecasting performance of the proposed deep learning model depends strongly on several hyperparameters, including the lookback length, number of CNN filters, kernel size, number of BiLSTM units, dense-layer size, dropout rate, learning rate, and batch size. Manual selection of these values can lead to suboptimal results and poor reproducibility. Therefore, Bayesian optimisation is adopted to automatically search for the best hyperparameter configuration.

Let the hyperparameter vector be defined as

$$\lambda = [L, F_c, q, U_1, U_2, D, \rho, \eta, B], \quad (20)$$

where L is the lookback length, F_c is the number of convolution filters, q is the kernel size, U_1 and U_2 are the numbers of units in the first and second BiLSTM layers, respectively, D is the number of dense neurons, ρ is the dropout rate, η is the learning rate, and B is the batch size.

The search space includes the lookback window length ($L \in \{8, \dots, 24\}$), convolution filters ($F_c \in \{16, 32, 64\}$), kernel size ($q \in \{2, 3, 5\}$), first BiLSTM units ($U_1 \in \{32, 64, 96\}$), second BiLSTM units

($U_2 \in \{16, 32, 64\}$), dense neurons ($D \in \{16, 32, 64\}$), dropout rate ($\rho \in [0, 0.15]$), learning rate ($\eta \in [10^{-4}, 3 \times 10^{-3}]$), and batch size ($B \in \{8, 16, 32\}$). In this study, Bayesian optimisation was conducted for eight trials to identify the best hyperparameter combination. Due to the high computational cost of training separate CNN-BiLSTM-Attention models for each CEEMDAN component, the Bayesian optimization procedure was limited to eight trials to obtain an efficient hyperparameter configuration given the available computational resources. We purposely limited the search budget to be computationally efficient, but given the same budget, Bayesian optimization systematically explores the hyperparameter space more efficiently than manual tuning.

For each trial, the selected hyperparameter set is used to train component-wise forecasting models over the training segment. The predicted validation components are then reconstructed by summation to obtain the validation wind power forecast. Therefore, the Bayesian optimisation objective is not based on the loss of a single decomposed component but on the reconstructed validation prediction. The objective function is formulated as

$$\lambda^* = \arg \min_{\lambda} \left[\frac{1}{N_v} \sum_{t=1}^{N_v} (\tilde{P}_t - \hat{P}_t)^2 \right], \quad (21)$$

where N_v is the number of validation samples, $\tilde{P}_t$ denotes the actual normalised wind power, and $\hat{P}_t$ denotes the reconstructed predicted wind power on the normalised scale during the validation period. In the implementation, the reconstructed prediction is clipped within $[0, 1]$ before computing the validation mean squared error. The hyperparameter set that minimises the validation reconstruction error is selected for the final training stage.

2.8 Component-Wise Forecast Reconstruction

After training the component-wise CNN-BiLSTM-Attention models, each trained model is used to forecast its corresponding decomposed CEEMDAN component in the testing region. The predicted components are collected as

$$\widehat{C}(t) = [\widehat{C}_1(t), \widehat{C}_2(t), \dots, \widehat{C}_K(t)]. \quad (22)$$

The reconstructed normalised wind power prediction is obtained by summing all predicted components:

$$\widehat{P}_t = \sum_{k=1}^K \widehat{C}_k(t). \quad (23)$$

Since the target wind power signal is normalised within $[0, 1]$, the reconstructed prediction is constrained to remain within this physical range:

$$\widehat{P}_t = \min \left(1, \max \left(0, \widehat{P}_t \right) \right). \quad (24)$$

Finally, the normalised prediction is transformed back to the original wind power scale using the inverse min-max transformation:

$$\widehat{P}_t = \widehat{P}_t (P_{\max} - P_{\min}) + P_{\min}. \quad (25)$$

This reconstruction process preserves the multi-scale characteristics of the original wind power signal while combining the individual predictions of all decomposed components.

2.9 Benchmark Models

Four popular deep learning models are selected as baseline models including LSTM, BiLSTM, GRU and CNN-LSTM to evaluate the forecasting ability of the proposed framework. These architectures are recurrent, bidirectional recurrent, gated recurrent and hybrid convolutional–recurrent architectures commonly used for time series forecasting. Their inclusion allows the comparison of the performance of the proposed CEEMDAN–CNN–BiLSTM–Attention framework with the existing deep learning forecasting architectures. The same chronological training-testing partition, normalized wind-power data, forecasting target and performance indicators were used for all benchmark models for a consistent comparative evaluation. The conventional LSTM, BiLSTM, GRU and CNN–LSTM models work directly on the normalized wind-power sequence, while the proposed framework includes the CEEMDAN decomposition and Bayesian optimization as constituent stages. Consequently, these stages were not applied to the conventional baselines, as their inclusion would transform the baseline models into different hybrid or decomposition-based forecasting architectures. An additional ablation experiment without Bayesian optimization is included to further quantify the influence of parameter optimization. An extended comparison with recent architectures such as TCN, Transformer, PatchTST and decomposition-based forecasting methods would further improve the evaluation. Such comparisons under a common hyperparameter-optimization budget are left to future work. For reproducibility purposes, all benchmark models used a lookback window of 24 steps. The LSTM, BiLSTM and GRU models used 2 recurrent layers of 64 and 32 units, while the CNN-LSTM model used a 32 filter Conv1D layer (kernel size 3) followed by LSTM layers of 64 and 32 units. All models were trained with the same evaluation protocol, using the Adam optimizer and Huber loss.

2.10 Performance Evaluation

The forecasting performance of the proposed model is evaluated using mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). These metrics are calculated using the actual and predicted wind power values in the original scale.

The MAE is defined as

$$MAE = \frac{1}{N} \sum_{t=1}^N |P_t - \widehat{P}_t|, \quad (26)$$

where N is the number of testing samples. The RMSE is calculated as

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (P_t - \widehat{P}_t)^2}. \quad (27)$$

The MAPE is given by

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{P_t - \widehat{P}_t}{\max(P_t, \epsilon)} \right|, \quad (28)$$

where ϵ is a small positive constant used to avoid division by zero when the actual wind power is close to zero. In the implementation, this constant is set to 10^{-9} . The coefficient of determination is computed as

$$R^2 = 1 - \frac{\sum_{t=1}^N (P_t - \hat{P}_t)^2}{\sum_{t=1}^N (P_t - \bar{P})^2}, \quad (29)$$

where $\bar{P}$ is the mean value of the actual wind power samples. Lower MAE, RMSE, and MAPE values indicate smaller forecasting errors, while an R^2 value closer to unity indicates stronger agreement between the actual and predicted wind power values.

To provide complementary evaluation measures that are less sensitive to near-zero observations, the symmetric mean absolute percentage error (sMAPE), normalized mean absolute error (nMAE), and normalized root mean square error (nRMSE) are also calculated as

$$\text{sMAPE} = \frac{100}{N} \sum_{t=1}^N \frac{2|P_t - \hat{P}_t|}{|P_t| + |\hat{P}_t| + \epsilon}, \quad (30)$$

$$\text{nMAE} = \frac{\text{MAE}}{P_{\max}^{\text{test}} - P_{\min}^{\text{test}}}, \quad (31)$$

$$\text{nRMSE} = \frac{\text{RMSE}}{P_{\max}^{\text{test}} - P_{\min}^{\text{test}}}, \quad (32)$$

where $P_{\max}^{\text{test}}$ and $P_{\min}^{\text{test}}$ denote the maximum and minimum actual wind-power values in the corresponding test set, respectively. Thus, nMAE and nRMSE are normalized by the range of the actual test observations rather than by the maximum value of the full dataset.

2.11 Analysis of Statistical Significance

Friedman and Nemenyi post-hoc tests are performed on block-wise forecasting errors to analyze whether the differences in forecasting performance among the models are statistically significant. One ranking unit is a block of 48 consecutive test samples, without overlap. In each block, the five forecasting models are ranked based on their block-wise MAE with the lowest MAE getting the rank 1. In case of ties, average ranks are assigned and the average rank of each model is obtained by averaging its ranks over all complete blocks. To assess the overall differences between the models we use the Friedman test, followed by the Nemenyi post-hoc test with significance level $\alpha = 0.05$. The critical difference is calculated using the studentized range distribution. Other than the Nemenyi procedure, no additional multiple comparison correction is applied. Since the blocks are taken from a continuous wind-power time series, there could be temporal dependence between consecutive blocks and the Friedman–Nemenyi results are therefore considered as a complementary statistical evidence and not as tests based on strictly independent realizations. In addition, the Diebold–Mariano (DM) test is used to compare the proposed model with each benchmark model pairwise. The comparison units are the individual testing samples and the loss function is the squared forecast error. The loss differential is, $d_t = e_{\text{Proposed},t}^2 - e_{\text{Benchmark},t}^2$. A negative DM statistic indicates that the proposed model has a lower squared-error loss. Serial dependence in the loss-differential series is captured by a Newey–West heteroskedasticity-and-autocorrelation-consistent (HAC) long-run variance estimator with Bartlett weights. The lag length is $\lfloor 4(n/100)^{2/9} \rfloor$, with n being the number of testing samples. Two-sided p -values are given for the pairwise comparisons.

3 Data Description

The first dataset used in this study is onshore wind power generation (WPG) data for Austria from the Open Power System Data platform (OPSD). The dataset is publicly accessible at <https://data.open-power-system-data.org/>. OPSD delivers high quality data on electricity generation and demand from a range of renewable energy sources. The data is aggregated from the European Network of Transmission System Operators for Electricity (ENTSO-E) at control area, country and bidding zone level. The dataset includes 2879 wind power samples at 15 min time resolution with timestamp information and corresponding onshore wind power measurements. Dataset 1 was extracted from the `AT_wind_onshore_generation_actual` field of the OPSD 15-min time-series file and covers 1 September 2020, 00:00 UTC to 30 September 2020, 23:30 UTC. The resulting series contains 2879 valid observations at 15-min resolution; the subsequent 23:45 UTC record contained a missing wind-power value and was therefore not included. Prior to the development of the model, the data set was analyzed for data quality and it was determined that there were no missing values, null values or major outliers, making it appropriate for forecasting analysis. Meteorological variables were not used as input features because the dataset represents the aggregated wind power generation of the whole 50 Hz control area. The second dataset was obtained from the National Energy System Operator (NESO) Open Data Portal and consists of operational metered wind power output for Great Britain. The dataset is publicly accessible at <https://www.neso.energy/data-portal/>. The dataset contains aggregated wind power generation measurements (MW) from England, Wales and Scotland derived from operational metering and records are presented by settlement date and settlement period. To get a full picture of national wind power production, the total wind power output is calculated by adding the operational metered generation from Scotland and England/Wales. The dataset is well suited for developing and evaluating the wind power forecasting models due to its high data quality, regular monthly updates and system-level coverage. Dataset 1 is recorded with a 15 min temporal resolution, Dataset 2 is recorded with a 30 min temporal resolution i.e., 48 settlement periods per day. The difference in temporal resolution allows for evaluation of the forecasting framework under different conditions of wind-power forecasting. Dataset 2 covers the observation period from 1 April 2025, to 5 April 2026, with measurements recorded at a 30-min temporal resolution.

The statistical features of the two wind power datasets used are summarized in Table 1, which highlights their different distributions and variabilities. Dataset 2 is substantially larger than Dataset 1, containing 17,713 samples compared to 2879 samples. It also exhibits considerably higher wind power values, with a mean of 8500.01 MW and a standard deviation of 4631.69 MW, indicating a much wider operating range. In contrast, Dataset 1 has a lower average power output (644.18 MW) but a more positively skewed distribution (skewness = 1.2976), suggesting frequent low-power observations with occasional high-power peaks. Dataset 2 shows a nearly symmetric distribution (skewness = 0.2193) and negative kurtosis (−1.1158), indicating a platykurtic distribution characterized by lighter tails relative to a normal distribution. The differences in sample size, statistical distribution and variability allow for the proposed forecasting framework to be tested with two different wind-power data characteristics. However, both datasets are based on aggregated wind-power generation, and thus the present evaluation does not verify generalization across geographically diverse regions or individual wind farms. Validation with additional regional and wind-farm level datasets is hence seen as an important direction in future work.

Table 1: Statistical characteristics of the two wind power datasets. Power-related statistics are reported in MW; standard deviation is calculated using the population convention (N denominator), and kurtosis is reported as Fisher excess kurtosis.

Statistic	Dataset 1	Dataset 2
Number of samples	2879	17,713
Mean	644.1806	8500.0116
Median	408.0000	7981.2900
Standard deviation	677.0636	4631.6940
Minimum	4.0000	196.8140
Maximum	2776.0000	18,426.3120
Range	2772.0000	18,229.4980
First quartile (Q1)	120.0000	4579.7060
Third quartile (Q3)	912.0000	12,425.8620
Interquartile range (IQR)	792.0000	7846.1560
Skewness	1.2976	0.2193
Kurtosis	0.8160	-1.1158

4 Results and Discussions

4.1 Comparative Analysis

In this section, we perform a comprehensive assessment of the proposed CEEMDAN-CNN-BiLSTM-Attention framework with Bayesian optimization on two public wind power datasets. The forecasting performance is evaluated first by visually comparing the actual and predicted wind power series, and then quantitatively by MAE, RMSE, MAPE, and R^2 . Additional analyses such as error distribution, scatter plots, sensitivity analysis, and Friedman Nemenyi statistical tests are performed to provide a comprehensive assessment of the proposed framework. The results are also compared with four benchmark deep learning models and the datasets with different statistical characteristics are considered to demonstrate the effectiveness and robustness of the proposed approach.

Fig. 2 compares the actual wind power output and the predictions of the proposed model and four benchmark deep learning models on Dataset 1. The proposed CEEMDAN-CNN-BiLSTM-Attention model almost follows the real wind power curve, and predicts the sharp fluctuations and peak values correctly, while the benchmark models show smoother responses and larger deviations, especially at the time of quick changes. These visual observations correspond well to the quantitative results that further verify the superior forecasting accuracy of the proposed framework.

Fig. 3 shows the actual wind power output compared with the predictions of the proposed model and four benchmark deep learning models on Dataset 2. The proposed CEEMDAN-CNN-BiLSTM-Attention model closely follows the actual wind power trajectory throughout the entire forecasting period, and successfully captures the rapid fluctuations and high-power peaks, while the benchmark models have relatively larger deviations, especially during abrupt transitions. The visual results are in line with the quantitative evaluation metrics, illustrating the better robustness and forecasting accuracy of the proposed framework for the larger and more complex Dataset 2.

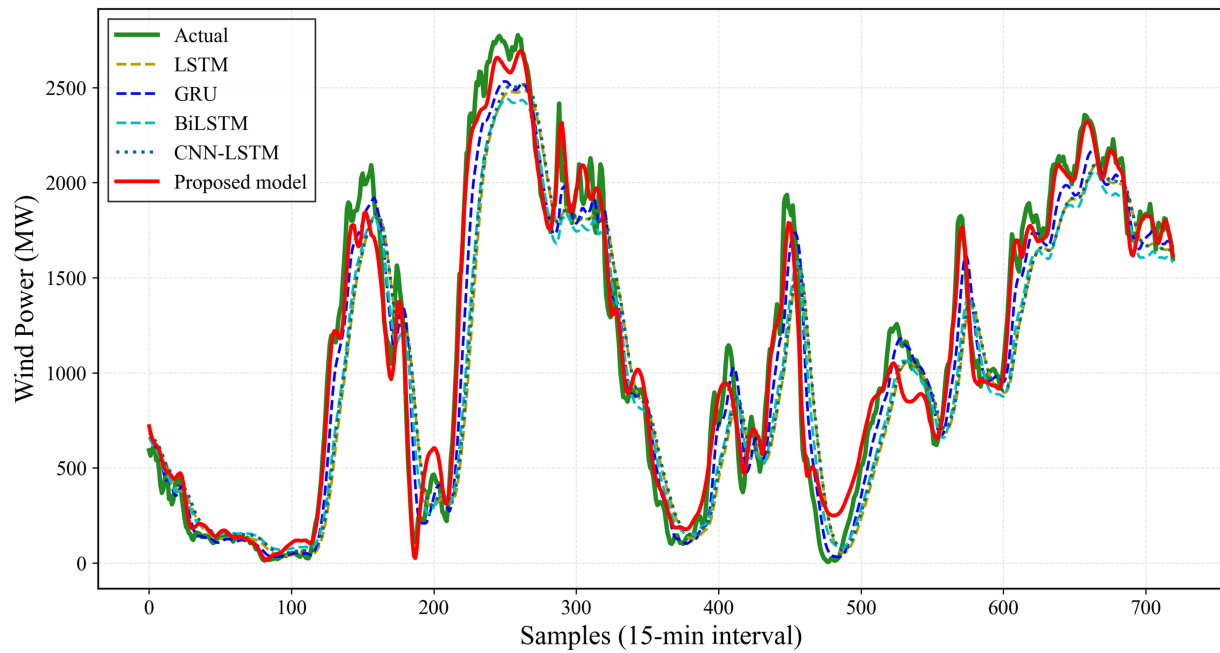


Figure 2: Prediction performance comparison for data set 1.

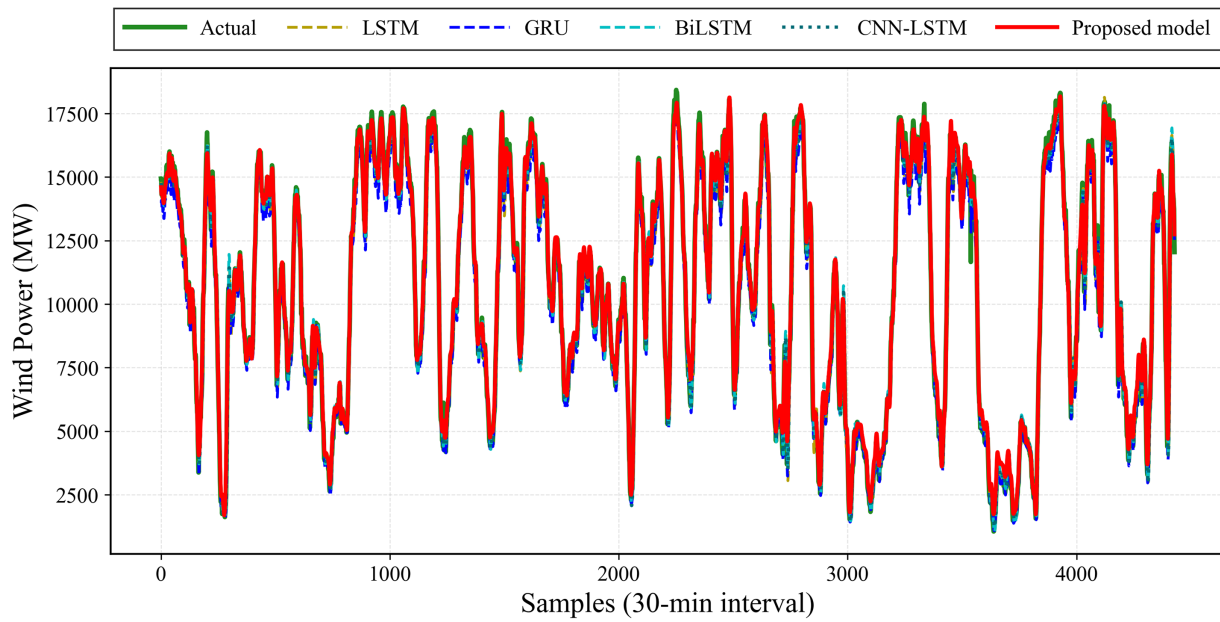


Figure 3: Prediction performance comparison for data set 2.

Figs. 4 and 5 show the comparison of forecasting errors between the proposed CEEMDAN-CNN-BiLSTM-Attention framework and the LSTM, BiLSTM, GRU and CNN-LSTM models for Dataset 1 and Dataset 2, respectively. In both datasets, the proposed model errors are more clustered around the zero-error reference line, showing less prediction bias and better forecasting stability. In comparison, the benchmark models have larger positive and negative error fluctuation, especially during the periods with rapid wind power variation, thus resulting in frequent high amplitude error spikes. For Dataset 1, the proposed model exhibits significant reduction of extreme forecasting errors along the entire prediction horizon, which indicates better tracking ability during the abrupt changes of wind power output. A similar trend is observed for Dataset 2 where the proposed framework maintains comparatively smaller error magnitudes even though the dataset size is larger and the wind power series is more variable. These visual observations agree with the quantitative evaluation metrics where the proposed model achieves the lowest MAE and RMSE, and the highest coefficient of determination (R^2) than all competing methods. In summary, the aforementioned error analysis indicates that the CEEMDAN decomposition, CNN feature extraction, BiLSTM temporal learning, attention mechanism for feature weighting, and Bayesian hyperparameter optimization have significantly improved the robustness, stability, and accuracy of the wind power forecasting at multiple temporal resolutions. The proposed framework employs a single-step-ahead forecasting strategy and is evaluated on two wind power datasets with different temporal resolutions, such that each prediction corresponds to the next sampling instant of the respective dataset.

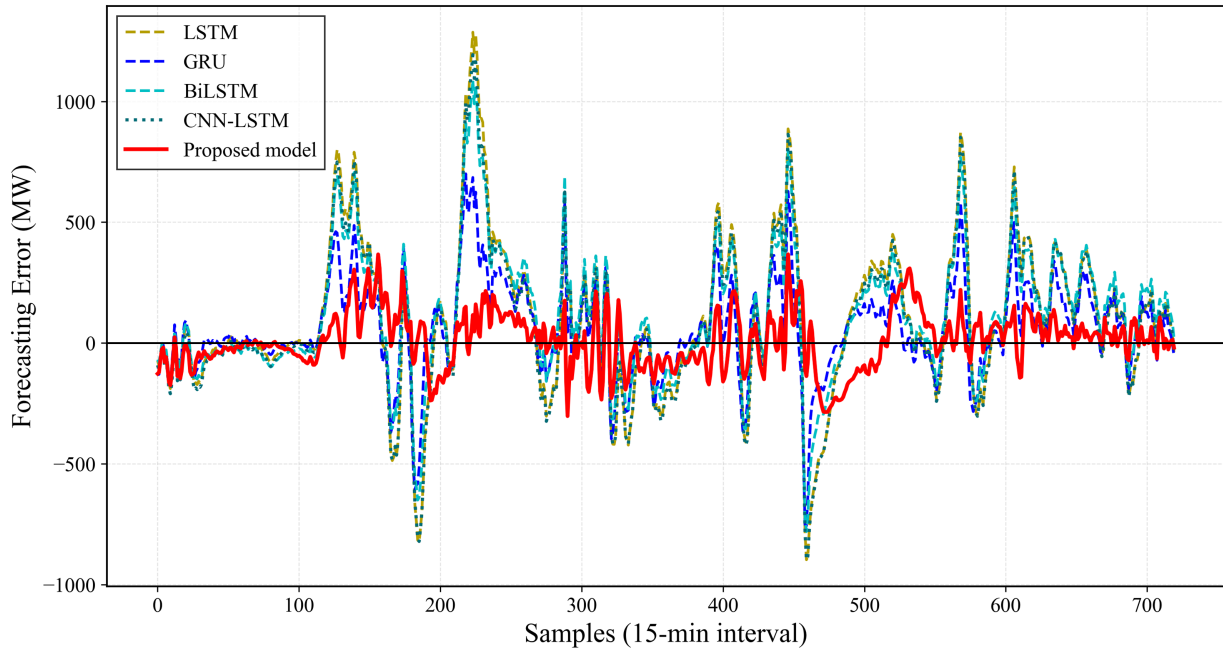


Figure 4: Forecasting errors for Dataset 1, defined as $e_t = P_t - \hat{P}_t$ (actual minus predicted). Positive errors indicate underprediction, whereas negative errors indicate overprediction.

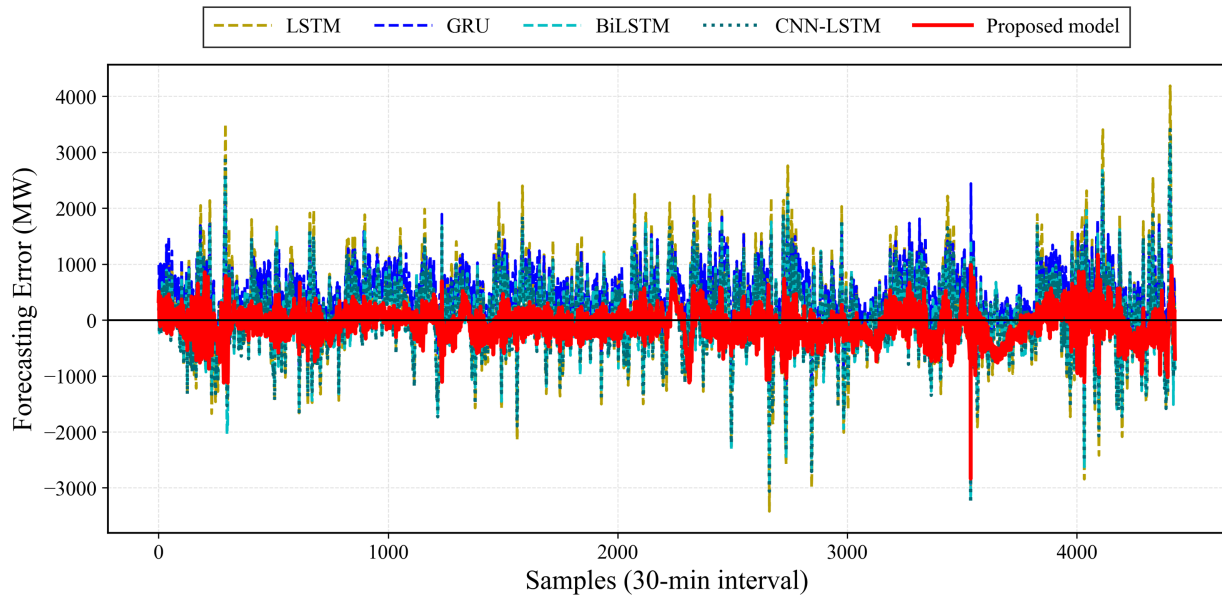


Figure 5: Forecasting errors for Dataset 2, defined as $e_t = P_t - \hat{P}_t$ (actual minus predicted). Positive errors indicate underprediction, whereas negative errors indicate overprediction.

The training and testing performance of the proposed CEEMDAN-CNN-BiLSTM-Attention framework and four benchmark deep learning models on Datasets 1 and 2 are shown in [Tables 2 and 3](#), respectively. For the Dataset 1, the proposed model achieves the minimum testing MAE (89.56 MW) and RMSE (115.36 MW) and the maximum coefficient of determination ($R^2 = 0.9796$) among the compared models. The MAE and RMSE in testing are larger than that in training, which indicates that there is a significant training-testing generalization gap. However, the proposed model still demonstrates good predictive performance in the test set. The proposed model does not give the lowest MAPE on Dataset 1, hence the performance gain is specifically expressed in the absolute-error and goodness-of-fit metrics and not in all evaluation metrics. The effectiveness of the proposed framework is more evident on Dataset 2, where it achieves the lowest testing MAE (230.19 MW), RMSE (299.59 MW), and MAPE (3.29%), and the highest R^2 value of 0.9957 as shown in [Table 3](#). The consistently small differences between the training and testing metrics indicate that the proposed model generalizes well without significant overfitting. In contrast, the testing errors of the benchmark models increase more noticeably, especially for LSTM and CNN-LSTM, which indicates weaker generalization ability. The improved forecasting performance is attributed to the synergistic combination of CEEMDAN-based signal decomposition, CNN feature extraction, BiLSTM temporal dependency learning, attention-based feature weighting, and Bayesian hyperparameter optimization. To make a more robust assessment, especially when close-zero wind-power observations happen, [Table 4](#) presents the testing performance with sMAPE, nMAE, and nRMSE. The lowest sMAPE of 19.24% and the lowest nMAE (0.0323) and nRMSE (0.0416) are achieved using the proposed model in Dataset 1 with 21.69% of the second best GRU model. The results indicate that the relatively high conventional MAPE of the proposed model on Dataset 1 is affected by the sensitivity of MAPE to the observations around zero and does not describe the overall error magnitude of the model. The proposed framework also consistently achieves the best values for Dataset 2, obtaining a sMAPE of 3.16%, nMAE of 0.0133 and nRMSE of 0.0172. Overall, the extra robust metrics further evidence that the proposed framework achieves better forecasting performance in both datasets while avoiding the potential misleading behavior of conventional MAPE near zero.

Table 2: Training and testing performance comparison for data set 1.

Model	MAPE (%)		RMSE (MW)		MAE (MW)		R^2	
	Train	Test	Train	Test	Train	Test	Train	Test
LSTM	68.89	64.22	180.94	333.12	116.72	240.77	0.8858	0.8300
BiLSTM	108.69	55.06	154.81	287.18	104.16	213.40	0.9164	0.8736
GRU	45.16	28.89	114.47	201.94	71.40	146.14	0.9543	0.9375
CNN-LSTM	94.47	67.78	175.16	318.85	115.00	234.17	0.8930	0.8442
Proposed	72.91	55.12	49.18	115.36	36.60	89.56	0.9918	0.9796

Table 3: Training and testing performance comparison for data set 2.

Model	MAPE (%)		RMSE (MW)		MAE (MW)		R^2	
	Train	Test	Train	Test	Train	Test	Train	Test
LSTM	9.64	6.73	693.35	773.33	522.32	585.30	0.9753	0.9713
BiLSTM	7.39	5.32	547.62	602.32	415.38	453.17	0.9846	0.9826
GRU	6.37	5.23	520.58	632.30	397.26	516.95	0.9861	0.9808
CNN-LSTM	8.49	5.65	578.75	634.97	440.07	479.02	0.9828	0.9806
Proposed	3.15	3.29	236.03	299.59	177.90	230.19	0.9971	0.9957

Table 4: Robust testing performance comparison using sMAPE, nMAE, and nRMSE for Datasets 1 and 2. nMAE and nRMSE are normalized by the range of the actual wind-power observations in the corresponding test set.

Dataset	Model	sMAPE (%)	nMAE	nRMSE
Dataset 1	LSTM	34.19	0.0869	0.1202
	BiLSTM	31.80	0.0770	0.1036
	GRU	21.69	0.0527	0.0729
	CNN-LSTM	34.24	0.0845	0.1150
	Proposed	19.24	0.0323	0.0416
Dataset 2	LSTM	6.72	0.0337	0.0445
	BiLSTM	5.33	0.0261	0.0347
	GRU	5.40	0.0298	0.0364
	CNN-LSTM	5.62	0.0276	0.0366
	Proposed	3.16	0.0133	0.0172

Fig. 6 shows the density-based scatter plots of predicted and actual wind power for Dataset 1 for LSTM, BiLSTM, GRU, CNN-LSTM, and the proposed CEEMDAN-CNN-BiLSTM-Attention models. The dashed diagonal line is the ideal prediction line where the predicted values equal the actual measurements. Among the benchmark models, the GRU has the best agreement with the ideal line, followed by the BiLSTM. The LSTM and CNN-LSTM have larger dispersion and more significant deviations, especially at higher wind power levels. The proposed model shows prediction points more closely aligned with the ideal line and with lower overall dispersion than the benchmark models, indicating improved agreement between the predicted and actual wind-power values. This observation is further supported by the corresponding coefficient of determination, where the highest value ($R^2 = 0.9796$) is obtained by the proposed model,

compared with 0.9375 for GRU, 0.8736 for BiLSTM, 0.8442 for CNN-LSTM and 0.8300 for LSTM. In addition, the proposed framework significantly alleviates the propagation of prediction errors under both low and high wind power output levels, which reveals its ability to accurately model the nonlinear variations of wind power series. These visual outcomes are in agreement with the quantitative performance metrics and validate that the presented framework provides the most reliable and robust forecasting performance over all models tested. Fig. 7 shows the performance of the proposed model compared to the benchmark models for dataset 2. The results in Fig. 7 are obtained using the original evaluation protocol and therefore should not be interpreted as leakage-controlled out-of-sample estimates. The leakage-controlled analysis on Dataset 1 shows that the reported performance can be reduced by excluding future information. Therefore, the results on Dataset 2 are considered as potentially optimistic and extending the leakage-controlled evaluation to Dataset 2 is left as future work.

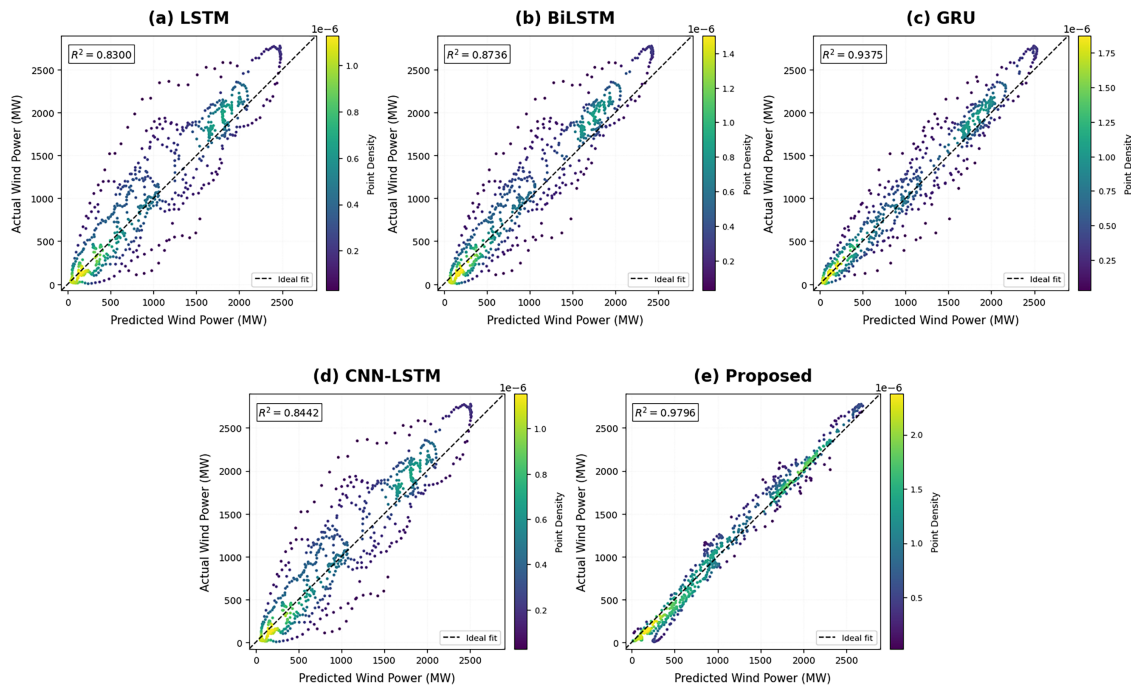


Figure 6: Scatter plots for Dataset 1. The color scale represents point density and is normalized independently within each model panel; therefore, color values indicate relative point concentration within a panel and are not intended for direct quantitative density comparison across models.

Fig. 8 shows the bar chart comparison of the testing error metrics, MAPE, MAE, and RMSE for the five forecasting models on (a) Dataset 1 and (b) Dataset 2. In Dataset 1, the proposed CEEMDAN-CNN-BiLSTM-Attention model has the lowest MAE and RMSE values among all models. The GRU model achieves the lowest MAPE value, indicating competitive accuracy in terms of percentage. However, the LSTM and CNN-LSTM models always have the largest prediction errors for all three evaluation metrics. For Dataset 2, the proposed model obtains the lowest MAPE (3.29%), MAE (230.19 MW) and RMSE (299.59 MW) among all the benchmark models, showing its better forecasting performance. However, the BiLSTM and GRU models perform relatively better than the LSTM and CNN-LSTM models but still perform significantly worse than the proposed framework in terms of absolute forecasting accuracy. The consistent decrease in error metrics across both datasets shows the effectiveness of the combination of CEEMDAN-based signal decomposition,

CNN-based feature extraction, BiLSTM-based temporal modeling, attention-based feature weighting and Bayesian hyperparameter optimization.

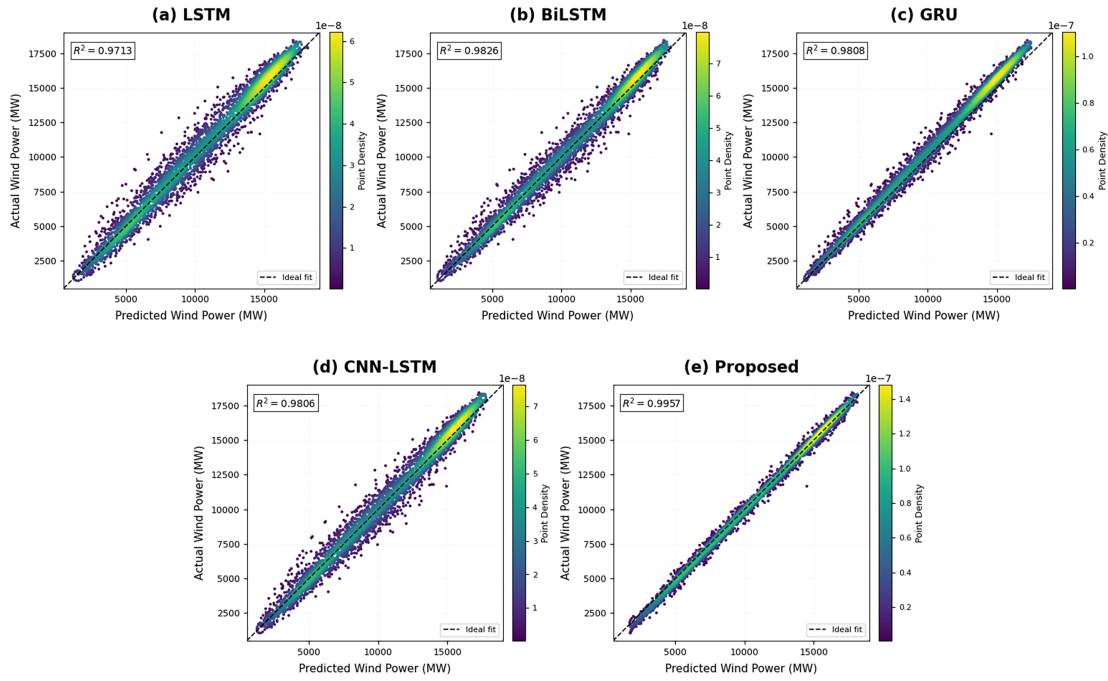


Figure 7: Scatter plots for Dataset 2 under the original evaluation protocol. The reported results are for full-series normalization and CEEMDAN decomposition prior to the chronological partition and should therefore be interpreted in light of the possible information leakage due to pre-processing discussed in [Section 4.2](#).

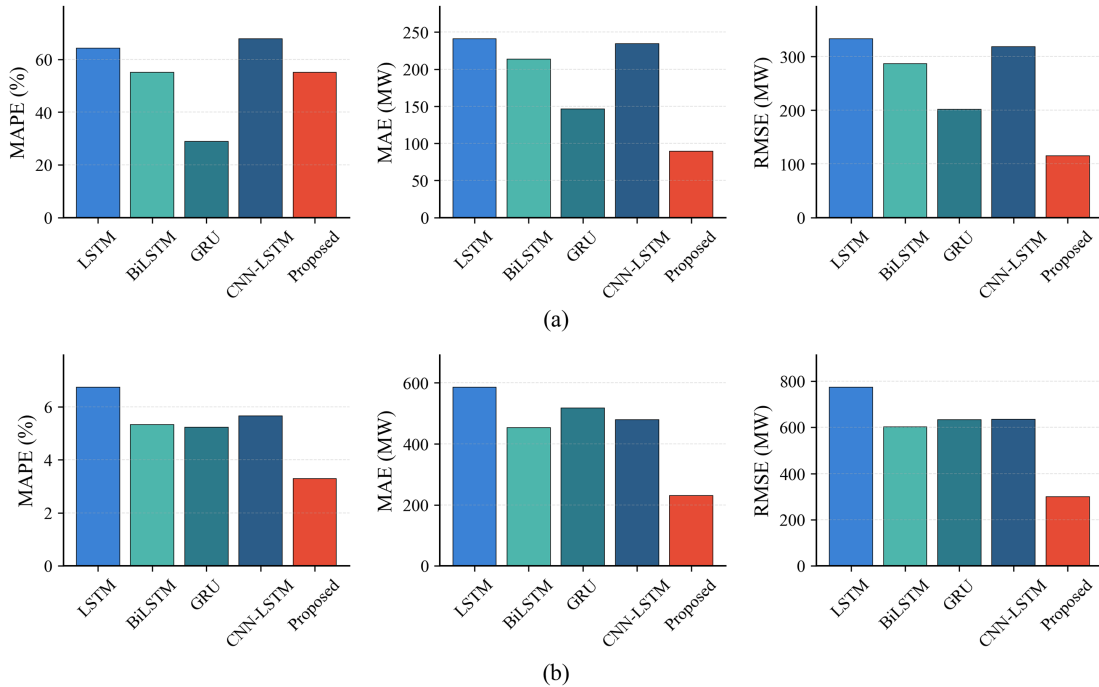


Figure 8: Bar plots of error matrices (a) Data set 1 (b) Data set 2.

Table 5 shows the block-wise sensitivity analysis of the forecasting performance on Dataset 1 using MAE, RMSE, and MAPE metrics. The Dataset 1 testing set contains 720 samples, which are divided into 15 non-overlapping blocks of 48 samples. Since all blocks contain an equal number of samples, the arithmetic means of the block-wise MAE and MAPE are identical to the corresponding overall testing MAE and MAPE reported in **Table 2**. In contrast, the mean of the block-wise RMSE values may differ from the overall testing RMSE because RMSE is a nonlinear metric. For Dataset 2, 92 complete blocks (4416 samples) are used, while the remaining 13 testing samples are excluded from the block-wise sensitivity analysis. For the absolute error metrics, the proposed CEEMDAN–CNN–BiLSTM–Attention model performs best with the lowest mean MAE (89.56 MW) and RMSE (106.09 MW), respectively, while the second-best GRU model has corresponding values of 146.14 and 182.95 MW. In addition, the proposed model has the smallest maximum MAE (155.60 MW) and RMSE (181.85 MW), which are substantially lower than those of the LSTM model (504.28 and 656.02 MW), indicating fewer large absolute forecasting errors. The proposed model also has the lowest standard deviations of MAE (39.05 MW) and RMSE (45.31 MW), indicating more consistent absolute-error performance across the testing blocks. For the MAPE metric, the GRU model has the lowest mean (28.89%), maximum (127.59%), and standard deviation (29.29%). The proposed model has the lowest minimum MAPE (2.03%) and the lowest median MAPE (18.60%). The relatively higher mean MAPE of the proposed model (55.12%) is mainly associated with low wind-power periods, during which percentage-based errors can become disproportionately large despite comparatively small absolute forecasting errors.

Table 5: Sensitivity analysis for Dataset 1.

MAE (MW)	Mean	Median	Maximum	Minimum	Std.
LSTM	240.77	231.23	504.28	23.11	129.02
BiLSTM	213.40	197.77	432.29	42.37	101.31
GRU	146.14	148.54	280.23	16.39	68.56
CNN-LSTM	234.17	215.05	475.37	35.87	118.85
Proposed	89.56	83.19	155.60	14.67	39.05
RMSE (MW)	Mean	Median	Maximum	Minimum	Std.
LSTM	292.77	263.78	656.02	32.42	158.91
BiLSTM	257.91	259.68	554.65	49.19	126.30
GRU	182.95	193.77	350.61	20.43	85.49
CNN-LSTM	283.75	250.47	602.12	44.20	145.43
Proposed	106.09	108.70	181.85	17.46	45.31
MAPE (%)	Mean	Median	Maximum	Minimum	Std.
LSTM	64.22	37.24	415.01	6.51	97.79
BiLSTM	55.06	33.23	270.89	8.06	66.73
GRU	28.89	24.52	127.59	5.16	29.29
CNN-LSTM	67.78	37.64	415.09	6.47	98.60
Proposed	55.12	18.60	373.47	2.03	97.49

Table 6 presents the sensitivity analysis of the forecasting performance for Dataset 2 with MAE, RMSE and MAPE statistics calculated over 92 complete non-overlapping testing blocks, comprising 4416 of the 4429 testing samples; the remaining 13 samples are excluded because they do not form a complete 48-sample block. The proposed CEEMDAN–CNN–BiLSTM–Attention framework achieves the lowest mean MAE (229.99

MW) and RMSE (278.25 MW), which are much better than the best competing model, BiLSTM, with mean values of 448.11 and 559.00 MW, respectively. In addition, the proposed model has the minimum median, maximum, minimum and standard deviation in both MAE and RMSE. The maximum MAE is only 570.53 MW with the standard deviation of 94.38 MW, which indicates not only higher accuracy in prediction, but also much higher stability and less variation under various operating conditions. The proposed model attains the least mean MAPE (3.30%), median MAPE (2.08%) and minimum MAPE (0.57%) for the percentage error metric, indicating its effectiveness over other models in terms of overall forecasting accuracy. The GRU model has the minimum maximum MAPE (11.42%) and the minimum MAPE standard deviation (1.37%) but its average absolute errors are still much larger than the proposed framework, which demonstrates the less effectiveness of the GRU model in minimizing the overall prediction deviations. The sensitivity analysis, in general, shows that the proposed model has the most reliable and consistent forecasting performance on almost all statistical indicators, and it has robustness for wind power forecasting at multiple temporal resolutions under different operating conditions. The sensitivity analysis is carried out by splitting the testing data into sequential blocks of 48 samples and calculating the MAE, RMSE and MAPE for each block separately. For Dataset 2, 4416 samples (92 complete blocks) are included in the block-wise analysis, while the remaining 13 samples are excluded. Consequently, the block-wise mean metrics may differ slightly from the corresponding overall testing results reported in Table 3. Although nonoverlapping blocks were used to reduce direct overlap among the statistical evaluation units, temporal dependence may still exist between successive blocks, as the observations are from continuous wind-power time series. Thus, the block-wise Friedman and Nemenyi analyzes are to be considered as complementary statistical tests, and not as evidence from strictly independent realizations.

Table 6: Sensitivity analysis for Dataset 2.

MAE (MW)	Mean	Median	Maximum	Minimum	Std.
LSTM	578.89	538.31	1429.21	206.50	188.90
BiLSTM	448.11	420.63	1040.43	214.64	139.01
GRU	513.95	513.07	900.92	109.80	173.33
CNN-LSTM	474.06	448.23	1145.36	185.38	149.35
Proposed	229.99	206.22	570.53	88.33	94.38
RMSE (MW)	Mean	Median	Maximum	Minimum	Std.
LSTM	719.39	681.62	1590.40	258.43	230.63
BiLSTM	559.00	531.61	1229.58	262.23	176.79
GRU	600.62	608.59	1002.26	140.80	178.96
CNN-LSTM	590.72	554.41	1298.97	243.98	183.63
Proposed	278.25	249.70	580.76	109.13	110.57
MAPE (%)	Mean	Median	Maximum	Minimum	Std.
LSTM	6.30	5.48	19.85	2.43	3.18
BiLSTM	4.99	4.42	15.48	1.59	2.61
GRU	5.02	4.66	11.42	3.02	1.37
CNN-LSTM	5.29	4.55	16.37	1.64	2.80
Proposed	3.30	2.08	22.58	0.57	3.58

Fig. 9 shows the Critical Difference (CD) diagram obtained from the Friedman test followed by the Nemenyi post-hoc analysis for Dataset 1. The proposed CEEMDAN-CNN-BiLSTM-Attention model exhibits the best average rank of 1.20, which reveals the highest overall forecasting performance among all the compared methods. The performance of the GRU model is the second best with an average rank of 1.80, while BiLSTM (3.47), CNN-LSTM (3.93) and LSTM (4.60) perform relatively worse. The calculated critical difference is 1.575. The horizontal bars of significance indicate that the proposed model and GRU are not significantly different from each other, and BiLSTM, CNN-LSTM, and LSTM constitute the second group that cannot be statistically separated. However, the difference between these two groups is bigger than the critical difference, indicating that the proposed framework and GRU are significantly better than the other benchmark models.

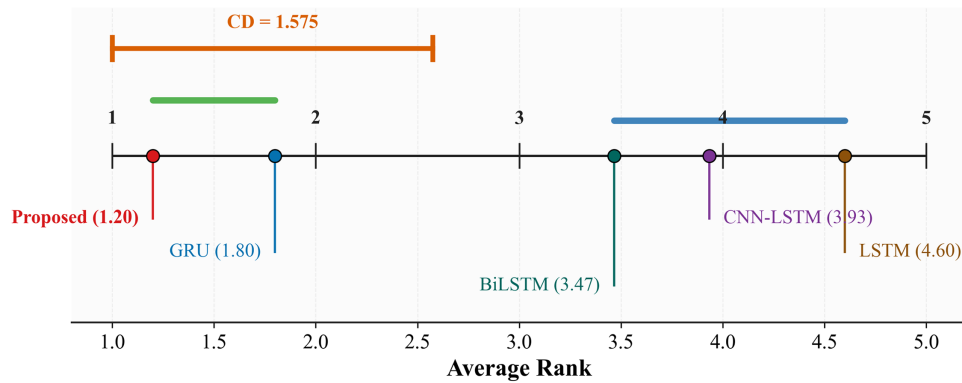


Figure 9: Critical difference (CD) diagram based on the Friedman and Nemenyi post-hoc tests for Dataset 1. Average ranks are calculated from block-wise MAE over non-overlapping blocks of 48 testing samples, with lower MAE corresponding to a better (lower) rank.

The CD diagram for Dataset 2 is shown in Fig. 10. In ascending order of ranks, the best average rank is 1.28 for the proposed CEEMDAN-CNN-BiLSTM-Attention model, as compared with 2.72, 3.03, 3.36, and 4.61 for BiLSTM, CNN-LSTM, GRU, and LSTM, respectively. The CD value is 0.636, and the average-rank differences between the proposed model and all benchmark models exceed this threshold. Thus, the Nemenyi comparison indicates differences in favor of the proposed model at the 5% significance level; however, these results should be interpreted cautiously because temporal dependence may remain between successive blocks. Moreover, BiLSTM, CNN-LSTM, and GRU constitute a statistically indistinguishable group, while LSTM exhibits the worst average rank. These results provide complementary statistical evidence supporting the favorable forecasting performance of the proposed framework on Dataset 2.

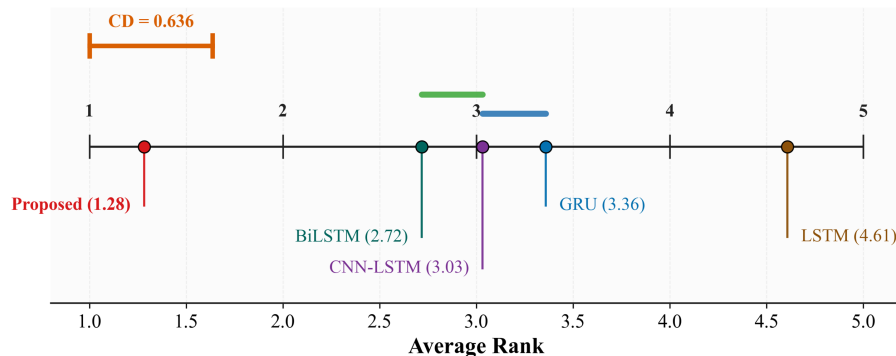


Figure 10: Critical difference (CD) diagram based on the Friedman and Nemenyi post-hoc tests for Dataset 2.

Table 7 shows the average model ranks from the Friedman test, along with the associated CD values from the Nemenyi post-hoc test. The proposed CEEMDAN–CNN–BiLSTM–Attention framework obtains the best performance with the lowest average rank on Dataset 1 (1.20) and Dataset 2 (1.28) among the evaluated models. However, for Dataset 1 the Nemenyi comparison indicates no significant difference between the proposed model and GRU, since the difference of their average-ranks is less than the critical difference. Moreover, the Friedman–Nemenyi results are interpreted as complementary statistical evidence rather than as significance tests based on strictly independent realizations, because successive blocks may still be temporally dependent.

Table 7: Average model ranks obtained using the Friedman test. Lower average rank indicates better forecasting performance. The critical difference (CD) is computed using the Nemenyi post-hoc test at the 5% significance level.

Dataset	Blocks	CD	Proposed	GRU	BiLSTM	CNN–LSTM	LSTM
Dataset 1	15	1.57	1.20	1.80	3.47	3.93	4.60
Dataset 2	92	0.64	1.28	3.36	2.72	3.03	4.61

The results of the Diebold–Mariano (DM) test provide additional statistical evidence on the forecasting performance of the proposed model as shown in Table 8. The DM test is conducted for the one-step-ahead forecasts using the full chronological test sequence rather than the block-level observations. All DM statistics are negative which implies that the proposed model has a smaller squared-error loss than each benchmark for both datasets. For Dataset 1, the DM statistics range from -4.4894 to -5.4001 and for Dataset 2 the magnitudes are even larger ranging from -12.1525 to -19.2972 . All pairwise comparisons are statistically significant ($p < 0.05$), proving that the proposed model has a statistically significant lower forecast loss than LSTM, BiLSTM, GRU and CNN–LSTM using the DM test.

Table 8: Diebold–Mariano test results comparing the proposed model with benchmark models.

Dataset	Comparison	DM Statistic	p -Value	Result
Dataset 1	Proposed vs. LSTM	-5.2068	<0.000001	Significantly better
Dataset 1	Proposed vs. BiLSTM	-5.2514	<0.000001	Significantly better
Dataset 1	Proposed vs. GRU	-4.4894	0.000007	Significantly better
Dataset 1	Proposed vs. CNN–LSTM	-5.4001	<0.000001	Significantly better
Dataset 2	Proposed vs. LSTM	-13.8551	<0.000001	Significantly better
Dataset 2	Proposed vs. BiLSTM	-12.1525	<0.000001	Significantly better
Dataset 2	Proposed vs. GRU	-19.2972	<0.000001	Significantly better
Dataset 2	Proposed vs. CNN–LSTM	-12.8814	<0.000001	Significantly better

To further investigate the contribution of the main trainable components of the proposed framework, we conducted a targeted component contribution analysis with Dataset 1 with the results summarized in Table 9. For the ablation configurations, “Without Attention” replaces the attention pooling layer with global average pooling while retaining the remaining architecture and optimized hyperparameters. “Without BiLSTM” replaces the two BiLSTM layers with conventional LSTM layers. For “Without Bayesian Optimization,” fixed manually specified hyperparameters are used: lookback window = 24, CNN filters = 32, kernel size = 3, recurrent units = 64 and 32, dense units = 32, dropout = 0.10, batch size = 16, and learning rate = 0.001. The MAE and RMSE increase from 89.56 MW and 115.36 to 95.91 MW and 121.18 MW respectively, and R^2

decreases slightly from 0.9796 to 0.9775, indicating that adaptive temporal feature weighting is beneficial for the absolute forecasting accuracy. The performance degradation is much larger when substituting the BiLSTM layers with conventional LSTM layers, which obtain MAE, RMSE, and R^2 of 187.53 MW, 232.01 MW, and 0.9175, respectively, proving the significance of bidirectional temporal feature learning in the proposed architecture. In the same way, the absence of Bayesian optimization and the adoption of manually specified hyperparameters raises the MAE and RMSE to 176.26 and 231.54 MW, respectively, while R^2 drops to 0.9178, corroborating the relevance of systematic hyperparameter optimization. The variant without attention has a lower MAPE (40.70%) than the full framework (55.12%), which is in line with the aforementioned sensitivity of MAPE to low-power observations in Dataset 1, but the full framework has the lowest MAE and RMSE, and the highest R^2 across the evaluated configurations. The results in general provide quantitative evidence that the attention mechanism, bidirectional temporal learning and Bayesian optimization contribute to the forecasting capability of the integrated framework in a complementary way.

Table 9: Targeted component contribution analysis for Dataset 1.

Configuration	MAE (MW)	RMSE (MW)	MAPE (%)	R^2
Without Attention	95.91	121.18	40.70	0.9775
BiLSTM replaced by LSTM	187.53	232.01	131.81	0.9175
Without Bayesian Optimization	176.26	231.54	84.91	0.9178
Full proposed framework	89.56	115.36	55.12	0.9796

4.2 Leakage-Controlled Experiment Results

To further study the robustness of the proposed framework against potential leakage of future information, we also conducted an experiment with leakage control on Dataset 1. For this experiment, the data was first split into training and testing sets chronologically with the same 75:25 ratio and the normalization parameters were only calculated on the training data. The CEEMDAN decomposition used for model training was performed only on the training data. During testing, the trained component models were kept fixed. For each test sample at time t , CEEMDAN was recomputed using only the observations available up to $t - 1$, and the most recent lookback window of each resulting component was used to generate the one-step-ahead prediction. Thus, neither the current target nor any future test observations were used during decomposition or prediction. This evaluation therefore offers a more rigorous assessment of the forecasting performance in a causal and leakage-controlled environment.

As can be seen in Table 10, the stricter leakage-controlled evaluation leads to larger forecasting errors compared to the original evaluation protocol. The MAE increases from 89.56 to 178.05 MW. The RMSE increases from 115.36 to 246.59 MW and the MAPE increases from 55.12% to 82.42%. At the same time, the R^2 drops from 0.9796 to 0.9068, suggesting a significant impact of the preprocessing and decomposition protocol on the reported forecasting performance. The proposed framework, however, still retains a R^2 above 0.90 under the leakage-controlled walk-forward evaluation, suggesting that it retains a significant predictive ability when strictly excluding future test information. These results lead to a more cautious assessment of the model generalization and point to the importance of causal pre-processing in the case of decomposition based forecasting models for practical real-time applications.

The actual and predicted wind power for the 720 testing samples in the leakage-controlled walk-forward evaluation are shown in Fig. 11. The predicted trajectory generally follows the main temporal variations of the actual wind power, including the fluctuations from near 0 MW to about 2800 MW, but with noticeable deviations in several rapid transitions and high-power periods. The proposed model achieved

an MAE of 178.05 MW, an RMSE of 246.59 MW, and an R^2 of 0.9068 with the proposed more rigorous evaluation protocol, which confirms that the model has high predictive power even without using any future information during the pre-processing phase. The remaining discrepancies, especially around sharp peaks and sudden changes, are due to the increased difficulty of causal walk-forward forecasting compared to the original full-sequence decomposition setting.

Table 10: Comparison of the original and leakage-controlled evaluation protocols for the proposed model on Dataset 1.

Evaluation Protocol	MAE (MW)	RMSE (MW)	MAPE (%)	R^2
Original evaluation	89.56	115.36	55.12	0.9796
Leakage-controlled walk-forward evaluation	178.05	246.59	82.42	0.9068

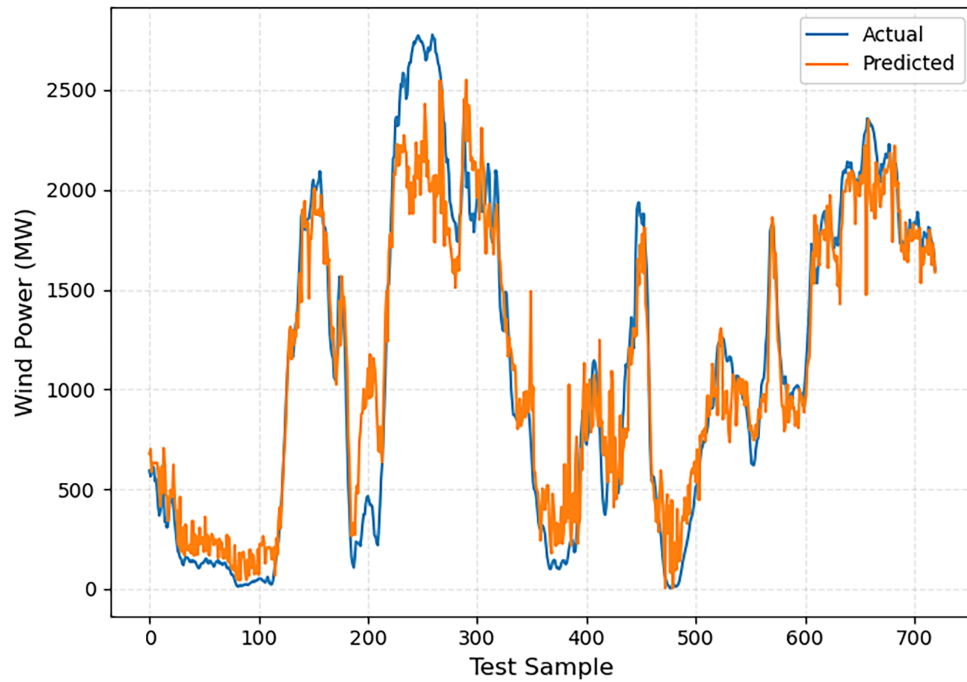


Figure 11: Actual and predicted wind power under the leakage-controlled walk-forward evaluation for Dataset 1.

The density based scatter plot between actual and predicted wind power values under the leakage-controlled walk-forward evaluation is shown in Fig. 12. Most of the observations lie near the ideal 45° line. The point density is particularly large in the lower power region below about 500 MW. Also, there is substantial agreement observed around 800–1100 MW and 1700–2200 MW. Results obtained show that the model explains about 90.68% of the variation of the observed wind power in the leakage-controlled setting. Although some intermediate and high-power operating points exhibit higher dispersion, the overall distribution indicates that the proposed framework still achieves a good agreement between predicted and actual values under the more rigorous causal evaluation protocol.

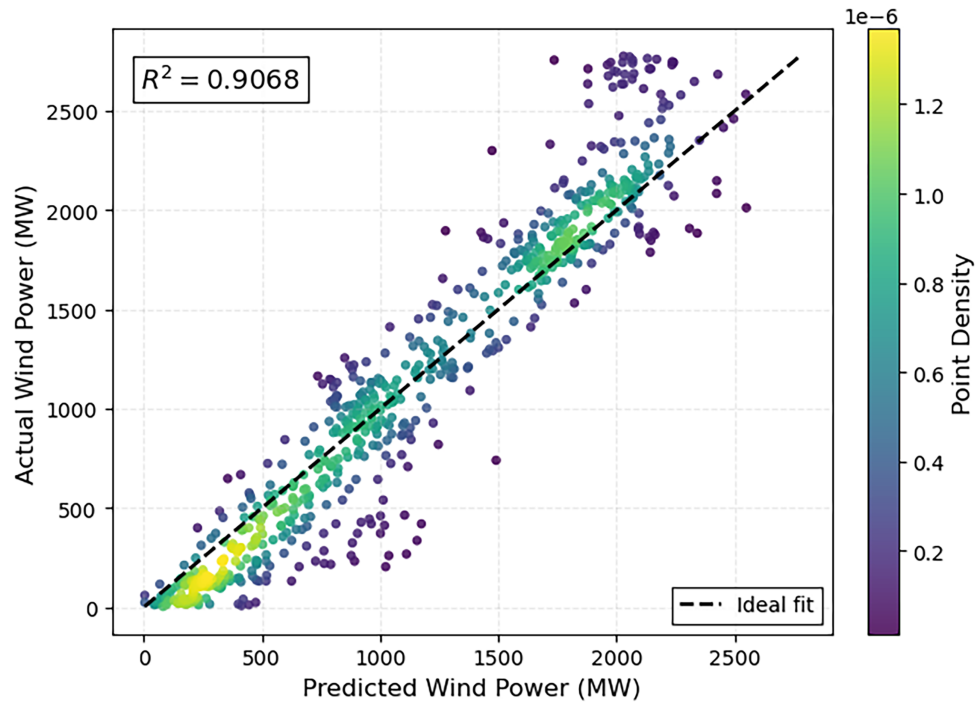


Figure 12: Scatter plot for Dataset 1 under leakage-controlled experiment.

All experiments were carried out on datasets in the original chronological order without random shuffling and were split with a ratio of 75% training and 25% testing. In Bayesian optimization, 65% of the data was used to train the model, 10% of the data was used for validation and 25% of the data was only used for testing. The use of a fixed random seed of 123 led to better experimental reproducibility. The CEEMDAN decomposition was performed using PyEMD, and all the components generated, including the residual, were kept for further prediction and reconstruction. We trained models using early stopping and used Bayesian optimization with eight trials to explore a nine-dimensional hyperparameter search space. The computational characteristics of the proposed framework were evaluated on a Windows 11 system equipped with an Intel Core Ultra 7 155H processor and 16 GB RAM, without TensorFlow GPU acceleration. For Dataset 1, CEEMDAN generated eight components, resulting in eight CNN-BiLSTM-Attention models with 185,857 trainable parameters each and 1,486,856 parameters in total. The total parameter storage is approximately 5.67 MB using FP32 representation. For 720 testing samples, the eight component models required a combined neural-network inference time of 2.52 s, corresponding to approximately 3.50 ms per forecast sample. The results demonstrate relatively low inference latency, while the offline training stage remains computationally more demanding because separate models are trained for individual CEEMDAN components and Bayesian optimization requires repeated model evaluations.

5 Conclusion

In this paper, a hybrid deep learning framework for wind power forecasting at multiple temporal resolutions is proposed by integrating CEEMDAN, CNN, BiLSTM, Attention mechanism and Bayesian Optimization. The proposed methodology unifies the signal decomposition, local feature extraction, bidirectional temporal dependency learning, adaptive feature weighting and automatic hyper-parameter optimization into one framework for forecasting. The proposed method first decomposes the original wind power signal into multiple intrinsic mode functions before prediction, which can greatly reduce the complexity of the signal

and learn nonlinear temporal patterns more effectively. We conducted a comprehensive assessment of the proposed framework on two real-world wind power datasets with different operating characteristics and temporal resolutions. The proposed framework achieved the best overall performance in terms of MAE, RMSE, and R^2 on both datasets compared with LSTM, BiLSTM, GRU, and CNN-LSTM, except that GRU achieved a lower MAPE on Dataset 1. On Dataset 1, the proposed model achieves the MAE of 89.56 MW, RMSE of 115.36 MW, MAPE of 55.12% and R^2 of 0.9796. The proposed framework further enhanced the forecasting performance on Dataset 2 with MAE of 230.19 MW, RMSE of 299.59 MW, MAPE of 3.29%, and R^2 of 0.9957. These results are consistent with the original evaluation protocol and should be considered in the light of its potential information leakage due to preprocessing. The framework achieved an MAE of 178.05 MW and a R^2 of 0.9068 under the leakage-controlled evaluation on Dataset 1, which indicates the lower but still significant predictive performance under the causal evaluation. The visual analysis, residual analysis, density scatter plots, error distributions and prediction trajectory comparisons all showed that the predicted and measured wind power outputs were in excellent agreement. To further validate the reliability of the proposed approach, statistical analyses were conducted by Friedman test and Nemenyi post-hoc test. Complementary statistical analyzes provided supporting evidence for the forecasting performance of the proposed framework. The DM tests showed significant pairwise differences when compared to the benchmark models, and the Friedman–Nemenyi test showed no significant difference between the proposed model and GRU on Dataset 1. Future work will consider probabilistic wind power forecasting, multi-step forecasting horizons, transfer learning across geographically diverse wind farms, uncertainty quantification, and lightweight transformer-based architectures for real-time deployment in large-scale renewable energy management systems.

Acknowledgement: The author would like to acknowledge the support provided by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia.

Funding Statement: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU264467].

Availability of Data and Materials: The datasets used in this study are publicly available from the Open Power System Data (OPSD) platform and the National Energy System Operator (NESO) data portal. The corresponding data sources and access links are provided in the manuscript.

Ethics Approval: This study did not involve human participants, animals, or personally identifiable information; therefore, ethical approval was not required. All analyses were conducted using publicly available wind-power datasets.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Hossain MA, Gray E, Lu J, Islam MR, Alam MS, Chakraborty R, et al. Optimized forecasting model to improve the accuracy of very short-term wind power prediction. *IEEE Trans Ind Inform.* 2023;19(10):10145–59. doi:10.1109/tii.2022.3230726.
2. Song Z, Gu Y, Liu H, Zou T, Lin Y, Ye K. Application of deep learning in wind, solar, and ocean energy: an analysis of prediction, optimization, and operation & maintenance. *Renew Sustain Energy Rev.* 2026;230:116663.
3. Alam MS, Al-Ismail FS, Salem A, Abido MA. High-level penetration of renewable energy sources into grid utility: challenges and solutions. *IEEE Access.* 2020;8:190277–99.
4. Faruque M, Hossain M, Alam SM, Negnevitsky M, Rahman M. A novel framework for short-term wind power prediction with RL-based hyper-parameter optimization. *IEEE Trans Consum Electron.* 2025;71(4):10467–78. doi:10.1109/tce.2025.3628745.

5. Liu J, Zang H, Sun K, Sun G, Cheng L, Wei Z. A mode-measurement-cooperated interaction framework based on federated split learning for regional wind power forecasting. *IEEE Trans Smart Grid*. 2026;17(4):3577–90. doi:10.1109/tsg.2026.3650939.
6. Wang Y, Zou R, Liu F, Zhang L, Liu Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl Energy*. 2021;304(1):117766. doi:10.1016/j.apenergy.2021.117766.
7. Chen K, Yu J. Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. *Appl Energy*. 2014;113:690–705. doi:10.1016/j.apenergy.2013.08.025.
8. Zhou J, Shi J, Li G. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers Manag*. 2011;52(4):1990–8. doi:10.1016/j.enconman.2010.11.007.
9. Xu Z, Kong Y, Shen A. Intelligent wind power forecasting for sustainable smart cities. *Appl Sci*. 2025;16(1):305. doi:10.3390/app16010305.
10. Han W, Ziyun Z, Jie Y, Ning Z, Jiawei Z, Yongqian L. A novel federated learning framework with adaptive parameter updater for wind power forecasting. *IEEE Trans Sustain Energy*. 2025;17(3):2016–27. doi:10.1109/tste.2025.3639633.
11. Kandemir E, Prasad DK, Hasan A, Kvamsdal T, Alaliyat SAA. Explainable deep learning with Shapley additive explanations for multivariate wind power forecasting. *Energy Convers Manag X*. 2026;30(3):101901. doi:10.1016/j.ecmx.2026.101901.
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. *Sci Rep*. 2025;15(1):16547. doi:10.1038/s41598-025-00741-9.
13. Zhao Y, Li L, Hu C, Tu Y, Wang Y, Liu Z. A hybrid model for wind power forecasting based on SMA-autoformer-BiLSTM with residual error correction. In: *Proceedings of the 2025 4th International Conference on Power System and Energy Technology (ICPSET)*; 2025 Jul 18–20; Chengdu, China. New York, NY, USA: IEEE; 2025. p. 285–92.
14. Zheng Q, Liu Y, Gu T. Wind power forecasting with a VMD-LSTM-Informer hybrid deep learning model. *Energy AI*. 2026;24:100765. doi:10.1016/j.egyai.2026.100765.
15. Balci M, Dokur E, Yuzgec U. A hybrid LSTM-single candidate optimizer model for short-term wind power prediction. *CMES Comput Model Eng Sci*. 2025;144(1):945–68. doi:10.32604/cmesci.2025.067851.
16. Liang J, Wang Q, Wang L, Zhang Z, Sun Y, Tao H, et al. Wavelet transform convolution and transformer-based learning approach for wind power prediction in extreme scenarios. *CMES Comput Model Eng Sci*. 2025;143(1):945–65. doi:10.32604/cmesci.2025.062315.
17. Hu H, Zheng T, Zhao X, Feng X, Li J, Zhu J. Short-term correlated probabilistic interval forecasting of wind-photovoltaic power based on improved ISODATA similar-day clustering and the IDBO-BiLSTM-copula method. *Electr Power Syst Res*. 2026;257:113015. doi:10.1016/j.epsr.2026.113015.
18. Yaghoubirad M, Azizi N, Farajollahi M, Ahmadi A. Deep learning-based multistep ahead wind speed and power generation forecasting using direct method. *Energy Convers Manag*. 2023;281:116760. doi:10.1016/j.enconman.2023.116760.
19. Huang L, Li L, Wei X, Zhang D. Short-term prediction of wind power based on BiLSTM-CNN-WGAN-GP. *Soft Comput*. 2022;26(20):10607–21. doi:10.1007/s00500-021-06725-x.
20. Faruque MO, Hossain MA, Islam MR, Alam SM, Karmaker AK. Very short-term wind power forecasting for real-time operation using hybrid deep learning model with optimization algorithm. *Clean Energy Syst*. 2024;9(3):100129. doi:10.1016/j.cles.2024.100129.
21. Alguhi AA, Al-Shaalan AM. LSTM-based prediction of solar irradiance and wind speed for renewable energy systems. *Energies*. 2025;18(17):4594. doi:10.3390/en18174594.
22. Xiao P. Short-term wind power prediction based on 1D-CNN with historical information reuse. In: *Proceedings of the 2026 International Russian Smart Industry Conference (SmartIndustryCon)*; 2026 Mar 23–27; Sochi, Russia. New York, NY, USA: IEEE; 2026. p. 619–24.
23. Ren Y, Suganthan PN, Srikanth N. A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods. *IEEE Trans Sustain Energy*. 2014;6(1):236–44. doi:10.1109/tste.2014.2365580.

24. Wang S, Zhang N, Wu L, Wang Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew Energy*. 2016;94:629–36. doi:10.1016/j.renene.2016.03.103.
25. Karijadi I, Chou SY, Dewabharata A. Wind power forecasting based on hybrid CEEMDAN-EWT deep learning method. *Renew Energy*. 2023;218:119357. doi:10.1016/j.renene.2023.119357.
26. Sun L. Adaptive multi-model ultra-short-term wind power forecasting for extreme weather via progressive outlier detection and TimeGAN-CEEMDAN. In: *Proceedings of the 2026 International Russian Smart Industry Conference (SmartIndustryCon)*; 2026 Mar 23–27; Sochi, Russia. New York, NY, USA: IEEE; 2026. p. 804–9.
27. Yang S, Yuan A, Yu Z. A novel model based on CEEMDAN, IWOA, and LSTM for ultra-short-term wind power forecasting. *Environ Sci Pollut Res*. 2023;30(5):11689–705. doi:10.1007/s11356-022-22959-0.
28. Li Z, Xu R, Luo X, Cao X, Sun H. Short-term wind power prediction based on modal reconstruction and CNN-BiLSTM. *Energy Rep*. 2023;9(2):6449–60. doi:10.1016/j.egyr.2023.06.005.
29. Yang J, Peng L, Luo L, Li W, Chen Y. Wind power forecasting using hybrid ICEEMDAN-KPCA and IWOA-BiLSTM models. *Int J Electr Power Energy Syst*. 2025;173:111445. doi:10.1016/j.ijepes.2025.111445.
30. Sun H, Cui Q, Wen J, Kou L, Ke W. Short-term wind power prediction method based on CEEMDAN-GWO-Bi-LSTM. *Energy Rep*. 2024;11(7):1487–502.
31. Yue X, Liu W, Bai Y, Yu Q, Wang R, Song Q. Multi-feature wind speed prediction using bidirectional temporal convolutional networks and attention mechanisms. *J Comput Sci*. 2026;97(1):102885. doi:10.1016/j.jocs.2026.102885.
32. Tian C, Niu T, Wei W. Developing a wind power forecasting system based on deep learning with attention mechanism. *Energy*. 2022;257:124750. doi:10.1016/j.energy.2022.124750.
33. Xiang L, Liu J, Yang X, Hu A, Su H. Ultra-short term wind power prediction applying a novel model named SATCN-LSTM. *Energy Convers Manag*. 2022;252(8):115036. doi:10.1016/j.enconman.2021.115036.
34. Yang T, Yang Z, Li F, Wang H. A short-term wind power forecasting method based on multivariate signal decomposition and variable selection. *Appl Energy*. 2024;360(15):122759. doi:10.1016/j.apenergy.2024.122759.
35. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing*. 2021;452:48–62.