AIMS Energy, 14(2): 335–357.
DOI: 10.3934/energy.2026015
Received: 18 October 2025
Revised: 20 January 2026
Accepted: 05 February 2026
Published: 16 March 2026

https://www.aimspress.com/journal/energy

Research article

Secondary decomposition-based EVMD-DWT-Transformer for wind speed prediction

Thi Hoai Thu Nguyen*, Xuan Bach Do, Trung Tuan Anh Nguyen and Phong Ky Pham

PGRE. Lab., School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam

* **Correspondence:** Email: thu.nguyenthihoai@hust.edu.vn; Tel: +84983533012.

Abstract: Accurate short-term wind speed forecasting is crucial for efficient grid management and the integration of renewable energy sources. This study introduces a hybrid forecasting model that combines enhanced variational mode decomposition (EVMD), discrete wavelet transform (DWT), and an encoder-only Transformer. The dataset is first preprocessed to handle missing values and outliers and then subjected to a two-step decomposition process: EVMD extracts the first nine intrinsic mode functions (IMFs), and DWT is applied to the last IMF from EVMD to further decompose high-frequency components and capture finer-scale variations. The encoder-only Transformer is then used to process and forecast all subseries obtained from both EVMD and DWT, excluding the final IMF from EVMD, and their outputs are aggregated to produce the final prediction. The model was evaluated using data from the Thai Hoa wind farm in Binh Thuan, Vietnam, and compared with six benchmarks (variational mode decomposition (VMD)-Transformer, VMD-long short-term memory (LSTM), Transformer, LSTM, and random forest (RF)) using mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and normalized RMSE (NRMSE) for both single and multi-step forecasting. Results show that the proposed hybrid model consistently outperforms all benchmarks across all evaluation metrics, achieving an RMSE of 0.535 m/s for 1-step and 0.834 m/s for 24-step forecasting, representing up to 1.9% improvement over the best baseline and over 63% improvement compared to the worst-performing model, thereby demonstrating superior accuracy across both forecasting horizons. By leveraging multi-level decomposition and attention mechanisms, the model effectively captures complex wind speed patterns, enhancing forecasting performance and supporting better integration of wind energy into power systems.

Keywords: enhanced variational mode decomposition (EVMD); discrete wavelet transform (DWT); secondary decomposition; encoder-only transformer; short-term wind speed forecasting

1. Introduction

Renewable energy sources, such as wind and solar power, are increasingly integral to modern power grids as nations strive to transition toward more sustainable and low-carbon energy systems. However, the integration of these energy sources presents substantial challenges due to their inherent variability and dependence on meteorological conditions. Wind speeds fluctuate, and solar radiation is influenced by cloud cover, resulting in intermittent and less predictable power generation [1]. These fluctuations can lead to grid instability, heightened reliance on backup power sources, and inefficiencies in energy dispatching [2]. To mitigate these challenges, precise forecasting techniques have become critical for effective grid management, facilitating more efficient scheduling of power generation, balancing of energy demand, and optimization of energy storage systems.

In the forecasting domain, forecasting methods are typically classified into four primary categories: (1) physical methods, (2) statistical methods, (3) artificial intelligence (AI) methods, and (4) hybrid methods. Among these, physical methods predominantly rely on numerical weather prediction (NWP) models or sky imagery for forecasting purposes. Notable models widely employed in this context include the weather research and forecasting (WRF) model [3] and the high-resolution limited area model (HIRLAM) [4].

Statistical methods, on the other hand, involve the application of various statistical approaches, such as the Box-Jenkins models (e.g., autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA)) and exponential smoothing. Liu et al. [5] conducted a comparative study of the SARIMA model against LSTM and gated recurrent units (GRU) for offshore wind speed forecasting. Utilizing a range of evaluation metrics, including R-squared (R^2), explained variance score (EVS), RMSE, MSE, MAE, and median absolute error (MedAE), the SARIMA model demonstrated superior performance relative to LSTM and GRU. This comparison highlights SARIMA's efficacy in handling seasonal data, such as wind speed, further reinforcing its suitability for such forecasting tasks.

However, the performance of physical methods and traditional statistical methods is inadequate for short-term forecasting due to the nonlinear nature of historical wind speed data. With the advancements of computer hardware, AI methods have garnered considerable attention in recent years. Initially, traditional machine learning techniques, such as multilayer perceptron (MLP) and support vector machines (SVM), were predominant in this field. For example, Liu et al. [6] proposed a Jaya-based SVM for short-term wind speed forecasting and achieved great accuracy when compared against various models. However, the emergence of deep learning, a subset of machine learning built upon artificial neural networks (ANN), has been a significant factor in the rapid proliferation of AI applications. Models such as convolutional neural networks (CNN) [7], LSTM networks [8,9], GRU [10], graph convolutional network (GCN) [11], and the more recent Transformer model have become integral to wind forecasting. For instance, Joseph et al. [12] introduced a novel bidirectional LSTM (BiLSTM) model for near real-time wind speed forecasting. When compared to standalone and hybrid models, including LSTM, recurrent neural networks (RNN), MLP, and RF, the proposed BiLSTM model demonstrated superior prediction performance, achieving the lowest relative RMSE,

ranging from 9.6% to 23.8%, and the lowest mean absolute percentage error (MAPE), ranging from 8.8% to 21.5%.

Hybrid methods focus on combining multiple forecasting techniques to enhance predictive performance. A prevalent approach within this category involves integrating a decomposition method with either a statistical model or a machine learning algorithm. Decomposition methods aim to partition data into smaller subseries, thereby improving forecasting accuracy, especially considering the sensitivity of certain machine learning and statistical models. Common decomposition techniques include singular spectrum analysis (SSA) [13], empirical mode decomposition (EMD) [14] and its variants, such as ensemble empirical mode decomposition (EEMD) [15] and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [16,17], as well as VMD [18] and DWT [19].

Li et al. [20] proposed a hybrid model that combines EMD with a causal convolutional network and a Transformer model. In this approach, EMD was used for data decomposition, while the CCTransformer—a Transformer model leveraging convolutional attention—was applied for ultra-short-term forecasting. The results demonstrated reductions in MAE, MSE, and RMSE compared to other models, thereby highlighting its effectiveness. Similarly, Gao et al. [21] developed a model that incorporated singular spectrum analysis-variational mode decomposition (SSA-VMD) for data decomposition and LSTM networks for forecasting. By optimizing VMD using SSA and employing LSTM for prediction, the proposed model achieved an RMSE of 2.03, significantly lower than that of the VMD-LSTM model, which yielded an RMSE of 3.8523.

Another hybrid modeling approach involves the integration of multiple decomposition techniques. Karijadi et al. [22] developed a model that combined CEEMDAN, empirical wavelet transform (EWT), and LSTM networks for wind power forecasting in France and Turkey. When compared to several single and hybrid forecasting models, their proposed approach outperformed all others. Similarly, Jin et al. [18] introduced a hybrid model that combined VMD, CEEMDAN, and the Informer model for ultra-short-term wind speed forecasting. The modes and IMFs derived from the decomposition processes were used as inputs for the Informer model. Their results highlighted the model's robust capability to capture long-term dependencies and deliver high forecasting accuracy.

While recent studies have demonstrated the effectiveness of integrating multiple advanced decomposition methods, such as combining CEEMDAN with EWT or VMD with CEEMDAN, these approaches often rely on complex or computationally intensive techniques for the second decomposition layer. To address this problem, we propose a different approach by combining an improved variational mode decomposition (EVMD) with DWT. DWT is a less computationally expensive method, which makes it particularly effective as a second decomposition stage for further refining signal components with minimal additional computational burden. This computational efficiency is particularly advantageous for large-scale or data-intensive forecasting tasks, where reducing model complexity is important. By combining this dual-layer decomposition with a Transformer model, we aim to enhance forecasting accuracy while maintaining a relatively lightweight model structure. The proposed model will be evaluated against single-layer decomposition methods, including VMD-LSTM and VMD-Transformer, as well as other standalone models, under both single-step and multi-step forecasting scenarios. The contributions of this study are as follows: (i) the development of a secondary decomposition algorithm utilizing EVMD and DWT; (ii) the application of an encoder-only Transformer model; and (iii) a comparative analysis of the forecasting results against VMD-LSTM, VMD-Transformer, and other standalone models, evaluated under both single-step and multi-step forecasting scenarios using performance metrics such as MAE, MSE, RMSE, and NRMSE.

2. Methodology

2.1. Outlier detection with interquartile range (IQR)

Before being decomposed, the dataset is processed for outliers using the IQR. The IQR is defined by the difference between the third quartile (Q_3) and the first quartile (Q_1). From there, the lower bound and the upper bound can be evaluated by:

$$LB = \mathrm{Q}_1 - 1.5 * IQR \tag{1}$$

$$UB = \mathrm{Q}_3 + 1.5 * IQR \tag{2}$$

Any data points higher or lower than the lower bound and upper bound are considered outliers. The replacement for these outliers is described as follows:

$$\begin{cases} Outlier = LB\, if\ Outlier < LB \\ Outlier = UB\, if\ Outlier > LB \end{cases} \tag{3}$$

2.2. Traditional variational mode decomposition (VMD)

Variational mode decomposition (VMD), introduced by Dragomiretskiy and Zosso in 2014 [23], is an adaptive, fully non-recursive signal decomposition technique. It decomposes a signal in the frequency domain into a series of IMFs while aiming to minimize the total bandwidth under the constraint that the sum of all mode components equals the original signal. The mathematical formulation of VMD is as follows:

$$\min_{u_k,\omega_k}\left\{\sum_{k=1}^{K}\partial(t)\left[\delta(t)+\frac{j}{\pi t}*u_k(t)\right]e^{-j\omega kt}{}_2^2\right\} \tag{4}$$

$$s.t.\sum_{k=1}^{N}u_k = f(t) \tag{5}$$

where ∂(t) represents the partial derivative with respect to time t, δ(t) is the Dirac function, ω_k is the center frequency of the k-th mode, $u_k(t),(\mathrm{k}=1,2,...,\mathrm{K})$ represents each decomposed mode component, j is the imaginary unit, and * is the convolution operator.

To determine the optimal decomposition, the constrained variational problem is transformed into an unconstrained problem using the Lagrangian multiplier λ(t) and a quadratic penalty factor α, formulated as follows:

$$L(u_k,\omega_k,\lambda) = \alpha\sum_{k=1}^{K}\left\|\partial(t)\left[\delta(t)+\frac{j}{\pi t}*u_k(t)\right]e^{-j\omega kt}\right\|_2^2 + \left\|f(t)-\sum_{k=1}^{K}u_k(t)\right\|_2^2 + \left\langle\lambda(t), f(t)-\sum_{k=1}^{K}u_k(t)\right\rangle \tag{6}$$

The alternating direction method of multipliers (ADMM) is used to find the "saddle point" minimum of the expanded Lagrangian expression. The best expression is IMF u_k and the central frequency ω_k. When the following constraint is met, the iteration ends:

$$\sum_{k=1}^{K} \frac{\left\| u_k^{n+1} - u_k^{n} \right\|_2^2}{\left\| u_k^{n} \right\|_2^2} < \varepsilon \tag{7}$$

Finally, based on the frequency domain characteristics of v, the frequency bands are divided, achieving adaptive decomposition of the signal.

$$u_k^{n+1}(\omega) = \frac{\hat{x}(\omega) - \sum_{i<k} u_i^{n}(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \tag{8}$$

2.3. Enhanced variational mode decomposition (EVMD)

A key limitation of traditional VMD is that the number of decomposition modes k must be predefined, which can significantly impact decomposition performance. To address this issue, Jiang et al. [24] proposed an enhanced VMD algorithm, termed EVMD, which incorporates the Hilbert Transform and empirical mode decomposition (EMD) to determine the optimal number of modes adaptively. The method can be described as follows:

Given a signal $x(t)$ decomposed into k subseries $s_j(t), j = 1,2,3,...,k$, each subseries can be represented in the following form using HT

$$s_j(t) = HT\left[s_j(t)\right] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{s_j(\tau)}{t-\tau} d\tau \tag{9}$$

From this representation, the instantaneous phase $\theta_j(t)$ and instantaneous frequency $\omega_j(t)$ can be derived as follows:

$$\theta_j(t) = \tan^{-1}\left(s_j(t) \Big/ s_j(t) \right) \tag{10}$$

$$\omega_j(t) = d\theta_j(t)/dt \tag{11}$$

The mean frequency of the signal, denoted as $\overline{\omega_j}(t)$, is then defined as:

$$\overline{\omega_j}(t) = \frac{1}{n}\sum_{j=1}^{n} \omega_j(t), j = 1,2,\ldots k \tag{12}$$

Subsequently, the variation coefficient (CV_k) of all $\left\{\overline{\omega_j}(t)\right\}$ values is computed as:

$$CV_k = \frac{\sigma_k}{u_k}; u_k = \frac{1}{k}\sum_{j=1}^{k}\overline{\omega_j}(t); \sigma_k = \frac{1}{k}\sum_{j=1}^{k}\left(\overline{\omega_j}(t) - u_k\right)^2 \tag{13}$$

where CV_k is used to measure the decomposition performance under different k. Generally, a larger CV_k means a clearer difference among different subseries, and the corresponding parameter k can be considered to be the optimum number of decomposition levels.

2.4. Discrete wavelet transform (DWT)

The discrete wavelet transform (DWT) [25,26] decomposes a signal through a sequence of filters, specifically high-pass (HP) and low-pass (LP) filters. The mathematical representation of these filters is as follows:

$$x_1(n) = \sum_{k=0}^{L-1} c_k x(n-k) \tag{14}$$

$$x_2(n) = \sum_{k=0}^{L-1} d_k x(n-k) \tag{15}$$

where $x_1(n)$ and $x_2(n)$ denote the low-pass and high-pass components, respectively, and c_k and d_k represent the coefficients of the LP and HP filters.

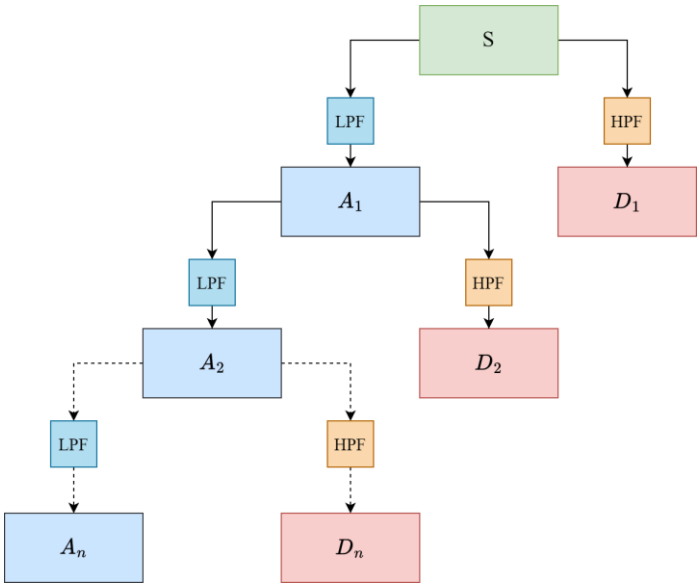

Figure 1. Algorithm of discrete wavelet transform [26].

Figure 1 illustrates the DWT decomposition process, where an input signal S is passed through LPF and HPF filters to obtain its low-frequency and high-frequency components. The low-frequency components retain the broader, smoother aspects of the signal, whereas the high-frequency components capture rapid variations and transient changes. The signal S can be expressed as the sum of its sub-signals:

$$S = D_1 + D_2 + D_3 + \ldots + D_n + A_n \tag{16}$$

2.5. Transformer network architecture

In this study, an encoder-only Transformer architecture is employed for wind speed forecasting. The Transformer model, originally proposed by Vaswani et al. [27], departs from traditional encoder-decoder architectures that rely on convolutional neural networks (CNNs) or recurrent neural networks (RNNs), instead utilizing self-attention mechanisms exclusively. The structure of the encoder-only Transformer is depicted in Figure 2.

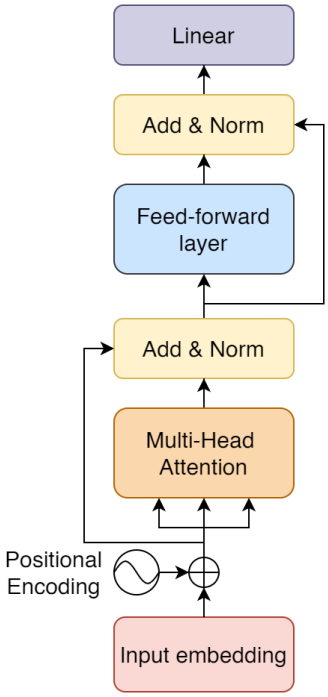

Figure 2. The structure of the encoder-only Transformer.

A key component of the Transformer model is the multi-head self-attention mechanism, which operates as follows:

Step 1: Computation of query, key, and value matrices

The query matrix Q, the key matrix K, and the value matrix V are obtained by three linear transformations. This can be realized by matrix multiplication as follows:

$$Q_i = QW_i^Q, K_i = KW_i^K, V_i = VW_i^V, i = 1, \ldots, n \tag{17}$$

where W_i^Q, W_i^K, and W_i^V is the weight matrix obtained by learning.

Step 2: Multi-head attention mechanism

The query, key, and value matrices are divided into multiple headers as follows:

$$split(Q,K,V)=[head_1,\ldots,head_n] \tag{18}$$

where $head_i$ denotes the first i header of the query, key, and value matrix.

Step 3: Scaled dot-product attention computation

For each attention head, attention scores are computed as follows:

$$Attention_i(Q,K,V)=softmax\left(\frac{Q_i.K_i^T}{\sqrt{d_k}}\right)V_i, i=1,\ldots,n \tag{19}$$

where d_k is the dimensionality of the key vectors, ensuring stable gradients during training.

For this calculation, the dot products of the query matrix and the key matrix, referred to as the similarity score, are divided by the squared root of the key vector dimension. Softmaft is applied directly to normalize the score, resulting in the score matrix. Finally, the score matrix is multiplied by the value matrix to compute the attention matrix.

Step 4: Output projection

After computing an n amount of attention matrices, where n is the amount of headers, these matrices are then concatenated through a concat operation, described as follows:

$$MultiHead(Q,K,V)=Concat(head_1,\ldots,head_n)W_o \tag{20}$$

where W_o is the output projection matrix. $Concat(head_1,...,head_n)$ indicates that the output signals of multiple heads are spliced together.

After this, a linear layer is applied to obtain the output of the multi-head attention mechanism.

2.6. The proposed model

This study introduces a hybrid forecasting framework integrating EVMD, DWT, and Transformer architectures for short-term wind speed prediction. The overall structure of the proposed model is illustrated in Figure 3.

Initially, the dataset is preprocessed and decomposed using EVMD, which decomposes the input signal into nine IMFs. Among these, the first eight IMFs are directly fed into the Transformer model for forecasting. In the second stage of decomposition, the ninth IMF undergoes further processing using DWT, which decomposes it into four additional subseries. This additional decomposition is intended to capture finer-scale fluctuations within the original signal, ensuring a more detailed representation of wind speed variations. After decomposition, each IMF is divided into train/val/test data at a proportion of 70/10/20 and then normalized using MinMaxScaler. These newly obtained subseries are then utilized as additional inputs to the Transformer model, alongside the previously obtained IMFs. Once all subseries have been processed, the Transformer model generates predictions

for each component. The final wind speed forecast is derived by aggregating the prediction results from both decomposition stages. By combining multi-level decomposition with an attention-based deep learning model, the proposed approach effectively captures intricate patterns in wind speed data, leading to improved forecasting performance.

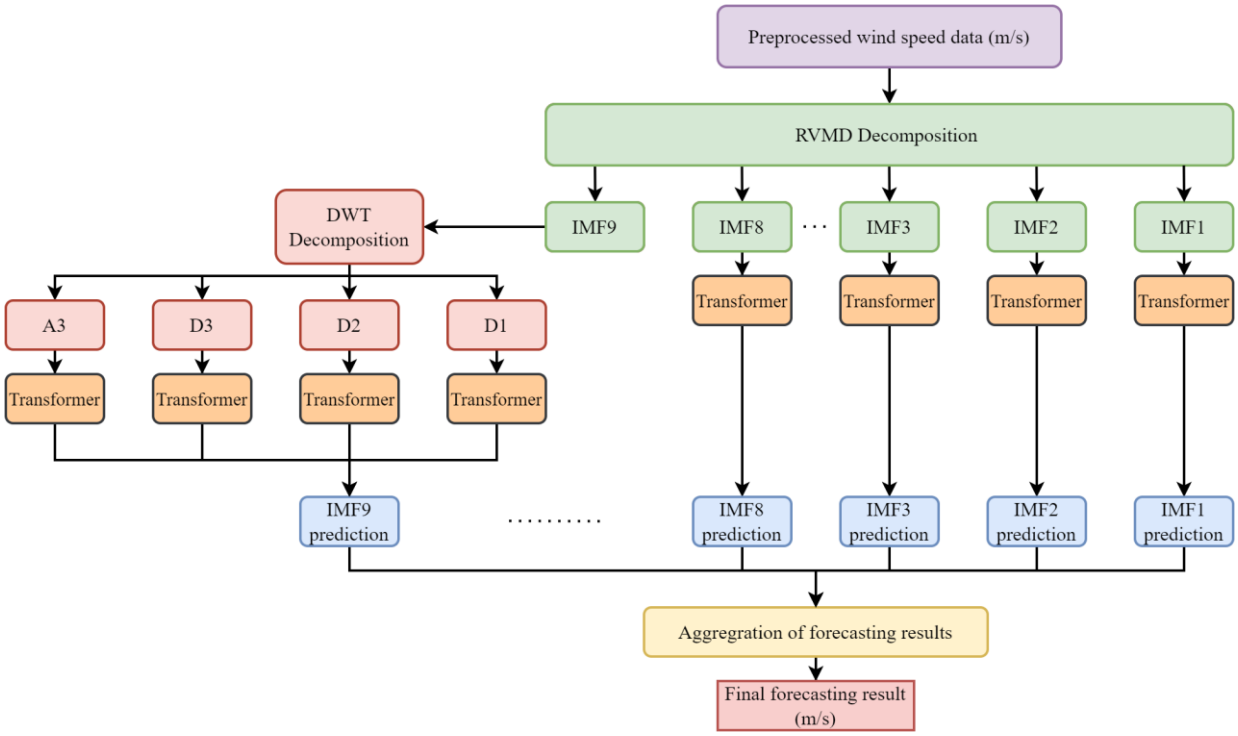

Figure 3. The structure of the proposed model.

2.7. Evaluation metrics

In this study, MAE, MSE, RMSE, and NRMSE are used to evaluate the forecasting performance of the proposed model. The equations below describe the formulas for the metrics:

$$MSE = \frac{\sum_{i=1}^{N}(A_i - F_i)^2}{N} \tag{21}$$

$$MAE = \frac{\sum_{i=1}^{N}|A_i - F_i|}{N} \tag{22}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(A_i - F_i)^2}{N}} \tag{23}$$

$$NRMSE = \frac{RMSE}{A_{max} - A_{min}} \tag{24}$$

where N is the number of sample data, A_i is the actual data, $\bar{A}_i$ is the mean of the actual data, and F_i is the forecasted data.

3. Data properties and processing

3.1. Data properties

This study employs a wind speed dataset collected from the Thai Hoa Wind Farm, situated in Binh Thuan Province, Vietnam (located approximately at 11.2637°N, 108.3599°E). The dataset spans a full year, from January 1, 2022, to March 31, 2024, with measurements recorded at an hourly interval. As a result, the dataset comprises a total of 19,704 data points. Figure 4 provides a visual representation of the dataset for better clarity and understanding of its characteristics, while Table 1 presents the statistical properties of the dataset.

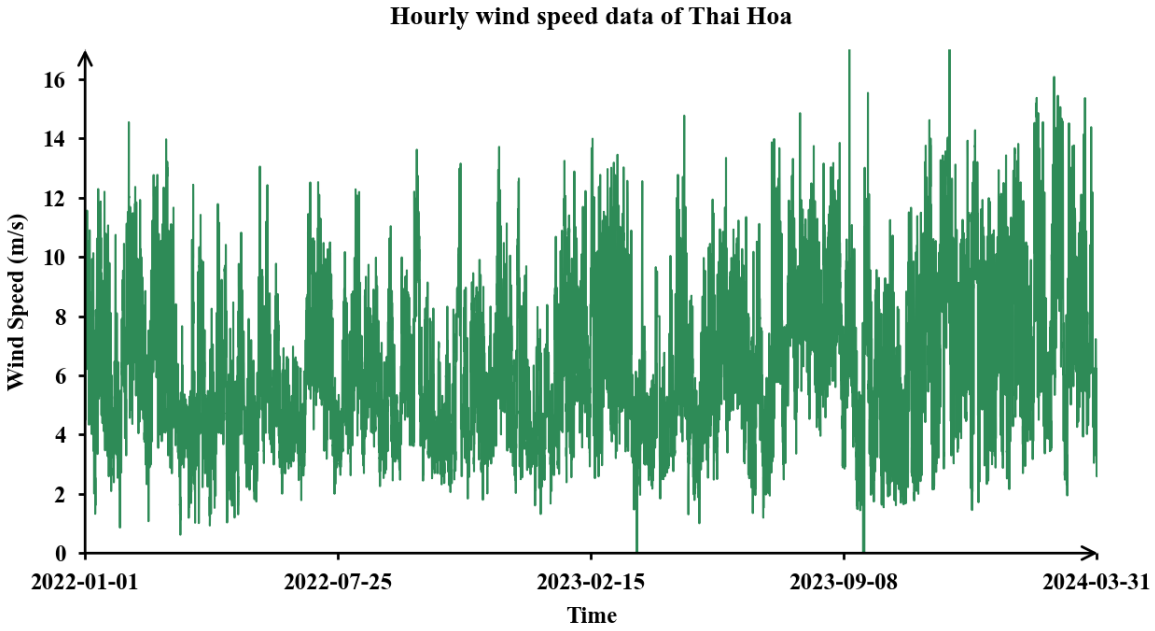

Figure 4. Wind speed data obtained from Thai Hoa wind farm.

Table 1. Statistical properties of the dataset.

Statistical properties	Value
Number of data points	19,704
Mean (m/s)	6.36
Standard deviation (m/s)	2.57
Minimum (m/s)	0.00
First quartile (m/s)	4.30
Median (m/s)	6.03
Second quartile (m/s)	8.09
Maximum (m/s)	19.13

Prior to applying decomposition techniques, the dataset undergoes preprocessing to address any missing values and ensure data quality. Specifically, the IQR method is utilized to detect and handle potential outliers, thereby enhancing the reliability of the dataset for subsequent analysis.

3.2. Data decomposition and processing

As described in the methodology, the preprocessed dataset is decomposed using a combination of EVMD and DWT to extract meaningful subseries for forecasting. Figures 5 and 6 illustrate the resulting decomposed components obtained from EVMD and DWT, respectively. Table 2 describes the parameters used for each decomposition method.

In the first stage, the EVMD algorithm decomposes the original time series into nine IMFs ($K = 9$), each corresponding to oscillatory components at distinct frequency bands under the imposed bandwidth constraint ($\alpha = 3000$). This configuration enables stable mode separation while avoiding excessive frequency overlap.

Among the extracted IMFs, the highest-frequency mode—which primarily captures rapid fluctuations and localized variations—is selected for secondary decomposition. In the second stage, this IMF is further decomposed using a discrete wavelet transform with a Daubechies-4 (db4) wavelet at three decomposition levels, producing four wavelet-based subseries that represent progressively finer temporal resolutions.

Through this two-stage decomposition process, a total of 12 subseries are obtained (eight EVMD-derived IMFs retained directly and four DWT-derived components). These decomposed components are subsequently used as parallel input features for the Transformer-based forecasting model.

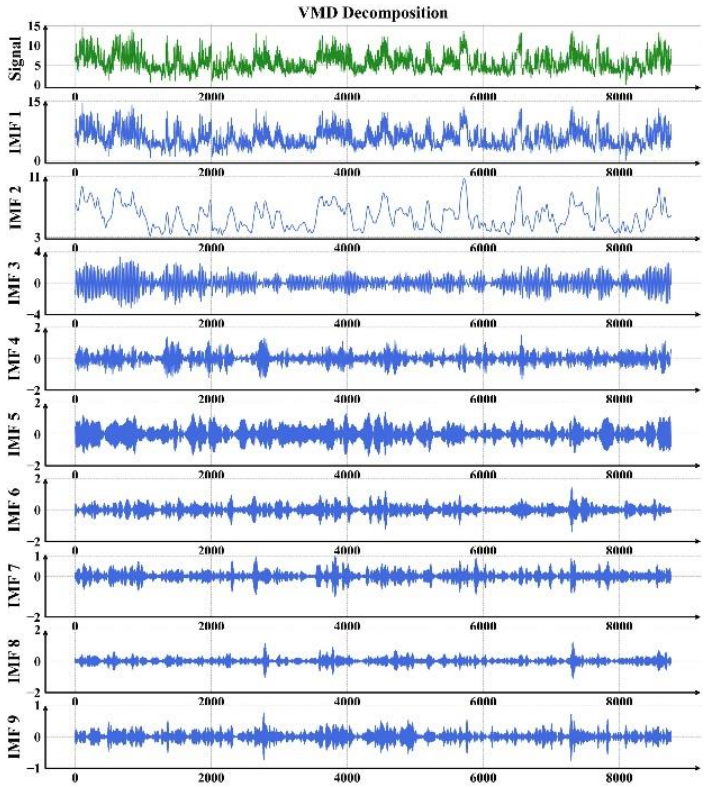

Figure 5. Decomposition results from EVMD.

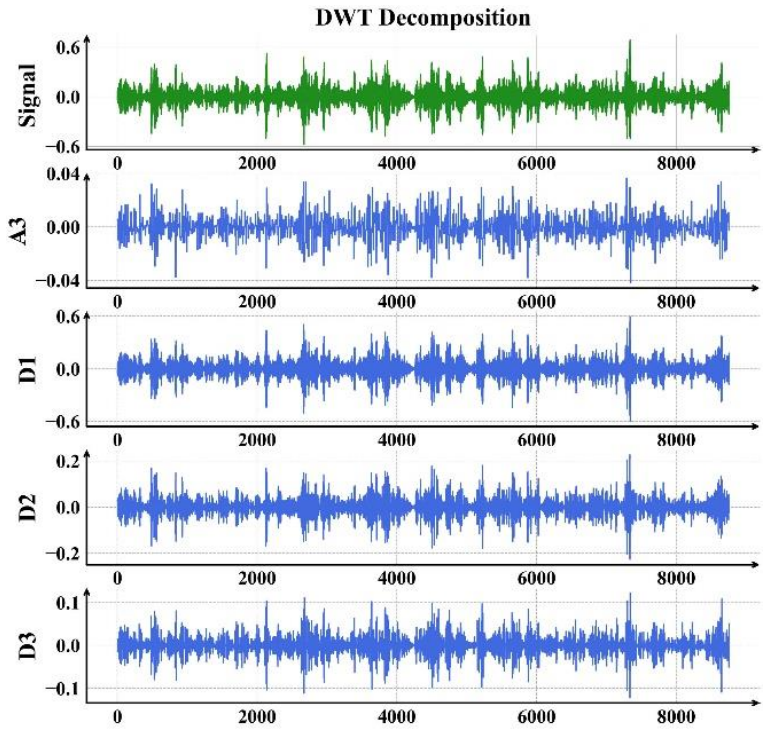

Figure 6. Decomposition results from DWT.

Table 2. Parameters for the EVMD and DWT.

Module	Parameters	Value	Description
EVMD	Number of modes (K)	9	Number of intrinsic mode functions extracted
	Penalty factor (α)	3000	Control bandwidth constraint of each mode
	Noise tolerance (τ)	0	Noise-tolerance parameter
	Initialization method	Uniform	Initialization strategy for center frequencies
	Convergence tolerance	1×10^{-6}	Stopping criterion for decomposition
DWT	Wavelet basis	db4	Mother wavelet used for decomposition
	Decomposition level	3	Number of multi-resolution levels

After the decomposition stage, each IMF is treated as an independent univariate time series and transformed into a supervised learning dataset using a sliding window strategy. Specifically, a rolling window with stride one is applied along the time axis. For a given time index t, the input sequence consists of the previous L observations $(x_{t-L}, \ldots, x_{t-1})$, where L denotes the lookback window length. The corresponding target sequence comprises the next H values $(x_t, \ldots, x_{t+H-1})$, where H is the forecasting horizon.

This window is then shifted forward by one time step to generate the next training sample, ensuring maximal utilization of the available data while preserving temporal order and causality. The process is repeated until the end of the time series is reached. All supervised samples are constructed first, after which the resulting dataset is split sequentially into training, validation, and testing sets according to predefined proportions (70%, 10%, and 20%, respectively). To better illustrate this process, Figure 7 illustrates the 3 rolling window iterations in a 2-step forecasting process with a lookback of 4 datapoints.

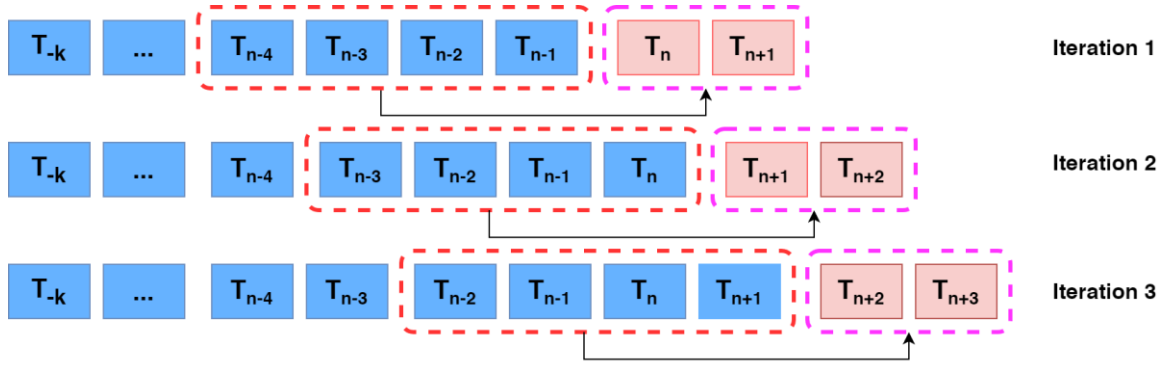

Figure 7. Illustration of the rolling window data construction procedure for multi-step forecasting.

4. Results and discussion

In this study, the EVMD-DWT-Transformer is compared with six other forecasting models using four evaluation metrics: MSE, MAE, RMSE, and NRMSE. The comparison is made for single-step forecasting and multi-step forecasting, each with two subsections: global forecasting performance and daily forecasting performance. The daily forecasting performance assesses model accuracy on a per-day basis by calculating error values from each set of 24 data points, corresponding to one day of hourly data. In contrast, the global forecasting performance evaluates overall accuracy by computing a single error value (per metric) based on the entire prediction set for each model. Since each model produces a sequence of error values (160 error points for 160 days in the 1-step forecasting scenario and 159 error points for 159 days in the 24-step forecasting scenario), the daily forecasting performance is evaluated using RMSE to summarize performance over time. On the other hand, the global forecasting performance is assessed using multiple metrics, including MSE, MAE, RMSE, and NRMSE, to provide a comprehensive evaluation across the entire dataset.

4.1. Hyperparameters of the proposed model

In this section, the hyperparameters of the proposed model are presented. A total of 12 Transformer models were trained, including eight models for predicting the eight modes produced by VMD and four models for predicting the modes derived from the DWT decomposition of the first VMD mode. For the Transformer model, input_dim denotes the input feature dimension, d_model represents the hidden model size, nhead indicates the number of attention heads, num_layers refers to the number of encoder layers, dropout is the regularization parameter, and epochs specifies the number of training epochs. Tables 3 and 4 provide a summary of the model hyperparameters corresponding to all modes. It should be noted that, in this study, all models were optimized empirically using prior experience and a trial-and-error approach to achieve the best possible results. The hyperparameter range of optimization is presented in Table 5.

Table 3. Hyperparameters for the Transformer models trained on VMD modes.

	VMD							
	Mode 1	**Mode 2**	**Mode 3**	**Mode 4**	**Mode 5**	**Mode 6**	**Mode 7**	**Mode 8**
input_dim	1	1	1	1	1	1	1	1
d_model	32	32	64	64	64	64	64	64
nhead	2	2	4	2	2	4	4	4
num_layers	2	3	2	3	3	1	1	1
dropout	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
batch_size	64	64	64	32	64	32	32	32
epochs	100	100	100	100	100	100	100	100

Table 4. Hyperparameters for the Transformer models trained on DWT modes.

	DWT			
	A3	**D3**	**D2**	**D1**
input_dim	1	1	1	1
d_model	64	64	64	64
nhead	4	4	2	4
num_layers	2	3	3	3
dropout	0.2	0.2	0.2	0.2
batch_size	64	32	32	64
epochs	100	100	100	100

Table 5. Hyperparameter optimization range.

Parameters	**Value**	**Description**
d_model	32,64,128,256,512	Model dimension
nhead	4,8	Number of attention heads
num_layers	2,3,4	Number of encoder layers
batch_size	32,64,128	Number of training samples processed in one backward/forward pass
dropout	0.05, 0.1, 0.15, 0.2	Regularization parameter

4.2. Global forecasting performance

This section evaluates the performance of six forecasting models, namely the proposed EVMD-DWT-Transformer, VMD-Transformer, VMD-LSTM, Transformer, LSTM, and RF, using four error metrics: MSE, MAE, RMSE, and NRMSE. Results are reported for 1, 5, 10, and 24-step forecasting on the entire test set. Tables 6–9 summarize the performance for the 1, 5, 10, and 24-step forecasts, respectively, while Figure 8 provides visualizations for both 1-step and 24-step forecasting results over a randomly selected 260-h segment to enhance interpretability.

Table 6. 1-step global forecasting performance of 6 models.

Metrics	Forecasting methods					
	Proposed	VMD-Transformer	VMD-LSTM	Transformer	LSTM	RF
MSE (m^2/s^2)	0.286	0.289	0.303	0.595	1.076	1.331
MAE (m/s)	0.412	0.414	0.425	0.574	0.775	0.828
RMSE (m/s)	0.535	0.538	0.551	0.771	1.037	1.154
NRMSE	0.034	0.034	0.035	0.049	0.066	0.074

Table 7. 5-step global forecasting performance of 6 models.

Metrics	Forecasting methods					
	Proposed	VMD-Transformer	VMD-LSTM	Transformer	LSTM	RF
MSE (m^2/s^2)	0.349	0.359	0.354	0.862	3.132	3.135
MAE (m/s)	0.454	0.460	0.463	0.695	1.351	1.357
RMSE (m/s)	0.591	0.599	0.595	0.928	1.770	1.771
NRMSE	0.038	0.038	0.038	0.059	0.113	0.113

Table 8. 10-step global forecasting performance of 6 models.

Metrics	Forecasting methods					
	Proposed	VMD-Transformer	VMD-LSTM	Transformer	LSTM	RF
MSE (m^2/s^2)	0.385	0.396	0.474	1.028	4.097	4.111
MAE (m/s)	0.476	0.483	0.536	0.744	1.574	1.575
RMSE (m/s)	0.621	0.630	0.688	1.014	2.024	2.028
NRMSE	0.040	0.040	0.044	0.065	0.130	0.130

Table 9. 24-step global forecasting performance of 6 models.

Metrics	Forecasting methods					
	Proposed	VMD-Transformer	VMD-LSTM	Transformer	LSTM	RF
MSE (m^2/s^2)	0.695	0.722	0.959	1.780	4.615	5.361
MAE (m/s)	0.615	0.628	0.732	0.995	1.684	1.802
RMSE (m/s)	0.834	0.850	0.979	1.334	2.148	2.315
NRMSE	0.053	0.054	0.063	0.085	0.138	0.148

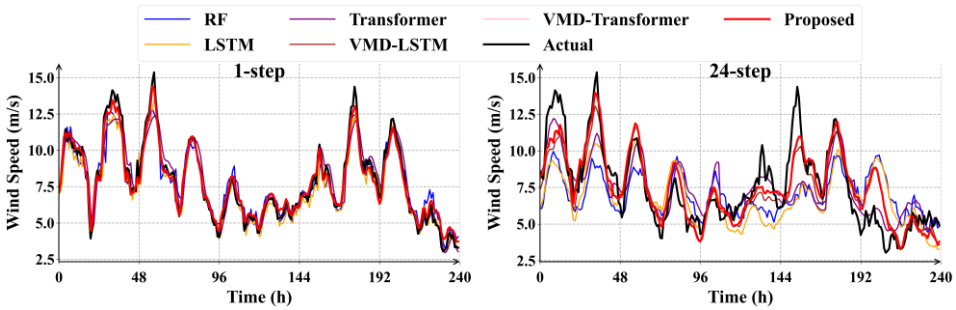

Figure 8. 1-step and 24-step forecasting results of 6 models.

Across all forecasting horizons, the proposed EVMD-DWT-Transformer model consistently outperforms all baseline methods across every evaluation metric. For the 1-step forecasting task, it achieves an MSE of 0.286 m^2/s^2, an MAE of 0.412 m/s, an RMSE of 0.535 m/s, and an NRMSE of 0.034, demonstrating strong accuracy in capturing short-term wind speed variations. As the forecasting horizon increases, the model maintains high stability and robustness. At the 10-step horizon, the proposed model records an MSE of 0.385 m^2/s^2, an MAE of 0.476 m/s, an RMSE of 0.621 m/s, and an NRMSE of 0.040, indicating only a moderate degradation in performance. For the more challenging 24-step forecasting task, it continues to achieve superior results, with an MSE of 0.695 m^2/s^2, an MAE of 0.615 m/s, an RMSE of 0.834 m/s, and an NRMSE of 0.053. These results highlight the model's strong generalization capability and resilience under increasing temporal uncertainty.

The visual analysis further supports these quantitative findings. In the 1-step forecasting scenario, most models are able to follow the general wind speed trends during periods of moderate and low variability. However, RF exhibits noticeable deviations, frequently overshooting and generating exaggerated peaks, particularly around the 50-h and 150-h marks. A similar behavior is observed in the 24-step forecasting task, where RF significantly underestimates the magnitude of the initial wind speed peak between 0 and 52 h. While the proposed model also shows minor underestimation during sharp peaks between 104 and 156 h, it remains the most accurate model overall, capturing both trend and magnitude more effectively than competing approaches.

Hybrid models incorporating VMD demonstrate a clear advantage over their standalone counterparts across all forecasting horizons. In the 10-step task, the VMD-LSTM model achieves an MSE of 0.474 m^2/s^2 and an RMSE of 0.688 m/s, representing a substantial improvement over the standalone LSTM, which records an MSE of 4.097 m^2/s^2 and an RMSE of 2.024 m/s. This performance gap widens further in long-term forecasting. At the 24-step horizon, VMD-LSTM achieves an MSE of 0.959 m^2/s^2, significantly outperforming both the standalone LSTM (4.615 m^2/s^2) and the standalone Transformer (1.780 m^2/s^2). These results confirm that signal decomposition via VMD plays a critical role in enhancing deep learning model performance, particularly under highly volatile wind conditions.

Among hybrid architectures, Transformer-based models exhibit greater effectiveness than LSTM-based models in leveraging decomposed signal components. Across all forecasting horizons, the VMD-Transformer consistently outperforms the VMD-LSTM. For example, at the 10-step horizon, the VMD-Transformer achieves an MSE of 0.396 m^2/s^2 and an RMSE of 0.630 m/s, outperforming the VMD-LSTM, which records an MSE of 0.474 m^2/s^2 and an RMSE of 0.688 m/s. This advantage becomes more pronounced in the 24-step task, where the VMD-Transformer achieves an MSE of 0.722 m^2/s^2, an MAE of 0.628 m/s, an RMSE of 0.850 m/s, and an NRMSE of 0.054. The proposed EVMD-DWT-Transformer further improves upon these results, delivering marginal yet consistent gains in accuracy and generalization, demonstrating the complementary benefits of integrating DWT with VMD-based decomposition.

In contrast, RF consistently ranks last or second to last across nearly all evaluation metrics. At the 10-step horizon, RF records an MSE of 4.111 m^2/s^2 and an RMSE of 2.028 m/s, performing similarly to the standalone LSTM but substantially worse than hybrid models. Its performance deteriorates further in long-term forecasting, reaching an MSE of 5.361 m^2/s^2 and an RMSE of 2.315 m/s at the 24-step horizon. Interestingly, the standalone LSTM slightly outperforms RF in the 24-step scenario, reversing their relative ranking from short-term forecasting. This shift highlights the LSTM's ability to capture long-term temporal dependencies, whereas RF's lack of temporal modeling capability leads to rapidly escalating errors as the forecasting horizon increases.

Regarding the effective forecasting horizon, the experimental results indicate a distinct performance threshold. The proposed model maintains high stability with minimal error accumulation up to the 10-step horizon, where the RMSE increases by only 0.086 m/s compared to the 1-step baseline (from 0.535 to 0.621 m/s). However, error accumulation accelerates beyond this point, with the RMSE rising to 0.834 m/s at the 24-step horizon. Therefore, while the model remains superior to baselines for day-ahead (24-step) forecasting, a maximum horizon of 10–12 steps is recommended for applications requiring high-precision control, whereas the 24-step horizon remains suitable for general scheduling and planning tasks.

4.3. Daily forecasting performance

In this section, the daily prediction error of each model is evaluated using the distribution of absolute errors. Since the data is at hourly resolution, daily errors are calculated by combining 24 consecutive predictions, resulting in 160 days of analysis per model. Figures 9 and 10 present histograms of the absolute error distributions for each model in the 1-step and 24-step forecasting scenarios, offering insights into their overall performance and error characteristics. Tables 10 and 11 offer more information about the statistics of the RMSE set produced from the results.

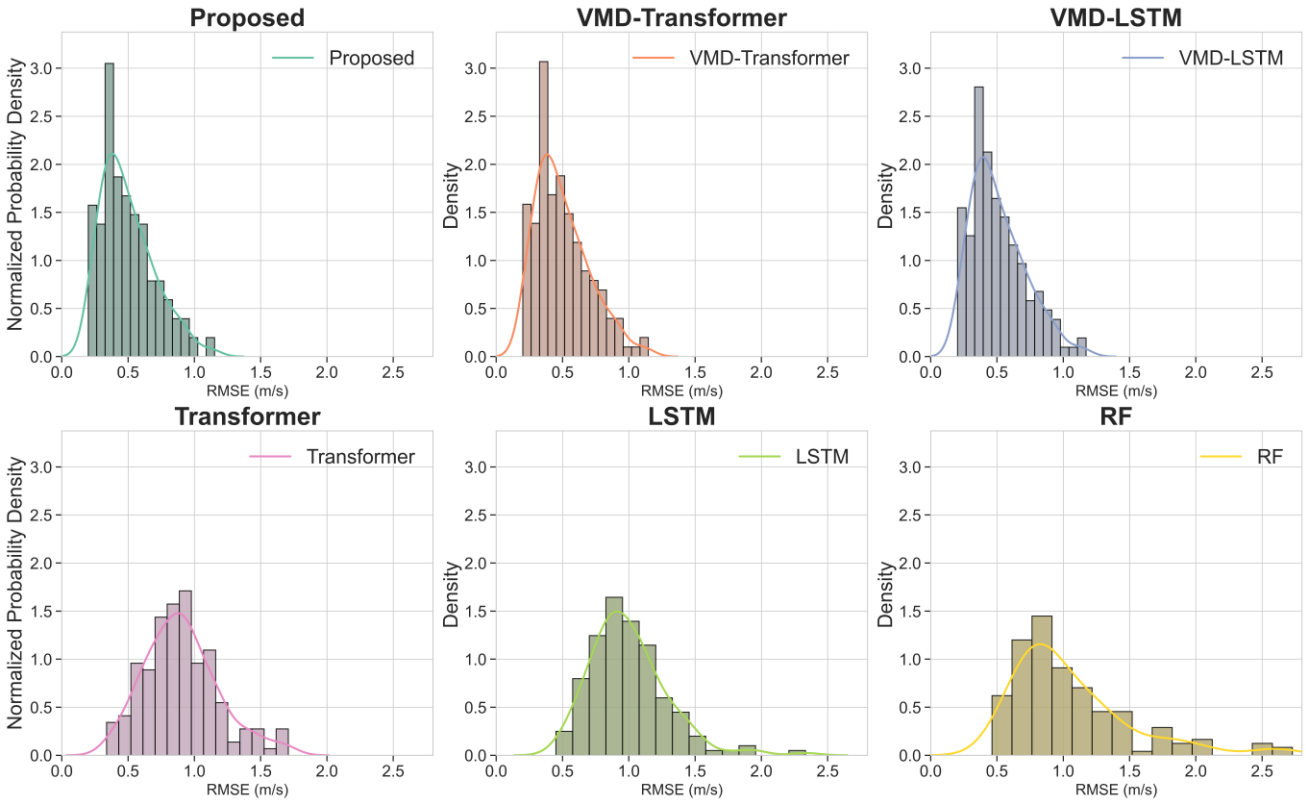

Figure 9. Error distribution of 6 models for 1-step forecasting.

Figure 10. Error distribution of 6 models for 24-step forecasting.

Table 10. Statistical summary of the RMSE from all models (1-step scenario).

Statistical value	Proposed	VMD-Transformer	VMD-LSTM	Transformer	LSTM	RF
Mean (m/s)	0.496	0.499	0.511	0.904	0.997	1.061
Standard Deviation (m/s)	0.201	0.202	0.206	0.279	0.290	0.457
Minimum (m/s)	0.200	0.200	0.203	0.338	0.451	0.462
25% (Q1) (m/s)	0.348	0.350	0.359	0.717	0.816	0.752
50% (Q2) (m/s)	0.454	0.459	0.467	0.884	0.955	0.928
75% (Q3) (m/s)	0.618	0.624	0.641	1.065	1.147	1.222
Maximum (m/s)	2.984	2.994	3.460	4.216	5.506	5.169

Table 11. Statistical summary of the RMSE from all 6 models (24-step scenario).

Statistical value	Proposed	VMD-Transformer	VMD-LSTM	Transformer	LSTM	RF
Mean (m/s)	1.396	1.411	1.519	1.565	2.198	2.321
Standard Deviation (m/s)	0.524	0.523	0.574	0.618	0.945	1.094
Minimum (m/s)	0.498	0.511	0.525	0.454	0.647	0.732
25% (Q1) (m/s)	0.991	1.000	1.084	1.151	1.496	1.485
50% (Q2) (m/s)	1.285	1.290	1.422	1.488	2.030	2.052
75% (Q3) (m/s)	1.756	1.786	1.907	1.890	2.663	2.968
Maximum (m/s)	2.984	2.994	3.460	4.216	5.506	5.169

Figures 9 and 10 present the RMSE distributions of six forecasting models for the 1-step and 24-step scenarios, respectively, across 160 daily predictions. For the 1-step forecasting scenario, the proposed model EVMD-DWT-Transformer, VMD-Transformer, and VMD-LSTM models exhibit compact, left-skewed distributions, with density peaks between 0.3 and 0.6 m/s, suggesting high accuracy and low variability. Among them, the proposed model shows the sharpest peak, followed by VMD-Transformer and VMD-LSTM. In contrast, the Transformer, LSTM, and RF models have broader distributions that extend to higher RMSE values, with noticeably flatter peaks and heavier right tails, highlighting increased prediction errors and instability. In the 24-step case, all models display right-shifted and more dispersed RMSE distributions due to the longer forecasting horizon. Nonetheless, the proposed model remains the most concentrated, followed closely by VMD-Transformer and VMD-LSTM. Transformer, LSTM, and especially RF exhibit wide spreads and flatter curves, indicating greater uncertainty and frequent large errors. These visual patterns confirm that hybrid models, especially the proposed model, maintain better forecasting reliability even under extended prediction horizons.

To further confirm the performance of the proposed model, Tables 10 andTable **11** provide a detailed statistical summary of the RMSE values for the six models under 1-step and 24-step forecasting scenarios. In Table 10, the proposed model achieves the lowest mean RMSE (0.496 m/s) and standard deviation (0.201 m/s), confirming its strong accuracy and consistency. Its IQR spans from 0.348 (Q1) to 0.618 (Q3), and its maximum RMSE remains modest at 2.984 m/s. The VMD-Transformer and VMD-LSTM also show competitive performance, with mean RMSEs of 0.499 and 0.511 m/s, and similarly narrow IQRs, indicating stable forecasts. In contrast, the Transformer, LSTM, and RF models yield much higher mean RMSEs (0.904, 0.997, and 1.061 m/s, respectively) and broader IQRs. RF performs the worst, with the highest variability (standard deviation = 0.457 m/s) and a maximum RMSE of 5.169 m/s, suggesting poor reliability in short-term forecasting.

Table 11 reveals performance reduction across all models in the 24-step scenario due to the increased forecasting complexity. Still, the proposed model remains the most accurate, with the lowest mean RMSE of 1.396 m/s, a relatively low standard deviation of 0.524 m/s, and a tight IQR of 0.991 to 1.756 m/s. The VMD-Transformer and VMD-LSTM follow closely with mean RMSEs of 1.411 and 1.519 m/s, respectively. Transformer, LSTM, and RF again exhibit higher average errors (1.565, 2.198, and 2.321 m/s) and significantly larger spreads, with RF showing the widest IQR (1.485 to 2.968 m/s) and the highest maximum RMSE (5.169 m/s). These results further validate the robustness and predictive stability of the proposed model under both short- and long-term horizons.

4.4. Computational complexity analysis

To evaluate the practical viability of the proposed framework, the computational cost was analyzed in terms of total training time and inference latency. All experiments were conducted in a Google Colab environment using a single NVIDIA T4 GPU. The corresponding results are reported in Table 12.

Table 12. Training and inference time for Transformer-based models.

Model	Proposed	VMD-Transformer	Transformer
Training time (min)	76.6	64.6	3.1
Inference time (ms)	3849.42	2732.31	217.62

The total training time denotes the cumulative time required to train all Transformer sub-models associated with the decomposed signal components, consisting of 9 modes for the VMD-Transformer and 12 modes for the proposed EVMD-DWT-Transformer. Due to the inclusion of additional DWT-based processing and a larger number of decomposed modes, the proposed model requires a longer training time (76.6 min) compared to the VMD-Transformer (64.6 min), representing an increase of approximately 18.5%. However, this overhead is incurred entirely during the offline training phase and does not impact the forecasting stage.

With respect to inference latency, the proposed EVMD-DWT-Transformer requires 3849.42 ms to generate a complete multi-step forecast, compared to 2732.31 ms for the VMD-Transformer and 217.62 ms for the standalone Transformer. The increased inference time is primarily attributed to the sequential reconstruction and aggregation of predictions from multiple decomposed components. Nevertheless, given that wind speed forecasting in this study is performed at an hourly resolution, the observed latency remains acceptable for offline analysis, scenario evaluation, and short-term operational planning.

Overall, the computational cost analysis indicates that the proposed framework achieves a favorable trade-off between predictive accuracy and computational efficiency. While more computationally demanding than single-model baselines, the EVMD-DWT-Transformer remains practical for research-oriented studies and decision-support applications where forecasting accuracy is prioritized over strict real-time constraints.

5. Conclusions

This study presents a novel secondary decomposition approach that integrates EVMD, DWT, and Transformer models for 1-step and 24-step wind speed forecasting. The proposed model was evaluated against VMD-LSTM, VMD-Transformer, and other standalone models using data obtained from the Thai Hoa power plant. The results demonstrate that the proposed model outperforms all other models, exhibiting superior performance across various evaluation metrics. Specifically, when evaluated for the entire dataset, the RMSE of the proposed model is 0.535 m/s for 1-step forecasting and 0.834 m/s for 24-step forecasting, both lower than those of any other models. Furthermore, single decomposition models continue to exhibit strong performance when compared to standalone models. Future work could focus on optimizing hyperparameters, enhancing DWT decomposition methods to better determine the optimal number of modes, or extending the approach from univariate to multivariate analysis.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research is funded by Hanoi University of Science and Technology (HUST) under project number T2025-PC-056.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

N.T.H.T.: Supervision, Project administration, Conceptualization, Methodology, Formal analysis, Writing—review & editing; D.X.B.: Conceptualization, Methodology, Data curation, Writing—original draft, Visualization; N.T.T.A.: Software development, Model implementation, Validation, Investigation; P.P.K.: Software development, Code optimization, Results interpretation, Visualization.

References

1. Karaman ÖA (2023) Prediction of wind power with machine learning models. *Appl Sci* 13: 11455. https://doi.org/10.3390/app132011455
2. Foley AM, Leahy PG, Marvuglia A, et al. (2012) Current methods and advances in forecasting of wind power generation. *Renewable Energy* 37: 1–8. https://doi.org/10.1016/j.renene.2011.05.033
3. Han Y, Mi L, Shen L, et al. (2022) A short-term wind speed interval prediction method based on WRF simulation and multivariate line regression for deep learning algorithms. *Energy Convers Manage* 258: 115540. https://doi.org/10.1016/j.enconman.2022.115540
4. Landberg L (1999) Short-term prediction of the power production from wind farms. *J Wind Eng Ind Aerodyn* 80: 207–220. https://doi.org/10.1016/S0167-6105(98)00192-5
5. Liu X, Lin Z, Feng Z (2021) Short-term offshore wind speed forecast by seasonal ARIMA—A comparison against GRU and LSTM. *Energy* 227: 120492. https://doi.org/10.1016/j.energy.2021.120492
6. Liu M, Cao Z, Zhang J, et al. (2020) Short-term wind speed forecasting based on the Jaya-SVM model. *Int J Electr Power Energy Syst* 121: 106056. https://doi.org/10.1016/j.ijepes.2020.106056
7. Zhang X, Zhao H, Yao J, et al. (2025) A multi-scale component feature learning framework based on CNN-BiGRU and online sequential regularized extreme learning machine for wind speed prediction. *Renewable Energy* 242: 122427. https://doi.org/10.1016/j.renene.2025.122427
8. Houndekindo F, Ouarda TBMJ (2025) LSTM and Transformer-based framework for bias correction of ERA5 hourly wind speeds. *Energy* 328: 136498. https://doi.org/10.1016/j.energy.2025.136498
9. Nguyen THT, Phan QB (2022) Hourly day ahead wind speed forecasting based on a hybrid model of EEMD, CNN-Bi-LSTM embedded with GA optimization. *Energy Rep* 8: 53–60. https://doi.org/10.1016/j.egyr.2022.05.110
10. Li S, Guo L, Zhu J, et al. (2025) Medium-term offshore wind speed multi-step forecasting based on VMD and GRU-MATNet model. *Ocean Eng* 325: 120737. https://doi.org/10.1016/j.oceaneng.2025.120737
11. Xu X, Hu S, Shao H, et al. (2023) A spatio-temporal forecasting model using optimally weighted graph convolutional network and gated recurrent unit for wind speed of different sites distributed in an offshore wind farm. *Energy* 284: 128565. https://doi.org/10.1016/j.energy.2023.128565

12. Joseph LP, Deo RC, Prasad R, et al. (2023) Near real-time wind speed forecast model with bidirectional LSTM networks. *Renewable Energy* 204: 39–58. https://doi.org/10.1016/j.renene.2022.12.123
13. Feng H, Jin Y, Laima S, et al. (2022) A ML-Based wind speed prediction model with truncated real-time decomposition and multi-resolution Data. *Appl Sci* 12: 9610. https://doi.org/10.3390/app12199610
14. Yue Y, Zheng W, Wu A, et al. (2025) Ultra-short-term wind speed forecasting based on secondary decomposition and Transformer-MLR combined model. *Electr Power Syst Res* 246: 111702. https://doi.org/10.1016/j.epsr.2025.111702
15. Nhat NNV, Huu DN, Nguyen TTH (2024) Short-term multi-step forecasting of rooftop solar power generation using a combined data decomposition and deep learning model of EEMD-GRU. *J Renewable Sustainable Energy* 16: 016101. https://doi.org/10.1063/5.0176951
16. Li Q, Wang G, Wu X, et al. (2024) Arctic short-term wind speed forecasting based on CNN-LSTM model with CEEMDAN. *Energy* 299: 131448. https://doi.org/10.1016/j.energy.2024.131448
17. Yan Y, Wang X, Ren F, et al. (2022) Wind speed prediction using a hybrid model of EEMD and LSTM considering seasonal features. *Energy Rep* 8: 8965–8980. https://doi.org/10.1016/j.egyr.2022.07.007
18. Jin Z, Fu X, Xiang L, et al. (2025) Informer learning framework based on secondary decomposition for multi-step forecast of ultra-short term wind speed. *Eng Appl Artif Intell* 139: 109702. https://doi.org/10.1016/j.engappai.2024.109702
19. Barjasteh A, Ghafouri SH, Hashemi M (2024) A hybrid model based on discrete wavelet transform (DWT) and bidirectional recurrent neural networks for wind speed prediction. *Eng Appl Artif Intell* 127: 107340. https://doi.org/10.1016/j.engappai.2023.107340
20. Li N, Dong J, Liu L, et al. (2023) A novel EMD and causal convolutional network integrated with Transformer for ultra short-term wind power forecasting. *Int J Electr Power Energy Syst* 154: 109470. https://doi.org/10.1016/j.ijepes.2023.109470
21. Gao X, Guo W, Mei C, et al. (2023) Short-term wind power forecasting based on SSA-VMD-LSTM. *Energy Rep* 9: 335–344. https://doi.org/10.1016/j.egyr.2023.05.181
22. Karijadi I, Chou SY, Dewabharata A (2023) Wind power forecasting based on hybrid CEEMDAN-EWT deep learning method. *Renewable Energy* 218: 119357. https://doi.org/10.1016/j.renene.2023.119357
23. Dragomiretskiy K, Zosso D (2014) Variational Mode Decomposition. *IEEE Trans Signal Process* 62: 531–544. https://doi.org/10.1109/TSP.2013.2288675
24. Jiang Y, Zhao N, Peng L, et al. (2019) A new hybrid framework for probabilistic wind speed prediction using deep feature selection and multi-error modification. *Energy Convers Manage* 199: 111981. https://doi.org/10.1016/j.enconman.2019.111981
25. Nabavi SA, Mohammadi S, Motlagh NH, et al. (2024) Deep learning modeling in electricity load forecasting: Improved accuracy by combining DWT and LSTM. *Energy Rep* 12: 2873–2900. https://doi.org/10.1016/j.egyr.2024.08.070
26. Arseven B, Çınar SM (2025) A novel hybrid solar radiation forecasting algorithm based on discrete wavelet transform and multivariate machine learning models integrated with clearness index clusters. *J Atmos Sol-Terr Phys* 267: 106417. https://doi.org/10.1016/j.jastp.2025.106417

27. Vaswani A, Shazeer N, Parmar N, et al. (2017) Attention Is All You Need. https://doi.org/10.48550/arXiv.1706.03762