Energy Reports 14 (2025) 3237–3253

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Transfer learning based-hybrid model for short-term wind speed forecasting

Lokmene Melalkia [a,*], Farid Berrezzek [a], Khaled khelil [a], Abdelhakim Saim [b,*], Radouane Nebili [a]

[a] Faculty of Science and Technology, Laboratory of Electrical Engineering and Renewable Energy (LEER Lab), University of Souk Ahras, Souk Ahras 41000, Algeria
[b] Nantes Université, Institut de Recherche en Energie Electrique de Nantes Atlantique, IREENA, UR 4642, Saint-Nazaire F-44600, France

ARTICLE INFO

Keywords:
Wind speed forecasting
CEEMDAN
Bi-GRU Encoder-Decoder
Transfer learning
Artificial intelligence
Renewable energy

ABSTRACT

Reliable integration of wind energy into the power grid relies on accurate short-term wind speed forecasting. To address this challenge, this paper introduces a hybrid forecasting framework that combines Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Bidirectional Gated Recurrent Unit Encoder-Decoder (Bi-GRU-ED), and transfer learning. CEEMDAN decomposes the original wind speed series into intrinsic mode functions (IMFs) and a residual, enabling the capture of nonlinear characteristics while reducing noise interference. The forecasting process begins with forecasting the residual component using the $\textit{Bi-GRU-ED}_{Res}$ model. Subsequently, transfer learning is employed to reduces computational demands and boosts performance by leveraging pre-trained knowledge across decomposition components to efficiently forecast the remaining IMFs, reducing training time without compromising accuracy. Finally, all forecasted components are aggregated to reconstruct the wind speed series. The proposed approach was validated on two datasets from the Maghreb region and compared with multiple benchmark models. Results demonstrate that the hybrid model achieves superior forecasting accuracy, confirming its suitability for short-term wind speed prediction and potential application in renewable energy systems.

1. Introduction

1.1. Research background

As global energy systems shift toward sustainability, wind energy has become a central pillar of this transition. Unlike fossil-based sources, wind is clean, abundant, and increasingly cost-effective. Yet, harnessing its full potential depends not only on infrastructure but also on accurate forecasting of wind speed, a variable known for its chaotic and non-stationary behavior. Inaccurate predictions can lead to energy imbalances, grid instability, and increased operational costs (Sun et al., 2024). These challenges have sparked the need for intelligent forecasting models capable of capturing the intricate dynamics of wind patterns in diverse environments (Joseph et al., 2024).

1.2. Literature review

Wind speed forecasting plays a pivotal role in the efficient integration of wind energy into power systems, mainly due to the stochastic and non-stationary nature of wind. Forecasting methods are typically classified based on the prediction horizon into four categories: very short-term (minutes to hours), short-term (hours to days), medium-term (days to weeks), and long-term (weeks to years) forecasts (Joseph et al., 2024). Similarly, methodologies can be broadly divided into physical, statistical, artificial intelligence, and hybrid approaches (Khelil et al., 2021).

Physical approaches, based on Numerical Weather Prediction (NWP), utilize topographical dataset, temperature, humidity, and pressure to forecast weather using mathematical formulas (Liu et al., 2021). Although advanced NWP models can accurately predict weather for specific locations, the dynamic nature of wind requires extensive datasets and lengthy computation times. While suitable for long-term wind speed forecasting, their reliance on large datasets, high processing demands, and complex models impedes their applicability for smaller-scale forecasting. Therefore, it is crucial to explore alternative methods to achieve a balance between precision and efficiency. The statistical approaches, including models like the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA), analyze temporal dependencies within historical data. (Erdem et al., 2014) explored Autoregressive Moving Average (ARMA) and

* Corresponding authors.
E-mail addresses: l.melalkia@univ-soukahras.dz (L. Melalkia), Abdelhakim.Saim@univ-nantes.fr (A. Saim).

https://doi.org/10.1016/j.egyr.2025.10.007
Received 1 November 2024; Received in revised form 28 August 2025; Accepted 19 October 2025
Available online 27 October 2025

Generalized AutoRegressive Conditional Heteroskedasticity models for both mean and volatility forecasting of wind speed, highlighting their effectiveness in structured time series analysis. However, traditional statistical models often fall short in handling the nonlinear and nonstationary patterns typical of wind behavior. More advanced versions, such as Nonlinear Autoregressive Exogenous Mode (NARX) (Cadenas et al., 2016) and Nonlinear Autoregressive Moving Average (NARMA) (Waheeb et al., 2019), have been proposed to capture complex patterns, but they are sensitive to data noise and outliers (Parmezan et al., 2019).

In contrast, artificial intelligence-based methods have demonstrated strong capabilities in modeling the complex and nonlinear characteristics of wind speed time series (Sun et al., 2024). Artificial Neural Networks (ANNs) have been widely adopted for their ability to approximate nonlinear functions using historical meteorological data, although their performance can be affected by local minima and overfitting issues (Parmezan et al., 2019). Support Vector Machines (SVM), known for their generalization capability, have also shown promise in short-term wind forecasting, especially when dealing with limited and noisy datasets (Parmezan et al., 2019). In (Li et al., 2016), an Extreme Learning Machine model (ELM) with an integrated error correction mechanism was proposed to enhance short-term wind power prediction, achieving rapid training and enhanced accuracy. (Shahid et al., 2021) employed a Stacked Long Short-Term Memory model (S-LSTM), while (Li et al., 2020) proposed an ensemble Gated recurrent unit (GRU) based interval prediction framework, both demonstrating the effectiveness of deep learning architectures in time-dependent forecasting. (Trebing and Mehrkanoon, 2020) extended this direction using multidimensional convolutional neural networks (CNN), emphasizing the role of spatial dependencies in addition to temporal modeling.

To overcome individual model limitations, hybrid approaches have emerged. These integrate multiple techniques, typically artificial intelligence-based models enhanced through preprocessing such as decomposition methods. (Shen et al., 2022) proposed a hybrid Convolutional Neural Network (CNN) with long short-term memory (LSTM) to forecast wind speed for autonomous sailboats. Similarly, (Khodayar et al., 2017) introduced a robust deep neural architecture for short-term forecasting, and Garg and (Garg and Krishnamurthi, 2023) built a Convolutional Neural Network (CNN) encoder–decoder with long short-term memory (LSTM) to improve sequential learning. Recently, (Houndekindo and Ouarda, 2025) introduced a novel approach using Gradient Boosted Trees (GBOOST), Long Short-Term Memory (LSTM) and Transformer-based bias correction using static covariates to enhance geographic adaptability in wind speed forecasting. Their work improved wind power forecasting across geographic regions and introduced new directions for integrating static covariates into forecasting pipelines. While hybrid models offer performance gains, they may still be computationally demanding.

Given the complexity and irregularity of wind speed signals, many studies incorporate signal decomposition as a preprocessing step to reduce data complexity and improve model accuracy. (Khelil et al., 2021) employed Genetic Algorithms (GA) to optimize Discrete Wavelet Filters, enhancing data decomposition. (Sun and Wang, 2023) applied complete ensemble Empirical mode decomposition with adaptive noise (CEEMDAN) with a chimpanzee optimizer to tune an extreme learning machine (ELM), which outperformed standalone models. As an improved version of the Empirical Mode Decomposition (EMD), the authors (Yan et al., 2022) proposed a hybrid wind speed prediction model that integrates the methodologies of Ensemble Empirical Mode Decomposition (EEMD) as a preprocessing technique with Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long short term memory (LSTM). Additionally, in (Wu et al., 2022), introduced a Multistep short-term wind speed forecasting method based on EEMD decomposition combined with an encoder-decoder architecture. Another variation of Empirical Mode Decomposition (EMD), known as Variational Mode Decomposition (VMD), was utilized in (Zhao et al., 2023), combined with Convolutional Neural Network (CNN) with Gated Recurrent Unit (GRU). Further innovations include (Wu et al., 2022), who introduced the hybrid Variational Mode Decomposition, Adaptive Differential Evolution and Temporal Fusion Transformer (VMD-A-DE-TFT) model that leverages Temporal Fusion Transformers for interpretable multi-horizon wind forecasting. (Zhang et al., 2025) developed a CNN-BiGRU architecture optimized by a dandelion algorithm and Symplectic Geometric Mode Decomposition (SGMD), achieving high accuracy through complexity-based signal separation. These models represent a shift toward interpretable and highly specialized deep learning architectures. However, the Ensemble Empirical Mode Decomposition (EEMD) algorithm was unable to fully eliminate Gaussian white noise during signal reconstruction, and Variational Mode Decomposition (VMD) typically exhibited reconstruction errors, as its decomposition of signals did not precisely match the original dataset during the reconstruction process (Yang et al., 2023). In a similar direction, (Sun et al., 2024) proposed a hybrid model that combines Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Grey Wolf Optimizer (GWO), and Bidirectional Long Short-Term Memory (Bi-LSTM), effectively enhancing signal-relevant components through noise-tolerant decomposition and metaheuristic tuning. While (Chen et al., 2025) addressed interval forecasting by proposing the Combined Method for Interval Forecasting (CMIF), integrating Hilbert transforms, multi-objective optimization, and neural models. This produced narrow prediction intervals with high coverage rates but faced challenges in differentiability and training.

Transfer learning, a growing trend in deep learning, has become vital in addressing data scarcity and model generalization challenges across various domains. The core advantage of this approach lies in its ability to transfer learned knowledge from a source domain to a target domain, which is especially beneficial when the target task suffers from limited training data. This capability has made it a valuable technique in wind speed forecasting. For instance, (Ji et al., 2022) proposed a Convolutional Neural Network and Gated Recurrent Unit-based transfer learning model (CNN-GRU-TL)) for short-term wind forecasting in canyon regions. (Liang et al., 2021) implemented Bidirectional long short-term memory (Bi-LSTM) models pre-trained on four wind farm locations and transferred them to a centralized control centre. However, the success of transfer learning in these applications depends heavily on the domain similarity between source and target domain as (Zhuang et al., 2021) and (Luo et al., 2023) pointed out, the risk of negative transfer persists when source and target distributions differ. Strategies like partial layer retraining (Zhuang et al., 2021) can reduce computation but may limit adaptability to the target task. Retraining all layers may offer superior performance by avoiding negative transfer but incurs more computational cost and training complexity (Luo et al., 2023). Thus, careful consideration of domain similarity and data availability is essential when applying transfer learning techniques to wind speed forecasting. Table 1 presents a synthesis of recent wind forecasting methods, outlining their architectures, key strengths, and unresolved limitations. This consolidated analysis underscores persistent research gaps particularly in decomposition reuse, transferability, and joint learning strategies.

1.3. Contributions and novelty

Building upon the existing research and the identified challenges in wind speed forecasting, this study introduces a novel hybrid framework, CEEMDAN-Bi-GRU-ED-TL, for enhanced short-term wind speed prediction. The model integrates three key components: **(i)** Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) for signal preprocessing, **(ii)** a Bidirectional Gated Recurrent Unit Encoder-Decoder (Bi-GRU-ED) network for temporal feature extraction, and **(iii)** transfer learning for efficient and scalable model training. The primary contributions of this research are as follows:

Table 1
Comparative summary of recent wind speed forecasting studies: techniques, strengths, and remaining gaps.

Authors	Technique / Architecture Used	Main Strengths	Identified Gaps (Addressed in This Work)
(Joseph et al., 2024)	Hybrid decomposition + Gated Additive Trees (GAT)	Enables automatic feature relevance selection and improved forecasting stability under temporal variability	Inflexible tree-based architecture lacks dynamic residual correction; absence of IMF-specific modeling impairs error isolation and component interpretability.
(Sun et al., 2024)	Temporal Convolutional Network (TCN) + Extended Kalman Filter (EKF) + Monte Carlo Method (MCM)	Achieves enhanced prediction accuracy at sub-hour scales by integrating filter-based uncertainty correction and temporal convolutions.	Poor robustness to high-frequency noise due to absence of signal preprocessing; lacks temporal granularity enhancement mechanisms.
(Zhang et al., 2025)	Convolutional Neural Network (CNN)+ Bidirectional Gated Recurrent Unit (Bi-GRU) + Symplectic Geometric Mode Decomposition (SGMD)+ Online Sequential Regularized Extreme Learning Machine (OSRELM) + Dandelion Optimizer	multi-scale hybrid learning framework separating high- and low-complexity components, achieves high accuracy through decomposition and stage-specific predictors; fast convergence.	Only considers wind speed as input; potential underutilization of high-frequency sub-signals, lacks generalization to multi-step horizons.
(Shen et al., 2022)	Convolutional Neural Network (CNN) + Long Short-Term Memory (LSTM)	Learns complex spatial-temporal wind dynamics jointly, improving prediction in topographically diverse or rapidly shifting environments.	High parameter count introduces overfitting risk; lacks hierarchical decomposition to localize forecasting complexity at multiple resolutions.
(Sun et al., 2024)	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) + Grey Wolf Optimizer (GWO) + Bidirectional Long Short-Term Memory (Bi-LSTM)	Combines noise-tolerant decomposition with population-based optimization to isolate and enhance signal-relevant components.	Independent model training for decomposed modes leads to redundant computation and suboptimal convergence; lacks joint learning strategy.
(Zhao et al., 2023)	Variational Mode Decomposition (VMD) + Convolutional Neural Network (CNN) + Gated Recurrent Unit (GRU)	Extracts multi-resolution temporal features and spatial dependencies across channels using hierarchical deep feature fusion.	Treats IMF predictors as isolated learners, resulting in inconsistent performance across modes; poor generalization to datasets with varying complexity.
(Wu et al., 2022)	Ensemble Empirical Mode Decomposition (EEMD) decomposition combined with an encoder-decoder.	Enables flexible sequence modeling for multi-horizon forecasts using attention mechanisms for temporal importance weighting.	Transformer requires dense training data, no mitigation for sparse regimes or temporal extrapolation challenges in small-sample domains.
(Wu et al., 2022)	Variational Mode Decomposition (VMD) + Temporal Fusion Transformer (TFT)	Improves model explainability and forecasting precision through frequency-aware attention and dynamic covariate encoding.	Does not constrain attention to denoised or decomposed features; overfitting risk persists on high-frequency, low-energy IMF components.
(Liang et al., 2021)	Bidirectional Long Short-Term Memory (Bi-LSTM) + Multi-Objective Optimization using Fuzzy Adaptive Differential Algorithm (MOO-FADA) + Transfer Learning (TL)	Applies evolutionary optimization to fine-tune cross-site knowledge transfer and enhance spatial generalization in forecasting.	Transfer mechanism is not modular; excessive parameter sensitivity during adaptation leads to poor reproducibility in operational settings.
(Ji et al., 2022)	Convolutional Neural Network (CNN) + Gated Recurrent Unit (GRU) + Transfer Learning (TL)	Utilizes pretrained hierarchical encoding for rapid convergence and robust predictions under limited-data conditions.	Lack of structural reusability between source and target domains; ignores inter-component dependencies during model transfer.
(Sun and Wang, 2023)	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) + Chimp Optimization + Extreme Learning Machine (ELM)	Improves noise resilience and resolution of input series via adaptive decomposition combined with low-latency learners.	Full re-training for each IMF increases computational burden; lacks shared parameter base or residual adaptation layers.
(Houndekindo and Ouarda, 2025)	Gradient Boosted Trees (GBOOST) + Long Short-Term Memory (LSTM) with Time-Invariant (TI) and Time-Resolved (TR) correction	Corrects geographic bias in wind speed time series using static covariates while preserving high-resolution prediction fidelity.	Emphasizes spatial-scale correction via static covariates but neglects temporal signal nonlinearity; unsuitable for high-resolution multi-horizon forecasting.

1. This study proposes a novel hybrid architecture that combines CEEMDAN, Bi-GRU-ED, and transfer learning to enhance short-term wind speed forecasting accuracy and robustness.
2. The CEEMDAN technique is employed to decompose the raw wind speed signal into intrinsic mode functions (IMFs) and a residual component, effectively reducing noise and isolating nonlinear patterns in the data.
3. A Bi-GRU-ED deep learning architecture is developed to extract complex temporal dependencies within the residual component, leveraging bidirectional processing and an encoder-decoder structure to improve sequence modeling.
4. Transfer learning is applied by reusing the trained Bi-GRU-ED model from the residual component to forecast the individual IMFs, reducing computational overhead while preserving model performance.

1.4. The main structure

The remainder of the paper is structured as follows: Section 2 outlines the structure of the proposed forecasting model and the material and methods employed in the study. Sections 3 and 4 describe the methods employed in the study and the experimental setup, respectively, while Section 5 discusses the experimental results along with a comparative analysis. Finally, the concluding remarks are presented in Section 6 and future research.

2. Structure of the proposed forecasting model

The main steps of the CEEMDAN-Bi-GRU-ED-TL forecasting scheme are outlined in Fig. 1 as follows:

1. A level-6 CEEMDAN decomposition is applied to break down the original wind speed $w[n]$ into a series of six *IMFs* ($IMF_1, \ldots, IMF_6$)and a residual component *Res*. Simultaneously, the other meteorological features (T, P *and* W_{10}) are normalized to the range $[0, 1]$ using Eq. (16).
2. The residual component Res (which captures the low-frequency trend) is forecasted using a Bi-GRU Encoder–Decoder (ED) model. The inputs to this model include the decomposed IMFs, residual and normalized meteorological features (T_n, P_n, w_{10n}), structured over a time window (10-time steps). The trained model is saved as *Bi-GRU-ED*$_{Res}$.
3. The *Bi-GRU-ED*$_{Res}$ model is fine-tuned for each of the six IMFs, resulting in six models *Bi-GRU-ED*$_l$ ($l = 1, \ldots, 6$). Each model forecasts the corresponding $\widehat{IMF}_l$ using the same type of input data (i.e., the six IMFs, the residual, and the normalized features), thereby leveraging the knowledge learned during residual forecasting.

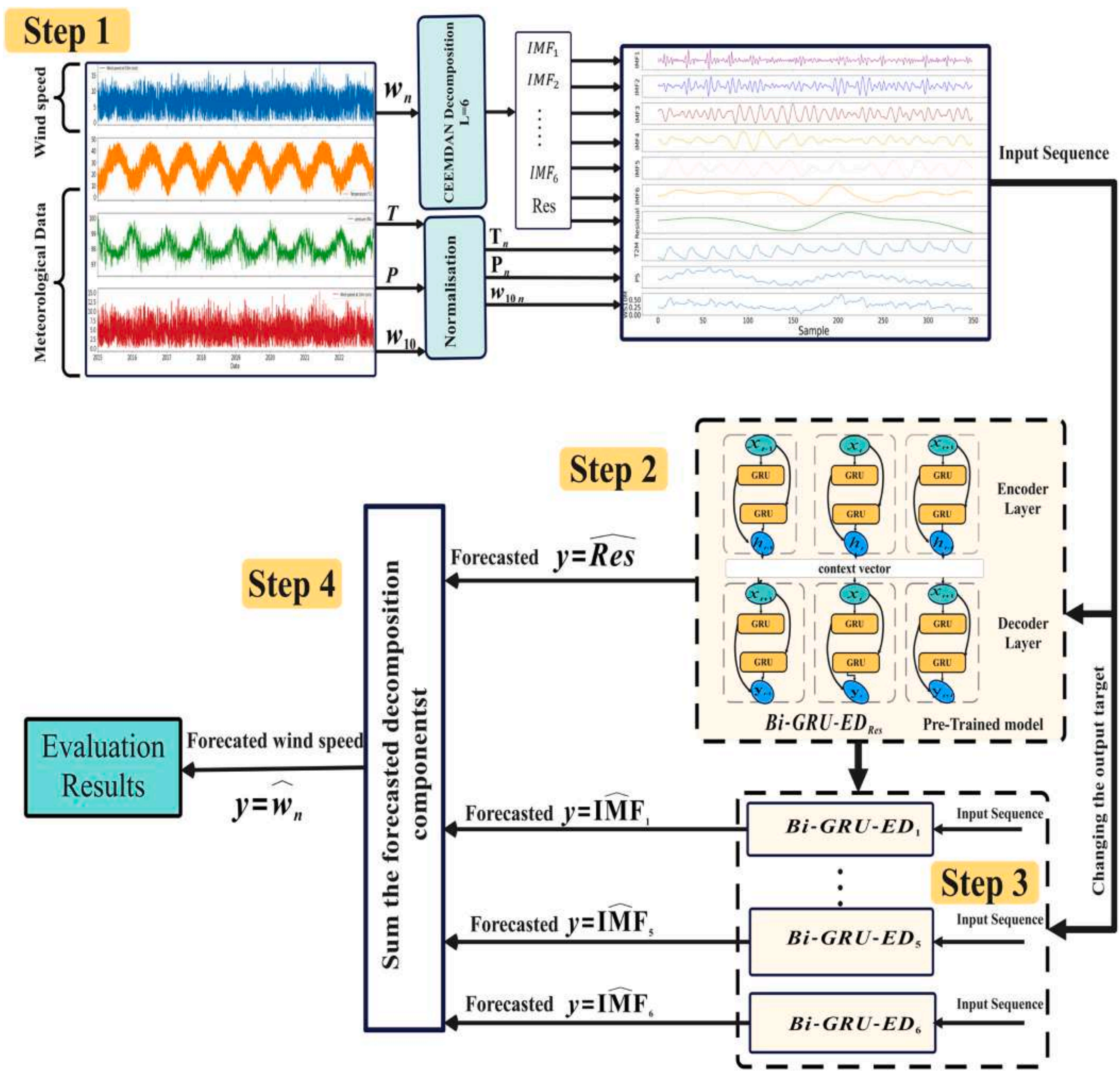

Fig. 1. The proposed CEEMDAN-Bi-GRU-ED-TL forecasting structure.

4. Finally, the forecasted $\widehat{Res}$ and $\widehat{IMF}_l \quad (l = 1, \ldots, 6)$ are aggregated to reconstruct the original forecasted wind speed signal $\widehat{w}[n]$, following the CEEMDAN additive reconstruction principle.

3. Material and methods

3.1. Dataset description

To assess the performance of the proposed forecasting scheme, two datasets from different sites were gathered. The initial dataset, designated as Dataset 1, spans from 2015 to 2023 and originates from Adrar, Sebaa, Algeria. Meanwhile, the second dataset, labelled as Dataset 2, covers the period from 2016 to 2021 and was acquired from Chennai in Tataouine, Tunisia. These two datasets include hourly recordings of wind speed (in meters per second) at both 50 m and 10 m, as well as measurements of pressure (P, in Pascals) and temperature (T, in °C). In total, Dataset 1 comprises 70,128 samples, while Dataset 2 contains 61,343 samples (Melalkia and Berrezzek, 2024). The locations and the correlations between the dataset variables are illustrated in Fig. 2.

3.2. The deep learning model architecture

3.2.1. Bidirectional gated recurrent unit (Bi-GRU)

The gated recurrent unit, an advancement introduced by Kyunghyang Cho in 2014 (Zhang et al., 2020), represents an enhanced version of the traditional Recurrent neural network. It addresses the prevalent issue of vanishing gradients encountered in conventional RNNs during the learning of long-term dependencies. The GRU noteworthy advantage is manifested in its gating mechanism, which incorporates two gates, the update gate z_t and the reset gate r_t. The update gate determines the extent to which historical information should be transferred to subsequent time steps, enabling the GRU to retain pertinent information and discard unnecessary or redundant inputs by regulating the flow of information. Conversely, the reset gate allows the model to decide how much of the previous information should be discarded or forgotten from the inputs. This gate permits the GRU to reset its state h_t, erasing only the information that is no longer accurate or relevant from earlier time steps. The mathematical formulation of the GRU is presented by Eq. (1) Eq. (2) Eq. (3) and Eq. (4) (Li et al., 2020).

$$r_t = \sigma(w_r \bullet [h_{t-1}, x_t] + b_r) \tag{1}$$

$$z_t = \sigma(w_z \bullet [h_{t-1}, x_t] + b_z) \tag{2}$$

Where σ is the sigmoid function, x_t is the input given to the gate, w_r and w_z are respectively the weight parameters for the reset and update gates. h_{t-1} is the previous hidden state, b_r and b_z, represent the bias of the reset and the update gates respectively.

- Candidate Hidden State

The candidate's hidden state $\widetilde{h}_t$ controls how much influence the previous hidden state has on the hidden state h_t using the reset gate value.

$$\widetilde{h}_t = \tanh(w_h \bullet (r_t \odot h_{t-1}, x_t) + b_h) \tag{3}$$

Where $\odot$ is the Hadamard product

- Hidden state

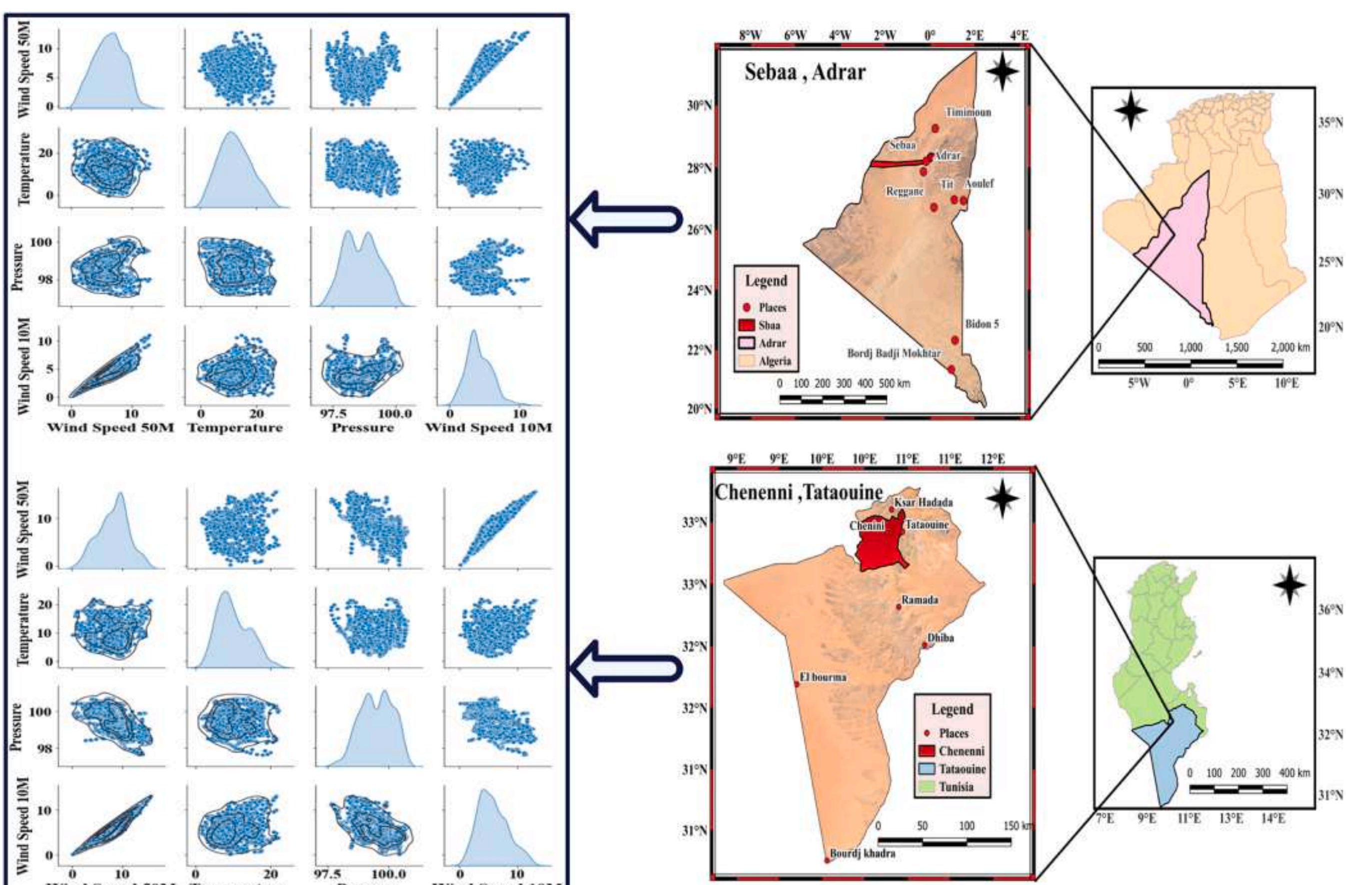

Fig. 2. Datasets description.

The current hidden state h_t is generated by the candidate hidden state $\widetilde{h}_t$ (Li et al., 2020):

$$h_t = (z_t \odot h_{t-1} + (1 - z_t) \odot \widetilde{h}_t) \tag{4}$$

In a simple GRU, the input sequence is processed sequentially in a unidirectional manner, moving from the beginning to the end of the input. An adapted version known as the Bidirectional GRU (Bi-GRU) has been introduced, which comprises two layers of GRU, as shown in Fig. 3. One GRU layer processes the input sequence in a forward direction which appears in Eq. (5), while the other GRU layer processes the input sequence in reverse as shown in Eq. (6) (Sun et al., 2024). The outputs from both layers are then concatenated using Eq. (7), resulting in a network capable of capturing both past and future contextual information simultaneously (Li et al., 2020).

$$\overrightarrow{h_t} = GRU_f\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{5}$$

$$\overleftarrow{h_t} = GRU_b(x_t, \overleftarrow{h_{t+1}}) \tag{6}$$

$$h_t = [\overrightarrow{h_t} \quad \bigoplus \quad \overleftarrow{h_t}] \tag{7}$$

3.2.2. Bidirectional GRU encoder-decoder architecture (Bi-GRU-ED)

In this implementation, the encoder-decoder bidirectional GRU model utilizes a bidirectional GRU as its encoder component as presented in Fig. 3. The input sequence x_t is analyzed by the encoder in two directions "forward and backward". This bidirectional operation allows the Bi-GRU to generate a context vector, which serves as a condensed representation of the temporal information present in the input dataset (Rai et al., 2022).

During the encoding process, the Bi-GRU network constantly receives input sequence x_t in succession. Each input step is processed individually, in the order it appears. The hidden state h_t of the Bi-GRU is updated when the model encounters each input. By applying this process to every element in the input sequence, the Bi-GRU is able to capture and integrate information from each time step. The final hidden state generated by the encoder typically serves as the context vector, which functions as the initial hidden state for the decoder.

The decoder, which is Bi-GRU as well, takes the context vector and the previously generated value as inputs and generates the output sequence, step by step, by considering the context vector and the previous outputs. To construct the output, the decoder uses encoded information from both the past and future contexts.

3.3. Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)

Empirical Mode Decomposition (EMD) is a popular method in signal processing, particularly for analyzing non-stationary and nonlinear signals. By breaking down the original signal into Intrinsic Mode Functions (IMFs), EMD enables the identification and isolation of distinct oscillatory modes present within the signal. This decomposition

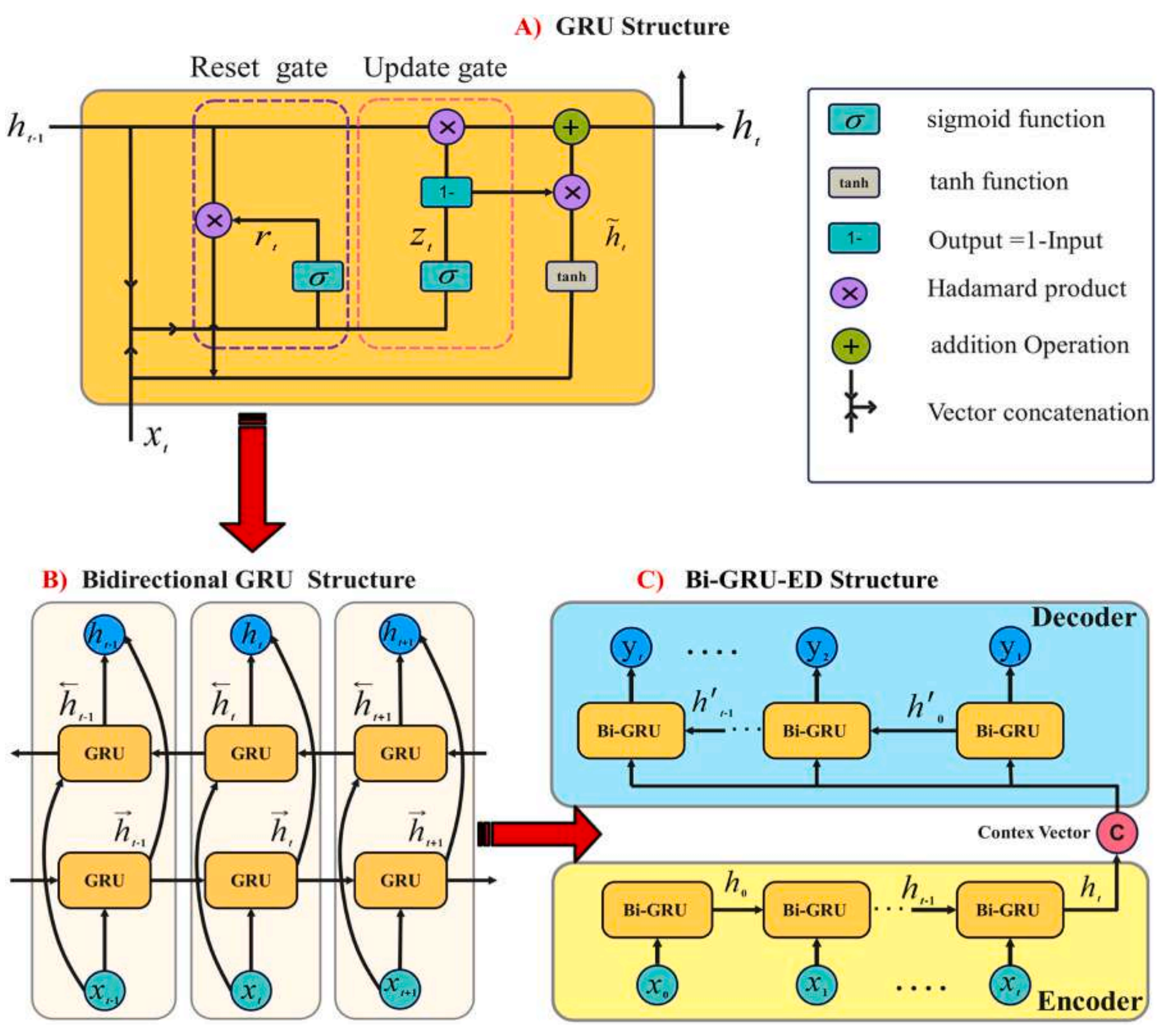

Fig. 3. Structure of GRU, Bi-GRU, Bi-GRU-ED models.

process is facilitated through the Hilbert–Huang Transform (Chen et al., 2019). An IMF is identified as a function that satisfies two criteria. The first criterion requires that the number of extrema (local maxima and minima) and zero crossings must be equal or differ by at most one. The second criterion states that the mean value of the envelope formed by the local maxima and the envelope generated by the local minima must be zero at any given point (Zhang et al., 2016). While EMD is acceptable for analysing time series signals, it suffers from a significant drawback known as mode mixing, which arises when a single IMF contains signals of different magnitudes or when a signal of similar scale appears across multiple IMF components. This phenomenon denotes a situation where an IMF produced by EMD decomposition comprises elements with varying frequencies, particularly within a short time, making it vulnerable to noise and outliers. To solve this issue, an improved variant of EMD known as Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) has been introduced (Torres et al., 2011). This method addresses certain drawbacks of EMD by generating multiple ensemble members and averaging the IMFs across signal trials. By doing so, it reduces sensitivity to specific noise patterns or outliers and mitigates mode mixing. Therefore, the ensemble averaging process aids in generating IMFs that are more stable and dependable (Sun et al., 2024). The CEEMDAN decomposition procedure, illustrated in Fig. 4, is applied through the following steps:

Step 1: Normal Gaussian white noise, denoted as $\omega^i(t)$ is added to the original time series dataset $y(t)$ during the ith iteration, as shown in Eq. (8) (Sun et al., 2024).

$$x^i(t) = y(t) + \varepsilon_0\omega^i(t) \quad ,i = 1 \quad ,2 \quad ,3\ldots\ldots,N \tag{8}$$

Here, ε_0 represents the noise weight coefficient, and N denotes the number of times the noise is added.

Step 2: The EMD method decomposes the time series signal $x^i(t)$ for N iteration, where the operator E represents the EMD decomposition operator. The first Intrinsic Mode Function $IMF_1(t)$ is derived by computing the average value of the decomposition outcome, as given in Eq. (9). The residual element $r_1(t)$ is obtained by subtracting the initial modal component from the original signal, as expressed by Eq. (10) (Sun et al., 2024).

$$IMF_1(t) = \frac{1}{N}\sum_{i=1}^{N} E_1\left[x^i(t)\right] \tag{9}$$

$$r_1(t) = y(t) - IMF_1(t) \tag{10}$$

Step 3: This step involves the addition of an adaptive white noise sequence to the residual component $r_1(t)$ creating a new time series signal $r_1(t) + \varepsilon_1 E_1(\omega_i(t))$, where E_1 is the first IMF obtained by EMD. This modified time series is then decomposed using EMD to obtain its mean value, leading to the extraction of the second $IMF_2(t)$ and the residual component $r_2(t)$. This process is described in Eq. (11) and Eq.

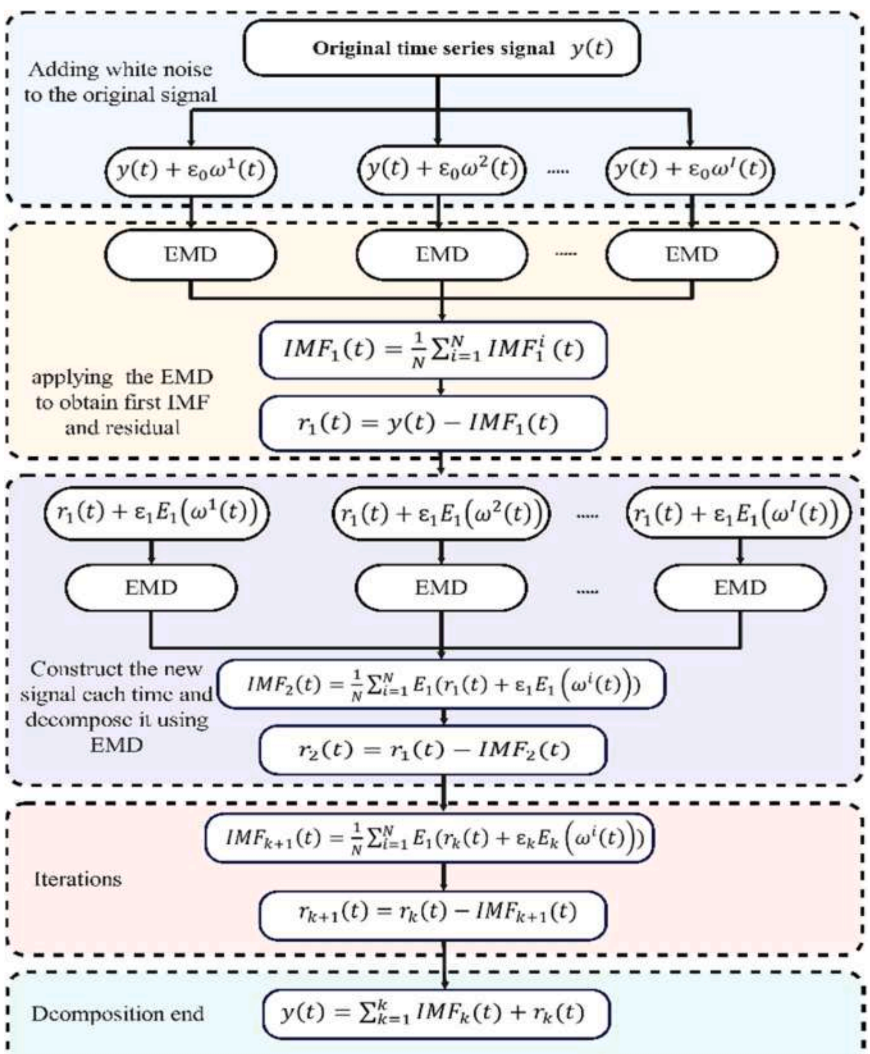

Fig. 4. CEEMDAN decomposition flowchart.

(12).

$$IMF_2(t) = \frac{1}{N}\sum_{i=1}^{N} E_1\left(r_1(t) + \varepsilon_1 E_1\left(\omega^i(t)\right)\right) \tag{11}$$

$$r_2(t) = r_1(t) - IMF_2(t) \tag{12}$$

Step 4: For $k = 2,\dots,K$, the kth $IMF_{k+1}(t)$ and $r_{k+1}(t)$ of CEEMDAN are obtained using Eq. (13) and Eq. (14).

$$IMF_{k+1}(t) = \frac{1}{N}\sum_{i=1}^{N} E_1\left(r_k(t) + \varepsilon_k E_k\left(\omega^i(t)\right)\right) \tag{13}$$

$$r_{k+1}(t) = r_k(t) - IMF_{k+1}(t) \tag{14}$$

Step 5: Step 4 is repeated until the resulting residual signal $r_k(t)$ reaches a point where further decomposition is no longer feasible (i.e., the residual does not contain at least two extrema). The number of obtained IMFs is indicated as k. At this stage, the original time series $y(t)$ is fully decomposed into its different components as identified in Eq. (15) (Sun et al., 2024).

$$y(t) = \sum_{k=1}^{k} IMF_k(t) + r_k(t) \tag{15}$$

3.4. Transfer learning

Transfer learning is a valuable technique employed to enhance forecasting accuracy. This method involves utilizing pre-trained models to transfer knowledge from previously learned tasks and apply it to related tasks (Yang et al., 2024). Transfer learning is typically classified into two major categories based on the consistency of the source and target feature space and label space:

- Homogeneous transfer learning, where the feature spaces X and the label spaces y between the source of the pre-trained model and the target in the new task are the same: $X^S = X^T, y^S = y^T$ otherwise $X^S \neq X^T$, $y^S \neq y^T$
- Heterogeneous transfer learning (Zhuang et al., 2021), this technique is often utilized when there is an insufficient dataset for the new task or when the task is complex, allowing for the efficient transfer of knowledge from one domain to another to improve learning performance.

In this study, a homogeneous transfer learning framework is adopted due to the strong similarity between the source and target tasks, both of which involve subcomponents of short-term wind speed forecasting. The source task (Task 1) involves training a Bi-GRU-ED model to forecast the residual component of the wind speed signal after decomposition. This training utilizes a multivariate dataset.

Once trained, the Bi-GRU-ED model is transferred to the target task, which focuses on forecasting the IMFs derived from CEEMDAN. Given the relatively limited data size associated with each IMF, fine-tuning is employed by retraining all layers of the pre-trained Bi-GRU-ED model. This approach offers several advantages: it eliminates the need to develop and train separate deep networks for each IMF, significantly reduces the computational burden, and provides a more stable and effective initialization of model parameters based on prior learning (Liang et al., 2021).

4. Experiment setup

To handle the complex and highly non-linear nature of wind speed signals, the **CEEMDAN** technique was applied to decompose the wind speed series $w[n]$ into a series of IMFs representing different frequency components and one residual component *Res* reflecting the overall trend. The CEEMDAN method involves several key parameters that directly influence the decomposition quality and downstream forecasting performance. In our study, the maximum number of extracted IMFs was empirically limited to six (N = 6), which provided a sufficient representation of the signal's main oscillatory modes while minimizing the extraction of high-frequency noise or redundant components. This decision was guided by both theoretical insights into mode decomposition and preliminary experiments evaluating forecasting accuracy. The ensemble size was set to 100, consistent with recommendations in the literature, to ensure decomposition stability without incurring excessive computational cost. The amplitude of the added white noise was chosen as 0.2 times the standard deviation of the original signal, which helps reduce mode mixing while preserving meaningful signal features. Cubic spline interpolation was used for envelope construction during the sifting process, as it provides smooth and accurate envelope fitting commonly adopted in CEEMDAN implementations. Lastly, the maximum number of sifting iterations was limited to 500 per mode to prevent over-decomposition and ensure convergence. These parameter choices strike a balance between decomposition quality and computational efficiency and were validated through preliminary trials on our dataset. This configuration results in six *IMFs* ($IMF_1, \dots, IMF_6$) and a residual *Res*, as illustrated in Fig. 5.

In parallel with decomposition, the other features temperature, pressure and wind speed at 10 m/s (W_{10}) are normalized (T_n, P_n, w_{10n}) by rescaling the variables to the range [0,1], as defined in Eq. (16). This preprocessing step prevents the models from exhibiting bias towards a specific range of values (Kiran and Vasumathi, 2020) and ensures an effective training process. Moreover, normalizing the dataset guarantees that each feature holds an equal influence on the overall model performance.

$$X_i^* = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{16}$$

After CEEMDAN decomposition and features normalisation, to forecast the residual component, a *Bi-GRU-ED*$_{Res}$ model is employed to forecast the residual component, which typically reflects the low-frequency trend. The input to this model is prepared using a sliding time window approach. Since forecasting performance is sensitive to the selected time step size, a trial-and-error procedure was conducted, and the optimal historical length was found to be 10. Accordingly, each input sample is constructed by combining the normalized features (T_n,P_n, w_{10n}) and the decomposed components across 10 consecutive time steps, forming a 10×10 input matrix $x^{(i)}$ where $i = 1, .., 10$, and each sequence corresponds to time steps $t-9, t-8, \dots, t$. This structured input format preserves short-term temporal dynamics while avoiding overfitting or feature scarcity. The resulting inputs are then used to forecast the residual at time $t+1$, as illustrated in Fig. 6. The same framework is subsequently extended via transfer learning to forecast each individual IMF component by changing the output target.

The prepared input data is first used to train the *Bi-GRU-ED*$_{Res}$ model to forecast the residual component. It is noteworthy that forecasting the residual first enables the encoder-decoder network to capture this crucial low-frequency trend and leverage it for accurate forecasting (Li and Xing, 2021). During training, the *Bi-GRU-ED*$_{Res}$ model learns the temporal dynamics inherent in the residual, and once trained, it is saved as a pre-trained model, retaining the learned weights and biases. This trained model serves as the foundation for a transfer learning strategy instead of initializing each IMF forecasting model from scratch, the pre-trained *Bi-GRU-ED*$_{Res}$ model is fine-tuned to forecast each intrinsic mode function (*Bi-GRU- ED*$_l$ $: l = 1, \dots, 6$). This approach accelerates training, reduces computational cost, and improves learning efficiency by reusing previously acquired temporal knowledge. Finally, the forecasted $\widehat{Res}$ and $\widehat{IMF}_l$ $(l = 1, \dots, 6)$ are aggregated to reconstruct the complete forecasted wind speed signal $\widehat{w}[n]$, yielding the final short-term forecast. Fig. 7 illustrates the suggested CEEMDAN-Bi-GRU-ED-TL forecasting model architecture.

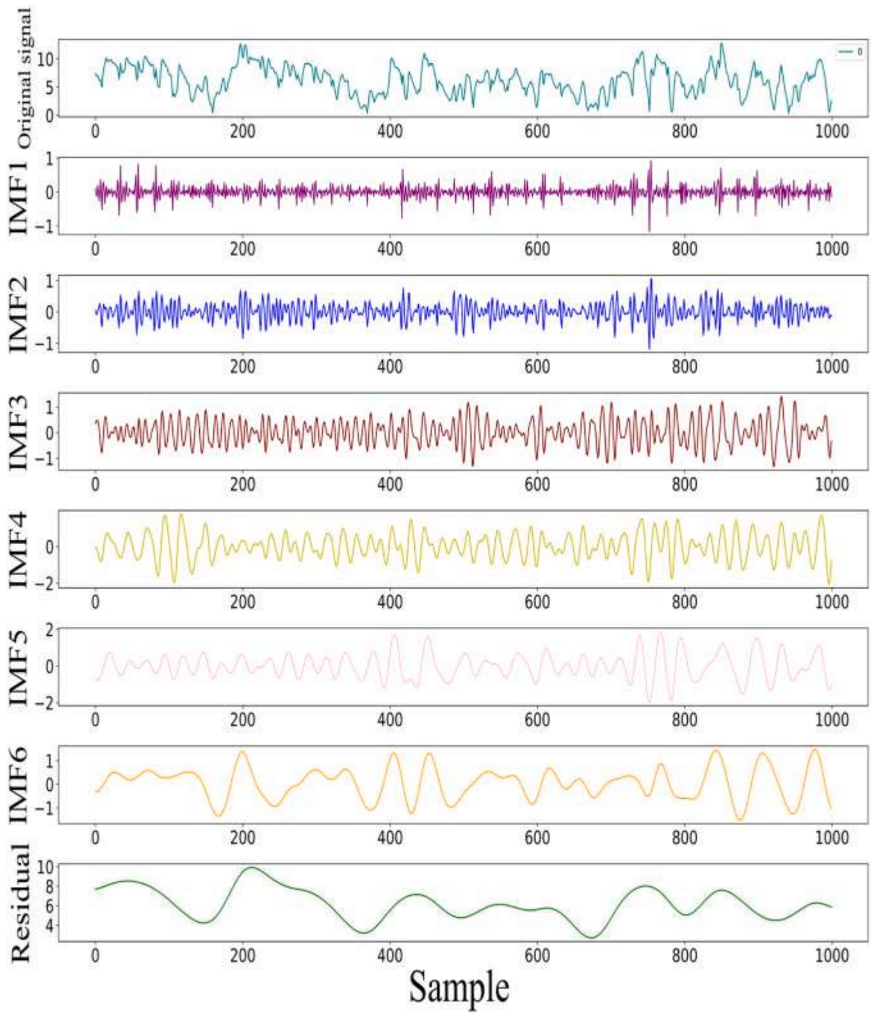

Fig. 5. CEEMDAN decomposition for wind speed in Adrar dataset.

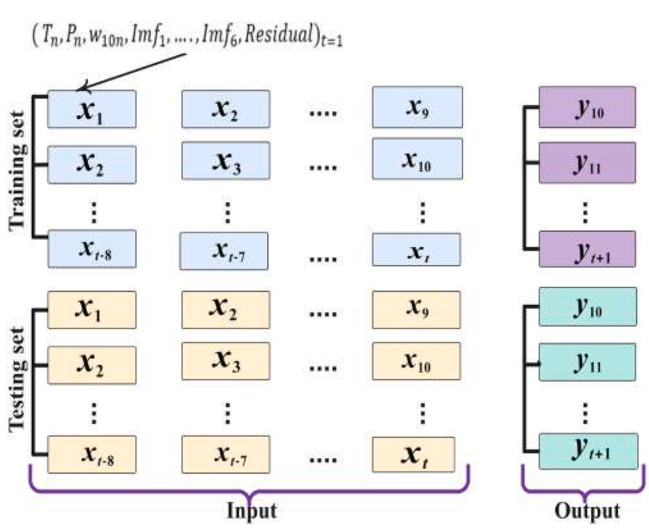

Fig. 6. Training and testing dataset structure.

4.1. Datasets splitting

Assessing models on time series datasets presents unique challenges due to their non-linear variability and intrinsic temporal dependencies. To overcome these challenges and improve model performance, it is crucial to implement proper dataset splitting techniques. Such methods ensure greater effectiveness in mitigating overfitting concerns and enhance the reliability of the model evaluation process (Hasanov et al., 2022). Table 2 presents statistical details regarding the distribution of training and testing samples, where 80 % of the data is allocated for training and the remaining 20 % is reserved for testing. Additionally, k-fold cross-validation with k = 8 splits is employed during the evaluation process.

4.2. Model hyperparameters

In general, choosing the hyperparameters of machine learning models manually is challenging and demands a substantial understanding of the problem at hand. In this study, the GridSearchCV tool is utilized for hyperparameter tuning. This tool is widely used in machine learning to determine the optimal combination of hyperparameters for a given model by exhaustively evaluating all provided values for each hyperparameter (Pu, n.d.). Table 3 presents the hyperparameters employed in the proposed model.

4.3. Evaluation metrics

When evaluating the efficacy of the proposed wind speed forecasting model, various error metrics are utilized, including the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2):

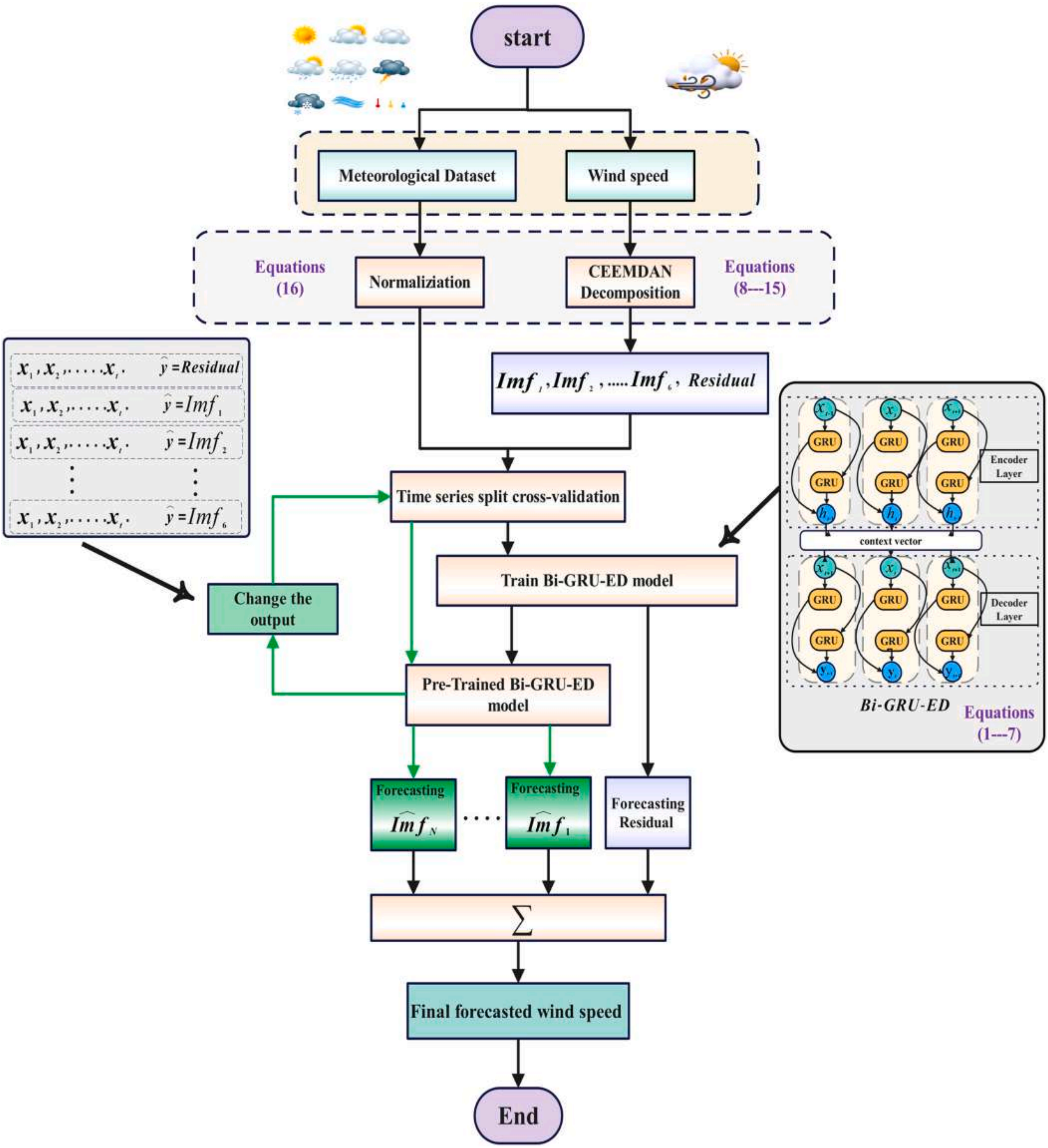

Fig. 7. Flowchart of the proposed forecasting architecture.

Table 2
Datasets statistical details.

Sites		Number of samples	Statistical indicator			
			Max (m/s)	Min (m/s)	Mean (m/s)	Std (m/s)
Site 1 (Adrar)	All samples	70128	18.37	0.6	6.64	2.65
	Training	58432	18.37	1.5	6.65	2.82
	Testing	11686	16.05	0.6	5.84	2.26
Site 2 (Tataouine)	All samples	61343	19.83	0.4	6.02	3.73
	Training	51111	18.50	1	7.18	4.22
	Testing	10222	19.18	0.4	6.23	3.79

Table 3
hyper-parameters of the *Bi-GRU-ED$_{Res}$* model.

Hyperparameter	value
number of input variables	10
time steps of input variables	10
Epochs	250
Optimizer	Adam
Batch size	260
Epochs	250
Learning rate	0.001
Layers	3
activation function	(relu-relu-linear)

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2} \tag{17}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t| \tag{18}$$

$$MAPE = \frac{100\%}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{19}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{N}(y_t - \overline{y}_t)^2} \tag{20}$$

Where y_t and $\hat{y}_t$ represent the actual and the forecasted values respectively.

5. Results and discussion

Employing the two datasets, the effectiveness of the proposed scheme, CEEMDAN-Bi-GRU-ED-TL, is evaluated from four perspectives:

1. The Bi-GRU-ED model will be compared with eight different wind speed forecasting models for one time-step forecasting. These models include SVR, RNN, CNN, LSTM, GRU, Bi-LSTM, Bi-GRU, and CNN-LSTM. All models will be trained under uniform conditions, without incorporating CEEMDAN decomposition, specifically focusing on one-hour-ahead forecasting.
2. The CEEMDAN-Bi-GRU-ED-TL model will be compared with the top-performing models that incorporate EEMD and CEEMDAN decomposition techniques.
3. A comparison will be made between the forecasted IMFs of the CEEMDAN-Bi-GRU-ED scheme with and without transfer learning.
4. Various forecasting time steps will be evaluated for the Bi-GRU-ED model, the CEEMDAN-Bi-GRU-ED model without transfer learning, and the proposed CEEMDAN-Bi-GRU-ED-TL model with transfer learning.

5.1. Models without decompositions

The forecasting results depicted in Fig. 9 suggest that the trends of the eight models generally align closely with the actual values. Notably, the GRU, LSTM, Bi-GRU, Bi-LSTM, CNN-LSTM, and Bi-GRU-ED models exhibit minimal discrepancies compared to the actual dataset values. Conversely, the SVR, CNN, and RNN models display notable deviations at specific points, as observed in the zoomed-in plots, indicating larger forecasting errors. The significant error observed in the CNN model could stem from its inability to maintain spatial invariance in the input dataset. The RNN model encounters challenges related to vanishing and exploding gradients, which affect its capability to handle long sequences effectively. Additionally, the SVR model faces limitations in wind speed forecasting due to its nonlinear nature. Notably, the zoomed-in plots demonstrate that the forecasting curve of the Bi-GRU-ED model closely tracks the actual values, underscoring its superior predictive performance. As shown in Table 4 and Fig. 8, the Bi-GRU-ED model achieves the best results among all baseline models, with the lowest MAE values of 0.4161 and 0.4426, RMSE values of 0.6451 and 0.6653, and MAPE values of 10.27 % and 11.13 % for Dataset 1 and Dataset 2, respectively. It also records the highest R^2 values of 0.9268 and 0.9278, indicating strong forecasting capability and robustness to fluctuations. In contrast, the SVM model yields the weakest performance, with MAE values of 0.6573 and 0.6815, RMSE values of 0.8107 and 0.8255, MAPE values of 18.79 % and 21.24 %, and the lowest R^2 values of 0.8681 and 0.8577. The remaining models, including RNN, CNN, LSTM, GRU, Bi-LSTM, Bi-GRU, and CNN-LSTM, fall between these two extremes. Among them,

Table 4
Comparison of Bi-GRU-ED model performances with the benchmark models without decomposition.

Models	Dataset 1 MAE	RMSE	MAPE	R^2	Dataset 2 MAE	RMSE	MAPE	R^2
SVR	0.6573	0.8107	18.79	0.8681	0.6815	0.8255	21.24	0.8577
RNN	0.5702	0.7551	14.87	0.8704	0.5933	0.7703	15.6	0.8771
CNN	0.5399	0.7347	12.59	0.8906	0.5748	0.7581	14.46	0.8731
LSTM	0.4997	0.7069	11.25	0.9081	0.5654	0.7523	13.68	0.8994
GRU	0.4872	0.698	10.88	0.9126	0.5455	0.7386	12.97	0.9094
Bi-LSTM	0.4579	0.6764	10.59	0.9187	0.486	0.6972	12.53	0.9113
Bi-GRU	0.4469	0.6685	10.54	0.9208	0.473	0.6877	11.98	0.9184
CNN-LSTM	0.4346	0.6592	10.39	0.9293	0.4544	0.6741	11.47	0.9206
Bi-GRU-ED	**0.4161**	**0.6451**	**10.27**	**0.9268**	**0.4426**	**0.6653**	**11.13**	**0.9278**

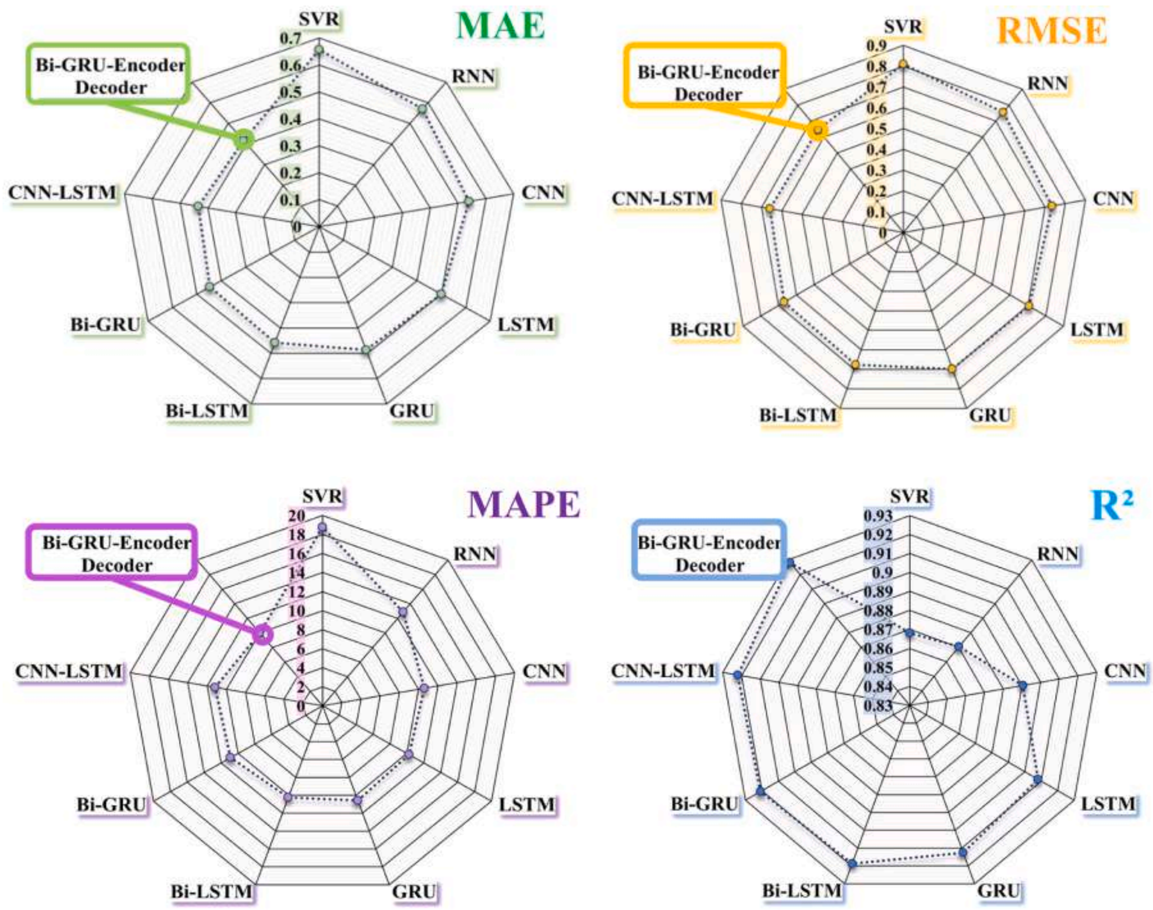

Fig. 8. Bi-GRU-ED model performances compared to other benchmark models.

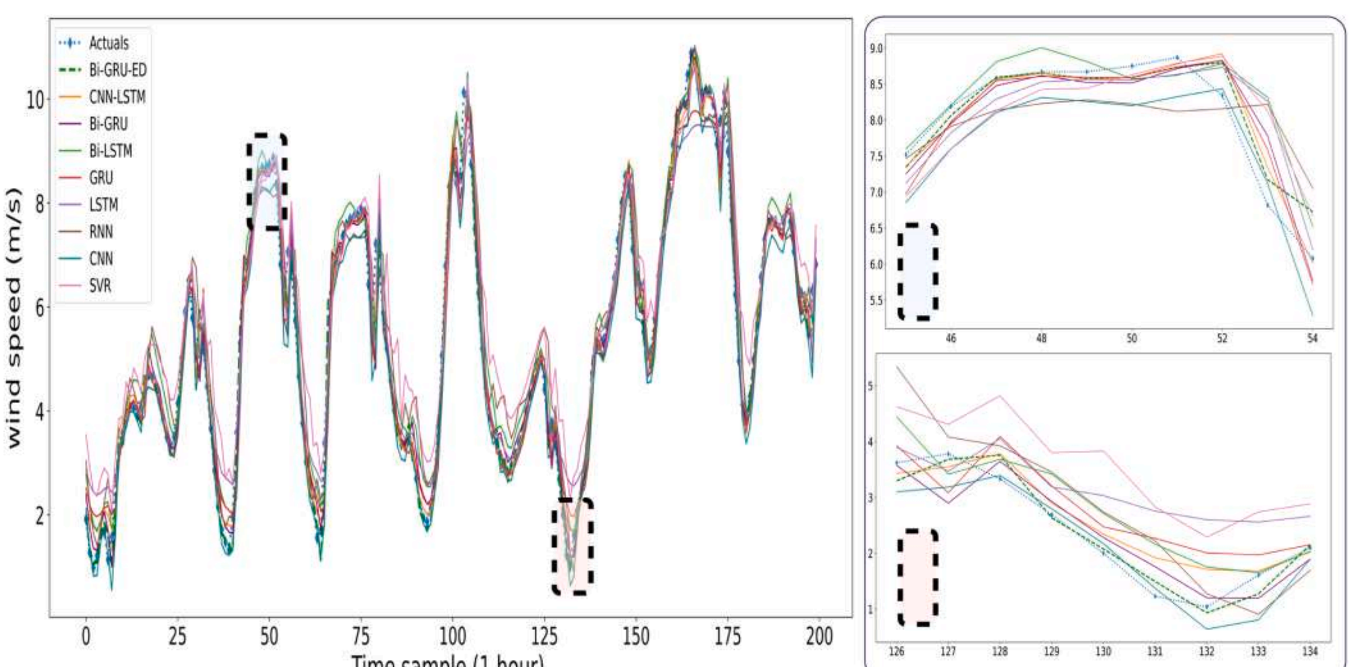

Fig. 9. wind speed forecasting results without decomposition for dataset 1.

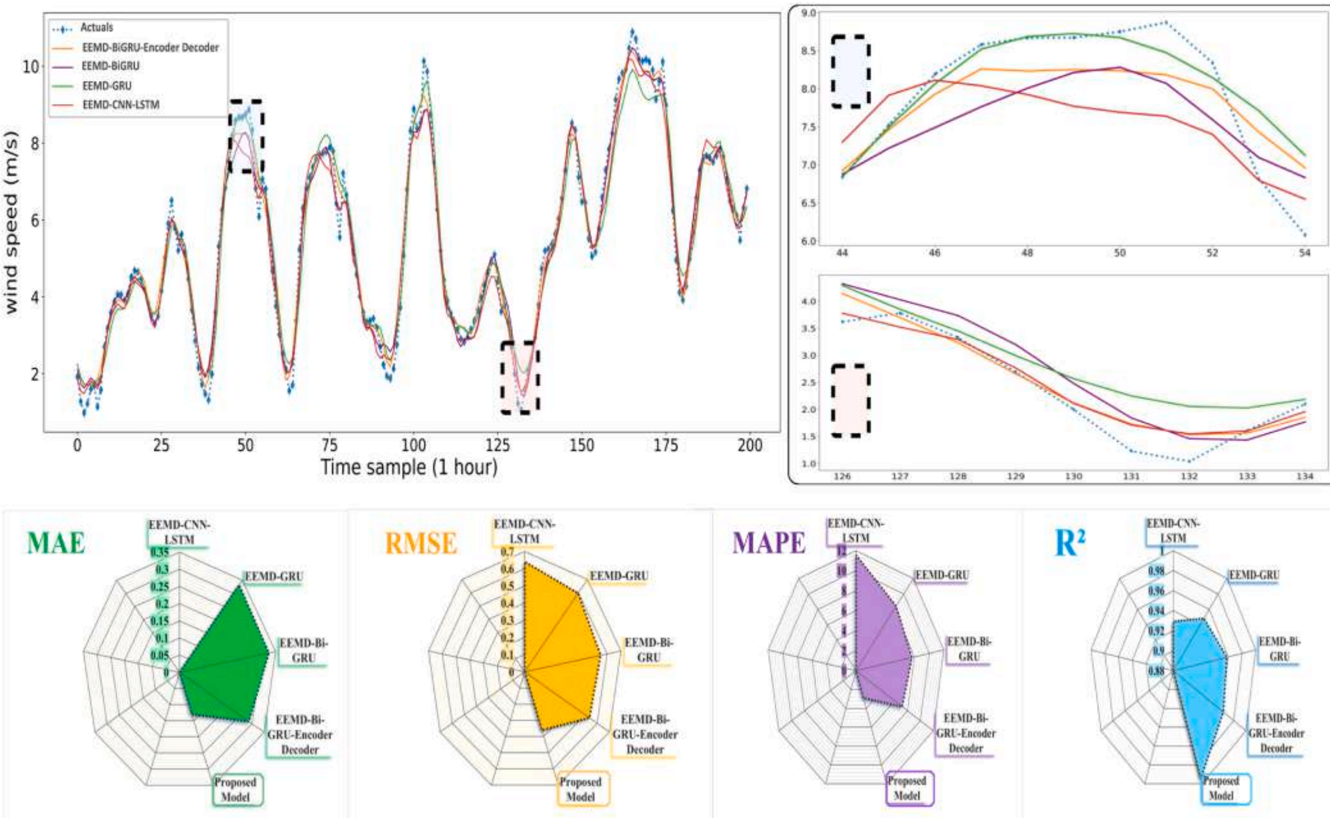

Fig. 10. Forecasted wind speed results utilizing EEMD decomposition for dataset 1.

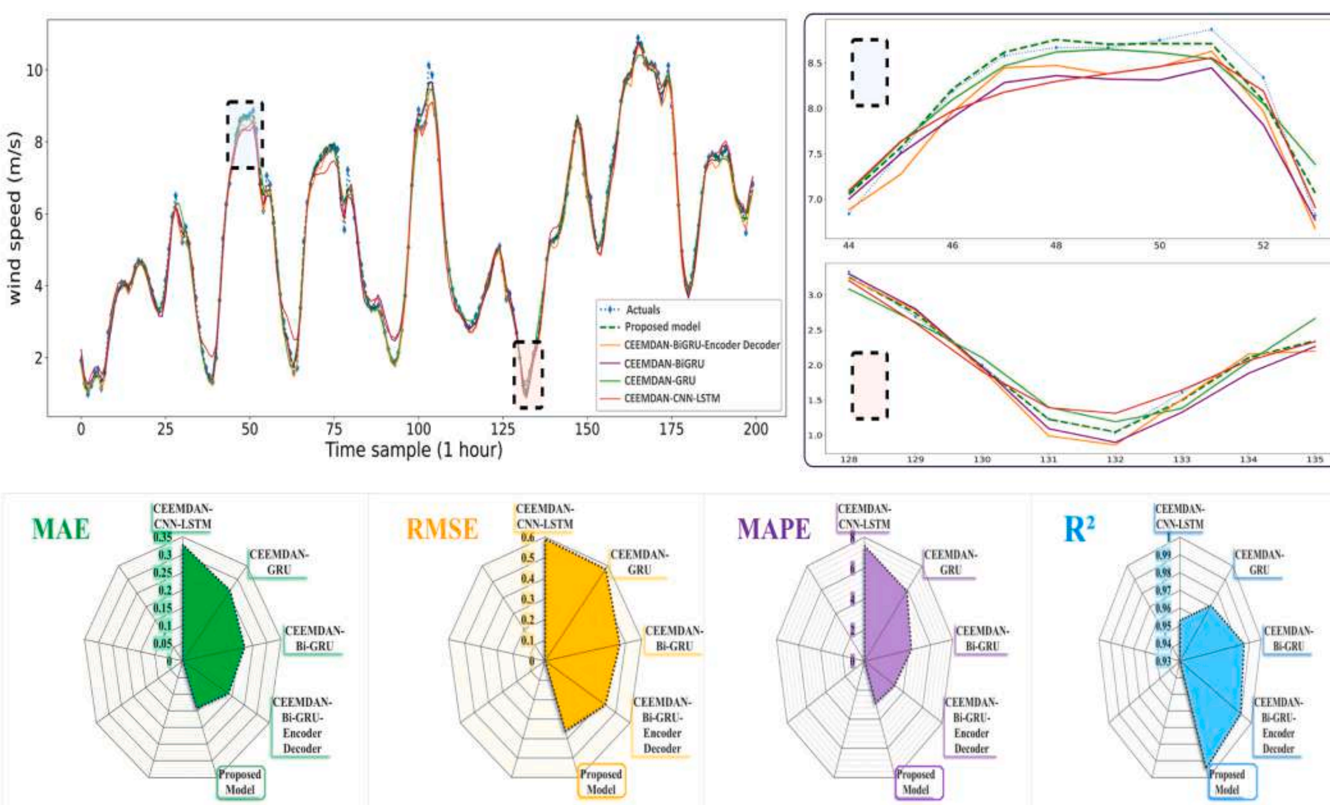

Fig. 11. Forecasting results using the CEEMDAN decomposition for dataset 1.

Table 5
Forecasting performance comparison: proposed model versus EEMD and CEEMDAN-based models.

	Dataset 1				Dataset 2			
Models	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE	R^2
EEMD-CNN-LSTM	0. 3987	0.6308	10.16	0.9221	0.3854	0.6137	9.98	0.9395
EEMD-GRU	0.3328	0.5954	8.44	0.9481	0.3494	0.5824	8.78	0.9459
EEMD-Bi-GRU	0.3254	0.5526	7.64	0.958	0.3201	0.5363	7.823	0.9587
EEMD-Bi-GRU-ED	0.2873	0.536	7.17	0.9627	0.2741	0.5123	7.208	0.9698
CEEMDAN-CNN-LSTM	0.3281	0.5902	7.43	0.9529	0.3041	0.5423	7.122	0.9607
CEEMDAN-GRU	0.2581	0.5787	5.88	0.9709	0.2311	0.4806	4.598	0.9792
CEEMDAN-Bi-GRU	0.2202	0.4653	4.25	0.9852	0.2	0.4472	4.001	0.9894
CEEMDAN-Bi-GRU-ED	0.1844	0.4255	3.11	0.9896	0.1765	0.4125	2.991	0.9912
CEEMDAN-Bi-GRU-ED-TL	**0.142**	**0.3612**	**2.90**	**0.9947**	**0.1305**	**0.3549**	**2.832**	**0.9955**

the bidirectional architectures Bi-LSTM, Bi-GRU, and Bi-GRU-ED as well as the CNN-LSTM hybrid, consistently deliver improved accuracy over their unidirectional counterparts, highlighting the benefit of incorporating both past and future temporal information.

5.2. Models with decompositions

The performance of the leading models, which did not involve decomposition, is improved with the incorporation of EEMD decomposition Fig. 10 and CEEMDAN decomposition Fig. 11. The analysis showed that the proposed model exhibited lower error rates across both datasets, followed by the CEEMDAN-Bi-GRU-ED according to the local zoomed-in plots depicted in Fig. 10 and Fig. 11 as they approach the line of actual values compared to the other models as indicated in Table 5. Notably, the proposed model showed remarkable advancements by attaining the lowest values of MAE, RMSE, and MAPE among all evaluated models. Specifically, it achieved 0.142, 0.3612, and 2.9 % for Dataset 1, and 0.1305, 0.3549, and 2.832 % for Dataset 2, respectively.

Furthermore, it reveals the highest R^2 values, 0.9947 and 0.9955, for both datasets, suggesting minimal errors and strong capability in managing unexpected wind speed variations. In addition, the CEEMDAN-Bi-GRU-ED model displayed the second-best performance in terms of MAE, RMSE, and MAPE, registering 0.1844, 0.4255, and 3.1 % for Dataset 1, and 0.1765, 0.4125, and 2.991 % for Dataset 2. However, it also achieved high R^2 values of 0.9896 and 0.9912 for Dataset 1 and Dataset 2, respectively. These results indicate that the proposed framework offers the most accurate forecasting, likely due to the transfer learning mechanism, which involves updating the internal state following the forecasting of each IMF and leveraging the learned weights and biases from the pre-trained model.

The analysis of the forecasting results presented in Table 4 indicates that models without decomposition achieve better results on dataset 1 compared to dataset 2. This difference could be attributed to the larger sample size of dataset 1, which typically enhances the forecasting accuracy (Wei et al., 2017). Conversely, upon applying decomposition techniques as shown in Table 5, the trend is reversed, which might be due to the fact that EEMD and CEEMDAN decompositions break down signals into IMFs an approach particularly advantageous for smaller datasets. This decomposition enables a more detailed and nuanced analysis, facilitating a clearer understanding of underlying trends and anomalies within the data, and improving model training. Consequently, it effectively mitigates the impact of random fluctuations in the dataset, leading to enhanced forecasting accuracy.

Additionally, as depicted in Table 5, the EEMD-CNN-LSTM model demonstrates the least satisfactory performance, characterized by the highest error rates: MAE values of 0.3987 and 0.3854, RMSE values of 0.6308 and 0.6137, MAPE values of 10.16 % and 9.98 %, and R^2 values of 0.9221 and 0.9395 for Dataset 1 and Dataset 2, respectively. It is apparent that models enriched with CEEMDAN generally outperform those employing EEMD, owing to CEEMDAN capacity to manage nonlinear and non-stationary data and effectively distinguish noise from the original wind speed signal (Sun and Wang, 2023). It is also worth noting that bidirectional models (EEMD-Bi-GRU, EEMD-Bi-GRU-ED, CEEMDAN-Bi-GRU, CEEMDAN-Bi-GRU-ED) perform better than their unidirectional counterparts, indicating that considering both past and future context boosts forecasting accuracy.

5.3. Transfer learning impact on forecasted decomposition components

Table 6 presents a detailed comparison of forecasting performance between the CEEMDAN-Bi-GRU-ED model with and without transfer learning, across both datasets. The results clearly show that integrating transfer learning significantly improves accuracy in most decomposition components. Specifically, reductions in MAE and RMSE are observed for the majority of IMFs, indicating enhanced learning efficiency and generalization. Notably, for Dataset 1, IMF2 and IMF3 show substantial MAE reductions from 0.1633 to 0.0453 and from 0.0199 to 0.0131, respectively. Similar trends are seen in Dataset 2, where IMF2 drops from 0.1578 to 0.0414 in MAE and from 0.3972 to 0.2028 in RMSE. These improvements validate the effectiveness of transfer learning in forecasting subcomponents of the wind speed signal, particularly those with lower frequency and smoother patterns. However, a slight degradation is observed in IMF1, where MAE increased from 0.1021 to 0.1211 and RMSE from 0.32 to 0.348 in Dataset 1. In Dataset 2, MAE improved (from 0.1224 to 0.0802), but RMSE increased slightly. This variability can be attributed to IMF1's representation of high-frequency

Table 6
Impact of transfer learning on forecasted IMFs and residual components for both datasets.

Component	Dataset 1				Dataset 2			
	CEEMDAN-Bi-GRU-ED without transfer learning		CEEMDAN-Bi-GRU-ED with transfer learning (Proposed model)		CEEMDAN-Bi-GRU-ED without transfer learning		CEEMDAN-Bi-GRU-ED with transfer learning (Proposed model)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Residual	0.0828	0.1665	0.0828	0.1665	0.0493	0.1403	0.0493	0.1403
IMF1	0.1021	0.32	0.1211	0.348	0.1224	0.2827	**0.0802**	0.3001
IMF2	0.1633	0.4041	**0.0453**	**0.2129**	0.1578	0.3972	**0.0414**	**0.2028**
IMF3	0.0199	0.1412	**0.0131**	**0.1386**	0.0192	0.1292	**0.0167**	**0.1246**
IMF4	0.0304	0.1678	**0.0281**	**0.1005**	0.187	0.1358	**0.0201**	**0.0856**
IMF5	0.0098	0.082	**0.0067**	**0.07761**	0.014	0.1184	**0.0127**	**0.0931**
IMF6	0.0094	0.0912	**0.0087**	**0.0892**	0.0098	0.1214	**0.0086**	**0.0902**

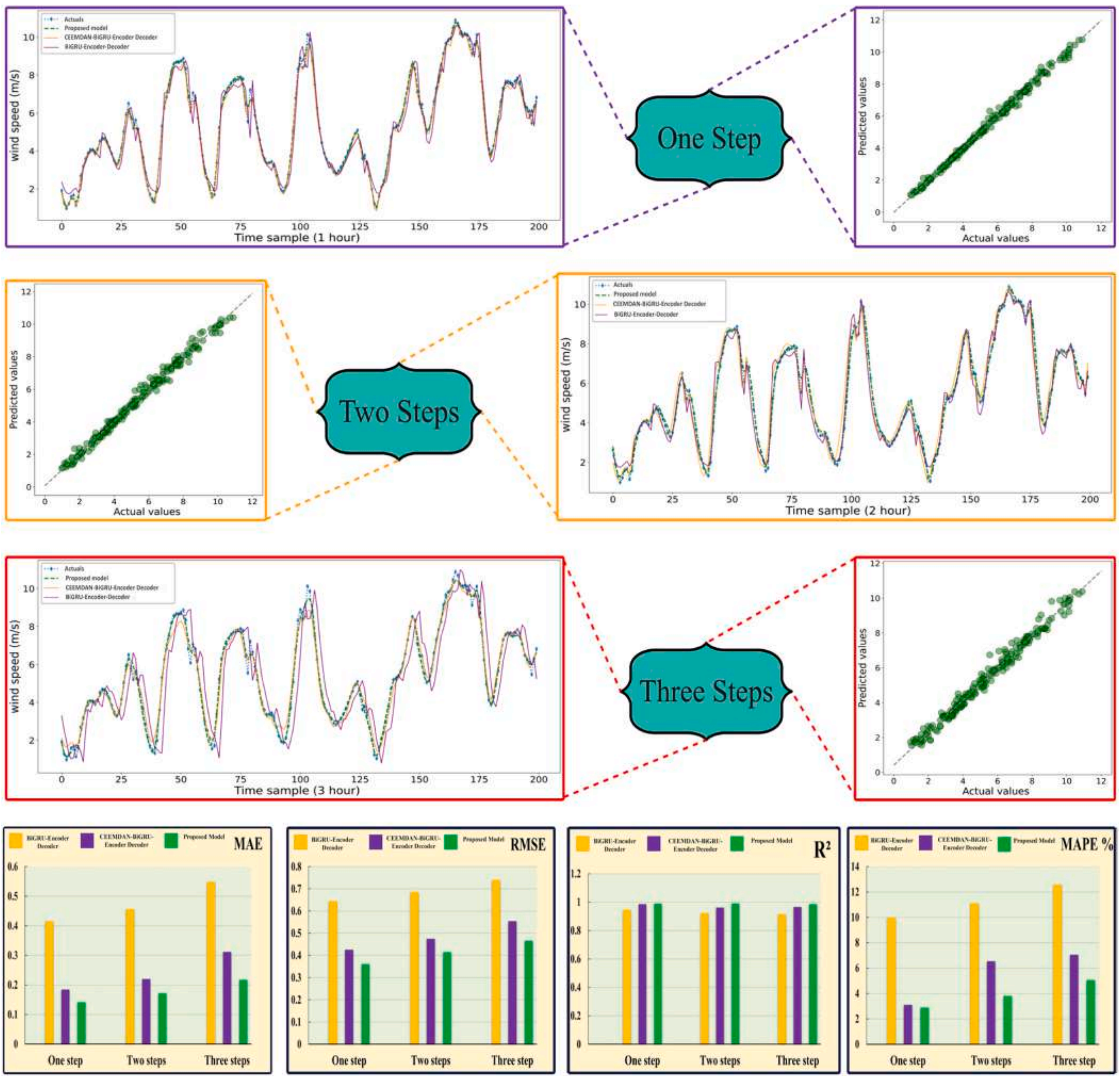

Fig. 12. Comparison of different timestep forecasts.

fluctuations, which are more prone to noise sensitivity and irregularity, reducing the benefits of transfer learning in this specific case. Overall, these results demonstrate that transfer learning contributes substantially to the proposed framework's effectiveness, offering a scalable strategy for improving accuracy while reducing the need to train multiple deep models from scratch. This validates transfer learning not only as an efficiency booster but also as a key architectural contribution to enhancing wind speed forecasting at the decomposition level.

5.4. Comparison of the forecasting models for different timesteps

Fig. 12 and Table 7 present the results obtained from forecasting at different timesteps using the Bi-GRU-ED model without decomposition, followed by its integration with CEEMDAN, and finally, the proposed model. For a one-hour forecasting timestep, the proposed model demonstrates superior performance, exhibiting minimal errors compared to other approaches. However, as the forecasting horizon extends to two hours ahead, a slight increase in errors is observed in the CEEMDAN-Bi-GRU-ED and the proposed model forecasts compared to the Bi-GRU-ED model. Specifically, for Dataset 1, the proposed model achieves an MAE of 0.1822, RMSE of 0.4049, MAPE of 3.808 %, and R^2 of 0.9894. For Dataset 2, the corresponding values are MAE of 0.1658, RMSE of 0.3964, MAPE of 3.658 %, and R^2 of 0.9915.

Additionally, when considering a forecasting timestep of 3, the slight decrease in performance observed in the CEEMDAN-Bi-GRU-ED model compared to the timestep = 2 results highlights the influence of each technique on model performance enhancement. It is important to note that forecast accuracy may diminish as the forecasting horizon extends, which could be attributed to the increasing uncertainty and variability of wind speed conditions over time. Moreover, the forecasting results of the developed model underscore the effectiveness of the decomposition method and transfer learning in substantially enhancing the model

Table 7
Comparative analysis of forecasting performance at various timesteps: proposed model versus models without transfer learning and CEEMDAN decomposition.

		Dataset 1				Dataset 2			
steps	models	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE	R^2
One step (t + 1)	Bi-GRU-ED	0.4161	0.6451	10.27	0.9268	0.4426	0.6653	11.13	0.9278
	CEEMDAN-Bi-GRU-ED	0.1844	0.4255	3.1	0.9896	0.1765	0.4125	2.991	0.9912
	Proposed model	**0.142**	**0.3612**	**2.9**	**0.9947**	**0.1305**	**0.3549**	**2.832**	**0.9955**
Two steps (t + 2)	Bi-GRU-ED	0.4666	0.6857	10.65	0.9122	0.4802	0.6897	11.95	0.9068
	CEEMDAN-Bi-GRU-ED	0.2281	0.4749	5.15	0.9713	0.2201	0.4671	4.369	0.9848
	Proposed model	**0.1822**	**0.4049**	**3.808**	**0.9894**	**0.1658**	**0.3964**	**3.658**	**0.9915**
Three steps (t + 3)	Bi-GRU-ED	0.5587	0.7408	15.01	0.8853	0.5402	0.7417	13.6	0.8756
	CEEMDAN-Bi-GRU-ED	0.312	0.5546	8.58	0.9606	0.2945	0.5422	6.47	0.9619
	Proposed model	**0.2478**	**0.4767**	**6.678**	**0.9712**	**0.2195**	**0.4572**	**5.09**	**0.9803**

capacity to forecast across diverse forecasting horizons, thereby mitigating the impact of increasing errors over longer time spans.

6. Conclusion

Accurate short-term wind speed forecasting is critical for renewable energy integration but remains a complex task due to the non-linearity, non-stationarity, and limited availability of wind data. This study introduces a novel hybrid deep learning framework "CEEMDAN-Bi-GRU-ED-TL" that strategically combines **(i)** Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) for multi-scale signal decomposition, **(ii)** a Bidirectional GRU-based Encoder–Decoder (Bi-GRU-ED) for temporal sequence modeling, and **(iii)** a homogeneous transfer learning strategy that reuses the residual forecasting model to enhance the forecasting of each IMF.

The main contribution of this work lies in the innovative integration of signal decomposition, deep recurrent architectures, and transfer learning into a unified forecasting pipeline. Unlike conventional approaches that require training separate models for each decomposed component, the proposed framework efficiently reuses and fine-tunes a pre-trained residual model across all IMFs, thereby reducing computational complexity and enhancing generalization. Experimental results on two real-world datasets confirm its effectiveness. Among non-decomposed baselines, the Bi-GRU-ED model achieved the best performance. Incorporating CEEMDAN further improved accuracy, reducing MAE from 0.4161 to 0.1844 and RMSE from 0.6451 to 0.4255 on Dataset 1. With the addition of transfer learning, the proposed model achieved further gains reaching an MAE of 0.142 and RMSE of 0.3612, while also lowering MAPE to 2.9 % and increasing R^2 to 0.9947. Similar improvements were observed on Dataset 2. These results underscore the effectiveness of the proposed scheme.

This framework is among the first to combine CEEMDAN, Bi-GRU-ED, and transfer learning, providing an effective solution to common challenges in wind forecasting.

Future research will explore:

- Integration of attention mechanisms to enhance temporal interpretability.
- Extension to multi-step and probabilistic forecasting.
- Deployment in real-time smart grid environments using edge computing.
- Application to other renewable energy sources.

CRediT authorship contribution statement

Abdelhakim Saim: Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis. **khelil khaled:** Writing – review & editing, Resources, Data curation. **Farid Berrezzek:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Lokmene Melalkia:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Radouane Nebili:** Writing – review & editing.

Ethical approval

This research did not require ethical approval.

Consent to participate

Informed consent was not required for this study.

Consent to publish

All authors have read and agreed to the manuscript being submitted for publication.

Funding

No funding has been received.

Declaration of Competing Interest

The authors declare that they have no conflict interests regarding the publication of this paper.

Acknowledgements

This work is supported by the Directorate General of Scientific Research and Technological Development (DGRSDT), Algeria. It has also been partially supported by the National Research Agency (ANR), LEAP RE "MiDiNA—Microgrids Development in North Africa" project, grant number "ANR-23-LERE-0002-01"

Data availability

Data will be made available on request.

References

Cadenas, E., Rivera, W., Campos-Amezcua, R., Cadenas, R., 2016. Wind speed forecasting using the NARX model, case: La Mata, Oaxaca, México. Neural Comput. Appl. 27, 2417–2428.

Chen, J., Sun, H., Xu, B., 2019. Improvement of empirical mode decomposition based on correlation analysis. SN Appl. Sci. 1, 1–16.
Chen, X.J., Han, T., Cheng, P., Da, X., 2025. A novel interval prediction method in wind speed based on deep learning and combination prediction. Sci. Rep. 15, 1–16.
Erdem, E., Shi, J., She, Y., 2014. Comparison of two ARMA-GARCH approaches for forecasting the mean and volatility of wind speed. Springer Proc. Phys. 155, 65–73.
Garg, S., Krishnamurthi, R., 2023. A CNN encoder decoder LSTM model for sustainable wind power predictive analytics. Sustain. Comput. Inform. Syst. 38, 100869.
Hasanov, M., Wolter, M., Glende, E., 2022. Time series data splitting for Short-Term load forecasting. PESS + PELSS 2022; Power Energy Student Summit 1–6.
Houndekindo, F., Ouarda, T.B.M.J., 2025. LSTM and Transformer-based framework for bias correction of ERA5 hourly wind speeds. Energy 328, 136498.
Ji, L., Fu, C., Ju, Z., Shi, Y., Wu, S., Tao, L., 2022. Short-Term canyon wind speed prediction based on CNN—GRU transfer learning. Atmosphere (Basel) 13, 1–15.
Joseph, L.P., Deo, R.C., Casillas-Perez, D., Prasad, R., Raj, N., Salcedo-Sanz, S., 2024. Multi-Step-Ahead wind speed forecast system: hybrid multivariate decomposition and feature Selection-Based gated additive tree ensemble model. IEEE Access 12, 58750–58777.
Khelil, K., Berrezzek, F., Bouadjila, T., 2021. GA-based design of optimal discrete wavelet filters for efficient wind speed forecasting. Neural Comput. Appl. 33, 4373–4386.
Khodayar, M., Kaynak, O., Khodayar, M.E., 2017. Rough deep neural architecture for Short-Term wind speed forecasting. IEEE Trans. Ind. Inform. 13, 2770–2779.
Kiran, A., Vasumathi, D., 2020. Data mining: Min–max normalization based data perturbation technique for privacy preservation. Advances in Intelligent Systems and Computing. Springer Singapore.
Li, C., Xing, W., 2021. Natural language generation using deep learning to support MOOC learners. Int. J. Artif. Intell. Educ. 31, 186–214.
Li, C., Tang, G., Xue, X., Saeed, A., Hu, X., 2020. Short-Term wind speed interval prediction based on ensemble GRU model. IEEE Trans. Sustain. Energy 11, 1370–1380.
Li, Z., Ye, L., Zhao, Y., Song, X., Teng, J., Jin, J., 2016. Short-term wind power prediction based on extreme learning machine with error correction. Prot. Control Mod. Power Syst. 1, 4–11.
Liang, T., Zhao, Q., Lv, Q., Sun, H., 2021. A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers. Energy 230, 120904.
Liu, M., De, Ding, L., Bai, Y.L., 2021. Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. Energy Convers. Manag. 233, 113917.
Luo, T., Tang, Z., Liu, J., Zhou, B., 2023. A review of transfer learning approaches for load, solar and wind power predictions. Proc. 2023 Panda Forum Power Energy, PandaFPE 2023, 1580–1584.
Melalkia, L., Berrezzek, F., 2024. Comparative Meteorological Datasets for Algeria and Tunisia: Key Data for Wind Speed and Power Prediction. 4TU.ResearchData.
Parmezan, A.R.S., Souza, V.M.A., Batista, G.E.A.P.A., 2019. Evaluation of statistical and machine learning models for time series prediction: identifying the state-of-the-art and the best conditions for the use of each model. Inf. Sci. (Ny) 484, 302–337.
Pu, T., n.d. GridSearchCV Method for TinyML.
Rai, A., Shrivastava, A., Jana, K.C., 2022. A robust auto encoder-gated recurrent unit (AE-GRU) based deep learning approach for short term solar power forecasting. Optik (Stuttg) 252, 168515.
Shahid, F., Zameer, A., Iqbal, M.J., 2021. Intelligent forecast engine for short-term wind speed prediction based on stacked long short-term memory. Neural Comput. Appl. 33, 13767–13783.
Shen, Z., Fan, X., Zhang, L., Yu, H., 2022. Wind speed prediction of unmanned sailboat based on CNN and LSTM hybrid neural network. Ocean Eng. 254, 111352.
Sun, H., Cui, Q., Wen, J., Kou, L., Ke, W., 2024. Short-term wind power prediction method based on CEEMDAN-GWO-Bi-LSTM. Energy Rep. 11, 1487–1502.
Sun, P., Wang, J., Yan, Z., 2024. Ultra-short-term wind speed prediction based on TCN-MCM-EKF. Energy Rep. 11, 2127–2140.
Sun, W., Wang, X., 2023. Improved chimpanzee algorithm based on CEEMDAN combination to optimize ELM short-term wind speed prediction. Environ. Sci. Pollut. Res. 30, 35115–35126.
Torres, M.E., Colominas, M.A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise. ICASSP IEEE Int. Conf. Acoust. Speech Signal Process. Proc. 4144–4147.
Trebing, K., Mehrkanoon, S., 2020. Wind speed prediction using multidimensional convolutional neural networks. 2020 IEEE Symp. Ser. Comput. Intell. 2020, 713–720 (SSCI).
Waheeb, W., Ghazali, R., Shah, H., 2019. Nonlinear autoregressive moving-average (NARMA) time series forecasting using neural networks. 2019 Int. Conf. Comput. Inf. Sci. ICCIS 2019, 1–5.
Wei, Y., Wang, H., Shi, S., Gao, H., Li, J., 2017. Any-Time Methods for Time-Series Prediction with Missing Observations. Proc. - 2017 IEEE 6th Int. Congr. Big Data, BigData Congr. 2017 427–430.
Wu, B., Wang, L., Zeng, Y.R., 2022. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. Energy 252, 123990.
Wu, H., Meng, K., Fan, D., Zhang, Z., Liu, Q., 2022. Multistep short-term wind speed forecasting using transformer. Energy 261, 125231.
Yan, Y., Wang, X., Ren, F., Shao, Z., Tian, C., 2022. Wind speed prediction using a hybrid model of EEMD and LSTM considering seasonal features. Energy Rep. 8, 8965–8980.
Yang, Q., Lin, Y., Kuang, S., Wang, D., 2024. A novel short-term load forecasting approach for data-poor areas based on K-MIFS-XGBoost and transfer-learning. Electr. Power Syst. Res. 229, 110151.
Yang, S., Yuan, A., Yu, Z., 2023. A novel model based on CEEMDAN, IWOA, and LSTM for ultra-short-term wind power forecasting. Environ. Sci. Pollut. Res. 30, 11689–11705.
Zhang, C., Wei, H., Zhao, J., Liu, T., Zhu, T., Zhang, K., 2016. Short-term wind speed forecasting using empirical mode decomposition and feature selection. Renew. Energy 96, 727–737.
Zhang, X., Li, J., Liu, X., Chen, Z., 2020. Improved EEMD-based standardization method for developing long tree-ring chronologies. J. For. Res. 31, 2217–2224.
Zhang, X., Zhao, H., Yao, J., Wang, Z., Zheng, Y., Peng, T., Zhang, C., 2025. A multi-scale component feature learning framework based on CNN-BiGRU and online sequential regularized extreme learning machine for wind speed prediction. Renew. Energy 242, 122427.
Zhao, Z., Yun, S., Jia, L., Guo, J., Meng, Y., He, N., Li, X., Shi, J., Yang, L., 2023. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. Eng. Appl. Artif. Intell. 121, 105982.
Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2021. A comprehensive survey on transfer learning. Proc. IEEE 109, 43–76.