Earth Systems and Environment
https://doi.org/10.1007/s41748-025-00714-y

ORIGINAL ARTICLE

A Comparative Study on Novel Hybrid Approaches Based on CEEMDAN, Random Forest, Deep Learning Methods for Predicting Daily Wind Speed

Amin Gharehbaghi[1] · Redvan Ghasemlounia[2] · Farshad Ahmadi[3] · Rasoul Mirabbasi[4] · Ali Torabi Haghighi[5]

Received: 12 April 2025 / Revised: 24 May 2025 / Accepted: 27 June 2025

Abstract

In this study, different kinds of hybrid Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) algorithms with forecasting models including Random Forest (RF), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) neural networks, are developed to estimate the mean daily wind speed at the height of 2 m in Ağrı city (WS_{st12}), Turkey. In these hybrid models, different layer networks of single and integrated LSTM and GRU models include general single LSTM, general single GRU, simple coupled LSTM-GRU, and novel coupled LSTM with GRU through Addition layer (i.e., LSTM + GRU model) structures are applied. The most effective parameters on the WS_{st12}, from a list of on-site potential meteorological parameters and wind speed values in its adjacent cities of Ağrı province from Jan 2015–Dec 2019 through the Pearson correlation coefficient method, are determined. In the hybrid CEEMDAN and DNNs-based models, State activation functions (*SAF*), numbers of hidden neurons (*NHN*), dropout rates (*P-rate*), and network structural architect (*NSA*) as the meta-parameters are tuned for lessening the impact of overfitting/underfitting dilemmas and improving modeling performance. According to the comparison plots, performance evaluation measures, and total learnable parameter (*TLP*), the novel developed hybrid CEEMDAN-RF-(LSTM + GRU) model is confirmed as the best approach with an R^2 of 0.86 while, in the optimal scenario using the RF model, R^2 was 0.47.

✉ Ali Torabi Haghighi
ali.torabihaghighi@oulu.fi

Amin Gharehbaghi
amin.gharehbaghi@hku.edu.tr;
gharehbaghi.amin@gmail.com

Redvan Ghasemlounia
redvan.ghasemlounia@gedik.edu.tr

Farshad Ahmadi
f.ahmadi@scu.ac.ir

Rasoul Mirabbasi
mirabbasi_r@yahoo.com

[1] Faculty of Engineering, Department of Civil Engineering, Hasan Kalyoncu University, 27110 Şahinbey, Gaziantep, Türkiye

[2] Faculty of Engineering, Department of Civil Engineering, Istanbul Gedik University, 34876 Istanbul, Türkiye

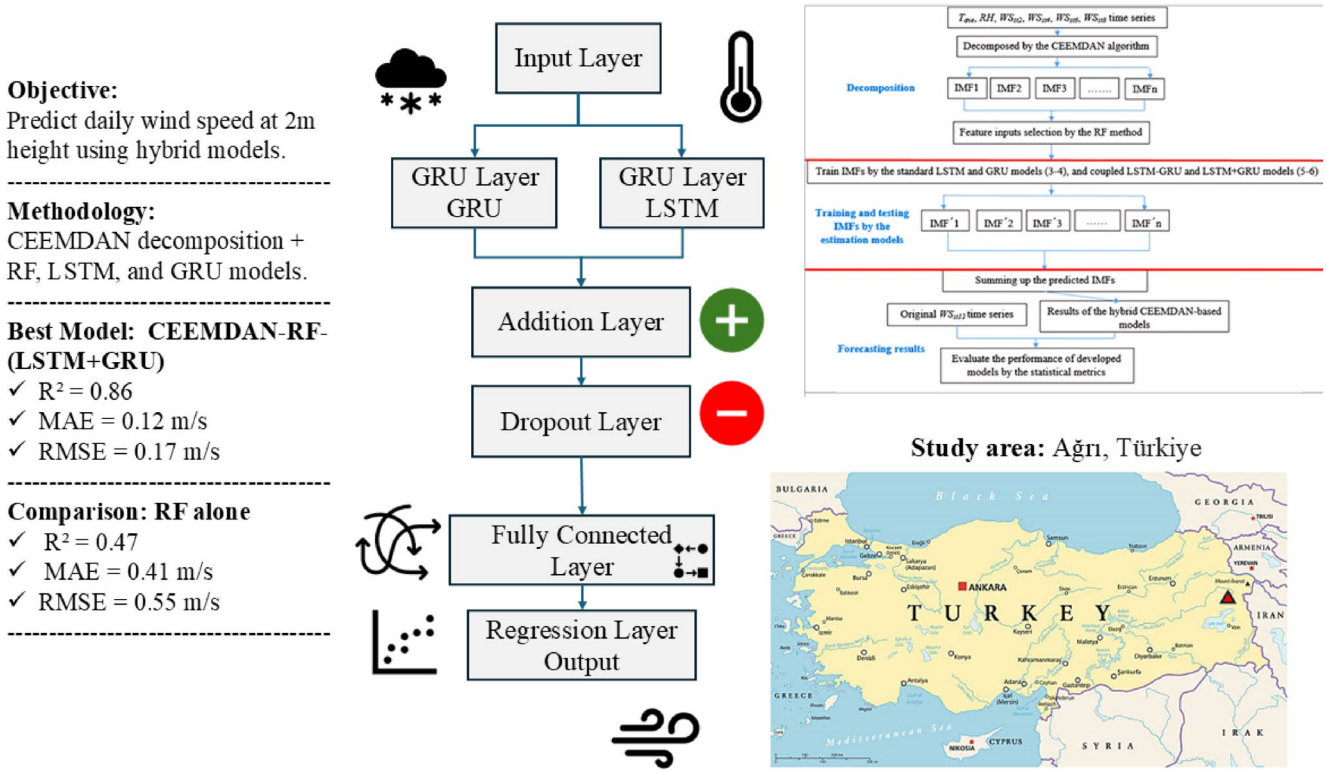
[3] Department of Hydrology and Water Resources Engineering, Shahid Chamran University of Ahvaz, 6135783151 Ahvaz, Iran

[4] Department of Water Engineering, Shahrekord University, Shahrekord, Iran

[5] Water, Energy, and Environmental Engineering Research Unit, University of Oulu, Oulu, Finland

Published online: 26 July 2025

Graphical Abstract



Based on the graphical snapshot, this study focuses on estimating daily mean wind speed at a 2-meter height in Ağrı, Turkey, using hybrid data-driven models. The research integrates the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) algorithm with advanced forecasting techniques, including Random Forest (RF), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) neural networks. The modeling framework explores various configurations, such as standalone LSTM and GRU, coupled LSTM-GRU structures, and a novel LSTM + GRU model using an Addition layer to enhance predictive accuracy.

Keywords Mean daily wind speed · CEEMDAN algorithm · Novel hybrid deep neural network model · TLP criterion parameter

1 Introduction

1.1 Background

Wind energy is a worthwhile and hygienic source in the energy sector. It has a significant role in sustainable worldwide energy development due to the growing demand for renewable energy sources. As an inexpensive and clean energy source, wind power has been reflected as an operative opportunity to substitute traditional fossil fuel energy (Ahmadi and Khashei 2021).

A precise and dependable estimation of wind speed (*WS*) is significant for the operational employment and management of wind energy when integrating wind power generation. Studies show that wind energy will provide 22% of the universal power source in 2030 (Zhang et al. 2021), hence, it is imperative to develop operative methods for precise prediction of *WS*. Nevertheless, *WS* is an intricate and completely non-linear process since it hangs on various meteorological and environmental parameters (Hu et al. 2021). Because *WS* is characteristically restricted by the temporal and spatial aspects, available difficulties such as inaccessible climatic data in some districts, and costly expenditures of direct measurement, scientists have to seek out subsidiary possibilities (Band et al. 2022). For these reasons, numerous flexible and strong non-linear estimation methods have been suggested for forecasting complicated phenome in the field of hydrology science such as Data-Driven methods (DDMs), Artificial Intelligence (AI)-based systems.

The AI methods are recognized as dominant surrogate schemes in estimating hydrological variables, particularly

the tortuous and highly nonlinear dynamic behavior of *WS*. These statistical methodologies have special pros such as no prerequisite information of inner relationships amid nonlinear variables with intricate relationships of any studied system, simple resolutions, and factual cheap calculations (Lin et al. 2022; Gharehbaghi and Ghasemlounia 2022). The DDMs have been frequently exploited as a robust system in different fields of science to estimate highly nonlinear multivariate hydraulic and hydrological parameters such as dimensions of flow separation zone (Gharehbaghi et al. 2023a, b), discharge coefficient of streamlined weirs (Gharehbaghi et al. 2023a, b), monthly reservoir inflows (Ahmadi et al. 2024), monthly runoff (Parsaie et al. 2024), and groundwater level prediction (Ghasemlounia et al. 2021; Lin et al. 2022; Gharehbaghi et al. 2022).

A perfect estimation of *WS* oscillations is an arduous undertaking because of the non-stationary and unsteady temporal mannerism and complex interactions between *WS* and other atmospheric factors. Also, as multivariate hydrological temporal parameters normally cover both deterministic and stochastic elements, choosing traditional DDMs is typically farfetched. Although conventional standalone DDMs could usually capture the nonlinear relationship between past and coming time series data, they have characteristically certain downsides such as high spatial-temporal variations, weights fine-tuning, and falling simply into the local optimal and overfitting, which cause to decrease in their predicting ability (Wang et al. 2021). So, it is hard to distinguish which approach has the superlative prediction ability since each model has its own pluses and minuses, and numerous approaches are also site-dependent (Band et al. 2022).

To overcome these problems and breaks, different hybrid DDMs have been recently expanded. The hybrid models as effective schemes are based on a combination of over one technique to take advantage of their superior single traits to improve the estimation facility of the classical individual AI techniques in modeling engineering problems. The most of these hybrid systems employ decomposition-based approaches with other forecasting models that firstly decompose original spatial-temporal data into different relatively more stationary sub-series; and then, generate predictive models for each sub-series so that be modeled more efficiently.

The majority of measured hydrological data have some noises that thwart the suitable transference of info to the models. Thus, signal pre-processing approaches have been utilized to address this problem. In the sphere of signal pre-processing, nonstationary signals are extensively decomposed adjustably by diverse mode decomposition (MD) algorithms such as CEEMDAN (Torres et al. 2011). CEEMDAN as an upgraded type of EMD (Huang et al. 1998) is adaptive and proficient for solving mode mixing problems without redundant noise in the reassembled signal, yet it cannot be realized by parallel computing (Torres et al. 2011). This algorithm disassembles adaptively an original time series into an efficient group of smooth and stationary Intrinsic Mode Functions (IMFs) components with diverse frequency bands and a residue component (R) on the basis of the local properties of time series. Then, each element is separately predicted through estimation models. In terms of *WS* prediction, different MD algorithms have been frequently integrated with other estimation models such as diverse ANN models by Qu et al. (2016); Zhang et al. (2017); Wang et al. 2017a, b; Yu et al. (2017); Liu et al. (2018); He et al. (2018), and have been reviewed in-depth by Heinermann & Kramer (2016), Manero et al. (2018), and Qian et al. (2019).

Of the single usual DDMs, the Random Forest (RF) approach developed by Breiman (2001) is very well-known and prevalent. It is a kind of ensemble technique based on decision trees (DTs) developed to address problems based on clustering and regression via the development of decision trees.

As a substitute for classical ANNs, innovative and specific system of data analytics such as Deep Neural Network (DNN) based models have been developed to estimate accurately the long-term non-stationery and unsteadiness behaviors of *WS*. Normally, DNN models have central self-coiled cells which beget to remember the previous information and generate it resourceful for adopting time series-based problems. They appraise the relationships amongst model inputs/output through an iterative characteristic of the learning process (Jiang et al. 2021). It should to noted that the location and number of atmospheric stations, opting for appropriate predictive parameters, and the obtainability of useful climatological variables recorded are contemplated as the primary limiting complications in performing DDMs (Parsaie et al. 2024).

At present, different types of DNN and hybrid DDMs have been employed as innovative approaches for the estimation of *WS*. For instance, Nezhad et al. (2021) developed an innovative prediction model for *WS* evaluation near the Favignana island in Sicily, Italy, via sentinel family satellite images and MLMs. They hybridized the generalized regression neural network (GRNN) and the whale optimization algorithm (WOA). They reported that the developed model has satisfactory accuracy with an $RMSE = 0.0205$, $MAE = 0.015$, and $MAPE = 6.83$. Tian and Chen (2021) forecasted multi-step short-term *WS* using a multi-model fusion approach. The developed technique was an integration of two different mode decomposition with stochastic configuration network (SCN) and hybrid SVM-PSO models. The recommended technique outclassed compared to the other

forecasting models with an $RMSE = 0.57$ (m/s), $R^2 = 0.94$, $MAPE = 0.18\%$, and $MAE = 0.42$ (m/s). López and Arboleya (2022) estimated short-term *WS* via the linear regression and multivariable LSTM and NARX neural network models over complex territory in a high altitude of Andes Mountains, Ecuador. They stated that LSTM model worked better than the other developed methods with an *MSE* of 0.13 (m/s), *RMSE* of 0.36 (m/s), *R* of 0.99, and *MAPE* of 4.9% in winter at a height of 80 (m) AGL. Yang et al. (2022) developed a novel hybrid model containing the empirical wavelet transform algorithm (EWT), Q-learning, GRU, BiLSTM, and Deep Belief Network (DBN) for forecasting short-term *WS*. They reported that the suggested hybrid EWT-Q-GRU-BiLSTM-DBN model outperformed compared to the other developed models. In 1-step ahead forecasting, *MAE* for sites #1–4 resulted in 0.0829, 0.0661, 0.0906, and 0.0803 (m/s), respectively. Kosana et al. (2022) developed a state-of-the-art robust hybrid online model selection with the Q-learning technique i.e., OMS-QL model and prediction model pool (FMP) for estimation of *WS*. Based on the prediction results, the OMS-QL model outperformed with an $RMSE = 0.63$, $MAE = 0.34$, and $MAPE = 0.085$, yet the optimal FMP model resulted in $RMSE = 0.93$, $MAE = 0.66$, and $MAPE = 0.11$. Chen et al. (2022) estimated *WS* using a novel hybrid LSTM with Improved Sparrow Search Algorithm-BP (ISSA-BP) neural network models. They employed singular spectrum analysis (SSA) and the CEEMDAN algorithm as preprocessing methods to de-noise the original *WS*. In comparison with six different estimation models, the proposed model outdid with $RMSE = 0.051$, $MAPE = 0.929\%$ for dataset 1 and $RMSE = 0.086$, $MAPE = 0.966\%$ for dataset 2 by one-step predictions. Li et al. (2022) estimated multi-step-ahead *WS* via a hybrid ensemble patch transform (EPT) and CEEMDAN with temporal convolutional networks (TCN). In 1-step-ahead, the developed EPT-CEEMDAN-TCN model resulted in an $MAE = 0.28$, $RMSE = 0.401$, $MAPE = 0.075$, and $NRMSE = 3.08$.

1.2 Problem Statement

Given the significant status of *WS* in energy studies, precise estimation of *WS* by an appropriate scheme is a vital task. So, this study aims to estimate a long-term time series of mean daily wind speed at a height of 2 (m) in Ağrı city (WS_{st12}), Turkey, through some standalone and combined DDMs. Concerning, diverse hybrid CEEMDAN-RF-based with estimation models including different layer structures of coupled versions of LSTM and GRU neural networks are developed using observed meteorological parameters and wind speed values in adjacent cities of Ağrı province. It is deliberately developed to intensify the prediction accuracy of the daily time series WS_{st12}. Accordingly, we do not restrain our probe merely to the ordinary DL networks structure.

The novelty of this study is the application of extraordinary and newly designed cutting-edge hybrid CEEMDAN-RF with different layer structures of DNN-based approaches with the module of seq2seq forecasting regression for the first time to predict elaborate natural phenome such as time series WS_{st12} oscillations. To the best of the authors' knowledge, in spite of several analyses in the forecasting *WS* by the general structure of DL-based models, no research has been performed in the literature on the function of the proposed model.

1.3 Objectives

The contributions of the present research are as follows:

1. Determine the most operative variables on daily WS_{st12}, amid a list of on-site potential weather parameters and wind speed values in its adjacent cities with time series attribute through a feature selection technique.
2. Development of state-of-the-art coupled CEEMDAN-RF-based algorithms with different layer structures of coupled versions of LSTM and GRU neural network models for precise forecasting of daily WS_{st12} vacillations rhythm.
3. Determine the optimal rate of aimed meta-parameters for well-configuration and lessening the impact of overfitting/underfitting obstacles.
4. Assessment and comparison of the accuracy of modeling with counterparts in the validation stage to élite the best-developed model in offering better reliable and consistent performance vis statistical indices and comparison plots.

2 Study Site Description

Ağrı province is located in the Eastern Anatolian Region of Turkey with an area of 11,376 km^2 and 1640 m above sea level. It has a continental semi-arid and Mediterranean-effected climate. In Ağrı province, summers are usually warm, dry, and short with cool nights but winters are comparatively long and frozen. The average low and high temperatures are −16 °C in January and 30 °C in August, respectively. In this city snows a great deal, remaining for a mean of 4 months in winter. Peak rainfall occurs in April-May, and total average precipitation in a year and mean

daily sunshine durations are 512.9 mm and 6.2 (hours), respectively (Turkish State Meteorological Service).

In the current work, to estimate the mean daily wind speed at a height of 2 (m) in Ağrı city (WS_{st12}), 1675 daily climatological datasets in Ağrı city accompanied by observed mean daily wind speed in its adjacent cities including Van, Bitlis, Muş, Erzurum, Kars, and Iğdır cities for the period 1/6/2015–12/31/2019 are employed. The meteorological parameters used in Ağrı city are daily average air temperature (T_{ave}) (°C), relative humidity (RH) (%), and sunshine duration (SSD) (hours), provided by the Turkish State Meteorological Service. Figure 1 (A and B) depict the map of geographical position and observation climatological stations used in the study area, respectively.

3 Data Preprocessing

As the precision of modeling for a particular target generally has confidence in the pertinent choice of forecaster parameters, irrelevant groupings could beget the deficient efficiency of the technique (Gharehbaghi and Ghasemlounia 2022). Thus, a features choice system is functioned to specify operational variables and lessen convergence time. In the current investigation, this practice is operated using the Pearson correlation coefficient (PCC) method to distinguish operative predictors among potential ones in forecasting the WS_{st12}.

PCC is a widely used correlation statistic to compute the strength of relationship amid predictor and target variables in numerical datasets. PCC quantities for the potential in-site parameters including daily weather variables in Ağrı

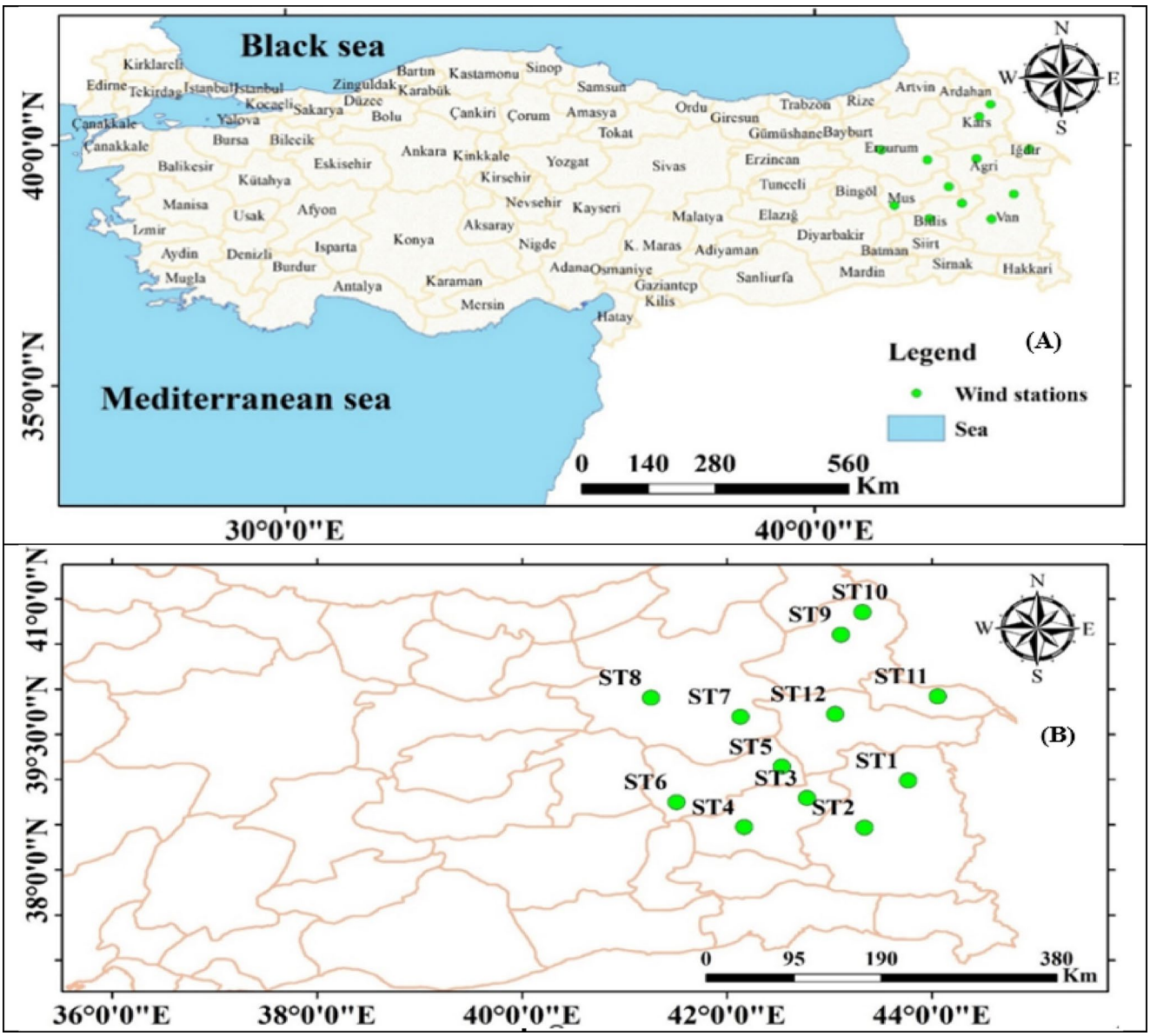


Fig. 1 Map of the (**A**) Geographical position, (**B**) observation climatological stations used in the study district

Table 1 The *PCC* value for the potential variables employed versus WS_{st12}

Variables	PCC value	Variables	PCC value
RH	0.57	WS_{st5}	0.61
SSD	0.25	WS_{st6}	0.29
T_{ave}	0.58	WS_{st7}	0.73
WS_{st1}	0.57	WS_{st8}	0.25
WS_{st2}	0.16	WS_{st9}	0.19
WS_{st3}	0.54	WS_{st10}	0.11
WS_{st4}	0.13	WS_{st11}	0.16

city along with observed wind speed in its adjacent cities with respect to the WS_{st12}, for a model targeting a 95% confidence level are provided in Table 1.

According to Table 1, T_{ave}, *RH*, WS_{st1}, WS_{st3}, WS_{st5}, and WS_{st7} are selected as the final effective predictors owing to their high *PCC* values (more than 0.5). Likewise, it is clear that the WS_{st8} is the most influential parameter on WS_{st12}. Consequently, WS_{st12} in its ultimate functional formula can be expressed as:

$$WS_{st12} = f(T_{ave}, RH, WS_{st1}, WS_{st3}, WS_{st5}, WS_{st7}) \quad (1)$$

Equation (1) reflects the temporal prearrangement of multi-predictor/target variables and is authorized for forecasting the WS_{st12} by DDMs. However, due to the byzantine and non-linear relationships of the variables represented in Eq. (1), reliable and robust DDMs are required to weigh the data series. To do so, all datasets for the period Jan 2015–Dec 2019 are firstly normalized to create unit variance and zero average as suggested by Lawrence et al. (1997). Then, they are chronologically riven into two subgroups without randomization, where the first 80% of the set is used for calibration of the models, and the rest 20% is also consecutively utilized for validation. Table 2 presents some statistics indices of effective predictors in Eq. (1) for forecasting the WS_{st12}.

4 Methods

4.1 CEEMDAN Algorithm

EMD as an empirical method is on the basis of local features of the signal, which disintegrates the signal into different frequency elements termed intrinsic mode functions (IMFs). It discards the restraints of Fourier transform and is appropriate for the analysis of nonlinear and non-stationary signals (Li et al. 2022). Nevertheless, the modal aliasing quandary typically occurs in the splitting process which somewhat begets the unsatisfying outcomes. In this sense, CEEMDAN as an advanced and highly preferred algorithm was developed on the basis of EMD and EEMD. Owing to its specific benefits, it can precisely realize signal decomposition by lessening mode aliasing difficulty originated by EMD decomposition and eliminating the reformation error originated by EEMD (Torres et al. 2011). Technically, instead of the white Gaussian noise, CEEMDAN adds adaptive white noise with normal distribution into the preliminary signal in each period of disintegration and computes an exclusive residue for each mode (Torres et al. 2011). Furthermore, it has a data-oriented structure and does not need a fundamental function, so it can offer an enhanced spectral decomposition and feature information extraction to the modes with low computation cost (Chen et al. 2022). The flow chart of CEEMDAN algorithm is depicted in Fig. 1. In this Figure, E_k ($W^{(i)}$) is the *k*-th IMF of the white Gaussian noise disintegrated by EMD; E_k (●) shows the operator which creates the *k*-th mode disintegrated by EMD, $W^{(i)}$ signifies the *i*-th added white Gaussian noise with unit variance and zero mean, and the coefficient $\beta_k = \varepsilon_0$ std (r_k) characterizes the choice of SNR at each stage with *k*-th signal residual element after decomposition (r_k). The *N* is the total ensemble number and *M* (●) shows the operator which creates the local mean of the signal (Bai et al. 2018).

4.2 Random Forest (RF)

To generate a regression tree in RF, repetitive dissection and several regressions are performed and the decision process is iterated at each inner node of the root node consistent with the tree instruction till a stop circumstance is procured (Breiman 2001). Then, RF incorporates a group of independently elected and dispersed decision trees and attains the ultimate individual estimation by an average of the yields of all DTs in the forest (Svetnik et al. 2003). In this model, each node has *d* attributes of the base DTs that $\sqrt{d}$ attributes are chosen subjectively from the characteristics set, and

Table 2 Statistical indices of variables used in Eq. (1) in the period Jan 2015–Dec 2019

Variable	Min	Max	Mean	STDV*	Skewness	CV*
T_{ave} (°C)	− 29.2	27.9	1.3	0.77	0.82	0.58
RH (%)	83.3	94.5	90.6	4.61	− 1.23	0.051
WS_{st1} (m/s)	0.4	9.7	2.66	1.12	1.65	0.42
WS_{st3} (m/s)	0	10	2.78	1.35	1.53	0.48
WS_{st5} (m/s)	0	2.4	0.71	0.29	0.81	0.41
WS_{st7} (m/s)	0.1	5.7	1.33	0.83	1.27	0.62
WS_{st12} (m/s)	0	4.9	1.3	0.77	0.82	0.58

**STDV* and *CV* indicate the standard deviation and coefficient of variation, respectively

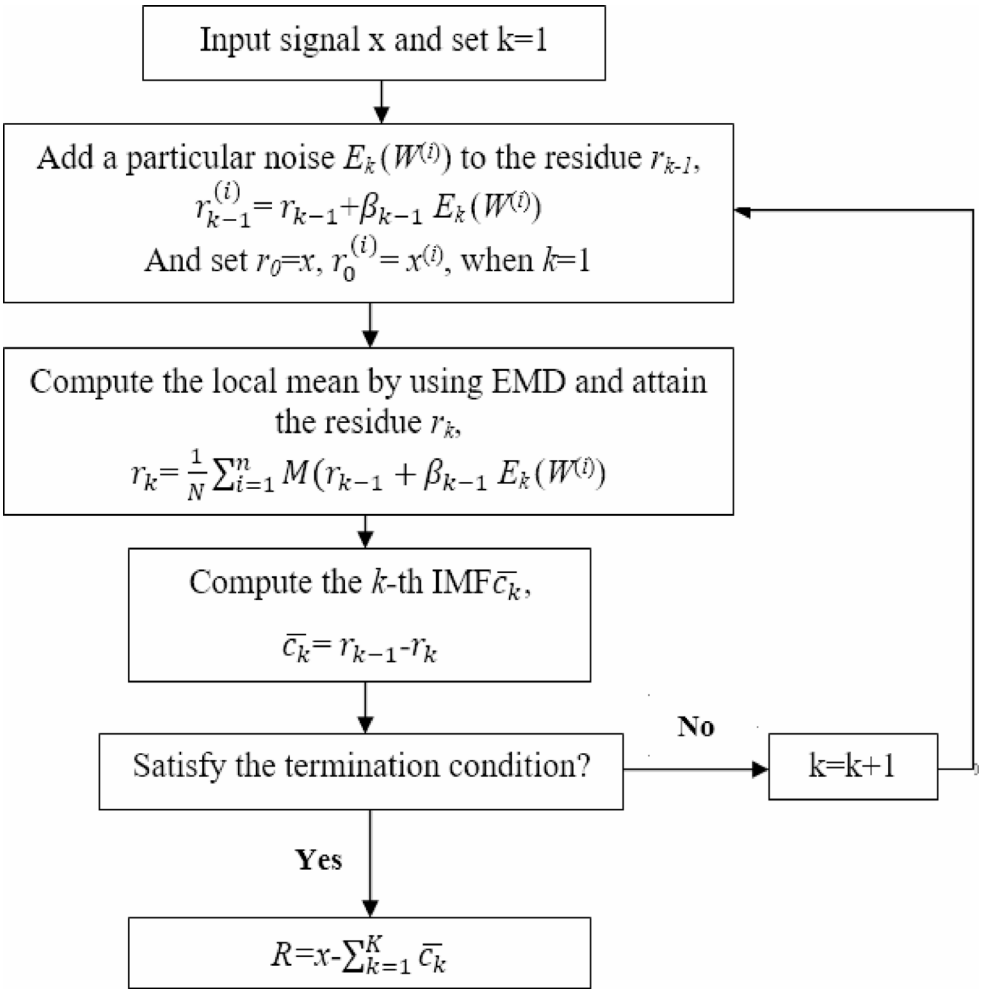


Fig. 2 Flow chart of CEEMDAN algorithm (Bai et al. 2018)

subsequently, the ideal characteristic is picked by the Gini impurity index (I_G) for the excellent split threshold (Strobl et al. 2008). I_G can be computed as follows (Breiman 2001):

$$IG\ (f)\ =\ 1 - \sum_{i=1}^{m} \bar{f}_i^2 \tag{2}$$

where f_i signifies the possibility of group i at node m. The optimum split threshold is determined by the smallest amount of I_G. In this work, the number of DTs is determined by a trial-and-error process.

4.3 LSTM and GRU Neural Networks

LSTM was developed by Hochreiter and Schmidhuber (1997) to preclude the waning gradient dilemma in traditional recurrent neural network (RNN) models. LSTM by using an outstretched memory cell unites prior times series-based info to capture the long-term trends of sequential data for forecasting the next state of a target. Memory blocks in LSTM contain CEC (Constant Error Carousel) cell, along with input (i), forget (f), and output (o) gates (Graves and Schmidhuber 2005). For a given time step t, f_t is an important state that specifies whether the present data should be overlooked or recalled. i_t brings forth the data up-to-date and $\widetilde{C}_t$ remembers the new information and the new memory cell state is upgraded by C_t. Finally, o_t regulates the output activations (Chen et al. 2022). These processes are expressed by the following Equations (Hochreiter and Schmidhuber 1997):

$$f_t\ =\ s\,(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \tag{3}$$

$$i_t\ =\ s\,(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{4}$$

$$\widetilde{C}_t =\ tan\,h\,(Wcx\,xt\ +\ Wch\,ht-1+\ bc) \tag{5}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \tag{6}$$

$$h_t = O_t \odot tanh\,(c_t) \tag{7}$$

$$O_t\ =\ s\,(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{8}$$

where x_t and y_t show the sequential input and output sequence at time (t). W and b are the corresponding weight matrix and bias of gates, respectively, and σ is the sigmoid function. The $\odot$ is dot product operation. The architecture of an LSTM memory cell is shown in Fig. 3.

GRU model was introduced by Cho et al. (2014) that operates a single hidden state in preference to two states as per LSTM, to hand out info over time steps. As a modification of LSTM, GRU does not include isolated memory cells, so its simple structure makes a faster training process. The h_t and C_t combine into one in the GRU. In the GRU cell, there are two control gates: update gate (z_t) and reset gate (r_t) are directly applied to h_{t-1} to compute $\widetilde{C}_t$ (candidate state) (Yang et al. 2022). In the structure of GRU cell as depicted in Fig. 4, update Equations are computed as follows (Cho et al. 2014):

$$r_t = s\left(W_{xr}x_t + W_{hr}h_{t-1} + b_r\right) \tag{9}$$

$$z_t = \sigma\left(W_{xz}x_t + W_{hz}h_{t-1} + b_z\right) \tag{10}$$

$$\widetilde{C}_t = tan\,h\left(Wxc\,xt + Whc\left(rt \odot ht - 1\right) + bc\right) \tag{11}$$

$$Ct = (1 - zt) \odot Ct - 1 + zt \odot \widetilde{C}_t \tag{12}$$

$$h_t = C_t \tag{13}$$

If r_t and z_t are set to 1, the model is considered the traditional RNN. Both GRU and LSTM operate comparatively in the same manner, but GRU has a small number of learnable parameters with a condensed building. Therefore, it can meet rapidly, if the dimension of data is not too large (Yang et al. 2022).

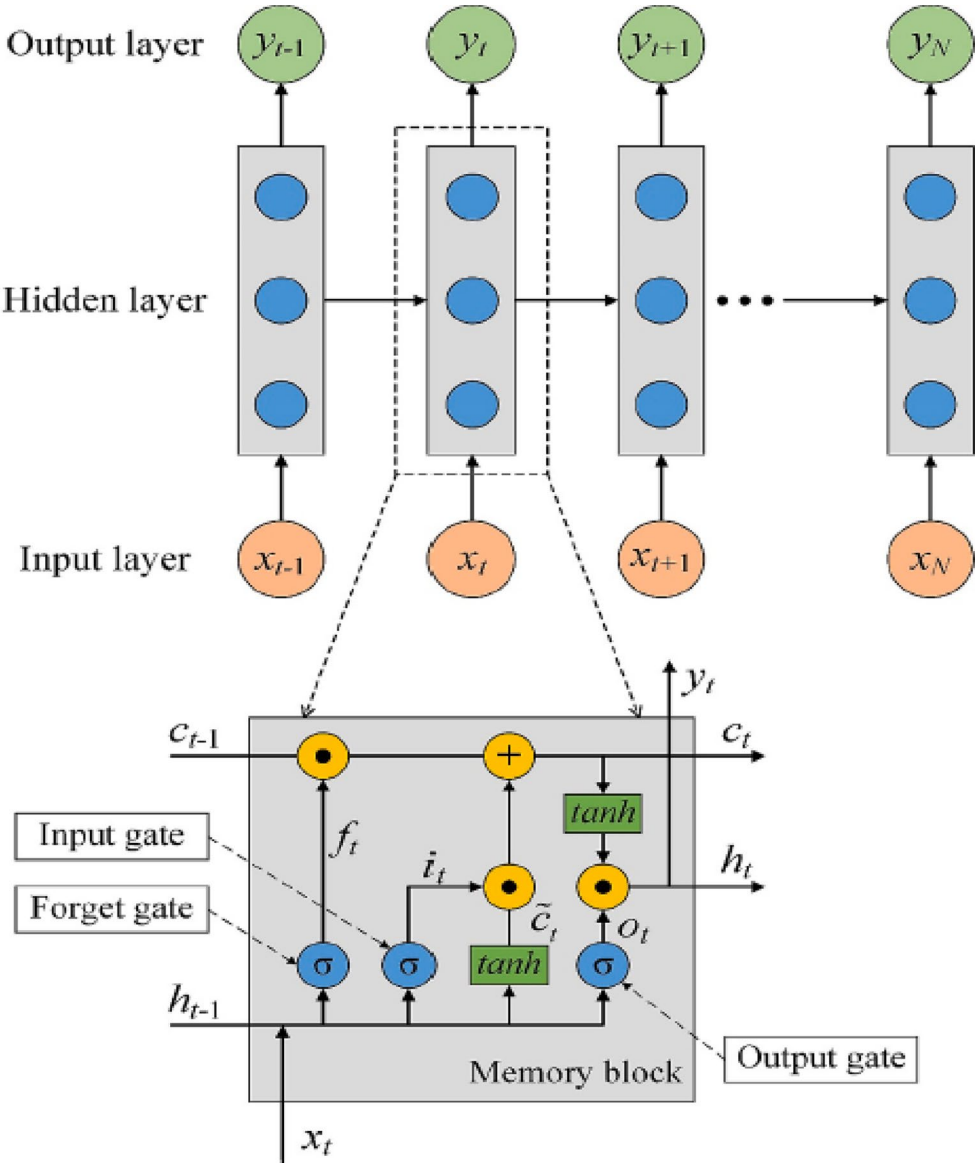


Fig. 3 Structure of an LSTM memory cell (Fang et al. 2021)

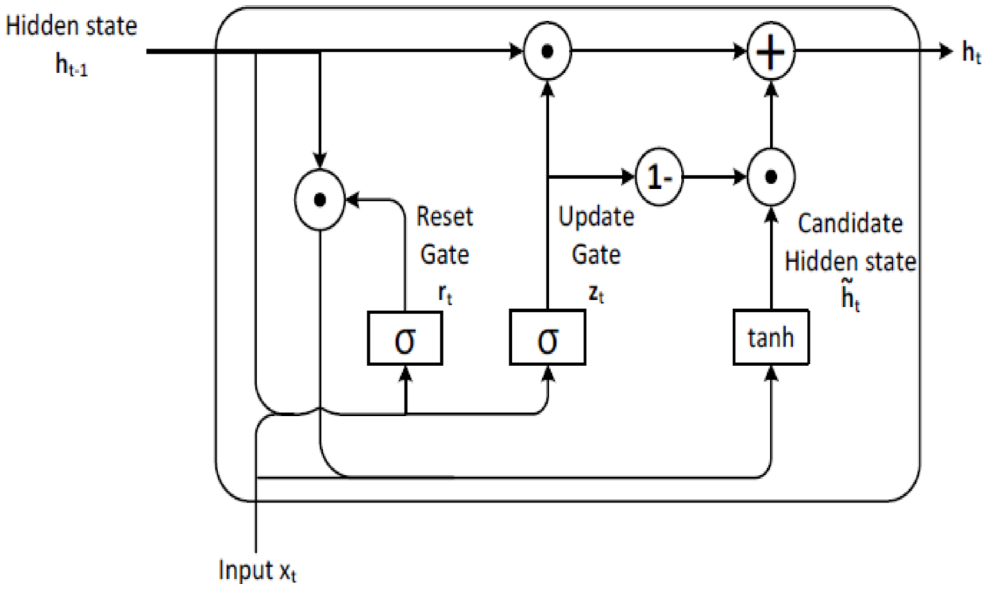


Fig. 4 Architecture of a GRU memory cell

4.4 Framework of the Modeling Process

In the procedure of forecasting time series WS_{st12}, as the first, the typical standalone RF as the model (1) is adopted by contemplating the mean squared error (*MSE*) value obtained in calibration and validation datasets. The ideal number of trees is employed as modeling the meant parameter via RF with the aim of no alteration in *MSE* by intensifying the number of trees.

To develop the combined models, firstly, the data decomposition scheme via CEEMDAN algorithm is operated. In this regard, the effective multivariate meteorological predictor matrix variables are decomposed simultaneously by CEEMDAN algorithm into their band-limited subsequence IMFs and R of train and test sets sub-signals, independently, with different frequency characteristics. It brings forth to enhance forecasting capabilities by diminishing the error accretion, non-stationary, complexity, and convergence time affected by too various elements in model development.

Then, the extracted IMFs and R are applied to RF to pick the best band-limited signals. Next, the outputs of CEEMDAN-RF are incorporated as input matrices to reform a new time series for training by the different layer network structures of coupled versions of LSTM and GRU neural network models with the sequence output mode to create various hybrid CEEMDAN-RF-DL models. In this direction, at the start, the following layer network structures are designed by MATLAB 2024a as displayed in Fig. 5 (A-D).

The total layers in the structures (1), (2), (3), and (4) are 5, 5, 6, and 7, respectively. At last, by integrating the output of CEEMDAN-RF algorithm with designed DL-based structures (1–4), four different hybrid decomposition-DL-based approaches namely, CEEMDAN-RF-LSTM, CEEMDAN-RF-GRU, CEEMDAN-RF-(LSTM-GRU), and CEEMDAN-RF-(LSTM+GRU) models are developed to estimate the WS_{st12} in the study area.

In these hybrid models, to reach apposite prediction results, reduce the overfitting influence, and enhance the qualification of developed structures, the algorithm tuning scheme by applying different kinds of meta-parameters is utilized. As there aren't definite yardsticks to determine the well-chosen meta-parameters for a given dataset and model, this manner ponders as a time-consuming and strenuous task. In this context, plentiful scenarios are categorized to clarify apt meta-parameters by a trial-and-error practice.

In these hybrid models, the Input Layer takes the output of CEEMDAN-RF algorithm and the Addition Layer (+) sums feedback from compound neural network layers element-wise (MathWorks, Inc. 2021a). Besides, for fine-tuning the *SAF*, diverse integrations of *tanh* and *softsign* in both GRU and LSTM layers with *σ* gate activation function are applied. As well, to tune *NHN*, different values in each structure are examined. In weighing up large data, the Drop-out Layer presents an operative controlling system based on routinely giving up some neurons with a specified possibility rate of *P* in a layer distributed at each training repetition (Srivastava et al. 2014). In this work, quantities including 0.2, 0.5, and 0.8 are employed for fine-tuning *P-rate.* The Fully Connected layer (FC) is operated to augment the fitting potential (MathWorks, Inc. 2021a). In all structures in the calibration stage, "auto" and 1 are applied in its input and output sizes, respectively, to fine-tune intuitively its dimensions. The Regression Output Layer is applied in the last layer of all structures to determine the "*half-mean-squared-error loss*" for regression purposes through the loss function. To recondition the network bias and weights in all structures, the "*Adam*" optimization algorithm with 250 maximum repetitions is used. In an attempt to preclude the

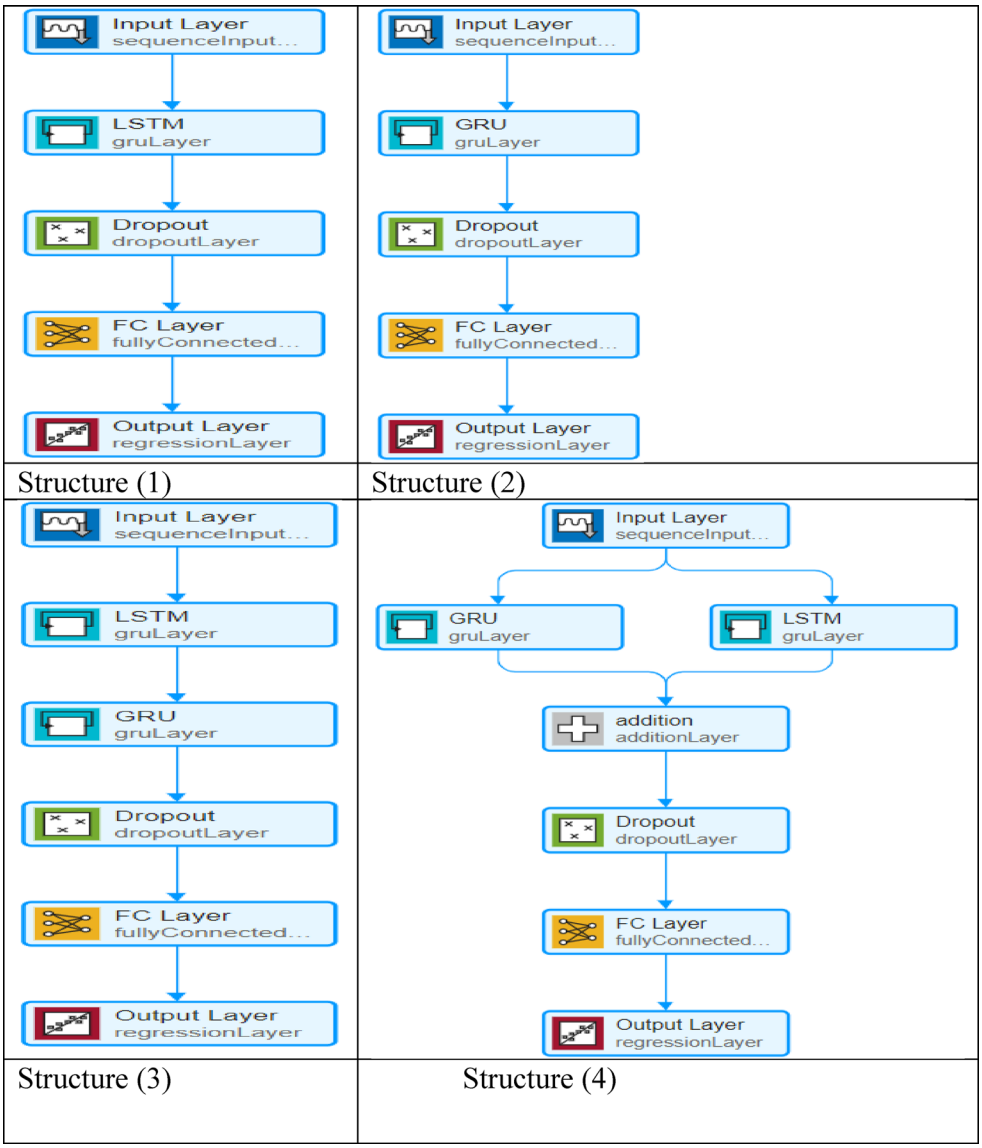


Fig. 5 Designed DL-based layer structures. (**A**) General single LSTM layer network structure (1), (**B**) General single GRU layer network structure (2), (**C**) Simple coupled LSTM-GRU layer network structure (3), (**D**) Coupled LSTM with GRU models by Addition layer (i.e., LSTM + GRU layer network structure (4))

gradients from vanishing and decrease the adverse impact of padding hitches, training options are set as instructed by Ghasemlounia et al. (2021).

To end with, the ensemble ultimate predicting results are obtained by summing up individually the sub-results of the relevant estimated IMFs sub-signals using the predictive DL-based models.

5 Performance Evaluation Metrics

R^2 (Eq. 14), *RMSE* (Eq. 15), *MAE* (Eq. 16), and *AICc* (Eq. 17) are computed to assess the precision and efficiency of the models developed in forecasting the temporal WS_{st12}.

Determination coefficient (R^2);

$$R^2 = \left(\frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N\sigma_x\sigma_y} \right) \tag{14}$$

Root mean square error (RMSE);

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}} \tag{15}$$

Mean absolute error (MAE);

$$MAE = \sqrt{\frac{\sum_{i=1}^{N}|(y_i - x_i)|}{N}} \tag{16}$$

Table 3 Results of models under ideal meta-parameters in the forecasting WSst12

Model	Model ID	AICc	R^2	MAE (m/s)	RMSE (m/s)	Convergence time (s)
Single RF	1	1553.6	0.47	0.41	0.55	0.4
CEEMDAN-RF	2	1253.2	0.78	0.25	0.36	15
CEEMDAN-RF-LSTM	3	1170.5	0.83	0.24	0.31	65
CEEMDAN-RF-GRU	4	1158	0.83	0.23	0.3	62
CEEMDAN-RF-(LSTM-GRU)	5	1258.5	0.77	0.24	0.34	161
CEEMDAN-RF-(LSTM+GRU)	**6**	**766.1**	**0.86**	**0.12**	**0.17**	**115**

*The bolded values specify the results of the best model

Corrected Akaike′s Information Criterion (AICc) Index;

$$AICc = \frac{(N1\mathrm{n}(\sigma_{\varepsilon}^2)(N-k-1)) + 2kN}{N-k-1} \quad (17)$$

where N is the number of datasets, k is the number of parameters, X_i and y_i are the observed and estimated WS_{st12} at time i. σ_{ε}, σ_x and σ_y are the residuals' standard deviation, standard deviation of measured and estimated WS_{st12}, respectively, μ_x and μ_y are the average of measured and estimated WS_{st12}, respectively. Smaller quantities for *MBE*, *RMSE*, *AICc* together with bigger quantities for R^2 show a superior forecasting performance.

6 Results and Discussion

6.1 Results of Models Developed

In the developed hybrid models, the number of trees in RF model was chosen such that growing the number of trees from the ideal number had no remarkable impact on the accuracy of models. With this in mind, in the RF model the number of trees, minimal number of samples for splitting in interior nodes, and minimal number of samples on each node are obtained 100, 2, and 1, respectively. After frequent trials, the results of all models under ideal meta-parameters in the validation phase are tabularized in Table 3.

According to Table 3, through a precision assessment of the single classic RF as the benchmark model and hybrid models, it can be concluded that operating the CEEMDAN algorithm improves noticeably the ability of the standalone RF model. It provides the convenient subcategories of the primitive measured series that can be employed effectively as the new predictors to intensify the model's capability in predicting the meant target parameter by mining appropriate info caused by these new sub-series. Moreover, the model 6 is contemplated as the best approach developed due to smaller quantities that indicate better prediction performance metrics equated to the other models.

Table 4 Characteristics and ideal meta-parameters of the hybrid DL-based models

Model	SAF	*P*-rate	NHN	TLP
CEEMDAN-RF-LSTM	tanh	0.5	32	4385
CEEMDAN-RF-GRU	softsign	0.5	32	3297
CEEMDAN-RF-(LSTM-GRU)	tanh-softsign	0.5	32	10,625
CEEMDAN-RF-(LSTM+GRU)	**tanh-softsign**	**0.5**	**32**	**7649**

*The phrase "*tanh-softsign*" signifies that the sort of *SAF* in LSTM and GRU layers are selected *tanh* and *softsign*, respectively. Also, the bolded values specify the characteristics of best model

6.2 Performance Comparison

In the hybrid DL-based models 3–6, under optimal meta-parameters in the training step, after 250 iterations the *Loss* and *RMSE* values were accomplished at 1E-7 and 1E-6, respectively. As a whole, these models in the training step were more exact than in the testing step. The characteristics of hybrid DL-based models under ideal meta-parameters in the validation phase are listed in Table 4.

According to Table 4, *tanh-softsign* combination as the suitable *SAF* in the models 5–6 brings forth learning intricate functions and not being disposed to the vanishing gradients dilemma in the training stage. *softsign* activation function intensifies the speediness of model training and *tanh* can take fittingly multipart relationships of long-term temporal data (Lin et al. 2022).

As well, of these DL-based models, the novel CEEMDAN-RF-(LSTM+GRU) model 6 outperformed strikingly that its superiority can be justified based on *TLP* values listed in Table 4. *TLP* factor is a significant criterion to measure the real physical capacity of DL-based models. A suitable layer network structure along with optimal *NHN* directed to a well-balanced *TLP* value that can reduce the overfitting problem (Gharehbaghi et al. 2022). From this perspective, in the same *NHN*, due to the over-extra *TLP* amount and consequently superfluous capacity in the model 5, it has comparatively forgotten the training dataset, causing it

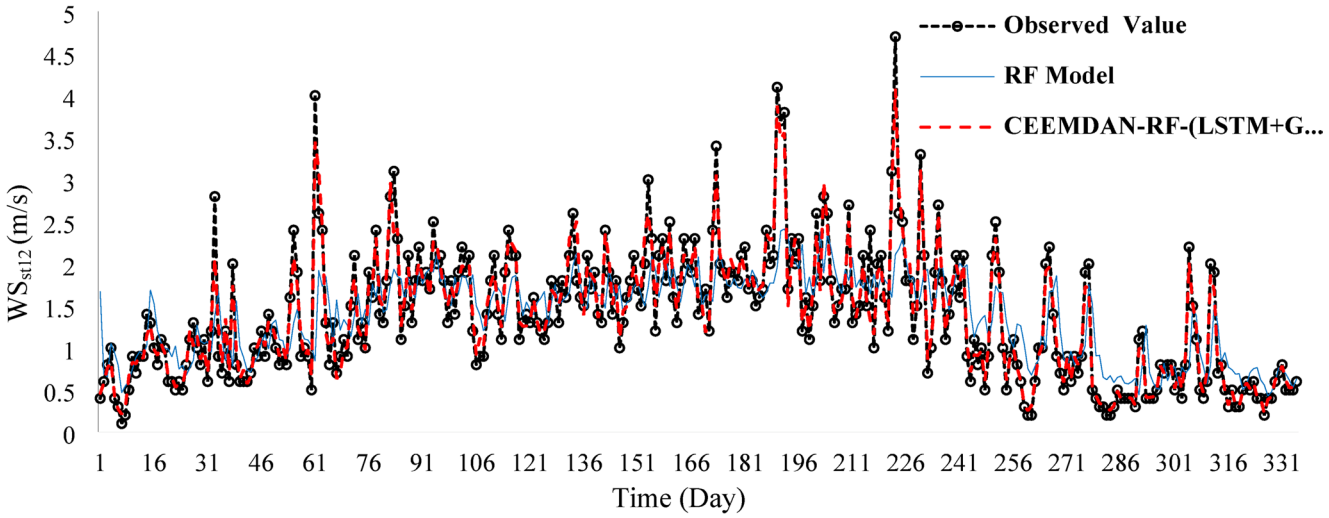


Fig. 6 Comparison of predicted and measured WS_{st12} in the testing stage (335 days)

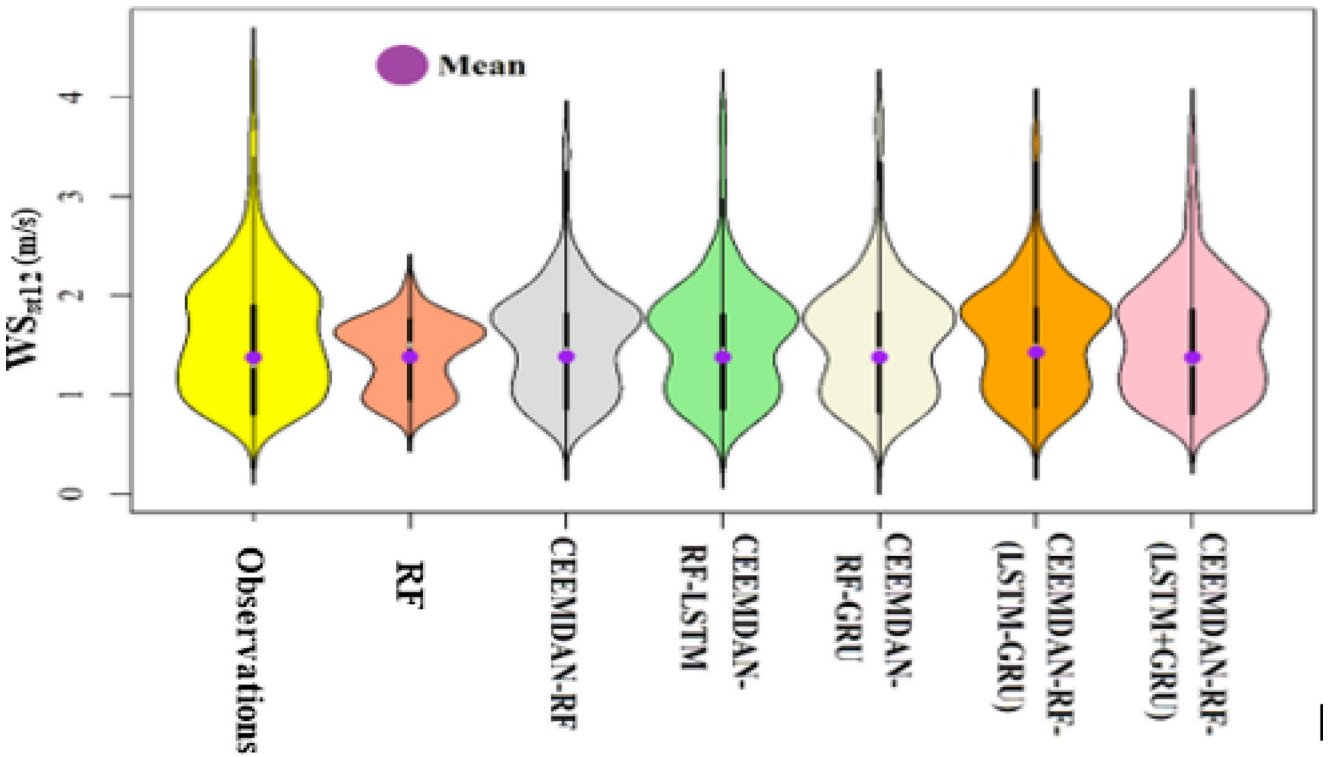


Fig. 7 Violin plots of the observed the WS_{st12} versus the forecasted under the ideal scenario in the testing stage

to overfit and bead the optimization. Likewise, as a consequence of the lesser *TLP* value in the models 3–4, they are underfitted. Nonetheless, owing to a balanced *TLP* in the model 6, it surpassed other models.

In respect of the convergence time and reiterations, the model 6 in the same *NHN* is more operational than the model 5 as a result of the apposite *TLP* that eases the model in realizing the ideal weight sets more swiftly– 250 repetitions in 115 s; Whereas, the model (5), for the maximum *TLP* value, necessitates more time to train– 250 reiterations in 161 s.

Figure 6 equates graphically the observed and forecasted time series time series WS_{st12} by CEEMDAN-RF-(LSTM + GRU) model 6 and standalone RF as the benchmark model under the ideal meta-parameters in the validation phase. According to this figure, the model 6 thanks to its ingenious layer's network structure could nearly reminisce properly the WS_{st12} distribution of time series observations

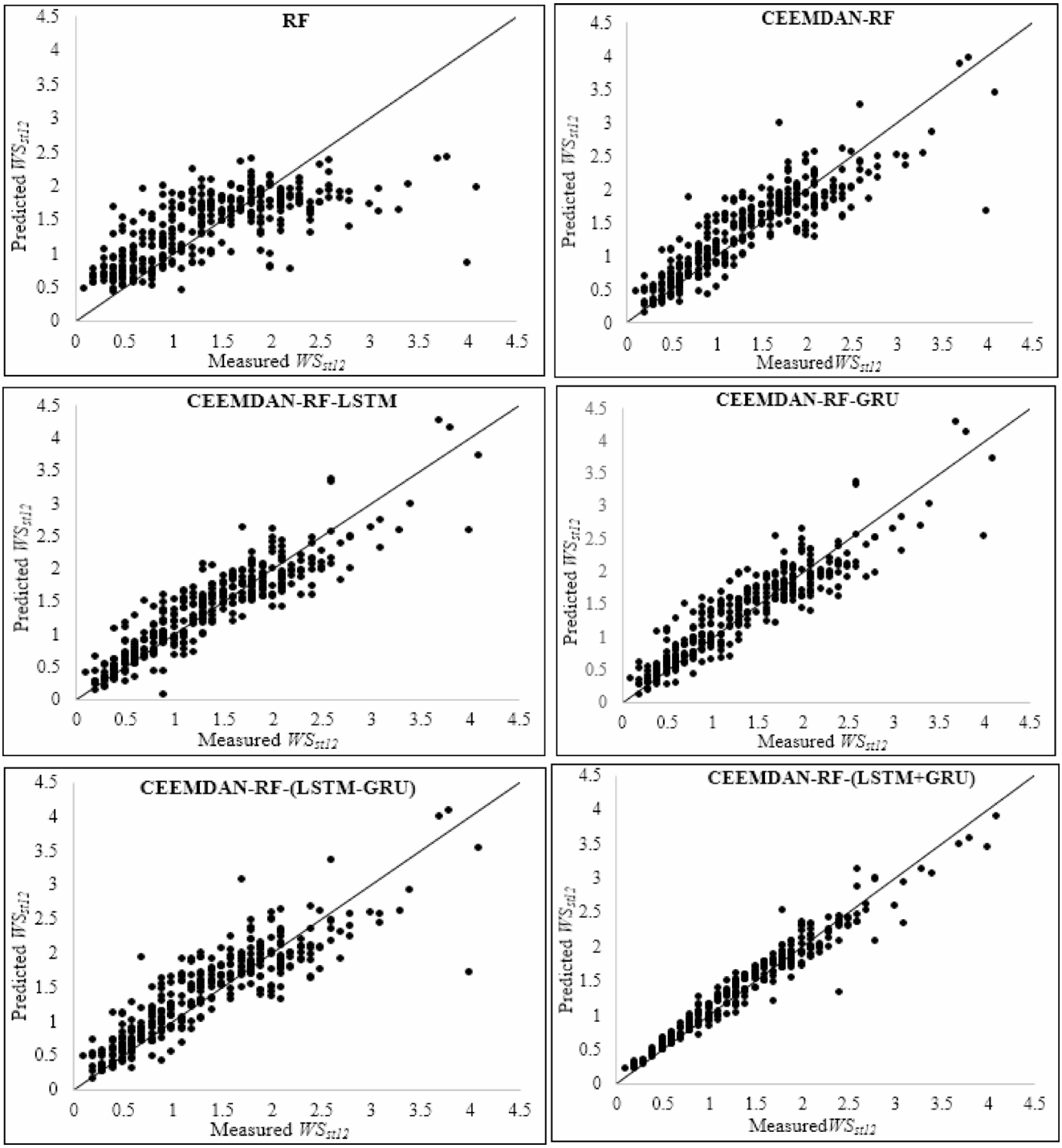


Fig. 8 Scatter plot between the measured and predicted WS_{st12} by the models developed under the optimal scenario in the testing phase

and successfully fit the fluctuations trend primarily in the head and deepest positions with high accordance.

To differentiate the best model, the observed and forecasted time series WS_{st12} in the validation stage are also graphically likened by the Violin and Scatter plots (Figs. 7 and 8, respectively). As per Fig. 7, according to the maximum, minimum, average, and distribution trend of forecasted data for the observed WS_{st12} by the model 6, it can be inferred that this model is appositely form-fitted compared to the other models. Model 6 thanks to its ground-breaking network architect can reminisce pertinently WS_{st12} of preceding observations and efficaciously capture the oscillations trend. The models 1, 2, and 5 in the same *NHN*, nonetheless, are abortive to fit the data and swerve too far from the

observed values; generally, in the zenith and beneath points that reveal high irregularities. Furthermore, it is apparent that there is no too profound divergence between the models 3 and 4, nonetheless, they both underrated the mean WS_{st12} matched to the observed one.

According to the fit line equation in the Scatter plot for the model 6, much of the data points are mostly close to the best fit line (i.e., Y = X) with lesser dispersal that verifies superiority and a high stability with an acceptable R^2 of 0.86.

7 Summary and Conclusion

A time series of 1826 mean daily wind speed at a height of 2 (m) from Jan 2015–Dec 2019 in Ağrı city (WS_{st12}), Turkey, was forecasted with measured mean daily of on-site potential climatic parameters (T_{ave}, *RH*) and wind speed observed in its adjacent cities by developing different hybrid CEEMDAN-RF algorithm with DNNs based models. These hybrid models are CEEMDAN-RF-LSTM, CEEMDAN-RF-GRU, CEEMDAN-RF-(LSTM-GRU), and CEEMDAN-RF-(LSTM + GRU) models. To diminish the impacts of underfitting/overfitting dilemmas and a well-configuration of the models developed, algorithm fine-tuning by *NHN*, *P-rate*, and *SAF*, as the meta-parameters in common with the trial-and-error method were executed. Moreover, the *TLP* parameter was computed as a critical norm for appraisal of the overfitting/underfitting circumstances and estimating the aptitude of hybrid DNN-based models. The most notable outcomes of the estimation process are:

1. All developed models were more precise in the calibration step than the testing step. In all developed hybrid CEEMDAN-RF with DDN-based models 3–6, the optimal *P-rate* and *NHN* values were obtained to be 0.5 and 32, respectively.
2. Since the standalone RF showed weak performance, integrating with the CEEMDAN algorithm began to increase the value of R^2 by 66% and reduced the value of *MAE* and *RMSE* by 40% and 35%, respectively. Also, integrating CEEMDAN-RF with the general single LSTM model caused to increase in the value of R^2 by 7.7% and reduced in the value of *AICc* and *RMSE* by 6.7% and 13.8%, respectively.
3. The novel suggested CEEMDAN-RF-(LSTM + GRU) model 6 in the same *NHN* owing to a balanced *TLP* value outperformed among all developed models. This model compared to the CEEMDAN-RF model caused an increase in the value of R^2 by 21.7% and reduced in the value of *AICc* and *RMSE* by 38.8% and 52.7%, respectively.
4. In the models 3–6, the CEEMDAN-RF-(LSTM-GRU) model 5 showed the weakest performance. The main reason for this is that in the same *NHN*, due to its over-redundant *TLP* amount and accordingly excess capacity, it has relatively forgotten the training dataset, causing it to overfit and drip somewhat the optimization.
5. The CEEMDAN-RF-GRU model 4 for the smallest *TLP* value, necessitates less time to train– 250 repetitions in 62 s which was pondered as the fastest model. However, the model 5 for the extreme *TLP* quantity, demands extra time to train– 250 echoes in 161 s.
6. The model 6 in the same *NHN* was more working than the model 5 on account of the pertinent *TLP* value, which facilitated the model in getting the optimal weight sets more swiftly– 250 echoes in 115 s. So, it can be decided that for accomplishing effective DNNs-based models, the ideal *NHN*, and *TLP* via the trial-and-error and algorithm fine-tuning processes should be operated.

The forward-thinking structure of model 6 is different from routine DNN models and is not normally operated in hydrology knowledge. It can be utilized as a smart prediction approach for the time series wind speed in numerous climatic conditions as its admirable performance and accuracy were statistically proven. Also, the model 6 is considered cheap-to-development, uncomplicated, and time-saving. Despite the worthy privileges of the model 6, it has some restrictions e.g. a necessity for the regular measurement of diverse climatic data records over a long-lasting time for estimating.

Albeit this research addressed the capability of various networks of DNNs-based models in predicting mean daily wind speed, the coming researche could be conducted by other sorts of strategies by integrating DNN models with bio-inspired optimization algorithms such as Cat Swarm Optimization (CSO), Fish Swarm Algorithm (FSA), etc. The outcomes ought to be likened to the conclusions of this study so that the superior model can be recognized.

Funding Open Access funding provided by University of Oulu (including Oulu University Hospital).

Data Availability Information on the precise computational approach used in this paper is accessible from the authors by formal request. The authors express their gratitude to the Turkish State Meteorological Service for providing the meteorological station data for this study. The developed codes and data used in this work are not private and can be presented upon formal requestethic.

Declarations

Conflict of interest The authors affirm that there is no conflict or competition of interest and all work is in compliance with ethical standards. Additionally, they assert that this research has not been published before, and is not under consideration for publication elsewhere.

This has been confirmed by all co-authors.

References

Ahmadi M, Khashei M (2021) Current status of hybrid structures in wind forecasting. Eng Appl Artif Intell 99:104133

Ahmadi F, Ghasemlounia R, Gharehbaghi A (2024) Machine learning approaches coupled with variational mode decomposition: a novel method for forecasting monthly reservoir inflows. Earth Sci Inf 17(1):745–760

Bai L, Han Z, Li Y, Ning S (2018) A hybrid de-noising algorithm for the gear transmission system based on CEEMDAN-PE-TFPF. Entropy 20(5):361

Band SS, Ardabili S, Mosavi A, Jun C, Khoshkam H, Moslehpour M (2022) Feasibility of soft computing techniques for estimating the long-term mean monthly wind speed. Energy Rep 8:638–648

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Chen G, Tang B, Zeng X, Zhou P, Kang P, Long H (2022) Short-term wind speed forecasting based on long short-term memory and improved BP neural network. Int J Electr Power Energy Syst 134:107365. https://doi.org/10.1016/j.ijepes.2021.107365

Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. ArXiv Preprint arXiv 14061078. https://doi.org/10.3115/v1/D14-1179

Fang Z, Wang Y, Peng L, Hong H (2021) Predicting flood susceptibility using LSTM neural networks. J Hydrol 594:125734

Gharehbaghi A, Ghasemlounia R (2022) Application of AI approaches to estimate discharge coefficient of novel kind of Sharp-Crested V-Notch weirs. J Irrig Drain Eng 148(3):p04022001

Gharehbaghi A, Ghasemlounia R, Ahmadi F, Albaji M (2022) Groundwater level prediction with meteorologically sensitive gated recurrent unit (GRU) neural networks. J Hydrol, 128262

Gharehbaghi A, Ghasemlounia R, Afaridegan E, Haghiabi A, Mandala V, Azamathulla HM, Parsaie A (2023a) A comparison of artificial intelligence approaches in predicting discharge coefficient of streamlined weirs. J Hydroinformatics 25(4):1513–1530

Gharehbaghi A, Ghasemlounia R, Latif SD, Haghiabi AH, Parsaie A (2023b) Application of data-driven models to predict the dimensions of flow separation zone. Environ Sci Pollut Res 30(24):65572–65586

Ghasemlounia R, Gharehbaghi A, Ahmadi F, Saadatnejadgharahassanlou H (2021) Developing a novel framework for forecasting groundwater level fluctuations using Bi-directional long Short-Term memory (BiLSTM) deep neural network. Comput Electron Agric 191:106568

Graves A, Schmidhuber J (2005) Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw 18(5–6):602–610

He Q, Wang J, Lu H (2018) A hybrid system for short-term wind speed forecasting. Appl Energy 226:756–771

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu W, Yang Q, Chen HP, Yuan Z, Li C, Shao S, Zhang J (2021) New hybrid approach for short-term wind speed predictions based on preprocessing algorithm and optimization theory, vol 179. Renewable Energy, pp 2174–2186

Huang NE, Shen Z, Long SR, Wu MC, Snin HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. A Math. Phys. Eng. Sci. https://doi.org/10.1098/rspa.1998.0193

Jiang P, Liu Z, Niu X, Zhang L (2021) A combined forecasting system based on statistical method, artificial neural networks, and deep learning methods for short-term wind speed forecasting. Energy 217:119361. https://doi.org/10.1016/j.energy.2020.119361

Kosana V, Teeparthi K, Madasthu S, Kumar S (2022) A novel reinforced online model selection using Q-learning technique for wind speed prediction. Sustain Energy Technol Assess 49:101780. https://doi.org/10.1016/j.seta.2021.101780

Lawrence S, Back AD, Tsoi AC, Giles CL (1997) On the distribution of performance from multiple neural network trials. IEEE Trans Neural Net 8(6):1507–1517

Li D, Jiang F, Chen M, Qian T (2022) Multi-step-ahead wind speed forecasting based on a hybrid decomposition method and Temporal convolutional networks. Energy 238:121981. https://doi.org/10.1016/j.energy.2021.121981

Lin H, Gharehbaghi A, Zhang Q, Band SS, Pai HT, Chau K-W, Mosavi A (2022) Time series-based groundwater level forecasting using gated recurrent unit deep neural networks. Eng Appl Comput Fluid Mech 16(1):1655–1672. https://doi.org/10.1080/19942060.2022.2104928

Liu H, Duan Z, Han FZ, Li YF (2018) Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm. Energy Conv Manag 156:525–541

López G, Arboleya P (2022) Short-term wind speed forecasting over complex terrain using linear regression models and multivariable LSTM and NARX networks in the Andes mountains, Ecuador, vol 183. Renewable Energy, pp 351–368

MATLAB User's Guide (2021a) The MathWorks Inc. (Deep Learning Toolbox). Natick, Massachusetts, United State; (2021). Computer Software. www.mathworks.com/

Nezhad MM, Heydari A, Pirshayan E, Groppi D, Garcia DA (2021) A novel forecasting model for wind speed assessment using Sentinel family satellites images and machine learning method. Renewable Energy 179:2198–2211. https://doi.org/10.1016/j.renene.2021.08.013

Parsaie A, Ghasemlounia R, Gharehbaghi A, Haghiabi A, Chadee AA, Nou MRG (2024) Novel hybrid intelligence predictive model based on successive variational mode decomposition algorithm for monthly runoff series. J Hydrol 634:131041

Qian Z, Pei Y, Zareipour H, Chen N (2019) A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. Appl Energy 235:939–953

Qu Z, Zhang K, Wang J, Zhang W, Leng W (2016) A hybrid model based on ensemble empirical mode decomposition and fruit fly optimization algorithm for wind speed forecasting. *Advances in Meteorology*, *2016*

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958

Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. BMC Bioinf 9(1):307. https://doi.org/10.1186/1471-2105-9-307

Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP (2003) Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci 43(6):1947–1958. https://doi.org/10.1021/ci034160g

Tian Z, Chen H (2021) Multi-step short-term wind speed prediction based on integrated multi-model fusion. Appl Energy 298:117248. https://doi.org/10.1016/j.apenergy.2021.117248

Torres ME, Colominas MA, Schlotthauer G, Flandrin P (2011) A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE Int. Conf. Acoust. Speech Signal Process. https://doi.org/10.1109/icassp. 2011.5947265

Wang C, Zhang H, Fan W, Ma P (2017a) A new chaotic time series hybrid prediction method of wind power based on EEMD-SE and full-parameters continued fraction. Energy 138:977–990

Wang D, Luo H, Grunder O, Lin Y (2017b) Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. Renewable Energy 113:1345–1358

Wang J, Li H, Wang Y, Lu H (2021) A hesitant fuzzy wind speed forecasting system with novel defuzzification method and multi-objective optimization algorithm. Expert Syst Appl 168:114364. https://doi.org/10.1016/j.eswa.2020.114364

Yang R, Liu H, Nikitas N, Duan Z, Li Y, Li Y (2022) Short-term wind speed forecasting using deep reinforcement learning with improved multiple error correction approach. Energy 239:122128

Yu C, Li Y, Zhang M (2017) Comparative study on three new hybrid models using Elman neural network and empirical mode decomposition based technologies improved by singular spectrum analysis for hour-ahead wind speed forecasting. Energy Conv Manag 147:75–85

Zhang W, Qu Z, Zhang K, Mao W, Ma Y, Fan X (2017) A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy Conv Manag 136:439–451

Zhang L, Wang J, Niu X, Liu Z (2021) Ensemble wind speed forecasting with multi-objective Archimedes optimization algorithm and sub-model selection. Appl Energy 301:117449