A hybrid model based on Discrete wavelet prediction (DWT), Bidirectional recurrent neural networks to wind speed prediction

Arezoo Barjasteh  
Kerman Branch, Islamic Azad University,

Seyyed Hamid Ghafoori (✉️ Ghafoori@iauk.ac.ir)  
Kerman Branch, Islamic Azad University,

Malihe Hashemipour  
Kerman Branch, Islamic Azad University,

Research Article

Keywords: Wind power prediction, deep learning, Non-stationarity, Discrete wavelet transformation (DWT), Gated Recurrent Unit (GRU) network

Posted Date: May 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2909112/v1

License: ☑️  This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

Additional Declarations: No competing interests reported.
A hybrid model based on Discrete wavelet prediction (DWT), Bidirectional recurrent neural networks to wind speed prediction

Arezoo Barjasteh
Department of Computer Engineering, Kerman Branch, Islamic Azad University, Kerman, Iran
Email: a.barjesteh@iauk.ac.ir

Seyyed Hamid Ghafouri*
Department of Computer Engineering, Kerman Branch, Islamic Azad University, Kerman, Iran
Email: Ghafoori@iauk.ac.ir

Malihe Hashemipour
Department of Computer Engineering, Kerman Branch, Islamic Azad University, Kerman, Iran
Email: M.hashemi@iauk.ac.ir

*Correspondence should be addressed to Seyyed Hamid Ghafouri
Department of Computer Engineering, Kerman Branch, Islamic Azad University, Kerman, Iran
76351-31167
Kerman, Iran
Tel: +98 (34) 3221-0051
Ghafoori@iauk.ac.ir
ABSTRACT

Wind speed is the main driver of wind power output, but its inherent fluctuations and deviations present significant challenges for power system security and power quality. Accurate short-term wind power forecasting is necessary to ensure the stability and integration of wind energy into the grid. Non-stationarity is a major challenge in analyzing wind speed data, and change-point detection are essential for optimal resource allocation. This paper addresses the issue of short-term wind power forecasting for stable and effective wind energy system operation. To predict non-stationary data and detect change points, non-stationary data must first be transformed into stationary data. Discrete wavelet transformation (DWT) is used to decompose wind speed traces into low- and high-frequency components for more accurate predictions using deep learning algorithms. The proposed approach uses a Gated Recurrent Unit (GRU) network, which has a concise network structure and requires less computational load, making it suitable for quickly predicting short-term and long-term dependencies in wind speed data. Experiments demonstrate that the proposed method outperforms other cutting-edge methods in terms of prediction accuracy.

Keywords: Wind power prediction, deep learning, Non-stationarity, Discrete wavelet transformation (DWT), Gated Recurrent Unit (GRU) network

1. INTRODUCTION

Along with the accelerated growth of contemporary industry, the degradation of the environment and the depletion of fossil fuels are becoming a growing concern. [1-3]. The search for clean and renewable energy is one of the most important issues in the world. Wind energy is garnering attention and wind energy production is expanding swiftly around the world [4, 5]. The wind is one of the most rapidly expanding energies due to its pure and unrestricted nature. However, naturally occurring fluctuations and variations in wind speed pose significant operational and organizing challenges for power systems [6]. Wind velocity is fundamental to wind energy production. However, when wind energy is coupled to the grid, its interruptions and fluctuations impact the equilibrium of electricity supply and demand, power system security, and power quality [7, 8]. Therefore, to guarantee the stability of the wind energy system and encourage the incorporation of wind energy on a large scale, it is essential to accurately foresee short-term wind power [9, 10]. wind power generation using Predicting wind speed over several periods enables energy planning techniques to preserve operational stability while enhancing the cost performance of the wind-powered microgrid [11–13]. Short-term wind prediction has a substantial impact on the generation of electricity and is also required for determining the scale of wind farms. An accurate wind prediction technique is required to decrease the cost of wind energy organization greatly [14, 15].

When the statistical characteristics of a time series remain constant over time, it is considered stationary time series. According to [16], stationary time series display periodic fluctuations and have a fixed correlation over time. When those requirements are not fulfilled in data, however, the time series exhibits non-stationary properties, which provides a significant challenge for many prediction applications [17]. Our investigation demonstrates that the wind speed data exhibit non-stationarity, with each location experiencing rapid fluctuations in wind speed and a mean that changes over time. This non-stationarity, characterized by varying wind speeds, can have a significant impact on prediction accuracy [18, 19]. Therefore, our findings strongly support the importance of change point detection in the decision-making process for optimal resource allocation.
To ensure accurate prediction of non-stationary data and detection of changepoints, it is necessary to convert the non-stationary data into stationary data. To achieve this, various techniques have been developed for analyzing non-stationary time series. These approaches are classified into two main types: temporal methods and spectral-temporal methods [20]. Spectral-temporal techniques are superior to temporal techniques in describing potential frequency shifts [21]. Two spectral-temporal methods are empirical mode decomposition (EMD) [22] and discrete wavelet transformation (DWT) [23]. Compared to other techniques used for analyzing time series data, DWT is a commonly used approach due to its superior performance in handling non-linear and non-stationary data, especially when compared to other decomposition techniques. DWT decompose a signal into its frequency components. In the case of wavelet transformation, time and signal frequency information are stored, making it a more powerful evolution for time-frequency analysis. The use of DWT is essential for achieving higher accuracy in wind speed prediction. To accurately forecast large-scale wind speeds, the prediction system must be able to handle multidimensional data and extract long-term dependencies. Additionally, it must be able to extract patterns from non-linear and non-stationary data with reasonable computational complexity. Traditional wind speed prediction methods often fail to extract patterns from non-linear and non-stationary data, leading to inefficient resource usage. Therefore, this work aims to improve resource efficiency by utilizing a deep-learning approach to predict wind speed.

The proposed procedure consists of two phases. Firstly, DWT [23–26] is employed to separate nonlinear and non-stationary wind speed sequences into multiple sub-bands with low- and high-frequency [27]. Therefore, multiple signal frequencies are generated in order to minimize signal-to-signal interaction. DWT decomposes wind speed traces into one approximate signal and some detailed signals. Second, each frequency signal is trained by the deep learning model, and finally, the trained model is applied to test values. Several deep learning algorithms for predicting linear and non-linear correlations in wind speed trace data have been introduced and have been widely used. One of the widely used deep neural networks for predicting future values using historical data is the recurrent neural network (RNN) [28]. Among the RNN models, the Long Short-Term Memory Model (LSTM) [29] and Gated-Recurrence Unit (GRU) [30] are the most popular ones, and they have been previously utilized for wind speed data prediction in various studies. We need deep learning algorithms with less computational load to deal with non-linear and non-stationary data. These algorithms need to find short-term and long-term dependencies. The GRU network is a more straightforward alternative to the LSTM network, with a compact network design that solves the issue of slow updating times in the LSTM. GRU outperforms LSTM because the average amount of data available for predictive model training is less [31]. Another advantage of GRU over LSTM is that GRU has fewer parameters than LSTM, allowing faster training and prediction with GRU [32]. Due to the rapid advent of requests, it is crucial to employ algorithms that can generate predictions rapidly. GRU is consequently preferable to other deep learning techniques.

RNNs are utilized for processing input sequences of varying lengths, which makes them ideal for time series forecasting. Nonetheless, many studies have demonstrated that stacking or combining multiple RNNs yields more precise predictions than a single LSTM network. As a result, we have selected an ensemble learning framework that incorporates LSTM and GRU to tackle these challenges. Although the two models are comparable, their data processing and calculation procedures differ, which may influence the weights when dealing with wind data. During our testing, we discovered that sometimes LSTM predictions were closer to the actual wind speed, while at other times, GRU predictions were superior. Certain temporal and spatial characteristics may affect the wind speed forecast more than others, and some factors may influence the wind speed for a brief period [33]. GRU outperforms LSTM in terms of parameter updating, CPU time convergence, and generalization [34]. LSTM is able to filter numerous features due to its controlled dissemination of memory content. Consequently, both models have advantages and disadvantages. In our setup, both models with distinct parameters are used to supplement each other and achieve the most accurate predicting results.
This study employs an extended version of LSTM and GRU models, namely bidirectional LSTM (BiLSTM) and bidirectional GRU (BiGRU) models, for wind speed prediction. Unlike LSTM and GRU, which only process inputs in a forward direction, BiLSTM and BiGRU process information in both forward and backward directions. These models utilize both forward and backward units and offer better performance than one-directional models by capturing the dependencies between past and future steps. The bidirectional models can automatically exploit the different significance levels of a wind speed sequence at various time points due to the varying impact of past steps on future wind speed. After signal decomposition with DWT, a hybrid deep learning model including BiLSTM and BiGRU is trained on each frequency sub-band, and the test values for each sub-band are predicted. In the end, the predicted values obtained from multiple signals are averaged to generate the final prediction outcomes. The outcomes indicate that our approach yields a lower prediction error when compared to other approaches. The contributions of this article can be summarized as follows:

1) A prediction model based on the combination of DWT and RNN models, including BiLSTM and BiGRU, is proposed for predicting linear and non-linear dependencies in wind speed data, significantly improving wind speed prediction accuracy. To the best of our knowledge, this is the first study that combines the DWT, BiLSTM, and BiGRU models to predict wind speed.

3) The DWT technique is used to extract patterns from non-stationary and non-linear data, allowing the BiLSTM and BiGRU models to learn both random and periodic fluctuations in the wind speed dataset.

3) The proposed method differs from other learning strategies that disregard the non-stationary character of wind speed data and the correlations between distinct sequences.

4) The real data is used to validate the hybrid model, and the experiment results show that the hybrid model exhibits a lower prediction error.

2-RELATED WORKS

Wind energy has shown high efficiency in the power system; therefore, an accurate and stable forecast of wind speed is of vital importance in the management of the power grid and the market economy. In this section, we will look at some wind speed prediction methods. A model for predicting wind speed was proposed in [36] that integrates EMD with novel RNNs and autoregressive integrated moving averages (ARIMA). EMD is employed to decompose the wind speed sequence in order to minimize the series' non-stationarity and complexity. Then, ARIMA is used for predicting the rest of the low-frequency subsequences and one residual, while LSTM is used for predicting the high-frequency sub-bands with high entropy. The final predictions are generated by combining the predictions of each sub-band. Combining statistical methods, noise processing, multi-objective optimization (MOO) techniques, and deep learning frameworks, The study [37] proposes a novel projection model. Utilizing the strengths of benchmark prediction models, this model addresses the nonlinear properties of wind speed series. Using wind speed data from three locations in China, the efficacy of this proposed model has been validated.

A new wind speed prediction method is suggested in [38] that combines integrated multi-model fusion and mode decomposition to enhance accuracy. Local mean decomposition and EMD are utilized to decompose wind speed into components. Odd and even sequences are extracted after each decomposition method to generate two new sequences. On the basis of the distinctive characteristics of the obtained components, the support vector machine (SVM) and stochastic configuration network are chosen as prediction models. In addition, SVM's parameters are optimized by the particle swarm optimization (PSO) algorithm. To obtain the final values, the predicted values of all decomposition components are implemented. A wind speed
prediction strategy for multi-wind farms is proposed in [39]. This method is applicable to a centralized control center for wind farms. This method entails pre-training multiple prediction models using historical data from different wind farms. Using transfer learning, these models are transferred to the central administration, and the wind speed of each farm is accurately predicted. The method applies the MOOFADA optimization algorithm for giving appropriate weights to different sets of prediction values in order to obtain optimal results.

For precise wind speed prediction, the study [40] presents a combined optimization model that utilizes a double-layer staged training network, variational mode decomposition (VMD), and a genetic algorithm (GA). The model preprocesses the original wind speed input with VMD and subsequently predicts each decomposed subsequence using the D-ESN model. Combining all of the subsequence predictions yields the ultimate prediction value. The study [41] proposes a serial-parallel dynamic network with dual dynamical features. The research develops a dynamic prediction model based on phase space reconstruction that incorporates the chaotic Coyote optimization algorithm and the enhanced complete ensemble EMD with the adaptive noise method. This method performs data preprocessing on the initial sequence, and the subsequence generated from decomposition is fed into the model for prediction. The study [42] presents a new RNN known as the time-frequency RNN (TFR) to improve the accuracy of short-term wind speed predictions. TFR design and DWT have been integrated to extract the time-frequency features. Then convolution processes are combined to improve the accuracy of the model. Using actual wind speed data from a monitoring site, the training time, the ability to predict, and the parameter sensitivity of the models are evaluated. A wind speed combined probability prediction model was proposed in [43] that uses the concept of quantile and integrates data denoising technology to build prediction components. A novel multi-objective marine predator for ensembling the prediction elements is created to overcome the limitations of conventional MOO algorithms. The study [44] proposed wind speed predictions by using a Legendre multiwavelet-based neural network (LMWNN) model. The model integrates the self-learning potential of neural networks with the exceptional properties of LMWNN. The model distributes weights within subintervals and learns input-output data pairs, thereby significantly reducing computing costs.

3- Backgrounds

In this portion, the instruments utilized in the research are discussed, namely DWT, LSTM, GRU, BiLSTM, and BiGRU.

3-1 Discrete wavelet transform (DWT)

DWT reveals various features of non-stationary signals, including trends, discontinuities, and repetitive patterns in signal analysis [17]. Due to the dynamic nature of wind speed data, the input data comprises components with various frequencies. DWT decomposes the input data into multiple low and high-frequency sub-bands, while the distributions of sub-bands are distinct [18]. BiGRU and BiLSTM are utilized to independently train each component of the model to improve its training performance. Consequently, a suitable decomposition algorithm divides the wind speed trace data into multiple signals with low and high- frequencies.

The DWT is a useful signal processing technique that works well with non-stationary and nonlinear signals [23]. The wavelet outputs capture the essential characteristics of the original signal and can help identify trends of non-stationary wind speed time series. To decompose the wind speed traces, Mallat's algorithm [24] is utilized. This algorithm employs high and low-pass filters to divide non-stationary wind speed traces into a number of sequences with varying frequencies. $dA$ and $dD$ in Equations (1) and (2) represented by [24] indicate the approximate and detail coefficients for the outputs of low and high-pass filters.
\[ dA = \sum_{k=-\infty}^{+\infty} X[k] \varphi_l[2n-k], \quad (1) \]
\[ dD = \sum_{k=-\infty}^{+\infty} X[k] \varphi_h[2n-k], \quad (2) \]

\( X \) is the original signal for wind speed traces, \( \varphi \) indicate the filter, \( l \) and \( h \) are the low and the high-pass filter. Figure 1 depicts the three-layer decomposition procedure. In the first layer, the original wind speed data \( X \) are decomposed by utilizing low and high-pass filters. The procedure yields sub-bands \( d1 \) and \( a1 \) that are both precise and approximative. \( a2 \) and \( d2 \) coefficients are obtained by passing the \( d1 \) sub-band via two low and high-pass filters In the second layer. In a comparable manner, the obtained \( d2 \) sub-band is passed by applying two low and high-pass filters in the third layer to generate two coefficients, \( a3 \), and \( d3 \). After the process of decomposition, various frequency-specific components can be obtained. After decomposition, the resulting pieces have varying lengths. Using the inverse DWT(IDWT) and approximative and specific coefficients, the data is reconstructed so that all sequences have the same length.

![Diagram of DWT with labels](image)

**Figure 1: A schematic diagram of DWT**

**3-2-LSTM MODEL**

The RNN models are unable to acquire long-term dependencies because of the well-known gradient vanishing problem [32]. This issue is resolved by LSTM, which often outperforms standard RNNs, by storing important information in the memory unit and removing unnecessary information. The principal procedure consists of substituting an LSTM block for the central layer of the RNN. LSTM is notorious for learning long-term dependencies, which RNNs cannot learn [32]. To predict subsequent time steps, it is necessary to revise the network’s weight values, which requires knowledge of the initial time steps. LSTM has been especially effective in the time-series domain due to its ability to store essential past information in the cell state and disregard irrelevant information [29]. LSTM uses three different gates to accomplish these complicated responsibilities. The first gate is the forget gate layer, which is responsible for determining which data will be removed from the cell's state. The LSTM gates equation is denoted by the following:

\[ f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3) \]
\[ i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (4) \]
\[ \tilde{c}_t = \tanh(W_c [h_{t-1}, x_t] + b_c) \]  
\[ a_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \]

The notation used in LSTM includes several variables, such as \( f_t, \sigma, W_f, h_{t-1}, x_t \), and \( b_f \). The forget gate \( f_t \) is determined by the sigmoid function \( \sigma \) and the weights \( W_f \) for forgetting neurons. The previous output \( h_{t-1} \) and the current input \( x_t \) are also used in the calculation. There are two steps in deciding what information to store in the cell state: the input gate layer and the \( \tanh \) function. The input gate is represented by it and is determined by the weights \( W_I \) for input neurons and the candidate for the cell state, \( \tilde{c}_t \), as well as the biases \( b_i \) and \( b_C \). Finally, the output gate layer \( o_t \) determines what information is passed to the output. \( b_o \) and \( W_o \) indicates the biases for the output gates and the output weights.

### 3-3 GRU MODEL

The GRU model is a shortened variant of the LSTM model that seeks to construct a faster and less complex recursive network. This is accomplished by integrating the forget and input gates into a single update gate, which eliminates one of the LSTM gates and differentiates the structure of the LSTM gates. As depicted in Figure 2, GRU consists of reset gates and update gates. Compared to LSTM, GRU is more computationally effective because it does not require a memory unit and also functions better when working with small datasets [32].

The update gate decides how much past information should be updated before passing to the next step. The gates of the GRU are defined as follows:

\[ z_t = \sigma(W_z x_t + U_z h_{t-1}) \]  
\[ r_t = \sigma(W_r x_t + U_r h_{t-1}) \]

The update gate is denoted by \( z_t \) at time step \( t \) and is determined by the input \( x_t \), holding values \( h_{t-1} \) for the previous \( t - 1 \) units, and corresponding weights \( W_z \) and \( U_z \). The reset gate, represented by \( r_t \), determines how much past information should be forgotten, and is determined by weights \( W_r \). The update and reset gates are used to store and filter information, and they effectively address the problem of vanishing gradients in RNNs. This allows relevant information to be stored in the memory cell and passed on to subsequent time steps in the network.

![GRU gate structure](image-url)
To capture the impact of both past and future records on the present record in a time series, the BiLSTM and BiGRU models are employed. BiLSTM contains two LSTM units, while BiGRU consists of two GRU units for forward and backward propagation [17]. Thus, during the training phase, both forward and backward models can learn using both past and future observations, resulting in a better comprehension of current data, as well as a reduction in errors and an increase in accuracy throughout the prediction phase. Both models possess two hidden layers that are connected in opposing directions while sharing an output layer that receives input from both past and future values. The structures of BiLSTM and BiGRU are illustrated in Figures 3 and 4, respectively.

The BiLSTM and BiGRU models consist of two major components that extract features for predicting wind speed over time. The forward component updates hidden states based on past wind speed data, and the backward component updates hidden states relying on future wind speed data. Employing bidirectional models permits the extraction of features that describe the interaction between input entities. The BiLSTM and BiGRU models are intended to significantly utilize incoming data and employ both forward and backward computation. For further analysis, extracted features feed the fully connected layer. Then, future wind speed is generated by using the output layer.

4-PROPOSED METHOD

There are two main reasons for using group learning with more than one interrelated model. Therefore, group learning leads to improved accuracy. A learning ensemble can make better predictions and perform better than any adjunctive model. Also, the robustness of a group learning model is higher than that of single models because group learning reduces the spread or dispersion of predictions and model performance. Ensemble learning is the concept of combining various forms of machine learning and deep learning models for prediction or classification. Our study presents a new approach to predicting wind speed.
by combining DW, BiGRU, and BiLSTM techniques. This innovative strategy combines the DWT data filtering method with BiLSTM and BiGRU networks to enhance the accuracy and efficiency of wind speed prediction. By segmenting the time series data into multiple predictable components with less non-stationarity, the DWT method notably improves prediction accuracy. The paper also includes a pseudocode of the proposed method in Figure 5, and the DWT method is used to decompose wind speed data into several sub-bands. This method distinguishes the high-frequency component of the main wind speed signal resulting from random shifts in load and oscillations, thereby resolving the problem of unpredictability in the wind speed sequence. The original wind speed data $X$ undergo decomposition through low and high-pass filters, resulting in three detailed sub-bands ($d_1$, $d_2$, and $d_3$) and one approximate sub-band ($a_1$) as illustrated in Figure 5. Using the sub-band values directly for training BiGRU is not appropriate since the values of the detailed and approximate sub-bands are different, and the Z-score method is used to normalize wind speed data.

The proposed method is illustrated in Figure 5. Patterns from nonlinear and non-stationary data is extracted by DWT. Afterward, BiLSTM and BiGRU learn random and periodic fluctuations-related features in the wind speed dataset. The prediction model comprises two elements: a BiLSTM or BiGRU and an output layer. In each iteration, BiLSTM and BiGRU receive the wind speed sequence data and generate predicted results for the output layer. The output layer provides final values for wind speed prediction. BiLSTM and BiGRU are employed to investigate dependencies, discover the hidden patterns of input data, and derive temporal features from wind speed data.

The study uses a hybrid set model and experiments with various configuration combinations to choose parameters such as the number of layers and neurons, and epoch size. The ensemble model consists of two levels: sub-model 1 is the LSTM, and sub-model 2 is the GRU model, both of which are used in the first level. The dataset is divided into training, validation, and test sets, and each set is used to train the blending ensemble model. The BiLSTM and BiGRU models are trained on the training data. After obtaining the LSTM and GRU validation predictions, they are combined into a new training dataset in the form of $p \times m$ (where $p$ represents the number of predictions and $m$ represents the number of models). After collecting the prediction results from the sub-models, they are combined to create a new training dataset for the meta-learner, also referred to as the second-level model. The meta-learner is a three-layer neural network with a
ReLu activation function and full connectivity. After the meta-learner has been trained, the test dataset is input back into the sub-models in order to produce intermediate test data for the meta-learner. Finally, the meta-learner utilizes these intermediate test predictions to generate the final wind speed predictions.

5- EXPERIMENTS AND RESULTS

4-1) Data Set

Two actual wind speed data are used to evaluate the performance of the proposed prediction method. The first is the wind speed data in the Sotaventogalicia site [45], and the second is the Kerman site dataset [46].

**Sotaventogalicia site**

Sotavento is located in the southwestern part of Europe, in Galicia, Spain, in "A Serra da Loba," with coordinates of 43.354377°N and 7.881213°W. Although the wind speed at the Sotaventogalicia wind farm is monitored every ten minutes, the hourly wind speed is used in this analysis. This study used hourly wind speed data between January 1, 2020, and February 30, 2022, as the dataset for short-term wind speed forecasts. 80% of the total samples are utilized for train sets, while 20% are used as test sets.

**Kerman site**

Kerman is located in the center of Iran in the middle east, at 30.2839° N and 57.0834° E. In this instance, hourly wind speed data from 2009 was used as the short-term wind speed prediction dataset. Similar to the Sotaventogalicia dataset, 80% of the entire 1200 samples are utilized for a train set, and 20% are used as test sets. The variable wind speed in this case study makes it challenging to make accurate forecasts.

4-2) Configuration

Prior to the training process, z-normalization is used for preprocessing the input data. The chosen parameters determined by the validation set can be found in Table 1. Using the grid search method, two parameters, window length (|w|) and batch size, were selected. For |w|, a network search was conducted on the range of \{12, 20, 28, 36, 52, 56, 60, 72\}, and it was found that the optimal performance was achieved at 28 based on the validation set. We also do a network search for a batch size of \{16, 32, 64, 128, 256\}, and the value of 32 shows the best performance and is used. Since the proposed model may depend on the long term, the |w| is set to 56. This number is much smaller than the found value of the optimal batch size parameters of LSTM and increases the speed of the proposed process.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w</td>
</tr>
<tr>
<td>Input layer size</td>
<td>48/64</td>
</tr>
<tr>
<td>Epoch</td>
<td>100</td>
</tr>
<tr>
<td>Batch size</td>
<td>32</td>
</tr>
<tr>
<td>Hidden layer size</td>
<td>256</td>
</tr>
</tbody>
</table>

4-3) Evaluation Metrics

The performance evaluation in this study uses three criteria: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). RMSE measures the square root of the average of the squared differences between predicted and actual values, normalized by the number of predictions.
\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \]  \hspace{1cm} (9)

\[ RMSE = \left[ \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right]^{\frac{1}{2}} \]  \hspace{1cm} (10)

\[ MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{|y_i|} \]  \hspace{1cm} (11)

The predicted value is represented by \( y_i \), while the actual value is represented by \( \hat{y}_i \) and \( n \) represents the test size.

4-4) Evaluation

The Keras library and Tensorflow framework were utilized to develop the proposed model in Python programming language. The experiments were performed on a computer equipped with an Intel Core i7-6500 CPU and Nvidia GeForce GTX 1080 graphics card.

<table>
<thead>
<tr>
<th>Category</th>
<th>name</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline methods</td>
<td>SVR</td>
<td>SVR is used for predicting wind speed using the wind speed time series</td>
</tr>
<tr>
<td></td>
<td>BPNN</td>
<td>The backpropagation neural network (BPNN) is used for predicting wind speed</td>
</tr>
<tr>
<td></td>
<td>LSTM</td>
<td>The model uses LSTM, a type of recurrent neural network, to make predictions</td>
</tr>
<tr>
<td></td>
<td>GRU</td>
<td>The model uses GRU, a type of recurrent neural network, to make predictions</td>
</tr>
<tr>
<td></td>
<td>BiLSTM</td>
<td>The wind speed time series is used by the model which utilizes BiLSTM to</td>
</tr>
<tr>
<td></td>
<td>BiGRU</td>
<td>make predictions about the wind speed.</td>
</tr>
<tr>
<td>DWT-based</td>
<td>DWT-SVR</td>
<td>DWT is used for decomposing data and SVR is used for predicting wind speed</td>
</tr>
<tr>
<td>methods</td>
<td>DWT-BPNN</td>
<td>DWT is used for decomposing data and BPNN is used for predicting wind speed</td>
</tr>
<tr>
<td></td>
<td>DWT-LSTM</td>
<td>DWT is used for decomposing data and LSTM is used for predicting wind speed</td>
</tr>
<tr>
<td></td>
<td>DWT-GRU</td>
<td>DWT is used for decomposing data and GRU is used for predicting wind speed</td>
</tr>
<tr>
<td></td>
<td>DWT-BiLSTM</td>
<td>DWT is used for decomposing data and BiLSTM is used for predicting wind speed</td>
</tr>
<tr>
<td></td>
<td>DWT-BiGRU</td>
<td>DWT is used for decomposing data and BiGRU is used for predicting wind speed</td>
</tr>
<tr>
<td></td>
<td>DWT-BiLSTM-BiGRU</td>
<td>DWT is used for decomposing data and combination of BiLSTM and BiGRU is</td>
</tr>
</tbody>
</table>

Table 2 provides a comprehensive overview of the different baseline models used in the study. The performance of our proposed method, named DWT-BiLSTM-BiGRU, is evaluated against twelve algorithms, including DWT-BiGRU, DWT-BiLSTM, DWT-GRU, DWT-LSTM, DWT-BPNN, DWT-
SVR, BiGRU, BiLSTM, LSTM, GRU, BPNN, and SVR. In our analysis, the comparison algorithms are divided into two sets. The first set comprises baseline algorithms and the second set comprises prediction algorithms that utilize the DWT method for decomposing wind speed data traces.

4-4-1) Sotaventogalicia site dataset

The dataset from the Sotaventogalicia site was utilized to perform the prediction, and Table 3 presents the results for MAPE, RMSE, and MAE. The best values for each metric are indicated in bold in the table.

<table>
<thead>
<tr>
<th></th>
<th>MAPE</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>28.56</td>
<td>2.67</td>
<td>1.93</td>
</tr>
<tr>
<td>BPNN</td>
<td>18.31</td>
<td>1.37</td>
<td>1.54</td>
</tr>
<tr>
<td>LSTM</td>
<td>10.09</td>
<td>0.92</td>
<td>0.67</td>
</tr>
<tr>
<td>GRU</td>
<td>9.87</td>
<td>0.88</td>
<td>0.64</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>9.56</td>
<td>0.85</td>
<td>0.61</td>
</tr>
<tr>
<td>BiGRU</td>
<td>8.98</td>
<td>0.84</td>
<td>0.57</td>
</tr>
<tr>
<td>DWT-SVR</td>
<td>11.42</td>
<td>1.08</td>
<td>0.95</td>
</tr>
<tr>
<td>DWT-BPNN</td>
<td>9.37</td>
<td>0.98</td>
<td>0.78</td>
</tr>
<tr>
<td>DWT-LSTM</td>
<td>4.12</td>
<td>0.43</td>
<td>0.31</td>
</tr>
<tr>
<td>DWT-GRU</td>
<td>3.78</td>
<td>0.39</td>
<td>0.28</td>
</tr>
<tr>
<td>DWT-BiLSTM</td>
<td>3.37</td>
<td>0.34</td>
<td>0.26</td>
</tr>
<tr>
<td>DWT-BiGRU</td>
<td>3.17</td>
<td>0.30</td>
<td>0.23</td>
</tr>
<tr>
<td>DWT-BiLSTM-BiGRU</td>
<td>2.81</td>
<td>0.25</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 3 shows the MAPE, RMSE, and MAE results of the prediction operations done with the Sotaventogalicia site dataset. The results show that DWT-BiLSTM-BiGRU outperforms the other models examined. The significantly reduced error of DWT-BiLSTM-BiGRU compared to other prediction algorithms is attributed to bidirectional learning and the utilization of DWT. Signal decomposition using DWT permits prediction models to recognize non-stationary characteristics and to capture the nonlinear and dynamic nature of wind speed data. The forward and backward learning mechanism permits the model to examine and extract temporal dependencies in historical data for weight adjustment. Bidirectional learning enables the proposed model to fully understand the relationship between distinct historical data points during the training phase by learning from both the past and the future. Figure 6 compares real and predicted wind speed data for DWT-based and baseline approaches using the Sotaventogalicia site dataset.
To facilitate the comparison of our approach with other methods, we normalized the evaluation metrics criteria values. The normalized values for different evaluation metrics are shown in Figure 7. The results indicate that our method performs better than the baseline approaches in terms of estimating CPU consumption in the Sotaventogalicia site test set, based on MAE and RMSE criteria. Additionally, a boxplot in Figure 7 illustrates the absolute wind speed prediction error for prediction models, using the Sotaventogalicia site data. The results show that the DWT-BiLSTM-BiGRU approach outperforms the other methods in terms of reducing absolute error.

Figure 6: Comparing the real and predicted Wind speed for DWT-based and baseline techniques with the Sotaventogalicia site dataset

Figure 7: the absolute error boxplot visualization for each wind speed prediction method using the Sotaventogalicia site dataset

4-4-2) Kerman site dataset
In this section, we present the results of our proposed method applied to wind speed data from the Kerman site dataset. Table 4 shows the prediction results in terms of various metrics. The results in Table 4 demonstrate that our method outperforms the other techniques in predicting wind speed for different wind speed patterns, with the lowest error rate. The results of the Kerman site dataset experiments show that the DWT-BiLSTM-BiGRU approach has better prediction accuracy than other compared methods. According to Table 4, the prediction accuracy of the DWT-BiLSTM-BiGRU approach is higher compared to those of other techniques.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAP E</th>
<th>RMS E</th>
<th>MA E</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>25.32</td>
<td>1.76</td>
<td>1.04</td>
</tr>
<tr>
<td>BPNN</td>
<td>15.41</td>
<td>1.05</td>
<td>0.82</td>
</tr>
<tr>
<td>LSTM</td>
<td>6.46</td>
<td>0.26</td>
<td>0.23</td>
</tr>
<tr>
<td>GRU</td>
<td>6.73</td>
<td>0.30</td>
<td>0.27</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>4.76</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>BiGRU</td>
<td>5.01</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td>DWT-SVR</td>
<td>9.78</td>
<td>0.51</td>
<td>0.39</td>
</tr>
<tr>
<td>DWT-BPNN</td>
<td>7.37</td>
<td>0.41</td>
<td>0.31</td>
</tr>
<tr>
<td>DWT-LSTM</td>
<td>3.54</td>
<td>0.17</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 4: The prediction results for the Kerman site Dataset
Table 5 demonstrates that DWT-BiLSTM-BiGRU has superior prediction accuracy compared to the other methods evaluated. Comparing DWT-BiLSTM-BiGRU to other methods of prediction demonstrates that the DWT-BiLSTM-BiGRU model provides substantially smaller MAPE, RMSE, and MAE values. Figure 8 depicts the comparison between actual and predicted wind speeds for both DWT-based and baseline methods using the Kerman site dataset.
Figure 8: Comparing the real and predicted Wind speed for DWT-based and baseline techniques with Kerman site data

The figures presented in this section demonstrate that the DWT-BiLSTM-BiGRU model outperforms other models in terms of prediction accuracy. Using a combination of DWT, BiLSTM, and BiGRU techniques, the model can detect temporal dependencies in wind speed data by examining past data in both backward and forward directions. Our model extracts non-linear features from future and historical wind speed data and achieves strong capability in comparison to other methods. The application of DWT distinguishes between high and low frequencies in the wind speed data, resulting in more accurate predictions, particularly when there are wind speed fluctuations. The BiGRU architecture with two hidden layers connected in opposing directions to a single output layer results in an accurate wind speed prediction. Overall, our proposed methodology achieves higher prediction accuracy than previous methods, particularly when load fluctuations occur. The comparison of our method to other approaches is illustrated in Figure 9, which shows the normalized values of MAPE, RMSE, and MAE. The results of Figure 8 demonstrate that DWT-BiLSTM-BiGRU performs better in predicting wind speed in the Kerman site dataset compared to traditional methods. Figure 9 also visualizes the absolute prediction error boxplot of the of prediction methods using the Kerman site dataset. The results of Figure 9 demonstrate that DWT-BiLSTM-BiGRU significantly outperforms other approaches in reducing absolute error.
Summarying the analysis results from Tables 4 and 5 and Figures 6–9, the following observations can be made:

- The conventional prediction techniques, SVR and BPNN, have the largest prediction errors when contrasted with all of the assessment measures. The results of Tables 4 and 5 and Figures 6–9 indicate that SVR and BPNN are unable to derive both linear and nonlinear characteristics from wind speed data. The poor capability of SVR and BPNN in extracting temporal features results in higher prediction errors compared to deep learning models. BiGRU and LSTM surpass SVR and BPNN at predicting future performance. Comparing BiLSTM and BiGRU with various baseline methods, it is evident that BiGRU and BiLSTM by incorporating bidirectional learning significantly reduce MAPE, MAE, and RMSE prediction errors.

- The contrast between DWT-based methods and those not using DWT revealed that the inclusion of DWT increases the precision of wind speed predictions. Evaluation metrics values for DWT-BiGRU and DWT-LSTM had been substantially lower than those of BiGRU and BiLSTM, according to the results. DWT reduced the prediction error values for both sets of data for all prediction models. By learning short- and long-term dependencies, DWT extracts temporal relationships between wind speed data records and diminishing significant fluctuations, thereby producing more accurate predictions. By understanding the nonlinear and dynamic nature of wind speed data and decomposing non-stationary data, it became apparent that DWT enhanced the accuracy of wind speed prediction.

- Figures 6 and 8 demonstrate how the bidirectional learning approach contributes to increasing the prediction accuracy. Comparing the accuracy of DWT-BiLSTM and DWT-LSTM and DWT-BiGRU and DWT-GRU reveals that the bidirectional-based models have significantly lower MAPE, MRSE, and MAE for all datasets. Bidirectional learning assigns different weights to the wind speed at different points in time and takes into account the variable effects of predicting future wind speed.

- The hybrid model, which combines both BiGRU and BiLSTM, shows better performance in terms of MAPE, MAE, and RMSE than either of the individual models. When compared to DWT-BiLSTM and
DWT-BiGRU, DWT-BiLSTM-BiGRU produces significantly lower prediction error rates for all three criteria. The lower absolute error rate of DWT-BiLSTM-BiGRU, as shown in Figures 7 and 9, is due to its superior performance in capturing long-term dependencies.

5- CONCLUSION

This study presents a method for predicting wind speed that incorporates the DWT, BiLSTM, and BiGRU models. The non-stationary and nonlinear wind speed data are decomposed into predictable sub-bands using DWT, while the BiLSTM and BiGRU models are used to discover long-term dependencies in order to predict future wind speed. The proposed approach was evaluated through the Sotaventogalicia and Kerman real-time wind speed datasets. The results demonstrate that the basic techniques fail to learn nonlinear and non-stationary data, particularly when random fluctuations occur. However, the proposed method achieves better results than other baseline methods in both datasets. Exploring the use of deep reinforcement learning techniques by investigating the impact of including other meteorological variables, such as temperature, humidity, and pressure, could also be a promising direction for future research.

References


45. [dataset] [56] https://www.sotaventogalicia.com/en/technical-area/monitored-data/