Short-term Spatio-temporal Wind Speed Prediction Based on VMD and Attention Mechanism

Yingnan Zhao1, Guanlan Ji2, Fei Chen3, Peiyuan Ji4

1Nanjing University of Information Science and Technology, Nanjing, 210044, China

Email: zh\_yingnan@163.com

**Abstract:** Improving short-term wind speed prediction accuracy and stability is still challenging and plagues wind forecasting researchers. In this paper, leveraging both temporal and spatial correlations of wind speed, we propose a new VMD-Attention-based Spatio-temporal network (VASTN) method. VASTN is a hybrid prediction model of wind speed that integrates variational mode decomposition (VMD), Squeeze-and-Excitation Network (SENet), and attention mechanism (AM) based Bidirectional long short-term memory (BiLSTM). Initially, VASTN applies VMD to decompose the wind speed matrix into a series of Intrinsic Mode Functions (IMF). Each IMF uses an improved CNN algorithm based on channel AM, also known as SENet, to extract the spatial features at the bottom of the model. Secondly, it combines BiLSTM and AM at the top layer to extract the aggregated spatial features and capture the temporal dependencies. Finally, VASTN accumulates the prediction results of each IMF to obtain the predicted wind speed. This method applies VMD to reduce the randomness and instability of the original data, and then it introduces AM to increase the prediction accuracy by mapping weight and parameter learning. The results of experiments on real-world data illustrate the superiority of VASTN compared with previous related algorithms.

**Keywords:** short-term wind speed prediction; VMD; Attention Mechanism; SENet; BiLSTM

# Introduction

As one of the most environmentally friendly renewable resources, wind energy has obvious advantages such as cleanliness, low cost, and sustainability. It has become the mainstream of current renewable energy sources and is developing rapidly worldwide [1]. However, wind power's intermittency, randomness, and instability pose challenges to wind power generation systems [2].

Accurate wind speed prediction is significant to the configuration, scheduling, maintenance, and planning of wind energy conversion systems [3]. There are four main current mainstream wind speed algorithms: 1) Physical methods, 2) Statistical methods, 3) Spatial correlation methods, 4) Artificial intelligence (AI) methods. Physical method, e.g., numerical weather prediction (NWP), is to establish a model for physical or meteorological information such as temperature, air pressure, humidity, topography, and air density in the area to predict wind speed for a long time [4]. However, this method requires complex calculations, and the performance is not very good in short-term wind speed forecasts[5]. Statistical methods adopt mathematical equations to predict wind speed based on a large amount of historical data [6]. Its representative methods include autoregressive moving average model (ARMA), moving average model (MA), autoregressive model (AR), and autoregressive integrated moving average model (ARIMA) [7]. Spatial correlation methods focus on the interaction of wind speed at different wind speed observation sites to predict [8]. AI methods do not rely on the precise model of the object. By using historical data, they learn the relationships between inputs and outputs. Therefore, they are suitable for random nonlinear systems. With the development of deep learning, algorithms such as multi-layer perception (MLP) [9], recurrent neural network (RNN) [10], convolutional neural network (CNN) [11], gated recurrent unit (GRU), long short-term memory (LSTM), etc., have been employed for short-term wind speed prediction

Traditional artificial intelligence methods are easy to implement and have a wide range of adaptations. However, conventional artificial intelligence methods still have room for improvement due to wind speed's randomness and intermittent characteristics. In this regard, the academic community has proposed a hybrid prediction model based on modal decomposition methods to obtain relatively stable wind velocity sub-sequences, such as empirical mode decomposition (EMD) that is good at processing nonlinear non-stationary signals [12]; Ensemble empirical mode decomposition (EEMD) is an improved method for EMD mixing phenomenon [13]; CEEMD (Complete Ensemble Empirical mode decomposition) is a further improvement based on EEMD, which makes up for the unclean noise removal of EEMD problem[14]; VMD method is robust to sampling and noise, and overcomes the shortcomings of EMD to a certain extent. Comparative studies show that the prediction effect of using the VMD hybrid model is better than that of a single prediction model [15-17], so this paper uses VMD to decompose wind speed series.

At present, most researchers pay attention to the selection of input variables in the prediction process while ignoring the different impact levels of various features extracted from the model on the output. In addition, most of the AI wind speed prediction methods only focus on extracting temporal features. In recent years, research on the capture of spatial features has gradually become popular. Therefore, this paper proposes a new method of predicting wind speed based on the attention mechanism of CNN and the BiLSTM deep learning network based on VMD decomposition. First, the spatio-temporal wind speed of the wind farm is used as input data to construct a CNN architecture, SENet, that uses the attention mechanism to weigh the channel features. While extracting the spatial features, it can pay more attention to the effectiveness and reduce the impact of useless features. Then it transfers the data containing spatial features to the BiLSTM-Attention layer to extract temporal features. Different weights are assigned to the hidden state of BiLSTM through mapping weights and parameter learning to strengthen the influence of important information. This method alleviates the problem of insufficient accuracy of the existing wind speed prediction AI model.

The rest of the paper is as follows: Section 2 gives the basic principles of the algorithm framework involved; Section 3 illustrates a detailed introduction to the VASTN network architecture; Section 4 discusses the experimental results and compares them with other typical algorithms; Section 5 summarizes the entire paper.

# Background theories

## Variational mode decomposition

VMD is a new signal decomposition technology proposed by Konstantin Dragomiretskiy and Dominique Zosso [18], which is mainly applied to decompose the input signal into discrete numbers of sub-signals called modes. VMD decomposes the original signal into a bandwidth with a center pulsation according to the specified number of modes. Then, it uses the alternate direction multiplier method (ADMM) to update each mode and its corresponding center pulsation, gradually demodulating each mode to the corresponding baseband. Finally, it extracts each mode and the related center pulsation together.

Assuming that the data to be processed is, each mode is and its center frequency is . The specific process of VMD is as follows:

1) Calculate the analytic signal related to each mode through the Hilbert transform to obtain the unilateral spectrum.

2) Through exponential mixing of the analytical signal of each mode and the corresponding center frequency, the frequency spectrum of each mode is modulated to the baseband.

3) The signal is demodulated according to Gaussian smoothness to obtain the gradient square -norm. Its variational constraint model is the following:

In the formula, the number of modes to be decomposed is , corresponds to the -th mode and is its center pulsation respectively.

4) Introduce the quadratic penalty function and the Lagrange multiplier to obtain the optimal solution of the variational constrained model mentioned above. Transforming the constrained variational problem into a non-constrained variational problem, and the resulting Lagrangian function is thereby as follows:

5) Finally, the ADMM method is used to solve the unconstrained variational problem, and and are continuously updated during the solution process. Finally, each can complete the frequency band division according to the frequency domain characteristics of , and realize the adaptive decomposition of the signal. The new and of the solution are as follows:

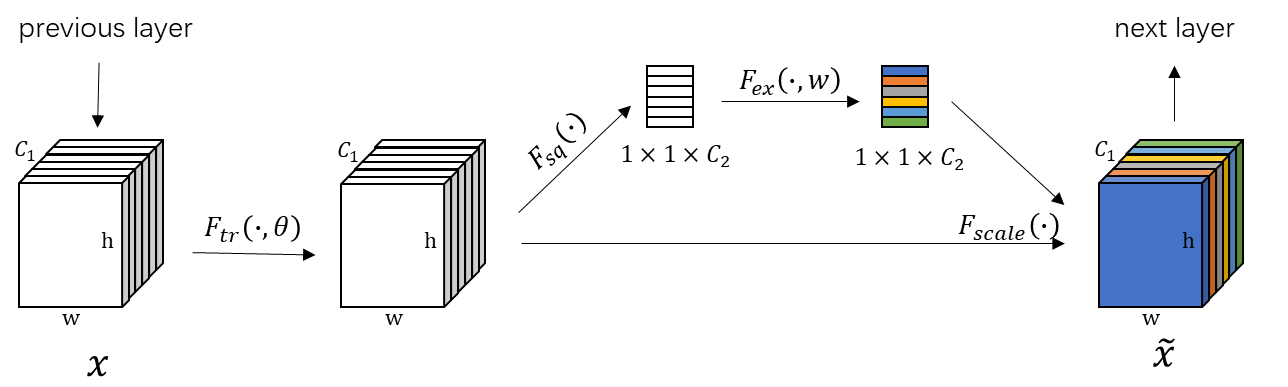
and do not stop updating until they satisfy the following requirement:

is a given accuracy that is greater than 0.

## SENet

Squeeze-and-Excitation Networks (SENet), proposed by Jie H and his team is a CNN-based attention mechanism that learns different channels' features. SENet term the "Squeeze-and-Excitation" (SE) block of learning the relationship between the channels of the CNN convolution kernel. It utilizes the channel attention mechanism to recalibrate channel-wise feature responses by explicitly modeling the interdependencies adaptively. At last, SENet combines SE block with the general CNN network to form the SENet architecture [19].

CNN is a feedforward neural network that uses convolutional calculations and has a deep structure [20]. As one of the representative models of deep learning, it employs trainable convolution kernels, local pooling operations, and fully connected operations to alternately apply the results of forwarding propagation and backpropagation to the raw input to extract the spatial features of the original input data [21]. Multiple fields have widely applied CNN to extract spatial features due to its excellent performance, e.g., machine vision [22], voice recognition [23], text processing [24], etc.



**Figure 1:** A typical squeeze and excitation block.

Figure 1 illustrates the structure of an SE block, where is a convolution operation, and its input and output are as the following:

The formula of is defined as follows, where represents the -th 2D matrix in 3D matrix , is a 2D spatial kernel that acts on the corresponding channel of , it represents a single channel of .

SENet proposes a *squeeze* step to alleviate the problem of not using the correlation between different channels. A global average pooling is employed to compress the spatial features to focus on the channel information. The formula (8) converts the input of to the output of

The *excitation* operation is to capture the dependencies of the channels fully. It makes use of a simple gating mechanism with sigmoid activation:

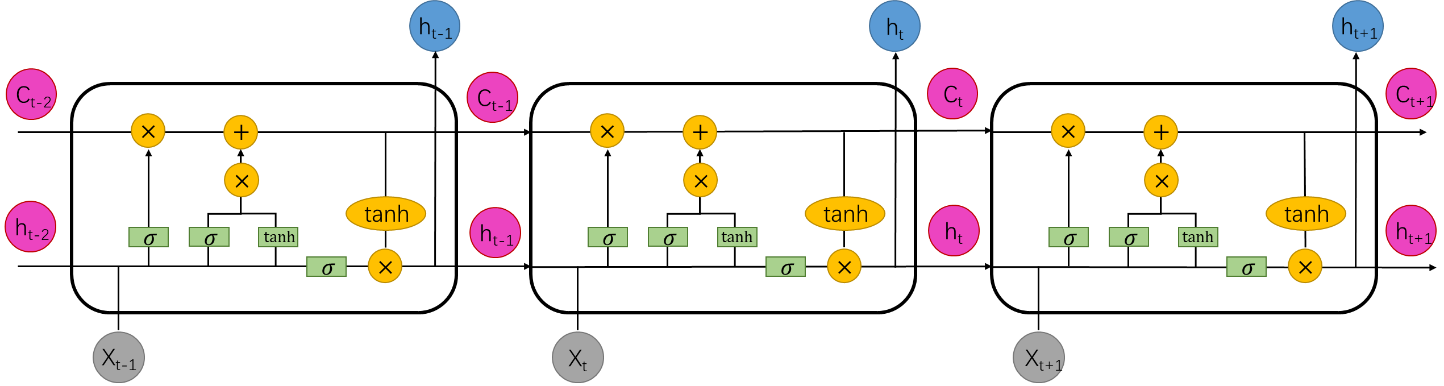
In the formula (9), is the output of *squeeze*, , are fully connected layer operation, means a ReLU layer operation and then go through the sigmoid function to get which is to characterize the weights of feature maps in matrix . This weight is learned through the previous fully connected layers and nonlinear layers to be trained end-to-end.

Finally, the SE block assigns the calculated channel weight to each matrix as formula (10), where is the weight of -th matrix .

## Bidirectional Long short-term memory

As the main component of BiLSTM, LSTM is proposed to solve the problem of gradient vanishing and long-term dependence of recurrent neural network (RNN). Due to its superiority in dealing with sequential data, LSTM has been widely used in video analysis [25], speech recognition [26], signal analysis [27], etc.

To solve the long-term dependence problem of RNN, LSTM introduces a gate mechanism to control the circulation and loss of features. Figure 2 shows the basic structure of the continuous LSTM unit.



**Figure 2:** Structure of LSTM units

Each LSTM unit uses three gates to determine the retained or lost information as the following six formulas, where represents the matrix of the inputs, means a bias vector and is an activation function:

1). forget gate: The forget gate determines how much of the unit state at the previous moment is retained to the current moment

2). input gate: The input gate determines the amount of network input saved to the unit state at the current moment

3). output gate: The output gate controls the proportion of the current output value output by the control unit status to the LSTM

4). Unit status update value :

5). Unit state :

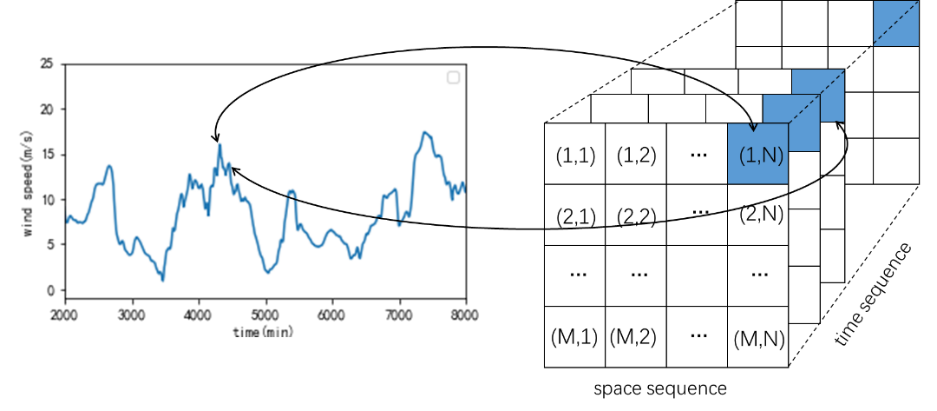
6). output value :

BiLSTM is an extension of regular LSTM, which consists of forward LSTM and backward LSTM (i.e., left-to-right and right-to-left). Firstly, it applies LSTM to the forward input sequence. Then it passes the reverse input sequence into the LSTM model. Therefore, the model's accuracy can be improved because the LSTM applied twice can greatly enhance the long-term dependence [28]. Recent studies have shown that this method can also promote wind speed prediction [29].

# Methodology

## Data process

How to retain the spatio-temporal correlation of data without increasing the amount of data is a challenging problem. Reference [30] proposes a practical solution to organize the spatio-temporal wind speed data into two-dimensional called spatial wind speed matrix (SWSM). Usually, our data set comes from a two-dimensional array, where it can be represented by an grid and each wind speed station can be represented by coordinates, as depicted in Figure 3(a). For each wind speed station, the wind speed in the time dimension is a one-dimensional time series, as illustrated in Figure 3(b). In this way, multiple SWSM that are continuous in time can represent the three-dimensional distribution of all wind speeds both in time and space.



(a) (b)

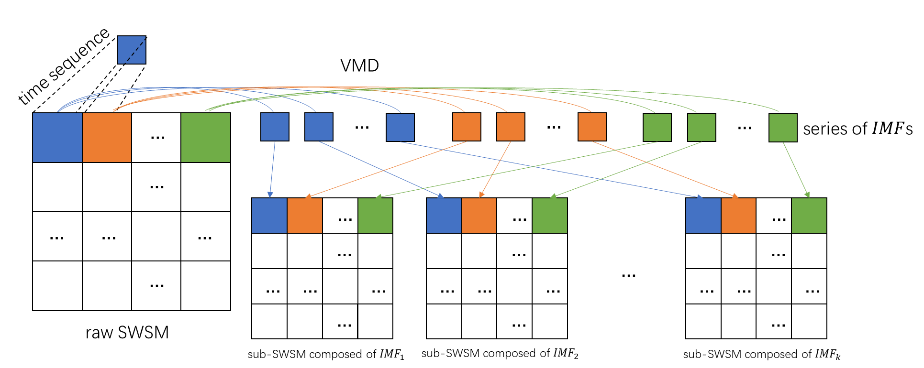
**Figure 3:** (a) A wind speed time series. (b) A series of SWSM

Another problem is that VMD is generally used to decompose time series, and it is difficult to act on three-dimensional spatio-temporal data directly. Here we have made some improvements to SWSM to solve it as represented in Figure 4, where is the number of modes:

1) After setting up the entire SWSM, extract and store the wind speed of each time series separately.

2) Decompose each time series separately with VMD, and get , , … for each time sequence

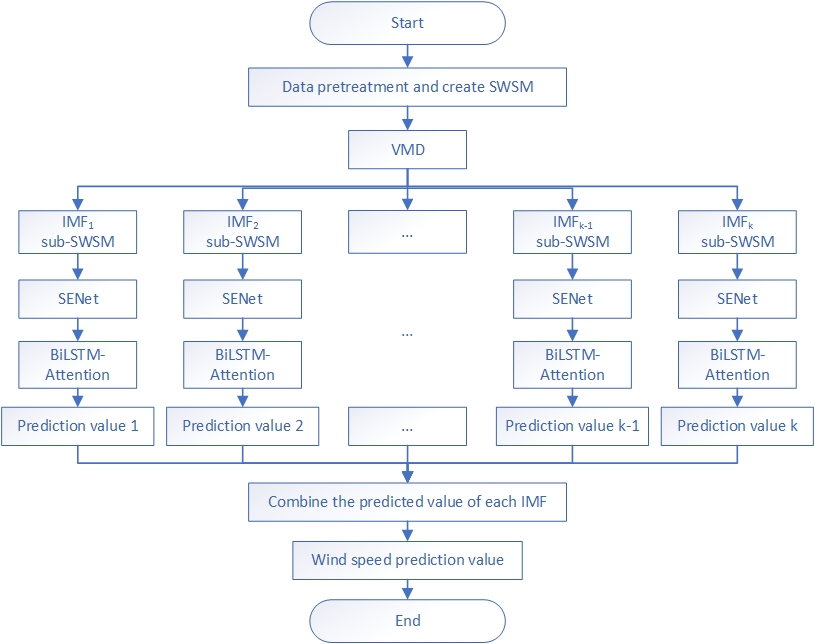
3) The s are formed into sub-SWSMs according to the corresponding positions



**Figure 4:** Steps to generate sub-SWSM composed of IMFs

## The basic strategy for VASTN

In the deep hybrid architecture, removing irrelevant data can significantly reduce unnecessary training time and prevent the experiment from overfitting. Therefore, the VASTN algorithm first clears the useless data in the selected original data and only retains the wind speed and corresponding time and geographic location information to ensure the data set's temporal and spatial correlation. Then it uses VMD to decompose the original SWSM with solid nonlinearity and randomness into a series of sub-SWSM composed of IMF. For each sub-SWSM component, we use the SENet-BiLSTM-Attention hybrid framework for model training and prediction and get the predicted value of each IMF, respectively. Finally, the predicted value of each IMF is superimposed to obtain the final predicted value of wind speed. The flow chart of VASTN's entire procedure is shown in Figure 5:



**Figure 5:** The flowchart of VASTN

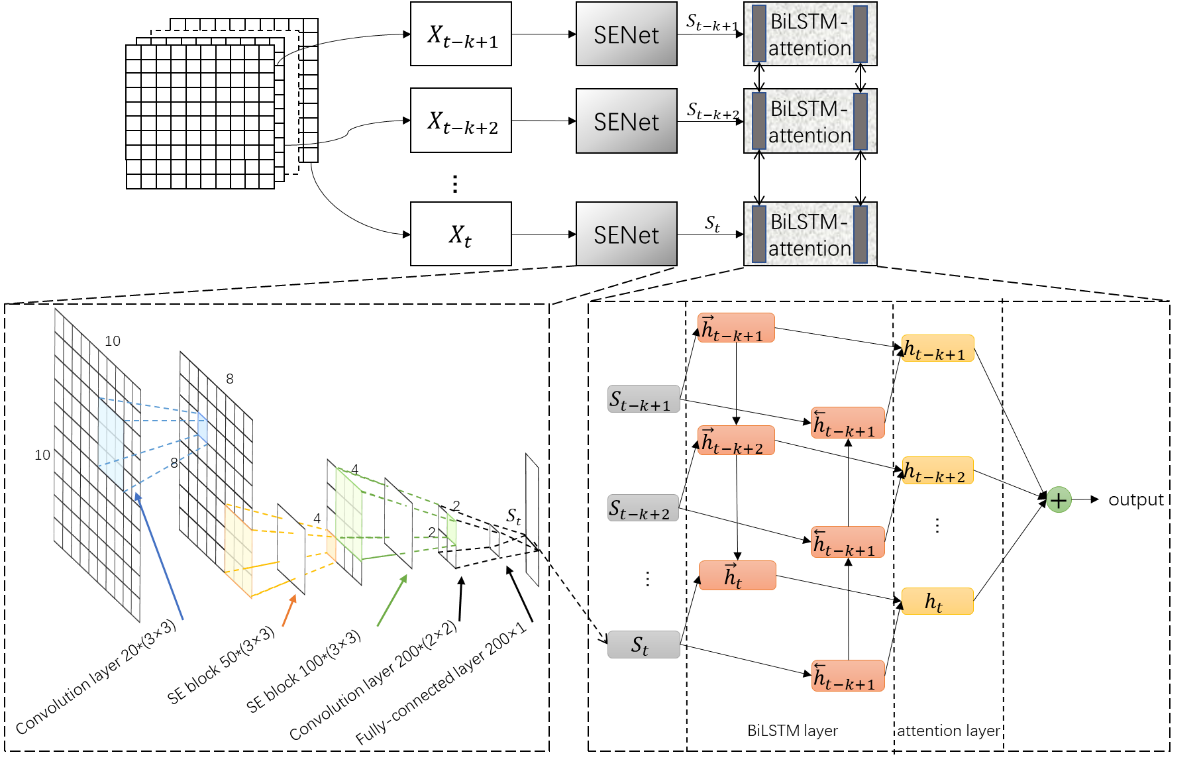
## Network Architecture

After the data processing mentioned above and the employment of VMD, we need to extract the features of the sub-SWSM, respectively, i.e., the extraction of spatial features and the capture of temporal features.

SENet is employed to extract the spatial features. It has the following advantages: 1) kernels' topology-preserving allows CNNs to thoroughly and directly utilize spatial attributes from a single SWSM [31]. 2) SENet can learn to use global information to selectively emphasize channels' information and suppress useless information without increasing too much computational complexity [19].

To capture the extracted spatial features in time sequences, we use the combination of BiLSTM and attention mechanism as the time model in this article whose advantages are as follows: 1) LSTM can analyze the dependencies within the sequence by capturing both the long-term and short-term time characteristics to meet different forecasting needs. 2) BiLSTM traverses time series in both directions, making the extraction of time features more effective than general LSTM [32]. 3) The combination of attention mechanism and BiLSTM enables the model to focus on more important features, decreasing training time complexity while increasing prediction accuracy.

The overall network framework proposed by the paper is shown in Figure 6(a), where Figure 6(b) illustrates the part of a sub-SWSM composed of IMFn. At the bottom of the model, it uses CNN to extract the spatial characteristics of a single wind speed matrix. Then SE block emphasizes the critical channel information and weakens the useless information, thereby analyzing the spatial correlation in the wind farm. In the upper layer of the model, we apply BiLSTM to bidirectional capture of the temporal features among the previously extracted spatial features. At the same time, employ the attention mechanism to improve efficiency and accuracy. Finally, the top layer can output the current IMF forecast value, and it is added together with other IMF forecast results to get the final wind speed forecast value.



(b)

(a)

**Figure 6:** (a) The process of the network framework (b) A series of SWSM

The entire mixed model can be jointly trained with a loss function mean squared error (MSE), which is defined as follows

where is the number of the whole data, represents the -th actual wind speed, and is the predicted value of -th wind speed.

The training error propagates down the model backward during model training, i.e., the backpropagation (BP) algorithm. The error difference propagates from the BiLSTM-Attention layer to the SENet layer to the underlying CNN. In this way, each layer can adjust parameters through training error guidance. Another potential benefit is that the spatial model (SENet) can adapt according to temporal information. Similarly, the temporal model (BiLSTM-Attention) can also adjust according to spatial information [30]. In this way, space and time learning can be coupled together so that the entire framework can learn spatio-temporal features together.

# Case study

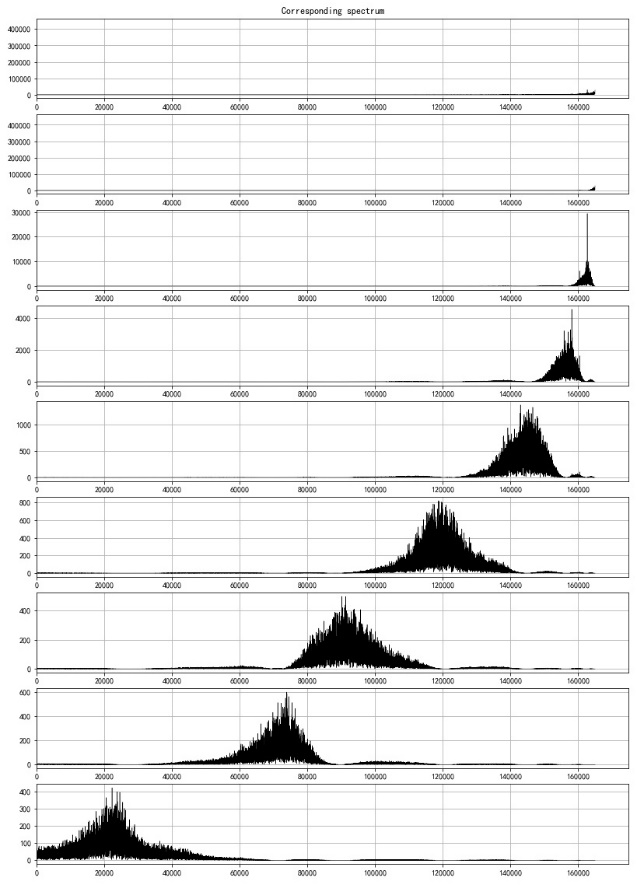
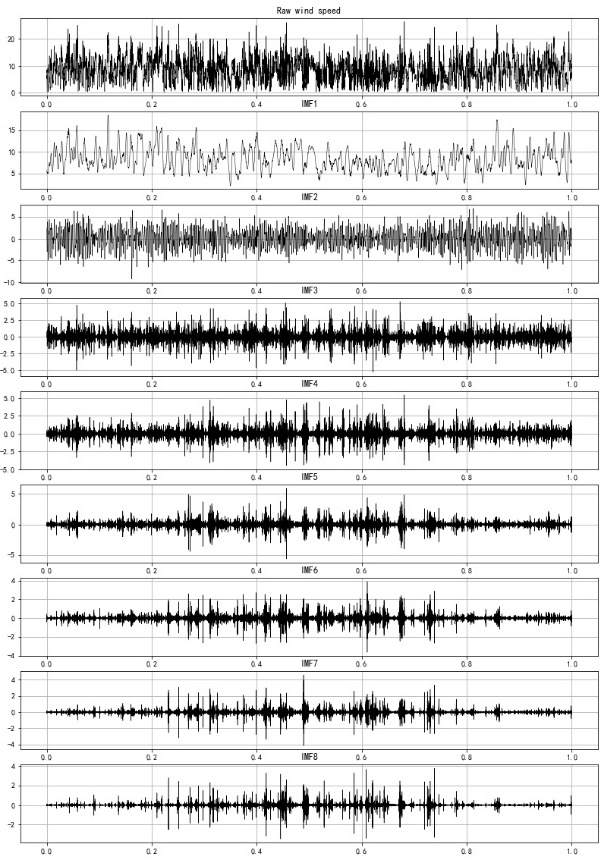
## Data Set

The National Renewable Energy Laboratory provides the Wind Integration National Dataset (WIND) Toolkit is a dataset that realistically reflects the simulated wind plants' ramping characteristics, spatial and temporal correlations, and capacity factors. It contains wind speed data with the 5-min resolution for more than 126,000 land-based and offshore wind power production sites over the continental United States for the year 2007-2013 [33].

This paper selects a 10×10 wind turbine matrix located in New Mexico in 2012 to obtain 100 wind speed data sets with 52560 frames, where the highest wind speed is 32.77m/s, and the lowest wind speed is 0.02m/s.

## Data Decomposition

The VMD algorithm involves the following parameters: the number of modes , the penalty term , the fidelity coefficient and the convergence tolerance level . Studies have shown that and can usually take default values [34], so the difficulty and focus of the VMD algorithm lies in how to select the appropriate and . In this paper, the values of K and α refer to the reference method [35], and we carry out multiple experiments to determine , , and . From the perspective of the frequency spectrum of the decomposition mode, too little K can lead to under-segmentation of the signal, and too many modes may lead to pattern repetition or additional noise [35]. Figure 7(a) shows the decomposition result, and Figure 7(b) represents no apparent modal aliasing phenomenon in the decomposition result which means the decomposition achieves a relatively ideal effect.



(a) (b)

**Figure 7:** (a) VMD diagram of a time series (b) Spectrogram of each IMF

## Evaluation Criteria

This paper applies the root mean square error (RMSE), mean absolute error (MAE), mean absolute percent error (MAPE), and Pearson correlation (COR) to compare the prediction results of the VASTN model and other algorithms to evaluate the prediction performance. The evaluation criteria are defined as

where n represents the number of predicted data, represents the -th actual wind speed, and is the predicted value of -th wind speed, represents the covariance between and , represents the standard deviation of . The larger COR indicates the better the predicted performance, and the smaller RMSE, MAE, MAPE means the worse the predicted performance.

## Environment and Settings

The experiment uses the Keras2.4.2 framework with the TensorFlow2.4 backend to implement the network. The experimental server environment is CPU E5-2689 16-core 16G, GPU NVIDIA GeForce RTX 3070, 8G video memory, 200GB solid-state drive.

The data set is divided into three parts in a prediction task: the first 60% frames are the training set, the subsequent 10% frames are the validation set, and the remaining 30% frames are the test set. Among them, the training set is to train the model. The validation set is applied to monitor the quality of the model, save the best model in real-time during training, and use the test set to evaluate the predicted performance of the model. At the same time, we employ early stopping in the validation set to improve model performance and prevent the model from overfitting. As mentioned above, there are 31536, 5256, and 15768 frames in the training set, validation set, and test set in this article. There may be a few differences in frame number in each set depending on the prediction horizon.

VASTN is trained by optimization algorithm Adagrad with the epoch for training set to 100. The deep hybrid framework shows in Table 1 that the model's optimization algorithm and hyperparameters are obtained from experiments.

**Table 1:** Hybrid deep framework configuration

|  |  |  |
| --- | --- | --- |
| Index | Type | Configurations |
| 1 | Convolution layer | kernels：20；kernel size：；stride： |
| 2 | SE block layer | filter：；ratio： |
| 3 | SE block layer | filter：；ratio： |
| 4 | Convolution layer | kernels：200；kernel size：；stride： |
| 5 | Fully connected layer | units：200 |
| 6 | BiLSTM-Attention layer | hidden units：200 |

## Comparison Algorithm

To verify the superior performance of the proposed VASTN, we compare it with the following wind speed prediction AI models: LSTM, MLP, RNN, and predictive spatio-temporal network (PSTN) [30]. At the same time, to verify the effectiveness of each component of the hybrid framework, we also set up some sub-models of VASTN for comparison, i.e., SENet-BiLSTM-Attention, BiLSTM-Attention, and VMD-CNN-LSTM model.

Different from other typical algorithms, PSTN is a spatio-temporal wind speed prediction model. Its model framework consists of two parts: CNN and LSTM, where CNN extracts the spatial features of the data and then uses LSTM to capture the temporal features and finally obtains the predicted value.

Tables 2 to 5 illustrate the comparison of RMSE, MAE, MAPE, and COR of different prediction models. VASTN has clearly shown superiority in each prediction horizon of various evaluation criteria. When the prediction horizon is 10, 20, 30, 60, 120min, the RMSE of VASTN is 13%, 20%, 16%, 21%, 19% lower than PSTN, respectively, with an average reduction of 17%. And the average optimization of MAE and MAPE was 18% and 28%, respectively, which indicates that VASTN has considerable advantages compared to the traditional wind speed prediction model.

In addition, for the sub-model SENet-BiLSTM-Attention that does not use VMD decomposition, VASTN optimizes the RMSE, MAE, and MAPE by an average of 20%, 16%, and 25%. Correspondingly, the VMD-CNN-LSTM sub-model is 14%, 13%, and 21% better than PSTN. It shows that VMD decomposition can better eliminate the randomness and unevenness of wind speed for short-term wind speed forecasting to obtain better forecast results. Also, BiLSTM has specific optimizations in RMSE, MAE, and MAPE compared to general LSTM. Similarly, VASTN also has 6%, 6%, and 10% improvement in these three aspects compared to the VMD-CNN-LSTM sub-model of the algorithm that does not use the attention mechanism, which proves the optimization of the model after the introduction of the attention mechanism. Compared with the spatio-temporal model PSTN, the best-performing model in the time sequence model, i.e., BiLSTM-Attention, is 5%, 4%, and 6% improvement in RMSE, MAE, and MAPE. It indicates that the model's performance using the temporal and spatial characteristics is generally better than the time sequence model, reflecting that the spatio-temporal data contains more features conducive to wind speed prediction than pure time-series data.

**Table 2:** Comparison of RMSE for different prediction models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Prediction Horizon (min) | | | | |
| 10 | 20 | 30 | 60 | 120 |
| LSTM | 0.616 | 0.961 | 1.214 | 1.812 | 2.549 |
| MLP | 0.621 | 1.033 | 1.227 | 2.222 | 2.582 |
| RNN | 0.678 | 1.000 | 1.280 | 1.871 | 2.552 |
| PSTN | 0.590 | 0.946 | 1.158 | 1.687 | 2.383 |
| BiLSTM-Attention | 0.613 | 0.949 | 1.212 | 1.804 | 2.541 |
| VMD-CNN-LSTM | 0.520 | 0.776 | 1.013 | 1.378 | 2.177 |
| SENet-BiLSTM-Attention | 0.578 | 0.883 | 1.389 | 1.664 | 2.363 |
| VASTN | **0.517** | **0.753** | **0.972** | **1.339** | **1.933** |

**Table 3:** Comparison of MAE for different prediction models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Prediction Horizon (min) | | | | |
| 10 | 20 | 30 | 60 | 120 |
| LSTM | 0.346 | 0.628 | 0.836 | 1.331 | 1.98 |
| MLP | 0.367 | 0.735 | 0.859 | 1.694 | 1.949 |
| RNN | 0.416 | 0.673 | 0.893 | 1.369 | 1.976 |
| PSTN | 0.362 | 0.654 | 0.805 | 1.225 | 1.831 |
| BiLSTM-Attention | 0.354 | 0.610 | 0.834 | 1.319 | 1.97 |
| VMD-CNN-LSTM | 0.332 | 0.532 | 0.714 | 1.023 | 1.653 |
| SENet-BiLSTM-Attention | 0.355 | 0.593 | 0.796 | 1.210 | 1.813 |
| VASTN | **0.331** | **0.513** | **0.691** | **0.990** | **1.493** |

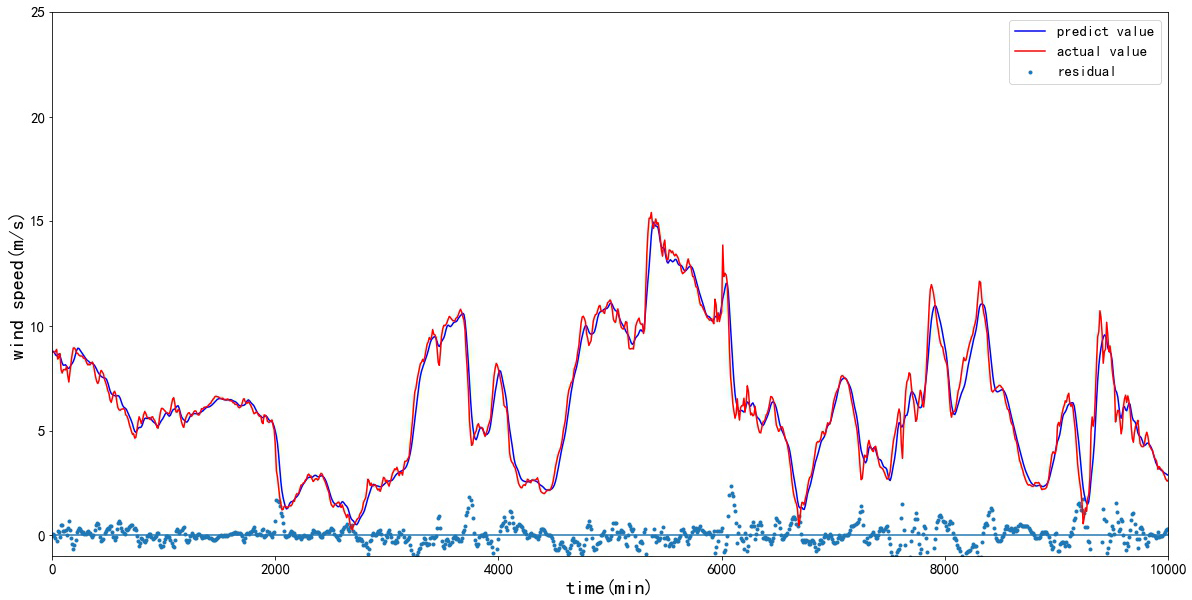
**Table 4:** Comparison of MAPE for different prediction models

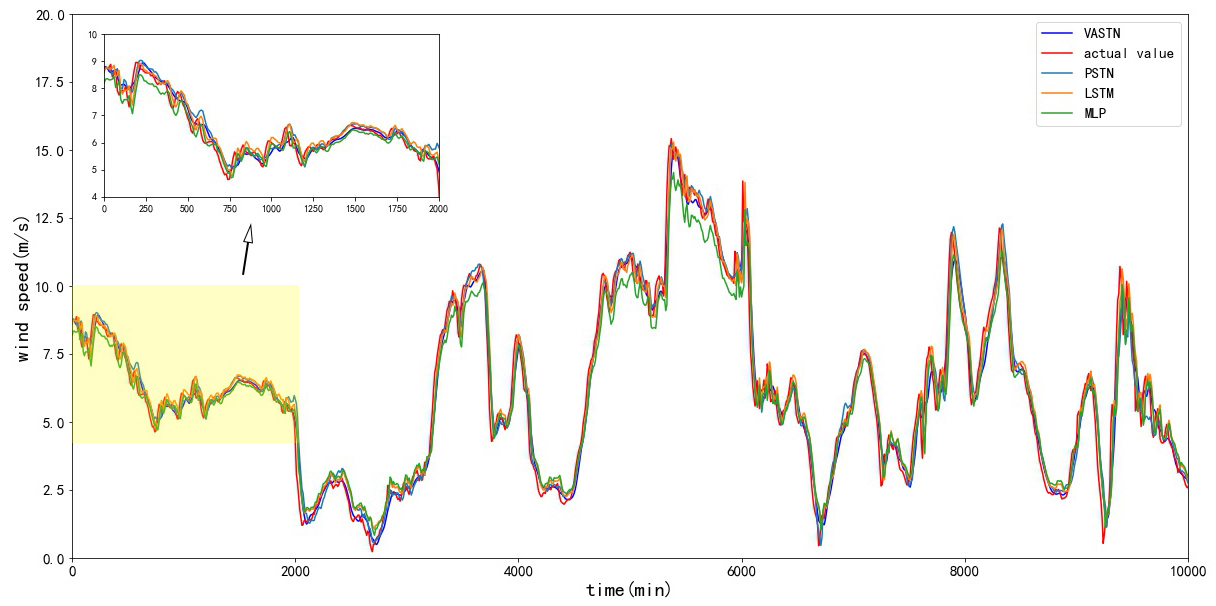
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Prediction Horizon (min) | | | | |
| 10 | 20 | 30 | 60 | 120 |
| LSTM | 6.476 | 12.872 | 17.789 | 30.553 | 49.325 |
| MLP | 7.614 | 17.633 | 18.815 | 26.769 | 49.009 |
| RNN | 6.677 | 13.552 | 15.918 | 25.894 | 47.391 |
| PSTN | 6.931 | 12.922 | 17.079 | 25.595 | 46.504 |
| BiLSTM-Attention | 7.351 | 12.315 | 17.954 | 30.256 | 48.089 |
| VMD-CNN-LSTM | 6.451 | 9.957 | 13.238 | 20.826 | 35.851 |
| SENet-BiLSTM-Attention | 6.637 | 12.084 | 15.581 | 26.710 | 43.513 |
| VASTN | **5.947** | **9.522** | **13.123** | **20.421** | **29.500** |

**Table 5:** Comparison of COR for different prediction models

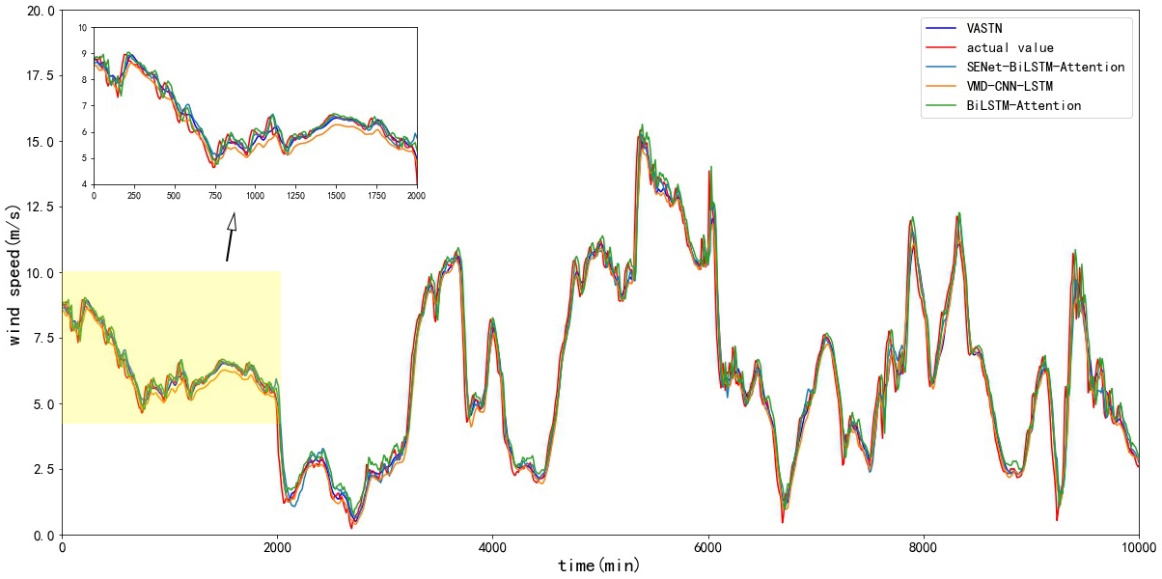
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Prediction Horizon (min) | | | | |
| 10 | 20 | 30 | 60 | 120 |
| LSTM | 0.989 | 0.975 | 0.961 | 0.905 | 0.802 |
| MLP | 0.990 | 0.976 | 0.959 | 0.906 | 0.802 |
| RNN | 0.989 | 0.975 | 0.959 | 0.906 | 0.801 |
| PSTN | 0.990 | 0.977 | 0.962 | 0.917 | 0.832 |
| BiLSTM-Attention | 0.990 | 0.974 | 0.962 | 0.908 | 0.802 |
| VMD-CNN-LSTM | 0.992 | 0.983 | 0.971 | 0.946 | 0.864 |
| SENet-BiLSTM-Attention | 0.991 | 0.978 | 0.964 | 0.921 | 0.830 |
| VASTN | **0.993** | **0.984** | **0.973** | **0.949** | **0.895** |

Figure 8 shows the prediction results of the 20-minute prediction horizon of the VASTN model proposed in the paper. It is evident that the predicted value of the VASTN model is consistent with the actual value. The residual analysis results show that the prediction residuals of the model can be uniformly and randomly distributed on both sides of the zero baselines when there is no apparent fluctuation in wind speed. In contrast, only when the wind speed fluctuates wildly may it deviate farther. These indicate that the modeling process rarely produces systematic errors. Figure 9 shows the comparison of 20-minute wind speed prediction results of different typical models and VASTN, including VASTN, LSTM, MLP, RNN, PSTN. Figure 10 indicates the comparison of the prediction results of VASTN and its sub-models, including SENet-BiLSTM-Attention, BiLSTM-Attention, and VMD-CNN-LSTM model. Figures 9 and 10 also show the superiority of the VASTN algorithm in short-term wind speed prediction.



**Figure 8:** 20min VASTN prediction results and residual chart

**Figure 9:** 20min typical model comparison chart



**Figure 10:** 20min sub-model comparison chart

# Conclusion and Future work

This paper proposes a new hybrid deep framework VASTN for short-term wind speed prediction, which integrates the learning of wind speed's temporal and spatial characteristics into a unified framework. First, the randomness of wind speed is reduced by VMD decomposition. Then the spatial characteristics of the wind speed matrix are extracted using SENet, and the time characteristics are weighted and extracted by BiLSTM-Attention. At last, it merges the prediction results of each IMF to obtain the wind speed prediction value. The experiments on real data sets illustrate the superiority of the VASTN algorithm over the traditional short-term wind speed prediction algorithm. It also proves the effectiveness of VMD, attention mechanism, and the extraction of temporal and spatial features of wind speed in short-term wind speed prediction.

In future work, we will use the improved VMD algorithm to adjust and select the optimal VMD parameters automatically. In addition, we will also try to use other attention mechanism methods, i.e., spatial attention or channel and spatial hybrid attention, to improve the accuracy and efficiency of the algorithm.

# References

1. W. Zhou, C. Z. Lou, Z. S. Li, L. Lu, & H. X. Yang, "Current status of research on optimum sizing of stand-alone hybrid solar–wind power generation systems", *Applied Energy*, vol. 87, no. 2, pp. 380-389, 2010
2. S. Fan, J. R. Liao, R. Yokoyama, L. N. Chen, & W. J. Lee, "Forecasting the wind generation using a two-stage network based on meteorological information," *IEEE Transactions on Energy Conversion*, vol. 24, no. 2, pp. 474-482, 2009
3. Q. H. Hu, R. J. Zhang, & Y. C. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy*, vol. 85, JAN, pp. 83-95, 2016
4. L. L, "Short-term prediction of the power production from wind farms," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 80, no. 1-2, pp. 207-220, 1999
5. G. Giebel, R. Brownsword, G. N. Kariniotakis, M. D. Denhard, & C. Draxl, "The State of the Art in Short-Term Prediction of Wind Power A Literature Overview," 2011
6. H. Liu, H. Q. Tian, & Y. F. Li, "Comparison of two new arima-ann and arima-kalman hybrid methods for wind speed prediction," *Applied Energy*, vol. 98, pp. 415-424, 2012
7. G. Y. Wu, Y. Xiao, & S. S. Weng, "Discussion about short-term forecast of wind speed on wind farm," *Jilin Electric Power*, vol. 181, no. 5, pp. 21-24, 2005
8. T. G. Barbounis, & J. B. Theocharis, "A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation," *Information Sciences*, vol. 177, no. 24, pp. 5775-5797, 2007
9. H. Liu, H. Q. Tian, & Y. F. Li, "Comparison of new hybrid feemd-mlp, feemd-anfis, wavelet packet-mlp and wavelet packet-anfis for wind speed predictions," *Energy Conversion & Management*, vol. 89(jan.), pp. 1-11, 2015
10. T. G. Barbounis, J. B. Theocharis, M. Alexiadis, & P. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Transactions on Energy Conversion*, vol. 21, no. 1, pp. 273-284, 2006
11. B. Li, E. Crisostomi, M. Raugi, & M. Tucci, "Wind Power Forecast Using Wind Forecasts at Different Altitudes in Convolutional Neural Networks," *2019 IEEE Power & Energy Society General Meeting (PESGM)*, IEEE, 2020
12. Z. H. Wu, and N. E. Huang, "Ensemble Empirical Mode Decomposition: a noise-assisted data analysis method, "*Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1-41, 2007
13. S. X. Wang, N. Zhang, L. Wu, & Y. M. Wang, , "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and ga-bp neural network method," *Renewable Energy*, vol. 94(aug.), pp. 629-636, 2016
14. M. E. Torres, et al. "A complete ensemble empirical mode decomposition with adaptive noise." *IEEE International Conference on Acoustics IEEE*, 2011.
15. T. L. Fu, "A hybrid prediction of wind speed based on variational mode decomposition method and long short-term memory," *2020 International Conference on Computer Engineering and Application (ICCEA)*,2020.
16. G. Sopena, J. Manuel, V. Pakrashi, and B. Ghosh, "Multi-step ahead wind power forecasting for Ireland using an ensemble of VMD-ELM models," *2020 31st Irish Signals and Systems Conference (ISSC)*, 2020.
17. Z. X. Sun, and M. Zhao, "Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis," *IEEE Access PP.99*, pp. 1-1
18. D. Zosso, K. Dragomiretskiy, et al. "Variational Mode Decomposition". *IEEE Transactions on Signal Processing A Publication of the IEEE Signal Processing Society*, 2014.
19. H. Jie, S. Li, S. Gang, et al, "Squeeze-and-Excitation Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, PP(99), 2017
20. X. X. Zhu, R. Z. Liu, Y. Chen, X. X. Gao, et al. "Wind speed behaviors feather analysis and its utilization on wind speed prediction using 3D-CNN," *Energy* 2021.
21. S. Ji, W. Xu, M. Yang, et al. "3D Convolutional Neural Networks for Human Action Recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 221-231, 2013
22. A. G. Howard, M. L. Zhu, B. Chen, et al, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", 2017.
23. R. V. Sharan, S. Berkovsky, S. Liu, "Voice Command Recognition Using Biologically Inspired Time-Frequency Representation and Convolutional Neural Networks," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society. IEEE*, 2020
24. I. Almakky, V. Palade, A. Ruiz-Garcia, "Deep Convolutional Neural Networks for Text Localisation in Figures From Biomedical Literature," *2019 International Joint Conference on Neural Networks (IJCNN). IEEE*, 2019
25. Y. Bin, Y. Yang, F. Shen, et al, "Describing Video With Attention-Based Bidirectional LSTM," *IEEE Transactions on Cybernetics,* *PP(7)*, pp. 1-11, 2019
26. Z. Wang, T. Zhang, Y. Shao, et al, "LSTM-convolutional-BLSTM encoder-decoder network for minimum mean-square error approach to speech enhancement," *Applied Acoustics*, 172(2):107647, 2021
27. L. Liao, J. Liu, X. Wu, et al, "Time Difference Penalized Traffic Signal Timing by LSTM Q-network to Balance Safety and Capacity at Intersections," *IEEE Access,* *PP(99)*, pp. 1-1, 2020
28. P. Baldi, S. Brunak, P. Frasconi, et al, "Exploiting the past and the future in protein secondary structure prediction," *Bioinformatics*, 15(11):937, 1999
29. S. Siami-Namini, N. Tavakoli, A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," *2019 IEEE International Conference on Big Data (Big Data)*, *IEEE*, 2019
30. Q. Zhu, J. Chen, D. Shi, et al, "Learning Temporal and Spatial Correlations Jointly: A Unified Framework for Wind Speed Prediction," *IEEE Transactions on Sustainable Energy*, pp. 1-1, 2019
31. R. Y. Li, W. Sheng, F. Y. Zhu, et al. "Adaptive Graph Convolutional Neural Networks," 2018.
32. S. Siami-Namini, N. Tavakoli, A. S. Namin "A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM," 2019.
33. C. Draxl, A. Clifton, B. M. Hodge, et al, "The Wind Integration National Dataset (WIND) Toolkit". *Applied Energy*, vol. 151(aug.1), pp. 355-366, 2015
34. D. Yang, F. Feng, Y. Hao, et al, "A VMD sample entropy feature extraction method and its application in planetary gearbox fault diagnosis," *Journal of Vibration and Shock*, vol. 37, no. 16, pp. 198-205, 2018
35. J. Fang, "Research on fault diagnosis of marine gearbox based on variational mode decomposition," Wuhan University of Technology, 2017