Short-term wind speed prediction based on combinatorial prediction model

Mingguang Li [1, #], Runyi Huang [2, #, *], Yumiao Yang [3, #]

[1] School of computing Nantong Institute of Technology Jiangsu, China

[2] School of computing Zhuhai College of Science and Technology Guangdong, China

[3] School of Computer and Information Engineering Harbin University of Commerce Harbin, China

* Corresponding Author Email: HRY04202023@stu.zcst.edu.cn

Abstract. Improving the accuracy of wind speed forecast can increase wind power generation and better achieve wind energy grid connection. Therefore, a two-stage wind speed prediction model based on Ensemble Empirical Modal Decomposition (EEMD) and the combination prediction of Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), eXtreme Gradient Boosting (XGBOOST), Gate Recurrent Unit (GRU), Temporal Convolutional Network (TCN) is proposed. First, the original wind speed series is separated into Intrinsic Mode Functions (IMFs) using EEMD. Then, RNN, LSTM, XGBOOST, GRU, TCN multiple prediction models are established to learn features from each subsequence and superimpose the prediction results of subsequences. Finally, Particle Swarm Optimization (PSO) is applied to the results of multiple prediction models to assign weights, combined with weight superimposing sequences to achieve higher accuracy and more robust wind speed prediction. Simulation analysis using data from St. Thomas, Virgin Islands wind measurement station to validate the validity of the combined prediction model. The experimental simulation results show that the model proposed in this paper has a good result on increasing wind speed prediction accuracy.

Keywords: Wind speed prediction; EEMD; combination prediction; PSO.

1. Introduction

In the last few years, with the depletion of non-renewable energy and its increasingly serious environmental problems, the research on wind energy has gradually been called a hot spot. As wind energy has intermittence and volatility characteristics, direct use will have a adverse influnce on the safety and stable operation of the power system. Wind speed prediction can be used to reduce errors to improve wind power generation and realize the safety of wind power systems to achieve large-scale grid connection.

The hot research models for wind speed prediction are broadly categorized as physical, statistical, and artificial intelligence models. Among them, physical models are most befitting for long-term forecasting, but they require a lot of computational resources. Statistical models, such as Autoregressive (AR), are low-cost and easy to implement, but the main problem is that the prediction accuracy is not satisfactory [1]. Artificial intelligence models, such as Support Vector Machines (SVM), are mostly shallow learning algorithms difficult to dig into the deep traits of the input data, thus limiting the prediction accuracy of the models [2]. So deep learning models were born, such as Recurrent Neural Network (RNN), Long Short Term Memory (LSTM) , eXtreme Gradient Boosting (XGBOOST), Gate Recurrent Unit (GRU), and Temporal Convolutional Network (TCN)[3-5].

The characteristics of wind speed make it not have stationarity in time series, that is, it cannot predict the future result better by current characteristics. If only the forecast model is used, the forecast accuracy will be greatly reduced. Signal decomposition algorithms can effectively reduce the effect of non-stationary sequences, such as the Ensemble Empirical Mode Decomposition (EEMD) that introduces white noise obeying a normal distribution.

Therefore, a two-stage wind speed prediction model based on EEMD and the combination prediction of RNN, LSTM, XGBOOST, GRU, TCN is proposed. EEMD was used to separate the primordial wind speed into IMFs, and multiple prediction models of RNN, LSTM, XGBOOST, GRU, and TCN were established to learn features from each subsequence, and the prediction results of the subsequence were superimposed. On this basis, the PSO is applied to weight the superimposed five sequences to match the weights, combined with weight superimposing sequences to achieve wind speed prediction.

The main innovations of this study are as follows:

(1) A combined prediction model is proposed, which can effectively reduce the problem of improper adaptation of a single base model in predicting different time series types, so as to better capture the different characteristics of each phase of the time series and the hidden relationships of the data, and effectively enhance the accuracy of wind speed prediction .

(2) The PSO optimization algorithm is used to determine the weights, which further improves the overall predictive capacity of the model in the face of different time series and enhances the robustness of the model.

The rest of arrangements are as follows: Section II introduces the fundamental theories of the proposed model. Section III performs experimental simulation, and Section IV summarizes the entire paper.

2. Methods

2.1. EEMD

EMD is also called the empirical mode decomposition algorithm, which can decompose a sequence into a series of sub-sequences, and each sub-sequence is called an IMF component. However, this decomposition method also has some defects, that is, it is easy to have the problem of mode mixing [6]. There are two main cases of modal aliasing as follows.

One is that signals of different scales appear in the same signal component, and the other is that signals of the same scale appear in different signal components.To solve this kind of problem, we can use the improved algorithm of EMD, EEMD. The improved method of EEMD compared with EMD is to add uniform white noise to the signal in the decomposition process, and the artificial added noise can mask the noise of the signal itself. Thus, more accurate upper and lower envelope can be obtained. Finally, by averaging the decomposition results, the influence of noise on decomposition can be reduced. This is because the white noise has zero mean.

2.2. LSTM

Time series forecasting is still considered to be a challenging problem. Recurrent Neural Network (RNN) is one of the classical and commonly used algorithms in time series prediction. The main algorithm idea is as follows:

$$\text{Input layer:} \quad X_t \tag{1}$$

$$\text{Hidden layer:} \quad S_t = f(U \cdot X_t + W \cdot S_{t-1}) \tag{2}$$

$$\text{Output layer:} \quad O_t = g(V \cdot S_t) \tag{3}$$

However, due to the problems of gradient disappearance and gradient explosion when RNN deals with continuous data, it has been less used in practical problems and is replaced by LSTM. LSTM is a special RNN architecture developed to overcome the gradient vanishing problem of RNN. Compared to RNN, LSTM can retain data for a longer period of time, is a stable network for solving long-dependence problems, and is now widely used. The structure of LSTM also contains the chain form of RNN's recurrent neural network model, but there is a different structure in the recurrent unit. There is no longer a single neural network layer, but it is composed of four units[7]. The structure of LSTM neural network(as shown in Fig.1) is mainly composed of three parts, which are input gate, forget gate and output gate. The corresponding equations are as follows:

Input gate: $$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \tag{4}$$

Output gate: $$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \tag{5}$$

Forget gate: $$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \tag{6}$$

Temporary cell state: $$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \tag{7}$$

Cell state: $$C_t = I_t \odot \tilde{C}_t + F_t \odot C_{t-1} \tag{8}$$

Hidden state: $$H_t = O_t \odot \tanh(C_t) \tag{9}$$

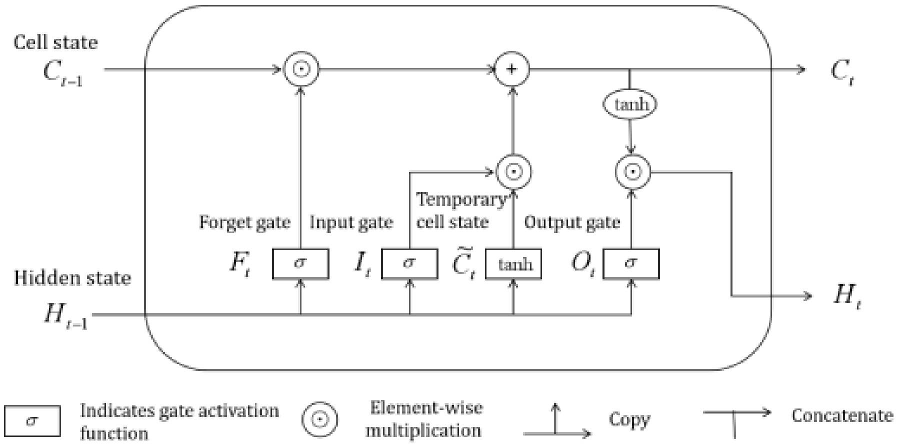

Fig 1. The architecture of LSTM.

Where:

$X_t \in \mathfrak{R}^d$ Input vector to the LSTM unit.
$F_t \in \mathfrak{R}^h$ Forget states activation vector
$I_t \in \mathfrak{R}^h$ Input gate's activation vector
$O_t \in \mathfrak{R}^h$ Output gate's activation vector
$H_t \in \mathfrak{R}^h$ Hidden state vector
$C_t \in \mathfrak{R}^h$ Cell state vector
W is the parameter weight matrix.σ and tanh is the Activation function [8].

2.3. GRU

Gated recurrent Unit GRUs are faster to train than LSTMS, can generalize with less data, and have similar results to LSTM. As a variant of LSTM, GRU optimizes the internal unit structure and the forget gate and output gate become the update gate [9], and the main mathematical principle is shown in Eq.10-Eq.13.

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{10}$$

$$r_t = \sigma(W_t x_t + U_t h_{t-1}) \tag{11}$$

$$\tilde{h}_t = \tanh\left(W x_t + U(r_t \circ h_{t-1})\right) \tag{12}$$

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t \tag{13}$$

Where X_t is the input variable at time t.h_{t-1} Is the output of the previously hidden layer.h_t Is the hidden layer output of this unit.W_z), W_r), W), U_z), U_r), U Is a trainable matrix.1 Is the identity matrix.$\circ$ Is a composite relation.

2.4. TCN

The TCN uses a one-dimensional convolutional network, which is composed of dilated convolution and residual modules. The extended convolution allows the input of the convolution to be interval sampled, and the residual module allows the network to transfer information across layers to avoid information loss due to too many layers.

For a one-dimensional sequence, the filter F= $(f_1, f_2, \dots\dots, f_K)$, the sequence X= $(x_1), x_2, \dots\dots, x_T)$. The mathematical principle is given in Eq.14

$$(F * X)_{(x_t)} = \sum_{k-1}^{K} f_k \, x_{t-(K-k)d} \tag{14}$$

Where k is the number of layers. A residual block includes two layers of convolution and nonlinear mapping, and the structure is shown in Fig.2.

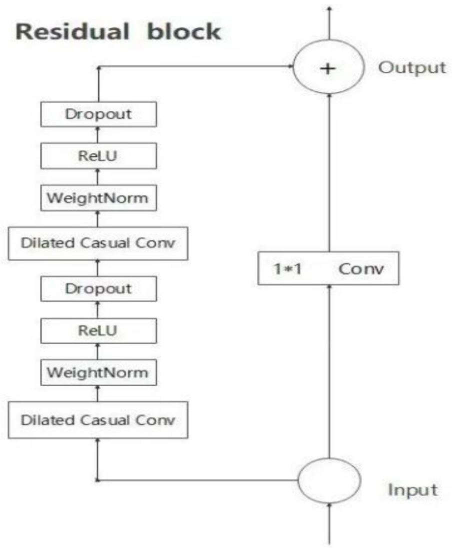

Fig 2. Schematic representation of the residual block structure of the TCN model.

2.5. XGBOOST

The core idea of extreme gradient boosting XGBOOST is to first build a tree to predict the value, then get a residual, build the next tree to learn the residual, and repeat the above steps until K trees are built, in this process, the tree is continuously optimized. Finally, the predicted values of each decision tree after each training are summed to obtain the predicted values of the optimized model. The mathematical principle is given in Eq. 15.

$$\hat{y_\imath} = \emptyset(x_i) = \sum_{k=1}^{K} f_k\,(x_i), f_k \in F \tag{15}$$

Where $\hat{y_\imath}$ Is the predicted value of the model for the ith sample.x_iIs the ith sample. Kis the number of classification regression trees.f_k Is the K-th tree model function.

2.6. PSO

PSO is an evolutionary algorithm that mimics the foraging behavior of birds and fish In other words, the particle swarm optimization algorithm can find the optimal solution in the solution set of multiple models, and it can find the global optimal solution by tracking the current optimal solution. This algorithm often shows its superiority in solving practical problems because of its easy implementation, high accuracy, fast convergence, and other advantages.PSO will first initialize a bunch of random particles. Then an iteration is performed to find the optimal solution. The position of each particle in the next generation is determined by its position in the current generation and its own velocity vector, which determines the direction and distance of each particle flight. During the flight, the particle will record the optimal position pbest, and the group will update the optimal position gbest.

$$v_i = v_i + c_1 \times rand() \times (pbest_i - x_i) + c_2 \times rand() \times (gbest_i - x_i) \tag{16}$$

$$x_i = x_i + v_i \tag{17}$$

Where i = 1, 2..., N), N is the total number of particles.v_i is the velocity of the particle. $rand()$Is a random number between (0,1). x_i Is the current position of the particle. c_1 and c_2 Is the learning factor. Usually, the learning factor is equal to two

2.7. The overall framework flow of the model

We proposed model in this study mainly consisting of two stages: data decomposition and combined prediction. Aiming at the volatility and instability of wind speed, a combined wind speed prediction model based on EEMD, PSO, TCN, LSTM, RNN, GRU, and XGBOOST is proposed in this study, which not only achieves higher wind speed prediction accuracy but also has stronger robustness. It solves the problem that the accuracy and robustness of wind speed prediction cannot be combined. Its execution flow chart is shown in Fig.3, and the specific operation process is as follows:

Step1: The original time series was decomposed into multiple subsequences by EEMD decomposition algorithm.

Step2: The time subsequence decomposed using EEMD is brought into GRU, LSTM, RNN, TCN, and XGBOOST algorithms for prediction, and the prediction results are summed and reconstructed.

Step3: The PSO algorithm is used to linearly combine the five predicted results by the EEMD decomposition algorithm to obtain the final prediction result.

3. Experiment

In this experiment, we propose a model based on the combination of the decomposition algorithm and combinatorial prediction. In order to verify the superiority of EEMD method, compared with EMD, CEEMDAN and VMD, MAE, MSE, RMSE and R^2 score four indexes were used to verify the prediction accuracy of the combined model. The effectiveness of the combined model was verified by comparing the combined prediction with the single model.

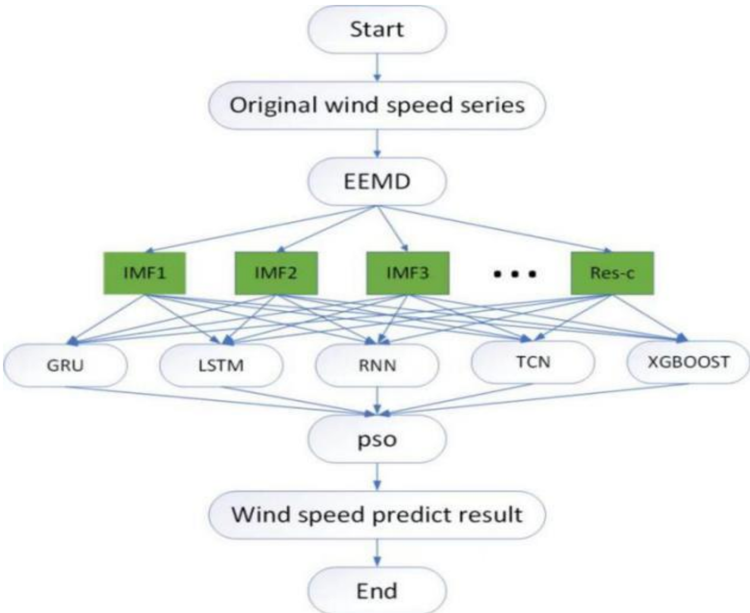

Fig 3. Flowchart of the overall framework of the model.

3.1. Case introduction

This experiment adopted the data from the St. Thomas Anemometer Station in the Virgin Islands from the official website of the United States Energy Administration. The sampling interval was 10 minutes since March 1, 2013, with a total of 3000 pieces of data. In this subject, we used the first 2500 pieces of data as the training set, and used the last 500 pieces of data as the test set.

3.2. Model evaluation index

Considering that the research of this topic is the prediction problem, so the accuracy of prediction is the focus of our research. In this project, we use mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and R^2 score as evaluation indexes [10]and the specific calculation formula is shown in Table I.

Table I. The calculation formulas of the four evaluation indicators.

Indicators	Calculation method
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

3.3. parameter setting

In this project, five prediction algorithms, TCN, RNN, GRU, LSTM, and XGBOOST, are used. The super parameters of the five prediction algorithms are shown in Table II.

Table II. Hyperparameters.

Model	Parameter	value
EEMD	None	
GRU\LSTM\RNN	Optimizer	Adam
	Epoch	100
	Batch_size	64
TCN	Optimizer	Adam
	Epoch	200
	Batch_size	50
	Nb_filter	10
XGBOOST	Booster	gbtree
	Learning_rate	0.2
	Gamma	0.1
	N_estimator	2000

3.4. Verify the validity of the decomposition algorithm

In this project, since the decomposition algorithm can decompose the disorganized data into several gentle data, the experiment adopts the combination of the decomposition algorithm and prediction algorithm to predict the final wind speed value. However, different decomposition methods have different influences on the accuracy of the experiment. Taking the combination of the GRU algorithm and four decomposition algorithms as an example, the accuracy of the four groups of algorithms is shown in Table III.

Table III. Post-combined error values of gru and decomposition algorithm.

GRU group	EMD_ GRU	EEMD_ GRU	CEEMDAN_ GRU	VMD_ GRU
MAE	0.4107	0.2798	0.4121	0.3249
MSE	0.3466	0.1646	0.3673	0.2119
RMSE	0.5887	0.4057	0.6060	0.4604
R^2	0.8641	0.9355	0.8560	0.9169

According to the evaluation indexes MAE, MSE, RMSE and R^2, among the GRU algorithm group, the accuracy of the experimental results predicted by the combination of the GRU algorithm and EEMD decomposition algorithm is the highest. Therefore, in the GRU group, the prediction result of GRU_EEMD is finally chosen as the optimal result. By this means, the optimal results predicted by the five algorithms and their decomposition algorithms are obtained. Similarly, the accuracy of the other four prediction algorithms and four decomposition algorithms is shown in Table IV.

Table IV. Combined algorithm error values.

Model	MAE	MSE	RMSE	R^2
EEMD_LSTM	7.41E-15	8.30E-29	9.11E-15	1.00
EEMD_RNN	0.1673	0.0629	0.2508	0.9753
EEMD_TCN	0.2349	0.1243	0.3525	0.9513
EEMD_XGBOOST	0.5086	0.3373	0.5808	0.8678

Therefore, five better combinations are selected, which are EEMD_GRU, EEMD_LSTM, EEMD_RNN, CEEMDAN_XGBOOST, and EEMD_TCN. The comparison between the five predicted results and the real values is shown in Fig.4.

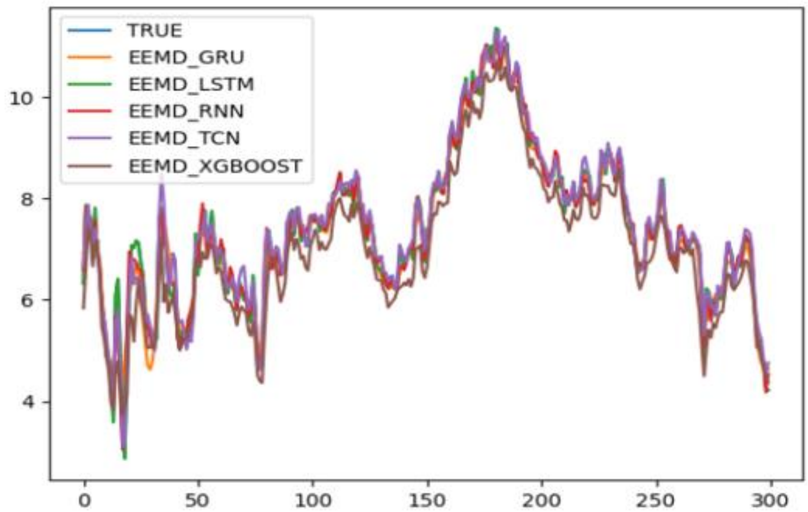

Fig 4. Comparison between the five predicted results and the real values.

As can be seen from Figure 4, the decomposition effect of EEMD decomposition algorithm is better. Therefore, our subsequent experiments will use the EEMD decomposition algorithm uniformly for prediction.

3.5. Combination prediction

The five groups of prediction data selected in the previous step were combined in a weighted manner, MSE was taken as the evaluation index, and used particle swarm optimization algorithm to solve the problem. The combined weight of the five groups of algorithms with the smallest error was obtained.

$$Y = a_1 * x_1 + a_2 * x_2 + a_3 * x_3 + a_4 * x_4 + a_5 * x_5 \quad (18)$$

Among them,x_1), x_2), x_3), x_4 and x_5 are the combined results of five prediction algorithms and the EEMD decomposition algorithm. The final weight results are shown in Table V.

Table V. Weights of combinations.

Model	EEMD_GRU	EEMD_ LSTM	EEMD_ RNN	EEMD_ XGBOOST	EEMD_ TCN
Weight	0.0688	0.9906	0.0189	-0.0919	0.0083

The proposed combined prediction model has strong robustness and can select a model with a good prediction effect for time series. The prediction accuracy is consistent with the highest prediction accuracy. In contrast, the new combined prediction method can also choose the prediction model with higher prediction accuracy,which can maintain the requirement of high precision while transforming the input data. The error values of the combined model are shown in Table VI.

Table VI. Error-values of the combined algorithm.

Combination	MAE	MSE	RMSE	R^2
Accuracy	0.0155	0.0004	0.0192	0.9998

The resulting prediction results against the true value are shown in Fig.5.

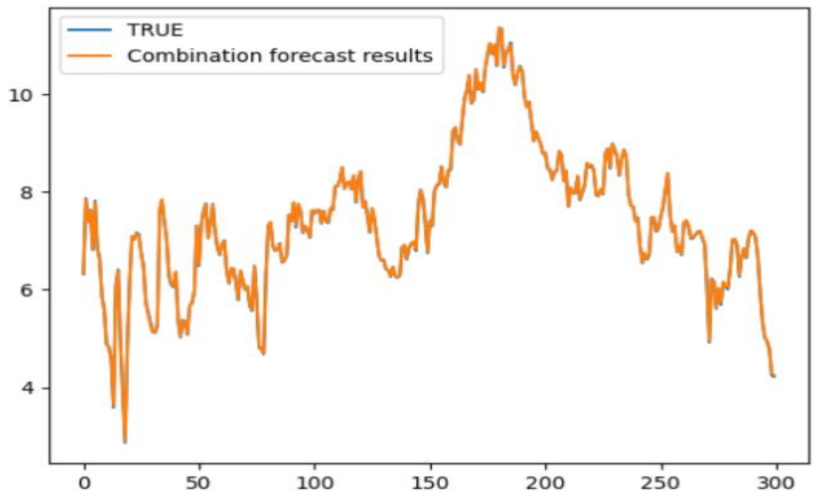

Fig 5. Predict results of the model.

4. Conclusions

In real life, wind power generation has accounted for 13% of the country's total power generation, but wind power generation is characterized by randomness and low density, so it is important to accurately predict the future wind speed to ensure continuous and stable progress of wind power generation.

Based on the two-stage wind speed prediction model of decomposition and combination prediction, we can draw these conclusions:

(1) By posing the raw data into multiple stationary data by a decomposition algorithm, and the idea of "divide and conquer" is adopted to make predictions one by one, which can lower the level of difficulty of prediction and improve the accuracy of the final prediction results.

(2) Using the combination result of multiple optimal prediction results as the best prediction result can improve the robustness of the model. Different weight values can be allocated with the PSO algorithm for different time series to increase the weight of the model with high accuracy and obtain the optimal result.

(3) Combined prediction improves the fitness of the model to new samples to a certain extent and avoids the occurrence of overfitting.

(4) According to the research results, we can see that when facing different wind speed sequences, we can predict the wind speed with high accuracy, which provides a theoretical basis for the stable wind power generation.

References

[1] Yadav G R, Muneender E, Santhosh M. Wind speed prediction using hybrid long short-term memory neural network-based approach[C]// 2021 International Conference on Sustainable Energy and Future Electric Transportation (SEFET). 2021.

[2] Ehsan M A, Shahirinia A, Zhang N, et al. Wind Speed Prediction and Visualization Using Long Short-Term Memory Networks (LSTM)[C]// 2020 10th International Conference on Information Science and Technology (ICIST). 2020.

[3] Zhou B, Ma X, Luo Y, et al. Wind Power Prediction Based on LSTM Networks and Nonparametric Kernel Density Estimation[J]. IEEE Access, 2019, 7:165279-165292.

[4] Li F, Ren G, Lee J. Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks[J]. Energy Conversion and Management, 2019, 186(APR.):306-322.

[5] Xu H, Chang Y, Zhao Y, et al. A novel hybrid wind speed interval prediction model based on mode decomposition and gated recursive neural network[J]. Environmental Science and Pollution Research:1-17.

[6] Kryuchkov Nikita P, Mantsevich Vladimir N, Yurchenko Stanislav O. Interacting Oscillators with Fluctuating Coupling: Mode Mixing without Cross-Correlations. [J]. Physical review letters,2022,129(3)..

[7] Surhone L M, Tennoe M T, Henssonow S F. Long Short-Term Memory[J]. Betascript Publishing, 2010.

[8] Pei S, Qin H, Zhang Z, et al. Wind speed prediction method based on Empirical Wavelet Transform and New Cell Update Long Short-Term Memory network[J]. Energy Conversion & Management, 2019, 196(SEP.):779-792.

[9] Niu Z, Zeyuan Y U, Bo L I, et al. Short-term wind power forecasting model based on deep gated recurrent unit neural network[J]. Electric Power Automation Equipment, 2018.

[10] P. Joshi, Python: Real World Machine Learning. Packt Publishing, 2016.