

# An Optimized Method of Wind Speed Prediction with Support Vector Machine and Genetic Algorithm

Guan-fa Li<sup>1</sup>, Wen-sheng Zhu<sup>2</sup>

<sup>1</sup>Science and Technology College Gannan Normal University, Ganzhou 341000, China

<sup>2</sup>Gannan Normal University, Ganzhou 341000, China

Received: July 30, 2020. Revised: March 5, 2021. Accepted: March 21, 2021. Published: March 30, 2021.

**Abstract**—Due to the randomness of wind speed and direction, the output power of wind turbine also has randomness. After large-scale wind power integration, it will bring a lot of adverse effects on the power quality of the power system, and also bring difficulties to the formulation of power system dispatching plan. In order to improve the prediction accuracy, an optimized method of wind speed prediction with support vector machine and genetic algorithm is put forward. Compared with other optimization methods, the simulation results show that the optimized genetic algorithm not only has good convergence speed, but also can find more suitable parameters for data samples. When the data is updated according to time series, the optimization range of vaccine and parameters is adaptively adjusted and updated. Therefore, as a new optimization method, the optimization method has certain theoretical significance and practical application value, and can be applied to other time series prediction models.

**Keywords**—Wind energy, Wind speed prediction, Support vector machine (SVM), genetic algorithm (GA)

## I. INTRODUCTION

WIND power not only provides clean energy for the society, but also brings a series of new problems to the stable operation of power system. Due to the randomness of wind speed and direction, the output power of wind turbine also has randomness. After large-scale wind power integration, it will bring a lot of adverse effects on the power quality of the power system, and also bring difficulties to the formulation of power system dispatching plan [1-4]. Wind speed prediction is of great significance to reduce the cost of wind power integration and to make a reasonable dispatching plan for the power sector.

Wind speed prediction is to predict the future wind speed

based on the existing data samples, and its prediction accuracy is the key problem. The development of short-term wind speed prediction is generally based on historical data, and its main methods are Kalman filter method [5], time series method [6], artificial neural network [7-8], support vector machine (SVM) [9], etc. Among them, SVM takes structural risk minimization as the goal, and the prediction results are generally more accurate than neural network algorithm based on empirical risk minimization [10]. In addition, SVM is based on small sample learning [11], so it has been widely valued. However, the setting of penalty parameters and kernel parameters has a great impact on the performance of SVM prediction model. In reference [12], genetic algorithm (GA) is used to optimize the parameters of SVM, which effectively improves the accuracy of short-term prediction. However, GA has the problems of slow convergence and premature convergence. In reference [13], Niche algorithm is used to optimize GA algorithm and SVM short-term prediction model is established. This method improves the premature convergence of GA algorithm and improves the prediction accuracy of the model, but the problem of slow convergence speed has not been effectively improved.

In this paper, the optimized genetic algorithm is used to support vector machine. The self-adaptive update of vaccine factor and parameter optimization range are studied based on the parallelism of genetic algorithm, the self-organization and self-adaptive mechanism in the process of optimization, the information processing mechanism of vaccine factor in immune algorithm, the maintenance mechanism of niche algorithm for population diversity, and the time-series variation of wind speed data. A SVM short-term wind speed prediction model based on optimization genetic algorithm is proposed. The example results show that the SVM based on the optimized genetic algorithm has higher prediction accuracy and shorter prediction time, which proves that the SVM optimized by the optimized genetic algorithm is feasible in short-term wind speed prediction.

## II. SUPPORT VECTOR MACHINE

The basic idea of support vector machine (SVM) is to find a real valued function  $g(x)$  on  $R^n$ , and infer the output value  $y$  corresponding to any input  $x$  with  $y=g(x)$ . For data samples  $(x_1, y_1), (x_2, y_2) \dots (x_m, y_m)$ ,  $i=1, 2, \dots, m$ ,  $m$  are the number of training samples, and  $R$  is a set of real numbers, so the linear regression function is shown in formula (1).

$$y(x) = \langle \omega, \varphi(x) \rangle + b \quad (1)$$

In this formula,  $\omega$  is the weight vector,  $\varphi(x)$  is the transformation of the  $n$ -dimensional vector  $x$  from nonlinear mapping to Hilbert space in two-dimensional space, and  $b$  is a constant.

According to the principle of structural risk minimization, formula (1) is transformed into formula (2),

$$\min_{\omega, b, \xi^*} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \quad (2)$$

$$\begin{cases} (\langle \omega, \varphi(x_i) \rangle + b) - y_i \leq \varepsilon + \xi_i \\ y_i - (\langle \omega, \varphi(x_i) \rangle + b) \leq \varepsilon + \xi_i^* \\ \xi_i^* \geq 0 \end{cases}$$

Among them, the value of relaxation variable  $\xi_i^*$  represents the distance of outliers at corresponding points, and the size of penalty parameter  $C$  represents the importance of outliers in regression model. The introduction of  $C$  value finds a balance between the generalization and accuracy of the regression model. The reasonable choice of  $C$  value has a profound impact on the generalization performance of the regression model.

In order to find out the optimization problem expressed by formula (2), Lagrange multiplier vector is introduced and transformed into formula (3).

$$\min_{a^* \in R} \frac{1}{2} \sum_{i,j=1}^m (a_i^* - a_i)(a_j^* - a_j) K(x_i, x_j) + \varepsilon \sum_{i=1}^m (a_i^* + a) - \sum_{i=1}^m y_i (a_i^* - a_i) \quad (3)$$

$$s.t. \begin{cases} \sum_{i=1}^m (a_i^* - a_i) = 0 \\ 0 \leq a_i^* \leq C \end{cases}$$

In formula (3),  $a^* = (a_1^*, a_2^*, \dots, a_m^*)^T$  is the Lagrange multiplier vector,  $K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$  is the kernel function, which represents the inner product value of the

transformation function.

Finally, the decision function is constructed by the optimal solution  $\bar{a}^* = (\bar{a}_1^*, \bar{a}_2^*, \dots, \bar{a}_m^*)^T$ , as shown in formula (4).

$$f(x) = \sum_{i=1}^m (\bar{a}_i^* - \bar{a}_i) K(x_i, x_j) + \bar{b} \quad (4)$$

Where  $\bar{a}_i^*$  and  $\bar{a}_i$  are Lagrange multiplier components in the open interval  $(0, C)$ . The value of  $\bar{b}$  is determined by the  $\bar{a}^*$  component between  $(0, C)$ . If  $\bar{a}_j$  is selected, its value is shown in formula (5).

$$\bar{b} = y_j - \sum_{i=1}^m (\bar{a}_i^* - \bar{a}_i)(x_i \cdot x_j) \quad (5)$$

There are polynomial kernel, Gaussian radial basis function (RBF) kernel and Sigmoid kernel in support vector machine. Because RBF kernel directly reflects the distance between two data vectors, it is widely used. Therefore, this paper selects Gaussian radial basis function as the kernel function of SVM regression model, and the formula of RBF kernel function is shown in formula (6).

$$K(x, x') = \exp\left(-\|x - x'\|^2 / \delta^2\right) \quad (6)$$

After the kernel function is determined, the parameters involved in wind energy regression fitting prediction by SVM are penalty parameter  $C$  and kernel parameter  $\delta$ .

## III. PARAMETER OPTIMIZATION BASED ON OPTIMIZED GENETIC ALGORITHM

Genetic algorithm (GA) is a stochastic global search and optimization algorithm suitable for complex system optimization, which originates from the simulation of biological evolution mechanism.

Niche algorithm, through the simulation of competition and cooperation among different species in the niche environment, compares the similarity between individuals in the population in a given environment, and adjusts the fitness of individuals, so as to achieve the purpose of maintaining the diversity of the population.

Vaccine factor in IA (Immune Algorithm) is a kind of feature information which contains one or more continuous genes extracted from the prior knowledge of the specific problem to be solved. It can also be regarded as an estimation of the best individual matching pattern to be solved.

The standard genetic algorithm has no restrictions on the fields and types of problems. It is a parallel search method that does not depend on specific problems. It has fewer parameters,

but it has obvious disadvantages. After several iterations, inbreeding will be formed among individuals with high fitness in the population, resulting in premature and reduced diversity. The initial value of the population is generated randomly, which takes a long time to get the result from the beginning to the end, and the efficiency is low.

In the process of iteration, niche algorithm adjusts individual fitness, which can effectively avoid the situation that individuals with larger fitness occupy the whole population quickly. The vaccine factor in immune algorithm is used to immunize the initial population, which can effectively reduce the evolution time of the population and improve the search efficiency. Therefore, niche algorithm and immune algorithm are introduced into genetic algorithm, and an optimized genetic algorithm is proposed. The principle of the algorithm is as follows.

The vaccine factor is used to guide the generation of the initial population, so that the individuals in the initial population have some segments of excellent individuals. After the genetic operators (selection, crossover and mutation) are calculated, the fitness distance between individuals is compared. If the distance is within the niche radius, the individuals with small fitness are punished to ensure the diversity of the population and avoid the premature phenomenon in the genetic algorithm.

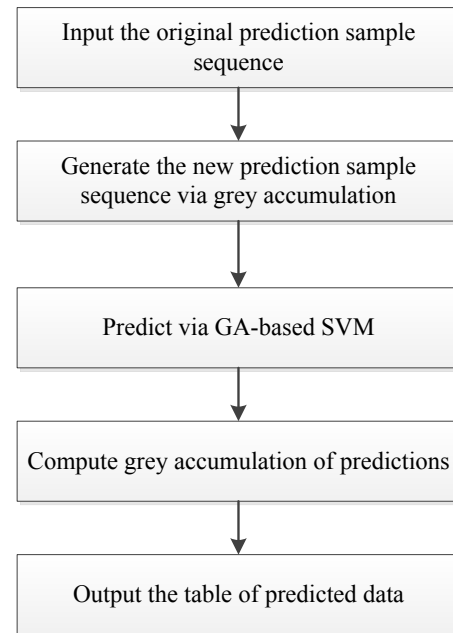
Optimization genetic algorithm has three core parts:

1) For the selection of niche radius  $L$ , the method of reference [13] is used to set. It can adaptively adjust its size according to the evolution of population.

2) In the optimization genetic algorithm, the individual with the largest fitness searched by niche genetic algorithm can be considered as the optimal value of parameters  $(C, \delta)$ , which is used as the initial vaccine to inoculate the initial population of the optimization genetic algorithm, and the optimal solution obtained is used as the vaccine factor.

3) The parameter optimization range can automatically update. the wind speed data changes according to the time series. When the data is updated, the parameter optimization range must also be updated. If the design of parameter  $(C, \delta)$  optimization range is larger, the optimization time of the parameter will increase, and if the design of parameter  $(C, \delta)$  optimization range is smaller, the parameter will converge to the local optimal solution. In order to achieve a balance between the search speed and the search accuracy of the algorithm, the position of the vaccine factor (optimal parameters  $(C, \delta)$ ) in the parameter range is determined at the initial stage of the algorithm. If it is within a certain range of the minimum value, the setting range of the parameters will be reduced correspondingly. If it is within a certain range of the maximum value, the range of the corresponding parameters will be increased, otherwise, the optimization range of the parameters will remain unchanged.

The wind speed can be predicted using the proposed method as shown in Fig.1.



**Fig.1** The proposed method for the prediction of wind speed

The implementation steps of the algorithm are as follows.

Step 1. Initialize the control parameters (population size, genetic algebra, crossover and mutation probability, etc.) and the optimization range of parameters  $(C, \delta)$ .

Step 2. Judge whether the vaccine factor reaches a certain extreme value of the optimization range of relevant parameters  $(C$  or  $\delta)$ , and adjust the optimization range accordingly.

Step 3. The vaccine factor is used to guide the generation of the initial population, and the fitness of individuals in the population is calculated.

Step 4. Judge the ending condition. If the ending condition is satisfied, store the code of the individual with the best fitness in the vaccine factor and output the optimization result. Otherwise, turn to step 5.

Step 5. Carry out selection, crossover, mutation and other genetic operations.

Step 6. Compare the Gauss distance  $d$  between two individuals. If  $d$  is within  $L$ , punish those with small fitness.

Step 7. Rank the adaptability of each individual according to the new fitness value, update  $M$  with the minimum Gauss distance of the first  $m$  individuals, and go to step 4.

#### IV. EXAMPLE ANALYSIS

In order to verify the effectiveness of the optimized genetic algorithm in the search accuracy and speed, the program algorithm is compiled, and the parameters are set as follows:

population size  $N=50$ , genetic number  $Gen=300$ , crossover probability  $P_c = 0.8$ , binary code length  $2 \times 8$ . According to the research of reference [14], when  $C$  is fixed and  $\delta$  is increased from 0 to a certain range, the test accuracy increases rapidly while the training accuracy decreases slightly. When  $\delta$  is fixed and  $C$  increases from 0 to a certain range, the accuracy of test and training increases rapidly with the increase of  $C$ . Therefore, according to the results of many experiments, the initial optimization range of parameters after data normalization is set as  $C$  [0,20] and  $\delta$  [0,8]

The hourly wind speed data measured by a 100 m high anemometer tower in a certain area is used, and the data collection time is from January 2020 to December 2020. 501 data were randomly selected according to the time sequence, the first 285 data were used to establish the prediction model, and the last 48 data were predicted. For the same wind speed data samples, GA optimization in reference [15], NGA optimization in reference [16], SVM prediction method in reference [17], and BP neural network prediction method are used for prediction.

The mean absolute error (MAE) and mean standard error (MSE) can be expressed as

$$MAE = \frac{1}{k} \sum_{i=n+1}^{n+k} \frac{|s_i - \hat{s}_i|}{s_i} \times 100\% \quad (7)$$

$$MSE = \sqrt{\frac{1}{k} \sum_{i=n+1}^{n+k} \left( \frac{|s_i - \hat{s}_i|}{s_i} \right)^2} \quad (8)$$

The prediction time and absolute error values of each method are shown in Table 1.

Table 1. The prediction times and errors of wind speed (m/s) forecasting by different algorithm

Index	GA-SVM	SVM	BPNN	NGA	GA
$T/s$	119.06	36.55	29.71	50.23	42.36
$e$	3.81%	9.05%	13.25%	16.25%	18.97%

It can be seen from Table 1 that the optimized genetic algorithm in this paper can find more suitable parameters for the given data samples, and the average absolute percentage prediction error  $e$  is significantly less than the  $e$  obtained by BP neural network and SVM with directly setting relevant parameters, and the SVM prediction error smaller than GA optimization and NGA optimization with changing niche radius. Compared with BP neural network and SVM without optimization, the optimization time is longer, but shorter than

GA and NGA.

At the same time, in order to verify the prediction accuracy of the model after the optimization range of vaccine factors and parameters is automatically updated with the change of data, 1-60 samples (20 data are updated each time and completed in three times) after the selection of training data are tested. The specific prediction results are shown in Figures 2 to 3, and the prediction time and prediction error are shown in Table 2.

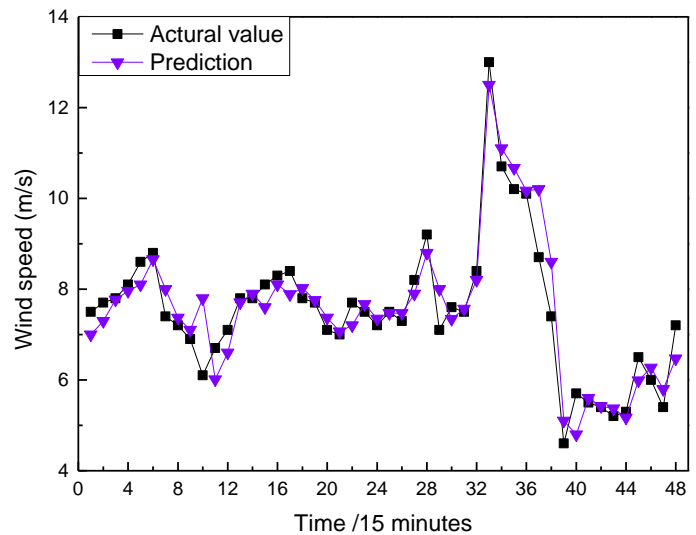


Fig.2 The prediction result of update data

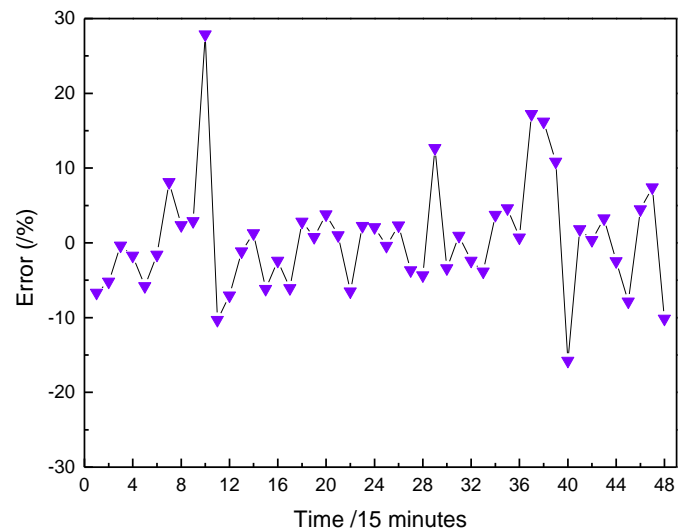


Fig.3 The prediction error for update data

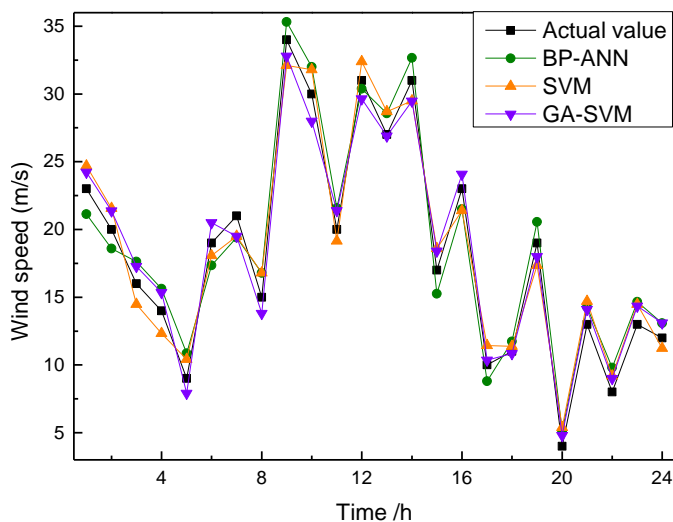
Table 2. The prediction error and times for data updated

Sample number	1-20	21-40	41-60

$T/s$	75.67	78.23	79.72
$e$	3.02%	7.63%	9.47%

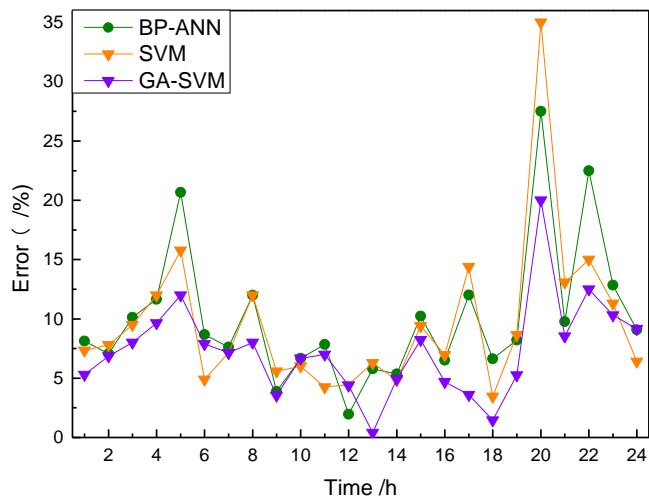
It can be seen from Fig. 2 ~ Fig. 3 and Table 2 that in the wind speed prediction model, when the wind speed data series is updated, the parameters optimization range and vaccine factors are automatically updated to guide the generation of the initial population, which can not only reduce the optimization time, but also reduce the error of wind speed prediction and improve the prediction accuracy.

The relative analysis from several prediction algorithms mentioned is done to evaluate superiority in Figure 4. The error of GA-SVM is the smallest in those prediction algorithms.



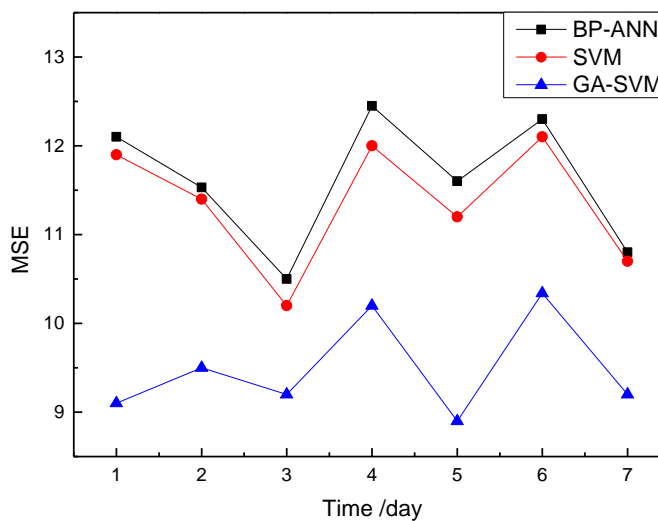
**Fig.4** The prediction value analysis from several prediction algorithms

The prediction error analysis from several prediction algorithms are shown in Figure 5. The error distribution range of GA-SVM is 0.38% -12.54%, BP-ANN is 2.16% - 18.31%, and SVM is 1.76% - 15.02%. Therefore, GA-SVM has higher prediction accuracy.



**Fig.5** The prediction error analysis from several prediction algorithms

The MAE and MSE of several prediction algorithms are shown in Figure 6 and Figure 7. The GA-SVM algorithm is lowest for MAE and MSE value than several prediction algorithms.



**Fig.6** MSE value of several prediction algorithms

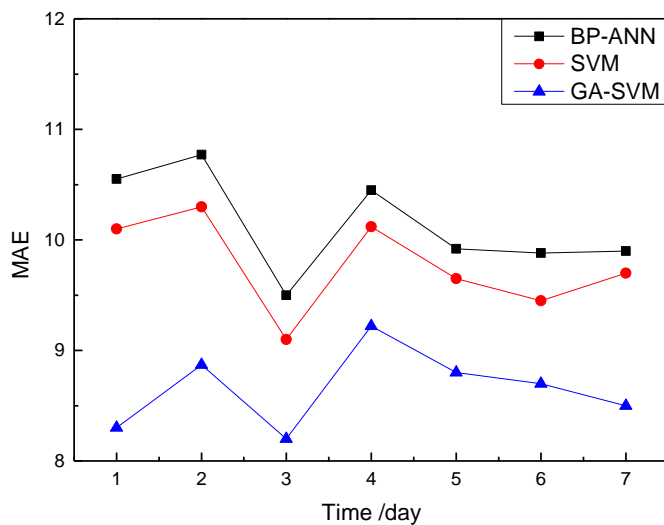


Fig.7 MAE value of several prediction algorithms

## V. CONCLUSION

In this paper, we will optimize the SVM based on the optimized genetic algorithm. In short-term wind speed forecasting, niche algorithm is used to maintain the diversity of population, combined with the global convergence ability of genetic algorithm, and the vaccine factor of immune algorithm is introduced to guide the generation of initial population, which can reduce the search time and ensure the diversity of population, avoid the premature and slow optimization of genetic algorithm, and improve the speed and accuracy of optimization.

Compared with other optimization methods, the simulation results show that the optimized genetic algorithm not only has good convergence speed, but also can find more suitable parameters for data samples. When the data is updated according to time series, the optimization range of vaccine and parameters is adaptively adjusted and updated. Therefore, as a new optimization method, the optimization method has certain theoretical significance and practical application value, and can be applied to other time series prediction models.

In the future, some novel models can be constructed by much related research and the accuracy of prediction results can be improved.

## Acknowledgement

This research is supported by Scientific and Technological Research Project of Jiangxi Provincial Department of Education "Application Research of Face Recognition based on Embedded Zerotree Wavelet" (Grant No.: J61635).

## REFERENCES

- [1] Lee J, Seo G, Mun J, et al. Thermal and Mechanical Design for Refrigeration System of 10 MW Class HTS Wind Power Generator. IEEE Transactions on Applied Superconductivity, 2020, 30(99), pp.1-5.
- [2] Le K X , Huang M J , Wilson C , et al. Tariff-based load shifting for domestic cascade heat pump with enhanced system energy efficiency and reduced wind power curtailment. Applied Energy, 2020, 257, pp. 106-115.
- [3] Aslam M . Analyzing Wind Power Data using Analysis of Means under Neutrosophic Statistics. Soft Computing, 2021, In Press.
- [4] Ohtake H . Future prediction of solar power and wind power resources of using climate models for 2050. Atomos, 2020, 62(2), pp.84-88.
- [5] Samadianfard S, Hashemi S, Kargar K, et al. Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. Energy Reports, 2020, 6, pp.1147-1159.
- [6] Xu W, Liu P, Cheng L, et al. Multi-step wind speed prediction by combining a WRF simulation and an error correction strategy. Renewable Energy, 2021, 163, pp.772-782.
- [7] Cai H , Jia X , Feng J , et al. Gaussian Process Regression for numerical wind speed prediction enhancement. Renewable Energy, 2020, 146(2), pp.2112-2123.
- [8] Vinothkumar T , Deeba K . Hybrid wind speed prediction model based on recurrent long short-term memory neural network and support vector machine models. Soft Computing, 2020, 24(7), pp.5345-5355.
- [9] Shivam K, Tzou J C, Wu S C. Multi-Step Short-Term Wind Speed Prediction Using a Residual Dilated Causal Convolutional Network with Nonlinear Attention. Energies, 2020, 13(7), pp.1772.
- [10] Chao A I , Wei G , Lijuan C , et al. Research on the Speed Control of Hydraulic Wind Turbine Based on Wind Speed Prediction. Journal of Mechanical Engineering, 2020, 56(8), pp.162.
- [11] Liu H , Duan Z . A vanishing moment ensemble model for wind speed multi-step prediction with multi-objective base model selection. Applied Energy, 2020, 261:114367.
- [12] Nie Y, Bo H, Zhang W, et al. Research on Hybrid Wind Speed Prediction System Based on Artificial Intelligence and Double Prediction Scheme. Complexity, 2020, 2020, pp.1-22.
- [13] K. G. Sheela, S.N. Deepa, "Neural network based hybrid computing model for wind speed prediction", Neurocomputing, vol. 122, no. 25, (2013) December, pp. 425-429.
- [14] Palomino K , Reyes F , José Núez, et al. Wind Speed Prediction Based on Univariate ARIMA and OLS on the Colombian Caribbean Coast. Journal of Engineering Science and Technology Review, 2020, 13(3), pp.200-205.
- [15] Mohamed S , Ndiaye D , Sidi M , et al. Neuron Network Prediction Feed-Forward Wind Speed Network on Mauritania's North Coast: Ballawack Case. International Journal of Innovative Technology and Exploring Engineering, 2020, 9(12), pp.2278-3075.
- [16] Mariia Ruda, Taras Boyko, Camillo La Mesa, Computer Simulation of the Influence of Wind Power Plants on The Compartments of The Complex Landscape System by The Method of Life Cycle Assessment , Engineering World, Volume 1, 2019, pp. 34-50.
- [17] Tuan Ngoc Anh Nguyen, Duy Cong Pham, Cong-Thanh Pham, Nguyen Huu Chan Thanh, D-axis Stator Current Control Methods Applied to PMSG-based Wind Energy Systems: A Comparative Study, WSEAS Transactions on Systems and Control, Volume 14, 2019, Art. #31, pp. 239-246.

**Creative Commons Attribution License 4.0  
(Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)