All Sciences Proceedings

2nd International Conference on Scientific and Academic Research

March 14-16, 2023 : Konya, Turkey

ICSAR

All Sciences Proceedings
http://as-proceeding.com/

© 2023 Published by All Sciences Proceedings

Application of Machine Learning on Wind Blowing Speed

Akin Ilhan*

[1] Department of Energy Systems Engineering, Faculty of Engineering and Natural Sciences, Ankara Yildirim Beyazit University, 06010, Ankara, Turkey, ORCID ID: 0000-0003-3590-5291

*(ailhan@ybu.edu.tr)

Abstract **–** The estimation of the meteorological parameters such as temperature, pressure, humidity, sea wave height, sea wave speed, and wind speed gains importance as it provides us with information about the unknown future values of these parameters. In this study, the measured sea shore wind speed data is predicted using a variety of machine learning algorithms. Accordingly, these wind speed data are obtained considering daily measured wind speed values of a sea shore region found in Turkey. A total of 1,415 data found in the wind speed data cluster has been utilized in forecasts depending on the historical-time series. The algorithms used for this purpose include long-short term memory, adaptive neuro-fuzzy inference system with fuzzy c-means (FCM), subtractive clustering (SC), and grid partitioning (GP). A cumulative of 66 different models have been structured using these four tools. The quality of the forecasts has been compared according to the statistical accuracy errors comprising of mean absolute error (MAE), root mean square error (RMSE), as well as the correlation coefficient (*R*). In this context, the analyses of the statistical errors of the computations have indicated that the best wind speed forecast for this sea shore region is obtained under the LSTM tool. Ultimately, the MAE, RMSE and *R* values of this best model were calculated to respectively correspond 0.2477 m/s, 0.3315 m/s, 0.9951.

Keywords – LSTM, ANFIS, Artificial Intelligence, Wind Speed Prediction, Time-Series

I. INTRODUCTION

With the increase in greenhouse gas emissions and the increase in global warming caused by these gases in recent years, and due to the rapid depletion of fossil fuels, power generation using renewable resources has shown a great increase throughout all World. For this reason, power plants powered by renewable resources have a very significant spreading potential compared to conventional fossil fuel powered power plants. Recently, most of the countries in all of the World have the trend of a rapid shift to renewable resource powered power plants, and in accordance with this, almost all of the World countries have at least one type of power plant operating with renewable source [1]. Wind energy, which is among these types, is one of the most important type today and is used in power generation in many countries where the wind is of sufficient quality. In this regards, World's total installed onshore and a recent technology offshore installed wind power are respectively presented in Figs. 1 and 2. As shown in these figures, the analysis of installed wind power of the World is given considering the last two decades, i.e., considering the years between 2001 and 2021 [2].

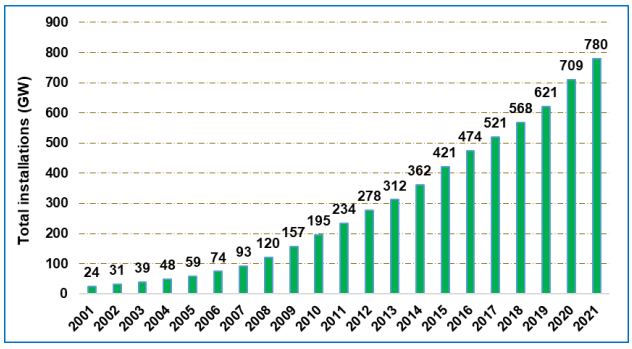

Fig. 1 Installed onshore wind power of the World

The onshore installations considering worldwide situation indicates that the power installations have rapidly increased from 24 GW to 780 GW, in a time interval of 20 years, as presented in Fig. 1. This corresponds to an enormous increase of about 32 times. On the other hand, the offshore technology which is a more recent technology is also rapidly increasing in terms of the installations, especially in the last decade. As demonstrated in Fig. 2, the total offshore installations were only 3 GW by the end of 2010, however, after 2010 till reaching the end of 2021, the cumulative offshore wind power installations have risen to 57 GW. This increase also corresponds to a record 19-fold increase. Investments in this technology are increasing rapidly as it provides the opportunity to capture better quality and high-speed wind energy in the open sea with this technology and to convert it into useful electrical energy [3]. And, with the development of technology, the tower heights and rotor diameters of offshore wind turbines in the open sea also increase very rapidly; resulting more capture of the wind energy transformed to the electrical energy.

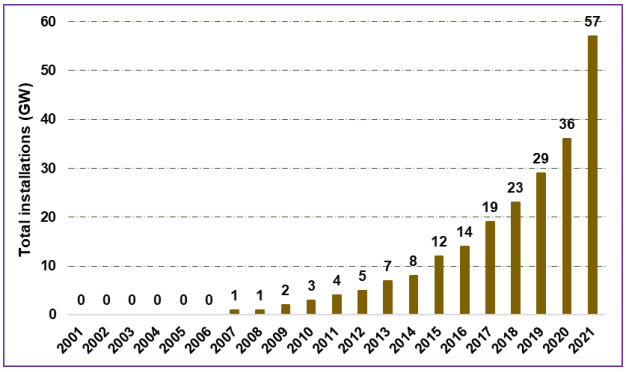

Fig. 2 Installed offshore wind power of the World

II. Materials and Method

A. *Long-short term methods (LSTM)*

The predictions studies using LSTM were initiated in 1997 [4]. LSTM is a type of recurrent neural network, in which the solutions are obtained in the algorithm by adding memory cells or the cell states, with constant errors. Thus, without disappearing the gradients, the errors could be reconstituted. In an LSTM, three different gates generally form the structure of the model, and those are the input, output as well as the forget gates. Different learning activities of LSTM are controlled by these three gates.

An LSTM layer consists of input and recurrent weights, besides the bias. These are the learnable weights of the algorithm. The abbreviations of *W*, *R*, and *b*, demonstrated in Eq. (1), respectively indicate these weights as well as the bias. The combinations of the input weights, recurrent weights and the bias are presented with these expressions in the matrices given in this equation.

$$W = \begin{bmatrix} W_i \\ W_f \\ W_g \\ W_o \end{bmatrix}, R = \begin{bmatrix} R_i \\ R_f \\ R_g \\ R_o \end{bmatrix}, b = \begin{bmatrix} b_i \\ b_f \\ b_g \\ b_o \end{bmatrix} \quad (1)$$

B. *Adaptive Neuro Fuzzy Inference System (ANFIS)*

In a wide range of engineering applications, adaptive neuro fuzzy inference system (ANFIS) is utilized, and it is one of the well-known universal estimator. A true continuous function is required for the ANFIS to be adapted to the problem, and it generates a high accuracy and success in a compact set-up. A grid structure that is in the form of Sugeno-type fuzzy systems forms the basis of ANFIS. But, the grid structure is reinforced with neural learning capabilities. Neural network learning algorithm is employed from input to output to create if-then rules utilizing proper membership functions (MFs). The development operation of the fuzzy inference system utilizing the adaptive neural network is known to be ANFIS [5].

B1. Subtractive Clustering (SC) Algorithm

In this method, each data point is considered a possible cluster center and the potential of each data point is evaluated by evaluating the density of data points surrounding the cluster center. Based on the location of the data point according to the other data points, the SC tool utilizes an adaptive process and supposed any data point could be the center of the cluster [5].

B2. Fuzzy C-Means (FCM) Algorithm

Fuzzy c-means (FCM) is a clustering technique allowing every data point to have multiple files and permits the data point to belong to different degrees of membership. In FCM, the objective function is needed to be minimized forming the basis of the algorithm. General and elaborate knowledge regarding the FCM tool of the ANFIS is found in the literature [6].

B3. Grid Partitioning (GP) Algorithm

Axis-paralleled partition is used to split the input data space to a rectangular subspace in the GP tool

of the ANFIS. The division of each input is achieved by splitting the inputs into MFs of identical shape. The number of the rules of the fuzzy if-then is determined using the mathematical term of M^n. In this term, while n indicates the input dimension, whereas, the M represents the amount of partitioned fuzzy subsets of each input variable [1]. Shortly in GP, in order to solve a technical or engineering problem, the functional decomposition is performed. Wide explanations of GP in the literature are also found with the applied studies executed in this direction [7].

C. *Statistical Error Analysis*

To obtain the quality of the artificial intelligence or machine learning computations, three statistical error parameters are used in the current study. Those include the mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (R). As MAE and RMSE statistical errors approach zero or as R correlation results approach 1, the quality of the estimations increase. In such situations, the prediction results will be more accurate and correlated with respect to the real measured or observed wind speed data. In this regards, Eq. (1) gives the mathematical relation of MAE, Eq. (2) presents the RMSE equation, and finally the correlation coefficient (R) is indicated by Eq. (3) [8].

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|p(i) - o(i)| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[p(i) - o(i)]^2} \quad (2)$$

$$R = \left(\sum_{i=1}^{N}[p(i) - \bar{p}][o(i) - \bar{o}]\right) / \left(\sqrt{\sum_{i=1}^{N}[p(i) - \bar{p}]^2}\sqrt{\sum_{i=1}^{N}[o(i) - \bar{o}]^2}\right) \quad (3)$$

In these three equations, the i, N, $o(i)$, and $p(i)$ abbreviations respectively correspond to the data order, the total members of the studied data set, the measured or observed instantaneous data member in the data set, the predicted instantaneous data member in the data set. On the other hand, $\bar{o}$ and $\bar{p}$ symbols only presented in Eq. (3), correspond to mean value of all measured or observed data in the cluster and average value of all predicted data in the cluster, respectively [8].

III. RESULTS

Recently, prediction of the atmospheric data has gained a lot of importance. In this way, unrealized weather events can be easily predicted and more importantly, measures can be taken against some disasters such as floods or hurricanes. Predictions related with the air temperature and precipitation data, in the travel of road vehicles and airplanes; estimates related with the sea and oceans such as sea wave height or speed as well as wind speed are frequently used today in the travels of ships or ferries. Of course, there are many other benefits of data estimates that are not covered here. Accordingly, the current study was performed utilizing a cumulative of 1,415 daily wind speed instantaneous data measured at the selected sea shore location found in Turkey. Among these data, a portion corresponding to 80% and 1,132 amounts was used to train the machine learning algorithms. Besides, the portion corresponding to 20% and 281 amounts was utilized for the testing of these algorithms. The algorithms considered in the machine computations of the current study include ANFIS-FCM, ANFIS-SC, ANFIS-GP and LSTM. Fig. 3 presents all measured instantaneous wind speed data, whereas, Fig. 4 shows the continuous functional distribution of the testing part corresponding to 20% of the total data cluster.

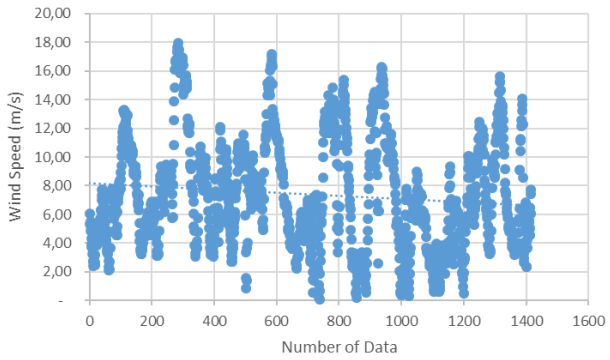

Fig. 3 The total data cluster including all instantaneous wind speed values

Fig. 4 The testing data cluster of instantaneous wind speed values

In LSTM type of computations, hidden layer (hl) is taken into account of the range $5 \leq hl \leq 300$. The maximum epoch number used for the amount

of iterations to converge the result in the computations, on the other hand is set to 300. In ANFIS type of modelling, the computations are performed according to the amount of the considered number of historical data to train the algorithm. Accordingly, the time-series functional relation is structured according to the selected number of historical data (*hd*). Three different tools of ANFIS are taken into account. In FCM of ANFIS algorithms, the *hd* is adjusted to be in the range of $3 \leq hd \leq 10$. Besides, the increment of $\Delta hd = 1$ is considered in this range. Whereas, the amount of the membership functions (*MFs*) is studied in the range of $2 \leq MFs \leq 10$, in which the increment of $\Delta MFs = 2$ has been considered in the membership functions. Therefore, for every *hd* value in the considered range of the number of historical data used for training, five different *MFs* is studied for FCM tool of the ANFIS. Conversely, in SC approach of the ANFIS, the *hd* is studied in a same way as in the case of FCM. But, the influence radius (*ir*) needed in this algorithm is considered for the range of $0.2 \leq ir \leq 0.9$. Ultimately, in GP algorithm of ANFIS, the *hd* and *MFs* are considered and studied within the ranges of $3 \leq hd \leq 4$ and $2 \leq MFs \leq 4$, respectively.

Among these four algorithms, a cumulative of 9, 40, 12, and 5 models have been designed and studied, respectively, in LSTM, FCM, SC, and GP tools. Accordingly, a total of 66 computation models have been generated; and the success of those designed models were compared using the statistical accuracy parameters including MAE, RMSE, and *R*. Initially, the LSTM modelling outcomes are exhibited in Table 1. Besides, the results of FCM of ANFIS have been demonstrated in Table 2. The computation outcomes of SC algorithm are presented in Table 3. And finally, for GP algorithm, the results of the computations have been indicated in Table 4. The setting parameters to initiate and continue of the computations for these cited algorithms are also revealed in those tables.

Table 1. The outcomes of LSTM computations

Hidden layer	Max. Epoch	MAE (m^3/s)	RMSE (m^3/s)	R
5	**300**	**0.2477**	**0.3315**	**0.9951**
10	300	0.2652	0.3537	0.9944
25	300	0.2552	0.3502	0.9946
50	300	0.2756	0.3964	0.9930
75	300	0.3147	0.4684	0.9902
100	300	0.4114	0.7165	0.9786
125	300	0.5334	0.8234	0.9697
150	300	0.5459	0.8986	0.9651
300	300	1.2203	1.5553	0.8906

Table 2. The outcomes of ANFIS-FCM computations

Historical data	MFs	Max. epoch	MAE	RMSE	R
3	2	100	0.25511248	0.34253408	0.99476200
	4	100	0.26921136	0.35271669	0.99444300
	6	100	0.30863154	0.40409985	0.99282700
	8	100	0.33235457	0.46840145	0.99023600
	10	100	0.38580892	0.52230578	0.98776800
4	2	100	0.25511249	0.34253408	0.99476200
	4	100	0.26921005	0.35271785	0.99444300
	6	100	0.30862469	0.40408246	0.99282800
	8	100	0.33254944	0.46866197	0.99022600
	10	100	0.36855887	0.51206622	0.98827500
5	**2**	**100**	**0.25511248**	**0.34253407**	**0.99476200**
	4	100	0.26921002	0.35271779	0.99444300
	6	100	0.30866813	0.40418039	0.99282400
	8	100	0.33227318	0.46830750	0.99023900
	10	100	0.36521278	0.47252814	0.99007900
6	2	100	0.25511249	0.34253408	0.99476200
	4	100	0.26921007	0.35271794	0.99444300
	6	100	0.30862567	0.40408565	0.99282700
	8	100	0.33225568	0.46828231	0.99024000
	10	100	0.38373863	0.51935526	0.98790900
7	2	100	0.25511248	0.34253408	0.99476200
	4	100	0.26921278	0.35271986	0.99444300
	6	100	0.30865989	0.40416182	0.99282500
	8	100	0.33226601	0.46829460	0.99023900
	10	100	0.38399584	0.51993631	0.98788300
8	2	100	0.25511247	0.34253409	0.99476200
	4	100	0.26921001	0.35271793	0.99444300
	6	100	0.30866547	0.40417555	0.99282400

	8	100	0.33222676	0.46823882	0.99024100
	10	100	0.38591886	0.52395578	0.98769300
9	2	100	0.25511248	0.34253408	0.99476200
	4	100	0.26920993	0.35271779	0.99444300
	6	100	0.30866667	0.40417836	0.99282400
	8	100	0.33222602	0.46823339	0.99024200
	10	100	0.36477537	0.47199906	0.99010100
10	2	100	0.25511248	0.34253409	0.99476200
	4	100	0.26920981	0.35271777	0.99444300
	6	100	0.30866953	0.40418278	0.99282400
	8	100	0.33226587	0.46829440	0.99023900
	10	100	0.38593886	0.52310847	0.98773500

It has been demonstrated that the FCM computations have resulted the discrepancies between two computations had mostly occurred with respect to the change of the *MFs* more than the change of the number of the historical data. Shortly, the statistical accuracy results were not influenced much according to the number of the historical data, but much influenced according to the *MFs*. In order to clearly reveal the differences between the statistical errors of the computations; the number of the significant digits after the decimal point for every computation is shown as 8.

In Table 3, i.e., considering SC computations of the ANFIS, while the alteration of the historical data didn't influence the statistical accuracy outcomes, the change of the influence radius has resulted some alterations of the statistical MAE, RMSE, and *R* results. For this reason, in Table 3, the statistical accuracy outcomes of *hd* values for SC computations, corresponding to only 3 and 4 have been revealed. In this regards, although the computations corresponding to *hd* in the range of $5 \leq hd \leq 10$ have been performed, it has been observed that further computations including *hd* values in this range didn't improve the computation results. Thus, only the statistical error results of *hd* in the range of $3 \leq hd \leq 4$ have been shown in this table. Already in this value range of *hd* values, statistical error results do not change when *hd* value is changed from *hd*=3 to *hd*=4; however, when the *ir* values are changed, it is seen in this table that statistical error results change slightly.

Table 3. The outcomes of ANFIS-SC computations

Historical data	Influence radius	Max. epoch	MAE	RMSE	R
3	0.2	100	0.5136	0.8800	0.9661
	0.3	100	0.4410	0.6360	0.9823
	0.4	100	0.3320	0.5229	0.9878
	0.6	**100**	**0.2895**	**0.3870**	**0.9933**
	0.8	100	0.2960	0.3968	0.9929
	0.9	100	0.2899	0.3915	0.9931
4	0.2	100	0.5136	0.8800	0.9661
	0.3	100	0.4410	0.6360	0.9823
	0.4	100	0.3320	0.5229	0.9878
	0.6	100	0.2895	0.3870	0.9933
	0.8	100	0.2960	0.3968	0.9929
	0.9	100	0.2899	0.3915	0.9931

Table 4. The outcomes of ANFIS-GP computations

Historical data	MFs	Max. epoch	MAE	RMSE	R
3	**2**	**30**	**0.8896**	**1.2055**	**0.9329**
	3	30	0.9128	1.2548	0.9281
	4	30	0.9359	1.2935	0.9233
4	2	30	0.8896	1.2055	0.9329
	3	30	0.9128	1.2548	0.9281

The best results of 66 computations of four algorithms provided in Tables 1, 2, 3, and 4 are shown by bold colour in these tables. In this context, these best results are indicated in Table 5, for quick visualization. Besides, the best overall result among these best results obtained from designed 66 models of four algorithms is indicated in Table 5, with bold colour, i.e., the statistical accuracy results of the LSTM tool is given in the table using bold colour. In this regards, the computation result obtained at 5 *hl* demonstrates that 0.2477 m/s MAE, 0.3315 m/s RMSE, and 0.9951 *R* statistical accuracy results have been obtained. The prediction performance of this cited model is on the other hand exhibited in Fig. 5. In this model of LSTM, while the real data cloud is displayed by blue colour, the orange colour depicts the corresponding forecasted data cloud of the all data cluster. Besides, it can be seen in this figure that two data functions oscillate and overlap each other quite harmoniously in this way.

Table 5. The best results

Algorithm	*hl*	*hd*	*MFs*	*ir*	Max epoch	MAE	RMSE	R
LSTM	**5**	-	-	-	**300**	**0.2477**	**0.3315**	**0.9951**
FCM	-	5	2	-	100	0.2551	0.3425	0.9948
SC	-	3	-	0.6	100	0.2895	0.3870	0.9933
GP	-	3	2	-	30	0.8896	1.2055	0.9329

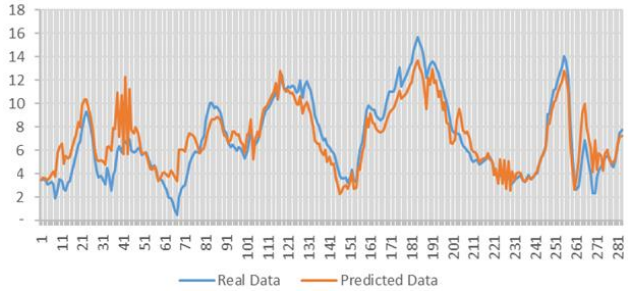

Fig. 5 Prediction performance of LSTM model

IV. DISCUSSION

Today, the estimation of atmospheric data has become so significant, in behalf of determining the unknown future value that makes our life easier in many areas or saves us from possible future troubles. Among the implemented physical data predictions, forecasting of wind blowing speed on the sea shores can be regarded as one of the most significant one. It is possible to obtain future wind blowing speed that has not occurred yet using historical time series of wind blowing data and benefiting from machine learning or artificial intelligence. On the other hand, in this approach, solving of complex mathematical or physical problems is not needed nor detailed technical knowledge or field experimentation are required. Besides, many adverse natural events can be prevented by estimating the future situation of wind blowing speed that may for instance cause possible floods or damages.

V. CONCLUSION

In this study, the artificial intelligence computations executed with the structured 66 models have demonstrated that the forecasts have been generally performed with low values of statistical accuracy errors. Besides, generally, the performance of the prediction data cloud was exhibited to be highly correlated with the real data cloud. Among whole estimations, particularly the computations actualized using LSTM tool have demonstrated that this approach can be easily and safely implemented at a high accuracy in predictions of wind blowing speeds at the sea shores.

REFERENCES

[1] A. Ilhan, "Forecasting of river water flow rate with machine learning," *Neural. Comput. Appl.*, vol. 34(22), pp. 20341–20363, Nov. 2022.

[2] (2020) Global Wind Energy Council website. [Online]. Available: https://gwec.net/

[3] M. Bilgili, H. Alphan, and A. Ilhan, "Potential visibility, growth, and technological innovation in offshore wind turbines installed in Europe," *Environ. Sci. Pollut. Res.*, Nov. 2022.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9(8), pp. 1735–1780, Nov. 1997.

[5] H. Z. Abyaneh, A. M. Nia, M. B. Varkeshi, S. Marofi, and O. Kisi, "Performance evaluation of ANN and ANFIS models for estimating garlic crop evapotranspiration," *J. Irrig. Drain Eng.*, vol. 137, pp. 280–286, May 2011.

[6] (2020) Mathworks Long Short-term Memory Networks website. [Online]. Available: https://www.mathworks.com

[7] K. M. Chandy and S. Taylor, *An Introduction to Parallel Programming*, Boston, USA: MA: Jones and Bartlett, 1992.

[8] M. Bilgili and B. Sahin, "Comparative analysis of regression and artificial neural network models for wind speed prediction," *Meteorol. Atmos. Phys.*, vol. 109, pp. 61–72, Oct. 2010.