Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 6 (2024) 2079

Journal of the Nigerian Society of Physical Sciences

Wind speed prediction in some major cities in Africa using Linear Regression and Random Forest algorithms

Timothy Kayode Samson [a], Francis Olatunbosun Aweda [b,*]

[a] Statistics Programme, College of Agriculture, Engineering and Science, Bowen University, Iwo, zip code: 540004, Nigeria
[b] Physics Programme, College of Agriculture, Engineering and Science, Bowen University, Iwo, zip code: 540004, Nigeria

Abstract

Globally, wind energy if properly harnessed, could serve as a source of energy generation in Africa. This study compared the performance of two Machine Learning (ML) algorithms (Linear regression and Random Forest) in predicting wind speed in five major cities in Africa (Yaoundé, Pretoria, Nairobi, Cairo and Abuja). Wind data were collected between January 1, 2000, and December 31, 2022, using the Solar Radiation Data Archive. The data preprocessing was carried out with 80% of the data used for training and 20% for validation. The performance of these ML algorithms was evaluated using Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R2). The result shows that Nairobi (3.814795 m/s) closely followed by Cairo (3.606453 m/s) has the highest mean wind speed while Yaoundé (1.090512 m/s) has the lowest. Based on the performance metrics used, the two Machine Learning algorithms were competitive. Still, the Linear Regression (LR) algorithm outperformed the Random Forest Algorithm in predicting wind speed in all the selected major African cities. In Yaoundé (RMSE = 0.3892, MAE= 0.3001, MAPE =0.5030), Pretoria (RMSE=1.2339, MAE=0.9480, MAPE=0.7450) Nairobi (RMSE= 0.4223, MAE =0.6499, MAPE =0.1872), Nairobi (RMSE=0.6499, MAE=0.5171, MAPE =0.1872), Cairo (RMSE =1.0909, MAE =0.8544, MAPE =0.3541) and Abuja (RMSE = 0.70245, MAE =0.5441, MAPE= 0.4515) the Linear regression algorithms was found to outperformed Random Forest Regression. Therefore, the Linear regression algorithm is more reliable in predicting wind speed compared with the Random Forest regression.

DOI:10.46481/jnsps.2024.2079

Keywords: Energy generation, Atmospheric parameters, Statistical models, Machine Learning Algorithms, African stations

Article History :
Received: 17 April 2024
Received in revised form: 13 June 2024
Accepted for publication: 14 August 2024
Published: 08 September 2024

Communicated by: Oladiran Johnson Abimbola

1. Introduction

Energy generation and usage have been a serious challenge throughout the entire world; thus, wind energy is one of the most important to any government for the comfort of the citizenry, which is known as unconventional energy. Emeis [1] posited that atmospheric parameters play an important role in electricity generation, with wind energy playing a major role in such generation globally. According to Routray *et al.* [2], wind speed prediction has been known to be integrated into grid energy for distribution, which can lead to citizen comfort. Routray *et al.* [2] discovered that wind speed variation influences the tasks involved in energy generation operations. However, in the management of wind speed energy, different efficiency levels influence the dynamic of some models used in the wind en-

*Corresponding Author Tel. No.: +234-803-256-6053.
Email address: aweda.francis@bowen.edu.ng (Francis Olatunbosun Aweda)

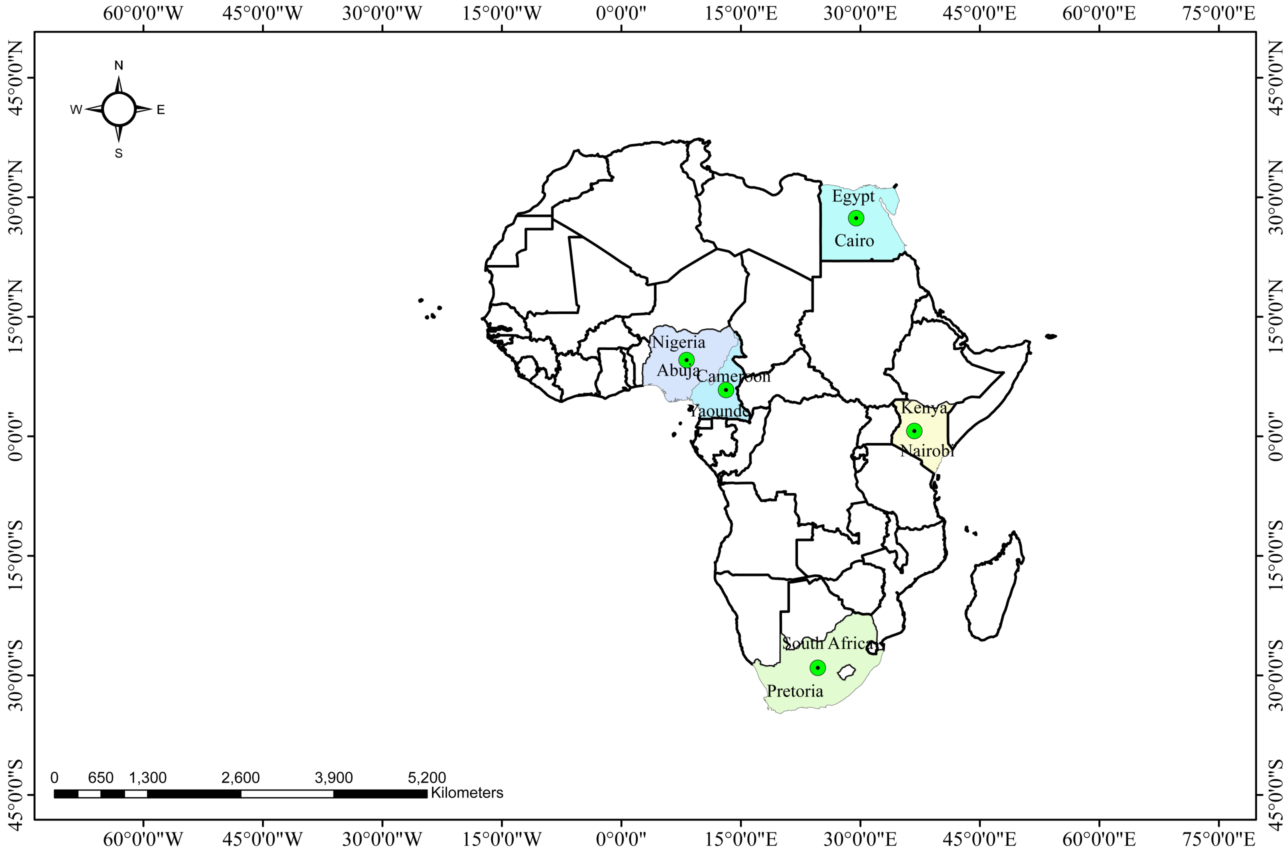

Figure 1. Map of African Countries showing the studied locations.

ergy prediction process in wind farms. However, the report has shown that in the long-term prediction of wind speed energy, short term varies in the interval of the wind speed energy, which leads to the prediction and forecast application, which sometimes are very complex in the application for future occurrence, but with the aid of machine learning, wind speed forecast will help in the future prediction [3], in other words by using optimize method [4]. Wind power optimization forecasting using CFD or artificial learning is based on various algorithms Saini *et al.* [5], more specifically, the use of deep learning algorithms as reported by Ibrahim *et al.* [6], which was also supported by Liu and Liang [7], that wind model forecast contribute to the development of any society. According to Gross *et al.* [8], methodologies in the prediction of wind speed application are based on the pattern of the wind speed, which contributed to the generation of electricity distribution for the comfort of the citizens, and there are some energy losses in the transmission of energy to different based stations across any country Mahmoud *et al.* [9].

One of the supervised Machine learning algorithms that can be used in wind speed prediction is the Linear regression algorithm. Linear regression algorithm is a supervised machine learning for predicting the target variable using the features or input variables. The goal of this algorithm is to establish the linear relationship between the target variable and the input variables or features. Wind energy contributes to energy generation and production Aweda *et al.* [10]. This study will further educate African countries on the importance of using wind energy as an alternative source of energy in Africa. This will be one of the steps in addressing the energy challenge in Africa. As reported by various studies Oyewole *et al.* [11]; Aweda *et al.* [10]; Aweda and Samson [12]; Okada *et al.* [13], wind speed has increased in some African countries, due to continental wind, tree plantations as well as the region placement on the earth, and if properly harnessed, it will be beneficial to regional development. Also, the use of Machine Learning is an emerging area of research which could also be explored in African countries. According to Nie *et al.* [14], wind power energy plays an important role in the generation of electricity type by improving the efficiency optimization of the wind equipment's power method. It has been reported that the environmental parameters depend on the wind speed, making the system application easily predictable [15–17]. According to research, renewable energy has contributed significantly to the development of any country based on wind speed ([18–21]). In recent years, Adams *et al.* [22] reported that green energy contributes to the development of wind energy. Authors such as (Routray *et al.* [2]; Oyewole *et al.* [11]; Aweda *et al.* [10]; Chen *et al.* [23]; Anjana [24]; Aweda and Samson [25]) have demonstrated that wind energy contributes to the development of any society. Furthermore, as a result of clean energy, revolutionary, and technological development, wind energy has generated human capacity growth in the development of any modern city.

To this end, wind energy and power generation have been reported to be preserved to overcome grid reliability issues (Chen *et al.* [23]; Anjana [24]; Jamil and Zeeshan [26]). As a result of the intensity and direction of development, this provides the amount of wind energy required for the scheduling and planning of power stations for the next day, which will be useful for future generation (Routray *et al.* [2]; Wang *et al.* [27]; Malik and Yadav [28]; Demetriou *et al.* [29]). According to Routray *et al.* [2], a variety of statistical time series models based on various methodologies have been used for wind speed prediction and analysis, as well as the use of various artificial intelligence methods to forecast the occurrence of wind speed features. Wind speed forecasting is based on popular Auto-regressive moving average (ARIMA), and Seasonal autoregressive moving average (SARIMA) models, as well as Multi-layer perception neural, SVM model, Autoregressive Integrated Moving Average (AIMA) model, and neural network, with PAERNN and PMERN, Ada Boost neural network, Back propagation neural network, Artificial Neural Networks and Decision Trees, among other models. However, the SARIMA model was also used in the prediction of rainfall as reported by (Murat *et al.* [30]; Murat *et al.* [31]). Due to the popularity of Machine Learning algorithms, different industries and organizations embrace them for prediction and forecasting purposes in various applications Routray *et al.* [2]. According to Chaudhary *et al.* [32], meteorological parameters such as air temperature, relative humidity, solar radiation, and wind speed are based on the approach of machine learning. More so (Routray *et al.* [2]; Aweda *et al.* [10]), reported that atmospheric data and meteorological parameters are important factors in predicting wind speed globally; however, if these parameters are modelled using machine learning and some statistical tools, they produce a good forecasting result. According to the report, Artificial Neural Networks (ANN), Long Short Term Memory Networks (LSTM), and Attention-Based Modeling are used for the prediction and forecasting of wind speed for wider use in engineering processes and procedures Makridakis *et al.* [33].

Various authors, including (Venkatakrishnan *et al.* [17]; Taher and Karimi [34]; Reddy and Selvajyothi [35]; Patel *et al.* [36]; Bhimarasetti and Kumar [37]), have reported that the generation of electricity requires a large amount of capital. Research has shown that wind speed prediction uses machine learning algorithms, statistical models, and some neural network models to forecast the features and tell the historical data about wind speed in African stations. Rao *et al.* [38] reported that wind speed data were collected on an hourly basis to predict what would happen in a feature in India; however, for this study, wind speed data were collected daily to determine what the effect of weed speed is daily in the selected African stations. (Mishra *et al.* [39]; Aweda and Samson [40]) reported that a series of data is available for the generation of wind speed that is based on the prediction precision using a machine-learning algorithm approach, whereas, in the old method, wind speed data operate based on other meteorological data such as temperature, relative humidity, and rainfall, which have a significant effect on the prediction of the feature of wind speed. According to (Routray *et al.* [2]; and Aweda *et al.* [10]), energy is one of the factors that contribute to the development of any economy. However, many African countries rely on energy for development and growth, and this adversary contributes to human development and the efficiency of any society's work. According to the report, wind energy has been classified as renewable energy for the development of any economy, thereby increasing the growth of any nation. Therefore, this study aims to predict wind speed using a machine-learning algorithm in selected African stations. More so, the objective of the research is to determine the feature occurrence and forecast of wind speed in the selected African stations for energy generation to reduce the lack of energy experienced in African stations. This study is very significant considering the importance of wind energy as a major source of renewable energy which is one of the cleanest and most sustainable ways of generating electricity with no toxic pollution or global warming emission. Wind is also abundant in some cities in Africa which could help some energy problems confronting Africa. The use of Random Forest has gained popularity in recent years due to its ability to capture non-linear relationships, its stability and interpretable abilities. For instance, Ho *et al.* [41] explored the use of Random Forests to predict short-term wind speed in coastal Taiwan. Similarly, other studies have also employed and established the forecasting accuracy of Random Forest in wind speed prediction (Santos *et al.* [42]; Zheng *et al.* [43]; Fitipaldi *et al.* [44]). Also, Aweda *et al.* [10] employed the use of machine learning algorithms including linear regression algorithms to investigate wind speed as a source of energy for some selected stations in Nigeria. The use of the Linear regression algorithm was also applied by Karasu *et al.* [45] to predict wind speed in Turkey. To the best of the researchers' knowledge, most of the studies that employed the use of these algorithms were carried out outside Africa with the exemption of that of Aweda *et al.* [10] that focused on some selected in Nigeria stations. The need to extend this to other countries in Africa is one of the major motivations for this study.

Furthermore, the demand for dependable methods to forecast wind speeds accurately, which is essential for maximizing the efficiency of renewable energy production most especially in Africa with an abundance of wind also necessitated this study. Energy production has been a major challenge in Africa and this has adversely affected socio-economic development in Africa. As a result, this study applied Linear Regression and Random Forest algorithms for predicting wind speed in selected major cities in Africa. This is believed will enhance energy generation in Africa through a reliable forecast of wind speed in these stations.

2. Research methodology

2.1. The study area and their locations

The African sub-region stations chosen for this study were divided into several divisions, and the coordinates of each sta-

Table 1. African Station division according to their regions and period of data collection.

Station	Division	Country	Longitude	Latitude	Period of Data
Abuja	Hinterland Region	Nigeria	07.399° W	09.077° N	1980 – 2022
Cairo	Nile Valley Region	Egypt	30.0444° N	31.2357° E	1980 – 2022
Nairobi	Highland Region	Kenya	1.2921° S	36.8219° E	1980 – 2022
Pretoria	Coastal Region	South Africa	25.7479° S	28.2293° N	1980 – 2022
Yaoundé	Hinterland Region	Cameroon	3.8480° N	11.5021° E	1980 – 2022

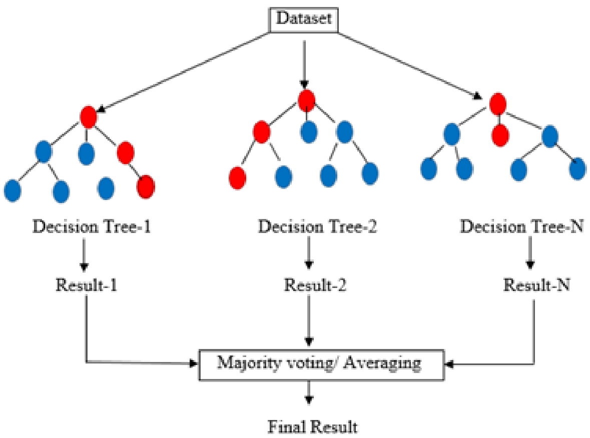

Figure 2. Pictorial illustration of Random Forest regression.

tion are listed and shown in Table 1 and Figure 1. These stations are classified as North, South, East, West, and Central Africa. Based on its region and location, this station was chosen to determine the rainfall pattern across African stations. The following locations were chosen for this study: Abuja (07.399° W, 09.077° N), Cairo (30.0444° N, 31.2357° E), Nairobi (1.2921° S, 36.8219° E), Pretoria (25.7479° S, 28.2293° N), Yaoundé (3.8480° N, 11.5021° E). However, the study spanned from 1980 to 2022 across all stations considered.

2.2. *The process of data collection*

Using the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) technique, monthly wind speed data for selected African stations were obtained from the HelioClim-1 (www.soda-pro.com) archives. HelioClim-1 data for the study years was accessed on May 1, 2023. This is a credible data source for wind speed and has been used by other researchers (Bright *et al.* [46]; Mamani & Hendrick [47]). The data was then collected and processed for a few minutes before being downloaded. The downloaded data was filtered using Excel packaged to separate the combined data. The data date was chosen to coincide with two solar circles to provide activities that occurred during the study period. The following are the coordinates for the stations used as shown in Table 1. However, the data for the stations used were obtained from the Solar Radiation Data Archive from 1980 to 2022, as reported by Gelaro *et al.* [48].

2.3. *Machine learning algorithms used*

The features used in this study are time series-based features using the past four months' data as the features while the target variable is the present month's wind speed. This means that in building these Machine Learning algorithms, the last four months' wind speeds were used to predict the present month's wind speed. The use of four previous lag values of wind speed is believed to help provide a more reliable and realistic forecast than using other variables such as pressure gradient, wind movement, ocean current, and air temperature among others which cannot be obtained in advance. Two supervised Machine Learning algorithms: Linear Regression and Random Forest regression. Random Forest (RF) is one of the supervised learning algorithms. Random Forest makes use of resembling methods, a method that combines the result of prediction from multiple decision trees on various sub-samples of the dataset and uses the majority vote to improve the predictive accuracy and control overfitting. Linear Regression also belongs to the class of supervised learning where the target feature, as well as the regressors, are also numerical. Linear regression algorithm predicts the target features based on the numeric independent features in this case the past four months' wind speed. The choice of Random Forest was premise on the fact that it has the potential to account for non-linear relationships between features and the target variable (Ho *et al.* [41]), unlike the linear regression algorithm that assumed a linear relationship between the features and the target variable (Aweda *et al.* [49]). The study considered these two Machine Learning algorithms because the study wants to investigate whether wind speed in the selected major cities in Africa can better be predicted by linear regression algorithm which expresses the present month's wind speed as a linear combination of the previous four-month values or by an algorithm that accounts for the non-linear relationship between the present month's wind speed values and the previous four months' values. This is one of the major motivations for this study. The loss function used in evaluating the performance of these ML algorithms Mean Square Error(MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2) were used. These performance metrics have also been used to evaluate the performance of Machine Learning algorithms in similar studies (Drisya *et al.* [50]; Ho *et al.* [41]; Fitipaldi *et al.* [44]; Routray *et al.* [2]). Each of these performance metrics is defined below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(W_i - \hat{W}_i\right)^2. \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(W_i - \hat{W}_i\right)^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|W_i - \hat{W}\right| \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{W_i - \hat{W}}{W_i}\right| \tag{4}$$

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{W}_i - \bar{W}_i\right)^2}{\sum_{i=1}^{n}\left(W_i - \bar{W}_i\right)^2}, \tag{5}$$

where, W_i is the actual wind speed, $\hat{W}_i$ is the predicted while $\bar{W}$ is the mean wind speed.

Figure 1 shows the pictorial representation of the Random Forest regression as provided by Khan *et al.* [51]. To assess the performance of these Machine Learning models on the unseen data, the hold-out method was used. The data were divided into training and testing sets with 80% of the data as training and 20% as test data.

2.4. Data preprocessing

Data preprocessing was carried out to identify any case of missing values and outliers. Appropriate Python libraries was used for this purpose. For instance, to check for missing values, .isnull().sum () was used and the result returned zero indicating that there is no missing value. The summary statistics were used to check for outliers and the result indicated that there is no evidence of outliers while data transformation was done using normalization. This was done using the Standard Scaler () after importing Standard Scaler from sklearn. preprocessing which is present in the scikit-learn library. This standardization transforms the data to have a mean of zero and a variance of 1.

3. Results and discussion

3.1. Descriptive statistics for wind speed in the selected cities

The result of descriptive statistics shows the mean wind speed in Nairobi (3.8148 m/s) was higher than that of other African cities. The maximum wind speed of 2.6900 m/s was obtained in Yaoundé while for Pretoria, Nairobi, Cairo and Abuja the maximum wind speed of 8.7600 m/s, 7.2700 m/s, 9.7300 m/s and 7.4500 m/s were obtained respectively. This implies that Cairo had the highest maximum wind speed while the minimum wind speed obtained in Yaoundé was lower than that obtained in other cities. The standard deviation for wind speed was higher in Pretoria (1.2963 m/s) was higher than that of other cities. This indicates more variation in wind speed in Pretoria than in other locations while in Yaoundé, wind speed varies less compared with other African cities considered. The histogram plots in Figure 3 reveal that wind speeds in Yaoundé, Pretoria, Cairo and Abuja were all positively skewed (skewed to the right) while that of Nairobi was skewed to the left. The plots of the variation of wind speed for selected African stations are presented in Figure 4 this shows no noticeable trend in wind speed.

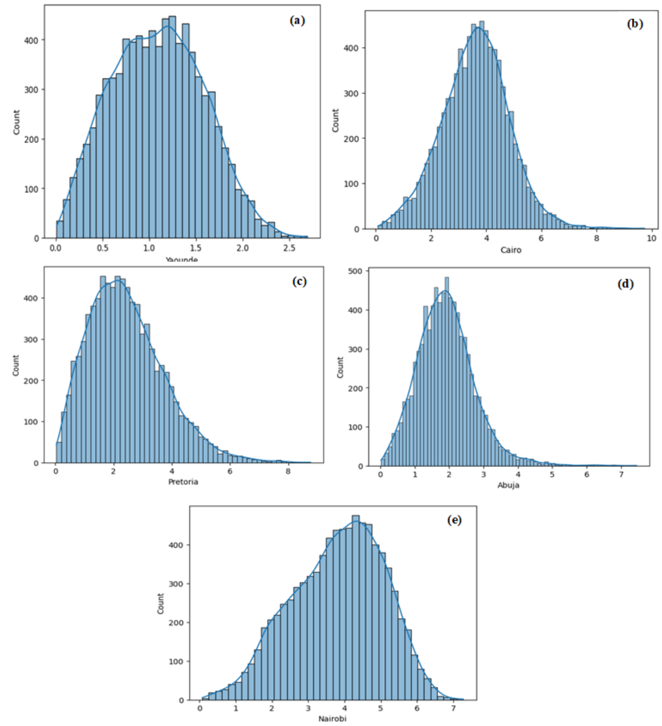

Figure 3. Histogram plot for selected African, (a) Yaounde (b) Cairo (c) Pretoria (d) Abuja (e) Nairobi.

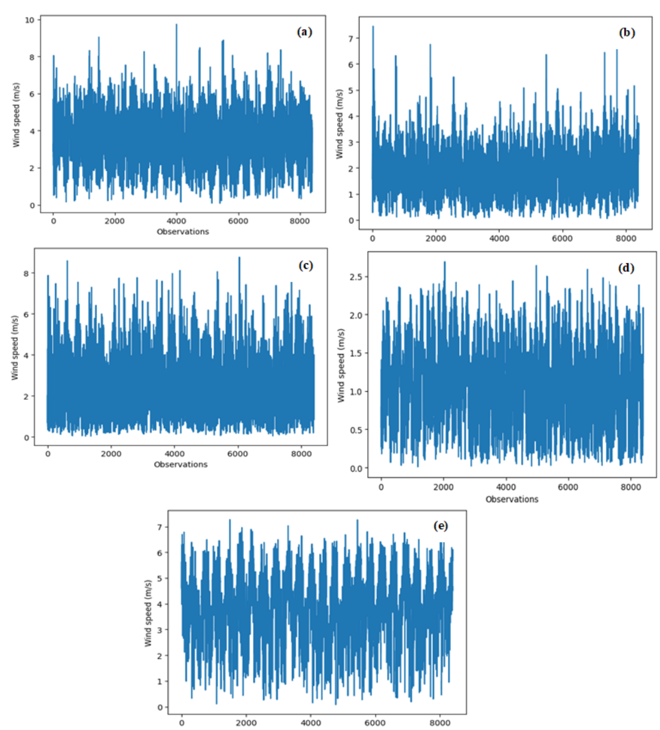

Figure 4. Variation of wind speed for selected African Station.

Table 2. Descriptive statistics for wind speed in the five selected cities in Africa.

Variables	African cities				
	Yaoundé	Pretoria	Nairobi	Cairo	Abuja
No of observations	8401	8401	8401	8401	8401
Mean	1.0905	2.4336	3.8148	3.6065	1.9086
Standard deviation	0.4885	1.2963	1.2541	1.2126	0.8268
Minimum	0.0100	0.0400	0.0800	0.0800	0.0200
25%	0.7200	1.4700	2.9200	2.8200	1.3400
50%	1.0900	2.2800	3.9300	3.6300	1.8600
75%	1.4500	3.2000	4.7500	4.3800	2.3900
Maximum	2.6900	8.7600	7.2700	9.7300	7.4500

3.2. *Monthly trend in wind speed in the selected major cities in Africa*

The average monthly wind speed in each of the selected cities was plotted and the Figures obtained are as shown in Figure 5. These Figures indicated that in Nairobi, Pretoria, and Abuja, wind speed decreased from January to February while in Yaounde and Cairo, the reverse was the case as there was a downward trend in wind speed between these two months. Wind speed was also found to be at its lowest value in December in both Abuja and Yaoundé while in Cairo and Nairobi, the lowest wind speed in the year occurred in September and June respectively. There was variation in the months of the year with the peak wind speed in the selected cities in Africa.

Furthermore, the results show that the wind speed pattern at each station is nearly identical, indicating that the wind speed at the chosen station moves at a similar rate, which can increase wind energy production across the entire nation. However, when wind energy output is increasing, short-term wind speed predictions are critical. Furthermore, the maximum wind speed for the stations ranged from 9.0 to 6.0 m/s, while the minimum wind speed ranged from 0.1 to 4.0. The minimum wind speed observed in Pretoria was in March (2.22 m/s), while the maximum wind speed observed was in October (2.49 m/s); however, the results revealed that the maximum wind speed observed in Pretoria could not be enough to power a turbine during such a period due to the low speed. In Yaoundé wind speed was lowest in December (1.07 m/s) and highest in April (1.12 m/s), indicating that the wind speed recorded in Yaoundé is low in comparison to other stations. The wind speed for Nairobi, as shown has a minimum value in April (3.75 m/s) and a maximum value in November (3.85 m/s). This demonstrates that Nairobi has a high wind speed when compared to other stations. The wind speed result for Cairo shows that the lowest value was recorded in September (3.525 m/s) and the highest value was recorded in May (3.701 m/s). The results show that Cairo has the highest wind speed when compared to other stations, which could be due to the proximity to the ocean and the resulting ocean breeze. The results of Abuja wind speed show that the minimum value of wind speed was recorded in December (1.82 m/s) and the maximum value of wind speed was recorded in February (1.99 m/s), indicating that Abuja wind speed was lower than that experienced at other stations. However, it demonstrates that Abuja's wind speed was insufficient to power any wind turbine.

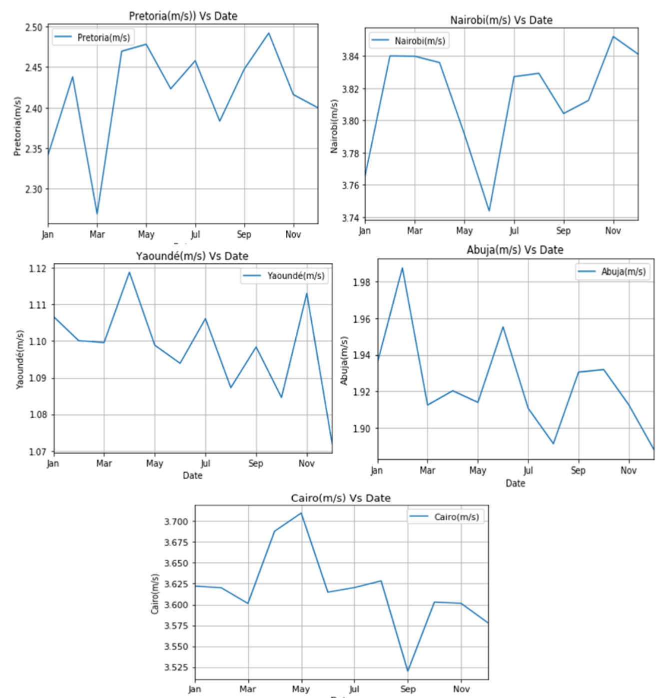

Figure 5. Monthly variation of wind speed trend in selected African stations.

3.3. *Comparative performance of linear regression and random forest algorithm*

Results in Table 3 show that the Linear regression algorithm outperformed Random Forest regression in all cities. Result in Table 3 presents the coefficient of determination obtained using the linear regression algorithm was higher than that of Random Forest (RF) regression in Yaoundé (R^2 = 0.4183 versus 0.36870), Pretoria (R^2 = 0.1195 versus 0.0674), Nairobi (R^2 = 0.7093 versus 0.6780), Cairo (R^2 = 0.1900 versus 0.1511) and Abuja (R^2 = 0.3217 versus 0.2673). This implies that the Linear regression algorithm provided a better fit to the data than the Random Forest regression. The result of the forecasting performance of these algorithms shows that the Linear regression algorithm gave the least values of Mean Square Error (MSE), Root Mean Square Error (RMSE),

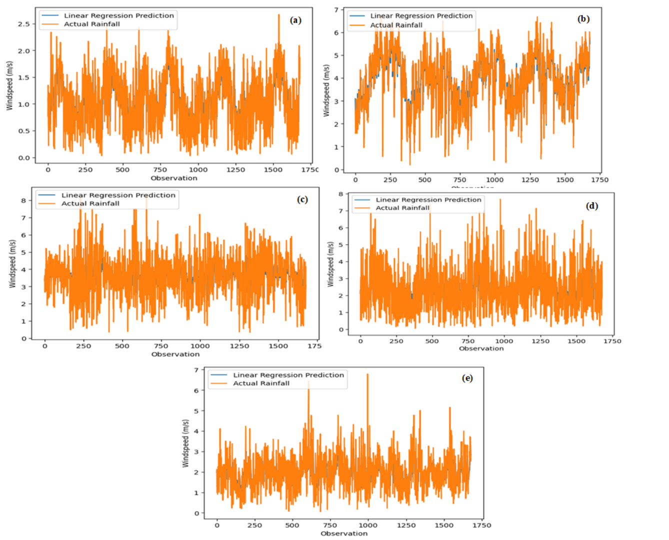

Figure 6. Graph of the actual and predicted wind speed for African stations based on Linear regression algorithm ((a) Yaoundé (b) Nairobi (c) Cairo (d) Pretoria (e) Abuja).

Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) in Yaoundé (MSE = 0.1515, RMSE =0.3892, MAE=0.3001), Pretoria (MSE= 1.5226, RMSE =1.2339, MAE = 0.9480, MAPE=0.7450), Nairobi (MSE =0.4223, RMSE=0.6499, MAE =0.5171, MAPE=0.1872), Cairo (MSE =1.1890, RMSE =1.0909, MAE =0.8544, MAPE = 0.3541) and Abuja (MSE=0.49343, RMSE =0.70245, MAE = 0.5441, MAPE = 0.4515). These results therefore established the superiority of the Linear regression algorithm over the Random Forest regression in predicting wind speed in these selected major cities in Africa. In terms of the predictive performance based on RMSE and MAE, the highest RMSE for the Linear regression algorithm was obtained in Pretoria (RMSE = 1.5226, MAE=1.2339) while the lowest RMSE and MAE was obtained in Yaoundé (RMSE = 0.1515, MAE=0.3001). This implies that Linear regression algorithm performed better in Yaoundé than other stations. Furthermore, the predictive performance of the Random Forest (RF) was better in Yaoundé (RMSE = 0.401090, MAE = 0.3119) and worst performance in Cairo (RMSE =1.266653, MAE = 0.8837) compared to other selected cities. Detailed results are presented in Table 3. The plots of the actual and predicted wind speed in these selected major cities in Africa based on the Linear regression algorithms are presented in Figure 6. This was done to visualize the performance of this algorithm.

Figure 6(a-e) show the actual and predicted wind speed based on the Linear regression algorithm. From the Figure, it can be deduced that there is a substantial agreement between actual and predicted wind speed using the Linear regression algorithm. The two values were almost indistinguishable as shown in Figure 6(a-e) indicating that the Linear regression algorithm has proven to be able to adequately capture the dynamic nature of wind speed in the selected major cities in Africa. The result

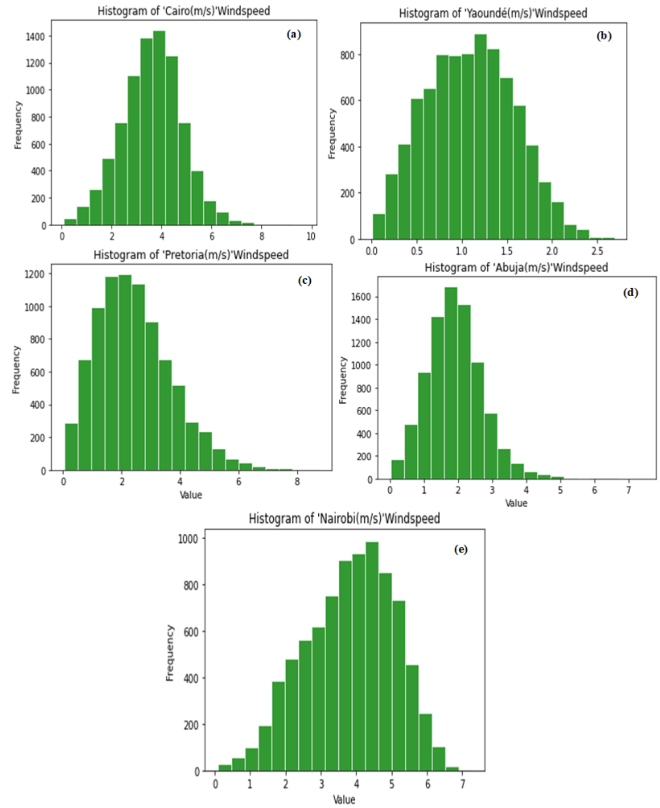

Figure 7. Histogram Representation of Wind Speed Variation Across the Stations ((a) Yaoundé) (b) Nairobi (c) Cairo (d) Pretoria (e) Abuja

of this study favoured the linear regression algorithm over Random Forest. One possible explanation for this could be that the linear regression is generally more interpretable than Random Forest. Random Forest fails to generalize well if the problem space is not stable. Though, the linear regression also performs poorly when the number of variables is too many this is not the case as only four features were considered and this same number of features was used in training the two Machine Learning models. Aside from the predictability advantage of the linear regression algorithm, the model also demonstrated better predictive accuracy in predicting wind speed in the selected cities in Africa over Random Forest as revealed by the performance metrics (MSE, RMSE, MAE, MAPE and R^2).

Findings showed wind speed in most of the selected cities were positively skewed which is consistent with that obtained in similar studies (Meenal *et al.* [52]; Samuel *et al.*[53]; Verma *et al.* [53]). This study also established the superiority of the linear regression algorithm over random forest regression which is corroborated by Verma *et al.* [54] which found that the linear regression performed better than random forest regression in wind speed forecasting in Brussel, Belgium. This finding is in line with that of studies by Meenal *et al.* [51] where the random forest regression was found to out-form linear regression and Support Vector Machine (SVM). This discrepancy may be due to the differences in the study area as the former study was carried out in India while the present study was carried out in

Table 3. Comparative result of Linear regression and Random Forest Regression.

Cities	Linear regression					Random Forest Regression				
	MSE	RMSE	MAE	R^2	MAPE	MSE	RMSE	MAE	R^2	MAPE
Yaoundé	0.1515	0.3892	0.3001	0.4183	0.5030	0.160873	0.401090	0.3119	0.3687	0.5258
Pretoria	1.5226	1.2339	0.9480	0.1195	0.7450	1.577885	1.256139	0.9769	0.0674	0.7628
Nairobi	0.4223	0.6499	0.5171	0.7093	0.1872	0.464514	0.681553	0.5466	0.6780	0.1995
Cairo	1.1890	1.0909	0.8544	0.1900	0.3541	1.125457	1.266653	0.8837	0.1511	0.3660
Abuja	0.49343	0.70245	0.5441	0.3217	0.4515	0.5330	0.73007	0.5654	0.2673	0.4645

MSE- Mean Square Error, RMSE- Root Mean Square Error, MAE- Mean Absolute Error, MAPE- Mean Absolute Percentage Error

Africa with its unique wind speed pattern. This finding that favoured linear regression was not corroborated by that of Ho *et al.* [41], Drisya *et al.* [50] and Samuel *et al.* [53] which favoured Random Forest. This discrepancy could be a result that these studies considered only random forests as the linear regression algorithm was not considered in their studies.

Figure 7(a-e) depicts the frequency distribution of wind speed for all stations considered. It was discovered that the highest frequency observed in all stations is approximately >1700. This demonstrates that not all of the stations are located in the same climate zone. The minimum value was recorded across all stations (Frequency < 200). This demonstrates that the stations are located in a variety of climate zones, with two to three of them located in the continent's coaster region due to their proximity to the ocean, which may influence the station's increased wind speed. Furthermore, the frequency distribution of wind speed in all considered stations revealed that Abuja had the highest frequency (Frequency>160). Pretoria shows a lower frequency distribution (Frequency < 200).

4. Conclusion

The use of wind speed prediction using machine learning over selected stations in Africa for this study represents the importance of wind speed around the world. This study demonstrated the use of linear regression algorithms and Random Forest (RF) in predicting wind speed in five major cities in Africa. Results indicated that these cities exhibited different patterns in wind speed as revealed by the month of the year with peak and lowest values of wind speed. The superiority of the linear regression over Random Forest has been demonstrated in this study with these Machine learning algorithms giving a reliable prediction of wind speed as indicated by the values of the performance metrics used. Nevertheless, there is a need for further studies on wind speed prediction using machine learning algorithms other than those considered in this study. Also, further studies are required in the area of the use of deep learning algorithms such as CNN, and LSTM among others in predicting wind speed in African countries. This study can also be extended to other cities in Africa aside from these five major cities considered in this study. Consequent to this study, it is very crucial that more cutting-edge research needs to be done in the area of renewable energy. This will help in policy formulation on how to fully harness the potential of wind power generation given the fact that inadequate power generation is a major problem in Africa. Therefore, as reported by Samson and Aweda [55], it would be recommended that the African government should make it a point of duty to invest in research for the development of the continent. Collaboration among governments, private sectors, and pertinent international organisations is essential to pool resources and expertise for the widespread implementation of wind energy projects in Africa. African nations ought to provide incentives, such as tax incentives, to stimulate private investment in wind energy initiatives.

Acknowledgement

Based on the data used in this study and the review made, the authors would like to express their gratitude to the MERRA-2 satellite for providing data and the anonymous reviewers for their contribution towards the good construction and modification made for the success of the manuscript.

References

[1] S. Emeis, "Current issues in wind energy meteorology", Meteorological Application **21** (2014) 803. https://doi.org/10.1002/met.1472.

[2] A. Routray, K. D. Mistry, S. R. Arya & B. Chittibabu, "Applied machine learning in wind speed prediction and loss minimization in unbalanced radial distribution system", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects **43** (2021) 1. https://doi.org/10.1080/15567036.2020.1859010.

[3] H. Acikgoz, C. Yildiz & M. Sekkeli, "An extreme learning machine based very short-term wind power forecasting method for complex terrain", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects **42** (2020) 2715. https://doi.org/10.1080/15567036.2020.1755390.

[4] Z. Tian, Y. Ren & G. Wang, "Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects **41** (2019) 26. https://doi.org/10.1080/15567036.2018.1495782.

[5] V. K. Saini, R. Kumar, A. Mathur & A. Saxena, "Short term forecasting based on hourly wind speed data using deep learning algorithms", 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, 2020, pp. 1-6. https://doi.org/10.1109/ICETCE48199.2020.9091757.

[6] M. Ibrahim, A. Alsheikh, Q. Al-Hindawi, S. Al-Dahidi & H. ElMoaqet, "Short-time wind speed forecast using artificial learning-based algorithms", Computational intelligence and neuroscience **2020** (2020) 1. https://doi.org/10.1155/2020/8439719.

[7] L. Liu, & Y. Liang, "Wind power forecast optimization by integration of CFD and Kalman filtering", Part A: Recovery, Utilization, and Environmental Effects, Energy Sources **43** (2021) 1880. https://doi.org/10.1080/15567036.2019.1668080.

[8] P. Gross, A. Boulanger, M. Arias, D. L. Waltz, P. M. Long, C. Lawson, R. N. Anderson, M, Koenig, M. Mastrocinque, W. Fairechio, J. A. Johnson, S. Lee, F. Doherty & A. Kressner, "Predicting electricity distribution feeder failures using machine learning susceptibility analysis", The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, Boston, Massachusetts, USA, 2006, pp. 1705-1711. https://api.semanticscholar.org/CorpusID:6175166.

[9] K. Mahmoud, M. Abdelnasser, H. Kashef, D. Puig & M. Lehtonen, "Machine learning based method for estimating energy losses in large-scale unbalanced distribution systems with photovoltaics", Artificial Intelligence and Sensor Informatics: Exploring Smart City and Construction Business Implications **6** (2020) 157. https://doi.org/10.9781/ijimai.2020.08.002.

[10] F. O. Aweda, S. Adebayo, A. A. Adeniji, T. K. Samson & J. A. Akinpelu, "Investigation of wind speed to generate energy using machine learning algorithms approach over selected Nigerian stations", Journal of Renewable Energy and Environment **10** (2023) 81. https://doi.org/10.30501/jree.2022.354698.1422.

[11] J. A. Oyewole, F. O. Aweda & D. Oni, "Comparison of three numerical methods for estimating weibull parameters using weibull distribution model in Nigeria", Nigerian Journal of Basic and Applied Sciences **27** (2019) 8. https://doi.org/10.4314/njbas.v27i2.2.

[12] F. O. Aweda & T. K. Samson, "Modelling the Earth's Solar Irradiance Across Some Selected Stations in Sub-Sahara Region of Africa", Iranian (Iranica) Journal of Energy & Environment **11** 2020 204. https://doi.org/10.5829/ijee.2020.11.03.05.

[13] M. Okada, T. Ichizawa, Y. Nakamura, K. Yamaguchi, R. Kodama, H. Kato, Y. Nagano, R. Ikeda, V. Q. Doan, H. Kusaka & N Ogasawara "Development of a wind power ramp forecast system by astatistical and meteorological approach", in Grand Renewable Energy proceedings Japan council for Renewable Energy, Yokohama, Japan, 2018, pp. 130–133. https://web.archive.org/web/20220506190938id_/https://www.jstage.jst.go.jp/article/gre/1/0/1_130/_pdf

[14] Y. Nie, H. Bo, W. Zhang & H. Zhang, "Research on hybrid wind speed prediction system based on artificial intelligence and double prediction scheme", Complexity **2020** (2020) 1. https://doi.org/10.1155/2020/9601763.

[15] J. M. Álvarez-Alvarado, J. G. Ríos-Moreno, E. J. Ventura-Ramos, G. Ronquillo-Lomeli & M. Trejo-Perea,"An alternative methodology to evaluate sites using climatology criteria for hosting wind, solar, and hybrid plants", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects **41** (2020) 1. https://doi.org/10.1080/15567036.2020.1772911.

[16] S. Louassa, O. Guerri, A. Kaabeche & N. Yassaa, "Effects of local ambient air temperatures on wind park performance: case of the Kaberten wind park", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects **45** (2023) 6082. https://doi.org/10.1080/15567036.2019.1673509.

[17] G. Venkatakrishnan, R. Rengaraj, K. Sathish, R. Dinesh & T. Nishanth, "Implementation of Modified Differential Evolution Algorithm for Hybrid Renewable Energy System", Journal of the Nigerian Society of Physical Sciences **3** (2021) 209. https://doi.org/10.46481/jnsps.2021.240.

[18] Y. Zhang, G. Pan, B. Chen, J. Han, Y. Zhao & C. Zhang, "Short-term wind speed prediction model based on GA-ANN improved by VMD", Renewable Energy **156** (2020) 1373. https://doi.org/10.1016/j.renene.2019.12.047.

[19] Y. Zhang, S. Gao, J. Han & M. Ban, "Wind speed prediction research considering wind speed ramp and residual distribution", IEEE Access **7** 2019 131873. https//doi.org/10.1109/ACCESS.2019.2940897.

[20] Y. Jiang, G. Huang, Q. Yang, Z. Yan & C. Zhang, "A novel probabilistic wind speed prediction approach using real time refined variational model decomposition and conditional kernel density estimation", Energy Conversion and Management **185** (2019) 758. https://doi.org/10.1016/j.enconman.2019.02.028.

[21] J. Wang & Y. Li, "An innovative hybrid approach for multi-step ahead wind speed prediction", Applied Soft Computing **78** (2019) 296. https://doi.org/10.1016/j.asoc.2019.02.034.

[22] S. O. Adams, D. A. Obaromi & A. A. Irinews, "Goodness of Fit Test of an Autocorrelated Time Series Cubic Smoothing Spline Model", Journal of the Nigerian Society of Physical Sciences **3** (2021) 191. https://doi.org/10.46481/jnsps.2021.265.

[23] Y. Chen, Z. Dong, Y. Wang, J. Su, Z. Han, D. Zhou, K. Zhang, Y Zhao & Y. Bao, "Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history", Energy Conversion and Management **227** (2021) 113559. https://doi.org/10.1016/j.enconman.2020.113559.

[24] R. Anjana, "Fuzzy and PI Based Speed Control of BLDC Motor using Bidirectional Converter for Electric Vehicle Application", Trends in Electrical Engineering **8** (2019) 35. https://doi.org/10.1109/GET.2016.7916647.

[25] F. O. Aweda, & T. K. Samson, "Comparative analysis of trigonometric and polynomial models in meteorological parameter prediction for Sub-Saharan West African stations", Acadlore Trans. Geosci. **3** (2014) 1. https://doi.org/10.56578/atg030101.

[26] M. Jamil & M. Zeeshan, "A comparative analysis of ANN and chaotic approach-based wind speed prediction in India", Neural Computing and Applications **31** (2019) 6807. https://doi.org/10.1007/s00521-018-3513-2.

[27] J. Wang, N. Zhang, & H. Lu, "A novel system based on neural networks with linear combination framework for wind speed forecasting", Energy conversion and management, **181** (2019) 425. https://doi.org/10.1016/j.enconman.2018.12.020.

[28] H. Malik & A. K. Yadav, "A novel hybrid approach based on relief algorithm and fuzzy reinforcement learning approach for predicting wind speed", Sustain Energy Technol Assess **43** (2021) 1. https://doi.org/10.1016/j.seta.2020.100920.

[29] D. Demetriou, C. Michailides, G. Papanastasiou & T. Onoufriou, "Coastal zone significant wave height prediction by supervised machine learning classification algorithms", Ocean Engineering **221** (2021) 108592. https://doi.org/10.1016/j.oceaneng.2021.108592.

[30] M. Murat, I. Malinowska, M. Gos & J. Krzyszczak, "Forecasting daily meteorological time series using ARIMA and regression models", International agrophysics **32** (2018) 253. https://doi.org/10.1515/intag-2017-0007.

[31] M. Murat, I. Malinowska, H. Hoffmann & P. Baranowski, "Statistical modelling of agrometeorological time series by exponential smoothing", International Agrophysics **30** (2016) 57. https://doi.org/10.1515/intag-2015-0076.

[32] A. Chaudhary, A. Sharma, A. Kumar, K. Dikshit & N. Kumar, "Short term wind power forecasting using machine learning techniques", Journal of Statistics and Management Systems **23** 2020 145. https://doi.org/10.1080/09720510.2020.1721632.

[33] S. Makridakis, E. Spiliotis & V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward", PloS one **13** (2018) e0194889. https://doi.org/10.1371/journal.pone.0194889.

[34] S. A. Taher & M. H. Karimi, "Optimal reconfiguration and DG allocation in balanced and unbalanced distribution systems", Ain Shams Engineering Journal **5** (2014) 735. https://doi.org/10.1016/j.asej.2014.03.009.

[35] M. S. K. Reddy & K. Selvajyothi, "Optimal placement of electric vehicle charging station for unbalanced radial distribution systems", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects **41** (2020) 1. https://doi.org/10.1080/15567036.2020.1731017.

[36] C. A. Patel, K. Mistry & R. Roy, "Loss allocation in radial distribution system with multiple DG placement using TLBO", IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2015, pp. 1-5. https://doi.org/10.1109/ICECCT.2015.7225932.

[37] R. T. Bhimarasetti & A. Kumar, "Distributed generation placement in unbalanced distribution system with seasonal load variation", 2014 Eighteenth National Power Systems Conference (NPSC), Guwahati, India, 2014, pp. 1-5. https://doi.org/10.1109/NPSC.2014.7103786.

[38] R. V. Rao, V. J. Savsani & D. Vakharia, "Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems", Computer-aided design **43** (2011) 303. https://doi.org/10.1016/j.cad.2010.12.015.

[39] S. Mishra, C. Bordin, K. Taharaguchi & I. Palu, "Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature", Energy Reports. **6** (2020) 273. https://doi.org/10.1016/j.egyr.2019.11.009.

[40] F. O. Aweda & T. K. Samson, "Relationship between air, temperature and rainfall variability of selected stations in Sub-Sahara Africa", Iranian (Iranica) Journal of Energy & Environment **13** (2022) 248. https://www.ijee.net/article_149738_4cc173522dfb8db1e1fe78bae3e13aae.pdf.

[41] C. Y. Ho, K. S. Cheng & C. H. Ang, "Utilizing the random forest method for short-term wind speed forecasting in the coastal area of central Taiwan", Energies **16** (2023) 1374. https://doi.org/10.3390/en16031374.

[42] L. B. dos Santos, L. S. Monteiro, R. Santos, M. A. Felipe Santos, M. C. C. de O. Carvalho & M. B. Figueredo, "Wind speed prediction model based on DWT and Randon Forest", Concilium **23** (2023) 347. https://doi.org/10.53660/CLM-2447-23S44.

[43] L. Zheng, Z. Shaohui, Y. Yingxin, S. Yi & G. Zhiqiu, "Short-term wind power prediction model based on WRF-RF model", 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, 2023, pp. 599-604. http://dx.doi.org/10.1109/ICCCBDA56900.2023.10154834.

[44] C. A. Fitipaldi, R. P. Monteiro & D. Pinheiro "Forecasting of wind power generation cristalândia wind farm using tree-based machine learning Approaches", XVI Brazilian Conference on Computational Intelligence (CBIC 2023), Salvador, 2023, pp. 1-6. http://dx.doi.org/10.21528/CBIC2023-028.

[45] S. Karasu, A. Altan, A., Z. Saraç & F. Hacioglu, "Estimation of wind speed by using regression learners with different filtering methods", 1st International Conference on Energy Systems Engineering, KBU—Karabuk, Turkey, 2017, pp. 7–11. https://www.researchgate.net/publication/322642431_Estimation_of_Wind_Speed_by_using_Regression_Learners_with_Different_Filtering_Methods.

[46] R. Mamani & P. Hendrick, "Weather research and forecasting model and MERRA-2 data for wind energy evaluation at different altitudes in Bolivia", Wind Engineering **46** (2022) 177. https://doi.org/10.1177/0309524X211019701.

[47] J. M. Bright, X. Bai, Y. Zhang, X. Sun, B. Acord & P. Wang, "Irradpy: Python package for MERRA-2 download, extraction and usage for clear-sky irradiance modelling", Solar Energy **199** (2020) 685. https://doi.org/10.1016/j.solener.2020.02.061.

[48] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. Randles, A. Darmenov, M. Bosilovich , R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. da Silva, W. Gu, G. K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, B. Zhao, "The modern-era retrospective analysis for research and applications, version 2 (MERRA-2)", Journal of climate **30** (2017) 5419. https://doi.org/10.1175/JCLI-D-16-0758.1.

[49] F. O. Aweda, J. A. Akinpelu, T. K. Samson, M. Sanni & B. S. Olatinwo, "Modeling and forecasting selected meteorological parameters for the environmental awareness in Sub-Sahel West Africa stations", Journal of the Nigerian Society of Physical Sciences **4** (2022) 820. https://doi.org/10.46481/jnsps.2022.820.

[50] G. V. Drisya, P. Valsaraj, K. Asokanb & K. S. Kumar, "Wind speed forecast using random forest learning method", International Journal on Computer Science and Engineering (IJCSE) **9** (2017) 362. https://doi.org/10.48550/arXiv.2203.14909.

[51] M. Y. Khan, A. Qayoom, M. S. Nizami, M. S. Siddiqui, S. Wasi & S. M. K. Raazi, "Automated prediction of Good Dictionary EXamples (GDEX): A comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques", Hindawi Complexity **2021** (2021) 1. https://doi.org/10.1155/2021%2F2553199.

[52] R. Meenal, P. A. Michael, D. Pamela & E. Rajasekaran, "Weather prediction using random forest machine learning model", Indonesian Journal of Electrical Engineering and Computer Science **22** (2021) 1208. http://dx.doi.org/10.11591/ijeecs.v22.i2.pp1208-1215.

[53] G. G. Samuel, G. F. Salimath, T. Porselvi & V. Karthikeyan, "Improved prediction of wind speed using machine learning", Journal of Physics", Conference Series **1964** (2021) 052005. https://doi.org/10.1088/1742-6596/1964/5/052005.

[54] A. Verma, K. G. Upadhyay & M. M. Tripathi "Development of artificial intelligent techniques for short-term wind speed forecasting", International Journal of Engineering Trends and Technology **69** (2021) 56. https://doi.org/10.14445/22315381/IJETT-V69I7P208.

[55] T. K. Samson & F. O. Aweda, "Seasonal autoregressive integrated moving average modelling and forecasting of monthly rainfall in selected African stations", Mathematical Modelling of Engineering Problems **11** (2024) 159. http://dx.doi.org/10.18280/mmep.110117.