Premier Journal of Engineering and Applied Sciences
Publication of Nigerian Society of Engineers, Ibadan Branch
P J E A S
Vol. 4, No. 2, 2023 ISSN:2734-3170

Comparison of Decision Tree and Support Vector Machine Models for Wind Speed Forecasting

K. O. Alawode*, Y. Ayanlade and O. M. Bada

Department of Electrical and Electronic Engineering, Osun State University, Osogbo, Nigeria

***Corresponding Author:** *alawodekehinde@uniosun.edu.ng*

ARTICLE INFO

Received: January, 2023
Accepted: April, 2023
Published: April, 2023

Keywords:
Decision Trees
Fine Trees
Gaussian SVM
Support Vector Machines
Wind Speed Forecasting.

ABSTRACT

For wind power plants to operate efficiently, accurate forecasts of prevailing wind speeds at a site are necessary. In this paper, models based on decision trees and support vector machines (SVM) were developed to forecast wind speed. The Nigeria Meteorological Agency's dataset comprising monthly averages of meteorological data from 2009-2016 for Osogbo, Southwestern Nigeria was used in this work for the development of machine learning models. Two decision tree (Fine Tree and Medium Tree) and two Support Vector Machine (SVM) models (Quadratic SVM and Medium Gaussian SVM models) were developed in MATLAB and trained to predict wind speed. Their performances in predicting wind speed using the 2017-2019 testing dataset were then compared. The Fine Tree model gave the best overall performance out of all the four models, achieving root mean square error of 0.0227, 0.0592 and 0.1581 respectively for 2017, 2018 and 2019. The mean absolute errors for each of the years 2017-2019 were 0.020, 0.0450 and 0.0850 respectively while the coefficients of determination were 0.9766, 0.9306 and 0.7206 respectively.

1. INTRODUCTION

In Nigeria, the power system is plagued by insufficient generation from the various hydro and thermal power plants along with poor transmission and distribution infrastructure. To overcome these challenges, there have been calls for greater inclusion of wind energy sources into the power mix via distributed generation thereby bringing the generation centres closer to the end users of electricity. This is achievable through the use of small-scale wind turbines with capacities ranging from a few kilowatts to a few megawatts. Wind energy is one of the major sources of clean and sustainable power generation. To effectively harness its power (in terms of the available output as well as the stable operation of wind power plants) depends on the availability of accurate wind speed forecasts. Accurate wind speed forecasts are needed for operators to predict and adjust the power output of turbines to achieve a balance in supply and demand of electricity, as well as integrate wind power into the grid more effectively (Ssekulima *et al.*, 2016).

Wind energy is stochastic, therefore modeling the wind variation at a particular site is an important step in siting wind power plants (Cetinay *et al.*, 2017). The wind power density (measured in watts per meter square) of a particular location is a measure of the wind energy flow rate per unit area irrespective of wind turbine size and is dependent on the location's mean wind speed value. According to the 2023 report of the International Renewable Energy Agency (IRENA, 2023), Nigeria has moderate wind potential with average wind speeds ranging between 2.1 m/s and 8 m/s across the country. Various studies have reported that lower wind speeds are generally experienced in the southern region of Nigeria while higher wind speeds are recorded in the north (Idris *et al.*, 2020; Adedipe *et al.*, 2018). Using data from the Nigerian Meteorological

Agency, Audu *et al.* (2019) reported an annual wind speed average of 4.77 m/s with a wind power density of 86.85 W/m^2 for Makurdi in northern Nigeria. The reported average wind speed is well above the cut-in speed of 3 m/s required for the operation of most wind turbines (Abolude and Zhou, 2017). The annual wind speed average recorded for Owerri was 3.36 m/s with an annual power density average of 23.23 W/m^2 (Oyedepo and Adaramola, 2012) while Uquetan *et al.* (2015) reported values of 1.3 m/s and 3.11 W/m^2 respectively for Calabar in Southern Nigeria.

Wind speed forecasting is usually done over different time horizons, ranging from very short-term, short-term, medium-term, and long-term (Souman *et al.,* 2010). Different techniques have been employed in forecasting wind speed. These include persistence modeling (Souman *et al.,* 2010), numerical weather prediction (NWP) models (Cheng *et al.,* 2017; Stathopoulos *et al.,* 2013), and statistical models like autoregressive integrated moving average (ARIMA), Auto-Regressive Integrated Moving Average with Exogenous inputs (ARIMAX) (Camelo *et al.,* 2018). Persistence models are generally limited in their forecast time horizon to very short term predictions (Axaopoulos and Tzanes, 2022). NWP models are known to be very complex to develop and also inexact because the equations used in simulating changes in the atmosphere are imprecise (Han *et al.,* 2022). The quality of forecasts by autoregressive models is affected by the presence of noise signals in the dataset (Chodakowska *et al.,* 2021).

In this era of big data (Rawat *et al.,* 2023), artificial intelligence and machine learning techniques such as Support Vector Machines (Zendehboudi, 2018; Santamaría-Bonfil, 2016), decision trees (Tronsco *et al.*, 2015), random forests (Liu *et al.,* 2023; Lahouar and Slama, 2017), neural networks (Men *et al.*, 2016), Adaptive Neuro-fuzzy Inference Systems (Chen and Folly, 2018) and Long Short-Term Memory (LSTM) networks (Shao *et al.*, 2021; Xie *et al.*, 2021) have also been applied severally. These methods handle large volumes of data from noisy, dynamic and nonlinear systems very well (Quej *et al.,* 2017).

Various studies focusing on wind speed forecasting in Nigeria have been conducted. Maduabuchi *et al.* (2023) presented a neural network model for forecasting wind speed and other environmental parameters for Enugu, Nigeria to estimate the renewable energy potential of the city. The model achieved a 96% correlation between the test dataset and the forecasts produced. Rurumah *et al.* (2021) forecasted the wind speed in Katsina using an artificial neural network model and utilized the result in developing a wind energy map. The model achieved a root mean square error of 8.9% and predicted annual wind speeds ranging between 0.9-13.1 m/s, with the highest values obtained in Katsina North. Egbunu *et al.* (2021) predicted a variety of climatic factors including wind speed using a random forest approach. The dataset used was that of average monthly climate records for ten years and a root mean square error of 8.2% was achieved in the prediction. Danladi *et al.* (2020) presented a Weibull model to characterize the wind speed as measured over Biu town in northern Nigeria. They also developed a neuro-fuzzy hybrid model to determine the potential for wind power generation at the Nigerian Army University, Biu.

A Weibull distribution model was developed by Odo *et al.* (2012) to predict the wind energy potential of Enugu. From the study, it was concluded that the developed model had a Weibull shape factor of 2.21 and a scale factor of 4.31m/s. Olaiya and Adeyemo (2012) presented the C5 decision tree and recurrent neural network models for wind speed forecasting (among other weather parameters) for Ibadan, Nigeria. The study concluded that with enough training data deployed in developing the models, the techniques are useful for weather prediction and climate change studies. Fadare (2010) presented a three-layer backpropagation artificial neural network model to forecast the monthly mean wind speed for various cities in Nigeria. The model achieved a Mean Absolute Percentage Error (MAPE) of 8.9%.

In this present study, support vector machines (SVM) and decision tree (DT) models are developed to forecast the wind speed. SVMs and DT models are generally interpretable which perform well when the available dataset is not too large, which is the case in this work. The study explores the wind energy potentials of Osogbo, Osun State, Nigeria and produce models that will be useful for proper planning towards exploiting them.

2. METHODOLOGY

This study was carried out in Osogbo, Osun State in Southwestern Nigeria. Osogbo is located between Latitude 7°48" N and Longitude 4°35" E with a total land area of 47 km^2. The map of the study area is shown in Figure 1. The steps involved in carrying out this work are summarized in Figure 2 and discussed in the following sub-sections.

Data Collection

A historical dataset comprising a monthly average of weather parameters like wind speed (m/s) at 10 m height, rainfall (mm), minimum temperature (°C), maximum temperature (°C), relative humidity, wind direction, solar radiation and sunshine hours for Osogbo was collected from the Nigeria Meteorological Agency (NIMET) Nigeria. The dataset which covered the period from 2009-2019 was loaded into MATLAB for preprocessing.

Data Preprocessing

Dataset preprocessing carried out involved replacing missing values with the mean of the dataset for the concerned weather variable, removing duplicated data, and expunging bad data. The dataset for sunshine hours contained many missing and duplicated values and was therefore excluded at this stage.

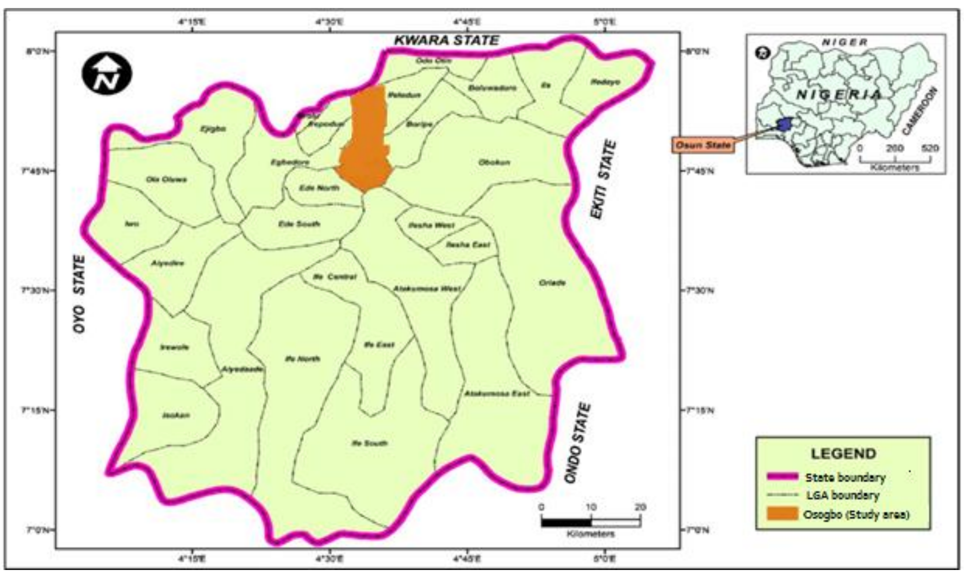


Figure 1: Study area map of Osogbo, Osun State.

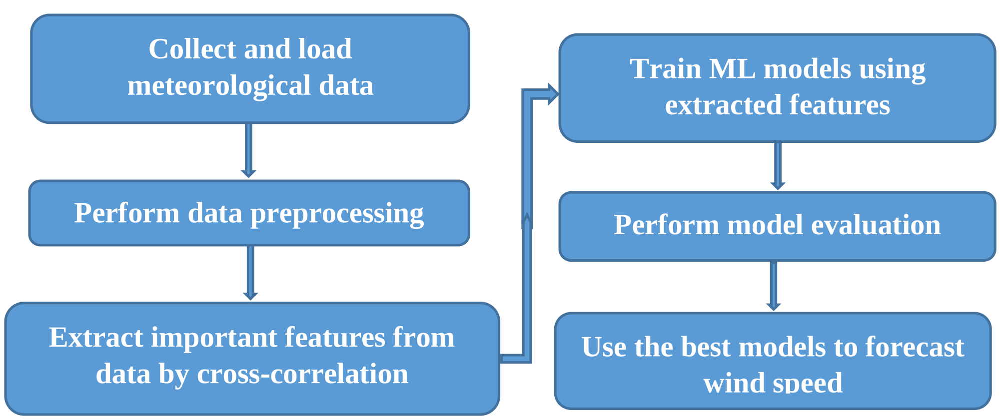


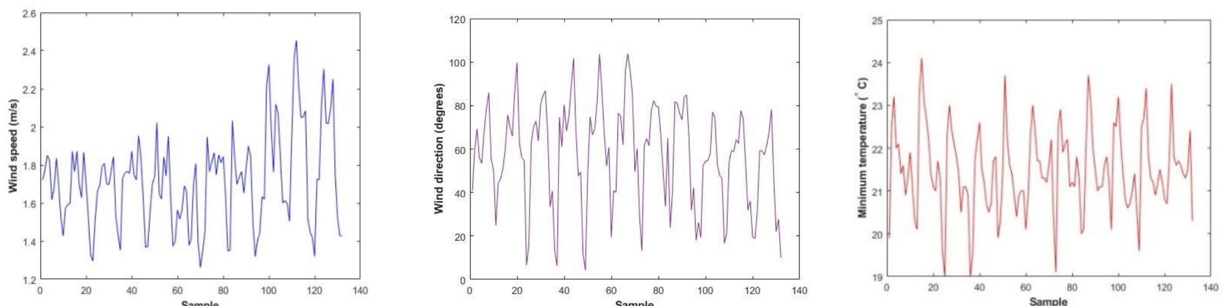
Figure 2: Steps for building wind speed ML forecasting models

Feature Extraction

Cross-correlation technique was used to extract the important features to build the machine learning models important from the data in order to forecast the wind speed. The cross-correlation coefficients obtained between wind speed and the other variables are presented in Table 1 with minimum temperature and wind direction variables having the highest correlation coefficient with respect to wind speed. The corresponding probability values (p-values) for these variables are less than 0.05 and therefore, significant. As a result, the minimum temperature and wind direction variables were chosen as input while wind speed was selected as the output variable. Plots of the mean monthly values of the wind speed, wind direction, and minimum temperature data samples are plotted in Figure 3(a)-(c) to show their variation over time. The 10 m height wind speed values ranged between 1.26 m/s - 2.46 m/s, the wind direction ranged between 4.3°-103.7° while minimum temperatures values were between 19℃ and 24.1℃.

Table 1: Correlation coefficients and p-values between wind speed and other weather parameters

	Rainfall	Maximum temp.	Minimum temp.	Relative humidity	Wind speed	Wind direction	Radiation
Correlation coefficient	0.049	0.057	0.222	0.024	1.000	0.375	-0.059
Corresponding p-value	0.612	0.564	0.022	0.809	0	0	0.546



(a) Wind speed (b) Wind direction (c) Minimum temperature

Figure 3: Variation of selected weather variables over time

Model Development

Decision trees and support vector machine models were selected for wind speed forecasting in this work.
Decision trees: Decision trees are widely used supervised learning algorithms for classification and regression tasks. They partition the feature space into hierarchical tree structures and produce an intuitive and interpretable representation of decision-making processes. When used for classification, the output of decision trees is a discrete value while their output is continuous-valued when they are used for regression (Rani et al., 2022). In this work, the regression learner app in MATLAB's Statistics and Machine Learning toolbox (The Mathworks, 2022) was used to build "fine tree" and "medium tree" models for wind speed forecasting. The minimum leaf size is a parameter that determines the size or depth of the decision tree model and has values of 4 and 12 for the fine tree and medium tree models respectively.

Support Vector Machines: Support Vector Machines are also supervised machine learning algorithms useful for solving both classification and regression problems. When the goal is to solve regression problems like wind speed forecasting involving the prediction of continuous values, they are specifically referred to as support vector regression models and they aim to find an optimal hyperplane that maximizes the margin from the predicted values in a high-dimensional feature space while allowing a predefined level of tolerance (epsilon) for errors or deviations (Keyvanpour and Shirzad, 2022). MATLAB's regression learner app was similarly used to develop the SVM regression models in this work. Depending on the type of kernel function used, different types of SVMs are available all of which vary in interpretability and model flexibility (The Mathworks, 2022), and as a compromise, the quadratic SVM and the medium Gaussian SVM models were used in this study.

Model Evaluation

To evaluate the performance of the decision tree and SVM models, different performance metrics were selected.

Root Mean Square Error (RMSE): RMSE is a measure of the average magnitude of the prediction errors. It calculates the square root of the average of the squared differences between the predicted and true values. RMSE is sensitive to outliers and gives higher weight to larger errors. It is calculated using Equation (1), where y_t and $y_t{}'$ are the true and predicted time sequence respectively and n is the number of samples. Smaller RMSE value indicates better model performance.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - y_t{}')^2} \quad (1)$$

Mean Absolute Error (MAE): MAE is a measure of the average absolute difference between the predicted and true values. It calculates the average of the absolute differences without considering the direction of the errors. MAE is less sensitive to outliers compared to RMSE, as it does not square the errors. It is calculated using Equation (2). Smaller MAE values indicate better model performance.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - y_t{}'| \quad (2)$$

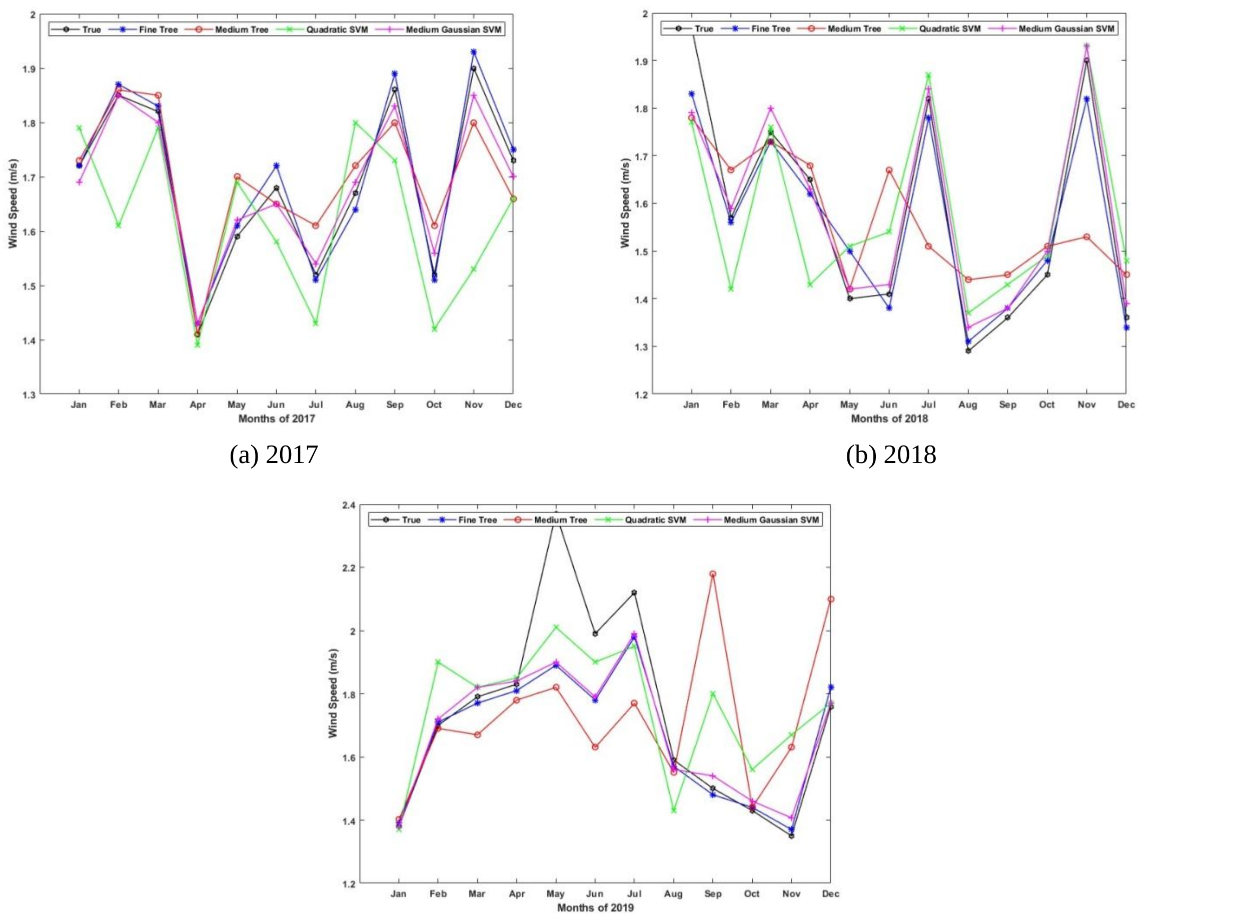
Coefficient of Determination (R-Squared): R-squared measures the proportion of the variance in the dependent variable that can be explained by the independent variables in a regression model. It indicates how well the model fits the data and its value ranges from 0 to 1. Higher R-squared values indicate a better fit. A value of 0 indicates that the model does not fit the data at all while a value of 1 indicates that the model fits the data perfectly. It is calculated using Equation (3).

$$R\text{-}Squared = 1 - \frac{\sum_{t=1}^{n}(y_t - y_t{}')^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \quad (3)$$

3. RESULTS AND DISCUSSION

The decision tree and SVM models were trained using values of wind direction and minimum temperature as inputs while the wind speed was taken as the output. To prevent overfitting, training of the models was done using five-fold cross-validation while the data for 2017 to 2019 were used for testing the performance of the models. Figure 4 (a)-(c) shows the plots of the actual and forecasted monthly average wind speed for 2017, 2018, and 2019 respectively. Although all four models captured the general trend in the test dataset, the Medium Gaussian support vector machine and Fine Tree models had the best predictions. The largest deviations between the true wind speed values and the values predicted were recorded for the medium tree and quadratic support vector machine models. Predicted values for the Medium Gaussian SVM tend to be higher than the true output values for most months of 2017-2019 while predictions by the Fine Tree model are generally underestimations of the true wind speed values. Comparing the accuracies of prediction across the years, all the models recorded the highest deviations from the true values from May to December of 2019.

To quantitatively evaluate the performance of the models on the testing dataset, the RMSE, MAE and R-Squared values between the true and predicted wind speeds for each model are shown in Table 2 with the lowest values attained shown in bold font. For the RMSE, the Fine Tree model gave the best performance for 2017 and 2018 testing data while for 2019, the Medium Gaussian SVM performed better than the other models. In terms of the MAE, the lowest values were obtained by the Fine Tree model for both 2017 and 2019 while Medium Gaussian SVM gave the lowest value for 2018. The Fine Tree and Medium Gaussian SVM models returned high coefficient of determination (R-squared) values for 2017 and 2018 which is an indication that both models fitted the data very well while the values for 2019 indicated a moderately good fit. The R-squared values also indicated that the Quadratic SVM and Medium Tree models failed to obtain a good fit of the data for 2017 and 2019 respectively. All the models did not achieve a good fit of the 2019 data- this is in agreement with the observed trend that the RMSE and MAE values were at their highest with respect to the 2019 data. In summary, the Fine Tree model gave the best overall performance out of all the decision tree and support vector machine models.



(a) 2017 (b) 2018

(c) 2019

Figure 4: True and predicted wind speeds for the year 2017-2019

Table 2: Performance evaluation of models

	RMSE			MAE			R-Squared		
Models	2017	2018	2019	2017	2018	2019	2017	2018	2019
Fine Tree	**0.0227**	**0.0592**	0.1581	**0.0200**	0.0450	**0.0850**	**0.9766**	**0.9306**	0.7206
Medium Tree	0.0654	0.1804	0.3202	0.0542	0.1408	0.2343	0.8062	0.3542	0.0000
Medium Gaussian SVM	0.0292	0.0606	**0.1544**	0.0267	**0.0425**	0.0856	0.9615	0.9271	**0.7337**
Quadratic SVM	0.1521	0.1193	0.1921	0.1208	0.1008	0.1500	0.0000	0.7179	0.5875

4. CONCLUSION

In this study, decision tree and support vector machine models were developed to forecast wind speed for Osogbo, Osun State in Southwestern Nigeria for possible wind power generation. The models utilized monthly averages of minimum temperature and wind direction as inputs. Of the decision tree models, the Fine Tree model gave the best overall performance in forecasting wind speed. The Medium Gaussian SVM model produced the best performance out of the support vector machine models. Although the wind speed values recorded for Osogbo at a height of 10 m are below wind turbines' cut-in speed of 3 m/s, it is expected that the wind speed values measured at higher altitudes would be up to the desired level for wind power generation. In future studies, the effect of hyperparameter optimization on the performances of the models will be investigated.

References

Abolude, A. and Zhou, W. (2017). A preliminary analysis of wind turbine energy yield. *Energy Procedia*, 138: 423-428.

Adedipe, O., Abolarin M. S. and Mamman, R. O. (2018). A Review of onshore and offshore wind energy potential in Nigeria. *IOP Conf. Series: Materials Science and Engineering* 413: 012039.

Audu, M. O., Terwase, A. S. and Isikwue, B. C. (2019). Investigation of wind speed characteristics and its energy potential in Makurdi, North Central, Nigeria. *SN Applied Sciences,* 1: 178.

Axaopoulos, P. J. and Tzanes, G. T. (2022). Wind energy potential (measurements, evaluation, forecasting), In Comprehensive Renewable Energy (Second Edition), Editor(s): Trevor M. Letcher, Elsevier, pp 79-103.

Camelo, H. N., Lucio, P. S., Junior, J. B. V. L. and De Carvalho, P. C. M. (2018). A hybrid model based on time series models and neural network for forecasting wind speed in the Brazilian northeast region. *Sustainable Energy Technologies and Assessments*, 28: 65-72.

Cetinay, H., Kuipers, F. A. and Guven, A. N. (2017). Optimal siting and sizing of wind farms. *Renewable Energy,* 101: 51-58.

Chang, W. Y. (2014). A literature review of wind forecasting methods. *Journal of Power and Energy Engineering,* 2: 161-168.

Chen, Q. and Folly, K. A. (2018). Wind power forecasting, *IFAC-Papers Online,* 51(28): 414-419.

Cheng, W. Y. Y., Liu, Y., Bourgeois, A. J., Wu, Y. and Haupt, S. E. (2017). Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation. *Renewable Energy,* 107: 340-351.

Chodakowska, E., Nazarko, J. and Nazarko, Ł. (2021). ARIMA models in electrical load forecasting and their robustness to noise. *Energies,* 14: 7952.

Danladi, A., Stephen, V., Obot, I., Oduma, F. O., and Maaji, N. S. (2020). Modeling and short-term forecasting of wind speed potentials for the generation of electricity for Nigerian Army University Biu. *Asian Journal of Emerging Research,* 2(1): 25-34.

Egbunu, C. O., Ogedengbe, M. T., Yange, T. S., Rufai, M. A. and Muhammed, H. I. (2021). Towards food security: the prediction of climatic factors in Nigeria using random forest approach. *Journal of Computer Science and Information Technology,* 7(4): 70–81.

Fadare, D. A. (2010). The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. *Applied Energy,* 87(3): 934–942.

Han,Y., Mi, L., Shen, L., Cai, C.S., Liu, Y., Li, K. and Xu, G. (2022). A short-term wind speed prediction method utilizing novel hybrid deep learning algorithms to correct numerical weather forecasting. *Applied Energy,* 312:118777.

Idris, W. O., Ibrahim, M. Z. and Albani, A. (2020). The status of the development of wind energy in Nigeria. *Energies*, 13(23): 6219.

IRENA (2023). Renewable Energy Roadmap: Nigeria, International Renewable Energy Agency, Abu Dhabi.

Keyvanpour, M. R. and Shirzad, M. B. (2022). Chapter 14 - Machine learning techniques for agricultural image recognition, In Editor(s): Mohammad Ayoub Khan, Rijwan Khan, Mohammad Aslam Ansari, Application of Machine Learning in Agriculture, Academic Press, Pages 283-305.

Lahouar, A. and Slama, J. B. H. (2017). Hour-ahead wind power forecast based on random forests. *Renewable Energy*, 109: 529-541.

Liu, B., Ma, X., Guo, J., Li, H., Jin, S., Ma, Y. and Gong, W. (2023). Estimating hub-height wind speed based on a machine learning algorithm: implications for wind energy assessment. *Atmospheric Chemistry and Physics*, 23: 3181–3193.

Maduabuchi, C., Nsude, C., Eneh, C., Eke, E., Okoli, K., Okpara, E., Idogho, C., Waya, B., and Harsito, C. (2023). Renewable energy potential estimation using climatic-weather-forecasting machine learning algorithms. *Energies*, 16(4): 1603.

Men, Z., Yee, E., Lien, F. S., Wen, D. and Chen, Y. (2016). Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. *Renewable Energy*, 87(1): 203-211.

Odo, F. C., Offiah, S. U. and Ugwuoke, P. E. (2012). Weibull distribution-based model for prediction of wind potential in Enugu, Nigeria. *Advances in Applied Science Research*, 3 (2):1202-1208

Olaiya, F. and Adeyemo, A. B. (2012). Application of data mining techniques in weather prediction and climate change studies. *International Journal of Information Engineering and Electronic Business*, 1: 51-59.

Oyedepo, S. O. and Adaramola, M. S. (2012). Analysis of wind speed data and wind energy potential in three selected locations in south-east Nigeria. *International Journal of Energy and Environmental Engineering*, 3(7):1–11.

Quej, V. H., Almorox, J., Arnaldo, J. A. and Saito, L. (2017). ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *Journal of Atmospheric and Solar-Terrestrial Physics*, 155: 62-70.

Rani, A., Kumar, N., Kumar, J., Kumar, J., and Sinha, N. K. (2022). Machine learning for soil moisture assessment. In: R. C. Poonia, V. Singh and S. R. Nayak (Eds.) Deep Learning for Sustainable Agriculture, 143-168.

Rawat, R., Oki, O. A., Sankaran, K. S., Olasupo, O., Ebong, G. N. and Ajagbe, S. A. (2023). A new solution for cyber security in big data using machine learning approach. In: S. Shakya, G. Papakostas, K. A. Kamel (Eds.) Mobile Computing and Sustainable Informatics. Lecture Notes on Data Engineering and Communications Technologies, 166: 495-505, Springer, Singapore.

Rurumah, A. S., Umar, F. A., Lawan, S. M., Kawu, S. J. Okpala, C. N. and Abidin, W. A. W. Z. (2021). Wind energy mapping based on quantum GIS, measured and predicted wind speed in Katsina State. *FUDMA Journal of Sciences*, 5(3): 18-28.

Santamaría-Bonfil, G., Reyes-Ballesteros, A. and Gershenson, C. (2016). Wind speed forecasting for wind farms: a method based on support vector regression. *Renewable Energy*, 85: 790-809.

Shao, B., Song, D., Bian, G. and Zhao, Y. (2021). Wind speed forecast based on the LSTM neural network optimized by the firework algorithm. *Advances in Materials Science and Engineering*, 2021: 4874757.

Soman, S. S., Zareipour, H., Malik, O. P. and Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. *In Proceedings of the 2010 North American Power Symposium*, Pp. 1 - 8.

Ssekulima, E. B., Anwar, M. B., Hinai, A. A., and El Moursi, M. S. (2016). Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review. *IET Renewable Power Generation*, 10(7): 885-989.

Stathopoulos, C., Kaperoni, A., Galanis, G. and Kallos, G. (2013). Wind power prediction based on numerical and statistical models. *Journal of Wind Engineering and Industrial Aerodynamics*, 112: 25-38.

The MathWorks Inc. (2022). MATLAB version: 9.12.0 (R2022a), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

Tronsco, A., Salcedo-sanz, S., Casanova-Mateo, C., Riquelmw, J. C. and Prieto, L. (2015). Local models-based regression trees for very short-term wind speed predictions. *Renewable Energy*, 81: 589-598.

Uquetan, U. I., Egor, A. O., Osang, J. E. and Emeruwa, C. (2015). Empirical study of wind energy potential in Calabar, Cross River State Nigeria. *International Journal of Scientific and Technology Research*, 4(10):113–121.

Xie, A., Yang, H., Chen, J., Sheng, L. and Zhang, Q. (2021). A short-term wind speed forecasting model based on a multi-variable long short-term memory network. *Atmosphere*, 12(5): 651.

Zendehboudi, A., Baseer, M. A. and Saidur, R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of Cleaner Production*, 199: 272-285.