Towards Effective Wind Power Prediction using Machine Learning Algorithms

Devi C. Arati[1, a)], Parvathy S. Menon[1, b)], Sikha O. K. [2,c)], Jithin Velayudhan[1,d)], Prabaharan Poornachandran[3,e)] and Arun K. Raj[1, f)]

[1]*Center for Computational Engineering & Networking, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham 641112, India.*
[2]*Department of Computer Science and Engineering, Amrita School of Computing, Coimbatore, Amrita Vishwa Vidyapeetham 641112, India*
[3]*Center for Internet Studies and Artificial Intelligence, Amrita Vishwa Vidyapeetham, Amritapuri 690546, India.*

[a)] cb.en.p2dsc22003@cb.students.amrita.edu, [b)] cb.en.p2dsc22011@cb.students.amrita.edu,
[c)] ok_sikha@cb.amrita.edu, [d)] v_jithin@cb.amrita.edu, e) prabha@amrita.edu, [f)] kr_arun@cb.amrita.edu
[f)] Corresponding author: kr_arun@cb.amrita.edu

Abstract. Wind energy data patterns are often volatile, highly random, and spasmodic. Further, geographical and/or meteorological dependence adds more complexity to data handling. Therefore, effective and accurate wind power prediction has always been the challenge, as it contributes to minimizing the hindrance in smart grid integration and brings out safety and reliability in its operation. Accordingly, this work emphasizes improving the prediction accuracy and determining whether using any of the measured data below the hub height is sufficient to accurately predict wind power generation at the hub, lowering the cost of wind data measurements. In this context, ten features comprising wind and meteorological parameters are analyzed at various heights (i.e., 10, 30, 50 m) and the hub level. Two-year wind power generation data from the SCADA system corresponding to two selected wind farms in China are used for training, validation, and testing the four machine learning models. Wind speed and direction and ambient temperature are the significant features, while ambient pressure and relative humidity demonstrated the least significance in influencing the prediction accuracy. The outcomes indicated that the measurement data corresponding to 30 m yielded wind power generation results comparable to the hub. The findings suggested that the random forest model resulted in more accurate predictions for low and high wind speed values, with R^2 values ranging from 0.9708 to 0.9942 and RMSE values from 0.0234 to 0.0397. The decision tree had the least performance accuracy at any wind speed range with an R^2 value as low as 0.7438 and RMSE as high as 0.1251.

INTRODUCTION

Renewable energy integration into the smart grid has been the motto for promoting sustainable development and carbon neutrality [1]. Energy security, decentralization, environmental benefits, and sustainability are the critical features of renewable energy that caused their widespread acceptance. Solar, hydroelectric, wind, biomass, and geothermal are the renewable resources currently in use or under development and are naturally replenished on a human timescale. Over the past few decades, wind energy has been in the spotlight as the cost of generating electricity from wind has decreased tremendously, making it more competitive with other forms of electricity generation. According to the International Energy Agency (IEA), the global installed capacity of wind energy generation has grown by 273 TWh in 2021, i.e., a 17% rise compared to 2020 [2]. China (70%), the United States (14%), and Brazil (7%) contributed to this rise, and this is a direct indicative of their stringent Net Zero policy reforms. Nevertheless, to achieve the ambitious Net Zero Emission goals by 2030, the average installed capacity needs to be increased several folds from the current capacity [2].

It is worthwhile to note that even though the cost of wind electricity generation has decreased in recent years, power generation using coal or natural gas is still cheaper unless wind power is supported with governmental subsidies or other forms of support. In addition to the capital cost, excessive land use, intermittency, technical challenges (design, installation, maintenance), and weather conditions hinder wind power generation. With such complexity, prediction/forecasting power generation are essential because it aids in optimizing the use of wind energy and enables energy companies, stakeholders, and grid operators to plan and manage the generated power and its integration into the grid effectively [3,4]. In turn, it helps to smoothen the grid operation during on and off-peak consumer load demands. Concurrently, accurate predictions can help optimize the operation of wind farms by identifying patterns in the data and providing recommendations for how to maximize electricity generation [5]. Also, predicting the availability of wind energy can improve economic viability allowing energy companies to estimate the cost incurred with power generation and distribution, which can aid in transitioning to a low carbon economy.

A common method used to improve the wind electricity generation prediction accuracy is to employ Machine Learning (ML) algorithms [3,6]. Using ML, large amounts of historical data, such as wind parameters (speed, direction), and ambient conditions (temperature, relative humidity, and other relevant factors), can be analyzed to build effective models that, in turn, can predict future wind power generations. Such developed models can accurately predict wind power generations over different locations, ranging from over a few hours ahead to several days or weeks in advance. Some of the different ML algorithms popularly used for wind power predictions are linear regression, neural networks, and decision trees.

Wickramasinghe *et al.* [7] compared wind energy prediction results obtained from five different ML and two statistical tools using the historical data (1 year) from a wind farm in Sri Lanka. The tools considered were Support Vector Regression (SVR), Gaussian Process Regression (GPR), Feed Forward Backpropagation Neural Network (FFBPNN), Cascade Forward Backpropagation Neural Network (CFBPNN), Recurrent Neural Network (RNN), Multiple Linear regression (MLR), and Power Regression (PR). The accuracy of the results was evaluated based on Root mean square error (RMSE), coefficient of determination (R^2), and Nash-Sutcliffe Efficiency (NSE). The study revealed that the FFBPNN model yields more accurate results with the slightest error. Also, the investigation to identify the pairwise correlation among ambient conditions, such as temperature, wind speed, and wind direction, suggested that the daily average wind energy strongly depends on wind speed and temperature. Similarly, Buturache and Stancu [8] reported a comparative study on wind power predictions based on Artificial Neural Networks (ANN), SVR, Random Forests (RF), and random trees. ANN showed better results than SVR and RF algorithms based on the dataset considered. Also, the study suggested that different methods would yield better results depending on whether the model is trained for short, medium, or long-term predictions. Shabbir *et al.* [9] compared the Support Vector Machine (SVM) algorithm's prediction accuracy with an in-house algorithm known as Elering, developed by the Estonian Energy Company. The authors considered one-month wind power generation data to predict the power generation for the next day. The findings showed that the SVM algorithm made more accurate predictions than the Elering, indicating the advantages of using ML for short-term forecasting.

Demolli *et al.* [10] examined the viability of using five different machine learning algorithms, namely, Least Absolute Shrinkage Selector Operator (LASSO), RF, SVR, extreme gradient boosting (XGBoost), and k Nearest Neighbor (kNN) regression for the long-term forecasting of wind power generation. Using the hourly data of wind speeds, the daily average wind speed dataset was generated, and the corresponding daily average power generation was estimated. In terms of R^2 and Mean Absolute Error (MAE), the results revealed that RF yielded more accurate results, while LASSO yielded poor results due to its linear bias. Furthermore, the study suggested that the developed ML models can be used to forecast wind power generation of wind farms before their establishment in an unknown geographical location if a similar dataset suitable to the site is available. A similar study performed by Alkesaiberi *et al.* [11] examined the performance of ML algorithms, such as SVR with different kernels, Ensemble learning (ES) models, and Bayesian Optimization (BO) to optimally tune hyperparameters of the Gaussian Process Regression (GPR) for wind power generation forecasting. The work comprised three stages: forecasting using univariate wind power time-series data, dynamic information incorporation to improve the forecasting accuracy, and introducing lagged measurements for capturing the time evolution into the design of the ML models considered. Apart from these, more input variables, such as wind direction and speed, are included to improve forecasting accuracy. The study highlighted that the dynamic models resulted in accurate forecasting results compared to static models, and the inclusion of suitable input parameters resulted in higher accuracy. Among the models considered, GPR and ES models obtained a reasonable prediction accuracy.

Several studies also focused on short-term wind power generation forecasting. In this context, Zheng and Wu [12] developed a new model comprising XGBoost, weather similarity analysis, and feature engineering. Also, the proposed model outcomes were compared with a simple XGBoost model and different traditional ML techniques, such as RF, SVR, and Classification and Regression Tree (CART). As expected, the results suggested that the developed model predictions were accurate due to the inclusion of weather similarity analysis and feature engineering. Further, the predictions can be improved by including spatial features (environment and geographic) pertinent to the farm location. In order to predict daily wind energy generation patterns, a similar comparison study of different ML algorithms (such as RF, ANN, and deep neural network) was conducted by Bochenek *et al.* [13] using the 2-year datasets obtained from the Institute of Meteorology and Water Management–National Research Institute in Poland. The outcomes recommended that the XGBoost resulted in more accurate forecasting results.

Another study reported by Singh *et al.* [14] investigated the forecasting accuracy of Gradient Boost Machine (GBM), RF, decision tree, and kNN models for short-term forecasting of the generated power from Turkish farms using the historical data comprising wind direction and wind speed. The findings revealed that the GBM produced more accurate results based on their respective MAE, Mean Absolute Percentage Error (MAPE), and RMSE scores. On the contrary, the decision tree resulted in poor forecasting accuracy. The possible cause of poor accuracy is that these models are not structurally stable and robust. Therefore, slight variations in the dataset can lead to different tree structures, resulting in other predictions during the validation. Further the authors [15] extended their study and the outcomes suggested that employing a minor reconstruction error can minimize the complexity involved with forecasting. On the contrary, Diop *et al.* [16] examined the prediction accuracy of decision trees and RF algorithms to predict wind power generation from the Taiba plant in Senegal. The models are trained using 1-year data from the plant. The short-term prediction results revealed that the decision tree and RF yield comparatively similar results with almost identical R^2 values.

Based on the cited literature and the author's knowledge, predicting wind power generation for short-term utilizing historical data has been ongoing research over the last few decades. In the aforementioned literature, several investigators have used different ML and statistical tools for developing accurate prediction models by considering a few or all meteorological parameters depending on the dataset availability, complexity, and desired outcomes. It is worth noting that the suitability of ML algorithms and their prediction accuracy varies based on the selected input parameters, desired output, dataset pre-processing, and analysis. Apart from these reported works, our analysis focuses on determining the best ML algorithm that yields accurate predictions based on the historical data (2 years) of two selected wind farms in China with a nominal wind generation capacity of 98.1 MW and 64.6 MW, respectively. Additionally, the current work focuses on determining the feasibility of using measurement data at heights lower than the hub level to predict wind power generation at the hub level accurately. Selected ML algorithms for comparison are Decision tree, Random Forest, Extreme Gradient Boosting), and Gradient Boosted Regression Trees. The selected input parameters are wind speed and directions at different heights (i.e., 10, 30, 50 m, and hub), ambient temperature, relative humidity, and atmospheric pressure, while the output parameter is the actual wind power generated.

This article is further divided into the following sections: Section II discusses the details of analyzing and pre-processing the dataset along with the four ML algorithms used for the investigation; Section III highlights the significance of the obtained results, provides comparative analysis, and comparison with the reported works from literature; Section IV provides an overall summary and outlines the future scope of the current work.

PROPOSED MODEL AND METHODOLOGY

Dataset Analysis and Pre-processing

This section highlights the process of estimating suitable input parameters that influences the actual power output of the wind turbines. The dataset analysis and prediction are made over a publicly available dataset acquired from two wind farms (represented hereafter as Site 1 and Site 2) located in China (accessed on 23rd November 2022) [17]. The dataset consists of ten features, namely, wind speeds (m/s) and wind directions (°) at 10 m, 30 m, 50 m, and at hub level measured from the ground, temperature (°C), ambient pressure (kPa), relative humidity (%) and power (MW). These are the ten selected input and output variables.

Prior to obtaining an accurate prediction model, analyzing the dataset is of utmost importance. Accordingly, an outline of steps undertaken for analyzing the dataset are discussed. Generally, wind farms are monitored using the Supervisory Control and Data Acquisition (SCADA) system. It connects the hardware and software for monitoring, controlling, and analyzing processes. All measurements were recorded as a time series dataset that logs the data every 15 minutes from 1st January 2019 to 31st December 2020. Based on the dataset, the nominal wind generation capacity varied from 98.1 MW for Site 1 to 64.6 MW for Site 2. Site 1 consists of 62 wind turbines, 50 of 1500 kW and 12 of 2000 kW, while Site 2 consists of 48 turbines, each with 200 kW capacity. The wind speed at the height of the wheel hub varied from 0 m/s to 31.1 m/s (Site 1) and 30.2 m/s (Site 2). The air temperature varied from −24.5 °C to 37.6 °C at Site 1 and 0 °C to 37.1 °C at Site 2. It is important to note that the height of the hub is 85.5 m and 90 m from the ground level for the turbines at Site 1 and Site 2, respectively. Weather conditions at different height levels exhibited a similar trend. The dataset was available as raw data and processed data. The processed data is considered for the study, which contains no null value. The obtained data holds 70,176 observations. The dataset is separated into two parts, namely, 70% of the dataset is considered for training, and the remaining is considered for testing.

Once the dataset is analyzed, pre-processing the data is another unavoidable step in machine learning. In pre-processing, dropping or modifying the data points occurs before the information is utilized to improve the model's efficiency. Often, in the SCADA system, the data points can be of a different magnitude order. Hence, it is essential to normalize all the values to the same magnitude. Normalization is achieved using the relation,

$$X' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Here X' is the scaled data in the range (0,1), x is the original data, x_{max} is the maximum value in the dataset and x_{min} is the minimum value in the dataset. It must be noted that from the dataset, atmospheric data (relative humidity, ambient temperature, and ambient pressure), and wind speed and wind directions at hub level and at different heights are considered for the investigation.

Selected Algorithms

This section briefly discussed the various algorithms used to predict the wind power production. Decision tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Gradient Boosted Regression Trees (GBRT) are considered for the analysis. These algorithms contain various hyperparameters that influence their performance. Optimum hyperparameters can be selected using GridSearchCV.

DT algorithm is one of the supervised machine learning algorithms that are intrinsically explainable [18]. It builds regression or classification models as a tree structure, making it self-explanatory. The model is trained to predict the value of a target variable or class by learning the decision rules derived from the past data (i.e., training data). It divides the data at each node repeatedly using metrics such as MSE, entropy, Gini index, etc. The decision tree-based model provides better prediction/classification accuracy using a more extensive training dataset. The current work employs the decision tree as a regressor with a maximum depth of five and a learning criterion as 'Friedman MSE'. A RF regressor is a homogeneous ensemble model that combines multiple decision trees to make the prediction. The model divides the original dataset into multiple sub-parts, and each sub-part is trained with a distinct decision tree. The decision/predictions from the individual trees were then given to a voting module, which uses a majority voting technique to provide a more stable and accurate prediction [19]. XGBoost is an optimized ensemble learning model that improves the performance of weak learners, such as CART. It employs a gradient-boosting mechanism with decision trees as weak learners [20]. The decision trees are arranged sequentially, and weights are assigned to all independent variables. The weights of variables associated with the incorrect prediction increase and these variables are subsequently given the second decision tree. These distinct classifiers/predictors are combined to form a more robust and precise model. GBRT, also known as shorter Gradient Boosting, is a non-parametric statistical learning approach for classification and regression. GBRT integrates multiple regression trees as the weak learners to improve the prediction [21]. A gradient-boosted trees model is constructed similarly to other boosting algorithms such as Adaboost, and it generalizes well than other models by optimizing the differentiable loss function.

RESULTS AND DISCUSSIONS

As discussed earlier, apart from determining the most suitable ML algorithm that can accurately predict the wind power generation for the short term for the considered Site 1 and Site 2 wind farms in China, this work attempts to determine the possibility of utilizing measurements at heights lower than at the hub level for accurate predictions. In practice, measuring the meteorological data (such as wind speed, ambient temperature, wind direction, relative humidity, and atmospheric pressure) using direct measuring tools at the hub level of the wind turbine is recommended. However, measurements at the hub level require the installation of Wind Measurement Towers (MET), which are thin, rigid structures. Due to its thin structure, the visibility of such towers often raises safety concerns, especially for low-flying aircraft. Also, if accurate wind power generation predictions can be made using the measurement data from lower heights (i.e., below the hub level), it would help to minimize the overall cost and have an impact on the farm economic evaluation process.

FIGURE 1, shows the measured values of daily wind speed data at different heights (i.e., 10, 30, 50 m, and at the hub) from Site 1 and Site 2. From the figure, it is clear that the wind speed profiles of different heights follow an identical pattern. Based on the observations, it can be inferred that the wind speed data at different heights are closely correlated, which indicates the possibility of using the measured data at lower heights for predicting the wind power generation at the hub. **FIGURE 2** illustrate the feature importance obtained from the trained/fitted DT model using

MSE. The feature importance plot indicates how significant each feature in the dataset is for the analysis. In DT, each node represents a conditional statement, while its branches indicate the outcomes of the conditional statement. In principle, the algorithm iterates from the highest node (root node) to the bottom-most node (leaf node). In turn, the leaf node indicates the decision formed after executing all the attributes in the above nodes. Analogous to the terminology, the terminal node indicates the critical feature contributing to the prediction. In contrast, feature relevance decreases as we move from the root node to the leaf node via the intermediate decision nodes. As observed in **FIGURE 2**, the DT model suggests that the wind speed data at the hub is the critical feature that influences wind power generation for both Site 1 and Site 2. This, in turn, agrees with the physics of power generation using wind turbines, reflecting that the wind energy available at the hub rotates the wind blades, leading to the maximum possible power generation. Also, the wind power curve depends mainly on the wind velocity at the hub. Hence, it is understood that using measurement data at a location below the hub height for prediction reduces the dependency on the actual power generation. Also, the second important feature that influences the wind power generation for the datasets obtained from Site 1 and Site 2 is the wind speed values at 50 m height. However, this work attempts to determine whether near-accurate predictions are possible with measurement data obtained at still lower heights. Hence, the further analysis compares the results corresponding to 30 m height and at the hub level for Site 1 and Site 2, respectively. Apart from wind speed and direction at different heights/altitudes, it is observed that power generation depends on ambient temperature, among other atmospheric variables. The observations highlight the negligible dependence of relative humidity and ambient pressure on wind power generation.

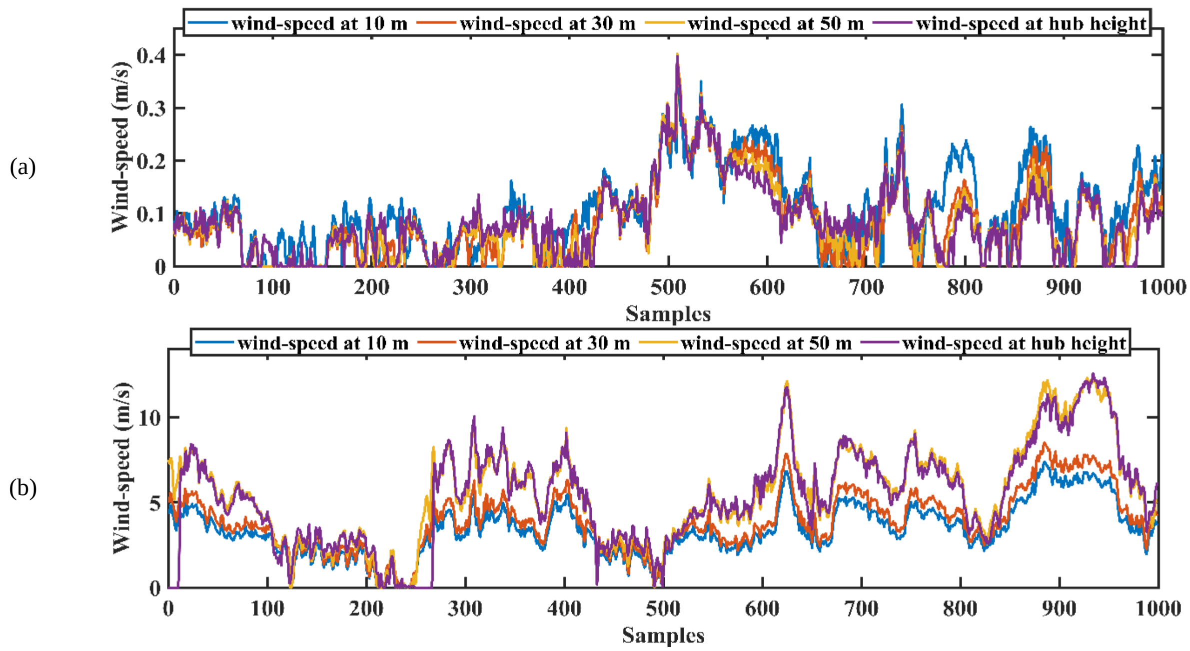

FIGURE 1. *Measured wind speed data corresponding to different heights for (a) Site 1, and (b) Site 2*

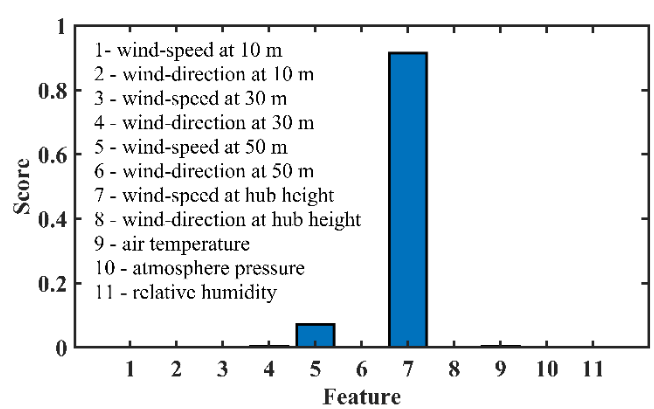

FIGURE 2. *Feature importance of input parameters for Site 1*

TABLE 1, discusses the resultant observations obtained from the four different ML models considered for this analysis. The performance of all models is analyzed based on statistical indicators such as MAE, RMSE, MSE, and R^2. The MAE reflects the average absolute difference between predicted and actual values. In contrast, RMSE is expressed as the square root of the average squared differences between the predicted and actual values. Generally, the lower the RMSE value, the better the model predictions. Also, if the RMSE values of the train and test samples are within a close range, then the model is regarded as the better model and without any overfitting. The MSE is the average square of the errors, and R^2 checks how well the model reproduces the observed outputs. In short, the model with the least error corresponds to the model yielding accurate predictions. Among these indicators, RMSE is considered the primary indicator for comparison. In RMSE, the errors are squared before being averaged and imposed with a high weight for large errors. Also, there is no significant variation between the error magnitudes if the RMSE values are close to MAE values, signifying the model's effectiveness and generalization. From **TABLE 1**, it is inferred that the RF model has the lowest RMSE value. In the reported works of Demolli *et al.* [10], the RF model had better wind power generation prediction accuracy than XGBoost, SVR, LASSO, and kNN models.

TABLE 1. *Performance evaluation of different ML algorithms at different heights and hub level*.

	Height	Algorithm	MSE	RMSE	MAE	R^2
Site 1	10 m	DT	0.01564	0.12509	0.08897	0.74387
		RF	**0.00158**	**0.03979**	**0.02618**	**0.97408**
		XGBoost	0.01305	0.11424	0.08185	0.78637
		GBRT	0.01233	0.11059	0.07693	0.79811
	30 m	DT	0.01080	0.10396	0.06782	0.82308
		RF	**0.00118**	**0.03445**	**0.02124**	**0.98056**
		XGBoost	0.00941	0.09704	0.06533	0.84585
		GBRT	0.01097	0.10474	0.07073	0.82042
	50 m	DT	0.00993	0.09965	0.06289	0.83744
		RF	**0.00107**	**0.03276**	**0.01951**	**0.98243**
		XGBoost	0.00865	0.09302	0.06084	0.85836
		GBRT	0.00933	0.09966	0.06414	0.83714
	Hub	DT	0.00977	0.09888	0.06169	0.83994
		RF	**0.00105**	**0.03252**	**0.01913**	**0.98268**
		XGBoost	0.00851	0.09229	0.05983	0.86056
		GBRT	0.08031	0.08961	0.05469	0.86855
Site 2	10 m	DT	0.01019	0.10096	0.06689	0.89425
		RF	**0.00073**	**0.02712**	**0.01587**	**0.99236**
		XGBoost	0.00686	0.08282	0.05575	0.92883
		GBRT	0.00904	0.09510	0.06305	0.90616
	30 m	DT	0.00899	0.09485	0.06122	0.90666
		RF	**0.00077**	**0.02777**	**0.01640**	**0.99199**
		XGBoost	0.00701	0.08377	0.05706	0.92719
		GBRT	0.00894	0.09459	0.06254	0.90717
	50 m	DT	0.00849	0.09217	0.05779	0.91185
		RF	**0.00065**	**0.02557**	**0.01462**	**0.99321**
		XGBoost	0.00620	0.07874	0.05248	0.93568
		GBRT	0.00939	0.09693	0.06646	0.90252
	Hub	DT	0.00767	0.08763	0.05598	0.92033
		RF	**0.00055**	**0.02349**	**0.01351**	**0.99427**
		XGBoost	0.00540	0.07350	0.04989	0.94395
		GBRT	0.00805	0.08977	0.06032	0.91639

FIGURE(S) 3 and **4**, shows the comparison of actual power generation with the predicted values for 250 sample datasets using the selected four ML models at Site 1 and Site 2, and also at 30 m and at hub level. For the comparison, 250 continuous samples are considered instead of all the data points for better visualization. It is noticeable that all four algorithms exhibit reasonably good prediction accuracy. However, based on the observations from **FIGURE(S) 3** and **4**, RF delivers the most accurate predictions for low and high wind speed values. It is substantiated by the low

RMSE and high R^2 values at all the different heights analyzed. This is because the RF model consists of multiple decision trees, and the output predicted from each decision tree is merged to obtain and stable and accurate prediction. It is known that by increasing the maximum depth of the tree, the prediction accuracy improves. Concurrently, it may lead to overfitting the dataset, which means that the model has a perfect fit with the dataset used for training but would not be able to generalize well with the testing dataset. Simultaneously, a low depth leads to under-fitting the prediction outcomes. For the current analysis, a maximum depth of 5 is found to be optimum. On the other hand, XGBoost and GBRT algorithms do not provide accurate predictions at low and high wind speed values. But XGBoost yields good results in the medium range of wind speeds followed by GBRT. Among all the four algorithms, the DT model exhibited poor performance with high prediction error, as highlighted by the statistical performance indicator values (i.e., high RMSE and low R^2 values) shown in **TABLE 1**. Reported works by Singh *et al.* [14]-[15] have also pointed out the poor prediction performance of DT models. A recent study by Kim *et al.* [22] also pointed out the possibility of accurately predicting the wind power generation at the hub using the measurement data obtained at a lower height than the hub. Based on the 1-year dataset obtained from the 14 wind farms across South Korea, the prediction accuracy of different ML models, namely, ANN, RF, kNN, and SVM, are compared. However, their work has considered only wind speeds at two different heights, i.e., 10 m and 80 m, respectively. The study suggested that the ANN model had better prediction accuracy, and using the data at 10 m, reasonable wind power generation at the hub can be estimated.

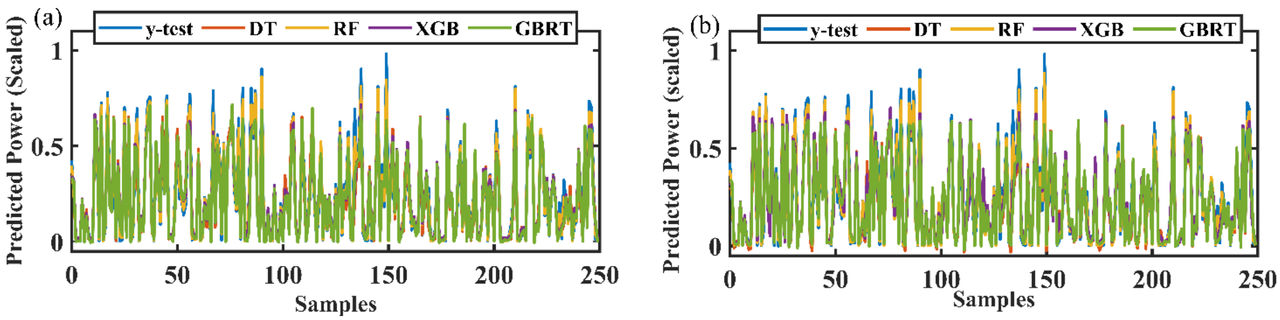

FIGURE 3. *Performance comparison of different ML algorithms for Site 1 using dataset corresponding to (a) Hub, and (b) 30 m*

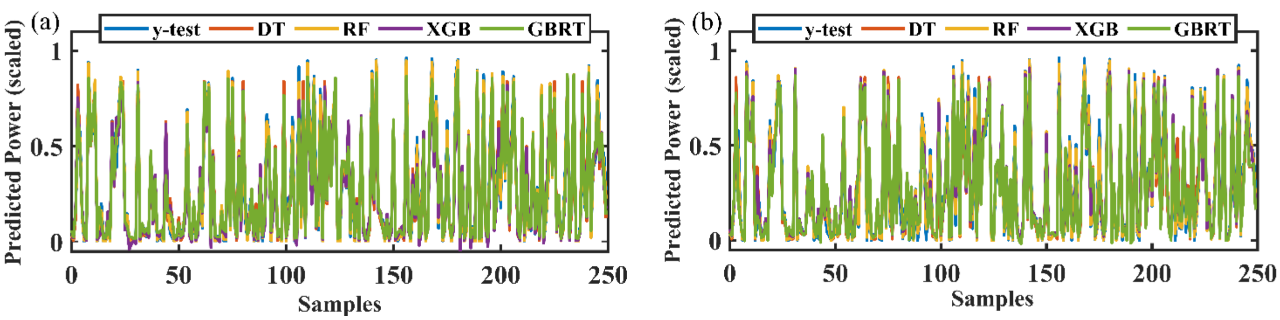

FIGURE 4. *Performance comparison of different ML algorithms for Site 2 using dataset corresponding to (a) Hub, and (b) 30 m*

CONCLUSION

Wind power generation and its prediction are crucial as they can guide the implementation of power-saving strategies, effective management, transmission, and distribution without disrupting normal grid operations. The current work considers the two-year historical data from two selected wind farms in China. The study attempts to determine the prediction accuracy of the wind power generation at the hub using the measurement data at three different heights (i.e., 10, 30, and 50 m) that are lower than the hub level. The analysis considered ML models, such as DT, RF, XGBoost, and GBRT, for prediction, and the effectiveness of prediction was analyzed based on statistical indicators, namely, MSE, RMSE, MAE, and R^2. The resultant findings revealed that wind power generation is closely dependent on the wind speed at the hub. As height increases, the effect of wind speed on power generation decreases. RF model had better prediction accuracy at low and high wind speed values, while XGBoost and GBRT exhibited accurate predictions for medium wind speed values. Also, the findings revealed that the wind speed at lower heights of 50 m or 30 m could also be considered for power prediction with decent accuracy rates depending upon the datasets from Site 1 and Site 2, respectively. In this work, the time frame was not a feature for power prediction. As an extension of this study, a seasonal analysis could be conducted considering the time data. Also, combinations of parameters that produce high or low power output can be determined. This can be used to determine and plan the maintenance periods of the wind farm in advance.

REFERENCES

1. Z. A. Khan, T. Hussain, I. U. Haq, F. U. M. Ullah and S. W. Baik, *Towards efficient and effective renewable energy prediction via deep learning*. Energy Reports, 8, 10230-10243 (2022).
2. International Energy Agency (IEA), https://www.iea.org/reports/wind-electricity (accessed on 18-12-2022).
3. K. Kiranvishnu, K. Sireesha, K. and J. Ramprabhakar, *Comparative study of wind speed forecasting techniques*. In 2016 Biennial International Conference on Power and Energy Systems: Towards Sustainable Energy (PESTSE). IEEE, 1-6 (2016).
4. P. P. Prabha, V. Vanitha and R. Resmi, *Wind speed forecasting using long short-term memory networks*. In 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). IEEE, 1310-1314 (2019).
5. V. Vanitha, D. Raphel and R. Resmi, *Forecasting of wind power using variational mode decomposition-adaptive neuro fuzzy inference system*. In 2019 Innovations in Power and Advanced Computing Technologies (i-PACT). IEEE, 1-4 (2019).
6. B. Bhardwaj and M. Ganesan, *Machine learning in wind energy: generation to supply. in machine learning techniques for smart city applications: trends and solutions*. Advances in Science, Technology and Innovation, Springer, Cham., 143-153 (2022).
7. L. Wickramasinghe, P. Ekanayake and J. Jayasinghe, *Machine Learning and Statistical Techniques for Daily Wind Energy Prediction*. Gazi University Journal of Science, 35(4), 1359-1370 (2022).
8. A. N. Buturache and S. Stancu, *Wind energy prediction using machine learning*. LCE, 12, 1-21 (2021).
9. N. Shabbir, R. AhmadiAhangar, L. Kütt, M. N. Iqbal and A. Rosin, *Forecasting short term wind energy generation using machine learning*. In 2019 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON). IEEE, 1-4 (2019).
10. H. Demolli, A. S. Dokuz, A. Ecemis and M. Gokcek, *Wind power forecasting based on daily wind speed data using machine learning algorithms*. Energy Conversion and Management, 198, 111823 (2019).
11. A. Alkesaiberi, F. Harrou and Y. Sun, *Efficient wind power prediction using machine learning methods: A comparative study*. Energies, 15(7), 2327 (2022).
12. H. Zheng and Y. Wu, *A xgboost model with weather similarity analysis and feature engineering for short-term wind power forecasting*. Applied Sciences, 9(15), 3019 (2019).
13. B. Bochenek, J. Jurasz, A. Jaczewski, G. Stachura, P. Sekuła, T. Strzyżewski and M. Figurski, *Day-ahead wind power forecasting in Poland based on numerical weather prediction*. Energies, 14(8), 2164 (2021).
14. U. Singh, M. Rizwan, M. Alaraj and I. Alsaidan, *A machine learning-based gradient boosting regression approach for wind power production forecasting: a step towards smart grid environments*. Energies, 14(16), 5196 (2021).
15. U. Singh and M. Rizwan, *Enhancing wind power forecasting from meteorological parameters using machine learning models*. Journal of Renewable and Sustainable Energy, 14(6), 063302 (2022).
16. S. Diop, P. S. Traore and M. L. Ndiaye, *Wind Power Forecasting Using Machine Learning Algorithms*. In 2021 9th International Renewable and Sustainable Energy Conference (IRSEC), IEEE, 1-6 (2021).
17. Y. Chen and J. Xu, *Solar and wind power data from the Chinese State Grid Renewable Energy Generation Forecasting Competition*. Scientific Data, 9(1), 1-12 (2022).
18. D. S. Sayad, *Decision-Tree Regression*. https://www.saedsayad.com/decision_tree_reg.htm
19. G. Iannace, G. Ciaburro and A. Trematerra, *Wind turbine noise prediction using random forest regression*. Machines, 7(4), 69 (2019).
20. D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang and Y. Si, *A data-driven design for fault detection of wind turbines using random forests and XGboost*. IEEE Access, 6, 21020-21031 (2018).
21. L. Cai, J. Gu, J. Ma and Z. Jin, *Probabilistic wind power forecasting approach via instance-based transfer learning embedded gradient boosting decision trees*. Energies, 12(1), 159 (2019).
22. J. Kim, A. Afzal, H. G. Kim, C. T. Dinh and S. G. Park, *Wind power forecasting based on hourly wind speed data in South Korea using machine learning algorithms*. Journal of Mechanical Science and Technology, 36(12), 6107-6113 (2022).