Long-Term Wind Power Forecasting Using Tree-Based Learning Algorithms

AMIRHOSSEIN AHMADI[1], (Member, IEEE), **MOJTABA NABIPOUR**[2],
BEHNAM MOHAMMADI-IVATLOO[3,4], (Senior Member, IEEE),
ALI MORADI AMANI[5], (Member, IEEE), **SEUNGMIN RHO**[6], (Member, IEEE),
AND MD. JALIL PIRAN[7], (Member, IEEE)

[1]Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran 1344893854, Iran
[2]Department of Mechanical Engineering, Tarbiat Modares University, Tehran 4816718986, Iran
[3]Department of Electrical Engineering, University of Tabriz, Tabriz 5166616471, Iran
[4]Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
[5]School of Engineering, Royal Melbourne Institute of Technology, Melbourne, VIC 3001, Australia
[6]Department of Software, Sejong University, Seoul 05006, South Korea
[7]Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea

Corresponding authors: Behnam Mohammadi-Ivatloo (bmohammadi@tabrizu.ac.ir) and Md. Jalil Piran (piran@sejong.ac.kr)

ABSTRACT The intermittent and uncertain nature of wind places a premium on accurate wind power forecasting for the reliable and efficient operation of power grids with large-scale wind power penetration. Herein, six-month-ahead wind power forecasting models were developed using tree-based learning algorithms. Three models were developed to investigate the impact of input data on forecasting accuracy. The first model was trained with the average and standard deviation of wind speed values measured at a height of 40 m with a 10-min sampling time. To evaluate the impact of sampling time on model performance, a second model was trained with wind speed values measured at a height of 40 m with 1-h, 12-h, and 24-h sampling times. To assess the effect of measuring height on model accuracy, the third model was trained with wind speed values measured at 40 m extrapolated from values measured at heights of 30 m and 10 m. Experiments revealed that using longer time intervals and height extrapolation leads to considerable accuracy degradation in forecasted models. Finally, to study the generalization ability of the forecasted models, they were tested against wind data measured at heights and locations different from what the models had been trained with. Simulation results substantiated that tree-based learning algorithms can be successfully adopted not only for long-term wind power forecasting, but for potential wind power forecasting at different heights and geographical locations.

INDEX TERMS Wind energy, long-term, wind power forecasting, machine learning, regression.

I. INTRODUCTION

The dramatic growth of the human population accompanied by the advent and advance of modern harvesting technologies has led to more intense exploitation of fossil fuel resources. This has resulted in gradual fossil fuel depletion and increased pollution density. These issues, besides the poor energy efficiency of conventional power systems, have motivated a new trend of generating power using renewable energies [1]. Wind energy is one of the most promising renewable energies experiencing an unprecedented proliferation

The associate editor coordinating the review of this manuscript and approving it for publication was B. Chitti Babu.

in modern power grids worldwide [2]. For all advantages, wind energy is contingent upon the highly variable, both geographically and temporally, nature of wind [3]. Hence, it carries a great deal of uncertainty, which makes wind power integration into the power grid a complicated task [4], [5]. Employing reserve power plants is a practical but expensive solution to compensate for the fluctuations of wind power [6]. Wind power forecasting (WPF) is a viable alternative solution for more reliability and, therefore, penetration of wind power into the power grid [7]–[11].

The WPF methods can be classified into either deterministic (also called point forecasting), which provide one output for a specific time horizon, or probabilistic (also called

interval forecasting), which provide a range of possible values at a given time [12]. Comprehensive reviews of state-of-the-art probabilistic WPF methods are presented in [13], [14]. The deterministic WPF can be categorized based on input data, forecast output, time-scale, and the forecasting method [15]. Generally, input data are historical weather records such as wind speed and direction, pressure, temperature, humidity, and the radiation measured at various time intervals and at different heights. In cases that the required data at a wind turbine height are not available, due to the lower height of anemometer towers of meteorological stations, wind data measured at lower heights can be extrapolated to the turbine height [16]. The WPF can be realized either directly or indirectly [17]. The former is to directly forecast wind power based on the data collected from SCADA systems, while the latter is to forecast wind speed first and then convert it into wind power forecast using power curves.

Although there is no standard definition for time-scale among scholars, a generally accepted definition can be very-short-term forecasting (few seconds to 30 min), used for wind turbine regulation and control strategies, electricity market clearing, and real-time grid operation; short-term forecasting (30 min to 6 hours) used for economic load dispatch planning, operational security in the electric market, and load decisions for increments; medium-term forecasting (6 hours to 1 day) used for decision-making of unit commitment, reserved requirement, and generator operation; and long-term forecasting (more than 1 day ahead) used to schedule maintenance and determine the long-term feasibility of the wind farm [18].

Deterministic WPF methods can be classified into five categories of persistence, physical, statistical, machine learning, and hybrid methods [19]. Persistence methods are simple yet accurate methods utilized as a benchmark among wind forecasting methods [20]. Physical methods are generally based on numerical weather prediction models and ordinarily perform well for long-term time horizons [21]. Statistical methods employ historical data to predict wind behavior and usually perform well for short-term time horizons [22]. Machine learning is a subdivision of statistical methods that can learn patterns from data and forecast accordingly [23]. Finally, hybrid methods attempt to provide better predictions by combining various forecasting methods using weighting, preprocessing or decomposition, feature selection or optimization, and postprocessing techniques [24], [25]. A review of hybrid wind forecasting is performed by Xiao *et al.* [26].

Over the past years, several studies have been conducted on the WPF using different forecasting methods and on various horizons. A comprehensive review of existing research and current developments in this area can be found in [18]. In recent years, the machine learning method has been extensively adopted (approximately 38% of the literature) to generate a more accurate forecast among which 48% is dedicated to very-short-term, 36% to short-term, 3% to the medium-term, and 13% to long-term [18]. Rodriguez *et al.* [27] addressed very-short-term WPF based on artificial neural networks and the records of wind power in the last 24 h. They investigated the correlation between weather variables and wind power in order to appropriately choose input variables. Jiajun *et al.* [28] proposed ultra-short-term wind prediction with wavelet transform, deep belief network and ensemble learning. They designed several case studies in order to explore the promotion of high dimensional feature extraction. Li *et al.* [29] presented short-term WPF based on the support vector machine with an improved dragonfly algorithm. The adaptive learning factor and differential evolution strategy were presented to enhance the performance of the traditional dragonfly algorithm and, thus, choose the optimal parameters of the support vector machine. Sun *et al.* [30] introduced short-term WPF by a synthetical similar time series data mining method. The clustering similar measure function was supposed to identify the similar wind speed days that are close in space distance and have a similar variance trend synthetically. Han *et al.* [31] proposed short-term WPF using an improved long short term memory network, where the variational mode decomposition technique was utilized to decompose the wind power signal. Abhinav *et al.* [32] presented short-term WPF based on a wavelet-based neural network, which was applicable to all seasons of the year. Chen and Liu [33] investigated medium-term WPF based on a multi-resolution multilearner ensemble and adaptive model selection. The method employed heterogeneous-based learners and datasets with different resolutions to guarantee diversity.

Results of another review revealed that most of the literature is dedicated to very-short-term and short-term WPF ($\approx 87.22\%$) and there are limited works on long-term WPF ($\approx 8.36\%$) [34]. Kanna and Singh studied [35] long-term WPF using an adaptive wavelet neural network. Lledo *et al.* [36] explicated seasonal forecasting of wind power generation using manufacturer-provided power curves. Alencar *et al.* [37] explored different models for WPF using artificial neural network models, autoregressive integrated moving average, and hybrid models, including forecasting using wavelets. Demolli *et al.* [34] investigated machine learning algorithms application' for indirectly forecasting a year-ahead wind power based on the daily mean wind speed and standard deviation measured at the height of 10 m extrapolated to the height of 50 m. Barbounis *et al.* [38] proposed 72-h ahead indirect WPF using meteorological data of four nearby locations and three types of local recurrent neural network models. Kusiak *et al.* [39] discussed models for 84-h ahead direct prediction of wind farm power using five different data mining algorithms and 10-min average data measured at the height of 60 m. Khan *et al.* [40] presented a new hybrid approach for a year-ahead direct WPF of hourly spaced wind turbine data using deep learning with a tensor flow framework and principal component analysis. Dumitru and Gligor [41] introduced long-term indirect WPF using feedforward artificial neural networks. Yan and Ouyang [8] developed a two-phase hybrid approach for three-month-ahead indirect WPF based on physical and

data mining methods and the sampling time of 15 min data. Khosravi *et al.* [42] explored a multilayer feed-forward neural network, support vector regression, and adaptive neuro-fuzzy inference system for 24-h ahead indirect WPF using data measured at 5-min, 10-min, 30-min, and 1-h intervals. Han *et al.* [43] proposed multi-step direct WPF based on a variational mode decomposition-long short-term memory. Finally, Hong and Rioflorido [44] introduced a hybrid deep learning neural network for 24-h direct WPF based on an hourly spaced wind power dataset.

It can be concluded from the literature that there exist limited works investigating long-term WPF using machine learning algorithms. Besides, the existing works neglect to consider the impact of input data on the performance of WPF. Overfitting, underfitting, and generalization are significant issues in machine learning that should be considered in the algorithm selection procedure. However, most of the published works split the dataset only based on time series, leading to totally dissimilar training and test sets with different characteristics, thus posing the overfitting problem on developed models. To deal with these problems, this article investigated machine learning-based six-month-ahead WPF based on wind records. Tree-based algorithms, including decision tree, bagging, random forest, boosting (AdaBoost), gradient boosting, and XGBoost, were selected to train accurate WPF models with generalization abilities. This selection was motivated by the fact that they benefit from various regularization (such as pruning) and ensembling (such as bagging and boosting) techniques, which empower them to better deal with overfitting and underfitting. The K-folds cross-validator is also adopted to split the initial dataset into several train-test subsets.

Clearly, the quantity and quality of the dataset have a profound impact on the performance of the WPF model. Indeed, the dataset should be a representative sample of the region's wind speed characteristics. Hence, the developed models algorithms were trained with the mean and standard deviation of wind speed and power values. The results corroborated that tree-based learning algorithms can be effectively used for long-term WPF. Experiments revealed that the height and time interval of wind speed records have profound impacts on WPF accuracy. Therefore, the forecasting accuracy of the proposed models was investigated using observations measured at various heights and time intervals. Besides, the possibility and effectiveness of applying the proposed models to a location different from the model-trained location were explored here. Experimental results demonstrated the generalization ability of the developed WPF models to predict the potential wind power at different heights and geographical locations even before wind turbines are installed.

The rest of this article is organized as follows: Section II reviews the background of the adopted algorithms and the performance indices. Data analysis, illustrative case studies, and simulation results are presented in Section III. Finally, conclusions are drawn in Section IV.

II. PRELIMINARIES
A. MACHINE LEARNING
Machine learning is a branch of artificial intelligence that can learn from datasets, identify patterns and forecast outcomes or behavior. Generally, learning models can be classified into four categories: supervised, unsupervised, semi-supervised and reinforcement learning. Supervised learning is the most common category of machine learning that can be further classified into classification and regression.

Tree-based methods are deemed to be one of the best and powerful supervised learning subsets used for classification and regression [45]. These algorithms employ predictive models with rapid performance, high accuracy, and easy interpretation. In contrast with linear models, the methods map nonlinear relationships perfectly and are capable of adapting to various kinds of problems in the machine learning area. Tree-based models include decision tree, bagging, random forest, boosting, gradient boosting and XGBoost, which are elaborated upon as follows.

1) DECISION TREE
Decision tree refers to tree-based algorithms that handle multi-output problems with little data preparation. It is like a map to find the probable consequences of a series of related choices. The goal is to make a model that predicts a target value by learning easy decision rules formed from the data features. The method usually starts with a single node, which branches into probable consequences. Next, each of those outcome nodes leads to extra ones, which are connected to other possibilities. This process has a tree shape to finally obtain the final result.

2) BAGGING
Bagging is an ensemble estimator that provides a better understanding of the bias and variance of the dataset and includes a random sampling of a small subset of the main dataset. It is necessary to note that the subset can be replaced, and the selection process of all samples in the dataset has an equal chance. In this method, samples are picked with replacement, and predictions are achieved through a majority voting mechanism. It can help us better realize the standard deviation and mean related to the main dataset.

3) RANDOM FOREST
Random forest is a flexible machine learning method that is made by a large number of decision trees. It is an ensemble learner for classification and regression by considering three steps, randomly selecting training data when making trees, choosing some subsets of features when splitting nodes, and employing only a subset of all features for splitting each node in each simple decision tree. During the training of the data, each tree learns from a random sample of the data points.

4) BOOSTING
Boosting method converts weak learners into powerful learners. In this method, each new tree fits on a modified version of the main dataset. The AdaBoost is a well-known subset

of boosting algorithms and begins by training a decision tree in which each sample is assigned an equal weight. Next, weights of the samples are changed after examining the first tree-based on correcting their past performance. During the process, these weights are adapted based on the current prediction error, so subsequent models focus more on difficult items.

5) GRADIENT BOOSTING
Gradient boosting method is like AdaBoost when it creates a predictor in the form of an ensemble of weak prediction models, normally decision trees. In contrast to AdaBoost, Gradient Boosting fits a new predictor to the residual errors by using gradient descent to find the failure in the predictions of the prior learner. Overall, the final model is capable of use as the base model to decreases errors over time.

6) XGBoost
XGBoost is an ensemble tree method like gradient boosting and is designed for better speed and performance. The method applies the principle of boosting for weak learners. The power of this method lies in its fast learning through parallel and distributed computing and also offering efficient memory usage. In-built cross-validation ability, efficient handling of missing data, regularization for avoiding overfitting, catch awareness, tree pruning and parallelized tree building are common advantages of the XGBoost method.

B. PERFORMANCE INDICES
The effectiveness of WPF methods can be assessed using various performance indices measuring accuracy, stability or direction. Nevertheless, most of the studies have evaluated forecast results only in terms of following accuracy metrics.

1) MEAN ABSOLUTE ERROR (MAE)
MAE measures the mean absolute of the difference between the observation and prediction, without deeming the errors' direction, whose formula can be expressed as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{1}$$

where N is the number of samples, y_i is the observation and $\hat{y}_i$ is the prediction.

2) ROOT MEAN SQUARE ERROR (RMSE)
RMSE measures the square root of the mean of the difference between the observation and prediction, whose formula can be expressed as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{2}$$

It shows how widely spread the predictions are from observations. Due to the large amplitude of the wind power, the

normalized RMSE (nRMSE) can also be used, whose formula can be expressed as

$$\text{nRMSE} = \frac{1}{P_{inst}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{3}$$

where P_{inst} denotes the installed wind farm capacity.

3) COEFFICIENT OF DETERMINATION (R^2)
R^2 is a statistical measure of fit, which shows how closely predictions match observations. It measures the ratio of the residual sum of squares to the sum of squares of the total deviations, whose formula can be expressed as

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{4}$$

where $\bar{y}$ denotes the mean of observations. It has a value within [0, 1] where the closer the value to 1, the higher the prediction accuracy.

III. SIMULATION RESULTS
Tree-based learning algorithms were used for forecasting six-month-ahead wind power based on the mean wind speed and standard deviation. In the following, data analysis is presented to analyze datasets for gaining insights that are advantageous in making strategic decisions on WPF. Next, illustrative case studies and simulation results are presented and investigated in detail.

A. DATA ANALYSIS
Over the last decade, wind power generation in Iran has experienced a remarkable increase from 203 MW in 2010 to 302 MW in 2019. Analysis of the wind energy potential in 26 stations, with a 33% energy efficiency, shows that electrical energy of 6500 MW can be produced [42]. The northeast region of Iran is a potential area for wind power generation with constant wind flow without extreme gusts. Hence, two sites in this region were considered, as displayed in Fig. 1: Ghadamgah (36.104° north and 59.066° east longitude) and Khaf (34.567° north and 60.148° east longitude). These regions are recognized for possessing a notable wind potential capacity accompanied by good infrastructural conditions and electrical grid connection, so they were selected for wind farm settlement.

Here, the dataset included 18 months of wind speed practical values with a 10-min sampling time as the inputs of the WPF model. A total of 236556 and 238002 observations measured at 40 m, 30 m, and 10 m heights were collected from Khaf and Ghadamgah wind farms, respectively. Figure 2 demonstrates the variation of the wind speed for selected wind farms at various heights, whose statistical information is shown in Table 1. Because of the intermittent nature of wind, it exhibits various fluctuations and spikes in different areas, heights, and lengths of time.

Probability density functions, such as Rayleigh or Weibull, are simple yet practical methods frequently adopted to
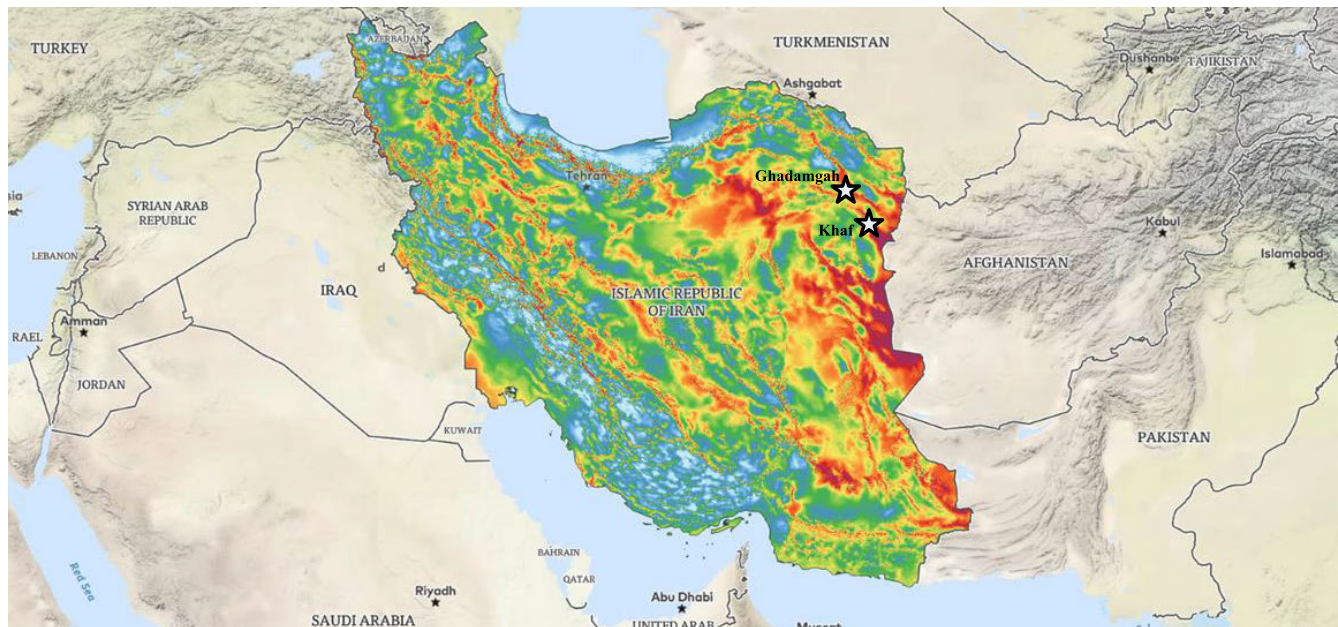
FIGURE 1. Locations of the meteorological stations within Iran wind map.
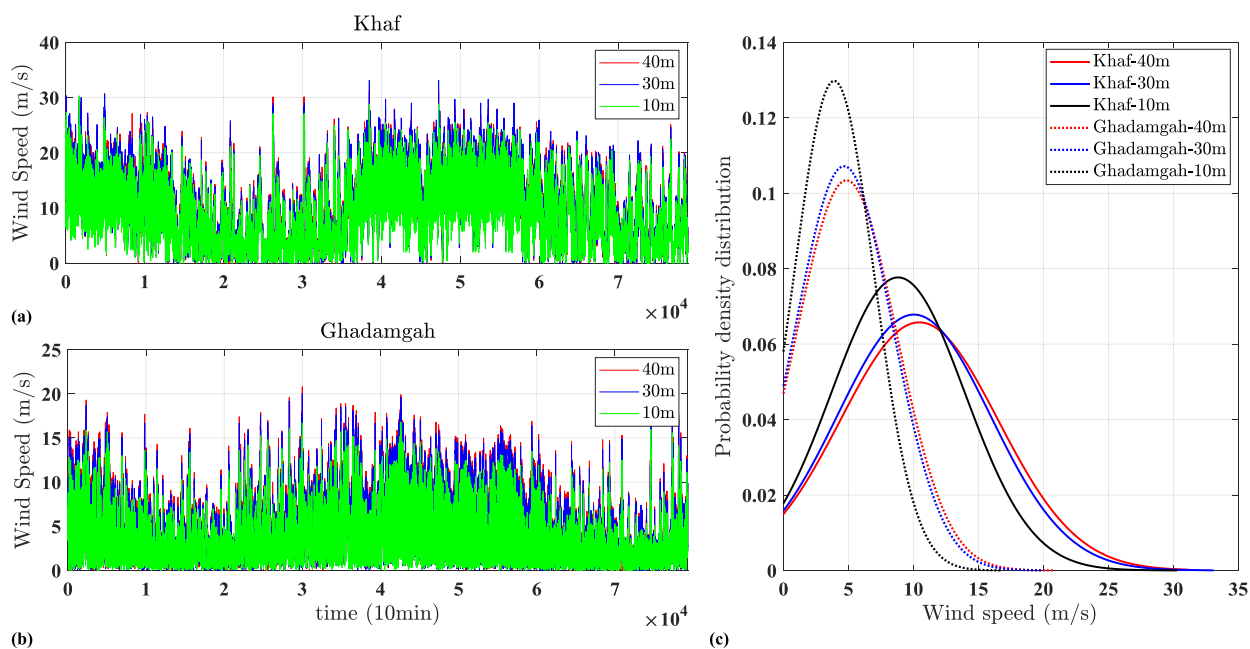

FIGURE 2. Variation of wind speed for selected wind farms.

describe the behavior of wind speed data in a period [46]. Figure 2c illustrates Weibull distributions for the selected locations at various measuring heights. Evidently, the peak of the curve for Ghadamgah is 0.128, whereas this value is 0.077 for Khaf. It means that the frequency of observing average levels of speed in Ghadamgah is higher than Khaf, whereas Khaf has higher wind speed values than Ghadamgah.

The power curve provides a suitable way to describe the correlation between the generated power and wind speed.

The power curve and technical specifications of 1 MW wind turbines installed in the Khaf wind farm are given in Fig. 3 and Table 2, respectively. This generally S-shaped curve contains three key speeds: cut-in speed (v_{ci}), which refers to the minimum speed at which the turbine delivers useful power; rated speed (v_r), which is the wind speed at which the generator produces maximum output power (P_r); and cut-out speed (v_{co}), which denotes the maximum wind speed at which the turbine is allowed to produce power. The power curve can

TABLE 1. Statistical information of four datasets.

Location	Time interval (10 min)	Samples	Train	Test	Height (m)	Min (m/s)	Mean (m/s)	Max (m/s)	Std
Khaf	08/07/2015 13:50-26/01/2017 10:50	78852	55197	23655	40	0	10.4516	31.2	6.06618
		78852	55197	23655	30	0	10.0382	33.1	5.88218
		78852	55197	23655	10	0	8.82163	30.3	5.13612
Ghadamgah	01/09/2015 11:40-05/03/2017 10:20	79334	55534	23800	40	0	4.85276	20.8	3.85746
		79334	55534	23800	30	0	4.66561	20	3.72295
		79334	55534	23800	10	0	3.90087	16.8	3.07386

TABLE 2. Technical specifications of the considered wind turbine.

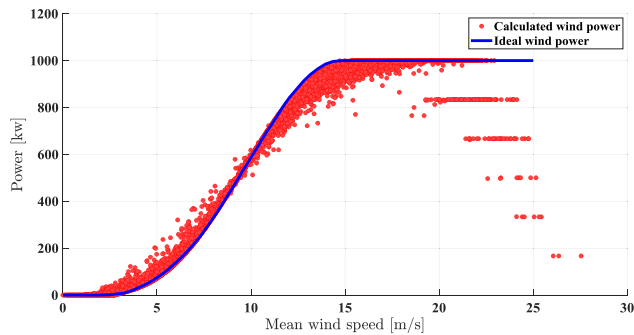| Characteristics	Values
Rated power (MW)	1
Hub height (m)	40
Rotor diameter (m)	54.2
Swept area (m^2)	2,300
Number of blades	3
Cut-in wind speed (v_{ci}) (m/s)	3
Rated wind speed (V_r) (m/s)	15
Cut-off wind speed (v_{co}) (m/s)	25

FIGURE 3. Power curve for a 1 MW wind turbine.

be modeled using

$$P_i(v) = \begin{cases} 0 & v < v_{ci} \\ \alpha_n v^n + \alpha_{n-1} v^{n-1} + \ldots + \alpha_0 & v_{ci} \le v < v_r \\ P_r & v_r \le v < v_{co} \\ 0 & v \ge v_{co} \end{cases} \quad (5)$$

where $\alpha_0, \ldots, \alpha_{n-1}$ and α_n are constants, and $P_i(v)$ is the corresponding output power. Accordingly, 1-h, 12-h, and 24-h wind power values can be computed by taking the average of output power values, as

$$P_t = \sum_{i=1}^{N} P_i(v) \Delta t = \frac{1}{N} \sum_{i=1}^{N} P_i(v) \quad (6)$$

where N denotes the number of samples in the considered time interval (1-h, 12-h, and 24-h), while Δt is the sampling time (10-min) [34]. For instance, Fig. 3 represents the obtained wind power values for 1-h interval.
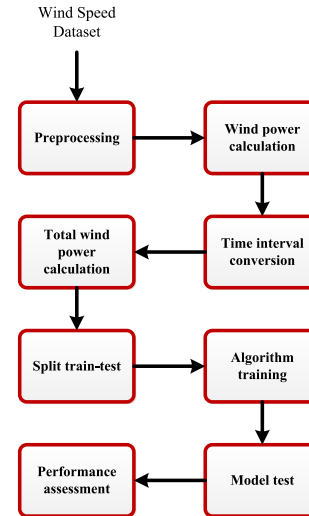

FIGURE 4. Flow chart of the proposed methodology.

B. METHODOLOGY

The proposed methodology is illustrated in Fig. 4 utilizing the trial-and-error obtained parameter settings listed in Table 3. First, data preprocessing was performed to clean the data and derive useful information. Next, the corresponding wind power was computed based on the wind power curve. Then, actual wind speed values with 10-min time intervals were converted into hourly mean wind speed values and standard deviations, and then, hourly total wind power was calculated by summing every 10-min sampled power values. Thereafter, the dataset was split into training and test sets using the K-folds cross-validator and based on the specified forecasting horizon. The training set was utilized to train the candidate algorithm and fit the model, and the test set was used to generate wind power values based on the WPF model. Finally, performance assessment was performed in terms of several well-known metrics.

Multiple case studies were designed to analyze the effectiveness of machine learning algorithms in long-term WPF and validate their generalization ability.

C. CASE 1

The first case was assigned to six-month-ahead WPF given 12 months of hourly mean wind speed values and standard deviation and their corresponding hourly total generated wind power values. The Ghadamgah dataset measured at

TABLE 3. Parameter settings for machine learning algorithms.

Algorithm	Parameters	Value
Decision tree	Max depth	10
Bagging	Max depth	10
	Estimator	Decision tree
	Number of trees	100
Random forest	Max depth	10
	Number of trees	100
AdaBoost	Max depth	10
	Estimator	Decision tree
	Number of trees	100
	Learning rate	0.1
	Loss function	Square
Gradient boosting	Max depth	10
	Number of trees	100
	Learning rate	0.1
XGBoost	Max depth	10
	Number of trees	100
	Learning rate	0.1
	Objective	Regressor: squared error

TABLE 4. Forecasting accuracy of proposed models for case study 1.

Algorithm	Time interval	MAE	RMSE	R^2
Decision tree	1-h	8.26	29.45	0.9996
	12-h	644.04	1304	0.9925
	24-h	1893.1	2984	0.9787
Bagging	1-h	7.24	25.15	0.9996
	12-h	482.69	1003	0.9950
	24-h	**1372.2**	2560	0.9854
Random forest	1-h	7.43	25.27	0.9996
	12-h	**479.67**	1026	0.9949
	24-h	1382.8	**2501**	0.9859
AdaBoost	1-h	8.43	25.5	0.9996
	12-h	557.25	1143	0.9934
	24-h	1571.4	2748	0.9843
Gradient boosting	1-h	6.34	25.7	0.9996
	12-h	613.98	1219	0.9929
	24-h	1792.9	3065	0.9804
XGBoost	1-h	**6.34**	**22.4**	**0.9997**
	12-h	484.73	**997**	**0.9952**
	24-h	1402.4	2546	**0.9867**

a height of 40 m with a 10-min time interval was converted into another dataset with an hourly time interval. Thereafter, in order to evaluate the effect of the time interval on the accuracy of the forecasted model, another two datasets with 12- and 24-hour time intervals were created based on the original dataset. Correspondingly, 1-h, 12-h, and 24-h total wind power values were computed. Therefore, three models were fitted using 70% of the data set for six-month-ahead WPF. Table 4 presents the calculated metrics for forecasted models. Evidently, all algorithms exhibited a powerful performance with a high coefficient of determination and small errors. However, XGBoost provided a more accurate model, while decision tree produced a less accurate model in terms of different performance indices. Moreover, performance indices revealed that longer time intervals lead to performance degradation due to the loss of representativity of the dataset. Nevertheless, all the forecasted models provided an acceptable performance against variations in the time interval, among which bagging and random forest showed better robustness evident in the lower variations of the corresponding metrics.

D. CASE 2
Although wind data at higher heights close to wind turbines' height are required to accurately forecast wind power, the height of anemometer towers of meteorological stations is usually 10 m. A conventional solution to this problem is extrapolating wind speed values measure at lower heights to desired heights applying [34]

$$v = v_0 \left(\frac{h}{h_0} \right)^{\alpha} \tag{7}$$

where v_0 is the wind speed observed at the anemometer height (h_0), v is the wind speed at the desired heights (h), and α is the power-law constant that depends on the surface roughness [16]. The second case aims to investigate the impact of extrapolation on forecasting models. In this regard, the Ghadamgah datasets, measured at a height of 10 m and 30 m with a 10-min time interval, were extrapolated to a height of 40 m by using (7) and $\alpha = 0.14$. The simulation results for various algorithms are listed in Table 5, which substantiate that the use of extrapolation causes performance degradation in the forecasted models. Nonetheless, the algorithms properly forecasted six-month-ahead wind power, among which XGBoost generated a model with higher accuracy and better robustness against data inaccuracies caused by extrapolation, whereas the decision tree produced the weakest performance.

Figure 5 demonstrates the WPF accuracy of machine learning algorithms using the Ghadamgah hourly dataset measured at the height of 40 m as the best case. As expected, the extrapolation slightly deteriorated the accuracy of WPF models, in which the more the difference between the measuring height and the desired height, the larger the forecasting error and the less the performance accuracy. However, all the algorithms yielded acceptable models for an entire gamut of wind speed values. Figure 6 displays hourly actual and forecasted power values of six-month-ahead, verifying that the algorithms can deal with the uncertain nature of inputs with high accuracy and without overfitting. Again, XGBoost yield a slightly superior model with lower MAE and RMSE and higher R^2, while the decision tree produced a slightly inferior model with higher MAE and RMSE and lower R^2.
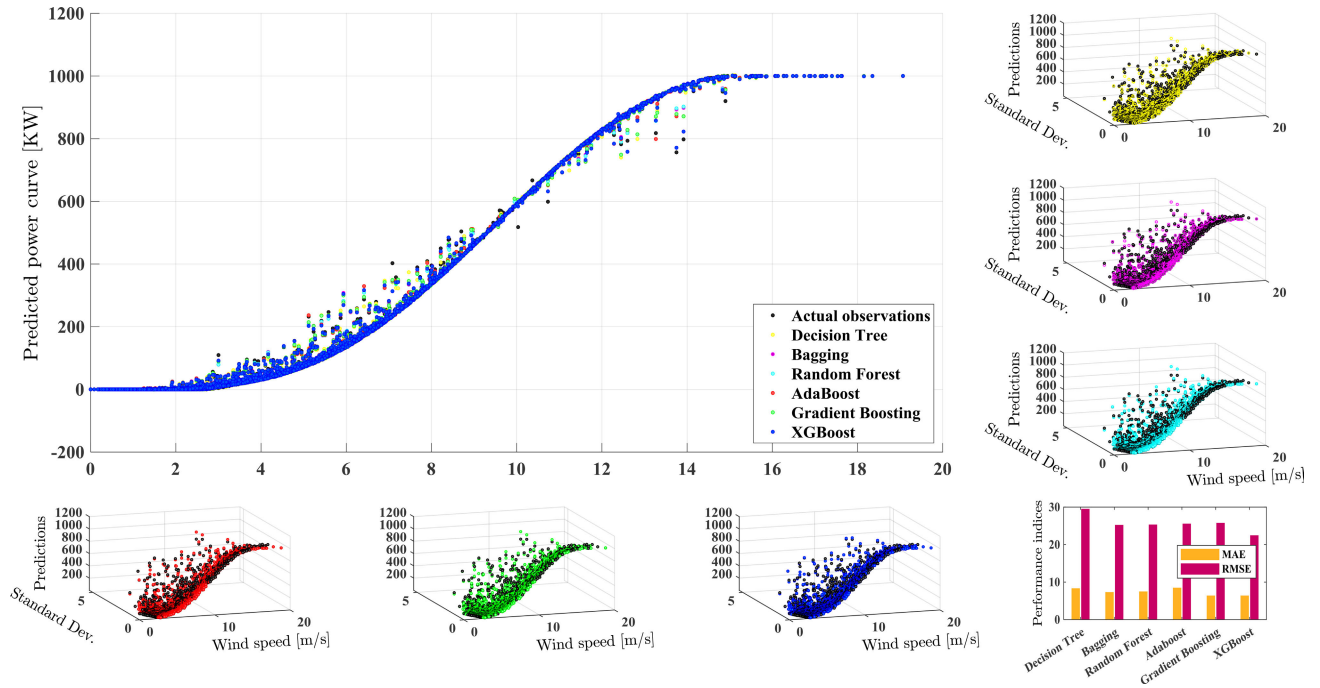
FIGURE 5. Forecasting performance of machine learning algorithms for case study 2.
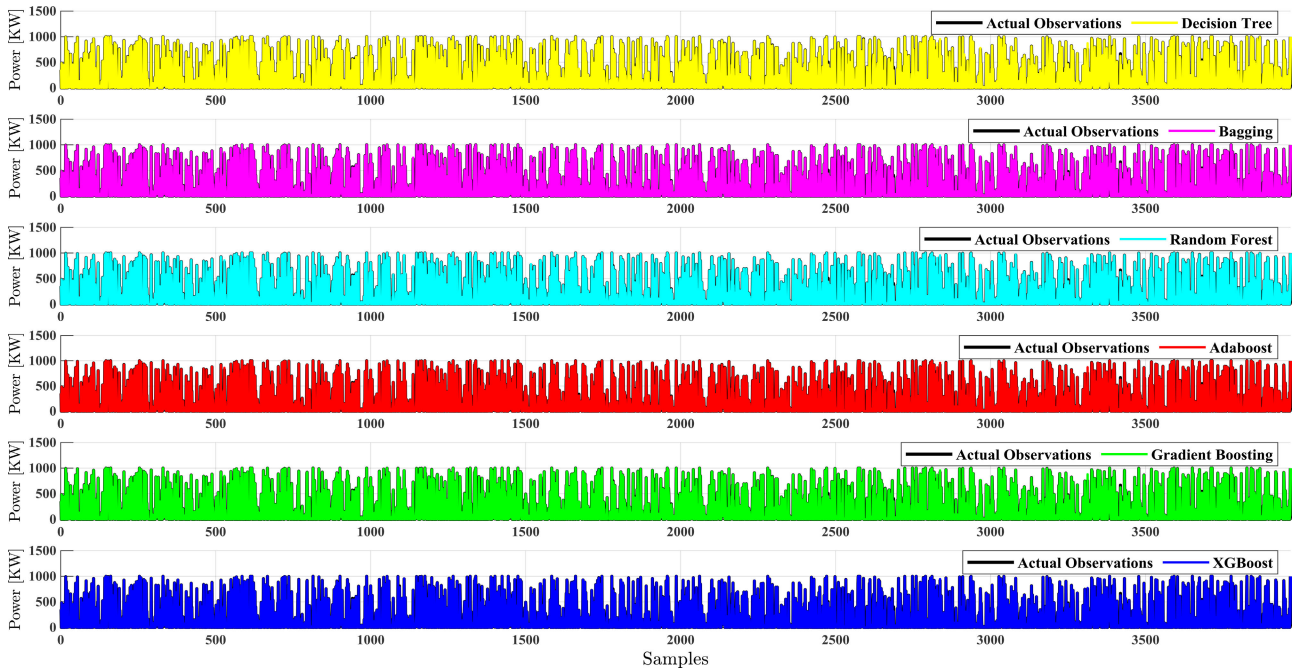

FIGURE 6. Hourly actual and forecasted power values of six-month-ahead for case study 2.

E. CASE 3

A generalizable model is a neither underfit nor overfit model aiming to make sensible predictions based on unseen validation datasets. Besides, knowing wind power at different heights is a prerequisite for having a more efficient wind farm by establishing wind turbines at various elevations.

Hence, this case explored the ability of the proposed models to forecast wind power at heights of 30 m and 10 m. Simulation results for various algorithms are summarized in Table 6, where the developed models manifest the generalization ability through predicting wind power at heights different from the model-trained height. For example, the model generated
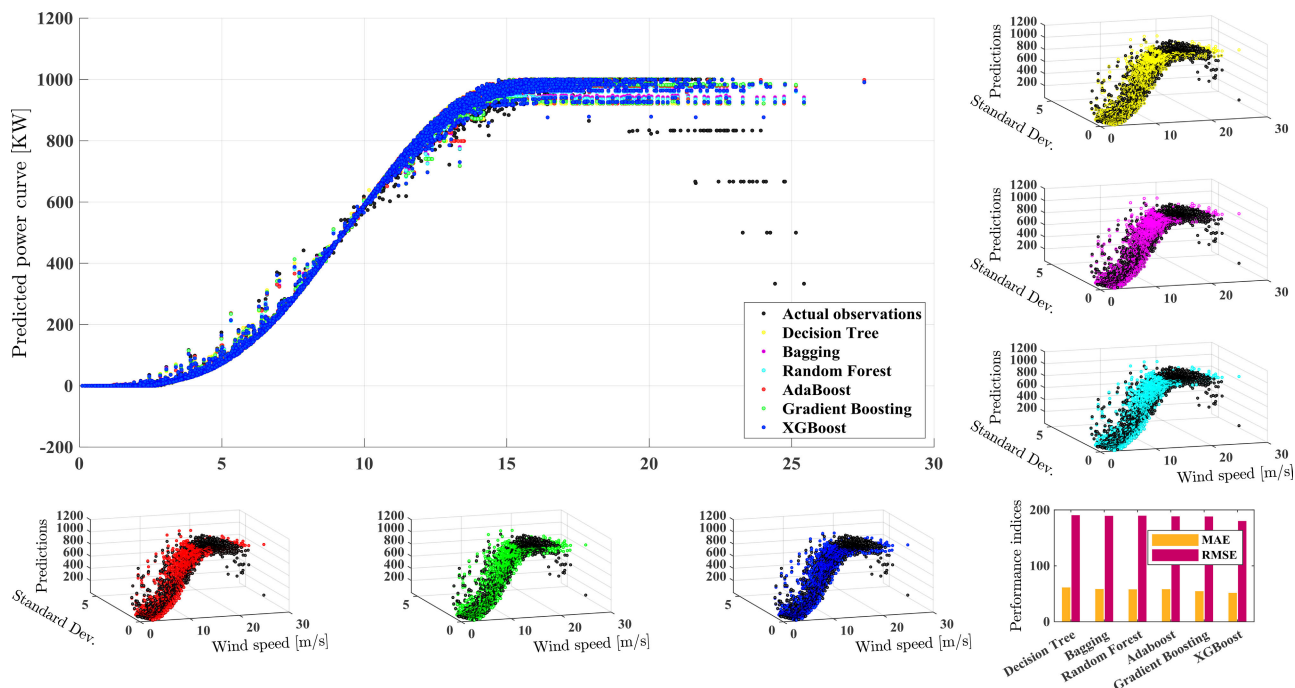
FIGURE 7. Forecasting performance of machine learning algorithms for case study 4.

TABLE 5. Forecasting accuracy of proposed models for case study 2.

Algorithm	Height	MAE	RMSE	R^2
Decision tree	40m actual	8.26	29.45	0.9996
	40m from 30m	174.12	305.9	0.9681
	40m from 10m	373.72	599.4	0.8778
Bagging	40m actual	7.24	25.15	0.9996
	40m from 30m	173.90	305.5	0.9682
	40m from 10m	373.39	598.3	0.8783
Random forest	40m actual	7.43	25.27	0.9996
	40m from 30m	173.89	305.5	0.9682
	40m from 10m	373.36	1892.1	0.8783
AdaBoost	40m actual	8.43	25.5	0.9996
	40m from 30m	174.07	305.6	0.9682
	40m from 10m	373.04	598.4	0.8783
Gradient boosting	40m actual	6.34	25.7	0.9996
	40m from 30m	174.01	305.7	0.9682
	40m from 10m	373.52	598.8	0.8781
XGBoost	40m actual	**6.34**	**22.4**	**0.9997**
	40m from 30m	**173.75**	**305**	**0.9683**
	40m from 10m	**372.62**	**597.2**	**0.8788**

TABLE 6. Forecasting accuracy of proposed models for case study 3.

Algorithm	Height	MAE	RMSE	R^2
Decision tree	40m	8.26	29.45	0.9996
	30m	38.22	65.47	0.9984
	10m	**292.15**	511.2	0.8334
Bagging	40m	7.24	25.15	0.9996
	30m	37.69	64.50	0.9984
	10m	292.2	510.8	0.8336
Random forest	40m	7.43	25.27	0.9996
	30m	37.6	64.31	0.9984
	10m	292.38	511	0.8335
AdaBoost	40m	8.43	25.5	0.9996
	30m	37.58	64.47	0.9984
	10m	292.34	511.3	0.8333
Gradient boosting	40m	6.34	25.7	0.9996
	30m	37.64	64.59	0.9984
	10m	292.3	510.9	0.8335
XGBoost	40m	**6.34**	**22.4**	**0.9997**
	30m	**37.28**	**63.25**	**0.9985**
	10m	292.3	**510.7**	**0.8337**

by the decision tree method has an MAE index of 8.26 at 40 m, 38.22 at 30 m, and 292.15 at 10 m for a 1MW wind turbine. This is where XGBoost possesses an MAE index of 6.34 at 40 m, 37.28 at 30 m, and 292.3 at 10 m for the same wind turbine. The results showed that considering MAE as an assessment criterion, XGBoost had a superior performance in the presence of dissimilar training and test sets with different characteristics, the same as the RMSE.

All the selected algorithms yield similar results for this case in terms of reported performance indices. This is due to their inherent features such as regularization and ensembling that enable them to successfully tackle variable and uncertain wind data with neither overfitting nor underfitting. Nevertheless, as expected, the more the difference between the training height and validation height, the more the forecasting error and the less the performance accuracy. This emphasizes the
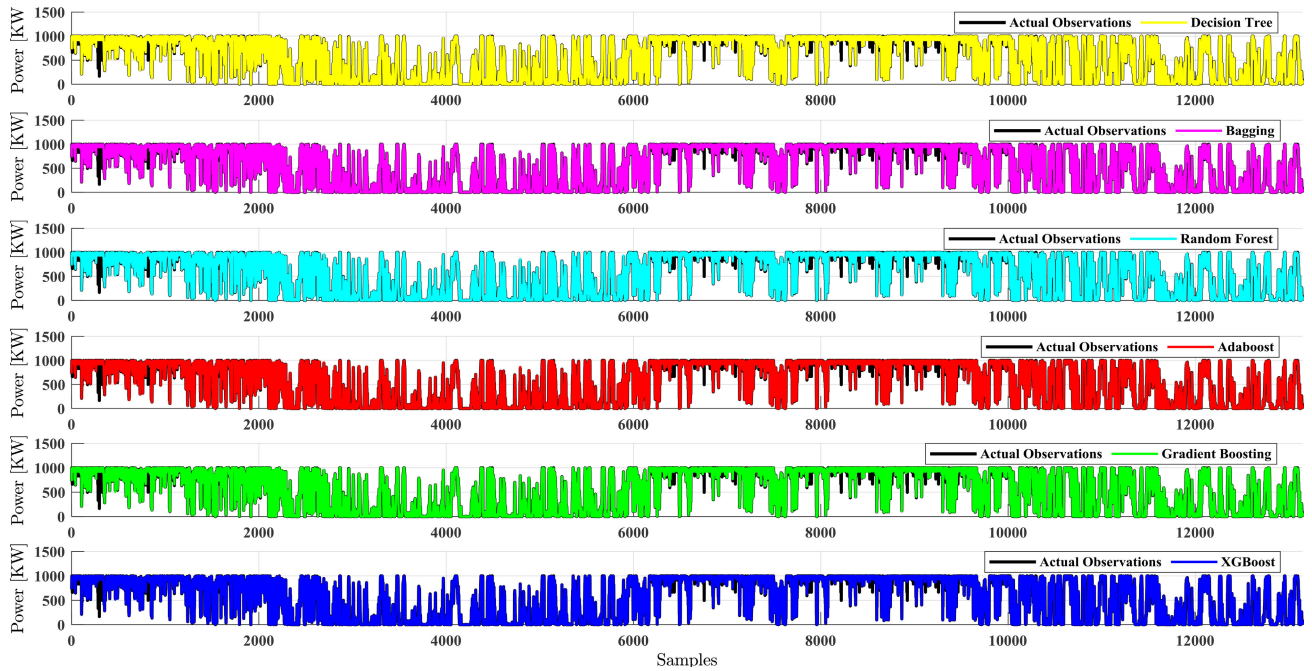
FIGURE 8. Hourly actual and forecasted power values of six-month-ahead for case study 4.

TABLE 7. Forecasting accuracy of proposed models for case study 4.

Algorithm	Location	MAE	RMSE	R^2
Decision Tree	Ghadamgah	8.26	29.45	0.9996
	Khaf	60.69	189.9	0.9938
Bagging	Ghadamgah	7.24	25.15	0.9996
	Khaf	58.02	188.7	0.9939
Random Forest	Ghadamgah	7.43	25.27	0.9996
	Khaf	57.31	188.9	0.9939
AdaBoost	Ghadamgah	8.43	25.5	0.9996
	Khaf	57.65	187.8	0.9939
Gradient Boosting	Ghadamgah	6.34	25.7	0.9996
	Khaf	54.05	187.4	0.9940
XGBoost	Ghadamgah	**6.34**	**22.4**	**0.9997**
	Khaf	**51.10**	**179.5**	**0.9945**

importance of data quality and its impact on the performance of developed models.

F. CASE 4

This case was designed to appraise the proposed models' predictions based on a new and previously unseen dataset. To do this, hourly wind measurements at a 40-m height in Khaf wind farm were applied to the trained models as test datasets. Figure 2 illustrates the variations of the new dataset besides its probability density function. Khaf and Ghadamgah, which belong to different geographical regions, possess different wind profiles, and Khaf exhibits high wind speeds with more fluctuations. Performance indices and experimental graphs are compared in Table 7 and Figs. 7 and 8, respectively. Overall, the results clearly show the reliability of previously

trained models to forecast wind power in Khaf. Based on Table 7, performance indices carry a slight deviation from previous metrics, while according to Figure 8, predictions represent observations very well. Figure 7 indicates that all the considered methods can predict the generated power with the accuracy of 5%-6% error in the case of MAE, and XGBoost has the best performance. This case revealed the intrinsic power of machine learning-based WPF models in making accurate predictions based on unseen datasets, which can be effectively adopted before the establishment of wind farms in another geographical region.

IV. CONCLUSION

This article proposed a comparative analysis for six-month-ahead WPF using tree-based learning algorithms. Overfitting and underfitting were considered in the algorithm selection and training procedure. Furthermore, datasets with sufficient samples were considered to bolster the prediction accuracy and generalization performance of the models. To increase the diversity of the training set and prevent overfitting, cross-validation was performed to split the initial dataset into several train-test subsets. Several models were trained with average and standard deviation values of the wind speed. The proposed models showed powerful performance in forecasting potential wind power in the Ghadamgah wind farm.

The intermittent and uncertain nature of wind poses serious problems for collecting sufficient representative datasets, which can degrade the prediction accuracy to a great extent and make it prone to overfitting. Hence, multiple case studies were defined to investigate the impact of the training dataset on forecasting accuracy. The simulation results revealed that using longer time intervals and height extrapolation leads to

considerable accuracy degradation in the forecasted models, where height extrapolation engenders more adverse effects in the forecasting accuracy of developed models. Indeed, using the standard deviation of the wind speed in the training procedure preserves the representativity of the dataset and reduces the adverse effects of increasing time intervals. For all negative effects of increasing the time interval and decreasing the measuring height on the performance of WPF models, simulation results demonstrate the robustness of the proposed models against data uncertainties.

To reinforce machine learning in industrial WPF applications, the models were proved on an unseen dataset collected from different heights, which is useful in choosing various wind turbine heights to be installed in wind farms. Finally, the generalizability of the models was confirmed in dealing with unseen datasets collected from Khaf wind farm. Experiments revealed the reliability of the models in predicting the potential wind power at different locations and heights.

ACKNOWLEDGMENT

REFERENCES

[1] S. P. Chowdhury, P. Crossley, and S. Chowdhury, "Distributed generation and Microgrid concept," in *Microgrids and Active Distribution Networks*, Stevenage, U.K.: IET, 2009, ch. 1, sec. 1, pp. 1–13.

[2] Y. Zhang, H. Sun, and Y. Guo, "Wind power prediction based on PSO-SVR and grey combination model," *IEEE Access*, vol. 7, pp. 136254–136267, 2019.

[3] D. B. Alencar, C. M. Affonso, R. C. L. Oliveira, and J. C. R. Filho, "Hybrid approach combining SARIMA and neural networks for multi-step ahead wind speed forecasting in Brazil," *IEEE Access*, vol. 6, pp. 55986–55994, 2018.

[4] S. Pei, H. Qin, Z. Zhang, L. Yao, Y. Wang, C. Wang, Y. Liu, Z. Jiang, J. Zhou, and T. Yi, "Wind speed prediction method based on empirical wavelet transform and new cell update long short-term memory network," *Energy Convers. Manage.*, vol. 196, pp. 779–792, Sep. 2019.

[5] A. Khosravi, L. Machado, and R. O. Nunes, "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil," *Appl. Energy*, vol. 224, pp. 550–566, Aug. 2018.

[6] M. E. Dittner and A. Vasel, "Advances in wind power forecasting," in *Advances in Sustainable Energy*, vol. 703, 1st ed. New York, NY, USA: Springer, 2019, ch. 3, sec. 2, pp. 37–57.

[7] X. Yang, Y. Yang, Y. Liu, and Z. Deng, "A reliability assessment approach for electric power systems considering wind power uncertainty," *IEEE Access*, vol. 8, pp. 12467–12478, 2020.

[8] J. Yan and T. Ouyang, "Advanced wind power prediction based on data-driven error correction," *Energy Convers. Manage.*, vol. 180, pp. 302–311, Jan. 2019.

[9] C. Wang, H. Zhang, and P. Ma, "Wind power forecasting based on singular spectrum analysis and a new hybrid Laguerre neural network," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114139.

[10] G. Wang, R. Jia, J. Liu, and H. Zhang, "A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning," *Renew. Energy*, vol. 145, pp. 2426–2434, Jan. 2020.

[11] Z. Niu, Z. Yu, W. Tang, Q. Wu, and M. Reformat, "Wind power forecasting using attention-based gated recurrent unit network," *Energy*, vol. 196, Apr. 2020, Art. no. 117081.

[12] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, Apr. 2014.

[13] J. Yan, Y. Liu, S. Han, Y. Wang, and S. Feng, "Reviews on uncertainty analysis of wind power forecasting," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 1322–1330, Dec. 2015.

[14] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, Jan. 2012.

[15] H. Liu, C. Chen, X. Lv, X. Wu, and M. Liu, "Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods," *Energy Convers. Manage.*, vol. 195, pp. 328–345, Sep. 2019.

[16] M. Gökçek, A. Bayülken, and Ş. Bekdemir, "Investigation of wind characteristics and wind energy potential in Kirklareli, Turkey," *Renew. Energy*, vol. 32, no. 10, pp. 1739–1752, Aug. 2007.

[17] Y. Ding, "A single time series model," in *Data Science for Wind Energy*, Boca Raton, FL, USA: CRC Press, 2019, ch. 2, sec. 1, pp. 17–55.

[18] M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Current perspective on the accuracy of deterministic wind speed and power forecasting," *IEEE Access*, vol. 7, pp. 159547–159564, 2019.

[19] Y. Wang, Q. Hu, D. Meng, and P. Zhu, "Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model," *Appl. Energy*, vol. 208, pp. 1097–1112, Dec. 2017.

[20] X. Zhao, S. Wang, and T. Li, "Review of evaluation criteria and main methods of wind power forecasting," *Energy Procedia*, vol. 12, pp. 761–769, Jan. 2011.

[21] J. Zhao, Z.-H. Guo, Z.-Y. Su, Z.-Y. Zhao, X. Xiao, and F. Liu, "An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed," *Appl. Energy*, vol. 162, pp. 808–826, Jan. 2016.

[22] E. López, C. Valle, H. Allende, E. Gil, and H. Madsen, "Wind power forecasting based on echo state networks and long short-term memory," *Energies*, vol. 11, no. 3, p. 526, Feb. 2018.

[23] H. B. Azad, S. Mekhilef, and V. G. Ganapathy, "Long-term wind speed forecasting and general pattern recognition using neural networks," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 546–553, Apr. 2014.

[24] R. Rahmani, R. Yusof, M. Seyedmahmoudian, and S. Mekhilef, "Hybrid technique of ant colony and particle swarm optimization for short term wind energy forecasting," *J. Wind Eng. Ind. Aerodynamics*, vol. 123, pp. 163–170, Dec. 2013.

[25] N. Ghadimi, A. Akbarimajd, H. Shayeghi, and O. Abedinia, "Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting," *Energy*, vol. 161, pp. 130–142, Oct. 2018.

[26] L. Xiao, J. Wang, Y. Dong, and J. Wu, "Combined forecasting models for wind energy forecasting: A case study in China," *Renew. Sustain. Energy Rev.*, vol. 44, pp. 271–288, Apr. 2015.

[27] F. Rodríguez, A. M. Florez-Tapia, L. Fontán, and A. Galarza, "Very short-term wind power density forecasting through artificial neural networks for microgrid control," *Renew. Energy*, vol. 145, pp. 1517–1527, Jan. 2020.

[28] H. Jiajun, Y. Chuanjin, L. Yongle, and X. Huoyue, "Ultra-short term wind prediction with wavelet transform, deep belief network and ensemble learning," *Energy Convers. Manage.*, vol. 205, Feb. 2020, Art. no. 112418.

[29] L.-L. Li, X. Zhao, M.-L. Tseng, and R. R. Tan, "Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm," *J. Cleaner Prod.*, vol. 242, Jan. 2020, Art. no. 118447.

[30] G. Sun, C. Jiang, P. Cheng, Y. Liu, X. Wang, Y. Fu, and Y. He, "Short-term wind power forecasts by a synthetical similar time series data mining method," *Renew. Energy*, vol. 115, pp. 575–584, Jan. 2018.

[31] L. Han, H. Jing, R. Zhang, and Z. Gao, "Wind power forecast based on improved long short term memory network," *Energy*, vol. 189, Dec. 2019, Art. no. 116300.

[32] R. Abhinav, N. M. Pindoriya, J. Wu, and C. Long, "Short-term wind power forecasting using wavelet-based neural network," *Energy Procedia*, vol. 142, pp. 455–460, Dec. 2017.

[33] C. Chen and H. Liu, "Medium-term wind power forecasting based on multi-resolution multi-learner ensemble and adaptive model selection," *Energy Convers. Manage.*, vol. 206, Feb. 2020, Art. no. 112492.

[34] H. Demolli, A. S. Dokuz, A. Ecemis, and M. Gokcek, "Wind power forecasting based on daily wind speed data using machine learning algorithms," *Energy Convers. Manage.*, vol. 198, Oct. 2019, Art. no. 111823.

[35] B. Kanna and S. N. Singh, "Long term wind power forecast using adaptive wavelet neural network," in *Proc. IEEE Uttar Pradesh Sect. Int. Conf. Electr., Comput. Electron. Eng. (UPCON)*, Varanasi, India, Dec. 2016, pp. 671–676.

[36] L. Lledó, V. Torralba, A. Soret, J. Ramon, and F. J. Doblas-Reyes, "Seasonal forecasts of wind power generation," *Renew. Energy*, vol. 143, pp. 91–100, Dec. 2019.

[37] D. B. de Alencar, C. de Mattos Affonso, R. L. de Oliveira, J. M. Rodríguez, J. Leite, and J. Reston Filho, "Different models for forecasting wind power generation: Case study," *Energies*, vol. 10, no. 12, p. 1976, Nov. 2017.

[38] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Trans. Energy Convers.*, vol. 21, no. 1, pp. 273–284, Mar. 2006.

[39] A. Kusiak, H. Zheng, and Z. Song, "Wind farm power prediction: A data-mining approach," *Wind Energy, Int. J. Prog. Appl. Wind Power Convers. Technol.*, vol. 12, no. 3, pp. 275–293, Apr. 2009.

[40] G. M. Khan, J. Ali, and S. A. Mahmud, "Wind power forecasting—An application of machine learning in renewable energy," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 1130–1137.

[41] C.-D. Dumitru and A. Gligor, "Daily average wind energy forecasting using artificial neural networks," *Procedia Eng.*, vol. 181, pp. 829–836, Jan. 2017.

[42] A. Khosravi, R. N. N. Koury, L. Machado, and J. J. G. Pabon, "Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system," *Sustain. Energy Technol. Assessments*, vol. 25, pp. 146–160, Feb. 2018.

[43] L. Han, R. Zhang, X. Wang, A. Bao, and H. Jing, "Multi-step wind power forecast based on VMD-LSTM," *IET Renew. Power Gener.*, vol. 13, no. 10, pp. 1690–1700, Jul. 2019.

[44] Y.-Y. Hong and C. L. P. P. Rioflorido, "A hybrid deep learning-based neural network for 24-h ahead wind power forecasting," *Appl. Energy*, vol. 250, pp. 530–539, Sep. 2019.

[45] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. Shahab, "Deep learning for stock market prediction," *Entropy*, vol. 22, no. 8, p. 840, Jul. 2020.

[46] A. Dolatabadi, B. Mohammadi-ivatloo, M. Abapour, and S. Tohidi, "Optimal stochastic design of wind integrated energy hub," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2379–2388, Oct. 2017.

ALI MORADI AMANI (Member, IEEE) received the B.Sc. degree in electrical engineering-control systems from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2000, and the M.Sc. degree from the University of Tehran, Tehran, in 2003. He is currently a Research Assistant with the School of Engineering, Royal Melbourne Institute of Technology. His main research interests include control of large scale and networked systems and fault-tolerant control.

SEUNGMIN RHO (Member, IEEE) received the B.S. degree in computer science and the M.S. and Ph.D. degrees in information and communication technology from the Graduate School of Information and Communication, Ajou University, South Korea, in 2001, 2003, and 2008, respectively. Before he joined the Computer Sciences Department, Ajou University, he spent two years in the industry. He visited the Multimedia Systems and Networking Laboratory, The University of Texas at Dallas, from 2003 to 2004. From 2008 to 2009, he was a Postdoctoral Research Fellow with the Computer Music Laboratory, School of Computer Science, Carnegie Mellon University. From 2009 to 2011, he was a Research Professor with the School of Electrical Engineering, Korea University. In 2012, he was an Assistant Professor with the Division of Information and Communication, Baekseok University. From 2013 to 2018, he was an Assistant Professor with the Department of Media Software, Sungkyul University. He is currently a Faculty Member with the Department of Software, Sejong University, South Korea. His current research interests include database, big data analysis, music retrieval, multimedia systems, machine learning, knowledge management, and computational intelligence.

AMIRHOSSEIN AHMADI (Member, IEEE) received the B.Sc. degree in electrical engineering-control systems from the University of Tabriz, Tabriz, Iran, in 2012, and the M.Sc. degree from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2015. He was an Instructor with the Industrial Control Laboratory, Amirkabir University of Technology. His main research interests include power electronics, microgrid control, renewable energies, optimization, and machine learning.

MOJTABA NABIPOUR received the B.Sc. degree in mechanical engineering from the Noshirvani University of Technology, Babol, Iran, in 2009, and the M.Sc. degree from Tarbiat Modares University, Tehran, Iran, in 2015. He is currently a Research Assistant with the Department of Mechanical Engineering, Tarbiat Modares University. His main research interests include additive manufacturing, composite materials, renewable energies, and machine learning.

BEHNAM MOHAMMADI-IVATLOO (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the University of Tabriz, Tabriz, Iran, in 2006, and the M.Sc. and Ph.D. degrees (Hons.) from the Sharif University of Technology, Tehran, Iran, in 2008 and 2012, respectively. He is currently a Professor with the Faculty of Electrical and Computer Engineering, University of Tabriz. He is also with Duy Tan University. His main research interests include economics, operation, and planning of intelligent energy systems in a competitive market environment.

MD. JALIL PIRAN (Member, IEEE) received the Ph.D. degree in electronics and radio engineering from Kyung Hee University, South Korea, in 2016. Subsequently, he continued his work as a Post-doctoral Research Fellow, in the field of resource management and Quality of Experience in 5G cellular networks and the Internet of Things, with the Networking Laboratory, Kyung Hee University. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Sejong University, Seoul, South Korea. He has published substantial number of technical papers in well-known international journals and conferences in research fields of 5G and 6G wireless communications, the Internet of Things (IoT), multimedia communication, cognitive radio networks, applied machine learning, security, and smart grid.

Dr. Piran has been an Active Delegate from South Korea in Moving Picture Experts Group (MPEG) since 2013 and an Active Member of the International Association of Advanced Materials (IAAM) since 2017, in the worldwide communities. He received the IAAM Scientist Medal of the year 2017 for notable and outstanding research in the field of new age technology and innovation in Stockholm, Sweden. Moreover, he has been recognized as the Outstanding Emerging Researcher by the Iranian Ministry of Science, Technology, and Research, in 2017. In addition, his Ph.D. dissertation has been selected as the Dissertation of the Year 2016 by the Iranian Academic Center for Education, Culture, and Research in the field of electrical and communications engineering.

• • •