



# Wind Energy Modelling and Machine Learning Approach to Study Wind Direction Effect

Muhammad Raza<sup>1\*</sup>, Adeel Tahir<sup>2</sup>, Zeshan Iqbal<sup>3</sup>, Zaheer Uddin<sup>3</sup>, Ejaz Ahmed<sup>4</sup>,  
Majid Hussain<sup>5</sup>, Arif A. Azam<sup>3</sup>, and Naeem Sadiq<sup>5</sup>

<sup>1</sup>Department of Physics, Karakoram International University, Gilgit-Baltistan, Pakistan

<sup>2</sup>Department of Physics, Federal Urdu University of Arts, Science & Technology,  
Karachi, Pakistan

<sup>3</sup>Department of Physics, University of Karachi, Karachi, Pakistan

<sup>4</sup>Department of Physics, Model College for Boys G-6/3, Islamabad, Pakistan

<sup>5</sup>Institute of Space Science & Technology, University of Karachi, Karachi, Pakistan

**Abstract:** Wind energy is one of the green renewable energy sources that is available everywhere. The wind-generated electrical energy is much less than the available wind potential. No windmill can even harness 50% of the available wind energy. A lot of research investigations are still needed to explore and harness the maximum energy from wind potential. This work is one of such series of research, in which we modeled wind speed using the Weibull distribution through a Python program that uses the least square method based on Python built-in functions and evaluated the shape and scale parameters of the distribution. The program also compares parameters calculated by other existing methods. The Python program based on the least square method fits the Weibull distribution well compared to the existing methods. The maximum value of scale parameters was found in June (more than 6.2), the corresponding value is also close in May, where it is more than 6.1; the other two months that follow June and May (July and September) have scale parameters near 5.7. It shows that the wind potential is maximum in June, and reasonable wind energy is available in May, July, and September. The effect of wind direction on the modelling of wind speed is also investigated. Perhaps it is the first study that involves wind direction in wind speed modelling. Two different Artificial Neural Network Architectures were studied with and without wind directions in the input. It was found that the results improve if wind direction is also taken in the list of input parameters. The Root Mean Square Error is the least (RMSE = 0.7224) for the model which includes wind direction in the input layer, the performance indicator (0.5219) is also the best for this architecture as compared to the other three.

**Keywords:** Weibull Distribution, Wind Energy Modelling, Python Programming, Least Squares Method, Machine Learning, Artificial Neural Networks (ANN), Wind Direction.

## 1. INTRODUCTION

Wind speed prediction through meteorological parameters is a critical area of research in renewable energy, particularly for assessing wind energy potential. The Weibull distribution has emerged as a widely adopted statistical model for characterizing wind speed data due to its flexibility in capturing the stochastic nature of wind patterns. Numerous studies have explored methods to estimate the Weibull

parameters (shape  $k$  and scale  $c$ ) and evaluate their accuracy, often leveraging computational tools and regional wind datasets. Elahi *et al.* [1] developed a Python library, *windz*, incorporating six methods, viz., Method of Moments (MoM), Empirical Method (EM), Energy Pattern Factor Method (EPFM), Maximum Likelihood Method (MLM), Modified Maximum Likelihood Method (MMLM), and Maximum Entropy Principle (MEP), to analyze wind speed data from three Pakistani cities

Received: July 2025; Revised: August 2025; Accepted: September 2025

\* Corresponding Author: Muhammad Raza <[raza.physics@kiu.edu.pk](mailto:raza.physics@kiu.edu.pk)>

(Harnai, Jacobabad, Talagang) [1]. Their findings highlighted the Modified Maximum Likelihood Method (MMLM) as the most efficient, while the graphical method was deemed the least effective. Similarly, Zahid *et al.* [2] introduced a computer package for wind speed modelling, emphasizing the utility of MMLM and MLM through statistical error tests (Mean square, Chi-square, R-square). Regional Studies in Pakistan provided information on the viability of site-specific wind energy. Khan *et al.* [3] compared the coastal wind profiles of Gwadar and Pasni, revealing greater stability in the Gwadar (parameter of form  $k = 4 - 6$ ), but greater wind energy density in Pasni. Khan *et al.* [3] also compared four methods (MoM, EM, EPFM, MLM) using Jiwani's wind data (1998 - 2007) and found that while all methods yielded nearly identical scale parameters ( $c$ ), MLM provided the best agreement with measured mean wind speeds [4, 5]. However, in a separate study analyzing Ormara's data, Khan *et al.* [4] concluded that MoM, EMP, and MLM outperformed MMLM in estimating shape parameters ( $k$ ), with differences within 0.3. These discrepancies underscore the context-dependent efficacy of parameter estimation methods.

Recent studies have proposed novel approaches to improve accuracy. Uddin and Sadiq [6] introduced the Method of Quartiles (MOQ), which calculates  $k$  and  $c$  using the first and third quartiles. This method has demonstrated lower values of the Akaike Information Criterion (AIC) and reduced overestimation of wind density compared to traditional methods, positioning MOQ as superior to cities such as Karachi, Hyderabad, and Quetta. Rajput *et al.* [7], through advanced computational efficiency, suggested a simplified gamma function formula for estimating scale parameters, achieving an error  $< 0.2\%$  and validating MLM as the most accurate method for Hyderabad wind potential. They classified Hyderabad as the most promising wind power site in Pakistan, followed by Karachi, due to consistent wind speeds that exceed the turbine cutting thresholds. On the other hand, cities like Peshawar and Lahore showed insignificant potential except for specific months. The integration of programming tools significantly enhanced validation processes.

Rehman *et al.* [8] developed a Python-based Newton-Gauss technique, showing minimum RMSE and AIC values for four cities, thus offering

a robust alternative to established methods. These innovations highlight the growing role of computational tools in improving parameter estimation accuracy. They approved the Python programs to automate the estimation of parameters, emphasizing the importance of software in modern wind energy research.

Uddin *et al.* [9] employed a least squares method (LSM) to evaluate Lagrange multipliers for MEP, reaching better adjustments than classic MEP through Python-based error metrics (RMSE, Chi-Square). While the Weibull distribution is widely used, Sadiq [10] demonstrated that extreme (generalized) value distributions have better modeled continuous and seasonal winds of Karachi with 95% confidence. This discovery emphasizes the need to select a specific context model, particularly in coastal urban environments.

The forecast of wind speed evolved significantly with the integration of machine learning techniques (ML) and deep learning (DL), particularly when combined with weather parameters. Although probabilistic models such as the Weibull distribution remain fundamental, recent studies emphasize the role of data-oriented approaches to improve accuracy and address the stochastic nature of wind patterns in Machine Learning and Hybrid Models for Wind Speed Prediction. Artificial Neural Networks (ANNs) and Genetic Expression Programming (GEP) emerged as robust tools for short-term wind speed forecast. Ghorbani *et al.* [11] compared ANN, GEP, multiple linear regression (MLR), and the persistence method using autocorrelation functions to determine input delays for wind speed data per hour from Kersey, Colorado. Their findings revealed that ANN and GEP performed comparably, with MLR surpassing MLR and persistence. Similarly, Demolli *et al.* [12] demonstrated the transferability of ML models (for example, random forest, increased gradient) for long-term wind energy forecasting in geographical places, highlighting their usefulness in the pre-assessment of the feasibility of wind farms in undeveloped areas.

Hybrid models increase predictive accuracy. Zhu *et al.* [13] suggest an amalgamation of convolutional neural networks and a Multilayer Perceptron for the wind farm data. Hur [14] proposed a two-stage method using an Extended

Kalman filter to estimate, and then a neural network for short-term forecasting of the wind speed. These hybrid approaches illustrate practical applications in wind power systems. Li and Jin [15] have also developed a hybrid model that combines resource selection, machine learning, and multi-objective optimization. Its application to the wind speed data from Penglai, China, generated unique objective models. Advanced uncertainty modelling related to convective wind was addressed by Ehsan *et al.* [16] by comparing Quantile Regression Forests and Bayesian Additive Regression Trees for the tempest. The QRF stood out in points and PIS estimates, reducing systematic errors and supporting emergency preparation during the extreme climate.

Regional studies emphasize the adaptability of ML models to local weather conditions. Türkan *et al.* [17] predicted the wind speed to 30 m high using 10 m data in Türkiye, identifying Support Vector Machines (SVM) as the most accurate method. In Malaysia, Hanoon *et al.* [18] evaluated the regression of the Gaussian process (GPR), increased the trees (BT), and SVR in 14 stations, with GPR reaching higher accuracy, despite the challenges in the correlation force. These studies emphasize the importance of adapting models to regional wind regimes and integrating pre-processing techniques to improve performance.

While existing studies validate ML/DL models for wind speed forecasting, challenges persist in dealing with non-stationary data and incorporating various weather variables (e.g., temperature, humidity). Hanoon *et al.* [18] emphasized the need to integrate optimization algorithms to improve the correlation between the expected and observed values. These gaps present opportunities for new structures, such as the ANN-Weibull hybrid proposed in this study, which takes advantage of the meteorological parameters for probabilistic forecasting. Han *et al.* [19] proposed a hybrid model that integrated research and weather forecast simulations (WRF), decomposition of the empirical mode of the complete set (Ceemdan), and a bidirectional CNN LSTM network optimized by attention mechanisms.

Cadenas and Rivera [20] employed Autoregressive Integrated Moving Average (ARIMA) models to predict wind speed in Oaxaca, Mexico, incorporating wind direction along with

temperature and moisture, and reported greater accuracy in the capture of seasonal wind variability. Similarly, Foley *et al.* [21] used Support Vector Machines (SVM) to predict the production of wind energy in Ireland, emphasizing the direction of the wind as a critical factor to resolve the dynamics of the coastal wind. Their work has shown that directional data reduced RMSE by 18% compared to the models excluding this variable.

Lydia *et al.* [22] developed a wind speed forecast system using SVM and K-Nearest Neighbors (K-NN) with inputs, including wind direction, but its structure lacked the nonlinear modelling capacity of the RNAs to deal with complex interactions between variables. Liu *et al.* [23] proposed a hybrid forecast model that combines transformations of Wavelets and Gradient reinforcement machines, but their approach did not explore ANS' potential to autonomously learn hierarchical characteristics from directional and temporal data. The absence of an ANN-based framework in these studies represents a difference in the literature, as ANS specifically suits non-linear relationships and sequential dependence, adapted to the sequential dependence contained in meteorological datasets. The task addresses this difference by developing an ANN-based model that integrates the air direction with temperature, humidity, and the precursor air speed values, using Python libraries such as TensorFlow and causes for implementation. Our approach not only creates fundamental insights of pre-studies but also introduces a novel application of ANNs to increase prediction accuracy in air speed modelling.

## 2. MATERIALS AND METHODS

Weibull Distribution is one of the most widely used distributions for modelling wind speed distributions. Shape ( $k$ ) and a scale parameter ( $c$ ) characterized this distribution through its cumulative distribution function (CDF) as:

$$F(v; k, c) = 1 - e^{-\left(\frac{v}{c}\right)^k} \quad (1)$$

Where  $F(v; k, c)$  is a function that provides the cumulative probability of observing a particular wind speed, and  $v$  is the wind speed. The shape parameter determines the shape of the distribution (e.g., spread or skewness) through any of the  $k < 1$  (high frequency of low wind speeds),  $k = 1$

(represents exponential distribution; wind speeds are randomly and uniformly distributed over time), or  $k > 1$  (high frequency of high wind speeds). The scale parameter ( $c$ ) represents the characteristic wind speed, indicating the range or magnitude of wind speeds. Higher values of  $c$  signify stronger average wind speeds in the region being analyzed. The Probability Density Function (PDF) can be presented as [24]:

$$W(v; k, c) = \left(\frac{k}{c}\right) \left(\frac{v}{c}\right)^{k-1} \exp\left(-\left(\frac{v}{c}\right)^k\right) \quad (2)$$

Where  $W(v; k, c)$  represents the likelihood of wind speed occurring at a specific value, showing how wind speeds are distributed over the random variable  $v$  (e.g., time, measurement). To estimate scale and shape parameters, different methods have been developed. The following methods are utilized in our study.

### 2.1. Method of Moments (MoM):

This method utilizes the mean and variance of the data to estimate the Weibull parameters [25]. The shape parameter  $k$  can be estimated by solving the following equation:

$$\sigma = c \left[ \Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right]^{\frac{1}{2}} \quad (3)$$

Where  $\sigma$  standard deviation of the wind speed data and  $\Gamma$  is the Gamma function. Once  $k$  is obtained, the scale parameter  $c$  is also calculated by using the following equation:

$$\bar{v} = c \Gamma\left(1 + \frac{1}{k}\right) \quad (4)$$

### 2.2. Empirical Method (EM)

This method relies on empirical relationships derived from the mean and standard deviation of the wind speed data [26]. To estimate the shape parameter ( $k$ ), the following empirical formula is used.

$$k = \left(\frac{\sigma}{\bar{v}}\right)^{-1.086} \quad (5)$$

Where,  $\sigma / \bar{v} = C$ , which is known as the coefficient

of variation. Once  $k$  has been estimated, the scale parameter ( $c$ ), also calculated through the relation:

$$c = \frac{\bar{v}}{\Gamma\left(1 + \frac{1}{k}\right)} \quad (6)$$

We used Python to calculate  $\Gamma$ .

### 2.3. Maximum Likelihood Method (MLM)

The MLM estimates the parameters  $k$  and  $c$  by maximizing the likelihood function [27]. The optimization conditions can be used to find the values of  $k$  and  $c$ .

$$k = \left[ \frac{\sum_{i=1}^n f_i v_i^k \ln(v_i)}{\sum_{i=1}^n v_i^k} - \frac{\sum_{i=1}^n f_i \ln(v_i)}{\sum_{i=1}^n f_i} \right]^{-1} \quad (7)$$

$$c = \left( \frac{1}{\sum_{i=1}^n f_i} \sum_{i=1}^n f_i v_i^k \right) \quad (8)$$

Where  $f_i$  represents the frequency or number of observations (e.g., recorded wind speed occurrences) in the  $i^{\text{th}}$  bin or interval.

### 2.4. Least Squares Method (LSM)

The least squares method is used by minimizing the square difference between recorded and fitted values [28]. It is an iterative method; the iteration stops once the termination criteria are reached. Python built-in library for least square fit was used to model the Weibull distribution. The program finds the new values of the probability distribution  $p$  (predicted) and compares them with the corresponding known values  $p$  (recorded). The sum of squares errors is calculated by the following formula.

$$SSE = \sum \{p(\text{recorded}) - p(\text{predicted})\}^2 \quad (9)$$

The program determines the scale and shape parameters by minimizing SSE.

### 2.5. Neural Network Architecture

The Artificial Neural Network method is a machine learning technique that discovers the hidden relationship between an output variable and a set of input variables. It works on a pattern like the way our brain. It makes use of neurons that link output and



input variables through one or more hidden values in a hidden layer between the output and input variables [29]. In ANN methods, the data is divided into three parts; the main part consists of more than 50% of the data is used to train the network. Once the network is trained, the second part of the data is used for testing the trained network. Finally, the last part of the data is used to validate the network. An overall check is also carried out on the complete set of data. The goodness of the network is measured by a couple of statistics: the correlation coefficient, the Root Mean Square Error, and the Performance of the network. If the correlation coefficient in training, testing, validation, and overall check is significant, the network is considered reliable; further reliability check is done with the lowest value of RMSE, and performance coefficient.

In general, there are three layers, known as input, hidden, and output layers. The data is fed into input layers having one or more input parameters, and the values in the hidden layer are generated using random weights and input values fed to an activation function. The hidden layer then generates output values using an activation function. A comparison is made between generated output and recorded output; if the difference is insignificant, the network is considered as trained; otherwise, a backward feedback error analysis generates new weights in the input layer for new values generated for the hidden layer. The process continues till the network is well-trained. The neurons in the hidden layer may be varied for optimized results.

In the present study, the input variables were Temperature and Humidity, whereas the output variable was wind speed. A hidden layer with 10 neurons was used to train the network. To notice the influence of the wind direction on the wind speed. The wind direction is also incorporated in the input parameters. A precursor network is also trained in this study with the same parameters. One of the four ANN networks is shown in Figure 1.

## 2.6. Wind Energy Potential

Wind energy potential mainly depends on the wind speed of the potential site; the other factors on which it depends are the area swept by the wings of the wind turbine and the air density of the location. The air density varies from site to site and depends on the altitude and air temperature. If  $A$  is the area

swept by the rotors in ( $\text{m}^2$ ),  $\rho$  is the air density in ( $\text{kg m}^{-3}$ ), and  $v$  is the wind speed in ( $\text{m s}^{-1}$ ), the power in watts is given by the following formula [30]:

$$P = \frac{1}{2} A \rho \int_0^{\infty} v^3 f(v) dv = \frac{1}{3} A \rho \bar{v}^3 \quad (10)$$

Here  $\bar{v}^3$  is the average value of the cube of the wind speed. To make this power independent of the area of the rotors, we find fractional power or power density (average power per unit area in ( $\text{watts m}^{-2}$ )) and is given by:

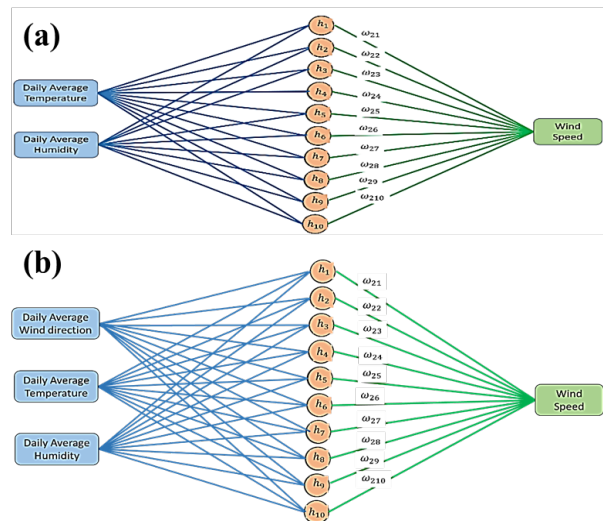
$$p = \frac{P}{A} = \frac{1}{3} \rho \bar{v}^3 \quad (11)$$

Wind potential for Karachi City for each month from January to December was calculated using the above formula.

## 2.7. Statistical Errors for Validation of Designed Models

The validity of designed models has been checked by the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Chi-Square Test ( $\chi^2$ ), and Kolmogorov-Smirnov (KS) Test are calculated by using Equations (12) to (15) [31].

$$MAE = \left( \frac{1}{n} \right) \sum |v_i - \hat{v}_i| \quad (12)$$



**Fig. 1.** The ANN architecture with (a) Temperature and Relative Humidity, (b) Temperature and Relative Humidity, and Wind direction in the input layer.

$$RMSE = \sqrt{\left[\left(\frac{1}{n}\right) \sum (v_i - \hat{v}_i)^2\right]} \quad (13)$$

$$\chi^2 = \sum \left[ \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} \right] \quad (14)$$

and

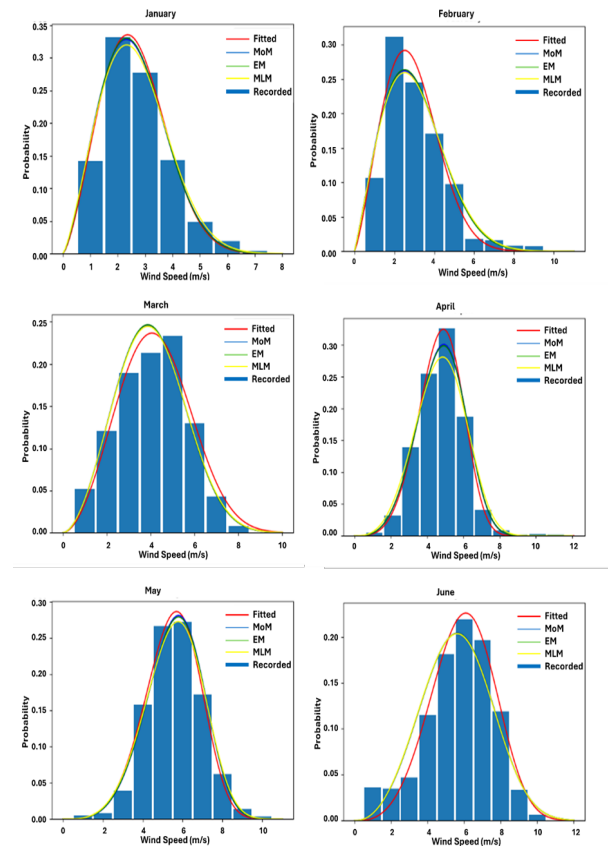
$$KS = \max |F_1(v) - \hat{F}_1(v)| \quad (15)$$

### 3. RESULTS AND DISCUSSION

The demand for renewable energy has been increasing day by day. Different available Renewable potentials around the world have been exploited and harnessed to generate electricity. Wind is one of the renewable energies that has potential in almost every part of the world. Weibull distribution with two parameters is frequently used in modelling wind speed distributions. To estimate the parameters of the Weibull distribution, various techniques and methods have been developed. In this study, we used the method of least squares to calculate the parameters using a built-in Python function. A Python program was developed to carry out the calculations and to estimate the parameters. The built-in function works well if initial conditions are close to the real values of the distribution parameters, and the convergence becomes faster. To do so, we used the fact that most of the scale parameters are greater than 2, and the value of the shape parameter can be written as a function of the scale parameter. This formula was given as the input parameter.

The program also computes the parameters using the Method of Moments (MoM), the Empirical Method (EM), and the Maximum Likelihood Method (MLM) to compare the results. Table 1 gives the estimated values of the scale and shape parameters computed by these methods. Both the values of scale and shape parameters are higher for estimation by the least square method, except in one or two months. Hussain *et al.* [32] also studied the wind speed distribution of the coastal region of Karachi. They also used the maximum likelihood method for estimation of shape and scale parameters. The average value of the shape parameter for 10 m height wind data was 3.3; the

corresponding average value in our study is 3.5. The PDF generated by the estimated values of the shape and scale parameters of the Weibull distribution is shown in Figures 2 and 3. The PDFs generated by the least square method are good representatives of wind speed distribution; these PDFs cover excellently the histograms generated by recorded wind speed data in almost all recorded data. In July, PDFs generated by all the methods coincide perfectly. Visually, it looks like the least method gives the best results. Now we have compared quantitatively the four different statistical errors that were computed to compare the performance of the methods used in this work for the calculations of the Weibull parameters and the generation of the PDF. Table 2 shows the values of errors (RMSE, MAE, Chi-Square, and Kolmogorov-Smirnov) from January to December. The least-square results give lower values of these statistical errors except for March, August, and October; however, the errors are not very far from each other. The Kolmogorov-Smirnov errors for the least square methods are lower except for March, August, and October.



**Fig. 2.** Comparison of PDFs generated (Jan-June) with the determined values of shape and scale parameters using four estimation methods and the recorded wind speed distribution histogram.

**Table 1.** The shape and scale parameters of the Weibull distribution were calculated from four methods for January to December for Karachi city.

|              | k        | c      | k      | c      | k         | c      |
|--------------|----------|--------|--------|--------|-----------|--------|
|              | January  |        | May    |        | September |        |
| Least square | 2.4245   | 2.9378 | 4.5632 | 6.0117 | 5.3670    | 5.7775 |
| MoM          | 2.3514   | 2.9291 | 4.5586 | 6.1212 | 5.4234    | 5.8488 |
| EM           | 2.3661   | 2.9288 | 4.5253 | 6.1238 | 5.3751    | 5.8518 |
| MLM          | 2.3110   | 2.9688 | 4.4140 | 6.1210 | 5.2400    | 5.8533 |
|              | February |        | June   |        | October   |        |
| Least square | 2.2954   | 3.2365 | 3.8950 | 6.5716 | 3.0384    | 4.2200 |
| MoM          | 2.1081   | 3.3881 | 3.3063 | 6.2710 | 2.9189    | 4.0384 |
| EM           | 2.1273   | 3.3882 | 3.3037 | 6.2712 | 2.9257    | 4.0380 |
| MLM          | 2.1030   | 3.4152 | 3.2930 | 6.2634 | 2.9170    | 4.0502 |
|              | March    |        | July   |        | November  |        |
| Least square | 2.8366   | 4.7310 | 4.9434 | 5.8680 | 2.3856    | 2.9270 |
| MoM          | 2.8108   | 4.5184 | 4.9640 | 5.8698 | 2.4324    | 2.8296 |
| EM           | 2.8195   | 4.5178 | 4.9229 | 5.8726 | 2.4437    | 2.8293 |
| MLM          | 2.8040   | 4.5309 | 4.9460 | 5.8717 | 2.3910    | 2.8750 |
|              | April    |        | August |        | December  |        |
| Least square | 4.4608   | 5.1896 | 5.1021 | 5.2883 | 2.2782    | 2.8058 |
| MoM          | 4.1712   | 5.2582 | 4.5405 | 5.2244 | 2.3243    | 2.7144 |
| EM           | 4.1413   | 5.2604 | 4.5035 | 5.2269 | 2.3394    | 2.7141 |
| MLM          | 3.8750   | 5.2578 | 4.5160 | 5.2252 | 2.3000    | 2.7599 |

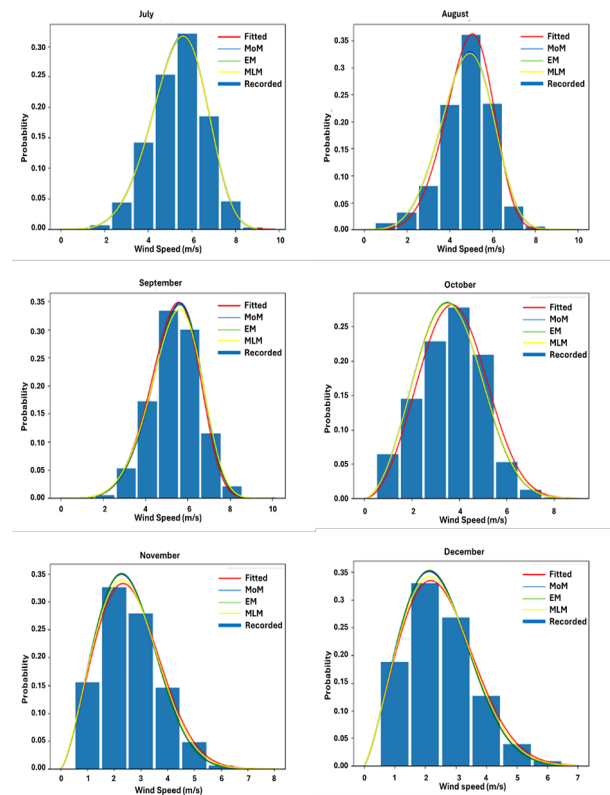
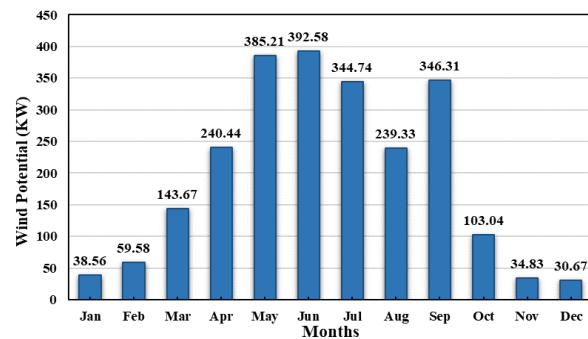
**Fig. 3.** Comparison of PDFs generated (July-Dec) with the determined values of shape and scale parameters using four estimation methods and the recorded wind speed distribution histogram.

Figure 4 shows the monthly wind potential for Karachi City. The wind potential was calculated using equation 11, for Karachi from January to December. The wind potential is maximum for June. From April to September, the wind speed is relatively high, so the wind potential is reasonably good for these months. In the months from October to March, wind speed is relatively low, so the wind potential compared to other months is low.

Wind modelling was done by many researchers in different studies by considering different meteorological parameters as input, but the wind direction was not considered in wind

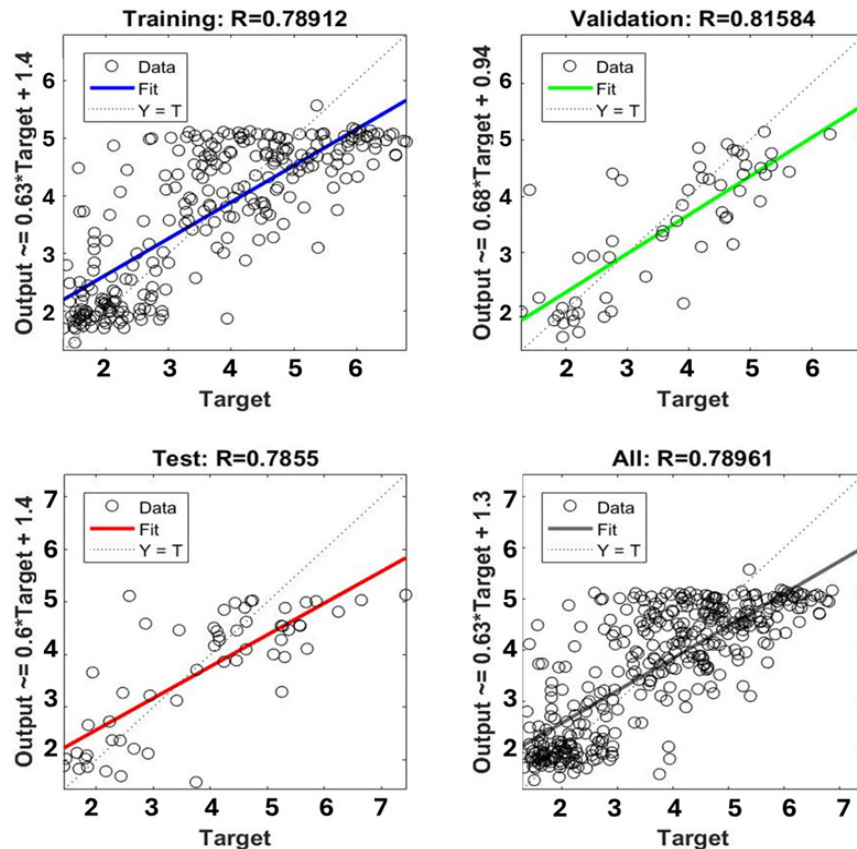
**Fig. 4.** Monthly wind potential in KW for Karachi city.

speed modelling. However, some researchers have used wind direction for wind energy analysis. Carta *et al.* [31] developed a joint probability distribution based on wind speed and wind direction for wind energy analysis [33]. Wang and Liu [34] used a finite mixture statistical distribution to assess wind energy potential using wind speed and wind direction data. They used a von Mises mixture distribution with various parameters for wind direction. Narain *et al.* [35] used the wake effect of wind turbines to investigate the effect of wind direction on wind energy potential. Mata *et al.* [36] studied the effect of wind direction on wind power output; they used three models for wind potential exploitation based on wind speed and wind direction shear. Models were based on blade elements and actuator disc representations.

We used an Artificial Neural Network to comprehend the effect of wind direction. We proposed two methods, one with temperature and humidity as input, and the second one as a precursor method. In each of the two methods, we further assessed each method with and without

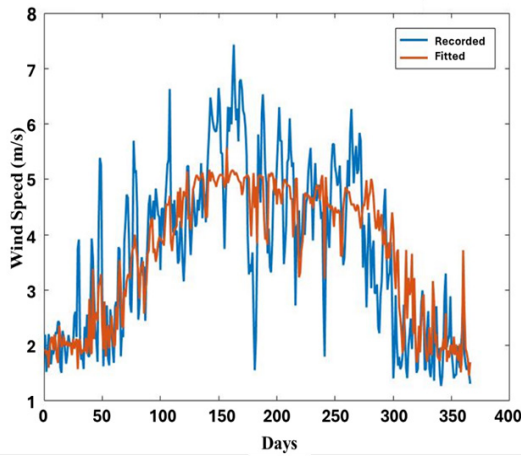
wind direction. All together four different ANN architectures were investigated in this work.

First, we consider the ANN architecture in which we modeled the wind speed distribution using temperature and moisture as input parameters. The results of the modelling are shown in the Figures. 5 and 6. Figure 5 depicts the regression analysis of the training, testing, validation, and overall performance of the model. Our results are not very promising as each R is less than 80%. Figure 6 portrays the comparison of the input wind speed distribution and the output generated by the ANN model. The conclusion made in Figure 5 is also verified by Figure 6; the overlap is not favorable. Ghorbani *et al.* [37] also used ANN to model wind speed distribution for Tabriz, Iran. The input parameters in ANN were Air temperature, air pressure, relative humidity, and precipitation. They used three different models having n, 2n, and 2n+1 neurons in the hidden layer. The coefficient of correlation for each of these models was above 95%. The wind pattern in Karachi does not show such promising results.



**Fig. 5.** Regression analysis of Testing, training, validation, and overall wind speed output, the input parameters are daily average temperature and relative humidity.



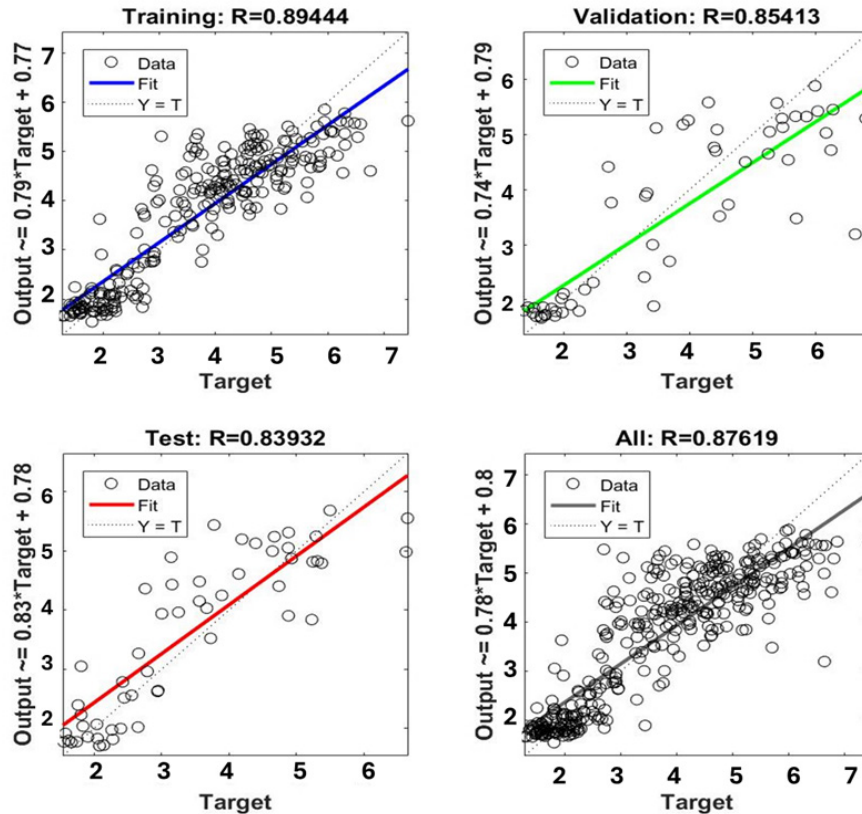


**Fig. 6.** Comparison of recorded and ANN modelled wind distributions by using daily temperature and relative humidity.

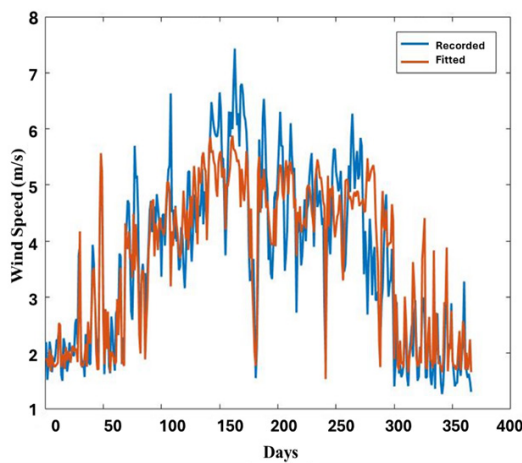
To investigate the impact of the wind direction in the ANN modelling, we added wind direction to the input parameters of the previous model. Figure 7 furnishes the  $R$  (correlation) values between the training (0.894), the test (0.839), and the validation

phases (0.854). Figure 7 depicts a stronger correlation between the predicted and observed wind speeds compared to the previous model ( $R = 0.785$ - $0.816$ ). Figure 8 illustrates similar results as shown in Figure 4, i.e., the comparison of recorded and modeled values of wind speed. The overlapping of recorded and modeled values is relatively better in Figure 8 compared to that in Figure 6 and demonstrates optimized modelling results due to wind direction incorporation in the list of input parameters.

In the second ANN architecture, we employed precursor values of temperature and Humidity to predict wind speed. The model's performance shows a moderate correlation during training, testing, and validation phases, with an overall correlation of  $R = 0.798$ . However, the test and validation metrics reveal relatively weaker performance, with  $R = 0.702$  for testing and  $R = 0.670$  for validation, as illustrated in Figure 9, indicating that this network underperforms compared to earlier models. Figure 10 further authenticates the results shown in Figure 9. Figure 10 shows the comparison of the input



**Fig. 7.** Regression analysis of testing, training, validation, and overall wind speed output; the input parameters are daily average temperature, relative humidity, and wind direction.

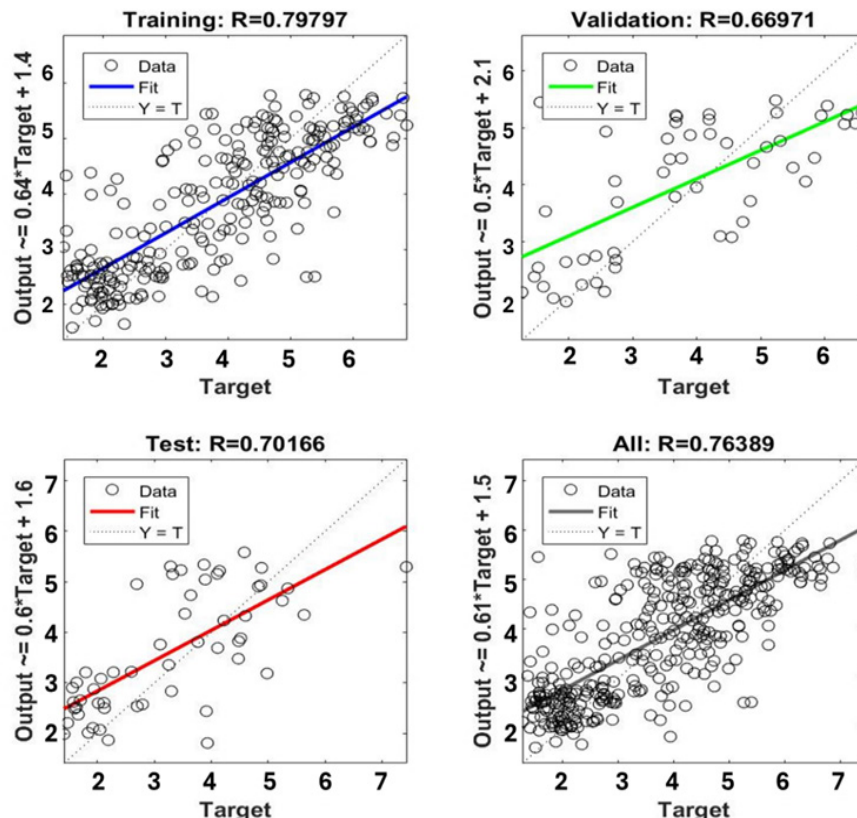


**Fig. 8.** Comparison of recorded and ANN modelled wind distributions by using daily temperature, relative humidity, and wind direction.

wind speed distribution and the output generated by the ANN model. Again, the results are not very promising as each  $R$  is less than 80%. The conclusion made in Figure 9 is also verified by Figure 10. Jamil and Zeeshan [38] studied wind speed modelling for

wind speed data of Gujarat, India, using ANN. They also used precursor values of wind speed to predict current wind speed. The coefficient of correlation between predicted and recorded values of wind speed was found to be above 98%. Again, the wind speed pattern of Karachi does not show such high values for the correlation coefficient.

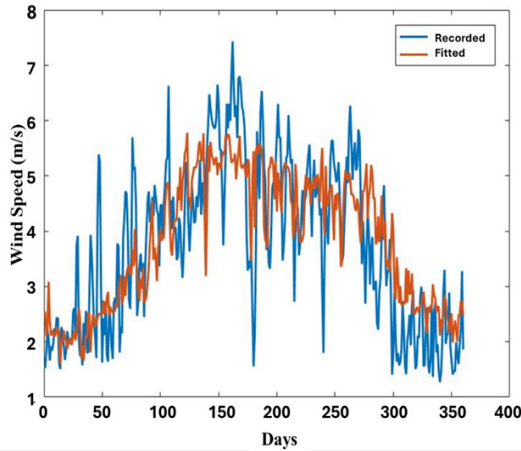
In the second case, the influence of wind direction in precursor modelling is depicted in Figures 11 and 10. We incorporated wind direction in the input list of the previous model. Figure 11 shows the  $R$  values for the training ( $R = 0.837$ ), testing ( $R = 0.807$ ), and validation ( $R = 0.803$ ). This attests to a significant improvement in the correlation by the inception of wind direction in the input parameters. Figure 12 illustrates the comparison of recorded and modeled values of wind speed. By comparing Figures 10 and 12, it is obvious that the overlapping of recorded and modeled values is exceptionally better than in Figure 10. It further elucidates the improvement in the results due to the inclusion of wind direction in the list of input parameters.



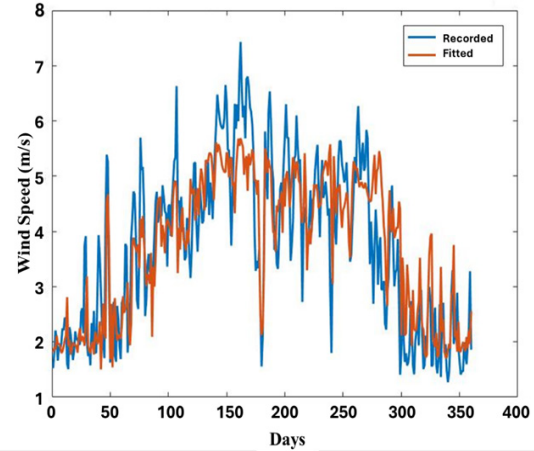
**Fig. 9.** Regression analysis of Testing, training, validation, and overall wind speed output, the input parameters are precursor values of daily average temperature and relative humidity.

The two cases scrutinized above for wind modelling with and without wind direction ascertain that wind direction is an important input parameter for wind speed forecasting. The addition of wind direction mitigates the Root Mean Square error and augments the performance of the modelling (see

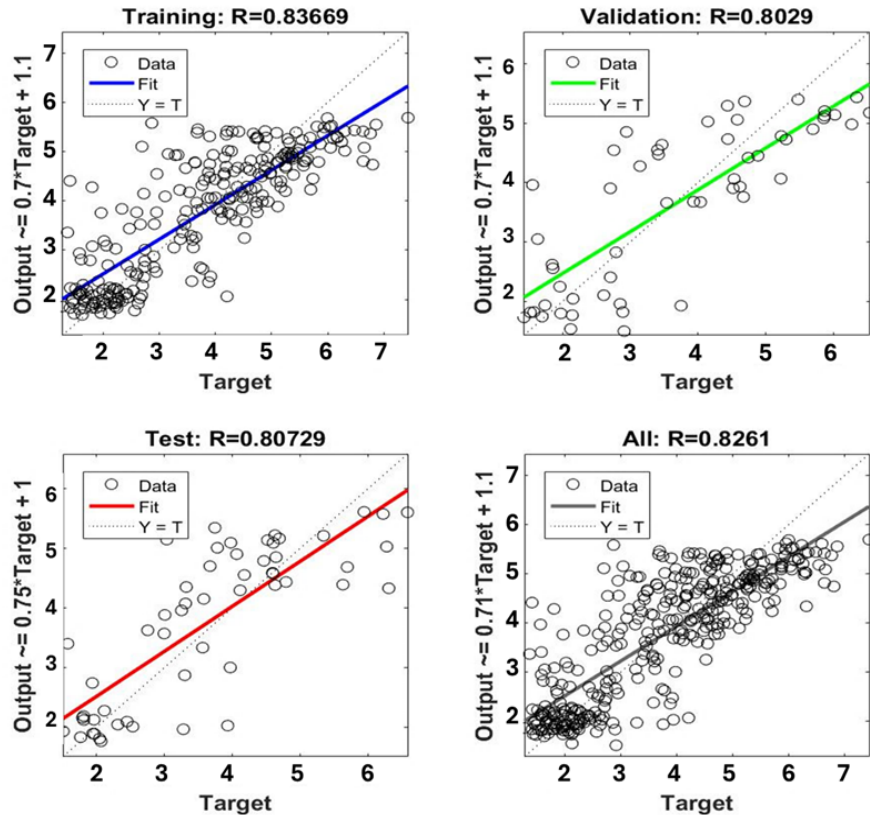
Table 2). Li and Shi [39] compared three different ANN models that are used to predict the wind speed of two sites in North Dakota. They used Back Propagation NN, Adaptive Linear Element NN, and Radial Basis Function NN. The RMSE values in the prediction of wind speed using three ANN models



**Fig. 10.** Comparison of recorded and ANN modelled wind distributions by using precursor values of daily temperature and relative humidity.



**Fig. 12.** Comparison of recorded and ANN modelled wind distributions, precursor values of daily temperature, relative humidity, and wind direction.



**Fig. 11.** Regression analysis of Testing, training, validation, and overall wind speed output. The input parameters are precursor values of daily average temperature, relative humidity, and wind direction.

**Table 2.** Four statistical errors were calculated in the PDFs generated by the shape and scale parameters of four estimation methods.

|          |         | Least square | MoM   | EM    | MLM   |           |         | Least square | MoM   | EM    | MLM   |
|----------|---------|--------------|-------|-------|-------|-----------|---------|--------------|-------|-------|-------|
| January  | RMSE    | 0.084        | 0.083 | 0.083 | 0.085 | July      | RMSE    | 0.07         | 0.07  | 0.071 | 0.07  |
|          | MABE    | 0.076        | 0.077 | 0.077 | 0.078 |           | MABE    | 0.047        | 0.046 | 0.047 | 0.047 |
|          | Chi     | 0.082        | 0.076 | 0.078 | 0.072 |           | Chi     | 0.088        | 0.091 | 0.085 | 0.088 |
|          | Kolmo-S | 0.142        | 0.142 | 0.142 | 0.142 |           | Kolmo-S | 0.176        | 0.172 | 0.18  | 0.176 |
| February | RMSE    | 0.073        | 0.095 | 0.093 | 0.096 | August    | RMSE    | 0.072        | 0.049 | 0.052 | 0.051 |
|          | MABE    | 0.05         | 0.054 | 0.053 | 0.054 |           | MABE    | 0.058        | 0.041 | 0.042 | 0.041 |
|          | Chi     | 0.05         | 0.047 | 0.046 | 0.047 |           | Chi     | 1.251        | 0.511 | 0.479 | 0.49  |
|          | Kolmo-S | 0.15         | 0.27  | 0.261 | 0.277 |           | Kolmo-S | 0.13         | 0.091 | 0.106 | 0.101 |
| March    | RMSE    | 0.141        | 0.132 | 0.132 | 0.132 | September | RMSE    | 0.057        | 0.06  | 0.058 | 0.051 |
|          | MABE    | 0.084        | 0.078 | 0.078 | 0.078 |           | MABE    | 0.043        | 0.045 | 0.044 | 0.038 |
|          | Chi     | 0.133        | 0.127 | 0.128 | 0.127 |           | Chi     | 0.289        | 0.306 | 0.28  | 0.219 |
|          | Kolmo-S | 0.405        | 0.379 | 0.379 | 0.38  |           | Kolmo-S | 0.128        | 0.14  | 0.132 | 0.102 |
| April    | RMSE    | 0.058        | 0.068 | 0.069 | 0.073 | October   | RMSE    | 0.108        | 0.102 | 0.102 | 0.103 |
|          | MABE    | 0.039        | 0.04  | 0.04  | 0.041 |           | MABE    | 0.072        | 0.065 | 0.065 | 0.065 |
|          | Chi     | 0.075        | 0.049 | 0.047 | 0.036 |           | Chi     | 0.151        | 0.145 | 0.146 | 0.143 |
|          | Kolmo-S | 0.129        | 0.207 | 0.212 | 0.225 |           | Kolmo-S | 0.282        | 0.278 | 0.277 | 0.279 |
| May      | RMSE    | 0.076        | 0.078 | 0.082 | 0.096 | November  | RMSE    | 0.085        | 0.084 | 0.085 | 0.084 |
|          | MABE    | 0.05         | 0.051 | 0.054 | 0.06  |           | MABE    | 0.078        | 0.076 | 0.076 | 0.077 |
|          | Chi     | 0.036        | 0.037 | 0.041 | 0.057 |           | Chi     | 1.501        | 1.821 | 1.85  | 1.62  |
|          | Kolmo-S | 0.162        | 0.166 | 0.185 | 0.25  |           | Kolmo-S | 0.156        | 0.156 | 0.156 | 0.156 |
| June     | RMSE    | 0.161        | 0.19  | 0.191 | 0.193 | December  | RMSE    | 0.079        | 0.078 | 0.078 | 0.078 |
|          | MABE    | 0.092        | 0.121 | 0.121 | 0.122 |           | MABE    | 0.073        | 0.072 | 0.072 | 0.072 |
|          | Chi     | 0.217        | 0.573 | 0.578 | 0.596 |           | Chi     | 0.143        | 0.168 | 0.171 | 0.155 |
|          | Kolmo-S | 0.497        | 0.508 | 0.51  | 0.521 |           | Kolmo-S | 0.142        | 0.142 | 0.142 | 0.142 |

were more than 1.25 in each case. However, the four different ANN architectures in this study show RMSE values between 0.72 and 0.96.

Table 3 quantifies the RMSE and performance values between different input configurations for wind speed prediction, reinforcing the critical role of wind direction in enhancing model accuracy. Notably, incorporating wind direction consistently reduces Mean Squared Error (MSE), a key indicator

of prediction error, across both precursor-inclusive and precursor-exclusive models. For instance, without precursors, adding wind direction (T, H & D) lowers MSE by 21.5% (from 0.9204 to 0.7224). With precursors, including wind direction (T, H, D & Precursor), reduces MSE by 12.9% compared to the precursor-only model (from 0.9633 to 0.8388). Table 4 shows the statistical testing and respective significance level.

**Table 3.** Comparison of various ANN architectures.

| Wind forecasting by                                         | RMSE   | Performance |
|-------------------------------------------------------------|--------|-------------|
| Temperature & humidity                                      | 0.9204 | 0.8471      |
| Temperature, humidity & wind direction                      | 0.7224 | 0.5218      |
| Precursor values of (Temperature & humidity)                | 0.9633 | 0.9280      |
| Precursor values of Temperature, humidity) & wind direction | 0.8388 | 0.7035      |



**Table 4.** Statistical testing and significance.

| Dataset 1 Statistics    |                                                                                 | Dataset 2 Statistics |        |
|-------------------------|---------------------------------------------------------------------------------|----------------------|--------|
| Count                   | 360                                                                             | Count                | 360    |
| Mean                    | 3.7208                                                                          | Mean                 | 3.7264 |
| Standard Deviation      | 1.2107                                                                          | Standard Deviation   | 1.4881 |
| Variance                | 1.4659                                                                          | Variance             | 2.2145 |
| <b>T-Test results</b>   |                                                                                 |                      |        |
| Difference Between Mean | -0.0057                                                                         |                      |        |
| T-Statistics            | -0.0561                                                                         |                      |        |
| Degree of Freedom       | 718                                                                             |                      |        |
| P-Value                 | 0.1329                                                                          |                      |        |
| <b>Interpretation</b>   | The difference between means is not statistically significant ( $p \geq 0.05$ ) |                      |        |

#### 4. CONCLUSIONS

A systematic investigation has been carried out to observe the impact of wind direction on wind speed modelling. The ANN was employed in modelling wind speed with two input parameters (temperature and humidity). In the second case, the precursor values of the input parameters were employed to forecast the current value of the wind speed. In the next step, input parameters were appended by an additional parameter, i.e., wind direction. In both cases, it was found that the results of ANN for training, testing, validation, and overall are significantly improved by the addition of wind direction with parameters temperature and humidity. Figures 7 and 11 depict the improved values of the correlation coefficient for all the categories. The best network in the forecasting of wind speed with three parameters (temperature, humidity, and wind direction) yields all correlation coefficients above 80%. This network has the lowest RMSE and improved performance. Wind speed is also modeled using the Weibull distribution. To find the scale and shape parameters, the least squares method is used with Python's built-in function. A Python program was developed to implement the algorithm, the program also computes both the parameters using three existing methods (method of moments, empirical method, and maximum likelihood method) to compare the result of the least square method. The modelling was carried out for each month of the year 2016, and in most of the months, the RMSE and Kolmogorov-Smirnov test values were optimized for the least square method. The PDFs drawn by the estimated values of shape and scale parameters for the four methods elucidate that the least square method's PDF explains well the

histogram generated by the recorded values of wind speed. We conclude that the least square method using Python's built-in function gives reliable results for wind speed modelling. The addition of wind direction in the input parameters to forecast improves the ANN results.

#### 5. CONFLICT OF INTEREST

The authors declare that there is no conflict of interest to submit/publish this research.

#### 6. REFERENCES

1. Z. Elahi, W. Gull, and Z. Uddin. Weibull distribution parameter estimation using spreadsheets. *Spreadsheets in Education* 1: 1-12 (2023). <https://sie.scholasticahq.com/article/72617>
2. M. Zahid, A. Rajput, S. Rehman, M. Adeel, and Z. Uddin. A computer package to model wind speed distribution using six different methods including maximum entropy principle. *International Journal of Energy* 45(3-4): 207-218 (2024).
3. J.K. Khan, M. Shoaib, Z. Uddin, I.A. Siddiqui, A. Aijaz, A.A. Siddiqui, and E. Hussain. Comparison of wind energy potential for coastal locations: Pasni and Gwadar. *Journal of Basic & Applied Sciences* 11: 211-216 (2015).
4. J.K. Khan, F. Ahmed, Z. Uddin, S.T. Iqbal, S.U. Jilani, A.A. Siddiqui, and A. Aijaz. Determination of Weibull parameter by four numerical methods and prediction of wind speed in Jiwan (Balochistan). *Journal of Basic & Applied Sciences* 11: 62-68 (2015).
5. J.K. Khan, Z. Uddin, I.S. Tanweer, F. Ahmed, A. Aijaz, and S.U. Jilani. An analysis of wind speed distribution and comparison of five numerical

- methods for estimating Weibull parameters at Ormara, Pakistan. *European Academic Research II* (11): 14007-14015 (2015).
6. Z. Uddin, and N. Sadiq. Method of quartile for determination of Weibull parameters and assessment of wind potential. *Kuwait Journal of Science* 50(3A): 105-119 (2023).
  7. A.A. Rajput, M. Daniyal, M.M. Zahid, H. Nafees, M. Shafi, and Z. Uddin. New approach to calculate Weibull parameters and comparison of wind potential of five cities of Pakistan. *Advances in Energy Research* 8(2): 95 (2022).
  8. S.U. Rehman, N. Sadiq, I. Tariq, M.M. Khan, M.M. Zahid, A.A. Rajput, and Z. Uddin. A new mathematical technique and its Python program to assess wind potential. *Beni-Suef University Journal of Basic and Applied Sciences* 13(1): 61 (2024).
  9. Z. Uddin, M. B. Khan, M.H., Zaheer, W. Ahmad, and M.A., Qureshi. An alternate method of evaluating Lagrange multipliers of MEP. *SN Applied Sciences* 1: 224 (2019).
  10. N. Sadiq. Seasonal and continual wind speed modelling for the coastal urban city, Karachi. *MAUSAM* 69(2): 289-296 (2018).
  11. M.A. Ghorbani, R. Khatibi, M.H. FazeliFard, L. Naghipour, and O. Makarynsky. Short-term wind speed predictions with machine learning techniques. *Meteorology and Atmospheric Physics* 128: 57-72 (2016).
  12. 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).
  13. Q. Zhu, J. Chen, L. Zhu, X. Duan, and Y. Liu. Wind speed prediction with spatio-temporal correlation: A deep learning approach. *Energies* 11(4): 705 (2018).
  14. S.H. Hur. Short-term wind speed prediction using Extended Kalman filter and machine learning. *Energy Reports* 7: 1046-1054 (2021).
  15. R. Li and Y. Jin. A wind speed interval prediction system based on multi-objective optimization for machine learning method. *Applied Energy* 228: 2207-2220 (2018).
  16. B.M.A. Ehsan, F. Begum, S.J. Ilham, and R.S. Khan. Advanced wind speed prediction using convective weather variables through machine learning application. *Applied Computing and Geosciences* 1: 100002 (2019).
  17. Y.S. Türkan, H.Y. Aydoğmuş, and H. Erdal. The prediction of the wind speed at different heights by machine learning methods. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA)* 6(2): 179-187 (2016).
  18. M.S. Hanoon, A.N. Ahmed, P. Kumar, A. Razzaq, N.A. Zaini, Y.F. Huang, M. Sherif, A. Sefelnasr, K.W. Chau, and A. El-Shafie. Wind speed prediction over Malaysia using various machine learning models: potential renewable energy source. *Engineering Applications of Computational Fluid Mechanics* 16(1): 1673-1689 (2022).
  19. Y. Han, L. Mi, L. Shen, C. S. Cai, Y. Liu, K. Li, and G. Xu. A short-term wind speed prediction method utilizing novel hybrid deep learning algorithms to correct numerical weather forecasting. *Applied Energy* 312: 118777 (2022).
  20. E. Cadenas and W. Rivera. Short-term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. *Renewable Energy* 35(1): 274-278 (2010).
  21. A.M. Foley, P.G. Leahy, A. Marvuglia, and E.J. McKeogh. Current methods and advances in forecasting of wind power generation. *Renewable Energy* 37(1): 1-8 (2012).
  22. M. Lydia, S.S. Kumar, A.I. Selvakumar, and E.P. Kumar. Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Conversion and Management* 112: 115-124 (2016).
  23. H. Liu, H.Q. Tian, and Y.F. Li. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. *Applied Energy* 98: 415-424 (2015).
  24. W. Qamar, M. Hussain, M.B. Zaheer, J. Akram, N. Sadiq, and Z. Uddin. Prediction of sunspot numbers via Weibull distribution and deep learning. *Astrophysics and Space Science* 370(7): 68 (2025).
  25. J. Merganič and H. Sterba. Characterisation of diameter distribution using the Weibull function: method of moments. *European Journal of Forest Research* 125(4): 427-439 (2006).
  26. K. Mohammadi, O. Alavi, A. Mostafaeipour, N. Goudarzi, and M. Jalilvand. Assessing different parameters estimation methods of Weibull distribution to compute wind power density. *Energy Conversion and Management* 108: 322-335 (2016).
  27. W.L. Hung and Y.C. Liu. Estimation of Weibull parameters using a fuzzy least-squares method. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 12(05): 701-711 (2004).
  28. P. Malik, A. Gehlot, R. Singh, L. R. Gupta, and A.K. Thakur. A review on ANN based model for solar radiation and wind speed prediction with real-time data. *Archives of Computational Methods in*

- Engineering* 29(5): 3183-3201 (2022).
29. A. Tahir, M. Ashraf, A. Razzak, S.M. Raza, and Z. Uddin. Temperature data of Hyderabad from the temperature of three neighboring cities using the ANN and the multiple regression methods. *Kuwait Journal of Science* 50(3A): 147-161 (2023).
  30. W. Shepherd and L. Zhang (Eds.). Electricity Generation Using Wind Power. *World Scientific* (2017).
  31. J.A. Carta, P. Ramirez, and C. Bueno. A joint probability density function of wind speed and direction for wind energy analysis. *Energy Conversion and Management* 49(6): 1309-1320 (2008).
  32. M. A. Hussain, S. Abbas, M.R.K. Ansari, A. Zaffar, and B. Jan. Wind speed analysis of some coastal areas near Karachi. *Proceedings of the Pakistan Academy of Sciences* 51(1): 83-91 (2012).
  33. M. Arashi, P. Nagar, and A. Bekker. Joint probabilistic modeling of wind speed and wind direction for wind energy analysis: A case study in humansdorp and noupoort. *Sustainability* 12(11): 4371 (2020).
  34. Z. Wang and W. Liu. Wind energy potential assessment based on wind speed, its direction and power data. *Scientific Reports* 11(1): 16879 (2021).
  35. A. Narain, S.K. Srivastava, and S.N. Singh. The impact of wind direction on wind farm power output calculation considering the wake effects of wind turbines. *Wind Engineering* 47(1): 74-85 (2023).
  36. S.A. Mata, J.P. Martínez, J.B. Quesada, F.P. Larrañaga, N. Yadav, J.S. Chawla, V. Sivaram, and M.F. Howland. Modeling the effect of wind speed and direction shear on utility-scale wind turbine power production. *Wind Energy* 27(9): 873-899 (2024).
  37. M.A. Ghorbani, R. Khatibi, B. Hosseini, and M. Bilgili. Relative importance of parameters affecting wind speed prediction using artificial neural networks. *Theoretical and Applied Climatology* 114(1): 107-114 (2013).
  38. M. Jamil and M. Zeeshan. A comparative analysis of ANN and chaotic approach-based wind speed prediction in India. *Neural Computing and Applications* 31(10): 6807-6819 (2019).
  39. G. Li and J. Shi. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy* 87(7): 2313-2320 (2010).