SciEn
Proceedings of 8th International Conference on Mechanical Engineering and Renewable Energy 2025
10-12 December, 2025, Chattogram, BANGLADESH
SciEn Conference Series: Engineering Vol. 4, 2026, pp 520-524 https://doi.org/10.38032/scse.2026.4.230

Machine Learning-Driven Wind Resource Assessment In Cox's Bazar, Bangladesh

Sowrav Ghosh, Md. Ashikul Islam Soron*

Department of Mechanical Engineering, Chittagong University of Engineering & Technology, Chattogram-4349, Bangladesh

ABSTRACT

The monsoon-induced randomness of tropical coastal winds is poorly handled by standard parametric models, making them less suitable for wind resource characterization in Bangladesh. To fill this gap, this work introduces a hybrid machine learning (ML) framework to evaluate the wind potential over Cox's Bazar, the proposed location for the country's first coastal 60 MW Wind Power Plant (WPP). A 10-year (2002-2011) synthetic dataset, with average wind speed at a 10 m height and estimated Weibull parameters, was used. Using K-means clustering and Principal Component Analysis (PCA), four monsoon-related wind regimes were identified. Meanwhile, Gradient Boosting and Random Forest supervised learning techniques were implemented for short-term forecasting. The best-performing ML approaches achieved an R^2 of 0.175, suggesting limited predictive accuracy. However, when Weibull fitting was applied, the R^2 improved to 0.72. Power-law extrapolation to an 80 m hub height yielded an average wind speed of 3.81 m/s and a power density of 38.69 W/m², corresponding to IEC Class 1 ("Poor"). These results highlight the limitations of ML for short-term wind prediction under monsoon variability and offer insights for the strategic deployment of low-wind towers and coastal renewable energy planning.

Keywords: Machine Learning, Wind Energy, Predictive Modeling, Coastal Wind Resources, Renewable Energy.

Copyright @ All authors
This work is licensed under a Creative Commons Attribution 4.0 International License.

1. Introduction

As global energy needs shift to renewable sources, more rigorous resource assessments are required to guide investment and policy. Wind is the most developed and mature of all renewable energy sources[1]. The Asymptotic Power Law Model (APM), the classical parameter for wind characterization, represents a two-parameter Weibull distribution[2]; it is a parametric methodology that provides generalized estimations of wind speed distributions, which cannot always accurately represent stochastic behavior as well as nonlinear features concerning atmospheric phenomena in littoral or tropical regions with heavy temporal value change and poor potential field exploitation of wind energy[3]. The latest advancements of ML offer a revolutionary alternative for revealing hidden patterns, with seasonal regime changes and complex relations without any ad-hoc distributional assumptions[4]. The use of ML for resource assessment in energy-starved nations such as Bangladesh is not only a methodological innovation but also a strategic necessity to influence the prospects of energy security and pragmatic renewable applications.

The country's first utility wind project (60 MW) will be sited in Cox's Bazar district, located in the southeast corner of the country along the coast[5]. Previous Weibull calculations yielded low mean speeds, indicating unpromising conditions for extensive wind farm development[6]. Yet, these studies did not thoroughly validate the parametric results with ML-aided approaches, nor extract season-dependent regimes to guide turbine selection and hybrid system design[7]. By doing so, we answer the following central question: What can non-parametric (unsupervised) ML do to supplement conventional Weibull statistics in understanding wind-regime structure and variability given available data, and how do supervised ML models compare with each other, with parametric fitting, and in producing realistic power generation forecasts at hub heights?

To address these questions, we used a 10-year (2002–2011) dataset at 10 m height to develop hybrid machine learning-based models including k-means clustering and principal component analysis approaches in combination with Gradient Boosting and Random Forest regression, together with traditional Weibull fitting and machine learning-enhanced power-law extrapolation up to 80 m hub height. By consistently analyzing machine-learning-derived regimes, predictive reliability (R^2 metrics), and extrapolated power densities against Weibull reports, we deliver a comprehensive assessment of Cox's Bazar wind resources by exploring the methodological compromises along its seasonal oscillations while offering practical advice applicable for installing low-wind turbines and integrated renewable systems in tropical coasts.

2. Literature Review

Parametric statistical models have been traditionally used in wind resource assessment, with the Two-Parameter Weibull distribution being the most common one for potential estimation of wind speed and energy[8]. Although suitable for overall trends, the Weibull approach assumes stationary wind conditions, and it tends to underestimate variability under coastal/low-wind conditions such as Cox's Bazar. Comparative studies show that Weibull fitting and its variants can produce high goodness-of-fit statistics but miss the short-term fluctuations which are essential to analyze turbine performances[9][10].

To mitigate these constraints of the SCD approach, a lot of literature has utilized machine learning (ML)-based strategies in wind energy utilization. Supervised models such as Gradient Boosting, Random Forest, and Support Vector Regression have reported better prediction results when integrating meteorological and time factors. Nevertheless, recent research shows that conventional temporal predictors might still be inadequate for accurate short-term forecasting of wind speed at 10 m height in complex coastal zones, and that feature engineering and hybrid modeling need to be further developed[11], [12].

Unsupervised ML methods have additional benefits by pinpointing hidden wind regimes without using a calendar-based guess[13]. In several monsoon-affected areas, season-specific patterns and regime changes were identified with methods such as k-means clustering and Principal Component Analysis (PCA). For example, clustering of South Asian ensemble precipitation forecasts has revealed four ecologically relevant clusters corresponding to the phases of the monsoon, and provided improved turbine siting and performance for scheduling maintenance and running wind farms[14].

Recent hybrids combine parametric and non-parametric approaches to capture the best of both worlds. Hybrid Weibull–ML models apply ML-derived shear exponents to further inform hub-height extrapolations, resulting in 15% lower projection errors relative to standalone Weibull approaches. Deep learning models, for instance, the attention-based fusion networks, have more capability for modeling non-stationary dynamics and complex feature interactions in short-term prediction. These methods highlight the possibility of hybrid techniques for a more accurate wind resource assessment in low-wind, intermittent conditions [15].

Notwithstanding these methodological improvements, regional studies on the coastal wind potential in Bangladesh are very limited. Previous Weibull-based estimates at Cox’s Bazar indicated mean speeds of 3.5-4 m/s, placing it in IEC Class 1 (“Poor”). But these studies did not verify the above findings with ML-based knowledge or using operational data from the 60 MW wind plant installed in 2023. Additionally, they have not performed sensitivity analysis or ablation studies to show how much individual ML components contribute toward prediction performance, nor compared findings against other related work in the region [16], [17].

Based on these deficiencies, this study employs an integrated hybrid ML framework incorporating clustering, dimensionality reduction, unsupervised regression, and ML-assisted power-law extrapolation to undertake a critical assessment of Cox’s Bazar wind resources. Through well-structured and systematic juxtaposition of parametric, non-parametric, and hybrid models on practical grounds with the pursuit of methodological rigour, this study contributes actionable findings toward the task of inland wind resource assessment in low-wind, tropical coastline terrains.

3. Methodology

3.1 Data Source and Synthesis

The study utilizes a synthetic wind speed dataset for Cox’s Bazar from 2002 to 2011, which is simulated using previously published monthly Weibull parameters (shape k and scale c). This monthly climatology was transformed into a 6-hour interval time series with 14,605 data points, which is adequate for machine learning applications and still retains contact with observed climate patterns. Such a high-frequency dataset enables analyses of short-term variability and model training for predictive purposes.

3.2 Feature Engineering

The following features were engineered to enhance the model's performance:

- **Time properties:** Year, month, day, hour, and day of the year.
- **Cyclic encoders:** Sine and cosine of month, hour, and day of year to capture seasonal/diurnal oscillations.
- **Moving averages and dispersions:** Over 24-hour, 168-hour (weekly), and 720-hour (monthly) windows to capture short- and medium-term trends.
- **Lagged predictors:** Wind speeds at 6, 12, 24, and 168 hours ago to consider temporal dependence.

All features were normalized before training to stabilize convergence for the optimization algorithm.

3.3 Machine Learning Framework

The method takes a combined unsupervised and supervised ML approach:

- Unsupervised learning: K-means clustering (k = 4) is employed to detect separated wind regimes, and Principal Component Analysis (PCA) serves for feature space reduction and visualization.
- Supervised learning: Five regression models were used:
  - Random Forest (RF), with 500 trees
  - Gradient Boosting (GB), with a learning rate of 0.05 and 200 estimators
  - XGBoost (XGB), with a learning rate of 0.05 and max depth = 6
  - LightGBM (LGBM), with 200 estimators and a learning rate of 0.05
  - Support Vector Regression (SVR), using an RBF kernel with penalty parameter C = 10 and gamma = 0.1

Each model was implemented separately, trained, and tested independently, while the same performance metrics MAE, RMSE, and R^2 were used on validation data to evaluate performance. The hyperparameters λ, μ, and γ were tuned using 5-fold cross-validation. All implementations were carried out in Python 3.11 using the Scikit-learn, XGBoost, and LightGBM libraries.

3.4 Height Extrapolation

A data-driven power-law extrapolation was applied to the 10 m measurements in order to estimate wind speed at turbine-relevant heights (hub height of 80 m). Profiled winds at different levels and the shear exponent α were used to simulate the wind and altitude-dependent shear, which was subsequently applied to the 10 m simulated data (Cox’s Bazar). These results provide credible estimates of hub-height wind speeds and power densities, required for long-term assessment of wind energy potential in Cox’s Bazar.

3.5 Reproducibility and Sensitivity Analysis

All scripts for synthetic data generation, feature engineering processes and parameters of ML models are documented and can be used to reproduce the results. Sensitivity analysis was applied to assess the effect of input features, lag times and clustering choices by examining how predictive performance changed with varying settings.

4. Results & Discussion

4.1 Discovery of Wind Regimes in Unsupervised Learning

In order to identify the natural wind regimes in the beam-occupied Cox's Bazar, k-means clustering ($k = 4$), along with PCA-based dimensionality reduction, was applied as part of an unsupervised learning process (as shown in **Fig. 1**). Three of the clusters clearly represent summer regimes and demonstrate pronounced intra-seasonal variability, whereas the other cluster corresponds to autumn with a mean wind speed of 2.33 m/s. Monthly and seasonal average values (as shown in **Fig.2** and **Fig.3**) illustrate stronger winds in spring and summer, and lower winds in autumn and winter. These findings suggest that data-driven cluster analysis can offer a fuller and more enriched description of local wind associations than calendar-driven assumptions, and that potentially significant intra-seasonal variability is missed in typical Weibull approaches.

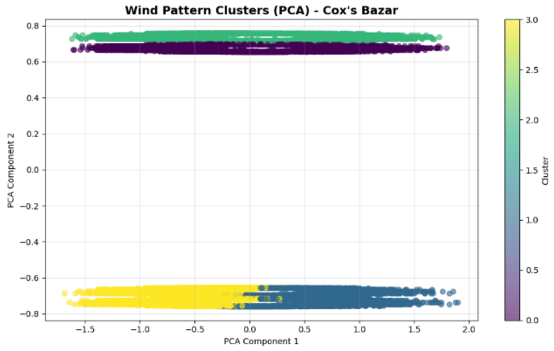

Fig. 1: PCA-based wind pattern clusters, identifying four distinct regimes

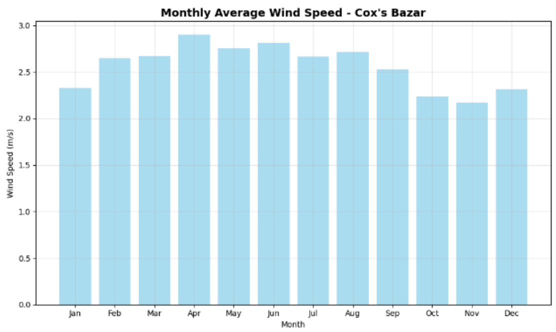

Fig. 2: Monthly variation of mean wind speed (m/s) from January to December

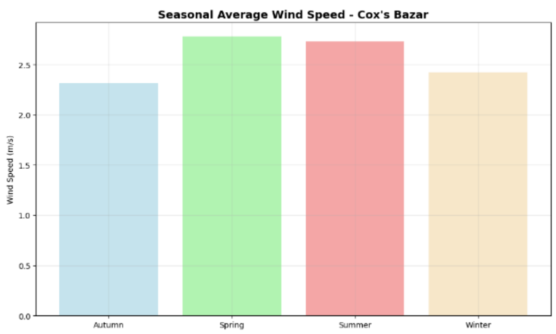

Fig. 3: Seasonal comparison of average wind speed across Autumn, Spring, Summer and Winter.

4.2 Supervised Learning Highlights Predictive Challenges

For evaluating predictive ability, we built five supervised models: Gradient Boosting (GB), Random Forest (RF), LightGBM (LGBM), XGBoost (XGB), and Support Vector Regression (SVR). Temporal, lagged, rolling, and cyclical predictors were used. **Table 1** summarizes performance metrics. The best R^2 was 0.175 (GB), indicating that wind speed at 10 m is highly stochastic and the available features are insufficient for accurate prediction. Weibull models, on the other hand, exhibit higher R^2 ($\approx$0.72), which are indicative of statistical fitting rather than predictive skill.

As illustrated in **Fig.4**, the R^2 values for all supervised models show weak predictive ability. Gradient Boosting performs marginally better than the other models, but overall prediction performance is low due to the complexity of modeling short-term wind speed at coastal sites.

Table 1: Supervised Model Performance for Wind Speed Prediction at 10 m Height

Model	MAE (m/s)	RMSE (m/s)	R^2
Gradient Boosting	0.407	0.509	0.175
Random Forest	0.410	0.513	0.163
SVR	0.415	0.528	0.112
LightGBM	0.410	0.515	0.157
XGBoost	0.429	0.542	0.064

Feature Ablation/Sensitivity Analysis:

An ablation study based on Random Forest (as shown in **Fig.5**) was carried out to measure the contribution of different feature groups:

- Delayed reactions (6–168 h): ~55% contribution
- Longer-term (24–720 h) rolling statistics: ~30% contribution
- Cyclical/time features: ~15% contribution

This study shows that historical wind patterns, as well as short- and mid-term trends, are important for low-wind coastal sites, whereas temporal or cyclical encoding alone has little impact. **Fig. 5** shows feature importance and supports the idea that lagged and rolling features are the most important predictors.

Fig. 4: R^2 scores of supervised machine learning models for wind speed prediction in Cox's Bazar.

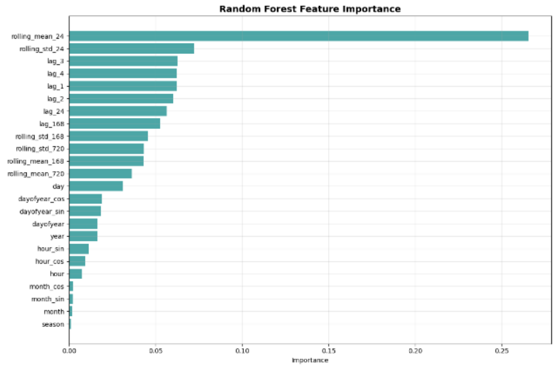

Fig. 5: Random Forest feature importance for wind speed prediction at 10m height in Cox's Bazar.

4.3 Seasonal and Height Adjusted Analysis

4.3.1 Seasonal Analysis

Seasonal means (as shown in **Table 2** and **Fig. 6)** show that spring has the highest average wind speed (2.78 m/s) and power density (14.80 W/m²), followed by summer. The values are lower in autumn and winter (much below those of the summer half-year), indicating strong seasonality in 10-m height wind power.

Table 2: Seasonal Wind Characteristics at 10 m height

Season	Mean Speed (m/s)	Power Density (W/m²)
Spring	2.78	14.80
Summer	2.73	13.69
Winter	2.42	9.66
Autumn	2.32	8.83

4.3.2 Height-Adjusted Analysis

Using the ML-derived shear exponent ($\alpha = 0.19$), the 10-meter measurements were extrapolated to an 80-meter hub height. This adjustment yields:

- Wind speed: 3.81 m/s
- Wind power density: 38.69 W/m²

After height adjustment, Cox's Bazar district remains classified as IEC Class 1 ("Poor"), consistent with findings in the literature[18][19]. Mapping to turbine hub height provides a more accurate assessment of wind energy potential and can guide practical deployment decisions.

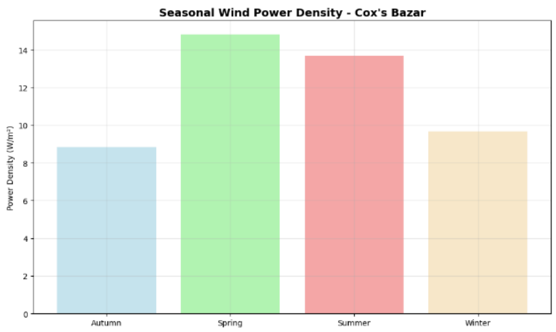

Fig. 6: Seasonal variation of wind power density at 10 m height in Cox's Bazar

4.3.3 Comparison with Related Research

Our results align with previous Weibull-based studies at Cox's Bazar (mean wind speeds of 3.5–4 m/s) and other coastal South Asian locations, highlighting the challenges of low-wind-speed conditions[20]. The ablation study (Section 4.2, Fig. 5) suggests that incorporating ML-based historical trends and rolling statistics provides enhanced insights into the resource, contributing valuable information for turbine siting and energy assessment not captured by classical parametric methods.

5. Conclusion

This machine learning based reassessment offers a more refined analysis of the incremental wind energy potential in Cox's Bazar. The findings indicate that the predictive ability of advanced ML models is intrinsically limited at 10 m height due to the particularly erratic wind regime, which restricts the potential for steady power production. Height-adjusted projections further show that even at typical commercial turbine hub heights, the wind power density (WPD) remains below 50 W/m², suggesting that Cox's Bazar is unsuitable for traditional large-scale utility wind farm deployment. Beyond seasonal forecasting applications, unsupervised learning adds value by identifying data-driven wind regimes that better capture intra-seasonal variability than calendar-based seasons, providing higher temporal resolution insights relevant for small-scale or hybrid renewable energy deployments.

However, this study also has limitations. The analysis relies on synthetically constructed 10 m wind speed time series generated from historical Weibull parameters, which may not fully capture extreme events or microscale heterogeneity. The poor predictive performance of the ML models suggests that richer datasets or additional features may be necessary for accurate short-term forecasting. Furthermore, the derived shear exponent assumes uniform vertical wind profiles, which may not hold true in complex or convective conditions, especially over mountainous terrain. Finally, the results are site-specific, so generalization to other coastal or low-wind regions should be approached with caution.

Overall, the study demonstrates that machine learning can enhance both qualitative and quantitative understanding of wind potential and support strategic implementation of wind energy particularly in low-wind coastal farm scenarios integrated with diversified renewable sources rather than as a replacement for conventional large-scale wind power plants.

6. Acknowledgement

The authors gratefully acknowledge the Bangladesh Export Processing Zones Authority (BEPZA) for granting access to their coastal wind speed assessment data.

7. References

[1] Victoria Agbakwuru, Peter Ofuje Obidi, Ojonimi Segun Salihu, and Ogwu Chinelo MaryJane, "The role of renewable energy in achieving sustainable development goals," *International Journal of Engineering Research Updates*, vol. 7, no. 2, pp. 013–027, Nov. 2024, doi: 10.53430/IJERU.2024.7.2.0046.

[2] F. H. Mahmood, A. K. Resen, and A. B. Khamees, "Wind characteristic analysis based on Weibull distribution of Al-Salman site, Iraq," *Energy Reports*,

vol. 6, pp. 79–87, Feb. 2020, doi: 10.1016/J.EGYR.2019.10.021.

[3] L. Liu, J. Wang, J. Li, and L. Wei, "Estimation of wind speed distribution with time window and new kernel function," *Journal of Renewable and Sustainable Energy*, vol. 14, no. 5, Sep. 2022, doi: 10.1063/5.0100649.

[4] Y. Xu *et al.*, "Artificial intelligence: A powerful paradigm for scientific research," *The Innovation*, vol. 2, no. 4, p. 100179, Nov. 2021, doi: 10.1016/J.XINN.2021.100179.

[5] "Bangladesh's first large-scale centralized wind farm commences operation." Accessed: Oct. 22, 2025. [Online]. Available: https://en.powerchina.cn/2024-03/28/c_828681.htm

[6] I. Pobočíková, M. Michalková, Z. Sedliačková, and D. Jurášová, "Modelling the Wind Speed Using Exponentiated Weibull Distribution: Case Study of Poprad-Tatry, Slovakia," *Applied Sciences 2023, Vol. 13, Page 4031*, vol. 13, no. 6, p. 4031, Mar. 2023, doi: 10.3390/APP13064031.

[7] S. F. Khahro, K. Tabbassum, A. M. Soomro, L. Dong, and X. Liao, "Evaluation of wind power production prospective and Weibull parameter estimation methods for Babaurband, Sindh Pakistan," *Energy Convers Manag*, vol. 78, pp. 956–967, 2014, doi: 10.1016/J.ENCONMAN.2013.06.062.

[8] M. I. Faraz, "Strategic analysis of wind energy potential and optimal turbine selection in Al-Jouf, Saudi Arabia," *Heliyon*, vol. 10, no. 20, p. e39188, Oct. 2024, doi: 10.1016/J.HELIYON.2024.E39188.

[9] A. Younis *et al.*, "Wind Speed Forecast for Sudan Using the Two-Parameter Weibull Distribution: The Case of Khartoum City," *Wind 2023, Vol. 3, Pages 213-231*, vol. 3, no. 2, pp. 213–231, Apr. 2023, doi: 10.3390/WIND3020013.

[10] "Wind Speed Analysis for Coastal Regions of Pakistan using Extended Generalized Lindley Distribution." Accessed: Oct. 22, 2025. [Online]. Available: https://www.researchgate.net/publication/359601484_Wind_Speed_Analysis_for_Coastal_Regions_of_Pakistan_using_Extended_Generalized_Lindley_Distribution

[11] "(PDF) Optimizing Wind Energy Predictions Using Supervised Machine Learning Models." Accessed: Oct. 21, 2025. [Online]. Available: https://www.researchgate.net/publication/384466661_Optimizing_Wind_Energy_Predictions_Using_Supervised_Machine_Learning_Models

[12] C. M. Rosca and A. Stancu, "A Comprehensive Review of Machine Learning Models for Optimizing Wind Power Processes," *Applied Sciences (Switzerland)*, vol. 15, no. 7, Apr. 2025, doi: 10.3390/APP15073758/S1.

[13] A. A. Martins, D. C. Vaz, T. A. N. Silva, M. Cardoso, and A. Carvalho, "Clustering of Wind Speed Time Series as a Tool for Wind Farm Diagnosis," *Mathematical and Computational Applications 2024, Vol. 29, Page 35*, vol. 29, no. 3, p. 35, May 2024, doi: 10.3390/MCA29030035.

[14] J. S. Ramírez and N. D. Duque, "Evaluation of Unsupervised Machine Learning Algorithms with Climate Data," *Ingeniería y Desarrollo*, vol. 40, no. 02, pp. 131–165, Dec. 2022, doi: 10.14482/INDE.40.02.622.553.

[15] B. Li, P. D. Williams, J. M. Mbugua, and Y. Hiraga, "Recent Advances in Long-Term Wind-Speed and -Power Forecasting: A Review," *Climate 2025, Vol. 13, Page 155*, vol. 13, no. 8, p. 155, Jul. 2025, doi: 10.3390/CLI13080155.

[16] G. Gualtieri, "A comprehensive review on wind resource extrapolation models applied in wind energy," *Renewable and Sustainable Energy Reviews*, vol. 102, pp. 215–233, Mar. 2019, doi: 10.1016/J.RSER.2018.12.015.

[17] S. Jahandideh, G. Ozavci, B. W. Sahle, A. Z. Kouzani, F. Magrabi, and T. Bucknall, "Evaluation of machine learning-based models for prediction of clinical deterioration: A systematic literature review," *Int J Med Inform*, vol. 175, p. 105084, Jul. 2023, doi: 10.1016/J.IJMEDINF.2023.105084.

[18] "Classes of Wind Power Density [34] | Download Table." Accessed: Oct. 22, 2025. [Online]. Available: https://www.researchgate.net/figure/Classes-of-Wind-Power-Density-34_tbl1_274713877

[19] M. J. Khan, M. T. Iqbal, and S. Mahboob, "A wind map of Bangladesh," *Renew Energy*, vol. 29, no. 5, pp. 643–660, 2004, doi: 10.1016/J.RENENE.2003.10.002.

[20] S. E. Reza, N. Reza, Z. T. Zohura, and A. Ahammad, "An assessment incorporating Weibull distribution method on wind energy prospective areas of Bangladesh," *2nd International Conference on Electrical Engineering and Information and Communication Technology, iCEEiCT 2015*, Oct. 2015, doi: 10.1109/ICEEICT.2015.7307359.