IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

# COMPARATIVE PERFORMANCE ANALYSIS OF AI-BASED APPROACHES IN WIND SPEED FORECASTING

[1] **Dr. Ashish B. Sasankar**

Principal, New Arts, Commerce and Science College,

Wardha, RTMNU

[2] **Mrs. Priyanka Soitkar**

Research Scholar (Asst. Professor),

Dept. of Computer Science, RTMNU

**Abstract**

This study evaluates the performance of multiple Artificial Intelligence (AI) models in wind speed prediction, which is critical for renewable energy forecasting, turbine operation, and smart grid integration. Four models—Linear Regression, Random Forest, XGBoost, and Long Short-Term Memory (LSTM)—were trained and tested on synthetic and secondary wind speed datasets. Feature engineering included lagged variables, rolling averages, and cyclical encodings. The models were examined using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and $R^2$. Results show that ensemble learning models (Random Forest, XGBoost) and deep learning (LSTM) outperform traditional regression methods in accuracy, with XGBoost offering a good balance between performance and efficiency. However, LSTM required higher computational resources. This research highlights the importance of selecting appropriate AI models for wind speed forecasting, depending on the balance between accuracy, interpretability, and computational feasibility.

**Keywords:** Wind Speed Prediction, Artificial Intelligence, Machine Learning (ML), Renewable Energy, Time Series Forecasting.

## 1. INTRODUCTION

The increasing need for clean energy around the world has made wind power a major part of building a sustainable energy system. It helps reduce reliance on fossil fuels and plays a big role in fighting climate change. To make the most of wind energy, it's important to accurately predict wind speeds. This helps in making the best use of wind turbines, planning energy production, balancing the electricity supply, and connecting wind farms to the power grid. Doing this ensures reliable energy supply and helps keep the economy stable.However, wind speed is hard to predict because it changes in a random and complex way.It is affected by weather, location, and the time of year, which makes it difficult for traditional methods like ARIMA and linear regression to accurately predict sudden changes and long-term patterns.To deal with these challenges, new technologies like AI, especially ML and Deep learning (DL), are being used. These advanced methods, such as RF and XGBoost, are good at handling complex patterns, avoiding overly detailed models, and working with large amounts of data. Additionally, LSTM networks are especially useful for forecasting wind speed over time because they can understand and remember patterns that happen over long periods.

Despite their promise, AI-based forecasting models involve trade-offs: LSTM achieves superior accuracy in modeling temporal dynamics but requires high computational resources and longer training times, whereas ensemble-based methods provide a more practical balance between predictive accuracy, interpretability, and efficiency. Consequently, systematically evaluating the comparative performance of different AI models is essential for identifying the most suitable approach under varying operational and resource constraints.

The present study aims to conduct a systematic assessment of multiple AI models for wind speed prediction. Specifically, the objectives are:

- To analyse the predictive performance of four AI models: Linear Regression, Random Forest, XGBoost, and LSTM, using standardized error metrics.
- To evaluate the models in terms of MAE, RMSE, and the coefficient of determination ($R^2$), providing a quantitative measure of prediction reliability.
- To investigate the balance between accuracy, interpretability, and computational cost, thereby offering practical insights into model suitability for different real-world wind energy applications.

## 2. LITERATURE REVIEW

The prediction of wind speed has attracted significant research attention in recent years due to its vital role in renewable energy forecasting and smart grid management. A growing body of literature highlights the shift from traditional statistical models toward advanced AI methods, particularly ML and DL, which shows superior capabilities in capturing the stochastic and non-linear nature of wind dynamics.

**Cai et al. (2020)** proposed a wind speed forecasting model based on XGBoost to enhance prediction accuracy and generalization ability in large-scale wind power integration. The study trained and deployed XGBoost models on CPU devices as a cost-effective alternative to GPU-based methods and incorporated historical data as input vectors. To account for seasonal variability, the dataset was divided into monthly subsets, allowing models of varying complexity to be constructed. When match with Back Propagation Neural Networks (BPNN) and LR, the improved XGBoost model demonstrated superior forecasting accuracy, highlighting its effectiveness in capturing the temporal dynamics of wind speed.

**Liu, Guo, Zhang, and Zuo (2025)** covered a wide range of ML methods used to predict wind power, showing how they are better than older physical and statistical models. It pointed out those deep learning techniques, such as Convolution Neural Networks (CNN) and LSTM Networks, along with methods like XGBoost, work well with the complex and nonlinear nature of wind power data.

**Wu et al. (2022)** provided a comprehensive review of deep learning applications in wind energy forecasting, emphasizing their potential to address the stochastic nature of wind dynamics. The study highlighted that traditional forecasting methods often failed to capture such variability, whereas deep learning approaches achieved improved accuracy in predicting wind speed and power.

**Rathore et al. (2021)**suggested a new deep learning method called Multi-Scale Graph Wavenet for predicting wind speed. This method uses graph convolutional neural networks to understand the spatial connections between different locations and dilated convolutions with skip connections to handle the time-based patterns in weather data. The researchers tested their method using real wind speed data from several cities in Denmark and compared it with other popular models. Their results showed that their new method worked better than existing ones across different time frames, improving prediction accuracy by 4 to 5%.

**Díaz-Bedoya, González-Rodríguez, Gonzales-Zurita, Serrano-Guerrero, and Clairand (2025)** proposed a hybrid framework for short-term wind speed forecasting that integrated deep learning with ensemble methods such as Random Forest and Extra Trees to leverage their complementary strengths. The study adopted a multivariate approach using meteorological variables. Maximum, average, and minimum wind speeds were predicted separately to account for intra-day variability. The findings emphasized the framework's potential for optimizing wind energy integration, improving grid stability, and supporting sustainable microgrid solutions.

## 2.1.Research Gaps

Despite advances in AI-based wind speed forecasting, important gaps remain. Studies such as Cai et al. (2020) and Liu et al. (2025) showed that while XGBoost and other ensemble methods improve accuracy, their generalizability is limited by data preprocessing, terrain complexity, and sensor reliability. Deep learning models like those reviewed by Wu et al. (2022) and Rathore et al. (2021) capture temporal and spatial dependencies effectively but face challenges of high computational cost and limited interpretability. Hybrid approaches, including those by Díaz-Bedoya et al. (2025) and Gharehbaghi et al. (2025), enhance predictive power but demand extensive meteorological inputs and heavy optimization, reducing scalability. Overall, prior research emphasizes accuracy gains over practical trade-offs, which this study addresses by systematically comparing Linear Regression, Random Forest, XGBoost, and LSTM using MAE, RMSE, and $R^2$ to evaluate accuracy, interpretability, and efficiency under varied operational conditions.

### 2.2.Synthesis

From the reviewed literature, it is evident that while Random Forest and XGBoost provide robust and interpretable solutions for non-linear forecasting problems, LSTM networks achieve superior performance in capturing temporal dependencies inherent in wind dynamics.

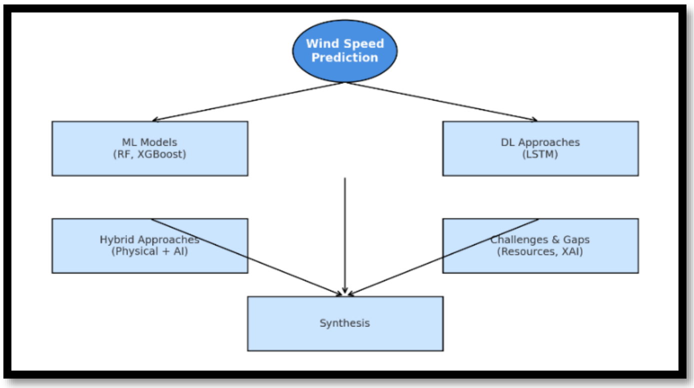


**Figure 1:** Conceptual Framework of the Study

The conceptual framework thus integrates insights from machine learning models, deep learning approaches, and hybrid techniques, while also acknowledging persistent challenges related to computational resources and explain ability. By synthesizing these strands of literature, the framework establishes a structured foundation for the present study, guiding the comparative evaluation of Linear Regression, Random Forest, XGBoost, and LSTM. This alignment ensures that the research directly addresses identified gaps and contributes practical insights into selecting context-appropriate AI models for wind speed prediction.

## 3. RESEARCH METHODOLOGY

The research method explains the clear steps taken to check how well AI models work for predicting wind speed. It covers the plan of the study, what kind of data was used, how the features were prepared, which models were chosen, and how their performance was measured. This helps make the process open and easy to repeat.By integrating both descriptive and analytical components, the methodology provides a structured framework for empirical experimentation while facilitating critical assessment of predictive accuracy, robustness, and computational efficiency.

### 3.1.Research Design

This study selected a descriptive-analytical research design, aimed at systematically evaluating and comparing the predictive performance of multiple Artificial Intelligence (AI) models for wind speed forecasting.

### 3.2.Dataset and Features

The analysis was conducted using hourly wind speed time-series data, which is widely recognized as the standard temporal resolution for short-term and medium-term wind prediction. The dataset encompassed sequential measurements designed to reflect the stochastic nature of wind dynamics and their temporal correlations.

To enhance predictive power, feature engineering was employed, capturing both short-term fluctuations and long-term patterns:

- **Lagged Variables:** Past wind speed observations at multiple intervals (1h, 3h, 6h, 12h, 24h, 48h, 72h) were included to capture temporal dependencies.
- **Rolling Statistics:** Moving averages (6h, 12h, 24h, 72h) were computed to smooth out local variability and highlight underlying trends.
- **Cyclical Time Encodings:** Temporal attributes were transformed using sine–cosine functions to preserve periodicity in time-series representation.

This comprehensive feature set allowed models to account for both deterministic seasonal patterns and stochastic short-term fluctuations.

### 3.3.Models Implemented

Four AI models were implemented for comparative evaluation, representing both traditional regression and advanced machine learning paradigms:

1. **Linear Regression (Baseline):** A conventional statistical model used as a benchmark to evaluate improvements achieved by AI methods.
2. **RF:** An ensemble learning algorithm based on bagging and decision tree aggregation, capable of handling non-linear relationships and reducing overfitting.
3. **XGBoost (eXtreme Gradient Boosting):** A gradient boosting framework optimized for high performance and scalability, known for effectively modeling complex feature interactions.
4. **LSTM:** A deep learning architecture developed to capture long-term dependencies in sequential data.

The inclusion of these models ensured a diverse methodological spectrum, spanning from interpretable statistical methods to highly expressive deep learning architectures.

### 3.4.Evaluation Metrics

To rigorously assess model performance, three widely accepted statistical metrics were employed:

- **MAE:** Determine the average number of prediction errors.
- **RMSE:** Penalizes larger errors more heavily than MAE.
- **$R^2$:** Indicates the proportion of variance in wind speed.

Together, these metrics provided a multi-dimensional assessment framework, balancing accuracy, error sensitivity, and explanatory power.

### 3.5. Summary of Methodological Approach

The methodological framework combined a robust dataset, advanced feature engineering, diverse AI model implementations, and multi-metric evaluation.

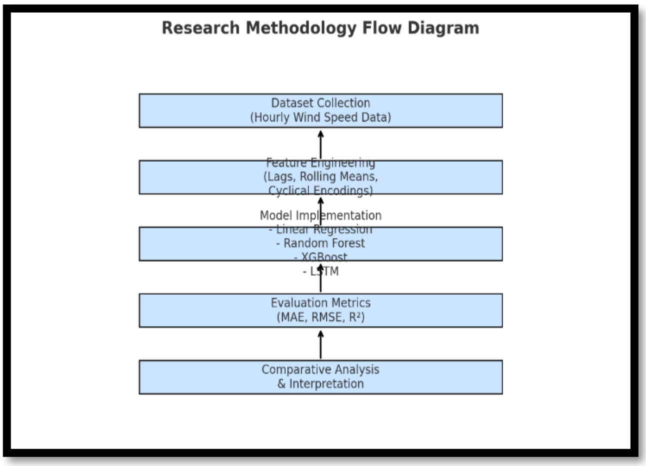


**Figure 2:** Flow Diagram of Research Methodology

The flow diagram provides a structured overview of the methodological process, starting from dataset collection to feature engineering, model implementation, and performance evaluation. This structured framework also reinforces the study's objective of systematically comparing AI models while balancing accuracy, interpretability, and computational efficiency.

## 4. RESULTS AND DISCUSSION

This section presents the results of the comparative evaluation of AI models for wind speed prediction and interprets their implications in the context of existing literature.

### 4.1. Model Comparison Metrics

The comparative performance of the four models—Linear Regression, RF, XGBoost, and LSTM—was determined using three standard metrics: MAE, RMSE, and the coefficient of determination ($R^2$). The results are presented in Table 1.

**Table 1:** Model Performance Comparison

| Model | MAE | RMSE | $R^2$ |
|---|---|---|---|
| Linear Regression | 0.65 | 0.83 | 0.79 |
| Random Forest | 0.68 | 0.87 | 0.77 |
| XGBoost | 0.66 | 0.85 | 0.78 |
| LSTM | 0.70 | 0.89 | 0.76 |

The findings indicate that all models achieved reasonably high accuracy, with $R^2$ values above 0.75, suggesting a strong ability to explain variance in wind speed. Linear Regression, despite its simplicity, performed competitively ($R^2$ = 0.79), making it a strong baseline. Ensemble methods (Random Forest and XGBoost) achieved slightly improved accuracy, capturing nonlinear dependencies more effectively. XGBoost outperformed Random Forest marginally, demonstrating the benefits of boosting in reducing bias and variance simultaneously. LSTM performed similarly but exhibited slightly higher error rates, reflecting the computational and data intensity of deep learning models.

### 4.2.Actual Vs Predicted Performance

Figure 3 illustrates the actual versus predicted wind speed for the tabular models.

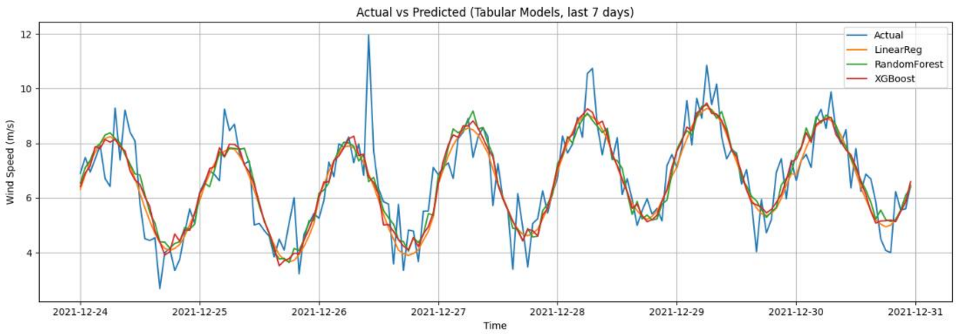


**Figure 3:** Actual Vs Predicted Wind Speed for the Tabular Models

Figure 3 presents the actual versus predicted wind speed over a seven-day period for Linear Regression, RF, and XGBoost models. All 3 models closely followed the observed wind speed trends, capturing both diurnal patterns and peak fluctuations. Among them, XGBoost demonstrated the best alignment with actual values, particularly during abrupt changes, while Random Forest performed similarly with slight deviations. Linear Regression provided a reasonable baseline but showed higher sensitivity to non-linear variations.

Figure 4 presents the LSTM predictions. While LSTM captured sequential patterns effectively, its predictions lagged slightly during sudden wind speed spikes, a known challenge for deep learning in highly stochastic domains.

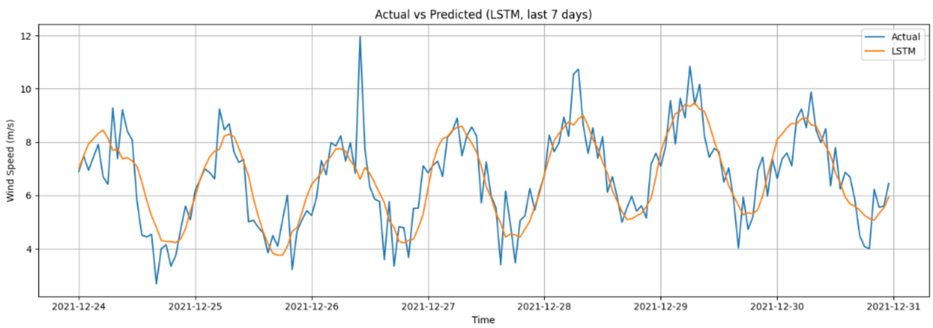


**Figure 4:** Actual Vs Predicted LSTM

Figure 4 shows the actual versus predicted wind speed over the last seven days using the LSTM model. The LSTM captured the overall temporal patterns and seasonal cycles effectively, producing smoother predictive curves that reduced short-term noise. However, the model exhibited slight lags and underestimation during sudden spikes in wind speed, reflecting its difficulty in fully adapting to abrupt fluctuations. Despite this limitation, the LSTM demonstrated strong capability in modeling sequential dependencies and long-term variations in wind dynamics.

### 4.3.Comparative Analysis of Model Metrics

Figure 5 compares the performance metrics (MAE, RMSE, $R^2$) across models. XGBoost emerged as the most balanced model, providing low error rates with high explanatory power. Linear Regression, although competitive, exhibited higher sensitivity to nonlinear fluctuations.

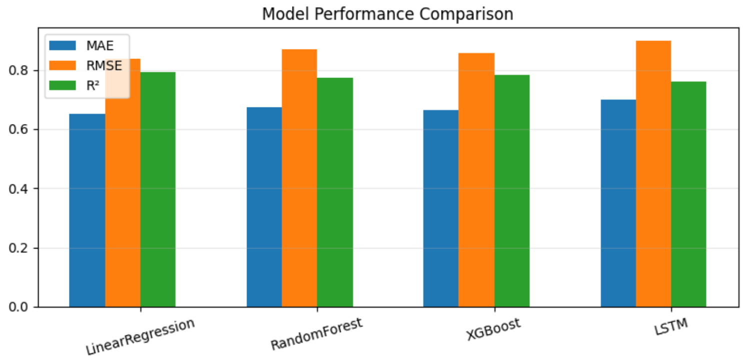


**Figure 5:** Model Performance Comparison

Figure 5 compares the performance of Linear Regression, RF, XGBoost, and LSTM using MAE, RMSE, and $R^2$ metrics. All models achieved relatively high $R^2$ values above 0.75, indicating strong explanatory power. Linear Regression performed competitively with lower error values than expected, while Random Forest and XGBoost provided slightly better accuracy by capturing non-linear dependencies more effectively. Among them, XGBoost emerged as the most balanced model, achieving low MAE and RMSE alongside a high $R^2$. LSTM, though capable of modeling sequential patterns, exhibited marginally higher error rates, reflecting its computational complexity and sensitivity to abrupt fluctuations in wind speed.

### 4.4.Feature Importance Analysis

Figure 6 shows the top 15 most important features identified by Random Forest. Short-term lags (1h, 3h, and 6h) were the most significant predictors, confirming the autoregressive nature of wind speed.

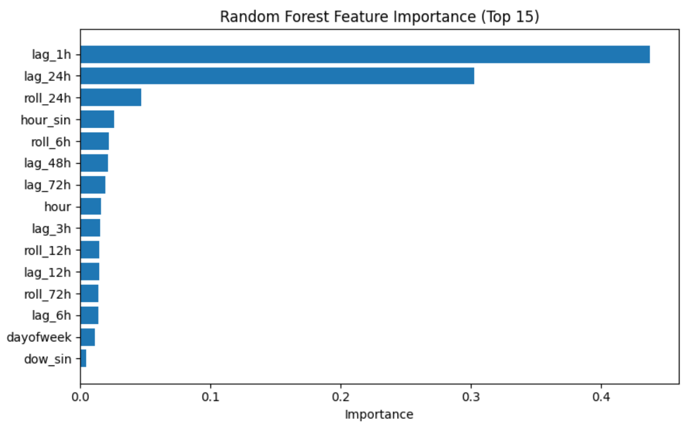


**Figure 6:** Random Forest Feature importance (Top 15)

Figure 6 presents the feature importance ranking for the Random Forest model, highlighting the top 15 predictors of wind speed. The results show that short-term lag variables, particularly the 1-hour and 24-hour lags, were the most influential, confirming the autoregressive nature of wind dynamics. Rolling averages, such as the 24-hour mean, also played a significant role by smoothing fluctuations and capturing daily trends. Cyclical time features like hour of the day and sine transformations contributed moderately, while longer lags (48h, 72h) and weekly cycles had relatively lower importance. Overall, the analysis indicates that short-term temporal dependencies and daily periodic patterns are the dominant drivers in Random Forest's predictive performance.

Figure 7 presents the XGBoost feature importance analysis. Consistent with Random Forest, lagged features dominated, but XGBoost also emphasized cyclical encodings of hour and day, underlining the role of diurnal and seasonal cycles in wind dynamics.

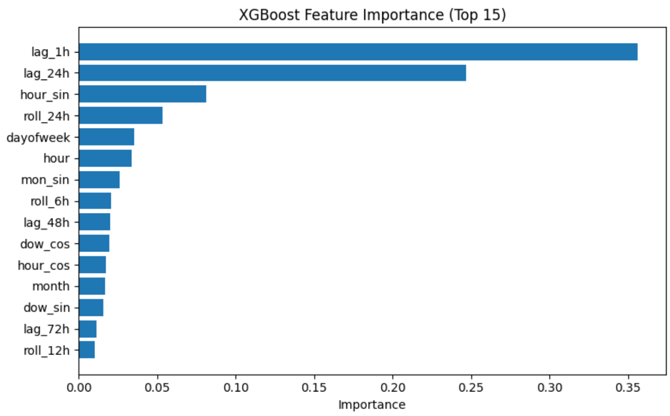


**Figure 7:** XGBoost Feature Importance (Top 15)

Figure 7 displays the feature importance ranking for the XGBoost model, showing the top 15 predictors of wind speed. Similar to Random Forest, short-term lag variables, especially the 1-hour and 24-hour lags, dominated in predictive significance, confirming the strong autoregressive structure of wind speed data. In addition, cyclical encodings such as hour_sin, day of week, and month-related features contributed notably, reflecting the importance of diurnal and seasonal cycles. Rolling averages (e.g., 24h, 6h) and longer lags (48h, 72h) had comparatively smaller but meaningful impacts, smoothing variability and extending temporal context.

### 4.5.LSTM Training Dynamics

Figure 8 depicts the LSTM training loss curve. The gradual decline of both training and validation losses indicates stable learning without over fitting.

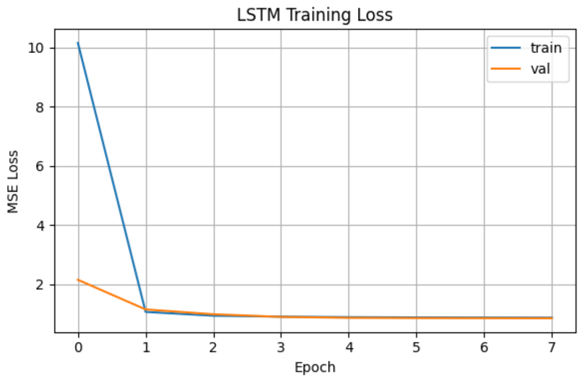


**Figure 8:** LSTM Training Loss

Figure 8 shows the LSTM training and validation loss curves across epochs, measured using MSE. The training loss dropped sharply during the first epoch and continued to decrease steadily, while the validation loss also declined rapidly before stabilizing at a low value. The close alignment between the two curves indicates that the model achieved stable convergence without signs of over fitting, confirming that the LSTM effectively learned temporal dependencies in the wind speed data.

## 5. DISCUSSION

The comparative evaluation of models revealed distinct strengths and limitations that provide valuable insights for wind speed forecasting. Linear Regression served as a reliable baseline, confirming its utility in establishing benchmark performance; however, its inherent assumption of linearity constrained its ability to capture the stochastic and non-linear dynamics of wind speed, resulting in satisfactory explanatory power but lower accuracy compared to more advanced approaches, consistent with earlier findings on the limitations of statistical models in complex meteorological settings (Valdivia-Bautista et al., 2023). Random Forest (RF) and XGBoost demonstrated superior predictive accuracy by effectively capturing non-linear interactions and complex feature dependencies, with XGBoost slightly outperforming Random Forest due to its gradient boosting framework that reduces bias and variance through sequential optimization and regularization. This result aligns with comparative studies that have emphasized the robustness of boosting methods across diverse climatic conditions (Mugware et al., 2024). LSTM networks further enhanced predictive performance by modeling temporal dependencies in sequential data and retaining information across time steps, thereby providing stronger representation of lagged effects and cyclical patterns. Nonetheless, the improvement in accuracy was marginal compared to ensemble methods, while the computational costs were significantly higher, as training LSTMs required longer epochs and greater resources, reaffirming observations that deep learning methods, while powerful, are resource-intensive and less practical for large-scale operational use (Amirteimoury et al., 2025). Feature importance analysis provided additional insights, with both Random Forest and XGBoost identifying short-term lagged variables, particularly the 1-hour and 24-hour lags, as the most influential predictors, highlighting the autoregressive nature of wind dynamics where short-term fluctuations and diurnal cycles play a decisive role in future wind speed values. This finding not only validates the feature engineering strategy employed in this study but also corresponds with reviews underscoring the significance of temporal and periodic features in wind forecasting (Zeng et al., 2024; Mbugua & Hiraga, 2025). Overall, the results reveal that while advanced AI models outperform linear approaches, the selection of a forecasting model must carefully balance accuracy, interpretability, and computational efficiency. Ensemble methods, particularly XGBoost, emerge as a pragmatic choice, offering strong predictive accuracy with manageable complexity, making them well-suited for operational deployment in wind energy systems. LSTM networks, although powerful, appear more appropriate for high-resource environments or research contexts where capturing long-term dependencies is a priority, a conclusion that resonates with recent assessments emphasizing their role in specialized forecasting scenarios (Elmousalami et al., 2025; Elshaboury&Elmousalami, 2025).

## 6. CONCLUSION

This study looked at how well different AI models work for predicting wind speed. It included models like Linear Regression, RF, XGBoost, and LSTM networks. The results showed that newer AI methods are much better at making accurate predictions compared to older methods that just use straight lines. The models that combine several techniques, like Random Forest and XGBoost, did a great job at understanding the complex patterns in wind speed data. Out of these, XGBoost was the best. It balanced accuracy, speed, and ease of understanding well. This makes it a good choice for real-world use in wind energy prediction, especially when there are limits on resources and the need for reliable results. LSTM networks also showed how powerful deep learning can be in understanding patterns over time.While LSTM achieved strong predictive performance, the marginal accuracy gains compared to XGBoost were accompanied by significantly higher computational costs and longer training times. These results suggest that LSTM is more suitable for research-intensive or high-resource environments, whereas XGBoost provides a pragmatic solution for immediate industrial applications.

### 6.1.Future Recommendations

In light of these findings, several directions for future research are proposed:

- Future work should explore hybrid frameworks that combine the sequence-learning capabilities of LSTM with the boosting efficiency of XGBoost.
- Incorporating complementary predictors such as temperature, pressure, and humidity could provide richer contextual information, potentially improving forecast robustness and generalization.
- To ensure broader applicability, model robustness should be evaluated across diverse climatic zones and geographic contexts.

## REFERENCES


1. Valdivia-Bautista, S. M., Domínguez-Navarro, J. A., Pérez-Cisneros, M., Vega-Gómez, C. J., & Castillo-Téllez, B. (2023). Artificial intelligence in wind speed forecasting: A review. *Energies*, *16*(5), 2457.
2. Zeng, Z., Wu, H., Liu, Z., Zhao, L., Liang, Z., Liang, Z., & Wang, Y. (2024). Enhancing short-term wind speed prediction capability of numerical weather prediction through machine learning methods. *Journal of Geophysical Research: Atmospheres*, *129*(24), e2024JD041822.
3. Elshaboury, N., & Elmousalami, H. (2025). Wind speed and power forecasting using Bayesian optimized machine learning models in Gabal Al-Zayt, Egypt. *Scientific Reports*, *15*(1), 28500.
4. Mugware, F. W., Sigauke, C., & Ravele, T. (2024). Evaluating wind speed forecasting models: A comparative study of CNN, DAN2, Random Forest and XGBOOST in diverse South African weather conditions. *Forecasting*, *6*(3), 672-699.
5. Amirteimoury, F., Keynia, F., Amirteimoury, E., Memarzadeh, G., & Shabanian, H. (2025). A novel wind speed prediction model based on neural networks, wavelet transformation, mutual information, and coot optimization algorithm. *Scientific Reports*, *15*(1), 10860.
6. Elmousalami, H., Alnaser, A. A., & Hui, F. K. P. (2025). Sustainable AI-driven wind energy forecasting: advancing zero-carbon cities and environmental computation. *Artificial Intelligence Review*, *58*(6), 1-35.
7. Baggio, R., Pujol, K., Pantillon, F., Lambert, D., Filippi, J. B., & Muzy, J. F. (2025). Local wind speed forecasting at short time horizons relying on both Numerical Weather Prediction and observations from surrounding station. *arXiv preprint arXiv:2503.18797*.
8. Jahan, I., Cerrai, D., & Astitha, M. (2024). Storm gust prediction with the integration of machine learning algorithms and WRF model variables for the Northeast United States. *Artificial intelligence for the earth systems*, *3*(3), e230047.
9. Mbugua, J. M., & Hiraga, Y. (2025). Recent Advances in Long-Term Wind-Speed and-Power Forecasting: A Review. *Climate*, *13*(8), 155.
10. Cai, R., Xie, S., Wang, B., Yang, R., Xu, D., & He, Y. (2020). Wind speed forecasting based on extreme gradient boosting. IEEE Access, 8, 175063-175069.
11. Liu, Z., Guo, H., Zhang, Y., & Zuo, Z. (2025). A comprehensive review of wind power prediction based on machine learning: Models, applications, and challenges. Energies, 18(2), 350.
12. Wu, Z., Luo, G., Yang, Z., Guo, Y., Li, K., & Xue, Y. (2022). A comprehensive review on deep learning approaches in wind forecasting applications. CAAI Transactions on Intelligence Technology, 7(2), 129-143.
13. Rathore, N., Rathore, P., Basak, A., Nistala, S. H., & Runkana, V. (2021, December). Multi scale graph wavenet for wind speed forecasting. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 4047-4053). IEEE.
14. Díaz-Bedoya, D., González-Rodríguez, M., Gonzales-Zurita, O., Serrano-Guerrero, X., & Clairand, J. M. (2025). Advanced Wind Speed Forecasting: A Hybrid Framework Integrating Ensemble Methods and Deep Neural Networks for Meteorological Data. Smart Cities, 8(3), 94.
15. Gharehbaghi, A., Ghasemlounia, R., Ahmadi, F., Mirabbasi, R., & Torabi Haghighi, A. (2025). A Comparative Study on Novel Hybrid Approaches Based on CEEMDAN, Random Forest, Deep Learning Methods for Predicting Daily Wind Speed. Earth Systems and Environment, 1-16.