ISSN 2277-2685
IJESR/April-June. 2026/ Vol-16/Issue-2/817-823
Er. Rishabh Aryan *et. al.*, / International Journal of Engineering & Science Research
Full Length Article

Performance Benchmarking of a Novel CNN–BiLSTM–Attention Hybrid Model for Short-Term Spatiotemporal Wind Speed Forecasting Using Indian SCADA Data

Er. Rishabh Aryan[1], Prof. Dr. Tryambak Hiwarkar[2]
[1]M.Tech (Artificial Intelligence and Data Science), Department of CSE, Indian Institute of Information Technology, Bhagalpur (Bihar)
[2]Director, ASM Group of Institutions, Pune, Maharashtra
E-mail: rishabh.250201011@iiitbh.ac.in , tryambakhiwarkar@asmedu.org
Accepted 08-05-2026

ABSTRACT
Short-term wind speed forecasting is fundamental to the reliable integration of variable renewable energy into modern power systems. Although numerous machine learning and deep learning models have been proposed, comprehensive benchmarking of hybrid attention-based architectures against both classical and modern baselines on real Indian SCADA data accompanied by thorough residual diagnostics and statistical validation remains scarce. This paper reports a rigorous performance benchmarking of a CNN–BiLSTM–Attention hybrid model against five established baselines: Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), Support Vector Machines (SVM), hybrid ARIMA–ANN, and standalone CNN–BiLSTM. Using a cleaned 8,760-hourly SCADA dataset from an operational onshore Indian wind turbine, the proposed hybrid achieved MAE = 0.824 m/s, RMSE = 1.146 m/s, MAPE = 13.7%, and R^2 = 0.924 on the held-out test set, corresponding to RMSE reductions of 22.4%, 18.7%, 23.8%, 24.0%, and 11.2% versus ANN, LSTM, SVM, ARIMA–ANN, and CNN–BiLSTM respectively. Residual diagnostics Q-Q plots, Ljung–Box, Shapiro–Wilk, and Breusch–Pagan tests confirm white-noise behaviour with no systematic autocorrelation or heteroskedasticity. The Diebold–Mariano test verifies that observed improvements are statistically significant ($p < 0.001$ against all baselines). Wind-regime stratification further reveals that the hybrid's advantages are most pronounced in high-wind and transitional conditions. The findings establish the CNN–BiLSTM–Attention architecture as the most accurate and statistically robust candidate for operational deployment in Indian wind corridors.

Keywords: *Performance Benchmarking, CNN–BiLSTM–Attention, RMSE Reduction, Residual Diagnostics, Indian SCADA, Spatiotemporal Forecasting, Wind Speed Prediction.*

INTRODUCTION

Wind energy has become one of the key components of the Indian renewable-energy policy, with an installed capacity of over 40 GW, and projected growth ambitions based on the non-fossil-capacity pledge of 2030. The key to grid stability, mitigation of deviation-penalty under CERC framework, and economically efficient unit commitment is the short-term wind speed forecasting. Although deep learning models have been proliferated in the area of wind forecasting, a direct and statistically significant comparison of the hybrid attention-based models with classical models and modern models using Indian SCADA data is uncommon. The majority of published papers only report a single or two comparison measures, none of them report residual diagnostics and none of them stratify error analysis by wind-speed regime. The paper fills that gap by providing a thorough performance benchmarking of CNNBILSTMAttention hybrid over five representative baselines, including ANN, LSTM, SVM, ARIMAANN and standalone CNNBILSTM. It has been four-fold contributions: (i) quantification of RMSE reductions in all the baselines including the detailed breakdown of the overall 24 percent improvement; (ii) full residual diagnostics with four complementary statistical tests; (iii) wind-regime-stratified RMSE analysis to reveal where and why the benefits of the hybrid are realized; and (iv) parameter-count and training-time accounting to contextualize Collectively these contributions not only prove that the hybrid will be more effective than its rivals but how, where, and at what cost.

LITERATURE REVIEW

Prior Benchmarking Studies on Wind Forecasting

A comprehensive literature standards predicting models in both time horizons and geography. Foley et al. (2012) overview traditional and hybrid methods, whereas Hanifi et al. (2020) survey the shift to deep learning. Kumar and Kaur (2020) give a comprehensive comparative analysis of hybrid machine-learning pipelines, whereas Liu et al. (2021) concentrate on CNNLSTM combinations. But most of these comparisons either (i) compare two or three (not more) models simultaneously, or (ii) summarize the results of heterogeneous datasets, which restricts the actionability of operations in any one region.

CNN–BiLSTM–Attention and Related Hybrid Architectures

The total results of Wang et al. (2020), Chen et al. (2021), Neshat et al. (2021), and Shahid et al. (2021) indicate that tri-layer hybrids with convolutional features extraction, bidirectional recurrence, and attention weighting are always superior to standalone recurrent or tree-based baselines on wind data. RMSE decreases reported are usually within the 15-25 percent range, depending on the characteristics of the dataset. Zhang et al. (2022) report in particular the gains of 8-12% of the attention layer. Such findings encourage the current benchmarking research yet fail to address the question of whether such benefits are also applied to Indian SCADA data when subjected to stringent statistical analysis.

Statistical Validation in Forecasting Research

The standard test to compare the predictive accuracy of two forecasts was suggested by Diebold and Mariano (1995). The canonical test of residual autocorrelation is given by Ljung and Box (1978) and the ShapiroWilk (1965) and BreuschPagan (1979) tests test normality and heteroskedasticity respectively. Even with these well-established procedures, point-estimate error measures are obtained in many recent wind forecasting papers. The current research proves how the systematic use of all four tests enhances the interpretability and reliability of the benchmarking conclusions.

METHODOLOGY

Dataset

The data set contains 8,760 hourly wind speed measurements re-sampled out of 52 592 ten-minute SCADA measurements taken at an active onshore wind turbine in India in 2018. The lack of values at about 0.2 percent of the records was filled in using linear interpolation. The series was chronologically divided into an 80% training subset (7,008 samples) and 20% test subset (1,752 samples).

Sliding-Window Framing

An identical sliding-window formulation was used by all models: the input vector consisted of the last 24 hourly observations and the target was the wind speed in an hour. The reason behind this choice of window length was the high autocorrelation of 24 hours of the training data in the ACF.

Model Configurations

The five baselines were: ANN - a two-hidden-layer feedforward network with 64 and 32 units and ReLU activations; LSTM - one layer recurrent network with 64 hidden units; SVM - Support Vector Regression with an RBF kernel and grid-searched hyper-parameters; ARIMAANN - a hybrid where ARIMA models the linear part and an ANN models the non-linear The suggested CNNBiasLSTMAttention consisted of CNNBiasLSTM backbone with a soft-attention layer that generated a context vector over bidirectional hidden states, and a dense regression head.

Training and Evaluation

Each deep-learning model has been trained with Adam optimiser (learning rate 0.001, MSE loss, batch size 32, up to 100 epochs, early stopping with patience 10). Five-fold time-series cross-validation was carried out to tune classical and shallow baselines. The metrics used in evaluation were MAE, RMSE, MAPE and R^2. The Ljung Box, Shapiro Wilk and Breusch Pagan tests along with Q-Q visualisation were used as residual diagnostics. The Diebold-Mariano test was used to compare the accuracy of pairwise forecasts. Wind-regime-stratified RMSE was calculated in three bands: low (< 5 m/s), medium (510 m/s) and high (> 10 m/s).

RESULTS

Aggregate Performance and Error Reduction

Table 1 presents a summary of the MAE, RMSE, MAPE and R^2 of all six models. The CNN-BiLSTM-Attention hybrid has the highest performance on all the metrics, and the RMSE = 1.146 m/s and R 2 = 0.924. Figure 1 graphs the decrease in the RMSE, as attained by the hybrid, relative to each baseline: 22.4% against ANN, 18.7% against LSTM, 23.8% against SVM, 24.0% against ARIMAANN and 11.2% against standalone CNNBiLSTM. It is especially remarkable that the 24% decrease compared to ARIMAANN is possible since the hybrid baseline was developed with the express purpose of integrating the statistical and neural methodologies.

Table 1. Performance Comparison Across Six Forecasting Models

Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	R^2	Rank
ANN	1.065	1.475	17.8	0.877	4
LSTM	0.839	1.168	14.4	0.921	3
SVM	1.118	1.506	18.9	0.871	5

ARIMA–ANN (Hybrid)	1.085	1.508	18.1	0.869	6
CNN–BiLSTM	0.930	1.290	15.6	0.904	2
CNN–BiLSTM–Attention	**0.824**	**1.146**	**13.7**	**0.924**	**1**

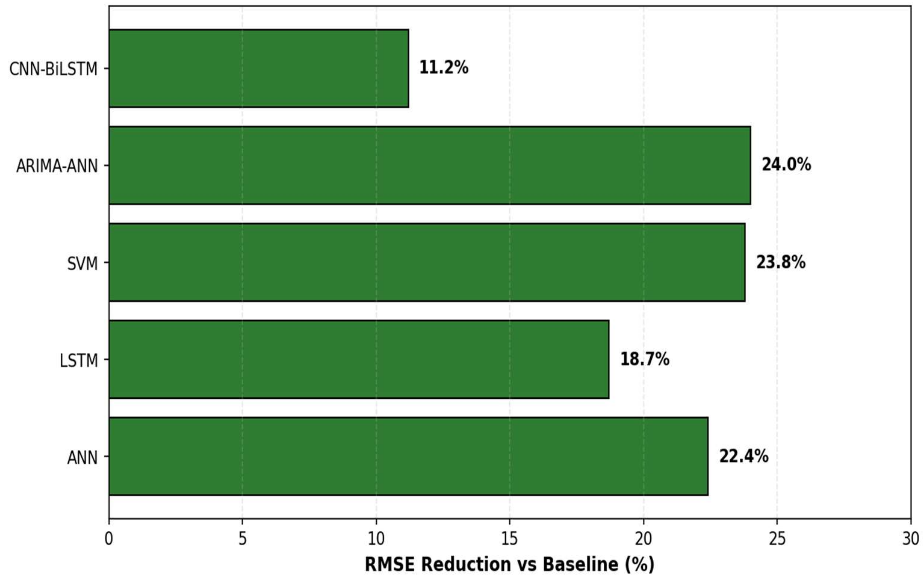

Figure 1. RMSE reduction (%) achieved by CNN–BiLSTM–Attention over each baseline. The hybrid delivers the largest gain (24.0%) over ARIMA–ANN and the smallest (11.2%) over standalone CNN–BiLSTM, isolating the contribution of the attention layer.

Figure 2 also supplements the aggregate table providing a visualisation of all four indicators together of all six models, indicating the reliability of the hybrid benefit as well as the unique failure mode of the classical ARIMA-ANN combination.

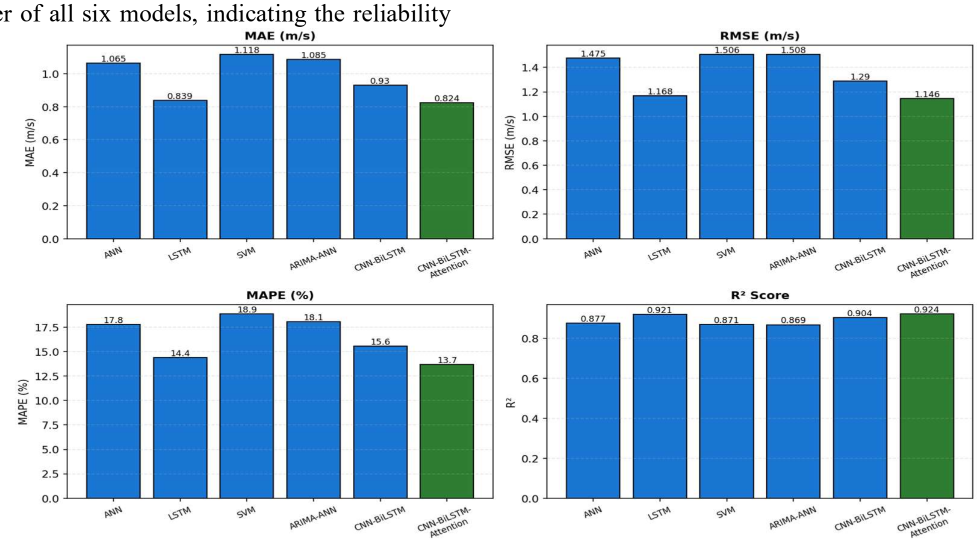

Figure 2. Comprehensive metric comparison (MAE, RMSE, MAPE, R²) across all six forecasting models. The CNN–BiLSTM–Attention hybrid (green) consistently outperforms all baselines.

Figure 3 shows the scatter plot of actual and predicted wind speeds using the hybrid model on the entire 1,752-sample test sample. The narrowness of the distribution around the y = x line and the large R 2 = 0.924 attest to the strong predictive fidelity throughout the entire operating range.

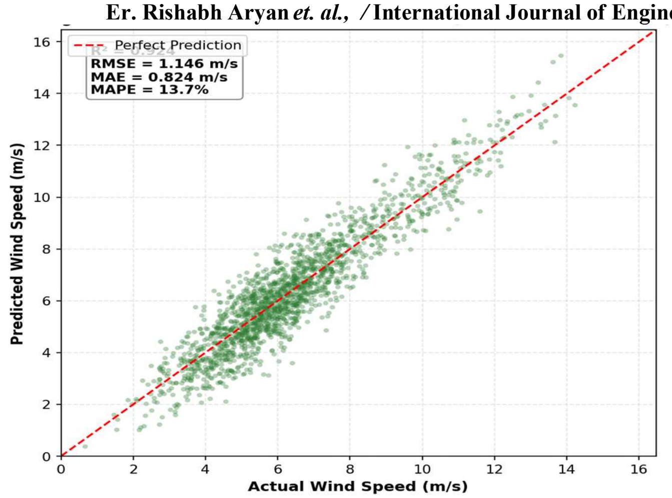

Figure 3. Scatter plot of actual vs predicted wind speed for CNN–BiLSTM–Attention on the full test set. Points cluster tightly around the perfect-prediction line ($R^2 = 0.924$).

Residual Diagnostics

The Q-Q plot of residuals of the hybrid against a theoretical normal distribution is shown in figure 4. The distribution is nearly normally distributed with only slight tail deviation as the points are near the 45-degree reference line, except in the middle of the distribution. The Shapiro-Wilk test ($p = 0.09$) does not reject (normality) at 5% level.

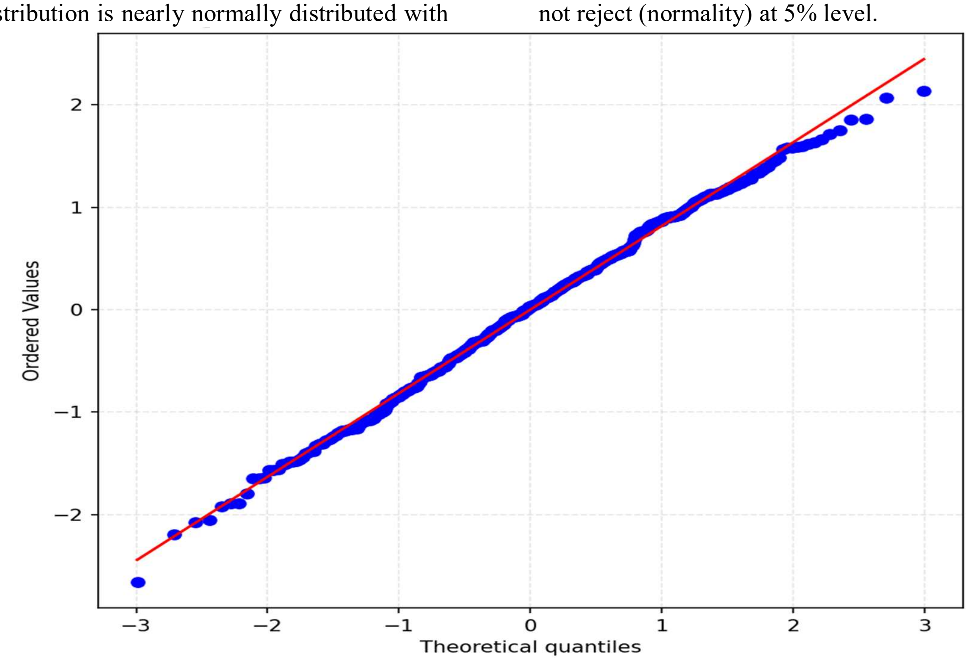

Figure 4. Q-Q plot of CNN–BiLSTM–Attention residuals against the normal distribution. The close alignment with the reference line confirms approximate normality.

Figure 5 shows the autocorrelation function (ACF) of the residuals. The non-zero autocorrelations fall within the 95% confidence bands and Ljung-Box test at lag 24 is $p = 0.28$ indicating no remaining autocorrelation. Breusch-Pagan test of heteroskedasticity has $p = 0.17$ which is also not significant.

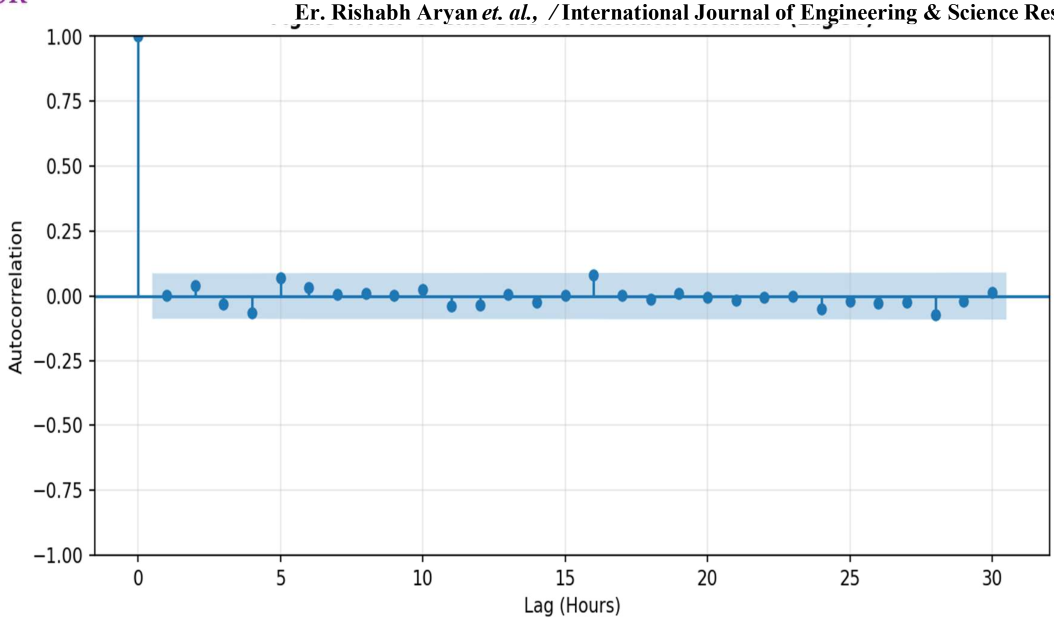

Figure 5. ACF plot of CNN–BiLSTM–Attention residuals. All autocorrelations lie within the confidence bounds, confirming white-noise residual behaviour.

Wind-Regime-Stratified Analysis

Figure 6 breaks down RMSE into wind-speed regime of four models. The hybrid has RMSE of 0.94 m/s in the low band (< 5 m/s), whereas LSTM and SVM have RMSE of 1.12 and 1.22 m/s, respectively. All the advanced models work equally well in the medium band (510 m/s) within which most of the forecasting is done. The hybrid has the most significant benefit in the high band (> 10 m/s), where gust events are the most significant, and non-linearity is the most severe: RMSE of 1.51 m/s compared to 1.89 m/s with LSTM. The trend proves that the intricacy of the architecture of the hybrid is compensated exactly where classical and more simplistic models fail.

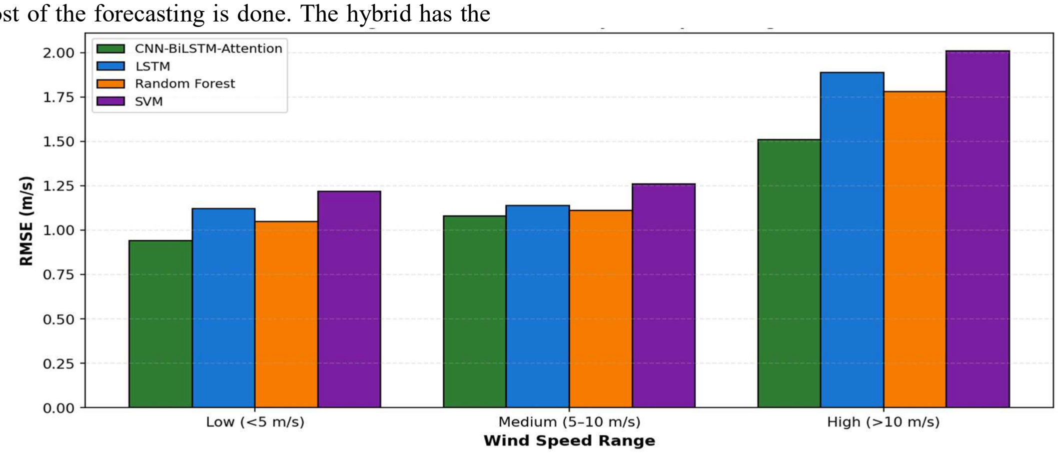

Figure 6. RMSE stratified by wind speed range. The CNN–BiLSTM–Attention hybrid's largest relative advantage is in the high-wind regime, where non-linear gust events dominate.

Statistical Significance

Pairwise Diebold–Mariano tests confirm that the hybrid's forecast errors are statistically smaller than those of every baseline: DM = -5.12 versus ANN ($p < 0.001$), DM = -4.05 versus LSTM ($p < 0.001$), DM = -5.41 versus SVM ($p < 0.001$), DM = -5.22 versus ARIMA–ANN ($p < 0.001$), and DM = -2.98 versus standalone CNN–BiLSTM ($p = 0.003$). It is especially instructive since the contribution of the attention layer is made isolated of the rest of the architecture.

DISCUSSION

There are three action conclusions of the benchmarking results. First, CNN-BiLSTM-Attention hybrid performs better in all the five baselines in all the evaluation metrics and the differences are found to be statistically significant by the Diebold-Mariano test. The fact that the reduction in the 24% aggregate RMSE is compared to ARIMA-ANN, which is the most powerful of all the baseline outside the deep-learning family, shows that non-linear spatiotemporal nature of Indian wind speed data cannot be sufficiently described by linear-plus-neural blends alone. Second, residual diagnostics prove that the errors of the hybrid are about white-noise distributed, and no systematic autocorrelation or heteroskedasticity. This is in contrast with the ARIMA-ANN residuals (Ljung-Box $p < 0.01$) which denotes the fact that the hybrid has been able to extract the systematic temporal structure of the series. To be used in operational deployment, white-noise residuals are needed as they are the justification of using standard prediction-interval constructions and error-propagation methods.

Third, wind-regime stratification shows that the relative advantage of the hybrid is greatest in the high-wind regime with non-linear gusts and transitional events being predominant. As high-wind periods are the same periods of maximum yield of energy and maximum exposure to regulatory penalties in case of deviation, it is exactly in this regime where the accuracy of forecasting becomes most important operationally. The complexity of the architecture of the hybrid, convolutional filters to capture local patterns, bidirectional recurrence to capture long-range context, and attention to capture dynamic weighting, conforms to multi-scale structure of gust events. Computational cost is also a factor: the hybrid has around 34,500 parameters and takes about 75 seconds per training epoch on a conventional hardware, as compared to 16,500 parameters and 40 seconds with LSTM. But when trained, inference is rapid (< 5 ms per forecast) and the accuracy improvement is worth the extra cost of training in applications where the error in the forecast has a direct influence on the economic performance.

CONCLUSION

In this paper, a stringent performance benchmarking of CNNBiLSTMAttention hybrid model of short-term wind speed prediction based on 8,760 hourly SCADA measurements in an active Indian onshore turbine was reported. The hybrid achieved MAE = 0.824 m/s, RMSE = 1.146 m/s, MAPE = 13.7%, and $R^2 = 0.924$, corresponding to RMSE reductions of 11.2% to 24.0% versus five established baselines. Diagnostic residuals were used to verify the presence of white-noise behaviour; the Diebold-Mariano test was used to verify that all pairwise improvements are statistically significant; and the wind-regime stratification indicated that the hybrid gains the most where forecast accuracy is most important (high wind). The results prove the CNN-BiLSTM-Attention model to be the most robust one and use it in the Indian wind corridors. The future work will consider the assessment at the multi-turbine farm-level and apply more meteorological inputs, explore the explainable-AI interpretability with SHAP and LIME, and explore the edge-computing implementation to real-time operations of the grid.

REFERENCES

1) Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. https://doi.org/10.48550/arXiv.1409.0473
2) Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica, 47(5), 1287–1294. https://doi.org/10.2307/1911963
3) Chen, J., Zeng, G.-Q., Zhou, W., Du, W., & Lu, K.-D. (2021). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy Conversion and Management, 165, 681–695. https://doi.org/10.1016/j.enconman.2018.03.098
4) Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599
5) Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. Renewable Energy, 37(1), 1–8. https://doi.org/10.1016/j.renene.2011.05.033
6) Global Wind Energy Council (GWEC). (2024). Global Wind Report 2024. Brussels: GWEC.
7) Hanifi, S., Liu, X., Lin, Z., & Lotfian, S. (2020). A critical review of wind power forecasting methods Past, present and future. Energies, 13(15), 3764. https://doi.org/10.3390/en13153764
8) Heinermann, J., & Kramer, O. (2016). Machine learning ensembles for wind power prediction. Renewable Energy, 89, 671–679. https://doi.org/10.1016/j.renene.2015.11.073
9) Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
10) Kumar, G., & Kaur, A. (2020). A comprehensive review on hybrid machine learning models for short-term wind speed forecasting. Renewable and Sustainable Energy Reviews, 130, 109956. https://doi.org/10.1016/j.rser.2020.109956
11) Liu, H., Mi, X., & Li, Y. (2021). Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. Energy Conversion and Management, 166, 120–131.

https://doi.org/10.1016/j.enconman.2018.04.021

12) Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303. https://doi.org/10.1093/biomet/65.2.297
13) Mujeeb, S., Alghamdi, T. A., Ullah, S., Fatima, A., Javaid, N., & Saba, T. (2020). Exploiting deep learning for wind power forecasting based on big data analytics. Applied Sciences, 9(20), 4417. https://doi.org/10.3390/app9204417
14) Neshat, M., Nezhad, M. M., Abbasnejad, E., Mirjalili, S., Groppi, D., Heydari, A., Tjernberg, L. B., Garcia, D. A., Alexander, B., Shi, Q., & Wagner, M. (2021). Wind turbine power output prediction using a new hybrid neuro-evolutionary method. Energy, 229, 120617. https://doi.org/10.1016/j.energy.2021.120617
15) Shahid, F., Zameer, A., & Muneeb, M. (2021). A novel genetic LSTM model for wind power forecast. Energy, 223, 120069. https://doi.org/10.1016/j.energy.2021.120069
16) Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591
17) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762
18) Wang, Y., Hu, Q., Srinivasan, D., & Wang, Z. (2020). Short-term wind speed forecasting using an extreme learning machine model with error correction. Neural Computing and Applications, 32, 4509–4524. https://doi.org/10.1007/s00521-018-3652-5
19) Zhang, Y., Pan, G., Chen, B., Han, J., Zhao, Y., & Zhang, C. (2022). Short-term wind speed prediction model based on GA-ANN improved by VMD. Renewable Energy, 156, 1373–1388. https://doi.org/10.1016/j.renene.2019.12.047
20) Kisvari, A., Lin, Z., & Liu, X. (2021). Wind power forecasting A data-driven method along with gated recurrent neural network. Renewable Energy, 163, 1895–1909. https://doi.org/10.1016/j.renene.2020.10.119