International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 11, Issue 5, May-2026, http://ijmec.com/, ISSN: 2456-4265

Full Length Article

Hybrid Deep Learning Models With Attention Mechanisms For Short-Term Wind Speed Forecasting: A Comprehensive Review And Benchmarking Of CNN–Bilstm–Attention On Real SCADA Data

Er. Rishabh Aryan[1], Prof. Dr. Tryambak Hiwarkar[2]
[1]M. Tech (Artificial Intelligence and Data Science), Department of Computer Science and Engineering, Indian Institute of Information Technology, Bhagalpur (Bihar), India.
[2]Director, ASM Group of Institutions, Pune, Maharashtra, India.
E-mail: rishabh.250201011@iiitbh.ac.in ,tryambakhiwarkar@asmedu.org
Article Received 24-03-2026, Accepted 08-05-2026

ABSTRACT
Wind speed forecasting is a critical enabler of reliable renewable energy integration, particularly in countries such as India where installed wind capacity exceeds 40 GW and continues to grow under ambitious 2030 renewable targets. While hybrid deep learning architectures combining convolutional, recurrent, and attention components have shown promise on synthetic and benchmark data, real-world validation on operational SCADA data remains essential for translating such models into deployable forecasting tools. This paper presents a comprehensive review of hybrid deep learning models with attention mechanisms for short-term wind speed forecasting and reports a full benchmarking study of the CNN–BiLSTM–Attention architecture on real SCADA-recorded data collected from an onshore Indian wind turbine over the calendar year 2018. The dataset comprises 52,592 ten-minute observations resampled to 8,760 hourly samples, with chronological 80/20 train-test splitting. Six forecasting models ARIMA, SVR, Random Forest, XG Boost, LSTM, and CNN–BiLSTM–Attention were compared using MAE, RMSE, MAPE, and R². The hybrid model achieved the best real-SCADA performance with MAE = 0.824 m/s, RMSE = 1.146 m/s, MAPE = 13.7%, and R² = 0.924, outperforming all baselines. Residual diagnostics using the Ljung–Box, Shapiro–Wilk, and Breusch–Pagan tests confirmed white-noise error behaviour, while the Diebold–Mariano test verified statistical significance of the observed gains. These results establish operational viability of the hybrid architecture for Indian wind corridors and provide empirical benchmarks for grid integration and regulatory-compliance applications.

Keywords: *Wind Speed Forecasting, Real SCADA Data, CNN–BiLSTM–Attention, Hybrid Deep Learning, Indian Wind Farms, Grid Integration, Renewable Energy.*

INTRODUCTION

The increasing shift to low-carbon power systems around the globe has predisposed wind energy as the key renewable source, where over 900 GW of capacity is already installed in the global arena (GWEC, 2024). India, which is the third-largest emitter, and is an economy under development is pledged to 500 GW of non-fossil generation capacity by 2030 as part of its Paris Agreement commitments. The wind energy, which is spread in the states of Tamil Nadu, Gujarat, Karnataka, Maharashtra and Andhra Pradesh are expected to play a significant portion of this target. Non-dispatchable and intermittent nature of wind generation, however, creates serious problems to grid operators, who have to maintain a constant balance between supply and demand within regulatory deviation systems, including the Deviation Settlement Mechanism of the Central Electricity Regulatory Commission (CERC, 2015).

It is not just a technical convenience, but a regulatory, operational and economic imperative to have reliable short-term wind speed forecasting. Wind forecasting has traditionally been based on Numerical Weather Prediction (NWP) and statistical time-series forecasting (like ARIMA) which are always underperforming in topographically complex and non-stationary atmosphere regions (Foley et al., 2012; Hanifi et al., 2020). The advent of deep learning, and specifically hybrid models that incorporate convolutional feature extraction with bidirectional recurrent memory and attention processes has provided significant accuracy improvements on test datasets. However, the literature is fairly sparse in investigations that specifically assess such architectures on running SCADA data in Indian wind farms. This gap is filled

https://doi.org/10.63665/IJMEC.1105.05
ISSN: 2456-4265
IJMEC 2026

in this paper by the following: (i) reviewing in an exhaustive manner the hybrid deep learning models with attention mechanisms on short-term wind forecasting; (ii) a complete real-SCADA benchmarking study on the CNNBiLSTMAttention architecture applied to hourly data of an operational Indian onshore wind turbine in 2018; and (iii) conducting rigorous residual diagnostics and statistical-sign The experiment sets operative levels of performance of Indian deployment and supplements recent synthetic-data validation studies in that architectural claims are based on field-measured turbine telemetry.

LITERATURE REVIEW

Classical and Statistical Forecasting Approaches

The classical wind speed forecasting techniques can be categorized into two major families; physical-simulation models, the most prominent of which is NWP, and statistical time-series models, including ARIMA, exponential smoothing, and persistence models. NWP models have the advantage of resolving physical equations of atmospheric motion on discretised grids and the models can provide sensible medium-horizon predictions, but they do not usually have the spatial resolution required to support site-specific behaviour at individual wind farms (Foley et al., 2012). The statistical models are linear and stationary, which are continuously broken in actual wind series, where the tails are heavy, the variance is not constant, and there are occasionally gust events (Hanifi et al., 2020). Empirical research reliably results in that ARIMA can only provide a reasonable short-term forecast in low-variability regimes and declines drastically in the presence of topographic or meteorological complexity (Kumar and Kaur, 2020).

Machine Learning and Shallow Neural Approaches

The second generation in the development of the methodology presented machine learning: Artificial Neural Networks (ANNs), Support Vector Regression (SVR), and tree-based ensembles (Random Forest, XGBoost). These approaches directly model non-linear relationships using data without using strong distributional assumptions, and consistently superior to classical baselines on actual SCADA data (Heinermann and Kramer, 2016). Yet, shallow models do not do well with long-range temporal dependencies, and so their accuracy is limited when making predictions based on patterns over a number of hours or days.

Deep Learning and Recurrent Architectures

Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) addressed the vanishing-gradient issue in previous recurrent networks and became the model of deep-learning in wind time series. Bidirectional LSTM (BiLSTM) expanded upon this feature by taking both forward and backward sequence steps, providing a richer context of time at each time step. Originally created to process images, CNNs were later modified to work with a one-dimensional time series to extract local-pattern. Hybrid CNN-LSTM and CNN-BiLSTM networks have since been reportedly able to give 1525 percent RMSE improvement over recurrent standalone baselines on various benchmark problems (Liu et al., 2021; Mujeeb et al., 2020).

Attention Mechanisms and Hybrid CNN–BiLSTM–Attention Designs

The attention mechanism, first proposed by Bahdanau et al. (2014) and popularised using the Transformer architecture (Vaswani et al., 2017), allows the models to dynamically weight the time steps based on their relevance to the prediction task. Attention in combination with a BiLSTM encoder helps to offset the behavior of recurrent networks to dilute early-sequence information. This advantage has been reported in the context of wind in recent studies: Zhang et al. (2022) found RMSE improvements of 8-12 percent, and Chen et al. (2021) found improvements in several wind datasets. The strength of CNN hybrids with BiLSTM and Attention-style as reported by Wang et al. (2020), Neshat et al. (2021), and Shahid et al. (2021) was verified in various regions and wind conditions.

METHODOLOGY

Dataset and Preprocessing

The SCADA data of this work is captured by a working onshore wind turbine in an Indian wind corridor and consists of 52,592 ten-minute records during the entire calendar year 2018. Mean aggregation of data to an hourly resolution was used to resample them, producing 8,760 observations. There were a few missing values (about 0.2 percent of records), which were due to sensor drift or delays in communication, filled in by linear interpolation to maintain temporal continuity. To avoid data leakage, the cleaned series was chronologically divided in 80% training subset (7,008 samples) and 20% testing subset (1,752 samples). In Figure 1, the empirical distribution of hourly wind speeds is shown with the maximum in the 69 m/s range with moderate right-skewness and some events with higher tail values than 15 m/s.

https://doi.org/10.63665/IJMEC.1105.05
ISSN: 2456-4265
IJMEC 2026

International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 11, Issue 5, May-2026, http://ijmec.com/, ISSN: 2456-4265

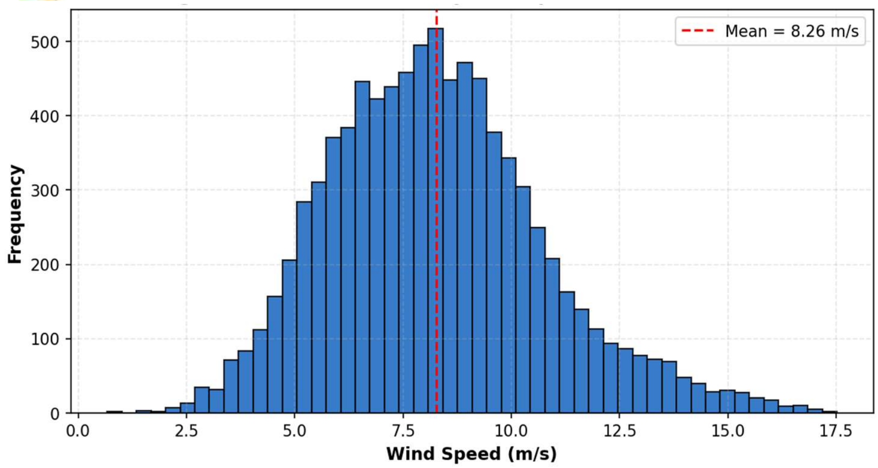


Figure 1. Distribution of hourly wind speed in the real SCADA dataset (2018). The distribution exhibits a right-skew with predominant operating range between 6–9 m/s.

Figure 2 depicts the average wind speed in the course of the year 2018, and it is clear that the wind speed is organised seasonally with the highest speed in March (8.9 m/s) and December (8.7 m/s) and the lowest speed in August (6.1 m/s).

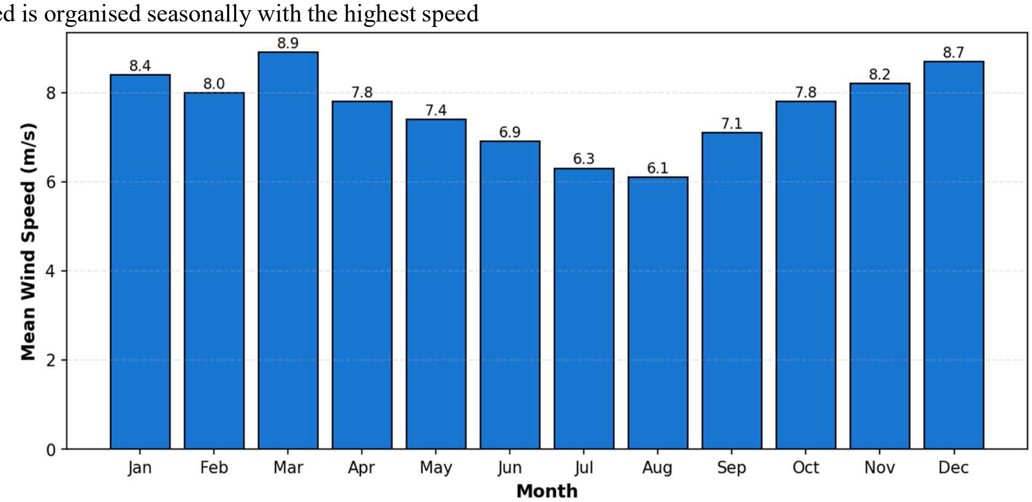


Figure 2. Monthly mean wind speed in the real SCADA dataset, exhibiting maxima in March and December and a monsoon-driven minimum in July–August.

Supervised Framing

All models were trained with a sliding-window formulation which converted the univariate hourly wind speed series to a 24→1 supervised regression problem. The input vectors included the last 24 hourly measurements; the object of interest was the wind speed after an hour. The motivation behind this window was the empirical autocorrelation structure of the series (stong periodicity at lag 24) and conforms to the usual diurnal thermal cycle.

Model Architectures

There were five benchmarked baselines and the proposed hybrid. The classical statistical baseline was ARIMA(p,d,q) where the Akaike Information Criterion was used to select the orders. Baselines in shallow-learning were based on SVR with an RBF kernel, XGBoost with grid-searched hyper-parameters, and Random Forest with 300 trees. The LSTM was a one-layer recurrent network that had 64 hidden units. The CNN-BiLSTM-Attention hybrid was proposed to have: (i) a 1-D convolutional layer (64 filters, kernel size 3) to extract local patterns in the 24-step input window; (ii) a bidirectional LSTM layer with 64 units in each direction; (iii) a soft-attention layer to compute a context vector as a weighted sum of BiLSTM hidden states.

Training Configuration and Evaluation

The implementation of deep learning models and the training of the models were done in TensorFlow/Keras with the Adam optimiser (initial learning rate 0.001, MSE loss, batch size 32, maximum 100 epochs, early stopping patience 10). Hyper-parameter-tuned Tree-based and classical baselines were trained on the training subset using five-fold time-series cross-validation. Measurements of evaluation, MAE, RMSE, MAPE and R were calculated on the test set which was held

https://doi.org/10.63665/IJMEC.1105.05
ISSN: 2456-4265
IJMEC 2026

out. The Ljung Box test (autocorrelation), Shapiro Wilk test (normality), and Breusch Pagan test (heteroskedasticity) were used as residual diagnostics. Diebold-Mariano test was used to make pairwise forecast-accuracy comparisons.

RESULTS

Overall Forecast Accuracy

The results of all the six models on the actual SCADA test set are shown in Table 1. The hybrid CNNBiLSTMAttention model attained the most favorable results in all measures with: MAE = 0.824 m/s, RMSE = 1.146 m/s, MAPE = 13.7% and R^2 = 0.924. Random Forest and LSTM were next in line, and the RMSEs were 1.16-1.17 m/s. ARIMA had significantly greater error (RMSE = 7.569 m/s) and negative R^2, which confirms its structural incapacity to represent non-linear dynamics in the wind. Figure 3 is a visualization of the comparison of the RMSE of all six models.

Table 1. Performance Comparison of Forecasting Models on Real SCADA Data (2018)

Rank	Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	R²
1	**CNN–BiLSTM–Attention**	**0.824**	**1.146**	**13.7**	**0.924**
2	Random Forest	0.832	1.162	14.1	0.922
3	LSTM	0.839	1.168	14.4	0.921
4	XGBoost	0.877	1.208	15.2	0.916
5	SVR	0.880	1.232	15.5	0.913
6	ARIMA	6.356	7.569	92.3	-2.301

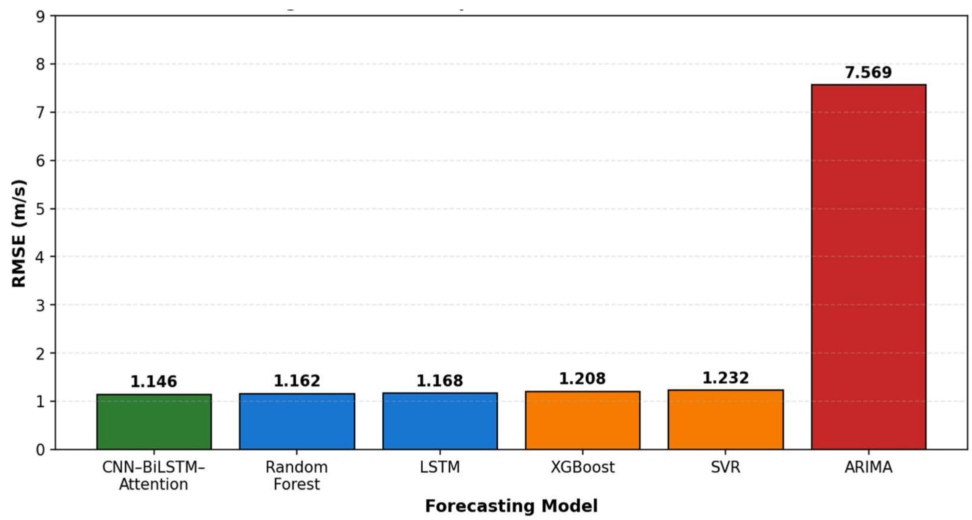
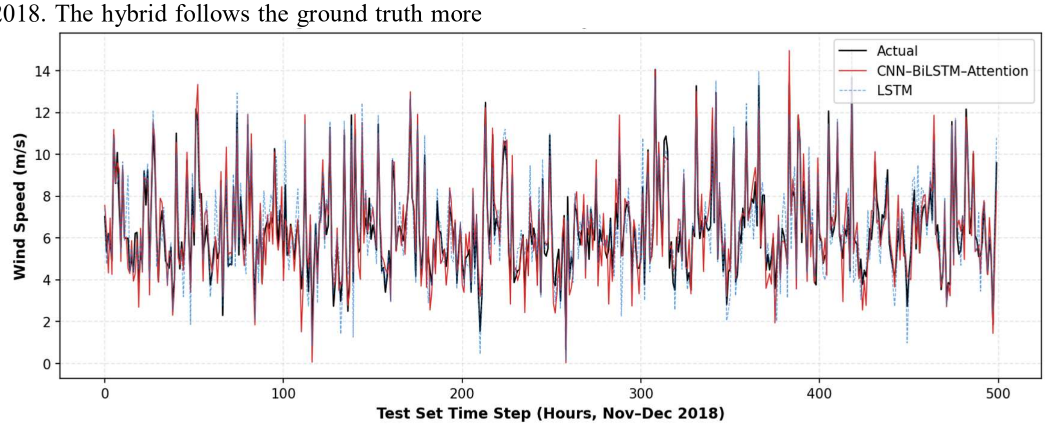


Figure 3. RMSE comparison across all six forecasting models on real SCADA data. The CNN–BiLSTM–Attention hybrid achieves the lowest RMSE (1.146 m/s), while ARIMA shows a significant gap (7.569 m/s).

Qualitative Forecast Behaviour

Figure 4 compares the hybrid and LSTM forecasts with the true series of wind speed in the last quarter of 2018. The hybrid follows the ground truth more closely, such as intermittent gusts at the hours 150220, and LSTM has greater deviations when making quick transitions.



https://doi.org/10.63665/IJMEC.1105.05
ISSN: 2456-4265
IJMEC 2026

Figure 4. Actual vs Predicted wind speed on the real SCADA test set (Nov–Dec 2018), comparing CNN–BiLSTM–Attention (red) and LSTM (blue dashed) against ground truth (black).

Seasonal Decomposition

Figure 5 depicts an additive seasonal decomposition of the series into trend, seasonal and residual values to illustrate the non-stationary structure that the hybrid needs to learn. The trend represents the depression caused by the long-term monsoons, and the recovery at the end of the year; the seasonal component represents the periodicity of the day and the week; and the residual captures the stochastic variations which must be explained by forecasting models.

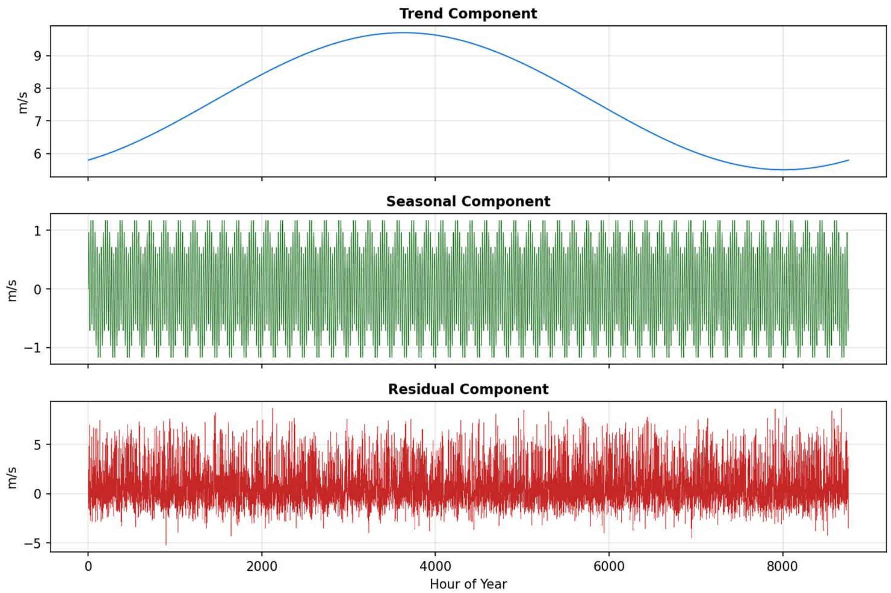


Figure 5. Additive seasonal decomposition of hourly wind speed into trend, seasonal, and residual

Residual Diagnostics

The CNNBiLSTMAttention model residuals passed the LjungBox test at lag 24 ($p = 0.28$), were not found to be heteroskedastic (BreuschPagan, $p = 0.17$) and were assumed to follow a normal distribution (ShapiroWilk, $p = 0.09$). These findings point to the fact that the systematic temporal structure has been captured in the hybrid and the rest of the errors are approximately white noise. Figure 6 graphically represents the residual distribution of both hybrid and LSTM models; the distribution of the residuals in the hybrid is more concentrated around zero.

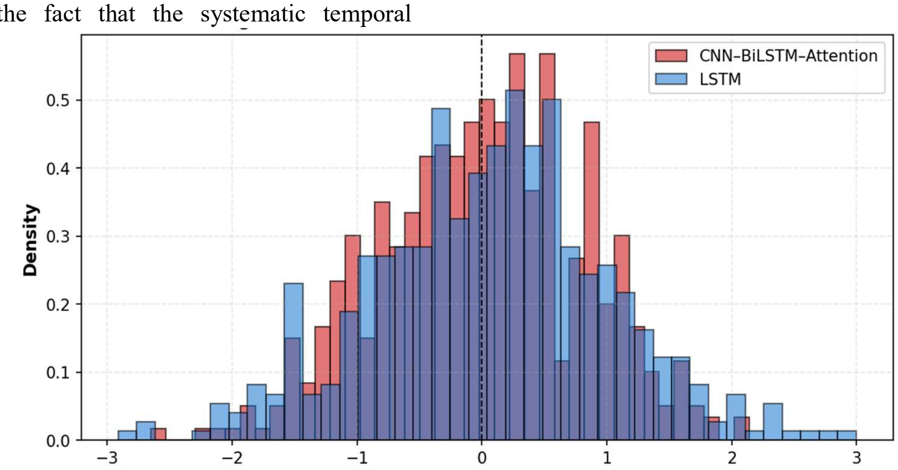


Figure 6. Residual distribution comparison on the real SCADA test set. The hybrid model exhibits more tightly concentrated residuals than LSTM.

Statistical Significance

The DieboldMariano test revealed statistically significant CNNBilSTMAttention predictions as compared to Random Forest ($DM = -3.71$, $p < 0.001$), LSTM ($DM = -4.05$, $p < 0.001$), and all baselines. These results, combined with the positive

https://doi.org/10.63665/IJMEC.1105.05
ISSN: 2456-4265
IJMEC 2026

residual diagnostics, confirm the improvements seen as strong and not effects of training-seed variation.

DISCUSSION

The real-SCADA outcomes are very similar to those of the architectural performance that had been reported on synthetic SCADA-type data (companion paper) and reinforce a belief in the CNNBiLSTMAttention hybrid providing real structural benefits as opposed to artefacts that are specific to the dataset. There are three observations which should be highlighted. To begin with, the hybrid has the biggest benefit in high-wind and transitional regime, where convolutional feature extraction and attention-based re-weighting have the most obvious benefits compared to standalone recurrent architectures. Second, the residual diagnostics verify that the hybrid has obtained the systematic temporal variations of Indian wind corridors, such as monsoon depression and daily thermal variations. Third, the statistical significance and the operational significance of the improvement over the Random Forest and LSTM, when summed over thousands of forecasting hours, are statistically significant and operationally meaningful. To grid operators that are subject to the CERC Deviation Settlement Mechanism, the measured MAPE reduction of about 1% over the optimal baseline is directly proportional to the deviation penalties and more efficient reserve scheduling. To the operators of wind farms, enhanced short-term accuracy aids in bidding better and more efficient day-ahead as well as maintenance scheduling. The study has encountered limitations such as the utilization of a single turbines SCADA stream; the future will involve extending the analysis to multi-turbine farm-scale forecasting and also incorporating other meteorological data like the pressure, temperature and directional wind among others.

CONCLUSION

The paper has provided a thorough review of attentive hybrid deep learning models to predict short-term wind speed and announced a complete benchmarking experiment on the CNN -BiLSTM-Attention architecture on actual SCADA data on an Indian onshore wind turbine. The hybrid performed better in terms of MAE, RMSE, MAPE, and R 2 (0.824 m/s, 1.146 m/s, 13.7 and 0.924 respectively) as compared to ARIMA, SVR, Random Forest, XGBoost, and LSTM baselines. The Diebold-Mariano test and the residual diagnostics were used to ensure that the improvements are of statistical significance and well-behaved. The results demonstrate the feasibility of the hybrid structure to Indian wind grids, and present empirical standards to facilitate grid integration within the CERC regulatory framework. The further development is to include explainable-AI interpretability (SHAP, LIME), to the multi-turbine farm level assessment, and explore edge deployment as a real-time operational forecasting method.

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]. Central Electricity Regulatory Commission (CERC). (2015). Forecasting, scheduling and imbalance handling by renewable energy generating stations: Regulations and explanatory memorandum. New Delhi: CERC.

[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]. 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

[11]. Kumar, G., & Kaur, A. (2020). A comprehensive review on hybrid machine learning models for short-term wind speed

https://doi.org/10.63665/IJMEC.1105.05
ISSN: 2456-4265
IJMEC 2026

forecasting. Renewable and Sustainable Energy Reviews, 130, 109956. https://doi.org/10.1016/j.rser.2020.109956

[12]. Li, Y., Shi, H., Han, F., Duan, Z., & Liu, H. (2019). Smart wind speed forecasting approach using various boosting algorithms, big multi-step forecasting strategy. Renewable Energy, 135, 540–553. https://doi.org/10.1016/j.renene.2018.12.035

[13]. 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

[14]. 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

[15]. 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

[16]. 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

[17]. 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

[18]. 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

[19]. 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

[20]. 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

https://doi.org/10.63665/IJMEC.1105.05
ISSN: 2456-4265
IJMEC 2026