The 12th Iranian Conference on Renewable Energy & Distributed Generation (ICREDG2025), 26 February, 2025, QUT, Iran.

Enhancing Wind Power Forecasting Through a Hybrid Deep Learning Approach: Long Short-Term Memory Integrated with Attention Mechanism

Amir Hossein Karamali
department of Electrical Engineering
Iran University of Science and Technology
Tehran,Iran
a_karamali@elec.iust.ac.ir

Ali Reza Danesh
Faculty of Engineering
Arak University
Arak,Iran
a.daneshnashalji.02@msc.araku.ac.ir

Mohsen Kalantar
department of Electrical Engineering
Iran University of Science and Technology
Tehran,Iran
Kalantar@iust.ac.ir

Mohammad Reza Moghadasian
Faculty of Engineering
Arak University
Arak,Iran
m.moghadasian.02@msc.araku.ac.ir

Saber Rezaei
department of Electrical Engineering
Iran University of Science and Technology
Tehran, Iran
s_rezaei@elec.iust.ac.ir

***Abstract*— Accurate prediction of wind energy generation plays a vital role in integrating renewable energy sources efficiently into electrical grids. This study explores the use of advanced artificial intelligence methods for forecasting wind power production at two meteorological stations located in Khomein and Saveh, Iran. A new hybrid model is proposed, combining Long Short-Term Memory (LSTM) networks with an attention mechanism. This approach surpasses traditional methods like Random Forest and XGBoost by better capturing intricate temporal dynamics and adjusting to fluctuating wind conditions. The results demonstrate the model's potential to improve grid stability and optimize energy management.**

Keywords*— *Time Series Forecasting, Machine Learning, Deep Learning, Attention Mechanism

I. Introduction

The rising global demand for clean and sustainable energy has intensified the focus on wind power as a viable alternative to fossil fuels. However, wind energy generation is inherently intermittent and unpredictable, posing significant challenges for grid stability and energy management. Reliable wind power forecasting plays a pivotal role in optimizing grid operations, energy storage planning, and cost reduction [1-3].
Over the years, various approaches to wind power forecasting have been developed, broadly categorized into physical, statistical, and machine learning methods. Physical models rely on meteorological data and parameters such as wind speed, temperature, pressure, and humidity to predict wind power generation. These models often use numerical weather prediction (NWP) systems to simulate atmospheric conditions and estimate future wind patterns. While physical models are effective for medium- and long-term forecasting, they require precise input data and are computationally intensive, making them less suitable for real-time or short-term predictions [4,5]. Additionally, their performance is highly dependent on the accuracy of the input meteorological data, which can be challenging to obtain in some regions [6,7].
Statistical approaches, conversely, have become popular because of their simplicity and capacity to model time-series data. Methods like Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) are commonly employed for short-term wind power prediction. These techniques identify linear relationships in historical wind power data and perform well under consistent wind conditions. Nonetheless, they face challenges in addressing the nonlinear and intricate patterns frequently found in wind power data, particularly in highly fluctuating environments [8]. To address these limitations, more advanced statistical techniques have been proposed. For example, Kalman filtering has been used to improve the accuracy of wind speed predictions by recursively updating estimates based on new observations [9]. Similarly, exponential smoothing methods have been applied to capture trends and seasonal variations in wind power data, providing more robust forecasts [10]. Another notable advancement in statistical forecasting is the use of Gaussian Processes (GP), which provide a probabilistic framework for modeling uncertainty in wind power generation. Gaussian Processes are particularly useful for generating prediction intervals, offering a range of possible outcomes rather than a single point estimate. This probabilistic approach is valuable for grid operators who need to account for the inherent uncertainty in wind power generation [11].
Despite the advancements in physical and statistical methods, the complexity and variability of wind power data have led researchers to explore more sophisticated approaches. The advent of artificial intelligence and machine learning has revolutionized wind power forecasting by enabling data-driven approaches that can learn and adapt to complex, nonlinear patterns. Algorithms such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Random Forests have demonstrated superior performance compared to traditional methods [12,13]. Among these, Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) networks have gained significant traction for their ability to model long-term temporal dependencies, making them particularly well-suited for wind power prediction tasks [14,15].

979-8-3503-6132-2/24/$31.00

In this study, a comprehensive exploration of artificial intelligence methods for wind power forecasting is conducted, with a focus on improving prediction accuracy and reliability. The research leverages advanced machine learning and deep learning techniques to address the challenges posed by the variability and complexity of wind power data. The key contributions and methodologies employed in this work are outlined below:

- Machine Learning Algorithms: Machine learning techniques, including Random Forests (RF) and XGBoost, were applied to predict wind power generation at two wind stations. These methods were chosen for their ability to handle complex datasets and provide robust predictions.
- Deep Learning Approach: A two-layer LSTM deep neural network was implemented to model temporal dependencies in wind power data for the same wind stations. LSTM networks are particularly effective in capturing long-term patterns in time-series data.
- Hybrid Model: A novel hybrid approach combining LSTM with an Attention Mechanism was proposed. This method demonstrated superior performance compared to traditional and standalone models, as it effectively focuses on important temporal features and identifies complex, nonlinear patterns in the data.
- Improved Forecasting Accuracy: The results indicate that the hybrid LSTM-Attention model significantly enhances forecasting accuracy, making it a valuable tool for grid management, energy storage planning, and optimizing the integration of renewable energy into power systems.

This research emphasizes the potential of integrating advanced machine learning and deep learning methods to tackle the challenges of wind power prediction, offering a reliable framework for future studies and real-world applications in the renewable energy field.

II. FOUNDATIONS AND PROBLEM FORMULATION

In this study, the dataset used for wind power forecasting spans a one-year period, with measurements recorded at 10-minute intervals. The data is collected from two meteorological stations located in Khomein and Saveh, both situated in the Markazi Province of Iran. These stations provide detailed and high-resolution wind-related data, including wind speed, direction, temperature, and other meteorological parameters, which are essential for accurate wind power prediction. The dataset's temporal resolution and geographical diversity make it a valuable resource for modeling and analyzing the variability and patterns of wind energy generation in the region.

The energy output of wind turbines can be determined using wind speed data and the power curves provided by wind turbine manufacturers [16]. This methodology involves integrating the turbine's power curve with time-series wind speed data to estimate the generated power. In this research, a 1 MW wind turbine is chosen for the analysis due to its widespread use in wind power plants and its suitability for regional wind conditions. The power curve corresponding to this turbine is shown in Fig. 1.

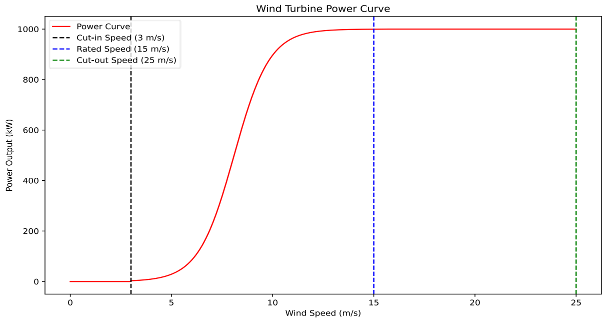

Figure 1. Power output curve of a 1 MW wind turbine. [16]

The technical Characteristics of the selected turbine are detailed in Table I.

TABLE I
Technical Characteristics of the Evaluated Wind Turbine [16]

Characteristics	Wind Machine
Rated power (kW)	1000.0
Hub height (m)	50.0
Cut-in wind speed (V_{ci})(m/s)	3.0
Rated wind speed (V_R)(m/s)	15.0
Cut-out wind speed (V_{co})(m/s)	25.0

To estimate the wind energy output, a polynomial equation of degree n is derived based on the turbine's power curve. This equation is valid between the cut-in speed (v_{ci}) and the rated speed (v_R), or between the cut-in speed and the cut-out speed (v_{co}). The equation is expressed as follows (Equation 1):

$$P_i(v) = \begin{cases} 0 & v < v_{ci} \\ a_n v^n + a_{n-1} v^{n-1} + \ldots + a_1 v + a_0 & v_{ci} \le v < v_R \\ P_R & v_R \le v < v_{co} \\ 0 & v \ge v_{co} \end{cases} \qquad (1)$$

Here, a_n, …, a_0 are regression coefficients derived from the power curve, v_{ci} is the cut-in speed, v_R is the rated speed, v_{co} is the cut-out speed, P_R is the rated power, and $P_i(v)$ represents the power generated at a specific wind speed .

The total energy output E over a specified time period can be calculated using Equation 2.

$$E = \sum_{i=1}^{N} P(v_i)\,\Delta t \qquad (2)$$

where N is the total number of time intervals in the considered period, $P(vi)$ is the power generated at wind speed vi, and Δt is the time interval (e.g., 10 minutes). This approach provides a robust framework for estimating wind energy production based on historical wind speed data and the operational characteristics of the turbine. By leveraging this methodology, the study ensures accurate and reliable predictions of wind energy output, which are critical for effective grid integration and energy planning [17].

III. METHODOLOGY

This section is divided into three main parts: machine learning, two-layer LSTM, and the proposed model. Each part is explained in detail below. The performance of these models is evaluated using standard metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) [5].

R-squared (R^2), as defined in Equation (3), represents the proportion of variance in the dependent variable that is predictable from the independent variables.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2} \tag{3}$$

MAE, as defined in Equation (4), quantifies the average of the absolute differences between the predicted and actual values. It offers a simple and direct way to understand the errors in the model's predictions.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{4}$$

RMSE , as defined in Equation (5), calculates the square root of the average squared differences between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{5}$$

A. Machine learning

Machine learning methods have been widely employed in wind power forecasting due to their ability to model complex, nonlinear relationships. In this study, two ensemble-based learning models Random Forest (RF) and Extreme Gradient Boosting (XGBoost) were implemented to predict wind power generation at two meteorological stations. The selection of these algorithms was based on their proven effectiveness in handling high-dimensional datasets and capturing intricate dependencies among input variables. The dataset comprises key meteorological parameters, including daily wind speed (dailyWS) and the daily standard deviation of wind speed (dailySD), recorded at 10-minute intervals over a one-year period from two meteorological stations in Saveh and Khomein, Iran. The primary target variable for prediction is daily power output (dailyPower). The dataset was split into training (80%) and testing (20%) subsets to ensure robust model evaluation.

a) Random Forest (RF): The Random Forest (RF) algorithm is an ensemble learning method that builds multiple decision trees during training and combines their outputs to improve prediction accuracy and reduce overfitting. This approach works by splitting the input dataset into smaller subsets and constructing individual decision trees for each subset. The final prediction is determined by aggregating the results from all trees, typically through majority voting (for classification) or averaging (for regression). This methodology simplifies complex problems by breaking them down into smaller, more interpretable components.

In the RF algorithm, a random subset of features (denoted as k) is selected from the feature space for each tree. Each tree is then trained using this subset along with a bootstrap sample of the training data. The generalization error and margin function in RF are defined as follows in Equation (6):

$$\begin{aligned} PE &= P_{X,Y}(mg(X,Y) < 0 \\ mg(X,Y) &= avg_k\left[I(h_k(X) = Y)\right] - \max_{j \neq Y} avg_k\left[I(h_k(X) = j)\right] \end{aligned} \tag{6}$$

Where X and Y are random vectors representing the input features and target variable, respectively, $mg(X,Y)$ is the margin function that measures the difference between the average votes for the correct class and the highest average votes for any incorrect class. Additionally, $I(\cdot)$ is the indicator function, and h_k represents the individual classifiers (decision trees) in the ensemble[18].

In this study, the RF model was configured with 10 trees and a random state of 50 to ensure reproducibility and stability in the results. This configuration allows the model to balance computational efficiency with predictive performance.

b) eXtreme Gradient Boosting (XGBoost): The eXtreme Gradient Boosting (XGBoost) algorithm is an advanced implementation of the gradient boosting framework, designed to build decision trees efficiently and in parallel. This approach enhances computational speed and scalability, making it a popular choice for both classification and regression tasks. The core idea behind XGBoost is to iteratively improve the model by minimizing an objective function, which consists of a loss function and a regularization term.The objective function in XGBoost, as defined in Equation (7), is given by:

$$F_{obj}(\theta) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{k}\Omega(f_k) \tag{7}$$

where $F_{obj}(\theta)$ is the objective function to be minimized, $L(y_i, \hat{y}_i)$ is the loss function that measures the difference between the predicted value$\hat{y}_i$ and the actual value y_i, and $\Omega(f_k)$ is the regularization term that penalizes model complexity to prevent overfitting. Here, θ represents the model parameters, K is the number of trees in the ensemble, and f_k denotes the k-*th* tree in the model. The regularization term $\Omega(f_k)$ is defined separately in Equation (8) as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2 \tag{8}$$

where T is the number of leaves in the tree, $\omega_j$represents the weight of the *j-th* leaf, and γ and λ are regularization parameters that control the trade-off between model complexity and accuracy [19]. In this study, the XGBoost model was configured with a learning rate of 0.1 and 500 trees to ensure a balance between computational efficiency and predictive performance. These hyperparameters were chosen to optimize the model's ability to generalize to unseen data while maintaining reasonable training times.

B. LSTM networks

LSTM networks are a variant of recurrent neural networks (RNNs) specifically designed to capture long-term dependencies in sequential data. Unlike traditional RNNs, LSTMs mitigate the vanishing gradient problem by incorporating memory cells and gating mechanisms that regulate information flow. These properties make LSTMs particularly effective in modeling time-series data, such as wind power generation [20].

In this study, a two-layer LSTM network is used to predict wind power generation. The model is trained on historical wind speed and direction data collected at 10-minute intervals from two meteorological stations in Saveh and Tafresh, Iran. As shown in Figure 2, the LSTM architecture captures temporal patterns, enabling accurate multi-step forecasts. By learning optimal gate weights during training, the LSTM retains and utilizes relevant information from input sequences, making it highly effective for wind power forecasting, where long-term trends are crucial. After training, the LSTM takes historical wind data as input and generates future wind power predictions. Based on the connections shown in Figure 2, the mathematical expressions can be expressed as follows:

$$\textit{Forget gate}: f_t = \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \tag{9}$$

$$\textit{Input gate}: i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \tag{10}$$

$$\textit{Candidate Cell State}: \hat{c}_t = \tanh(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\hat{c}}) \tag{11}$$

$$\textit{Cell State Update}: c_t = f_t.c_{t-1} + i_t.\tilde{c}_t \tag{12}$$

$$\textit{Output Gate}: o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \tag{13}$$

$$\textit{Hidden State}: h_t = o_t.\tanh(c_t) \tag{14}$$

In the LSTM architecture, weight matrices W_{fh}, W_{fx}, W_{ih}, W_{ix}, $W_{\tilde{c}h}$, $W_{\tilde{c}x}$, W_{oh}, and W_{ox} are associated with different gates and transformations. Specifically, W_{fh} and W_{fx} correspond to the forget gate, W_{ih} and W_{ix} to the input gate,to $W_{\tilde{c}x}$ and $W_{\tilde{c}h}$ the candidate cell state, and W_{oh} and W_{ox} to the output gate.

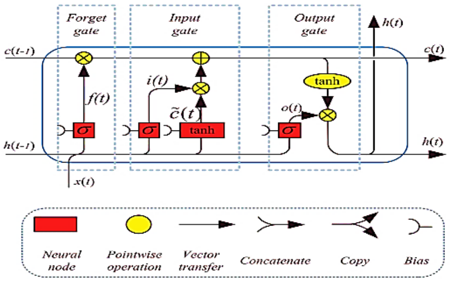

Figure 2. Architecture of LSTM with a forget gate [14]

These weights regulate the influence of the previous hidden state h_{t-1} and the current input x_t on each gate or transformation. The LSTM cell updates its internal states using the previous hidden state h_{t-1} and the current input x_t. Bias terms b_f, b_i, b_c, and b_o are added to the forget gate, input gate, candidate cell state, and output gate, respectively, to introduce shifts in activation functions and improve learning. The cell state c_t retains long-term information, while the hidden state h_t serves as the output of the LSTM cell at time t. The output gate activation o_t controls which parts of the cell state are passed forward. The candidate cell state $\tilde{c}_t$ provides potential updates to the cell state, and the input gate activation *it* regulates the flow of new information. The forget gate activation f_t determines what information from the previous cell state c_{t-1} should be retained or discarded [14].

C. Proposed Model

In the proposed model architecture (Figure 3) for predicting wind power on a daily basis, a combination of different layers is used, which operate as follows:

a) Input layer: The input data includes daily wind speed and daily wind power. This data is fed into the model as a daily time series. The dataset is split such that 80% of the annual data is used for training, and the remaining 20% is reserved for testing. The size of this layer is 128 units, representing the number of neurons used for initial data processing.

b) LSTM layers: Two LSTM layers, each with 128 units, are used to model temporal dependencies in the data. LSTMs are capable of learning both long-term and short-term patterns in time-series data. The output of these layers is a sequence of hidden states that capture temporal information

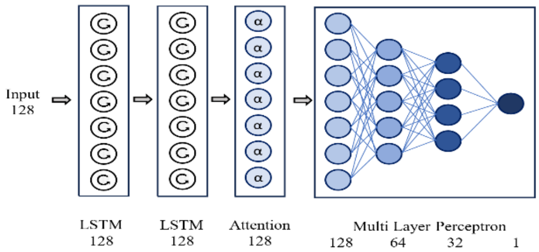

Figure 3.Hybrid LSTM-Attention model architecture

c)Attention Layer: The attention layer in this model is used to focus on the most important parts of the data and improve prediction accuracy. This layer takes the output of the LSTM layers, which consists of a sequence of hidden states represented as $H=[h_1,h_2,...,h_T]$, and computes attention weights for each hidden state. These attention weights indicate the importance of each part of the data for the final prediction. The weights are calculated using a small neural network (usually a single-layer perceptron) based on Equation 15:

$$e_t = \tanh(W_a h_t + b_a) \tag{15}$$

The attention weights α*tαt* are then computed using Equation 16:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{T}\exp(e_i)} \tag{16}$$

In these equations, W_a and b_a are learnable parameters, h_t is the hidden state at time t, and αt is the attention weight corresponding to that hidden state. Next, a context vector is created as the weighted sum of the hidden states, which provides key information for the subsequent layers. This context vector is calculated using Equation 17:

$$c = \sum_{t=1}^{T}\alpha_t h_t \tag{17}$$

The attention mechanism allows the model to better identify complex patterns and long-term dependencies in the data, ultimately improving the accuracy of wind power prediction [21].

d) Multi-Layer Perceptron (MLP) Layers: After the attention layer, a MLP with layers of 128, 64, and 32 units is used. These layers further process the extracted features, reduce data dimensionality, and refine the information for prediction. Finally, the output of these layers is passed to an output layer that predicts daily wind power [22].

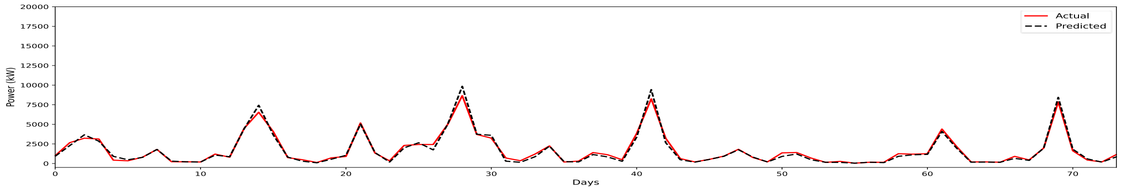

Figure 4.Daily wind power prediction for Khomein station, comparing actual (red) and predicted (black dashed) values

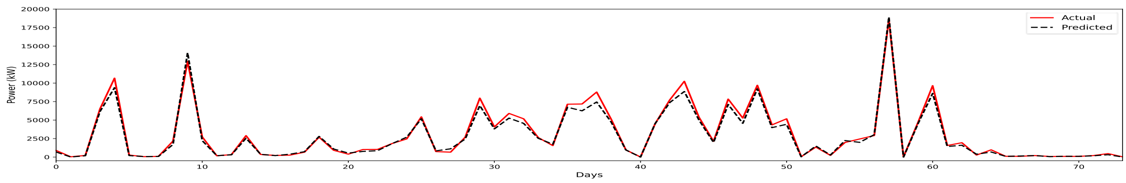

Figure 5.Daily wind power prediction for Saveh station, comparing actual (red) and predicted (black dashed) values

IV. RESULT AND DISCUSSION

In this section, the performance of the proposed hybrid LSTM-Attention model is evaluated using the test dataset, which consists of 20% of the annual data (73 days) from the Khomein and Saveh wind stations. The results are displayed in two figures, highlighting the model's effectiveness in forecasting daily wind power and its ability to capture the relationship between wind speed and power output. Figures 4 and 5 present the predicted versus actual daily wind power for the Khomein and Saveh wind stations, respectively. The predicted values are plotted against the actual values over the 73-day test period. The figures demonstrate that the proposed hybrid model closely tracks the trends and variations in the actual wind power data, showcasing its capability to manage the inherent variability in wind energy generation.

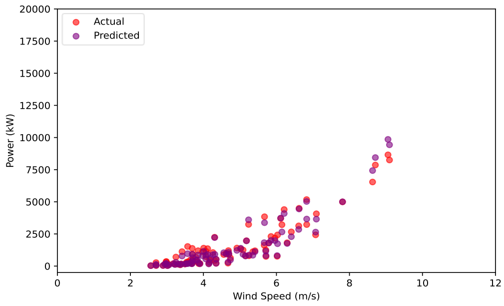

Figure 6. Scatter plot of wind power vs. speed for Khomein

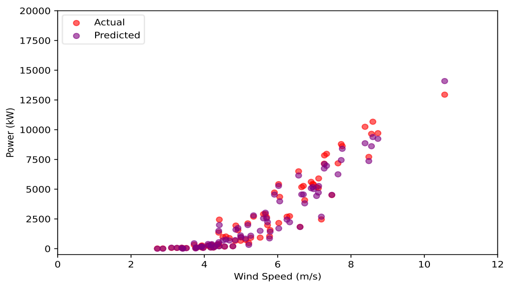

Figure 7.Scatter plot of wind power vs. speed for Saveh

Figures 6 and 7 illustrate scatter plots depicting the relationship between wind power and wind speed at the Khomein and Saveh wind stations. These visualizations demonstrate the model's capability to capture the inherent nonlinear correlation between wind speed and power output, which is essential for precise wind power forecasting. The findings confirm that the hybrid LSTM-Attention model successfully identifies and learns complex patterns within the data, even under fluctuating wind conditions.

Table II compares the performance of various models in wind power prediction at two stations, Khomein and Saveh. In machine learning-based methods, the dependency on local data characteristics is highly significant. At the Khomein station, the XGBoost model outperformed other machine learning models, such as Random Forest, while at the Saveh station, the Random Forest model achieved more accurate predictions.

However, our proposed model, which combines a Long Short-Term Memory (LSTM) network with an attention mechanism, demonstrated the best overall performance across both stations. This model effectively leveraged key feature extraction and temporal pattern recognition, significantly enhancing prediction accuracy. It has proven to be highly flexible and efficient for diverse geographical conditions and varied datasets.

TABLE II
Performance Comparison of Wind Power Prediction Models

Model	Wind Station	R^2	MAE	RMSE
RF	Khomein	0.9092	346.4	4638
XGBoost	Khomein	0.9319	315.1	3457
LSTM	Khomein	0.9521	281.2	1875
Proposed Model	**Khomein**	**0.9812**	**203.2**	**984**
RF	Saveh	0.9588	411.5	4052
XGBoost	Saveh	0.9306	523.3	8575
LSTM	Saveh	0.9612	343.0	3451
Proposed Model	**Saveh**	**0.9911**	**286.4**	**1930**

V. CONCLUSION

This study explores advanced machine learning and deep learning techniques for wind power forecasting at two meteorological stations. While traditional models like Random Forest and XGBoost perform well under certain conditions, their accuracy varies depending on local wind characteristics.

The proposed hybrid LSTM-Attention model significantly enhances forecasting precision by capturing long-term dependencies and dynamically assigning importance to key temporal features. The results suggest that integrating deep learning with attention mechanisms can substantially improve renewable energy forecasting.

References

[1] G. Zhang, Y. Wu, K. P. Wong, Z. Xu, Z. Y. Dong, and H. H.-C. Iu, "An advanced approach for construction of optimal wind power prediction intervals," *IEEE transactions on power systems,* vol. 30, no. 5, pp. 2706-2715, 2014.

[2] Z. Liang, J. Liang, C. Wang, X. Dong, and X. Miao, "Short-term wind power combined forecasting based on error forecast correction," *Energy conversion and management,* vol. 119, pp. 215-226, 2016.

[3] M. Monemi Bidgoli and R. Ghani, "Optimal Energy Management of Water-Energy Nexus in Multi-Carrier Systems Integrated with Renewable Sources," *Power, Control, and Data Processing Systems,* vol. 1, no. 1, 2024.

[4] A. Zendehboudi, M. A. Baseer, and R. Saidur, "Application of support vector machine models for forecasting solar and wind energy resources: A review," *Journal of cleaner production,* vol. 199, pp. 272-285, 2018.

[5] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Applied Energy,* vol. 304, p. 117766, 2021.

[6] C. Cakiroglu, S. Demir, M. H. Ozdemir, B. L. Aylak, G. Sariisik, and L. Abualigah, "Data-driven interpretable ensemble learning methods for the prediction of wind turbine power incorporating SHAP analysis," *Expert Systems with Applications,* vol. 237, p. 121464, 2024.

[7] M. Najjarpour, B. Tousi, and A. H. Karamali, "Optimizing Reactive Power for DG Units to Minimize Active Distribution Network Losses Using probabilistic Modeling," *Journal of Green Energy Research and Innovation,* 2024.

[8] K. Szostek, D. Mazur, G. Drałus, and J. Kusznier, "Analysis of the Effectiveness of ARIMA, SARIMA, and SVR Models in Time Series Forecasting: A Case Study of Wind Farm Energy Production," *Energies (19961073),* vol. 17, no. 19, 2024.

[9] J. Li and M. Miao, "Short-term wind power forecasting using interval A2-C1 type-2 TSK FLS method with extended Kalman filter algorithm," *Chinese Journal of Electrical Engineering,* 2024.

[10] A. Alkesaiberi, F. Harrou, and Y. Sun, "Efficient wind power prediction using machine learning methods: A comparative study," *Energies,* vol. 15, no. 7, p. 2327, 2022.

[11] H. Jin, L. Shi, X. Chen, B. Qian, B. Yang, and H. Jin, "Probabilistic wind power forecasting using selective ensemble of finite mixture Gaussian process regression models," *Renewable Energy,* vol. 174, pp. 1-18, 2021.

[12] C. Sasser, M. Yu, and R. Delgado, "Improvement of wind power prediction from meteorological characterization with machine learning models," *Renewable Energy,* vol. 183, pp. 491-501, 2022.

[13] A. Alkesaiberi, F. Harrou, and Y. Sun, "Efficient wind power prediction using machine learning methods: A comparative study," *Energies,* vol. 15, no. 7, p. 2327, 2022.

[14] A. Karamali, A. Daeichian, and S. Rezaei, "Using Long Short-Term Memory Networks as Virtual Wind Direction Sensors for Improved Wind Farm Turbines Orientation," in *2024 9th International Conference on Technology and Energy Management (ICTEM)*, 2024: IEEE, pp. 1-5.

[15] S. Hanifi, A. Cammarono, and H. Zare-Behtash, "Advanced hyperparameter optimization of deep learning models for wind power prediction," Renewable Energy, vol. 221, p. 119700, 2024.

[16] H. Demolli, A. S. Dokuz, A. Ecemis, and M. Gokcek, "Wind power forecasting based on daily wind speed data using machine learning algorithms," *Energy Conversion and Management,* vol. 198, p. 111823, 2019.

[17] M. Yang, B. Dai, J. Wang, X. Chen, Y. Sun, and B. Li, "Day-ahead wind power combination forecasting based on corrected numerical weather prediction and entropy method," *IET Renewable Power Generation,* vol. 15, no. 7, pp. 1358-1368, 2021.

[18] K. L. Jørgensen and H. R. Shaker, "Wind power forecasting using machine learning: State of the art, trends and challenges," in *2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE),* 2020: IEEE, pp. 44-50.

[19] Q.-T. Phan, Y.-K. Wu, and Q.-D. Phan, "A comparative analysis of xgboost and temporal convolutional network models for wind power forecasting," in *2020 International Symposium on Computer, Consumer and Control (IS3C),* 2020: IEEE, pp. 416-419.

[20] F. Shahid, A. Zameer, and M. Muneeb, "A novel genetic LSTM model for wind power forecast," *Energy,* vol. 223, p. 120069, 2021.

[21] S. Wang, J. Shi, W. Yang, and Q. Yin, "High and low frequency wind power prediction based on Transformer and BiGRU-Attention," *Energy,* vol. 288, p. 129753, 2024.

[22] S. Samadianfard *et al.*, "Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm," *Energy Reports,* vol. 6, pp. 1147-1159, 2020.