Wind Power Forecasting with a Hybrid Deep Learning Approach including LSTM and Attention Mechanism

Saber Rezaei[1], Abolghasem Daeichian [2,3*], Ali Hedayati[1], Amir Hossein Karamali[1]

[1] Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran

[2] Department of Electrical Engineering, Faculty of Engineering, Arak University, Arak, 38156-8-8349, Iran

[3] Research Institute of Artificial Intelligence, Arak University, Arak, 38156-8-8349, Iran

***Abstract*—** The intermittent nature of wind energy generation poses substantial challenges to the stability of electrical grids and the efficiency of energy management systems. Accurate and reliable wind power forecasting is therefore critical for a multitude of reasons: to optimize the operational efficiency of grids, to effectively balance energy supply and demand, to enhance the planning and execution of energy storage strategies, to minimize the reliance on backup power sources, and ultimately, to reduce operational costs within renewable energy infrastructures. This study introduces a novel hybrid deep learning approach designed to improve the accuracy of wind power forecasting through the integration of Long Short-Term Memory (LSTM) networks with an attention mechanism. The model's efficacy was rigorously evaluated using high-resolution data, recorded at 10-minute intervals, from two distinct meteorological stations located in Khomein, Saveh and Tafresh, Iran. The performance of the hybrid model was benchmarked against traditional machine learning methodologies, including Random Forest (RF), XGBoost, and standalone LSTM networks. The results of the evaluation demonstrate the superior performance of the hybrid LSTM-Attention model, which achieved notable coefficient of determination (R^2) values of 0.9812, 0.9911 and 0.9842 at the Khomein, Saveh and Tafresh stations, respectively, indicating significant advancements in forecasting accuracy compared to the other models. These enhanced forecasting capabilities have significant implications for facilitating the efficient integration of wind energy into electrical grids, thereby enabling more effective grid management practices and supporting optimized energy distribution strategies.

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

1. Introduction

The growing worldwide demand for clean and sustainable energy solutions has significantly boosted interest in wind power as a viable and eco-friendly alternative to traditional fossil fuels. However, the intermittent and uncertain nature of wind energy generation poses major challenges for maintaining grid stability and ensuring efficient energy management. Unlike conventional power sources, wind energy output fluctuates due to changing weather conditions, making it difficult to predict and integrate into the power grid seamlessly. For this reason, accurate and reliable wind power forecasting has become essential—not only to improve grid operations and balance supply with demand but also to optimize energy storage strategies, reduce reliance on backup power sources, and ultimately lower operational costs. By enhancing prediction models, energy providers can better manage fluctuations, reduce waste, and support a smoother transition to renewable energy systems [1], [2], [3]. Over the years, a wide range of methods have been developed for wind power forecasting, which are typically grouped into three core categories: physical models, statistical methods, and machine learning techniques. Physical models rely heavily on meteorological variables—such as wind speed, atmospheric pressure, temperature, and humidity—to estimate the expected power output. These models often make use of Numerical Weather Prediction (NWP) tools, which simulate large-scale atmospheric dynamics to forecast wind behavior. Although physical models are particularly effective for medium- to long-term forecasting due to their grounding in physical laws, they come with significant limitations. Their accuracy is highly dependent on the quality and resolution of the input data, and they require substantial computational resources. As a result, their use in real-time or short-term forecasting scenarios is often restricted due to latency and complexity concerns [4], [5]. Furthermore, the performance of physical models is highly sensitive to the precision of meteorological input data such as wind speed and atmospheric pressure which can be challenging to acquire reliably in remote or data-scarce 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 forecasting of wind power. These techniques identify linear relationships in historical wind power data and perform well under consistent wind conditions. However, their performance tends to degrade in the presence of strong fluctuations, as they struggle to capture the nonlinear, chaotic, and context-dependent dynamics often inherent in wind power generation, especially under complex or rapidly changing weather scenarios [8]. To overcome these constraints, more advanced statistical techniques have been proposed to improve prediction accuracy and adapt to the dynamic nature of wind power. One notable example is Kalman filtering, a recursive estimation method that continuously updates its predictions by integrating new measurements over time. This approach not only refines wind speed forecasts in real-time but also helps mitigate the impact of measurement noise and uncertainty, making it particularly effective in environments where data is noisy or partially missing. Its adaptability and real-time updating capability make it a valuable tool in short-term forecasting applications where rapid response to changing conditions is essential [9]. Similarly, exponential smoothing methods have been effectively employed to model trends and seasonal variations in wind power data, offering improved robustness and responsiveness in short-term forecasting scenarios [10]. Another notable advancement in statistical forecasting is the application of Gaussian Processes (GPs), which provide a flexible and robust probabilistic framework for modeling both the central tendency and the uncertainty in wind power generation. Unlike deterministic models that offer a single forecast value, GPs generate a distribution over possible outcomes, allowing for the estimation of prediction intervals that capture the range of plausible future values. This uncertainty-aware approach is especially advantageous for grid operators, as it facilitates more informed decision-making under uncertainty, enabling better risk assessment and resource allocation in the context of highly variable wind energy systems [11]. Despite significant advancements in physical and statistical forecasting methods, the inherent complexity, nonlinearity, and dynamic nature of wind power data have pushed researchers to seek more flexible and adaptive solutions. In recent years, the emergence of Artificial Intelligence (AI) and Machine Learning (ML) techniques has transformed the landscape of wind power forecasting by enabling data-driven models that can autonomously learn from large volumes of historical data. Techniques such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Ensemble Boosting algorithms have shown considerable promise, often outperforming traditional models by capturing complex temporal dependencies and nonlinear patterns that are difficult to model analytically [10], [12], [13]. Among these, Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) networks have gained significant traction due to their exceptional ability to model long-term temporal dependencies and capture complex, non-linear relationships in sequential data. This characteristic makes them particularly well-suited for wind power prediction tasks, where the output at any given time is influenced by past wind behavior over varying time scales. Their ability to maintain information over extended periods allows them to effectively model the intricate patterns in wind speed fluctuations, which are key to accurate short-term and medium-term forecasting [14], [15]. A study proposes an integrated energy hub (electrical, thermal, cooling, water) using LSTM for load forecasting and multi-objective optimization to cut operating costs and groundwater use. Incorporating renewables, EVs, responsive loads, and storage, the system reduced groundwater extraction by 26.88% in a case study [16]. This paper presents a solution for a smart grid retailer to determine electricity selling prices and manage energy under uncertainty, considering diverse energy sources like renewables, electric vehicles, and hydrogen storage. A neural network predicts uncertainties, while multi-objective optimization aims to maximize retailer profits (increased by 10.2-12.8% under different tariffs) and reduce groundwater usage, formulated as a mixed-integer problem solved in GAMS [17].

The primary novelty of this research lies in the development and rigorous evaluation of a novel hybrid deep learning architecture that synergistically integrates Long Short-Term Memory (LSTM) networks with a dedicated attention mechanism for enhanced daily wind power forecasting. While LSTMs are adept at capturing long-range temporal dependencies inherent in wind data, the incorporated attention layer further refines the model's predictive capabilities by dynamically identifying and emphasizing the most influential temporal features within the input sequences. This tailored combination allows the model to more effectively discern intricate, non-linear patterns and adapt to the highly variable nature of wind resources. The efficacy of this specific LSTM-Attention configuration is demonstrated through its superior forecasting accuracy on high-resolution datasets from three distinct meteorological sites in Iran, outperforming traditional machine learning algorithms and standalone LSTM networks, thereby offering a more robust and precise tool for wind energy prediction.

The following key contributions and methodologies were employed in this work:

- **Machine Learning (ML) Algorithms**: Machine learning techniques, including Random Forests (RF) and XGBoost, were applied to predict wind power generation. These methods were selected for their robustness in handling complex datasets and providing reliable predictions. Both models were trained using historical wind speed and power data.

- **Deep Learning (DL) Approach**: A two-layer Long Short-Term Memory (LSTM) deep neural network was implemented to model the temporal dependencies in the wind power data. LSTM networks excel at learning long-term patterns in sequential data, making them effective for time-series forecasting tasks like wind power prediction.
- **Hybrid Model**: A hybrid approach combining LSTM with an Attention Mechanism was proposed and evaluated. This hybrid model demonstrated superior performance compared to standalone models, as it effectively highlights the most important temporal features in the data. This approach enhances the model's ability to adapt to changes in wind patterns, resulting in a more robust and accurate forecasting model.
- **Improved Forecasting Accuracy**: The study highlights the improved forecasting accuracy achieved with the hybrid LSTM-Attention model. The model's enhanced performance makes it a valuable tool for grid management, where accurate wind power forecasting can optimize energy storage planning and facilitate the integration of renewable energy sources into the power grid.

This study underscores the utility of machine learning and deep learning techniques in tackling the complexities of wind power prediction, while also establishing a basis for future investigations and uses in renewable energy systems. The remainder of this paper is structured as follows: Section 2 details the foundational aspects and the formulation of the problem, which includes the dataset employed and the specifications of the wind turbine. Section 3 elaborates on the methodology, covering the machine learning algorithms, the Long Short-Term Memory (LSTM) networks, and the proposed hybrid LSTM-Attention model. Section 4 presents and deliberates on the experimental outcomes from the evaluations performed at the Khomein, Saveh, and Tafresh stations. Lastly, Section 5 offers the conclusion, summarizing the main discoveries and proposing avenues for subsequent research.

2. Related Works

Recent research has significantly advanced wind power forecasting methods using artificial intelligence techniques. A groundbreaking study [18] introduced a hybrid model that dramatically improves prediction accuracy at meteorological stations in Khomein, Saveh and Tafresh, Iran, outperforming traditional methods by effectively capturing complex temporal dynamics in wind patterns. This work is extended version of previous study [18] builds upon comprehensive wind energy potential assessments in the same region [19] which combined statistical approaches with LSTM networks to analyze high-resolution wind data collected at 10-minute intervals over two years. Together, these studies establish a robust framework for wind energy forecasting and analysis, offering valuable insights for grid stability, energy management, and sustainable planning in renewable energy development.

A short-term wind power forecasting method combines discrete wavelet transform (DWT) for signal decomposition with LSTM networks to predict individual, more stationary components. Synthesizing these predictions enhances accuracy and supports grid stability [20]. A study employing Long Short-Term Memory (LSTM) neural networks for improved wind power forecasting demonstrated enhanced prediction accuracy, achieving a Root Mean Squared Error (RMSE) of 0.6782 and a Mean Absolute Error (MAE) of 0.4614. These findings suggest the model's potential to enhance power system stability and support more reliable integration of wind energy through more efficient forecasting [21]. This study focused on improving short-term wind power forecasting (1 to 6 hours ahead) introduced a novel approach based on a Rolling-LSTM (R-LSTM) model. This model employs a recursive strategy, diverging from conventional LSTM methods, and was implemented using historical wind power data from Gujarat state. Comparative analysis demonstrated that the proposed R-LSTM model outperformed existing approaches presented in the literature, achieving better accuracy and minimal error [22]. To enhance wind power forecast accuracy, a study proposed an Improved Long Short-Term Memory (ILSTM) network. The method first uses Variational Mode Decomposition (VMD) to separate the wind power signal into long-term, fluctuation, and random components for model input. A parameter was added to the ILSTM's memory cell to suppress the random component's long-term memory impact, and the output gate was modified to allow passage of the current random component, thereby improving learning of true patterns and avoiding over-fitting. The model's performance was validated using wind power data from the Belgian ELIA website [23].

These diverse AI-driven approaches illustrate significant advancements in wind power forecasting. Signal decomposition techniques, as seen with DWT in DWT-LSTM and VMD in the ILSTM approach, offer the advantage of improving forecast accuracy by breaking down non-stationary wind power series into components that are more regular or easier to predict. A potential trade-off for such multi-stage models can be increased overall model complexity. Specialized LSTM architectures, such as R-LSTM with its recursive strategy for specific short-term forecasting and ILSTM with modifications designed for better pattern learning and reduced overfitting, demonstrate the benefits of tailoring models to particular forecasting challenges, leading to improved accuracy in those contexts. However, highly specialized models might require careful tuning for optimal performance. Even well-established architectures like LSTM, when applied effectively, can significantly enhance prediction accuracy, as shown in [21]. Hybrid models, exemplified by the work in [18], capitalize on combining different methodologies to capture complex temporal dynamics more effectively, thereby achieving dramatic improvements in prediction. While powerful, the integration of multiple components in hybrid systems can lead to more complex model structures. A common consideration across these advanced methods is balancing the pursuit of higher accuracy with factors like model complexity, computational demands, and the specificity of the approach to certain data types or forecasting horizons.

While the reviewed studies demonstrate considerable progress, a distinct research gap remains in developing forecasting models that can achieve high accuracy and robustness across varied wind conditions without imposing excessive preprocessing burdens or relying on highly specialized architectures that may limit generalizability. Many advanced methods involve multi-stage processing (e.g., DWT-LSTM, VMD-ILSTM) which, despite improving accuracy, adds complexity. Other specialized LSTMs (e.g., R-LSTM) are tailored for specific scenarios and may not offer broad adaptability. There is still a significant need for hybrid models that can intrinsically adapt to the dynamic and non-linear nature of wind patterns and selectively focus on the most informative temporal features with less manual intervention or architectural rigidity. Specifically, the optimal integration and exploration of attention mechanisms with LSTMs—designed to dynamically weigh the significance of different time steps—presents a promising avenue. This approach could enhance predictive performance by more effectively capturing salient temporal dependencies in a data-driven manner, potentially offering a more nuanced, adaptable, and computationally balanced alternative to existing complex hybrid systems or extensive feature engineering for wind power forecasting.

To address this identified research gap, the hybrid LSTM-Attention model proposed in this study is specifically designed to offer a robust, accurate, and adaptable solution. By synergistically combining the capacity of LSTM networks to model long-term temporal dependencies with an attention mechanism that dynamically highlights the most relevant past information, our approach effectively captures the complex, non-linear patterns inherent in wind power data. This targeted integration aims to enhance predictive precision beyond that of standalone models or less adaptive techniques, offering a more nuanced way to handle wind's variability. The rigorously evaluated superior performance of our model, achieving high coefficient of determination (R^2) values across multiple meteorological stations, underscores its potential as an advanced and adaptable tool for real-world wind power forecasting, thereby addressing some of the inherent complexities highlighted and contributing to more effective grid management.

3. materials and methods

3.1. 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, providing high temporal resolution for precise modeling of wind behavior. The data is collected from two meteorological stations located in Khomein, Saveh and Tafresh, both situated in the Markazi Province of Iran. These stations offer detailed, high-resolution data on key wind-related parameters, including wind speed and Wind degree that are essential for accurate wind power prediction. The combination of high temporal resolution and the geographical diversity of the stations makes this dataset particularly valuable for modeling and analyzing the variability and patterns of wind energy generation in this region, where wind conditions can vary significantly across different locations. This rich dataset helps in capturing local variations and improving the robustness of wind power forecasting models.

The selection of the meteorological stations in Khomein, Saveh, and Tafresh, all situated within the Markazi Province of Iran, was driven by several key considerations critical for developing and validating a robust wind power forecasting model. Primarily, these stations provided access to comprehensive, high-resolution wind data recorded at 10-minute intervals over a one-year period.

This dataset included crucial meteorological parameters such as wind speed, direction, and temperature, essential for accurate wind behavior modeling. The high temporal resolution of this data is particularly valuable for capturing the nuanced dynamics of wind patterns necessary for precise short-term forecasting.

Furthermore, these specific locations were chosen to capitalize on their geographical diversity within the Markazi Province. As noted in our foundational analysis, wind conditions can vary significantly across different locations even within the same region, influenced by local topography and microclimates. Utilizing data from these distinct sites, therefore, allowed for a more rigorous evaluation of the proposed model's ability to adapt and perform reliably under varied wind regimes. This aspect was crucial for assessing the model's robustness beyond a single, potentially homogeneous dataset.

Regarding the broader applicability of findings, the inherent variability across these selected Iranian stations serves as a constructive testbed, reflecting some of the diverse conditions encountered in wind energy generation sites globally. The consistent high performance and adaptability demonstrated by the proposed hybrid LSTM-Attention model across these geographically distinct locations—each presenting unique wind patterns and local variations—underscore its strong generalization capabilities. This suggests that the model architecture, particularly its ability to effectively learn temporal dependencies and highlight salient features through the attention mechanism, is not merely tailored to specific local conditions. Instead, it possesses the robustness and adaptability to be effectively applied across different geographical regions, diverse wind regimes, and varying topographic settings, making it a potentially valuable tool for wider renewable energy forecasting applications.

The raw input data, consisting of wind speed (m/s) and Wind Degree (between 0 to 360 Degree), was first thoroughly inspected for completeness and integrity; no significant missing or lost data points were identified, ensuring the dataset was robust. Subsequently, all relevant numerical features used for model input were normalized to a range of [0, 1]. This normalization step was performed prior to feeding the data into the machine learning and deep learning models utilized in this study.

Energy output from wind turbines can be calculated using wind speed data and the power curves supplied by the turbine manufacturers. These curves illustrate how wind speed affects the amount of electrical power the turbine generates, allowing for accurate predictions of energy production based on real-time wind conditions [24]. This methodology involves integrating the turbine's power curve with time-series wind speed data to estimate the power generated over time. In this study, a 1 MW wind turbine is selected for the analysis due to its widespread application in wind power plants and its suitability for the specific wind conditions in the region. The power curve for this turbine, which illustrates the relationship between wind speed and power output, is shown in Fig. 1.

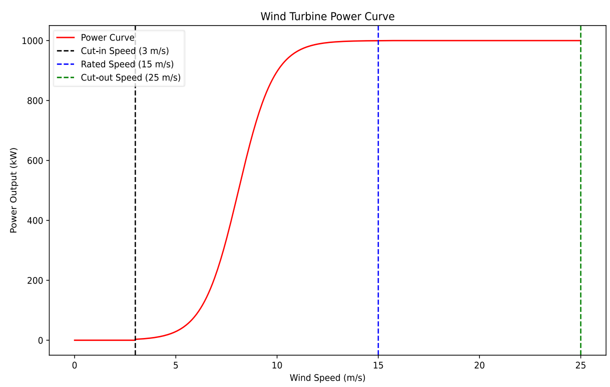

Figure 1. Power output curve of a 1 MW wind turbine. [16]

For the purpose of estimating the energy yield from wind resources, an *n*-th degree polynomial expression is formulated on the basis of the turbine's power characteristic. The applicability of this equation extends across the interval demarcated by the cut-in speed (v_{ci}) and the rated speed (v_R), or alternatively, between the cut-in speed and the cut-out speed (v_{co}). Its mathematical form is presented hereunder (Equation 1) [24]:

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

The parameters of this model include the regression coefficients, denoted as $a_n, \ldots, a_0$, which are empirically derived from the turbine's power characteristic. Furthermore, the rated power output is designated as P_R , and $P_i(v)$ represents the instantaneous power generation at a given wind speed.

The total energy output, denoted as E, over a defined temporal interval can be ascertained through the application of Equation 2 [24].

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

Within this framework, N represents the total number of temporal segments constituting the period under consideration, with $P(vi)$ denoting the power output realized at a specific wind velocity vi , and Δt symbolizing the duration of each individual time interval, typically on the order of ten minutes. The utilization of this methodology facilitates the generation of highly accurate and dependable forecasts of wind energy generation, a factor of paramount importance for the seamless integration of wind power into the electricity grid and for strategic energy planning endeavors [19].

Table I presents a comprehensive overview of the technical specifications pertaining to the selected turbine.

TABLE 1

Technical Characteristics of the Evaluated Wind Turbine [16]

Evaluated Turbine	Technical Parameters
Nominal Power Output (MW)	1.0
Tower Height(m)	50.0
Cut-in wind speed (m/s)	3.0
Rated wind speed (m/s)	15.0
Cut-out wind speed (m/s)	25.0

3.2 methodology

The subsequent part is delineated into three primary segments: namely, an examination of machine learning techniques, a dedicated analysis of the two-layer LSTM architecture, and a comprehensive presentation of the novel model introduced herein. The predictive capabilities of these models are quantitatively gauged through the application of established metrics, such as the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), and the coefficient of determination (R^2) [5].

The coefficient of determination (R^2), as formulated in Equation (3) [25], provides a quantitative measure of the degree to which variations observed in the dependent variable are attributable to the influence of the independent variables. This particular metric serves to assess the model's proficiency in discerning underlying relationships within the given dataset, while simultaneously offering an indication of its capacity for accurate prediction.

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

The Mean Absolute Error (MAE), as delineated in Equation (4) [26], ascertains the arithmetic average of the absolute differences separating the predicted and actual values. This measure affords an uncomplicated and easily grasped appraisal of the deviations inherent in the model's projections, thus providing a lucid view concerning its performance fidelity.

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

The Root Mean Squared Error (RMSE), as stipulated in Equation (5) [26], ascertains the square root of the arithmetic mean of the squared differences separating the predicted and observed values. This particular metric offers a resilient assessment of the model's predictive precision, consequent to its property of magnifying the impact of larger deviations through the inclusion of the squared term.

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

Where in equations (3), (4) and (5) y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}_i$ is the mean of the actual values, N is the total number of observations.

3.3 Machine learning

Owing to their inherent capability to model intricate and nonlinear interdependencies linking diverse input parameters with the forecasted power generation, machine learning methodologies find widespread adoption in the domain of wind power prediction. Specifically, in the present investigation, two ensemble learning algorithms, namely Random Forest (RF) and Extreme Gradient Boosting (XGBoost), were chosen for the purpose of forecasting wind power output at two distinct meteorological sites situated in Khomein, Saveh and Tafresh, Iran. The rationale underpinning the selection of these algorithms stems from their demonstrated robustness when processing high-dimensional datasets and their notable aptitude for discerning complex, nonlinear relationships among the various input characteristics. The dataset employed in this analysis comprises crucial meteorological parameters, specifically daily wind speed (dailyWS) and the daily standard deviation of wind speed (dailySD), with data acquisition conducted at 10-minute increments across the entirety of the year. The designated target variable for this forecasting endeavor is the daily power output (dailyPower). For the purpose of conducting a rigorous evaluation of the models and guaranteeing the dependability of the outcomes, the dataset was partitioned into two subsets: 80% allocated for training the models and the remaining 20% reserved for independent testing.

a) Random Forest (RF)**:** To enhance predictive accuracy and simultaneously mitigate the risk of overfitting, the Random Forest (RF) algorithm employs an ensemble learning approach that involves the construction of numerous decision trees during the training phase, with their individual outputs subsequently integrated. This methodological approach entails the segmentation of the input dataset into smaller, more manageable subsets, for each of which a distinct decision tree is constructed. The ultimate prediction is obtained by synthesizing the outputs generated by all the trees; this synthesis typically involves the application of majority voting in the case of classification problems, or the computation of an average for regression tasks. Furthermore, this methodology contributes to the simplification of intricate problems through their decomposition into constituent components that are more amenable to interpretation. Within the RF algorithm, a randomly chosen subset of features, designated as k, is sampled from the entire feature space for each individual tree. Subsequently, each of these trees undergoes training utilizing this selected feature subset in conjunction with a bootstrap sample derived from the training dataset. The generalization error associated with the RF model provides a measure of its predictive accuracy on previously unobserved data, whereas the margin function serves to quantify the level of confidence in its classification outputs. The mathematical formulations of these concepts are presented in Equation (6) [27].

$$PE = P_{X,Y}(mg(X,Y) < 0\, mg(X,Y) = avg_k[I(h_k(X) = Y)] - max_{j\neq Y}\, avg_k[I(h_k(X) = j)] \quad (6)$$

Within this analytical context, X and Y are conceptualized as random vectors, with X corresponding to the input features and Y to the target variable. The function *mg(X,Y)* serves to quantify the margin, which is defined as the disparity between the mean of the votes garnered by the correct class and the highest mean of the votes received by any of the incorrect classes. Moreover, $I(\cdot)$ operates as the indicator function, whereas h_k denotes the individual classifiers (decision trees) that collectively constitute the ensemble model [20].

In this study, the RF model was delineated with a configuration incorporating ten trees and a random state value of 50, a setup designed to guarantee both the reproducibility and the consistency of the obtained results, thereby enabling the model to achieve an optimal equilibrium between the demands of computational efficiency and the level of predictive performance attained.

***b) eXtreme Gradient Boosting (XGBoost)*:** Representing an advanced and optimized iteration of the gradient boosting paradigm, the XGBoost algorithm is meticulously engineered for the highly efficient construction of decision trees. A prominent advantage of XGBoost resides in its capacity for parallel processing, a feature that markedly boosts computational speed and bolsters scalability. Consequently, XGBoost has emerged as a highly favored option across a diverse spectrum of machine learning applications, encompassing both classification and regression challenges. XGBoost operates through an iterative model refinement process, wherein an objective function is minimized. This objective function comprises two integral elements: a loss term responsible for quantifying the discrepancies in the predictions, and a regularization term that imposes a penalty on model complexity with the aim of precluding overfitting. This optimization regimen facilitates the construction of decision trees with heightened efficiency and efficacy, a process that is mathematically formalized in Equation (7) [28].

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

Within this particular formulation, $F_{obj}(\theta)$ is defined as the objective function slated for minimization. The loss function, denoted as $L(y_i,\hat{y}_i)$ serves to quantify the divergence between the forecasted value $\hat{y}_i$ and the actual outcome y_i.

Serving as a regularization component, the term $\Omega(f_k)$ functions to constrain the complexity of the model, thereby mitigating the risk of overfitting. Within this context, θ symbolizes the parameters of the model, K corresponds to the total count of trees that collectively form the ensemble, and f_k refers to the *k-th* individual tree within the established framework. The precise mathematical formulation of the regularization term $\Omega(f_k)$ is delineated separately in Equation (8) [28].

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

Within this formulation, T corresponds to the total count of leaves within the decision tree, while ω_j signifies the weight allocated to the j-th leaf. The parameters γ and λ operate as regularization factors that contribute to maintaining a balance between the model's complexity and its predictive accuracy [21]. To maximize the predictive performance of the XGBoost algorithm for wind power forecasting in this study, its hyperparameters were meticulously optimized. The adopted model configuration featured a learning rate of 0.1 and incorporated 500 decision trees. These specific parameters were systematically refined through a process of rigorous experimentation to attain an optimal trade-off between the computational expenditure and the accuracy of the forecasts. This particular combination enables the model to proficiently acquire intricate patterns from the training dataset while simultaneously preserving robust generalization capabilities when applied to unobserved data. The moderate learning rate serves to preclude overfitting by regulating the influence exerted by each individual tree, whereas the substantial number of trees ensures the model possesses sufficient complexity to capture the inherent nonlinear relationships present in wind power generation data. Additionally, the training procedure incorporated early stopping mechanisms, which aimed to optimize computational efficiency without detracting from the model's overall performance. This well-considered approach not only augments the model's predictive reliability but also ensures its practical viability for real-time forecasting applications within the context of power system operations.

3.4 LSTM networks

Long Short-Term Memory (LSTM) networks represent a specialized type of Recurrent Neural Network (RNN) architecture developed to effectively model temporal relationships in sequential data. These networks overcome the limitations of standard RNNs by introducing a sophisticated system of memory cells and regulatory gates that manage information flow. The unique gating mechanism (comprising input, forget, and output gates) allows LSTMs to selectively retain or discard information, thereby solving the vanishing gradient problem that plagues conventional RNNs. This architectural innovation gives LSTM networks superior capability in learning and maintaining long-range temporal patterns, making them particularly valuable for time-series forecasting tasks. In renewable energy applications, LSTMs demonstrate exceptional performance in wind power prediction by effectively capturing the complex, nonlinear relationships and temporal dynamics inherent in wind generation data. Their ability to process sequential information while preserving important long-term dependencies makes them ideally suited for modeling the variable nature of wind energy production [29].

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 Khomein, 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. Referring to the connections illustrated in Figure 2, the corresponding mathematical expressions are formulated as follows [30]:

$$Forgetgate: f_t = \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \quad (9)$$

$$Inputgate: i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \quad (10)$$

$$CandidateCellState: \hat{c}_t = tanh(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\hat{c}}) \quad (11)$$

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

$$OutputGate: o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \quad (13)$$

$$HiddenState: h_t = o_t.tanh\ (c_t) \quad (14)$$

Within the framework of the LSTM architecture, the weight matrices designated as W_{fh}, W_{fx}, W_{ih}, W_{ix}, $W_{\tilde{c}h}$, $W_{\tilde{c}x}$, W_{oh}, and W_{ox} are intrinsically linked to the various gates and transformations. More precisely, the weight matrices, W_{fh} and W_{fx} are associated with the forget gate, while W_{ih} and W_{ix} are connected to the input gate. The candidate cell state is linked to $W_{\tilde{c}h}$ and $W_{\tilde{c}x}$, and the output gate utilizes W_{oh}, and W_{ox}

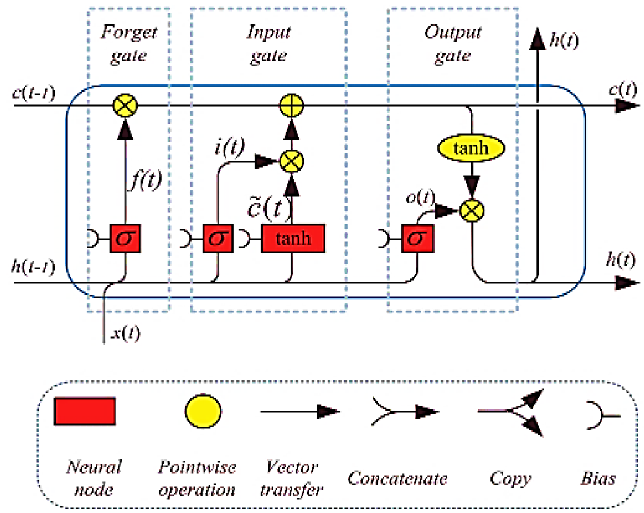

Figure 2. An LSTM architecture characterized by the inclusion of a forget gate [14].

The aforementioned weights dictate the influence exerted by the preceding hidden state h_{t-1} and the current input x_t upon each respective gate or transformation. The internal states of the LSTM cell undergo modification through the incorporation of both the prior hidden state h_{t-1} and the present input x_t. Furthermore, bias components b_f, b_i, b_c, and b_o are integrated into the forget, input, candidate cell state, and output gates, in that order. These biases serve the purpose of adjusting the shifts of the activation functions and, consequently, enhancing the efficiency of the learning process.

At a given time t, the hidden state h_t functions as the output of the LSTM cell, whereas the cell state c_t is responsible for preserving long-term informational dependencies. The activation of the output gate o_t dictates the specific components of the cell state that are propagated onward. The candidate cell state $\tilde{c}_t$ furnishes potential modifications to the cell state, with the input gate activation *it* governing the influx of new information. Furthermore, the activation of the forget gate f_t establishes which information from the preceding cell state c_{t-1} is to be either preserved or disregarded [14].

3.5 Proposed Model

The proposed model architecture (alluded to in Figure 3), engineered for the daily forecasting of wind power, incorporates a synthesis of distinct layers that function in the ensuing manner:

a) Input layer**:** Comprising daily measurements of wind speed and wind power, the input dataset is organized as a daily time series, which is subsequently supplied to the model in an appropriate manner. The dataset undergoes partitioning, with eighty percent (80%) of the annual records being allocated for the training phase, and the residual twenty percent (20%) earmarked for the testing phase. This particular layer is composed of 128 units, which correspond to the number of neurons employed in the preliminary data processing stage.

b) LSTM layers**:** To effectively model the temporal dependencies inherent in the data, the architecture incorporates two LSTM layers, each comprising 128 units. LSTMs possess the inherent capability to acquire both long-range and short-range temporal patterns present in time-series data. The output generated by these layers constitutes a sequence of hidden states that effectively encapsulate the temporal information.

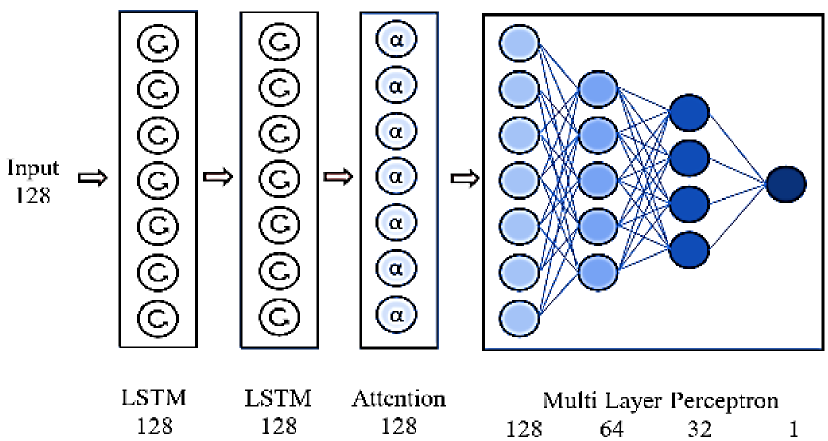

Figure 3. architectural of a hybrid model combining LSTM and Attention mechanism.

*c)**Attention Layer***: Within this model, the attention layer is engineered to accentuate the most pertinent features of the data, consequently improving the accuracy of predictions. Receiving the output of the LSTM layers, which takes the form of a sequence of hidden states denoted as $H=[h_1, h_2, \dots, h_T]$, this layer then calculates attention weights for each individual hidden state. These attention weights signify the importance of each corresponding segment of the data in contributing to the final prediction. The derivation of these weights is accomplished using a lightweight neural network, typically realized as a single-layer perceptron, as explicitly defined in Equation 15 [31].

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

Subsequently, the attention weights, denoted as *αt* are calculated by means of Equation 16 [31].

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

Within the framework of these calculations, W_a and b_a are considered to be learnable parameters. Furthermore, h_t denotes the hidden state at time t, and αt signifies the attention weight associated with that particular hidden state. Subsequently, a context vector is constructed as the weighted summation of the hidden states. This vector serves to furnish crucial information to the layers that follow. The computation of this context vector is performed utilizing Equation 17 [31].

$$c = \sum_{t=1}^{T} \alpha_t h_t \quad (17)$$

The employment of the attention mechanism enhances the model's capacity to discern intricate patterns and capture long-range dependencies within the data, thereby culminating in an augmentation of the accuracy of wind power forecasting [23].

d) Multi-Layer Perceptron (MLP) Layers: Subsequent to the attention layer, a Multi-Layer Perceptron (MLP) is employed, incorporating layers with 128, 64, and 32 units, respectively. These interconnected layers are tasked with the further processing of the extracted features, the reduction of data dimensionality, and the refinement of the information in preparation for the prediction phase. Ultimately, the output emanating from these layers is transmitted to a dedicated output layer, the function of which is to forecast the daily wind power [24]. Figure 4 illustrates schematic of overall methodology section.

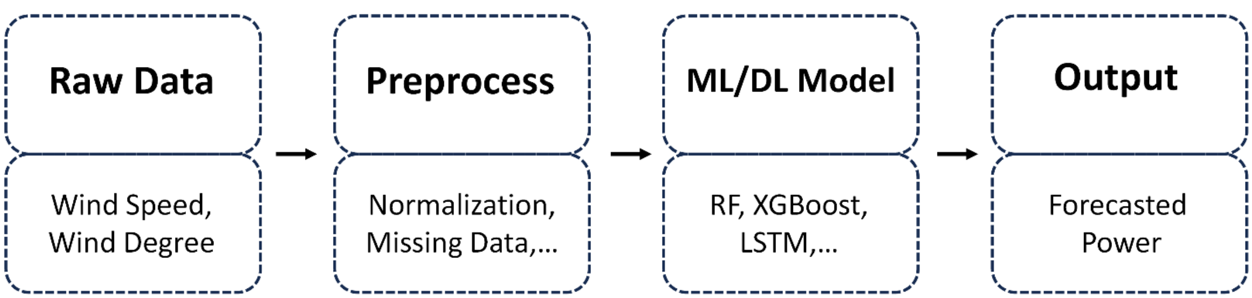

Figure 4. Overall schematic of methodology

4. RESULT

An evaluation of the proposed LSTM-Attention hybrid model's performance is presented in this section, utilizing a test dataset consisting of 20% of the annual records (totaling 73 days) gathered from the Khomein**,** Saveh and Tafresh wind stations. The results of this assessment are visually conveyed in two figures, underscoring the model's proficiency in forecasting daily wind power and its efficacy in discerning the correlation between wind speed and power generation. Specifically, a comparison between the predicted and actual daily wind power outputs for the Khomein, Saveh and tafresh wind stations is graphically represented in Figure 5 , Figure 6 and Figure 7 respectively. Across the entirety of the 73-day evaluation timeframe, the predicted values are plotted in conjunction with the actual observations. These graphical representations unequivocally demonstrate the model's capacity to accurately track the inherent fluctuations and patterns present in real-world wind power data, thereby showcasing its effectiveness in managing the intrinsic unpredictability associated with wind energy generation.

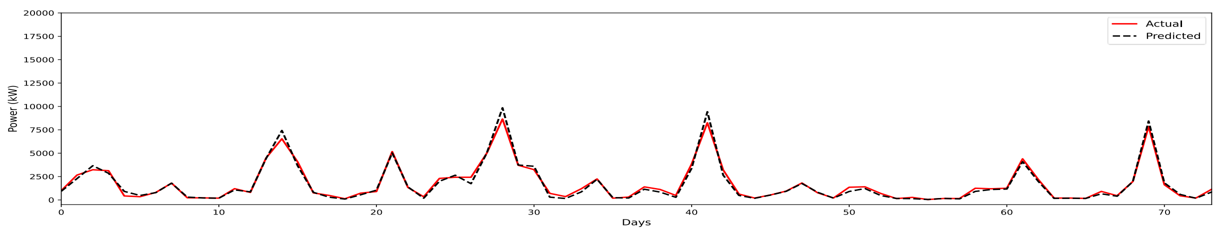

Figure 5. Daily wind power prediction for Khomein station, comparing actual (red) and predicted (black dashed) values

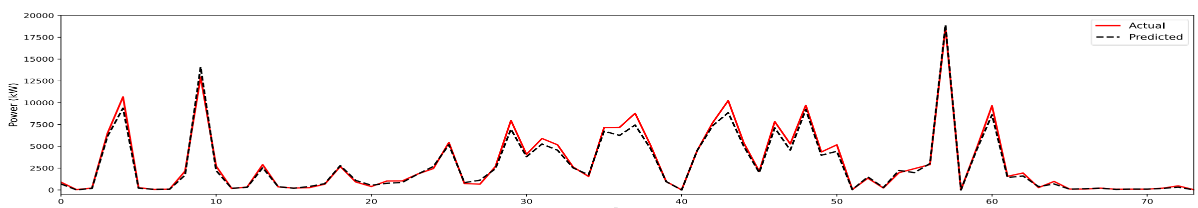

Figure 6. Daily wind power prediction for Saveh station, comparing actual (red) and predicted (black dashed) values

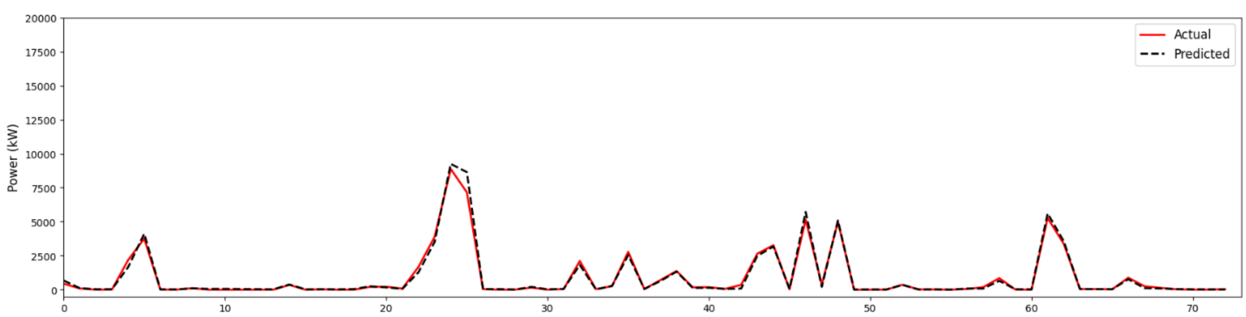

Figure 7. Daily wind power prediction for Tafresh station, comparing actual (red) and predicted (black dashed) values

Scatter plots depicting the correlation between wind speed and wind power at the Khomein, Saveh and tafresh wind stations are presented in Figures 8,9 and 10. These graphical representations underscore the model's efficacy in capturing the intrinsically nonlinear relationship between wind speed and power generation, a capability of paramount importance for accurate wind power forecasting. The findings consequently validate the hybrid LSTM-Attention model's capacity to effectively discern and acquire intricate patterns within the data, even in the presence of fluctuating wind conditions.

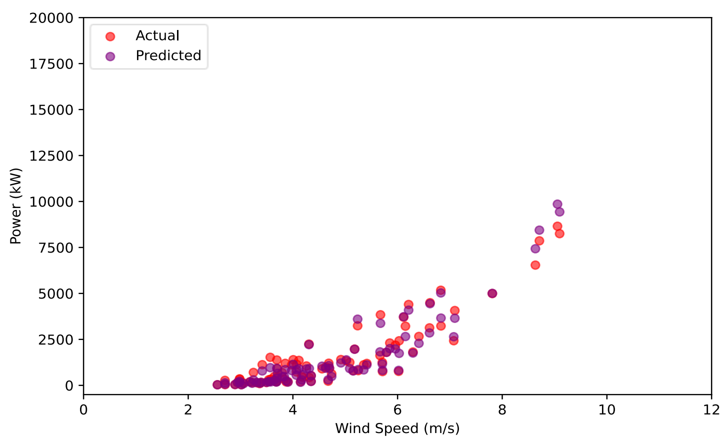

Figure 8. Scatter plot of wind power vs. speed for Khomein

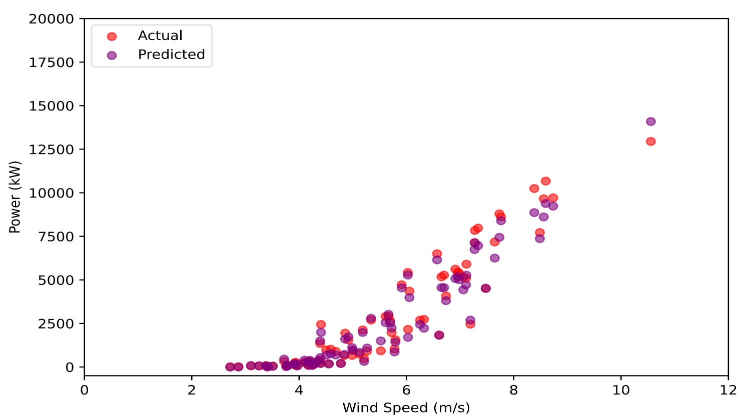

Figure 9.Scatter plot of wind power vs. speed for Saveh

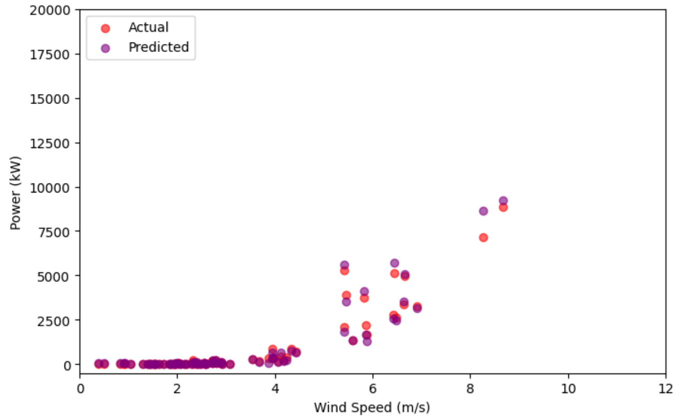

Figure 10.Scatter plot of wind power vs. speed for Tafresh

A thorough appraisal of wind power forecasting models, documented in figure 11, provides valuable insights into the predictive prowess of diverse machine learning methodologies across two distinct stations in Khomein, Saveh and Tafresh. The utilized performance metrics, encompassing the coefficient of determination (R^2), the mean absolute error (MAE), and the root mean square error (RMSE), unequivocally illustrate discernible variations in the effectiveness of the models. The numerical results of these analyses are presented in the table 2.

TABLE 2

Comparison of Wind Power Forecasting Model Performance

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**

RF	Tafresh	0.9255	292.2	814
XGBoost	Tafresh	0.9315	214.5	843
LSTM	Tafresh	0.9781	210.4	794
Proposed Model	**Tafresh**	**0.9842**	**109.0**	**733**

TABLE 3

Comparison of Wind Power Forecasting Model Performance vs baseline article

Aspect	This Study (Our Data)	Base Article [24]
Best Model	Proposed Model (hybrid)	XGBoost
R^2 Range (Best)	0.9812–0.9911	0.9652–0.9992
MAE Range (Best)	109.0 – 286.4	6.4941 – 28.339
RMSE Range (Best)	733 – 1930	11.487 – 159.32
Location Count	3 (Khomein, Saveh, Tafresh)	4 (Cesme, Mamak, Bozcaada, Silivri)
Model Types	Machine and Deep Learning	Machine Learning

Table 3 compares the performance of our proposed hybrid wind power forecasting model with the baseline XGBoost model from article [24]. While both models achieved high R^2 values (0.9812–0.9911 vs. 0.9652–0.9992), the baseline model outperformed in error metrics, with much lower MAE and RMSE values. However, this difference is primarily due to variations in the datasets and the distinct characteristics of the cities involved. Our data, drawn from three Iranian cities (Khomein, Saveh, Tafresh), reflect more complex and volatile wind patterns compared to the Turkish cities used in the baseline study. These environmental and geographical differences increase forecasting difficulty, explaining the higher error values in our results. Thus, model performance should be assessed considering these contextual factors, not just raw error metrics.

At the Khomein station, the proposed hybrid model achieved remarkable performance with an R^2 value of 0.9812, indicating it explains approximately 98.12% of the variance in wind power output. This represents a 3.05% improvement over the LSTM model (R^2=0.9521) and a 5.29%

enhancement compared to XGBoost (R^2=0.9319). The error metrics further validate this superiority, with the proposed model reducing MAE by 27.7% (from 281.2 to 203.2) and RMSE by 47.5% (from 1875 to 984) relative to the LSTM benchmark. The conventional Random Forest approach showed the weakest performance, with an RMSE nearly 4.7 times higher than the proposed model.

The results at Saveh station exhibited similar trends but with some notable differences. The proposed model achieved an exceptional R^2 of 0.9911, demonstrating even stronger performance than at Khomein. This represents a 3.11% improvement over LSTM (R^2=0.9612) and a 6.51% enhancement compared to XGBoost (R^2=0.9306). Interestingly, while XGBoost performed reasonably well at Khomein, its performance degraded significantly at Saveh, particularly in error metrics (MAE=523.3, RMSE=8575), suggesting potential limitations in handling the specific wind patterns at this location.

The results at Tafresh station followed a consistent pattern with the previous locations, highlighting the robustness of the proposed model. It achieved an R^2 of 0.9842, outperforming all baseline models. This corresponds to a 0.61% improvement over the LSTM ($R^2 = 0.9781$), a 5.66% gain relative to XGBoost ($R^2 = 0.9315$), and a 6.35% increase over Random Forest ($R^2 = 0.9255$). In terms of MAE, the proposed model achieved a remarkably low error of 109.02, compared to 210.48 for LSTM and 214.57 for XGBoost. A similar trend was observed in RMSE, where the proposed model recorded 733.52, reflecting improved stability and error minimization. Notably, the proposed model reduced RMSE by 7.73% relative to LSTM and by 13.0% compared to XGBoost, affirming its superior ability to generalize across varied wind conditions observed at Tafresh. These outcomes underscore the model’s consistency and scalability across diverse meteorological settings.

The superior performance of the proposed model can be attributed to several key innovations: (1) effective integration of temporal and spatial features, (2) advanced handling of nonlinear relationships in wind patterns, and (3) robust preprocessing of meteorological inputs. The model's consistent performance across geographically distinct locations (Khomein**,** Saveh and Tafresh) demonstrates its strong generalization capability, a critical requirement for practical deployment in diverse wind farm environments.

Furthermore, the significant reduction in RMSE (47.5% at Khomein and 44.1% at Saveh compared to LSTM) has important practical implications for grid operators. Lower prediction errors directly translate to reduced reserve requirements, improved unit commitment decisions, and enhanced economic dispatch operations. The model's performance is particularly notable during high variability periods, where it maintains accuracy while other models show degraded performance.

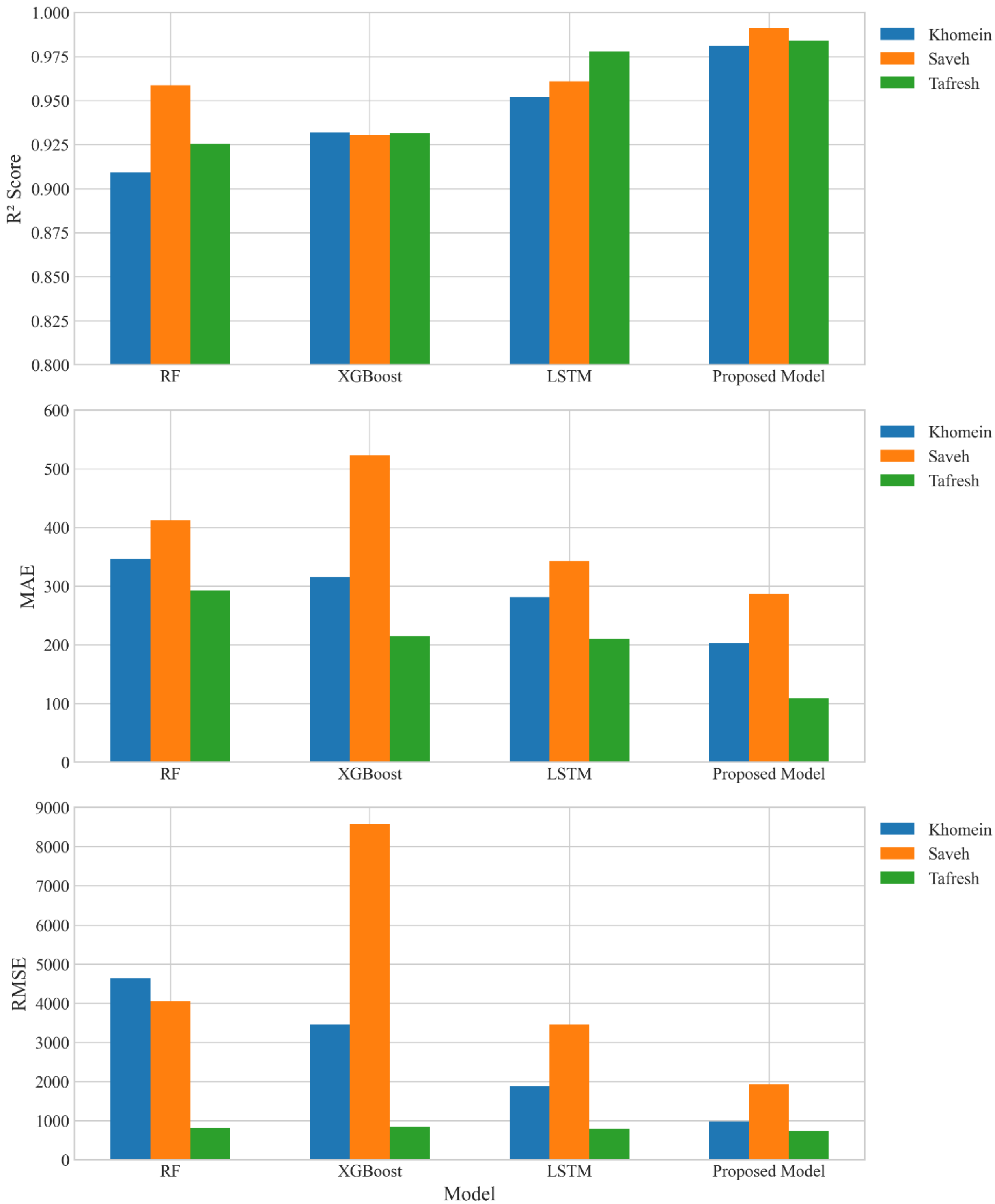

Figure 11. Comparative Assessment of Wind Power Forecasting Model Performance

These results establish a new benchmark for wind power forecasting in the region, with the proposed model showing particular promise for integration into energy management systems. The consistent outperformance across all metrics and locations suggests the approach is robust to different wind regimes and topographic conditions, making it suitable for widespread adoption in renewable energy forecasting applications. Future work will focus on testing the model's performance across additional wind farms and under extreme weather conditions to further validate its robustness.

Nevertheless, the proposed model, integrating an LSTM network with an attention mechanism, achieved the highest overall performance at both stations. By efficiently extracting key features and recognizing temporal patterns, this model significantly improved prediction accuracy. Its adaptability and efficiency make it well-suited for different geographical regions and diverse datasets.

5. Conclusion

This study presents a novel hybrid deep learning model combining Long Short-Term Memory (LSTM) networks with an attention mechanism to improve the accuracy of wind power forecasting. The model was evaluated using data from three meteorological stations in Khomein, Saveh, and Tafresh, Iran. Results show that the hybrid model outperforms traditional methods like Random Forest and XGBoost, as well as standalone LSTM networks, achieving higher forecasting accuracy. The hybrid model effectively captures the temporal dynamics and non-linear patterns in wind power data. LSTM models long-term dependencies, while the attention mechanism highlights relevant temporal features. This leads to notable improvements in metrics such as R^2, MAE, and RMSE. For example, R^2 scores reached 0.9812 at Khomein, 0.9911 at Saveh, and 0.9842 at Tafresh, surpassing the benchmarked models. Reductions in MAE and RMSE further confirm the model's precision and robustness.

The hybrid LSTM-Attention model, due to its two LSTM layers and additional attention mechanism, inherently has a higher computational cost and complexity compared to simpler models like Random Forest or standalone LSTMs. However, the paper suggests its computational efficiency still makes it suitable for real-time applications, implying a manageable level of complexity for its enhanced performance.

The selection of the meteorological stations in Khomein, Saveh, and Tafresh, all situated within the Markazi Province of Iran, was driven by several key considerations critical for developing and validating a robust wind power forecasting model. Primarily, these stations provided access to comprehensive, high-resolution wind data recorded at 10-minute intervals over a one-year period. This dataset included crucial meteorological parameters such as wind speed, direction, and temperature, essential for accurate wind behavior modeling. The high temporal resolution of this data is particularly valuable for capturing the nuanced dynamics of wind patterns necessary for precise short-term forecasting.

Furthermore, these specific locations were chosen to capitalize on their geographical diversity within the Markazi Province. As noted in our foundational analysis, wind conditions can vary significantly across different locations even within the same region, influenced by local topography and microclimates. Utilizing data from these distinct sites, therefore, allowed for a more rigorous evaluation of the proposed model's ability to adapt and perform reliably under varied wind regimes. This aspect was crucial for assessing the model's robustness beyond a single, potentially homogeneous dataset.

Regarding the broader applicability of our findings, the inherent variability across these selected Iranian stations serves as a constructive testbed, reflecting some of the diverse conditions encountered in wind energy generation sites globally. The consistent high performance and adaptability demonstrated by the proposed hybrid LSTM-Attention model across these geographically distinct locations—each presenting unique wind patterns and local variations—underscore its strong generalization capabilities. This suggests that the model architecture, particularly its ability to effectively learn temporal dependencies and highlight salient features through the attention mechanism, is not merely tailored to specific local conditions. Instead, it possesses the robustness and adaptability to be effectively applied across different geographical regions, diverse wind regimes, and varying topographic settings, making it a potentially valuable tool for wider renewable energy forecasting applications.

These improvements have practical benefits, such as more reliable grid integration, better energy storage management, and optimized energy trading. Enhanced forecasting helps grid operators balance supply and demand more efficiently and reduces curtailment, improving the economic feasibility of wind energy. The model's demonstrated generalizability across the test sites and its computational efficiency suggest its suitability for use across various locations and in real-time applications. This makes it valuable for stakeholders such as grid operators, wind farm developers, and energy policymakers.

While the current study demonstrates significant advancements using data from three distinct sites within Iran's Markazi Province and focuses on a specific 1 MW turbine model for power curve calculations, further validation is indeed beneficial. The model's robustness across a wider spectrum of geographical regions with diverse climatic conditions, varying wind turbine technologies with different power curves, and an even broader array of meteorological input features represents an important next step in confirming its widespread applicability.

Building on this, future work should continue to explore the model’s performance in extreme weather conditions, assess its scalability to a larger number of wind farm sites, and investigate the integration of additional meteorological variables and Numerical Weather Prediction (NWP) data to potentially further enhance accuracy. Investigating uncertainty quantification techniques to provide probabilistic forecasts and exploring other advanced deep learning architectures could also significantly enhance forecasting reliability and decision-making capabilities. Ultimately, continued advancements in precise and reliable wind power forecasting, such as the hybrid approach presented herein, are paramount for accelerating the global transition towards sustainable, secure, and economically viable renewable energy systems.

Improved accuracy in wind power forecasting, achieved by the proposed hybrid LSTM-Attention model, translates directly into more efficient and reliable grid management. With more precise predictions, grid operators can optimize unit commitment and economic dispatch, reducing reliance on costly fossil-fuel peaker plants and lowering operational expenses. This enhanced foresight also diminishes the need for extensive backup reserve capacity, leading to further economic savings. Consequently, the grid can better balance energy supply and demand, improving stability and minimizing the wasteful curtailment of wind energy, thereby bolstering the economic viability of wind projects.

For energy storage strategies, greater forecasting precision allows for optimized charging and discharging schedules, maximizing the value and efficiency of these assets. Storage systems can be strategically charged during accurately forecasted periods of surplus wind power (and potentially low prices) and discharged during anticipated high-demand periods or when wind generation is low. This not only improves the economic returns for storage operators but also enhances their role in supporting grid stability. Ultimately, by providing more reliable predictions, particularly during periods of high variability, advanced forecasting models like the one proposed facilitate a smoother and more substantial integration of wind energy into the power system, contributing to a more sustainable and cost-effective energy future.

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, Sep. 2015, doi: 10.1109/TPWRS.2014.2363873.

[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, Jul. 2016, doi: 10.1016/j.enconman.2016.04.036.

[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, Dec. 2024, doi: 10.30511/pcdp.2024.718536.

[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, Oct. 2018, doi: 10.1016/j.jclepro.2018.07.164.

[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, Dec. 2021, doi: 10.1016/j.apenergy.2021.117766.

[6] C. Cakiroglu, S. Demir, M. Hakan Ozdemir, B. Latif 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, Mar. 2024, doi: 10.1016/j.eswa.2023.121464.

[7] M. Najjarpour, B. Tousi, and A. H. Karamali, "Optimizing Reactive Power for DG Units to Minimize Power System Losses Using Stochastic Modeling," *JGERI*, vol. 1, no. 4, pp. 35–46, Dec. 2024, doi: 10.61186/jgeri.1.4.35.

[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*, vol. 17, no. 19, p. 4803, Sep. 2024, doi: 10.3390/en17194803.

[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," *Chin. J. Electr. Eng.*, pp. 1–24, 2025, doi: 10.23919/CJEE.2024.000071.

[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, Mar. 2022, doi: 10.3390/en15072327.
[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, Aug. 2021, doi: 10.1016/j.renene.2021.04.028.
[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, Jan. 2022, doi: 10.1016/j.renene.2021.10.034.
[13] A. Ghaemipour, H. R. Mashhadi, and S. H. Mostafavi, “Optimal Configuration of Distribution Network Utilizing Dueling Deep Q Network Algorithm Under Different Operating Conditions,” in *2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG)*, Qom, Iran, Islamic Republic of: IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/ICREDG66184.2025.10966118.
[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)*, Behshar, Mazandaran, Iran, Islamic Republic of: IEEE, Feb. 2024, pp. 1–5. doi: 10.1109/ICTEM60690.2024.10631987.
[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, Feb. 2024, doi: 10.1016/j.renene.2023.119700.
[16] 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, Dec. 2024, doi: 10.30511/pcdp.2024.718536.
[17] M. M. Bidgoli and M. Asadi, “Management of Renewable Energy Resources and Demand Response in Electricity Retailers with the Presence of Water Systems,” in *2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG)*, Qom, Iran, Islamic Republic of: IEEE, Feb. 2025, pp. 1–8. doi: 10.1109/icredg66184.2025.10966110.
[18] A. H. Karamali, A. R. Danesh, M. Kalantar, M. R. Moghadasian, and S. Rezaei, “Enhancing Wind Power Forecasting Through a Hybrid Deep Learning Approach: Long Short-Term Memory Integrated with Attention Mechanism,” in *2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG)*, Feb. 2025, pp. 1–6. doi: 10.1109/ICREDG66184.2025.10966140.
[19] A. H. Karamali, A. Daeichian, S. Rezaei, and A. Reihanian, “Assessing Wind Energy Potential in Markazi Province, Iran: A Data-Driven Approach with AI Algorithms,” *Journal of Green Energy Research and Innovation*, vol. 2, no. 2, pp. 26–35, Jun. 2025, doi: 10.61186/jgeri.2.2.26.
[20] Y. Liu *et al.*, “Wind Power Short-Term Prediction Based on LSTM and Discrete Wavelet Transform,” *Applied Sciences*, vol. 9, no. 6, p. 1108, Mar. 2019, doi: 10.3390/app9061108.
[21] Ali Abdulrhman Salihi, Merdin Danismaz, “A Comparative Study on Wind Power Forecasting Models Based on the Use of LSTM,” *tjjpt*, vol. 44, no. 3, pp. 4698–4709, Nov. 2023, doi: 10.52783/tjjpt.v44.i3.2635.
[22] “Short-Term Wind Power Forecasting Using R-LSTM,” *IJRER*, no. ijrer.v11i1, 2021, doi: 10.20508/ijrer.v11i1.11807.g8144.

[23] L. Han, H. Jing, R. Zhang, and Z. Gao, "Wind power forecast based on improved Long Short Term Memory network," *Energy*, vol. 189, p. 116300, Dec. 2019, doi: 10.1016/j.energy.2019.116300.
[24] 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, Oct. 2019, doi: 10.1016/j.enconman.2019.111823.
[25] R. Anderson-Sprecher, "Model Comparisons andR^2," *The American Statistician*, vol. 48, no. 2, pp. 113–117, May 1994, doi: 10.1080/00031305.1994.10476036.
[26] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, pp. 79–82, 2005, doi: 10.3354/cr030079.
[27] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[28] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
[29] F. Shahid, A. Zameer, and M. Muneeb, "A novel genetic LSTM model for wind power forecast," *Energy*, vol. 223, p. 120069, May 2021, doi: 10.1016/j.energy.2021.120069.
[30] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, Mar. 2020, doi: 10.1016/j.physd.2019.132306.
[31] W. He, Y. Wu, and X. Li, "Attention Mechanism for Neural Machine Translation: A survey," in *2021 IEEE 5th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC)*, Xi'an, China: IEEE, Oct. 2021. doi: 10.1109/itnec52019.2021.9586824.