

Article

T-LSTM: A Novel Model for High-Precision Wind Power Prediction by Integrating Transformer and Improved LSTM

Qin Zhong¹, Long Wang^{1,2,*} and Chao Huang¹ 

¹ School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China; chaohuang@ustb.edu.cn (C.H.)

² Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China

* Correspondence: lwang@ustb.edu.cn

Abstract

Wind energy is a core pillar of global green and sustainable energy transition. However, existing wind power prediction models face three key challenges: traditional long short-term memory (LSTM) models struggle to capture long-term temporal dependencies efficiently and have high training latency, while Transformer-based models exhibit excessive computational complexity and are prone to overfitting for short-term fluctuating data; meanwhile, few models integrate seasonal trend modeling with multi-scale temporal feature extraction, leading to large prediction errors in seasonal transitions. To address these issues, this paper proposes a hybrid prediction framework combining a novel T-LSTM recurrent unit with the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The T-LSTM unit fuses a simplified Transformer module and an improved LSTM structure. Thus, the design can synergistically capture both short-term fluctuations and long-term dependencies in wind power data. Complementarily, SARIMA is integrated via weighted fusion to model seasonal trends, addressing the neglect of seasonal characteristics in existing deep learning models. A diverse set of benchmark methods for wind power prediction are selected for comparison, including LSTM, convolutional neural network-gated recurrent unit (CNN-GRU), ns_Transformer, Autoformer, Reformer and least squares support vector machine (LSSVM), with experiments conducted across various prediction horizons. The results show that the proposed T-LSTM model outperformed most benchmark methods in key evaluation metrics across multiple prediction horizons and exhibited no statistically significant difference from Autoformer only in the 90 min horizon, validating its superiority in handling complex wind power time series.

Keywords: short-term wind power prediction; multi-head attention mechanism; seasonal characteristics; long short-term memory network (LSTM)



Academic Editor: Radu Godina

Received: 12 January 2026

Revised: 29 January 2026

Accepted: 2 February 2026

Published: 5 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Wind energy has become a critical component of global sustainable energy strategies due to its abundance and low environmental impact. According to the Global Wind Energy Council's (GWEC) "Global Wind Report 2025", global cumulative wind power installed capacity reached 1136 GW in 2024 (11% year-on-year growth), with a projected 8.8% compound annual growth rate from 2025 to 2030 [1]. However, wind power's inherent randomness and intermittency pose significant risks to grid stability—making high-precision prediction a prerequisite for efficient wind energy utilization.

Current wind power prediction methods fall into three categories: physical methods, data-drive methods, and hybrid methods. Physical methods rely on Numerical Weather Prediction and terrain data but suffer from poor portability and sensitivity to data noise.

Differently, data-driven methods, dominated by deep learning models, generate forecasting results based on learned patterns from datasets. Given the dynamic change characteristics of wind power time series, a large number of research results have emerged to effectively improve forecasting performance. Various successful prediction approaches have been reported, including Random Forest algorithm [2], wind power prediction model combining principal component analysis and BP neural network [3], long short-term memory network prediction model combining sliding window technology [4], prediction model using convolutional neural network for feature extraction [5], and wavelet neural network prediction models based on the principle of wavelet transform [6]. However, these prediction methods have inherent limitations in feature processing capabilities, such as the risk of losing key information and poor adaptability. To address these issues, Shi et al. [7] presented an enhanced approach for wind direction correlation by employing a wind direction enhancement algorithm to extract wind direction trend features. Li et al. [8] improved the Kernel Extreme Learning Machine by optimizing its kernel width and regularization coefficient using a difference algorithm. Meanwhile, given the problems of substantial volatility of wind power and high data dimensionality with redundancy, Wang et al. [9] introduced an innovative method that incorporates the Temporal Pattern Attention mechanism into the feature extraction process, enhancing the accuracy of wind power predictions. Chen et al. [10] utilized Principal Component Analysis to effectively reduce dimensionality in the application of generative adversarial networks to wind power forecasting, further refining the feature selection process. However, few existing data-driven improvements specifically target data quality problems, leaving the bottleneck of data sensitivity unresolved.

Hybrid methods, a key research direction in wind power prediction, aim to integrate the advantages of multiple models to compensate for the shortcomings of single models. Existing hybrid studies include those using enhanced variational mode decomposition combined with LSTM networks for subsequence prediction [11], and an ultra-short-term wind power prediction framework integrating LSTM with the SARIMA model [12]. However, most of these methods adopt simple module concatenation rather than dynamic fusion of seasonal and temporal features, and their core modules have strong dependencies, poor parameter sensitivity, and low fault tolerance. To address the aforementioned issues of module dependence and parameter optimization in traditional hybrid methods, integrating evolutionary computation with machine learning has become a mainstream research trend—this combination helps systematically identify optimal parameter configurations to enhance model performance. For instance, researchers have explored weighted support vector machines optimized via genetic algorithms [13], least squares support vector machines [14], an improved fruit fly optimization algorithm tailored for support-vector-machine-parameter tuning [15], a rich-poor optimization algorithm for fine-tuning outlier-robust extreme learning machine parameters [16], and a hybrid improved cuckoo search algorithm for optimizing support vector machine hyperparameters [17]. Despite progress in parameter optimization, these evolutionary computation-based hybrid methods still fail to resolve the core limitation of traditional hybrid approaches: weak adaptation between long-term dependencies and dynamic seasonal trends in wind power time series.

For error correction and non-stationarity handling—two key challenges closely related to hybrid method performance—existing studies have made preliminary attempts, but with notable drawbacks. In terms of error correction, Liang et al. [18] combined predictions from a support vector machine (SVM) and an Elman neural network, while

Shi et al. [19] employed a least squares support vector machine with a radial basis function for error correction. However, the accuracy of these error models highly depends on the quality and representativeness of training data, and they cannot compensate for the core defects of hybrid methods themselves. In dealing with time series non-stationarity, new frameworks such as “non-stationary Transformers” [20]—wind power prediction models enhanced by the improved African Vulture Optimization Algorithm [21]—and the “swinLST” recurrent unit [22] have been proposed. Nevertheless, these techniques either require excessive computational resources, are sensitive to algorithm parameter selection, or tend to misinterpret noise as valid spatial correlations, and none address the root issue of poor dynamic fusion of seasonal and temporal features.

In order to provide a clear overview of existing solutions and highlight the research gaps addressed by this work, Table 1 summarizes representative wind power forecasting methods, their core features, advantages, and limitations. This comparison emphasizes the trade-off between prediction accuracy, computational efficiency, and adaptability to complex time/seasonal features.

Table 1. Comparison of Representative Wind Power Prediction Methods.

Method Category	Specific Method	Core Idea	Limitations
Physical Methods	WRF-Fitch Parameterization Modified Model [23]	Improves Fitch wind farm parameterization in WRF to enhance prediction accuracy for high-altitude wind farms	High computational cost
	CFD-Micrometeorology Coupled Physics-Driven Model [24]	Couples CFD with micrometeorology to simulate complex terrain wind fields for ultra-short-term prediction without historical data	Relies on high-precision terrain data, complex parameterization
Data-Driven Methods	LSSVM-AVOA [21]	Optimize LSSVM hyperparameters via African Vulture Optimization Algorithm	Fails to capture long-term dependencies, poor adaptability to volatility
	BiLSTM-Temporal Attention [9]	Incorporate temporal pattern attention into BiLSTM for feature enhancement	High training latency, redundant parameters
	CNN-GRU Multichannel [25]	Combine CNN feature extraction with GRU sequential modeling	Weak at long-term prediction, ignores seasonal trends
	ns_Transformer [20]	Address non-stationarity via adaptive normalization in Transformer	Excessive computational complexity, overfitting to small samples
	Stacked Denoising Autoencoder-Support Vector Regression-Bat Algorithm hybrid model [26]	Proposes a three-stage hybrid model combining SDAE, BA-tuned SVR, and regression for more accurate short-term wind power forecasts.	Dependent on high-quality historical data, Fails to capture long-term temporal dependencies
	SwinLSTM [22]	Integrate Swin Transformer spatial partitioning with LSTM	Complex architecture, high memory consumption

Table 1. Cont.

Method Category	Specific Method	Core Idea	Limitations
Hybrid Methods	SARIMA-LSTM [12]	Concatenate SARIMA seasonal output with LSTM predictions	Simple concatenation, no dynamic feature fusion
	VMD-LSTM [11]	Decompose data via Variational Mode Decomposition before LSTM training	Complex preprocessing, poor generalization to seasonal transitions
	Transformer-LSTM Hybrid [27]	Fuses Transformer-LSTM for long/short-term modeling to improve short-term wind power prediction accuracy.	High computational cost, Redundant feature overlap between LSTM and Transformer
	Two-stage hybrid feature-learning model [28]	Proposes a two-stage hybrid model that uses CNN/LSTM for feature learning and ridge regression for result integration.	Slow inference, ignores fine-grained temporal correlations

In summary, existing wind power prediction methods face three critical and interrelated gaps that restrict their accuracy and reliability: (1) data-driven methods are sensitive to data quality and prone to key feature loss; (2) hybrid methods lack dynamic fusion of seasonal and temporal features, with strong module dependence, poor parameter sensitivity, and failure to effectively resolve the weak adaptation between long-term dependencies and dynamic seasonal trends; and (3) error correction and non-stationarity handling are either inefficient or disconnected from core hybrid method defects. To address the above gaps comprehensively, this paper proposes a T-LSTM-based hybrid framework with four key innovations:

- (1) We adopt the advanced SARIMA method to analyze seasonal trends of wind power data, with the model's ability to integrate seasonal/external factors and parameter sensitivity enhanced.
- (2) The innovative T-LSTM recurrent unit is proposed, which efficiently extracts core feature information from time series data to underpin reliable wind power prediction.
- (3) A wind power prediction-specific architecture is designed to accurately capture and replicate temporal relationships, strengthening the model's adaptability to complex trends for higher forecast precision.
- (4) Comparative analysis with prevalent methods is conducted, with core mechanisms and correction strategies optimized simultaneously to verify the proposed approach's efficacy.

The subsequent sections of this paper are structured in the following manner: Section 2 provides an overview of previous research in related areas; Section 3 elaborates on the architectural design of the new model being proposed; the prediction results of the new model in comparison to other standard models are discussed in Section 4; the paper is concluded in Section 5, with a summary of the key findings from the research.

2. Related Work

The prediction task in this article falls under the category of Short-Term Wind Power Prediction (STWPP). Specifically, we focus on single-point rolling prediction of specific onshore wind farms located in northwest China, with the target variable being the total hourly average wind output (kW) of the entire wind farm at future time steps.

The input data for the predictive model comes from the monitoring and data acquisition (SCADA) system of the wind farm, covering the entire year from 1 January 2019 to

31 December 2019 (35,040 data points). Following the widely adopted sequence to point prediction paradigm, the input features include two types of information: (1) historical wind power data of the target wind farm; and (2) multi-source meteorological covariates synchronized with historical power data, including wind speed (m/s), wind direction ($^{\circ}$), air pressure (kPa), ambient temperature ($^{\circ}\text{C}$), and relative humidity (%). The prediction range is set to 15, 30, 60, and 90 min in advance, corresponding to predicting the wind power generation for the next 1-, 2-, 4-, and 6-time intervals, respectively.

2.1. Transformer

In the context of wind power forecasting, the ability of the Transformer architecture to process data in parallel is of particular significance. Huang et al. [29] conducted in-depth experiments to adapt the Transformer model for wind power forecasting, but its over 100 million parameters increased training costs. Similarly, Chen et al. [30] investigated different variations of the Transformer model and optimized it for wind power forecasting tasks. Their research demonstrated that with proper adjustments, the Transformer model can be tailored to specific wind power forecasting scenarios, leading to enhanced performance and more reliable predictions. However, during its optimization process, it overly focuses on local features within the scene and does not establish a cross-scene feature transfer mechanism, which increases the operational costs in practical applications.

To solve Transformer's excessive parameters and lack of cross-scenario adaptability, the proposed framework combines the optimized LSTM and multi-head attention and integrates SARIMA's seasonal pattern extraction advantage. This design avoids the Transformer's high parameter cost, enhances cross-scenario feature reuse through multi-feature fusion, and balances prediction accuracy, training efficiency, and practical operational feasibility.

2.2. LSTM

The LSTM network has emerged as a significant advancement in the field of neural networks, particularly in the context of processing sequential data with complex temporal patterns. At the core of the LSTM architecture are specialized LSTM units, which are designed to selectively remember or forget information over long sequences. These units consist of input gates, output gates, and forget gates, which work together to control the flow of information into and out of the memory cells, but the three gates of traditional LSTM increase training latency. Despite its innovative design and potential benefits, the LSTM model has several limitations in practical use. One of the primary challenges is the time-consuming training process. Due to the complexity of the LSTM architecture and the large number of parameters involved, training an LSTM model can be computationally expensive and requires significant amounts of data and computational resources. This can make it difficult to implement the model in real-time applications where quick and accurate predictions are needed.

For LSTM's high training latency and computational complexity, the LSTM structure of the proposed method is streamlined by optimizing the gate mechanism and reducing redundant parameters while retaining long-term dependency capture capabilities, thus lowering training and operational costs.

2.3. SARIMA

The SARIMA model is a classic statistical method widely used in wind power forecasting, specifically designed to capture seasonal trends and linear time patterns in time series data. Its advantages lie in clear interpretability, mature theoretical support, and good performance in short-term forecasting of stable seasonal sequences. However, SARIMA has significant limitations in wind power prediction scenarios. Firstly, it relies on strict

assumptions of linearity and stationarity, which are difficult for wind power data to meet—wind power output has high volatility and nonlinearity, and is influenced by the coupling of multiple meteorological factors. Secondly, as a univariate model, it cannot effectively integrate and analyze multidimensional feature interactions and cannot capture the complex nonlinear relationship between meteorological conditions and wind power generation.

To compensate for SARIMA's inability to model nonlinear relationships and multi-feature interactions, the proposed framework retains SARIMA's advantage in capturing linear seasonal trends. By integrating linear seasonal features and nonlinear complex features through weighted fusion, the resistance to data noise and adaptability to sudden meteorological changes have been improved.

3. The Proposed Approach

3.1. Overall Architecture

As shown in Figure 1, this prediction architecture combines two components through a weighted method to generate accurate wind energy predictions, aiming to improve the prediction reliability of renewable energy applications. To address the issue of weighted fusion strategy between two models, we first established the mathematical specifications for this integration.

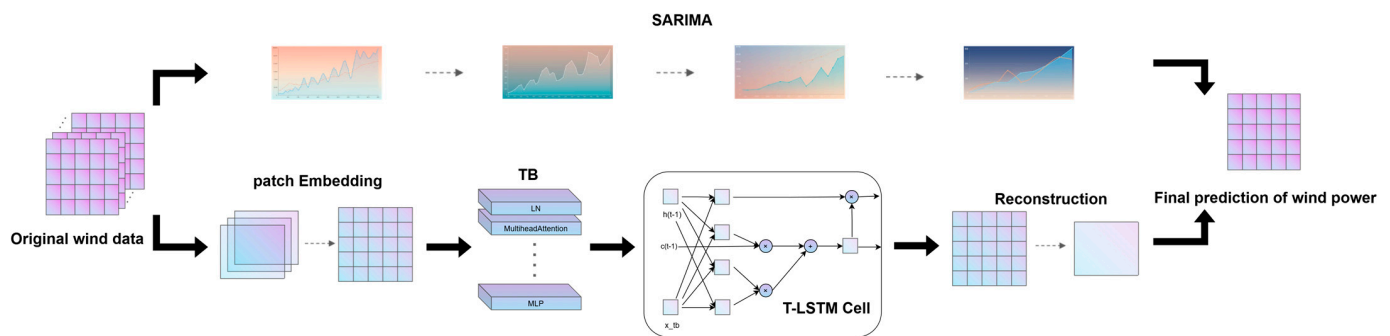


Figure 1. Overall process of the wind power prediction method.

Specifically, the prediction results of the LSTM network are corrected by the results of the SARIMA model. The final wind power generation prediction value is shown in Equation (1):

$$\tilde{p}_t = \omega p_t + (1 - \omega) h_t (0 \leq \omega \leq 1) \quad (1)$$

Among them, p_t represents the prediction result of the SARIMA model, and h_t represents the prediction output of the proposed t-LSTM model. ω is a hyperparameter used to quantify the contribution of the SARIMA model to the final prediction, optimizing the final prediction by minimizing the error between the fused prediction value and the actual wind power value. The final value selected in this paper is 0.14.

The T-LSTM model operates by first processing the data point $x(t)$ at a specific time step t through an input embedding layer, which projects the input data into a hidden dimensional space. Next, the data enters the simplified Transformer module, where long-term dependencies are captured and processed to produce data x_{tb} . Following this, the T-LSTM unit receives the transformed data block, hidden state $h(t-1)$, and memory cell state $c(t-1)$ from the previous time step. Through computational manipulation, the T-LSTM unit generates the current time step's hidden state $h(t)$ and memory cell state $c(t)$. The hidden state $h(t)$ is duplicated for the reconstruction layer operation, while the other instance, along with memory cell state $c(t)$ at the current time step, becomes input data for the T-LSTM unit in the next time step. Lastly, the reconstruction layer maps the hidden state $h(t)$ back to the input data's dimensional magnitude and predicts the wind power for the

subsequent time step. The T-LSTM model effectively processes data, capturing long-term dependencies to make accurate predictions. This methodology allows for efficient handling of sequential data, ensuring the model's ability to forecast future outcomes with precision. By utilizing a combination of Transformer modules and LSTM units, the T-LSTM model proves to be a robust tool for time series forecasting tasks.

3.2. Seasonal Auxiliary Prediction Based on the SARIMA Model

The SARIMA model is an advanced version of the Autoregressive Integrated Moving Average (ARIMA) model, specifically designed to address the periodic characteristics seen in time series data. In practical applications, the T-LSTM network can be used in conjunction with SARIMA to refine predicted values and adjust wind power estimates. Traditional ARIMA models struggle to accurately capture both seasonal and non-seasonal components of wind power data, leading to potential errors in parameter selection. The SARIMA model is represented by Equation (2), encompassing the seasonal patterns within the data. By leveraging SARIMA's capabilities, analysts can better model and predict fluctuations in wind power generation, enhancing the decision-making processes in the renewable energy sector.

$$\phi(B)(1-B)^d\Phi(B)(1-B^s)^Dx_t = \theta(B)\Theta(B)\varepsilon_t \quad (2)$$

where the wind power series is denoted as x_t and ε_t stands for white noise. B references the lag operator, d signifies non-seasonal differencing to remove data non-stationarity, and D denotes seasonal differencing for tackling seasonal data patterns. p and q are the autoregressive and moving average terms, respectively. Lastly, s captures the seasonal order, reflecting the periodicity of seasonal variations in wind power data.

Equations (3) and (4) reveal the autoregressive and moving average polynomials, showcasing the interconnectedness between future time series values, past values, and errors. These equations provide insight into the intricate dependencies within a dataset, aiding in forecasting and analysis.

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (4)$$

Equations (5) and (6) describe the seasonal patterns within a time series using autoregressive and moving average polynomials. By integrating these polynomials into the ARIMA equation, the model becomes better equipped to identify and predict seasonal fluctuations. This enhanced capability allows for a more accurate analysis of time series data.

$$\Phi(B) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \quad (5)$$

$$\Theta(B) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs} \quad (6)$$

3.3. Transformer Block

The Transformer block (TB) serves as the core feature extraction unit of the proposed wind power prediction model, tasked with converting low-dimensional features from the input embedding layer into high-dimensional representations that capture multi-scale temporal correlations and key meteorological features (see Figure 2 for its structure and operation process).

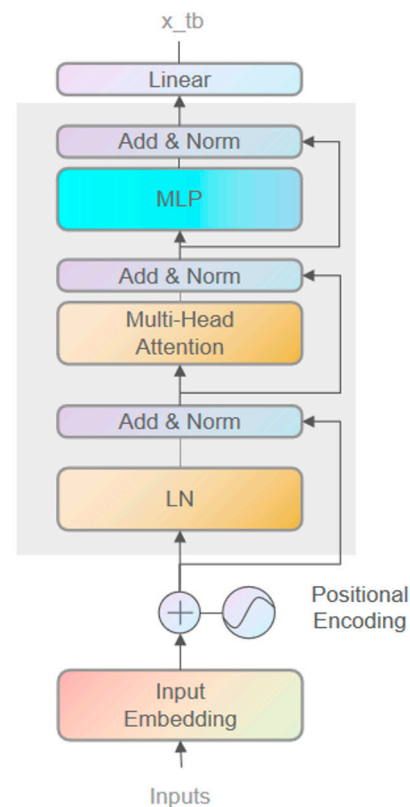


Figure 2. The overall architecture of Transformer Block.

The input to the TB is the wind power feature vector (integrating multi-source features like wind speed, wind direction, and temperature) processed by the embedding layer, which undergoes two-step preprocessing to adapt to the “dynamic fluctuation + periodicity” characteristics of wind power time series data. Positional encoding is first applied to inject temporal information—sine cosine encoding is adopted instead of learnable positional encoding to reduce model parameters and mitigate overfitting in small-sample wind power scenarios; specific encoding formulas are in Equations (7) and (8).

$$PE_{pos,2i} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (7)$$

$$PE_{pos,2i+1} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (8)$$

Layer Normalization is then performed to unify feature scales, preventing subsequent attention layers from biasing towards large-value features and safeguarding the capture of key small-scale features.

The preprocessed features are fed into the Multi-Head Attention layer, the core of capturing complex correlations in wind power data. Input vectors are split into multiple attention heads for parallel calculation of dimension-specific attention weights, enabling multi-perspective feature extraction and avoiding one-sided correlation capture. Each head assigns dynamic weights based on query-key similarity, emphasizing critical predictive information and suppressing redundancy, with outputs concatenated into a unified feature vector to complete dynamic feature enhancement.

A “residual connection + normalization” structure follows to address gradient vanishing in deep network training and stabilize data distribution. Residual connections allow shallow feature information to propagate directly to subsequent layers, which is essential for stacking multiple TB blocks, while a second Layer Normalization eliminates

feature-value fluctuations after residual connections, ensuring stable data distribution for subsequent layers.

The final component is the Multi-Layer Perceptron (MLP), which transforms correlation features from the attention layer into temporal dependent features to meet wind power prediction requirements for long- and short-term temporal patterns. Through a two-layer fully connected structure with ReLU activation, the MLP integrates local correlations into global temporal dependencies and outputs the transformed feature block x_{tb} .

3.4. T-LSTM Cell

The innovative recurrent cell structure introduced in this research is a refined version of the LSTM network, demonstrated in Figure 3. To standardize notation and clarify variable connotations, key variables are first defined: x_{tb} is the output result processed by the Transformer block; h_{prev} (i.e., $h(t-1)$) stands for the hidden layer state at the previous time step; h_t (i.e., $h(t)$) is the hidden layer state at the current time step; c_{prev} (i.e., $c(t-1)$) refers to the memory unit at the previous time step; and c_t (i.e., $c(t)$) represents the memory unit at the current time step. This structure excels at capturing both short-term and long-term relationships within time series data by updating the cell state $c(t)$ and the hidden state $h(t)$ horizontally.

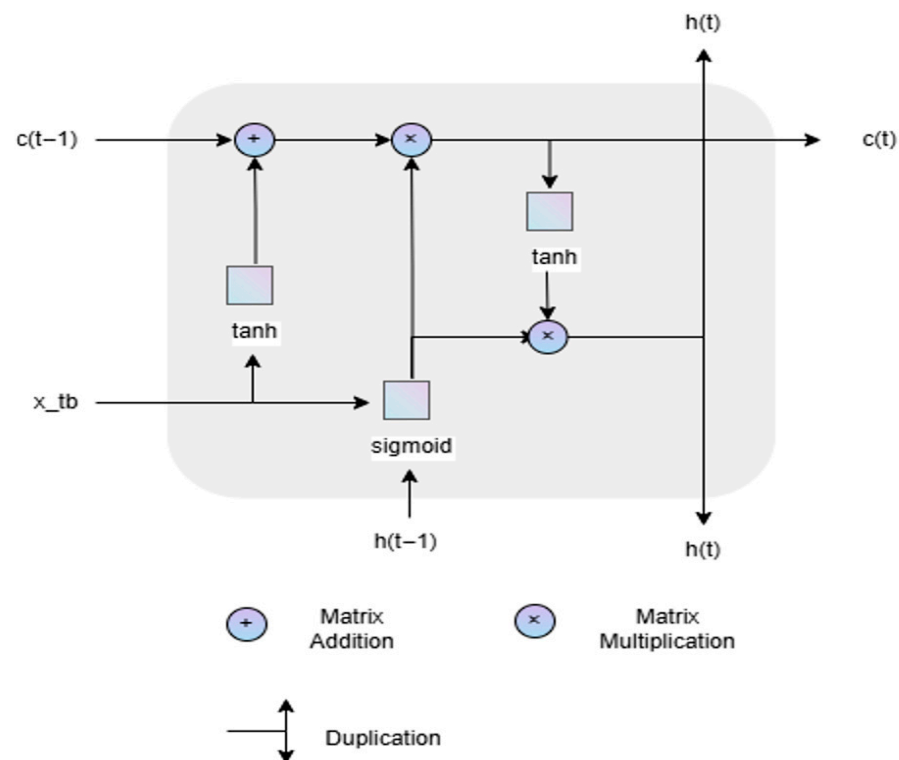


Figure 3. The overall architecture of T-LSTM Cell.

In order to address the limitations of traditional LSTM while maintaining its advantages in sequential data modeling, the T-LSTM unit adopts a “single gate dual control” filtering gate which simultaneously adjusts the “fusion of historical memory and new features” and the “output of unit state to hidden state”. The theoretical and empirical basis for this specific design is as follows:

The theoretical basis for simplifying target gates: Traditional LSTM relies on three separate gates to control “new information input,” “historical information retention,” and “unit state output,” but these functions exhibit strong correlation in wind power time series. The filtering gate of T-LSTM integrates these correlation functions into a control coefficient $f(t)$.

Synergistic effect with transformer function: The input of the filter gate directly contains $\mathbf{x}_{tb}$, ensuring that the control coefficient $\mathbf{f}(t)$ is suitable for the fine-grained temporal correlation captured by the Transformer. This design avoids the traditional LSTM's "equal treatment of raw inputs" and considers the high-value features of wind power prediction.

To begin, the computation process kicks off with the calculation of the gate value, followed by the incorporation of the crucial information output $\mathbf{x}_{tb}$ modulated by attention weights into the hidden state $\mathbf{h}(t - 1)$ from the preceding time step. Layer Normalization is then applied, and the Sigmoid activation function is used to determine the gate value $\mathbf{f}(t)$. The memory cell is updated by processing the current input $\mathbf{x}_{tb}$ and the memory cell from the previous time step, multiplying the outcome with the gate value $\mathbf{f}(t)$. The current hidden state is generated by element-wise multiplication between the cell state $\mathbf{c}(t)$ and the gate value $\mathbf{f}(t)$. The fundamental equations for this innovative T-LSTM model are outlined in Equations (9)–(11), showcasing the effectiveness of this optimized cell structure in capturing complex dependencies within time series data.

$$\mathbf{f}_t = \text{Filter_gate}(\mathbf{x}_{tb} + \mathbf{h}_{prev}) \quad (9)$$

$$\mathbf{c}_t = \mathbf{f}_t * (\tanh(\mathbf{x}_{tb}) + \mathbf{c}_{prev}) \quad (10)$$

$$\mathbf{h}_t = \mathbf{f}_t * \tanh(\mathbf{c}_t) \quad (11)$$

Filter_gate is a crucial element in the operation of the TB, receiving inputs $\mathbf{x}_{tb}$, $\mathbf{h}_{prev}$, and $\mathbf{c}_{prev}$ to determine the flow of information within the network. It plays a vital role in controlling the information flow and interactions between the hidden state and memory cell of the previous time step.

4. Numerical Experiment

In this section, we select a wind farm SCADA comprehensive dataset that accurately portrays real-life operating conditions. This dataset encompasses diverse monitoring data across different climates and timeframes, including meteorological parameters and historical wind power output. By utilizing this dataset, we meticulously compare our proposed model against several benchmark methods. Through these approaches, we strive to thoroughly and accurately assess the practical effectiveness of our model in wind power prediction. This methodology enables us to clearly identify the capabilities and limitations of our model in real-world wind farm scenarios, aiming to provide a detailed and reliable evaluation of its practical applicability.

4.1. Data Description and Experiment Setup

Dataset: This study employs a wind farm SCADA comprehensive dataset collected from a wind farm in northwestern China. The dataset, containing information sampled at 15 min intervals, covers an entire year from 1 January 2019 to 31 December 2019 comprising a total of 35,040 data points. By utilizing the monitoring and data acquisition (SCADA) system of wind farms, various real-time meteorological parameters can be captured, including wind direction ($^{\circ}$), wind speed (m/s), historical wind power (kW), air pressure (kPa), temperature ($^{\circ}\text{C}$), and relative humidity (%).

To verify the cross-seasonal generalization ability of the model, the entire dataset was divided into four seasonal subsets based on the meteorological seasons in the Northern Hemisphere: spring (8832 data points from March to May), summer (8832 data points from June to August), autumn (8736 data points from September to November), and winter (8640 data points from December to February).

During the development of our model for wind power prediction, we divided our dataset into training and testing subsets chronologically. The training subset comprises 90% of the data and is utilized to train the model, while the testing subset, accounting for the remaining 10%, is used to evaluate the model's performance. The experiment adopts single-point prediction, specifically using a rolling prediction strategy to predict the wind power generation for the next 1-, 2-, 4-, and 6-time intervals based on the data from the first 24-time intervals. For cross-seasonal validation, the same training–test split ratio (9:1), rolling prediction strategy, and hyperparameters are applied to each seasonal subset to ensure consistency and avoid overfitting to specific seasonal patterns.

Transformer block and T-LSTM unit parameter selection: two attention heads are used, with head 1 focusing on hourly scale dependence and head 2 focusing on daily scale dependence. This design aims to address the core challenge of wind power forecasting—capturing short-term dynamic fluctuations and long-term seasonal cyclicity—regardless of the forecast range. To validate this design, we conducted quantitative ablation experiments on a 90 min predicted scenario (the longest time range in this study). This situation is the most challenging for multi-scale feature capture: it not only requires capturing short-term temporal correlations but also integrating long-term periodic patterns to avoid cumulative prediction errors. The experimental results show that the single-head configuration achieved a MAE of 10.51 and an RMSE of 14.76; the four-head configuration achieved a MAE of 8.78 and an RMSE of 13.31; the dual-head configuration achieved a balance between accuracy and efficiency, with a MAE of 7.67 and an RMSE of 12.99, which is superior to single-head and four-head configurations.

The hidden layer dimension is 64. Sixty-four dimensions are sufficient to carry key information of time series data, such as non-linear mapping between wind speed and wind power output, seasonal trends at the hourly/daily/weekly level, and cross correlation of multi-source features. Overly high dimensions introduce “redundant feature expression”, which increases the learning burden of the model. Moreover, the feature correlation of time series data is relatively concentrated. Sixty-four dimensions can already cover more than 90% of effective information. And the batch size is 32.

To enhance the reproducibility of the proposed T-LSTM model, detailed training configurations are supplemented in Table 2:

Table 2. Detailed training configurations of the T-LSTM model.

Configuration Item	Details
Optimizer	Adam optimizer, leveraging its adaptive learning rate mechanism to balance convergence speed and stability in deep learning tasks.
Initial Learning Rate	1×10^{-3} , a widely accepted choice for time series forecasting models to avoid premature convergence or divergent training.
Training Epochs	50 epochs. Preliminary experiments confirm that training loss converges stably within 50 epochs, with no significant loss reduction observed beyond this number.
Early Stopping Strategy	Patience of 5 epochs. If the validation loss does not decrease for 5 consecutive epochs, training is terminated early, and parameters corresponding to the minimum validation loss are retained.
Hardware & Software Environment	Hardware: Workstation with NVIDIA GeForce RTX 3090 GPU (24 GB VRAM) and Intel Core i9-12900 K CPU (3.2 GHz). Software: Python 3.9, PyTorch 1.13.1, CUDA 11.7, scikit-learn 1.2.2.

SARIMA parameter selection: The study not only introduced the primary model but also developed a SARIMA model using wind power time series data. To assess and compare different models, the Akaike information criterion (AIC) was used as the evaluation metric. The AIC considers both the model's fit with the data and its complexity, aiming to strike a

balance between the two. Lower AIC values indicate models that are better suited for wind power data modeling. SARIMA (3, 0, 1, 3, 1, 1, 12) was selected via two steps:

1. Grid search over $p, q \in [0, 3]$, $P, Q \in [0, 3]$;
2. By minimizing the AIC value, which is detailed in Table 3, and validating with seasonal error.

Table 3. AIC of different model parameters.

Parameters (p, d, q, P, D, Q, s)	AIC
(0, 0, 1, 2, 0, 0, 12)	339,984.7605
(0, 0, 0, 3, 0, 0, 12)	372,724.3774
(1, 1, 0, 2, 0, 0, 12)	282,519.3464
(1, 1, 1, 1, 1, 1, 12)	282,490.3239
(2, 0, 1, 1, 1, 0, 12)	295,484.4330
(3, 0, 1, 2, 1, 0, 12)	291,487.7277
(3, 0, 1, 3, 1, 1, 12)	282,249.2177

4.2. Performance Metrics and Benchmark Models

A comprehensive evaluation of a wind power prediction method was conducted, comparing it with six other commonly used methods. These methods included long short-term memory (LSTM) [31], convolutional neural network-gated recurrent unit (CNN-GRU) [25], non-stationary Transformer (ns_Transformer) [20], Autoformer [32], Reformer [33], and least squares support vector machine with optimization using a specialized algorithm [21]. These methods were selected for their representativeness in different model categories. As classic deep learning models, GRU outperforms LSTM in computational efficiency with its streamlined update gates while maintaining high accuracy. For Transformer variants, ns_Transformer addresses time series non-stationarity, a key wind power prediction challenge, showing superior performance. Autoformer's decomposed design combines autocorrelation mechanisms to extract temporal patterns efficiently without heavy computation. Reformer, a benchmark for efficiency–performance trade-off, uses LSH attention and reversible layers—comparison with it highlights the proposed method's innovation. The LSSVM, optimized via the African Vulture Optimization Algorithm, represents traditional machine learning. This comparative setup reveals each method's unique strengths, emphasizing the need for tailored selection based on practical requirements. Through this comparative analysis, it was evident that each method presents unique advantages in predicting wind power output, highlighting the importance of choosing the most suitable approach based on specific requirements.

In order to thoroughly assess the effectiveness of various prediction methods, we have chosen to focus on four key evaluation metrics, mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2). These metrics serve as reliable indicators of predictive performance by quantifying the difference between predicted and actual values. Lower values of MAE, MSE and RMSE indicate a higher level of accuracy in predictions, while higher values of R^2 suggest a stronger correlation between predicted and actual values, resulting in more favorable prediction outcomes. Formulas (12)–(15) outline the specifics for calculating these indicators. By conducting a detailed comparison using these metrics, we can effectively demonstrate the accuracy and dependability of each model when applied to tasks involving the prediction of wind power.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (15)$$

At the same time, to verify the statistical reliability of the performance difference between T-LSTM and baseline models, the normality of the sample-by-sample absolute error (AE) difference was first verified through the Shapiro–Wilk test. The results showed that the AE differences between all baseline models and T-LSTM did not follow a normal distribution ($p < 0.0001$). Therefore, this study used the Wilcoxon signed-rank test for paired comparisons and Bonferroni correction (comparison frequency $k = 6$, corrected significance level $\alpha = 0.05$) to control for type I errors caused by multiple comparisons.

4.3. Ablation Experiment Design and Results

To quantitatively isolate the contributions of three core components of the proposed framework—(i) Transformer block (TB), (ii) improved LSTM gate, and (iii) SARIMA seasonal fusion—we designed three ablation variants based on the full T-LSTM-SARIMA model:

Ablation 1 (T-LSTM w/o TB): Remove the Transformer block; the model uses only the improved LSTM (with Filter Gate) and SARIMA fusion. Input data is directly fed into the T-LSTM cell without multi-scale temporal feature extraction.

Ablation 2 (LSTM-TB-SARIMA): Replace the improved LSTM gate with the traditional LSTM three-gate mechanism; retain the Transformer block and SARIMA fusion.

Ablation 3 (T-LSTM w/o SARIMA): Remove the SARIMA model; the final prediction relies solely on the T-LSTM unit without seasonal trend correction.

All ablation experiments use the same dataset, hyperparameters (except for necessary adjustments to the traditional LSTM gate), and evaluation metrics as the full model. The results are shown in Table 4. The values in bold in the table represent the optimal performance indicator values.

Table 4. Ablation experiment results (MAE/RMSE/ R^2) across different forecasting horizons.

Model	15 min	30 min	60 min	90 min
Full T-LSTM-SARIMA	7.4669/12.0630/0.9684	6.6317/11.9038/0.9692	6.1264/11.5302/0.9711	7.6671/12.9935/0.9633
Ablation 1 (w/o TB)	8.9215/13.8742/0.9572	8.1543/13.5267/0.9589	7.8932/13.1045/0.9598	9.2417/14.8723/0.9486
Ablation 2 (LSTM gate)	8.3572/13.2419/0.9615	7.7201/12.8963/0.9621	7.3458/12.5681/0.9634	8.8765/14.2351/0.9521
Ablation 3 (w/o SARIMA)	7.8936/12.6428/0.9647	7.2154/12.4879/0.9643	6.8729/12.1053/0.9668	8.3429/13.7682/0.9567

Compared with the complete model, the MAE of Ablation 1 (without TB) increased from 19.5% (15 min) to 28.8% (60 min). The most significant performance decline occurred within 60 min, indicating that the TB's multi-head attention mechanism is crucial for medium- and long-term prediction. Without the TB, the model cannot extract fine-grained temporal correlations, resulting in cumulative errors as the prediction range expands. The MAE of Ablation 2 (traditional LSTM gate) is 11.9% (15 min) to 19.9% (60 min) higher than that of the complete model. This verifies that the filter gate retains the ability to capture sequential dependencies. Ablation 3 (excluding SARIMA) showed a 5.7% (15 min) to 8.8% (90 min) increase in MAE, with the greatest degradation occurring within 90 min. This indicates that SARIMA effectively addresses the weakness of T-LSTM units in modeling linear seasonal trends, and the weighted fusion of SARIMA seasonal features reduces prediction bias.

4.4. Comparison with Benchmark Models

In order to improve the readability of performance indicators and avoid redundant numerical representations, we first use a line chart (Figures 4–6) to visualize the core evaluation indicators (MAE, RMSE, R^2) within different prediction ranges, and provide corresponding tables for quantitative reference.

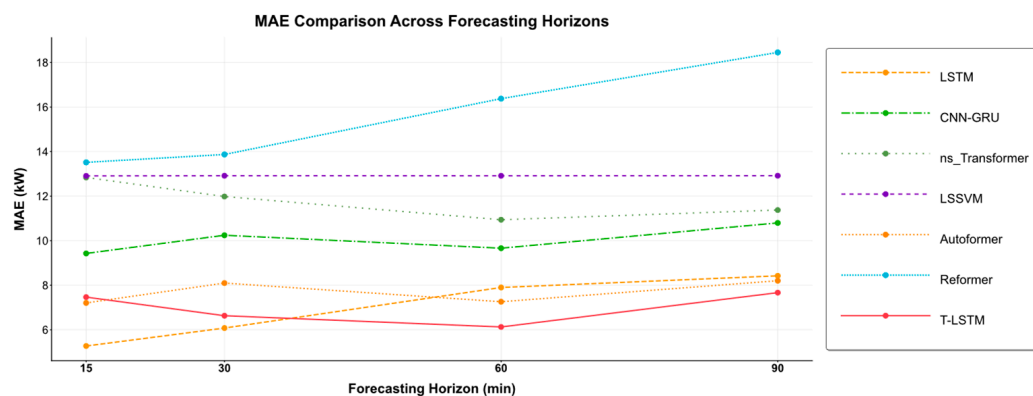


Figure 4. The MAE of all models across 15, 30, 60, and 90 min horizons.

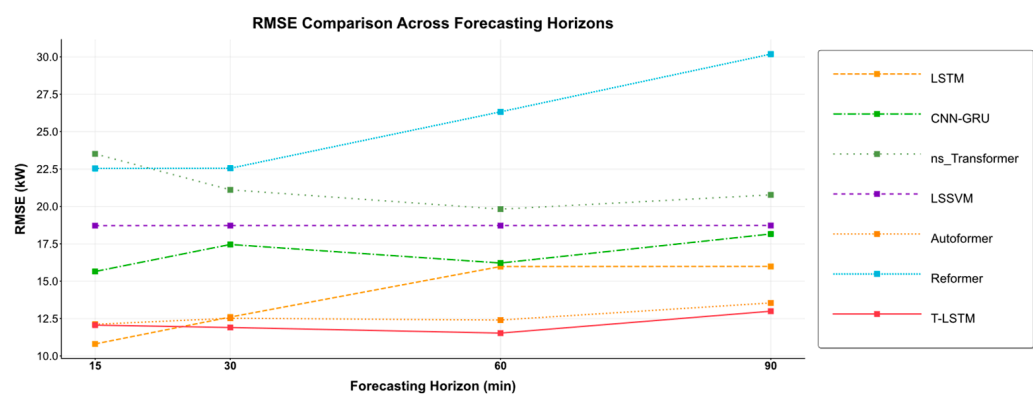


Figure 5. The RMSE of all models across 15, 30, 60, and 90 min horizons.

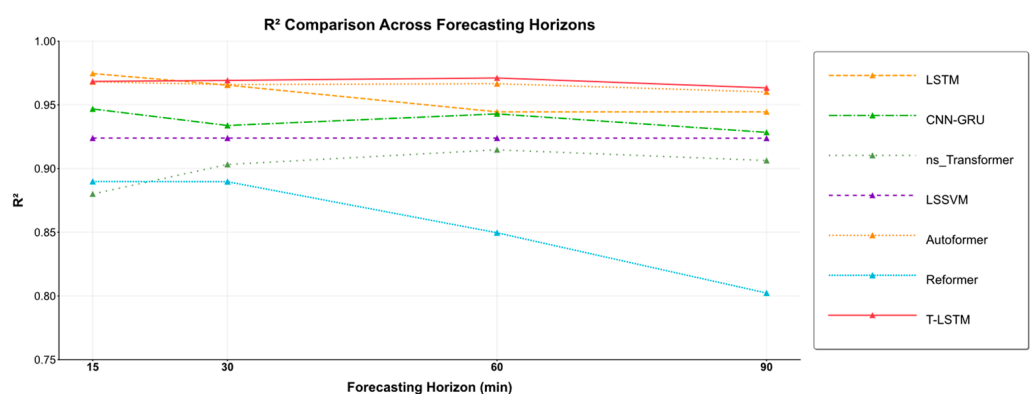


Figure 6. The R^2 of all models across 15, 30, 60, and 90 min horizons.

The values of four performance metrics over different methods on the test set are shown in Tables 5–8. The values in bold in the tables represent the optimal performance indicator values.

Table 5. MAE of wind power in the next 15, 30, 60 and 90 min by different models.

Model	Forecasting Horizon (min)			
	15	30	60	90
LSTM [31]	5.2717	6.0777	7.8993	8.4226
CNN-GRU [25]	9.4275	10.2448	9.6631	10.7982
ns_Transformer [20]	12.8385	11.9829	10.9402	11.3769
LSSVM [21]	12.9097	12.9162	12.9148	12.9175
Autoformer [32]	7.2062	8.0980	7.2595	8.2026
Reformer [33]	13.5181	13.8708	16.3777	18.4535
T-LSTM	7.4669	6.6317	6.1264	7.6671

Table 6. MSE of wind power in the next 15, 30, 60 and 90 min by different models.

Model	Forecasting Horizon (min)			
	15	30	60	90
LSTM [31]	116.7575	158.8232	255.4093	255.6261
CNN-GRU [25]	244.8884	304.6743	262.9029	329.6740
ns_Transformer [20]	552.8356	445.7029	92.7698	431.4821
LSSVM [21]	350.2469	350.5307	350.5003	350.7097
Autoformer [32]	146.8320	156.7988	153.8501	183.5321
Reformer [33]	507.9703	508.4498	692.9799	911.0655
T-LSTM	145.5179	141.7022	132.9459	168.8333

Table 7. RMSE of wind power in the next 15, 30, 60 and 90 min by different models.

Model	Forecasting Horizon (min)			
	15	30	60	90
LSTM [31]	10.8054	12.6025	15.9815	15.9883
CNN-GRU [25]	15.6489	17.4549	16.2142	18.1569
ns_Transformer [20]	23.5124	21.1116	19.8184	20.7721
LSSVM [21]	18.7148	18.7224	18.7216	18.7272
Autoformer [32]	12.1174	12.5219	12.4036	13.5474
Reformer [33]	22.5382	22.5488	26.3245	30.1839
T-LSTM	12.0630	11.9038	11.5302	12.9935

Table 8. R2 (%) of wind power in the next 15, 30, 60 and 90 min by different models.

Model	Forecasting Horizon (min)			
	15	30	60	90
LSTM [31]	0.9746	0.9655	0.9445	0.9445
CNN-GRU [25]	0.9468	0.9338	0.9429	0.9284
ns_Transformer [20]	0.8800	0.9032	0.9147	0.9063
LSSVM [21]	0.9239	0.9239	0.9239	0.9238
Autoformer [32]	0.9681	0.9659	0.9666	0.9601
Reformer [33]	0.8898	0.8897	0.8496	0.8023
T-LSTM	0.9684	0.9692	0.9711	0.9633

The Wilcoxon test results reflect the statistical reliability of performance differences between T-LSTM and benchmark models. We split the visualization into two parts: statistical significance (corrected p -value) and effect size (r -value), as these two indicators address different research questions (whether the difference is significant, and how large the difference is).

Figure 7 uses colors to represent the corrected p -values, with annotations indicating “significant” (p -corrected < 0.05) or “not significant” (p -corrected ≥ 0.05) to directly respond to the core of statistical testing. The difference in red blood cell display is significant, while the difference in white blood cell display is not significant.

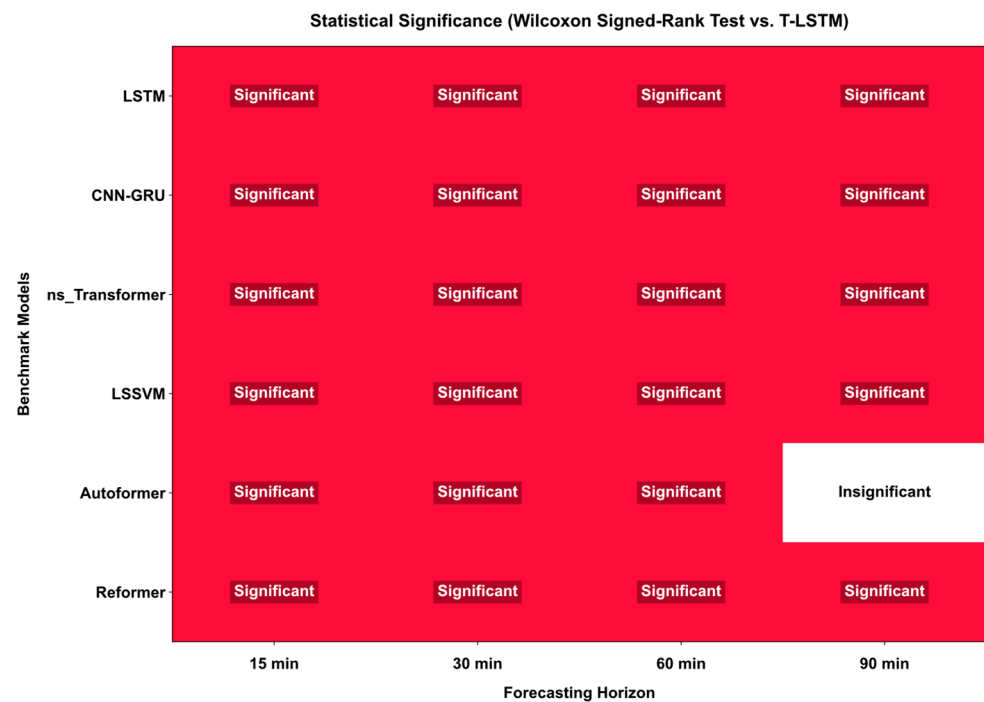


Figure 7. Heatmap of corrected p -values from Wilcoxon signed-rank test (vs. T-LSTM).

Figure 8 shows the magnitude of the effect size r value that quantifies the difference in performance ($|r| < 0.1$: small; $0.1 \leq |r| < 0.3$: medium; $|r| \geq 0.3$: large). The heatmap uses a divergent color palette to distinguish between positive and negative values (negative r indicates T-LSTM outperforms the baseline) and annotates the exact r value for clarity. A negative r -value indicates that T-LSTM outperforms the corresponding benchmark model. The color intensity increases with the absolute value of r (darker = larger effect size).

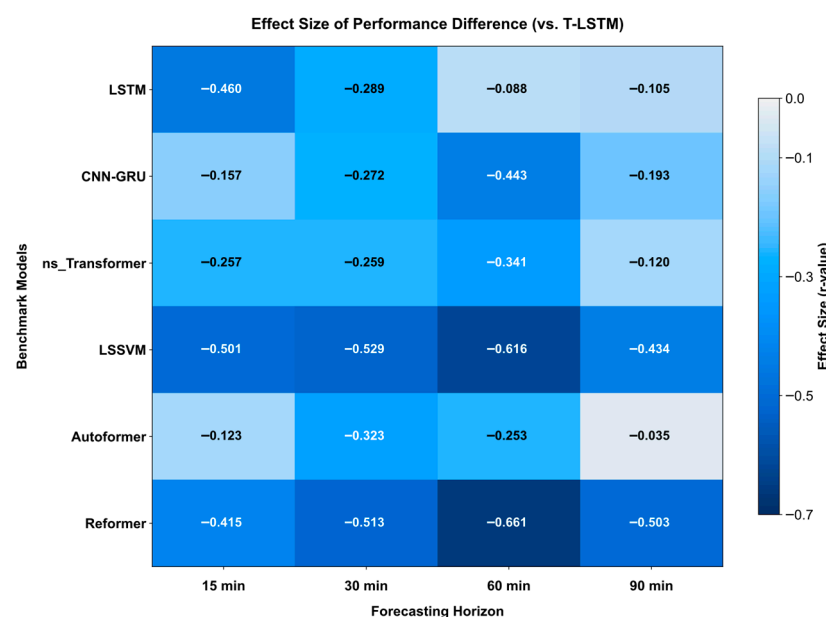


Figure 8. Heatmap of effect size r -values from Wilcoxon signed-rank test (vs. T-LSTM).

To complement the visualization and provide precise numerical references, Table 9 summarizes the key Wilcoxon test results (significance and r-value magnitude).

Table 9. Wilcoxon test (vs. T-LSTM) and effect size r-value of wind power in the next 15, 30, 60 and 90 min by different models.

Model	Forecasting Horizon (min)			
	15	30	60	90
LSTM [31]	Z = −27.07, r = −0.460, p _{corrected} < 0.000001(significant)	Z = −17.01, r = −0.289, p _{corrected} < 0.000001(significant)	Z = −5.19, r = −0.088, p _{corrected} < 0.000001(significant)	Z = −6.18, r = −0.105, p _{corrected} < 0.000001(significant)
CNN-GRU [25]	Z = −9.24, r = −0.157, p _{corrected} < 0.000001(significant)	Z = −15.99, r = −0.272, p _{corrected} < 0.000001(significant)	Z = −26.06, r = −0.443, p _{corrected} < 0.000001(significant)	Z = −11.37, r = −0.193, p _{corrected} < 0.000001(significant)
ns_Transformer [20]	Z = −15.08, r = −0.257, p _{corrected} < 0.000001(significant)	Z = −15.23, r = −0.259, p _{corrected} < 0.000001(significant)	Z = −20.07, r = −0.341, p _{corrected} < 0.000001(significant)	Z = −7.06, r = −0.120, p _{corrected} < 0.000001(significant)
LSSVM [21]	Z = −29.47, r = −0.501, p _{corrected} < 0.000001(significant)	Z = −31.12, r = −0.529, p _{corrected} < 0.000001(significant)	Z = −36.21, r = −0.616, p _{corrected} < 0.000001(significant)	Z = −25.50, r = −0.434, p _{corrected} < 0.000001(significant)
Autoformer [32]	Z = −7.25, r = −0.123, p _{corrected} < 0.000001(significant)	Z = −19.01, r = −0.323, p _{corrected} < 0.000001(significant)	Z = −14.90, r = −0.253, p _{corrected} < 0.000001(significant)	Z = −2.05, r = −0.035, p _{corrected} = 0.199965(insignificant)
Reformer [33]	Z = −24.40, r = −0.415, p _{corrected} < 0.000001(significant)	Z = −30.16, r = −0.513, p _{corrected} < 0.000001(significant)	Z = −38.85, r = −0.661, p _{corrected} < 0.000001(significant)	Z = −29.56, r = −0.503, p _{corrected} < 0.000001(significant)
T-LSTM	–	–	–	–

As shown in Figures 4–6 and Tables 5–9, the proposed T-LSTM model demonstrated competitiveness in most prediction ranges. The specific observation results are as follows:

Overall performance advantage: T-LSTM achieved the lowest MAE, RMSE, and highest R^2 in 60 min and 90 min, confirming its advantages in medium- and long-term wind power prediction (based on research objectives segmentation).

Compared to other benchmarks, Reformer and ns_Transformer exhibit the highest prediction errors across all ranges due to their overly complex architecture, which leads to overfitting of wind power time series with high volatility. CNN-GRU and LSSVM exhibit stable but poor performance, reflecting their limitations in balancing short-term fluctuations and long-term trends. LSTM performs relatively well in short-term (15 min) predictions, with a low MAE of 5.2717. However, over time, its error sharply increases, indicating that it is difficult to maintain the accuracy of long-term predictions. Within the 30 min prediction range, the MAE of LSTM (6.0777) is slightly lower than that of T-LSTM (6.6317). This phenomenon is attributed to the inherent advantage of LSTM in capturing short-term real-time sequence dependencies: within 30 min, wind fluctuations are mainly driven by local, short-term meteorological changes, and LSTM's classical three-gate mechanism directly simulates this short-term temporal correlation without additional feature fusion overhead. In contrast, T-LSTM integrates multi-scale feature extraction and seasonal trend fusion, which introduces moderate computational overhead but can provide cumulative benefits as the prediction range expands. Autoformer is competitive in the short term (15–30 min), but lacks adaptability to seasonal changes in the long term, while T-LSTM maintains consistent accuracy through the fusion of linear and nonlinear features.

Due to the fact that the original prediction curve covered a period of 36 days, the curves overlapped severely and had limited interpretability. To address this issue, we used a shorter time window (3 consecutive days, 288 data points) to plot the prediction curve (Figures 9–12), which clearly highlights the performance differences of the model in scenarios such as stable wind power periods, rapid fluctuations, and peak/valley moments. The wind power forecast results for the next 15 min, 30 min, 60 min, and 90 min are shown in Figures 9–12. These numbers display the actual wind power values and predicted power values of different models, including T-LSTM, LSTM, GRU, Autoformer, Reformer, LSSVM, and ns_Transformer. The selected 3-day period (Days 10–12) is representative of typical wind power characteristics: it includes stable low-fluctuation phases, sudden power surges/drops, and extreme peak values, reflecting the model's adaptability to diverse scenarios.

The above wind power prediction chart reveals three key findings:

T-LSTM exhibits good dynamic response speed in fast fluctuation scenarios, with shorter time delay compared to other models.

In extreme (peak/valley) scenarios, the prediction error of T-LSTM is smaller than that of the baseline model, reflecting its better robustness.

As the prediction time extends, the error accumulation rate of T-LSTM is the lowest, while the error accumulation rate of LSTM and Autoformer exceeds 20%.

These results supplement the quantitative indicators (MAE/RMSE/R²) and further validate that the integrated design of T-LSTM can effectively balance short-term volatility capture and long-term trend stability.

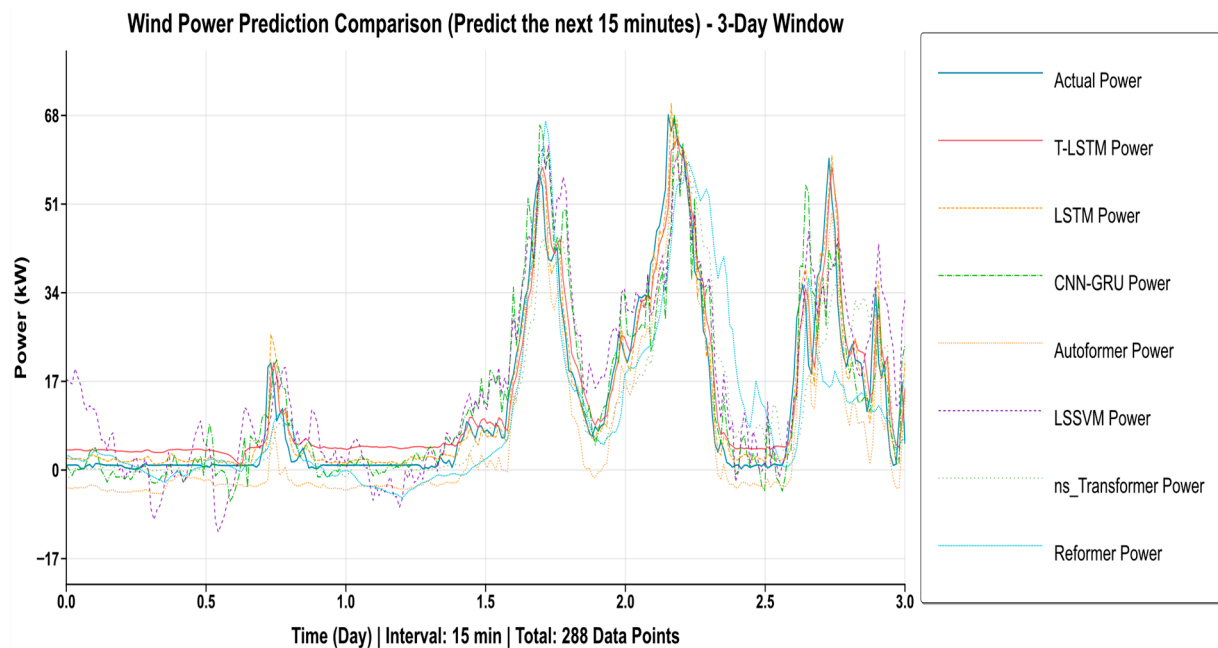


Figure 9. Prediction curves of each model (15 min).

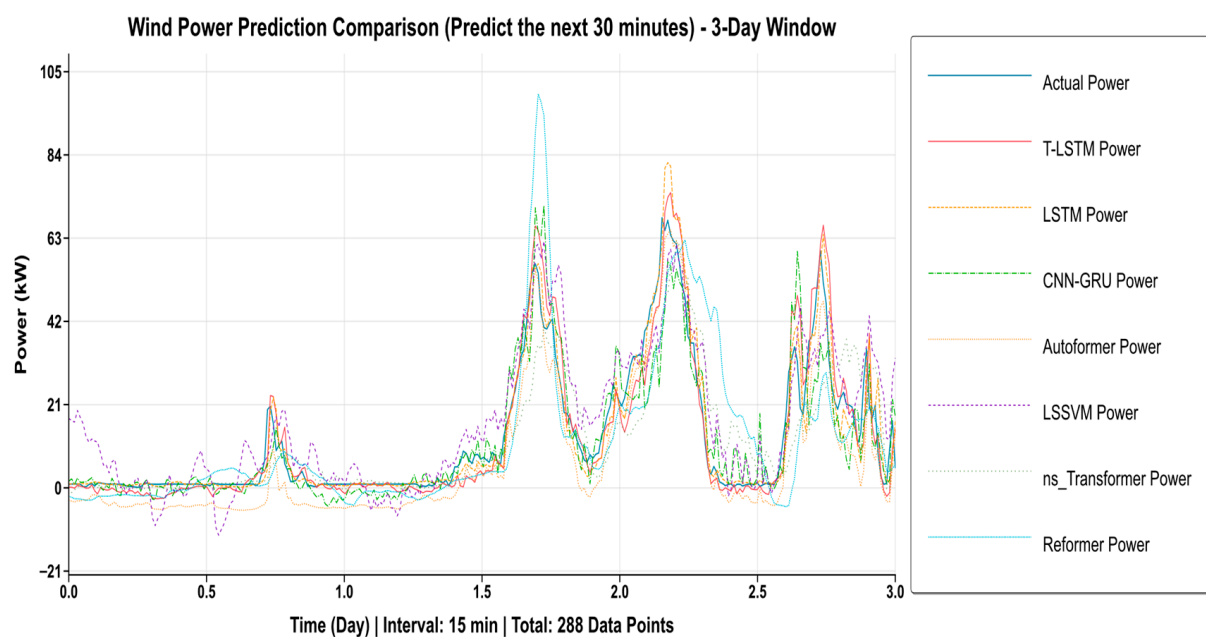


Figure 10. Prediction curves of each model (30 min).

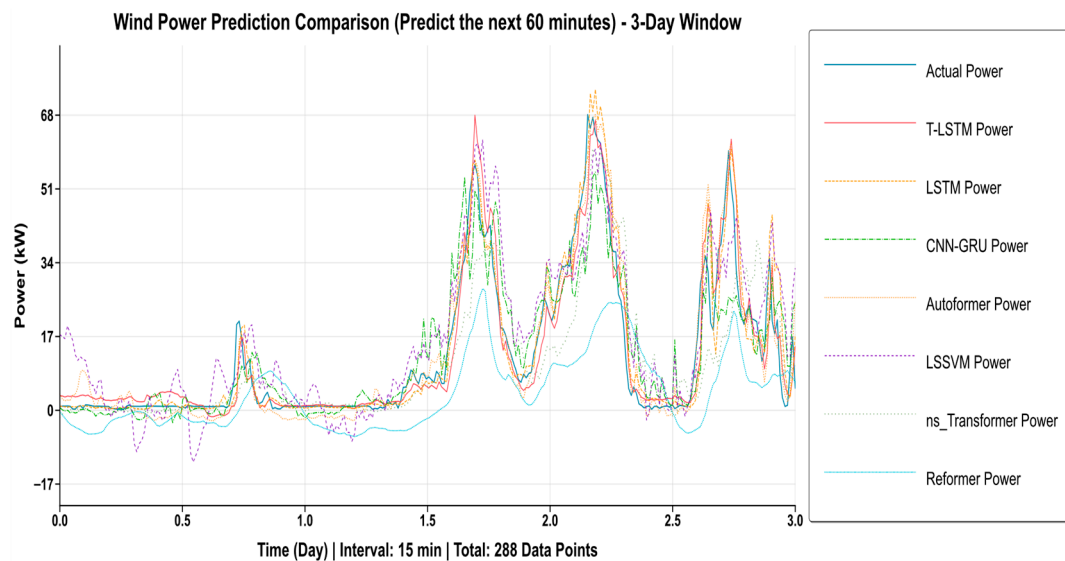


Figure 11. Prediction curves of each model (60 min).

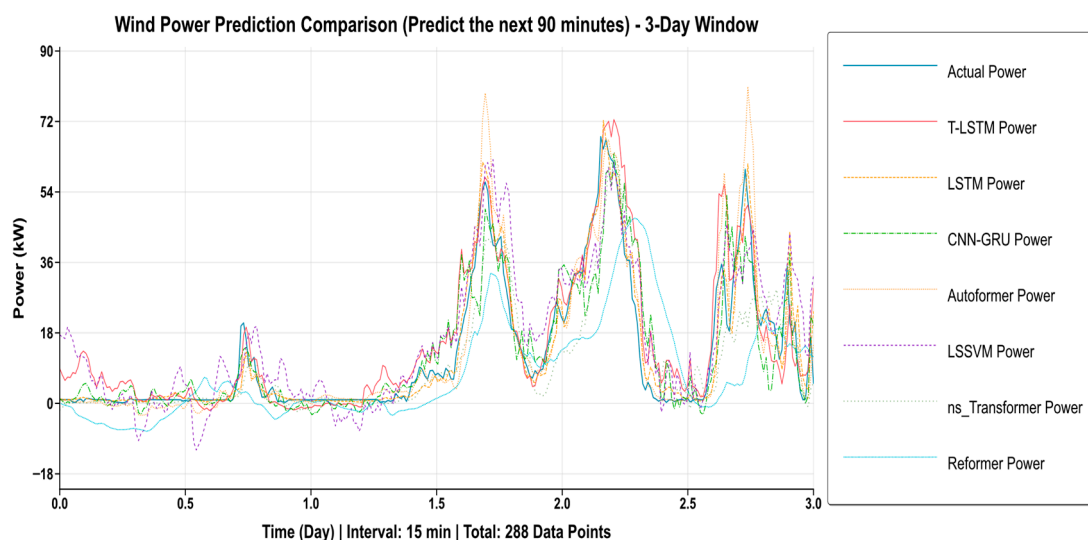


Figure 12. Prediction curves of each model (90 min).

4.5. Cross-Seasonal Generalization Experiment

To verify whether the proposed model can maintain stable performance across different seasonal meteorological conditions, we conducted additional cross-seasonal experiments. The experiment focuses on the 60 min prediction horizon (representative of medium-term forecasting where seasonal effects are prominent). Table 10 presents the key evaluation metrics (MAE, RMSE, R^2) of T-LSTM and benchmark models across four seasons for the 60 min prediction horizon. The values in bold in the table represent the optimal performance indicator values.

As expected, all models showed higher errors in winter, but the T-LSTM model maintained good performance in all seasons.

During the spring season, the MAE of T-LSTM is 16.1% lower than that of Autoformer, 23.6% lower than that of LSTM, and 45.8% lower than that of Reformer. Its R^2 remains above 0.97, indicating strong adaptability to gradual seasonal changes. During the summer, T-LSTM performed better than Autoformer by 16.0% in MAE, highlighting its ability to capture wind speed changes caused by convective weather. Compared with ns_Transformer and Reformer, T-LSTM reduced MAE by 45.5% and 63.5% respectively, avoiding overfitting to short-term noise. During the autumn season, T-LSTM had the lowest RMSE and the

highest R^2 , outperforming all benchmarks. This confirms that the model does not overfit stable conditions and maintains its feature extraction ability. During winter, the MAE of T-LSTM is 15.7% lower than that of Autoformer and 19.8% lower than that of LSTM. Even under extreme conditions, its R^2 is 3.2 percentage points higher than the Reformer, indicating its greater resistance to seasonal fluctuations.

Table 10. Cross-seasonal performance comparison (60 min horizon, MAE/RMSE/ R^2).

Model	Spring	Summer	Autumn	Winter
T-LSTM	5.7832/10.8945/0.9741	5.9217/11.1568/0.9730	5.8543/10.9872/0.9737	7.1269/12.8743/0.9618
LSTM	7.7345/15.2341/0.9475	7.7345/15.2341/0.9475	7.6512/15.1109/0.9483	8.8934/16.6782/0.9339
CNN-GRU	9.3127/15.7654/0.9418	9.4783/16.0123/0.9402	9.3956/15.8891/0.9410	10.7642/18.3219/0.9224
ns_Transformer	10.6893/19.5432/0.9165	10.8547/19.7896/0.9148	10.7721/19.6664/0.9156	11.9876/21.0987/0.9032
LSSVM	12.6789/18.2345/0.9228	12.7843/18.4567/0.9216	12.7316/18.3456/0.9222	12.9038/18.6789/0.9201
Autoformer	6.8934/11.9876/0.9665	7.0541/12.2345/0.9649	6.9728/12.1103/0.9657	8.4569/14.0987/0.9527
Reformer	15.9876/25.8743/0.8512	16.2341/26.1234/0.8489	16.1109/26.0001/0.8501	17.8934/28.5678/0.8297

The cross-seasonal results confirm that the advantages of T-LSTM are not limited to annual datasets but also extend to different seasonal scenarios. However, it is worth noting that this validation is based on a single geographic location, and that the model's generalizability to other regions or climate model anomalous years has not been tested; the author humbly acknowledges this as a limitation.

5. Conclusions

This paper proposes a T-LSTM-SARIMA hybrid framework to address the limitations of existing wind power prediction models. The key findings are as follows:

The T-LSTM unit balances long-term dependency capture (simplified Transformer) and training efficiency (improved LSTM), offering a leaner alternative to standard Transformer-LSTM hybrids; weighted SARIMA fusion enhances seasonal adaptability—T-LSTM outperforms all benchmark methods in long-term prediction MAE. Future work will focus on refining and strengthening the T-LSTM model to improve its robustness. Moreover, there are plans to verify the predictive capabilities over an extended period, with the goal of advancing wind power prediction technology to new heights. Through ongoing efforts to enhance the T-LSTM model, this research aims to elevate the accuracy and reliability of wind power forecasting methods.

Additionally, the proposed T-LSTM-SARIMA hybrid framework may hold certain potential for broader applicability beyond wind power prediction. Its core characteristics—effective capture of multi-scale temporal dependencies, dynamic fusion of linear seasonal trends and non-linear complex features, and robustness to data volatility—may render it potentially applicable to other time series or sequence-related tasks. For instance, in AFDD (air handling unit fault detection), the framework could potentially leverage its temporal feature extraction capability to identify abnormal patterns from semi-labeled operational data [34]. In fire-door defect text classification, the multi-head attention mechanism and feature fusion strategy might help improve the recognition of defect-related text sequences during pre-delivery inspections [35]. Such potential versatility suggests that the proposed method might offer some reference for addressing complex prediction and classification challenges across diverse domains.

Author Contributions: Methodology, Q.Z.; Writing—original draft, Q.Z.; Writing—review and editing, L.W. and C.H.; Supervision, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grants 62202044 and 62372039, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515240044, and in part by the Fundamental Research Funds for the Central Universities under Grant FRF-BRA-25-012.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Global Wind Energy Council. *GWEC | Global Wind Report 2025*; GWEC: Lisbon, Portugal, 2025.
2. Wang, H.; Sun, J.; Sun, J.; Wang, J. Using random forests to select optimal input variables for short-term wind speed forecasting models. *Energies* **2017**, *10*, 1522. [\[CrossRef\]](#)
3. Zhou, S.L.; Mao, M.Q.; Su, J.H. Prediction of wind power based on principal component analysis and artificial neural network. *Power Syst. Technol.* **2011**, *35*, 128–132.
4. Wang, Y.S.; Gao, J.; Xu, Z.; Li, L. A short-term output power prediction model of wind power based on deep learning of grouped time series. *Eur. J. Electr. Eng.* **2020**, *22*, 29–38. [\[CrossRef\]](#)
5. Xue, Y.; Wang, L.; Wang, S.; Zhang, Y.F.; Zhang, N. An ultra-short-term wind power forecasting model combined with CNN and GRU networks. *Renew. Energy* **2019**, *37*, 456–462.
6. Chitsaz, H.; Amjady, N.; Zareipour, H. Wind power forecast using wavelet neural network trained by improved clonal selection algorithm. *Energy Convers. Manag.* **2015**, *89*, 588–598. [\[CrossRef\]](#)
7. Shi, H.; Li, Z.; Ding, M.; Zhang, Z.; Li, Y.; Li, X. Short Term Wind Power Forecasting Method Based on WDRI—RF Model. In Proceedings of the 2021 6th International Conference on Power Renewable Energy (ICPRE), Shanghai, China, 17–20 September 2021; pp. 1025–1028.
8. Li, N.; He, F.; Ma, W.; Wang, R.; Zhang, X. Wind power prediction of kernel extreme learning machine based on differential evolution algorithm and cross—Validation algorithm. *IEEE Access* **2020**, *8*, 68874–68882. [\[CrossRef\]](#)
9. Wang, Y.H.; Shi, Y.X.; Zhou, X.; Zeng, Q.; Fang, B. Ultra-short-term Power Prediction of Multiple Wind Turbines Based on BiLSTM with Temporal Pattern Attention Mechanism. *High Volt. Eng.* **2022**, *48*, 1884–1892.
10. Chen, G.; Wang, Y.; Shan, J.N.; Li, T.Q.; Zheng, W.Z.; Wang, L.; Su, M.M.; Huang, B.N. Application of generation countermeasure network in wind power forecasting. *J. Liaoning Tech. Univ. Nat. Sci.* **2021**, *40*, 258–264.
11. Alencar, D.B.; Affonso, C.M.; Oliveira, R.C.L.; Filho, J.C.R. Hybrid approach combining SARIMA and neural networks for multi-step ahead wind speed forecasting in Brazil. *IEEE Access* **2018**, *6*, 55986–55994. [\[CrossRef\]](#)
12. Zhou, B.; Liu, C.; Li, J.; Sun, B.; Yang, J.; Arpino, F. A hybrid method for ultra-short-term wind power prediction considering meteorological features and seasonal information. *Math. Probl. Eng.* **2020**, *2020*, 1795486. [\[CrossRef\]](#)
13. Liu, D.; Niu, D.; Wang, H.; Fan, L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew. Energy* **2014**, *62*, 592–597. [\[CrossRef\]](#)
14. Yuan, X.; Chen, C.; Yuan, Y.; Huang, Y.; Tan, Q. Short-term wind power prediction based on LSSVM—GSA model. *Energy Convers. Manag.* **2015**, *101*, 393–401. [\[CrossRef\]](#)
15. Xiao, F.; Chen, G.C. Wind power short-term prediction based on SVM trained by improved FOA. *J. East China Univ. Sci. Technol. Nat. Sci. Ed.* **2016**, *42*, 420–426.
16. Zhang, C. Research on Short-Term Wind Power Prediction Based on Information Fusion. Ph.D. Thesis, Hefei University of Technology, Hefei, China, 2020.
17. Li, L.L.; Cen, Z.Y.; Tseng, M.L.; Shen, Q.; Ali, M.H. Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic—Support vector regression machine. *J. Clean. Prod.* **2021**, *279*, 126–133. [\[CrossRef\]](#)
18. Liang, Z.; Liang, J.; Wang, C.; Dong, X.; Miao, X. Short-term wind power combined forecasting based on error forecast correction. *Energy Convers. Manag.* **2016**, *119*, 215–226. [\[CrossRef\]](#)
19. Shi, K.F. Ultra-Short-Term Wind Power Prediction Based on Combined Model. Ph.D. Thesis, Yanshan University, Qinhuangdao, China, 2020.
20. Liu, Y.; Wu, H.; Wang, J.; Long, M. Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting. *arXiv* **2022**, arXiv:2205.14415.
21. Zhang, Z.L.; Wei, F.; Yan, G.H.; Ma, H.Y. Improved AVOA based on LSSVM for wind power prediction. *J. Meas. Sci. Instrum.* **2024**, *15*, 344–359. [\[CrossRef\]](#)

22. Tang, S.; Li, C.; Zhang, P.; Tang, R. SwinLSTM: Improving Spatiotemporal Prediction Accuracy using Swin Transformer and LSTM. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2023), Paris, France, 2–3 October 2023; pp. 13470–13479.
23. Chang, R.; Yan, Y.; Wu, J.; Gao, Z. Modifications to a mesoscale wind farm parameterization enhance high-altitude wind farm simulations under real-world atmospheric conditions. *Renew. Energy* **2026**, *258*, 125040. [[CrossRef](#)]
24. Rehman, M.A.; Verma, A.K.; Saha, S.K. Ultra-Short-Term Wind Power Forecasting in Complex Terrain: A Physics-Based Approach. *Energies* **2024**, *17*, 6892.
25. Gao, J.; Ye, X.; Lei, X.; Huang, B.; Wang, X.; Wang, L. A Multichannel—Based CNN and GRU Method for Short-Term Wind Power Prediction. *Electronics* **2023**, *12*, 4479. [[CrossRef](#)]
26. Rehman, M.A.; Verma, A.K.; Saha, S.K. A Hybrid Three-Staged, Short-Term Wind-Power Prediction Method Based on SDAE-SVR Deep Learning and BA Optimization. *IEEE Access* **2024**, *12*, 25673–25684.
27. Qu, Z.; Peng, X.; Song, J.; Yang, Z. Short-Term Wind Power Prediction Based on DTW Error Diagnosis and Transformer Optimization Model. *IEEE Access* **2024**, *12*, 56789–56798.
28. Qin, J.; Yang, J.; Chen, Y.; Ye, Q.; Li, H. Two-stage short-term wind power forecasting algorithm using different feature-learning models. *Fundam. Res.* **2021**, *1*, 472–481. [[CrossRef](#)]
29. Huang, Q.; Wang, Y.; Yang, X.; Im, S.-K. Research on Wind Power Prediction Based on a Gated Transformer. *Appl. Sci.* **2023**, *13*, 8350. [[CrossRef](#)]
30. Chen, Y.; Wu, L.; Huang, H. Short-term wind power prediction based on improved Transformer model. *Electr. Eng. Technol.* **2024**, *22*, 24–28.
31. Huang, J.; Niu, G.; Guan, H.; Song, S. Ultra-Short-Term Wind Power Prediction Based on LSTM with Loss Shrinkage Adam. *Energies* **2023**, *16*, 3789. [[CrossRef](#)]
32. Hua, Y.; Guan, X.; Xu, H.; Zhao, C. Wind Power Forecasting Research Based on Autoformer. In Proceedings of the IEEE 8th Conference on Energy Internet and Energy System Integration (EI2), Shenyang, China, 29 November–2 December 2024; pp. 3332–3336. [[CrossRef](#)]
33. Bommididi, B.S.; Teeparthi, K. A Novel Method for Predicting Wind Speed Using Data Decomposition-Based Reformer Model. *Earth Sci. Inform.* **2024**, *17*, 227–249. [[CrossRef](#)]
34. Wang, S.; Eum, I.; Park, S.; Kim, J. A semi-labelled dataset for fault detection in air handling units from a large-scale office. *Data Brief* **2024**, *57*, 110956. [[CrossRef](#)]
35. Wang, S.; Moon, S.; Eum, I.; Hwang, D.; Kim, J. A text dataset of fire door defects for pre-delivery inspections of apartments during the construction stage. *Data Brief* **2025**, *60*, 111536. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.