Electrical Engineering
https://doi.org/10.1007/s00202-025-03295-1

ORIGINAL PAPER

Advancing short-term wind power forecasting by AI-driven models for improved accuracy

B. Kishore[1] · K. Kabilan[1] · L. S. Rahul[1] · G. Prajwal Priyadarshan[1] · E. P. Vishnutheerth[1] · Rahul Satheesh[1] · Mohan Lal Kolhe[2]

Received: 27 May 2025 / Accepted: 17 July 2025

Abstract
Accurate short-term wind power forecasting is crucial for the effective incorporation of renewable energy into modern power systems. Building upon previous hybrid deep learning approaches, this study presents an advanced, AI-driven framework that extensively compares machine learning and deep learning models for short-term wind power forecasting utilizing SCADA data, including variables such as wind speed, direction, theoretical power, and actual power output at 10 mins intervals. Existing approaches suffer from inadequate preprocessing of noisy SCADA data, lack of anomaly detection mechanisms, and the absence of physics-informed constraints ensuring compliance with aerodynamic principles. Moreover, many models fail to capture complex nonlinear and temporal dependencies, limiting their forecasting reliability. This study proposes an AI-driven forecasting framework that compares ML and DL models such as XGBoost, LightGBM, Random Forest, Transformer variations, and Temporal Fusion Transformer with processed data parameters. Anomaly detection is performed using physics-informed neural networks and long short-term memory networks ensuring physical plausibility of forecasts under Betz' law. Gaussian noise augmentation, normalization, and dimensionality reduction using principal component analysis, which reduced the input feature space while retaining over 90% of the variance, are used for preprocessing. XGBoost demonstrates the highest forecasting accuracy among all evaluated models, attaining an R^2 score of 0.9890, a mean squared error of 0.0015, a root-mean-squared error of 0.0385, and a mean absolute error of 0.0151 following preprocessing. These findings emphasize the efficiency of ensemble learning, especially gradient boosting, in addressing complex nonlinearities present in wind power generation. These results indicate that a data-driven and preprocessing-enhanced methodology significantly improves prediction reliability, thus facilitating more intelligent and sustainable energy systems.

Keywords Machine learning models · Renewable energy · Short-term prediction · Temporal fusion transformer (TFT) · Variational mode decomposition (VMD) · Wind power forecasting

1 Introduction

The global shift toward cleaner and more sustainable energy systems has highlighted renewable energy sources, especially wind energy. Wind power provides an eco-friendly and cost-effective substitute for fossil fuels, making substantial contributions to global energy supply objectives [1]. The intrinsic instability and unpredictability of wind present significant issues for energy system operators, particularly in real-time supply and demand balancing. Short-term wind power forecasting, spanning minutes to several hours ahead, is essential for maintaining grid stability, enhancing turbine efficiency, and facilitating strategic energy trading strategies [2]. Conventional forecasting methods, such as statistical models and physics-based simulations, often fall short when dealing with high-frequency, non-stationary time series data. Conversely, data-driven methodologies, especially machine learning (ML) and deep learning (DL) models, have shown significant capability in capturing temporal relationships and nonlinear dynamics from historical data. The literature presents a wide array of methodologies, from conventional statistical models to sophisticated hybrid neural networks

✉ Rahul Satheesh
s_rahul1@cb.amrita.edu

[1] Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore, India

[2] Department of Engineering Sciences, University of Agder, Kristiansand, Norway

Published online: 07 August 2025

and ensemble learning techniques, each tackling the intrinsic issues associated with the stochastic characteristics of wind.

A substantial body of research has demonstrated the effectiveness of hybrid neural network architectures for wind power forecasting. These models include the advantages of convolutional and recurrent layers, facilitating the extraction of both spatial and temporal characteristics from wind farm data. The use of supplementary meteorological and environmental data, including sun irradiance, has significantly improved forecasting accuracy. Hybrid models that combine artificial neural networks, LSTM, and support vector regression, utilizing solar data and wind measurements, have demonstrated improved forecasting accuracy, especially in regions with considerable weather variability [3]. Transfer learning and variational autoencoders have been employed to enhance forecasting models for diverse wind farm sites, illustrating the promise for scalable and transferable solutions [4].

Optimization algorithms have become essential to recent progress in wind power forecasting. Methods including the sparrow search algorithm, kernel-based extreme learning machines, and particle swarm optimization have been utilized to optimize model parameters and improve convergence speed and resilience, even amidst non-stationary data [5]. Hybrid frameworks that integrate many ML techniques utilize the complimentary advantages of each, yielding models that generalize effectively across numerous datasets and operational contexts [6]. Clustering techniques have been employed to partition data into homogenous groups, which, when integrated with neural networks and optimization algorithms, provide more accurate modeling of local wind patterns and diminish overall forecast error [7]. Deep learning models, especially those utilizing attention mechanisms, have become prominent for their ability to represent complex temporal and spatial connections. Attention-based RNNs, CNNs, and hybrid transformer designs have exhibited significant enhancements in accuracy and computational efficiency relative to traditional models [8]. The integration of diverse data sources, encompassing meteorological, operational, and environmental characteristics, significantly amplifies the predictive capability of these models [9].

Hybrid models that include machine learning and deep learning techniques have consistently surpassed single-model methodologies. Feature selection techniques and ensemble learning methodologies, including stacking and bagging, have demonstrated enhancements in both accuracy and robustness [10]. Attention-based deep neural networks, when integrated with meteorological data and teaching-learning-based optimization, have achieved superior outcomes in short-term wind power forecasting [11–13]. Comparative analyses underscore the superiority of attention-based RNNs compared to conventional LSTM and GRU models, especially regarding their capacity to concentrate on pertinent input characteristics and handle long-range dependencies [14]. The integration of these methods with deep learning time series models and sophisticated feature selection techniques has enhanced predicting results [15, 16]. Ultra-short-term forecasting has particularly benefited from hybrid CNN-LSTM models and the explicit modeling of atmospheric uncertainties through variational mode decomposition [17, 18]. Thorough evaluations of deep learning methodologies in wind forecasting highlight the inclination toward hybrid and ensemble models, which amalgamate the advantages of several techniques to tackle the intricate, nonlinear, and non-stationary characteristics of wind power data [19]. Transformer-based designs, including memory-efficient models and those employing convolutional distillation and series decomposition, have established new standards for mid-term and short-term wind power forecasting by adeptly capturing long-term trends and periodic patterns. Adaptive hybrid models integrating CNNs, Informer topologies, and sophisticated feature extraction approaches have exhibited robust performance in both short-term and ultra-short-term contexts, highlighting the continuous advancement of hybrid and attention-based forecasting methodologies [20].

Ensemble learning methods, including XGBoost, CatBoost, Random Forest, and linear regression, have demonstrated efficacy in short-term wind power forecasting by more effectively capturing nonlinearities and volatility compared to singular models [21]. Ultra-short-term forecasting (minutes to hours) has progressed with hybrid deep learning models, specifically CNN-LSTM combinations with attention mechanisms, demonstrating robust performance, while issues like as data gaps and swiftly changing weather conditions remain prevalent [22]. Transformer-based models improve mid-term forecasting through memory-efficient architectures and frequency-aware attention, facilitating precise, low-resource predictions [23]. CNN-Informer hybrids enhance accuracy by effectively extracting multi-scale features and managing long-sequence data [24]. Conventional statistical models retain their significance, particularly when combined with high-resolution, fully observable wind flow data, providing interpretable and dependable forecasts [25].

The study in [26] proposed an LSTM-based model optimized through sparrow search and firefly algorithms, enhancing predictive performance by fine-tuning hyperparameters. Furthermore, [27] presented a comprehensive review of wind forecasting techniques, highlighting how statistical and machine learning approaches can be effectively integrated to improve accuracy and address grid integration challenges. The study in [28] introduced a deep learning-based model incorporating an attention mechanism to enhance short-term wind power forecasting at Senegal's flagship wind farm, significantly improving prediction accuracy by focusing on relevant temporal features. In [29], an optimized multilayer perceptron neural network (MLP-NN) approach is presented, utilizing machine learning techniques to fine-tune model parameters, thereby achieving better forecasting performance through enhanced learning efficiency. Furthermore, [30] proposed a hybrid model combining generative adversarial networks (GANs) with long short-term memory (LSTM) networks to generate realistic wind power scenarios and improve sequential prediction capabilities, effectively capturing the complex and volatile characteristics of wind data. Furthermore, [31] demonstrated the efficacy of machine learning algorithms in enhancing wind power forecasting, emphasizing their significance for renewable energy applications. Accurate short-term wind power forecasting remains challenging due to the non-stationary and noisy nature of wind data. Existing models often lack robust preprocessing strategies to extract meaningful features and fail to incorporate anomaly detection or physical consistency checks. There is a clear need for forecasting frameworks that integrate advanced signal decomposition, dimensionality reduction, and anomaly detection to enhance both prediction and reliability. The major contributions of the paper are:

1. To develop an anomaly detection model using PINN and LSTM. The main aim is to develop AI-driven machine learning (Random Forest, LightGBM) and deep learning (Transformer) models for short-term wind power forecasting.
2. To employ preprocessing techniques such as variational mode decomposition (VMD), Gaussian noise injection, and PCA to extract multi-scale, temporal, and nonlinear features.
3. To demonstrate predictions with scalability through comprehensive validation, making the framework suitable for real-world applications.

In order to address the limitations identified in existing studies such as lack of robust preprocessing, the absence of anomaly detection, and neglect of physical consistency, this paper provides a integrated techniques like VMD, PCA, and Gaussian noise augmentation for effective feature extraction. Additionally, this paper employ's LSTM-based anomaly detection and PIN to ensure data integrity and adherence to physical laws. The selection of models such as XGBoost, LightGBM, and Transformer variants was driven by their proven ability to handle nonlinear temporal patterns, making them well suited to overcome the shortcomings of earlier approaches.

2 Methodology

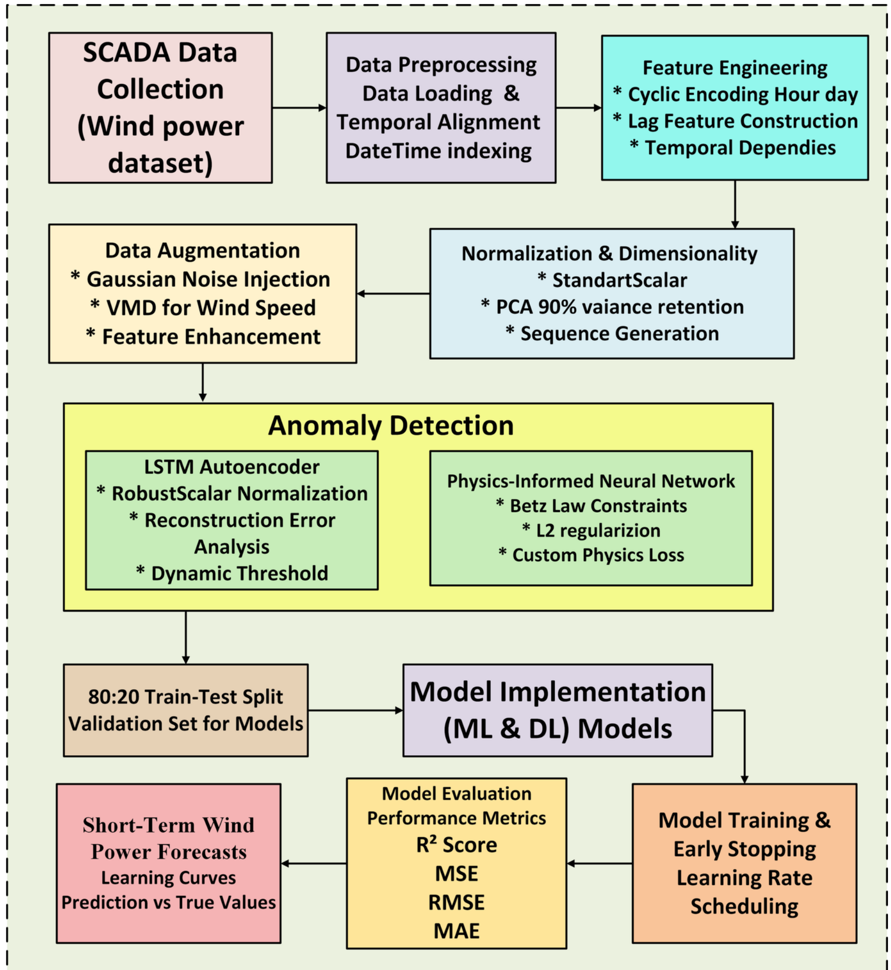
This section comprises the methodological framework used for the short-term wind power forecasting, as shown in Fig. 1. The overall process involves a number of phases, starting with data collection and processing of wind turbine data, followed by large-scale data preprocessing and feature engineering to prepare the dataset for predictive modeling. Under data preprocessing, missing values are handled and normalized in the values under the same range after which 80:20 train–test split is performed. Augmentation procedures are used to improve generalizability. In the following phase, a variety of machine learning and deep learning architectures, including ensemble models and Transformer-based architectures, are implemented and optimized for forecasting tasks. The performance of the models is then evaluated using popular statistical measures, while anomaly detection methods with data-driven models and physics-informed models are employed to ensure reliability and physical validity of the predictions. In this study, we employed cyclic encoding due to its proven effectiveness in handling periodic time features, as it preserves the continuity between the end and start of cycles. Cyclic encoding is widely recognized as the most appropriate method for representing temporal cyclicity in time series forecasting. Thus, sinusoidal encoding of hour/day-of-week captured temporal periodicity more effectively than one-hot encoding.

2.1 Dataset description

The data used in this research are wind turbine operation data obtained from a Supervisory Control and Data Acquisition (SCADA) system which is available on Kaggle [32]. The data are gathered at 10 mins intervals and include essential meteorological and turbine-generated information vital for effective wind power forecasting. It includes various wind conditions and fluctuations in power output, forming a basis for model training and assessment.

This study employs a dataset consisting of four critical input features and one target variable. The initial characteristic, date/time, denotes the timestamp linked to each reading and is essential for recording and analyzing temporal patterns and trends in wind power generation. Wind speed (m/s), recorded at the turbine's hub height, is a vital predictor, as wind power output fluctuates roughly in a cubic relationship with wind speed. The theoretical power curve (kWh) shows the predicted power output based on manufacturer parameters related to wind speed, providing as a standard for assessing the turbine's operational efficiency. Wind direction, measured in degrees, is an important parameter as variations in wind direction can affect the yaw control system and, in turn, alter the turbine's power generation efficiency. The target variable, LV active power (kW), shows the actual power output generated by the turbine, providing predictive modeling and allowing for a direct comparison between theoretical and actual power generation. This collection provides high-resolution time series data with accurate meteorological

Fig. 1 Methodology block diagram



characteristics, facilitating comprehensive modeling of wind power. The incorporation of theoretical power curves facilitates a thorough comparison examination of theoretical and actual power generation. This dataset is utilized to improve forecasting precision and enable effective wind energy management.

2.2 Data preprocessing

Data preprocessing is designed to ensure consistency, robustness, and extraction of informative characteristics from the raw time series dataset, and there are no missing values in the dataset. Two main pipelines were developed to serve the diverse modeling needs: (i) ML regression model and (ii) a Transformer-based model.

2.2.1 Data loading and temporal alignment

The raw time series data are initially loaded from CSV format. Given the time-dependent nature of the dataset, the "Date/Time" column is converted into the appropriate datetime format and subsequently set as the index. This transformation permits direct temporal indexing and facilitates subsequent time series operations.

2.2.2 Feature engineering

a) Cyclic Encoding of Temporal Features: Hour-of-day and day-of-week features are transformed into cyclic features using sine and cosine functions, preserving the periodicity of the data and preventing artificial discontinuities between cycle endpoints [33]. This cyclic representation prevents the artificial discontinuity between the end and start of the cycles.
b) Lag Feature Construction: Lag features are generated for the target variable to capture temporal dependencies. For instance, by using the previous observation as an input, both the Transformer and recurrent neural network pipelines (attention-based) embed short-term temporal dynamics directly in the feature space.

2.2.3 Normalization and dimensionality reduction using PCA

All models leverage normalization using a StandardScaler to scale the features, ensuring that each attribute contributes equally to model training. In DL-based transformer models, principal component analysis (PCA) is applied to reduce original features to 6 principal components while preserving 90% variance, eliminating multicollinearity. Figure 2 reveals 2 dominant components explain, more than 60% variance, indicating that wind speed and temporal cycles (hour/day) drive most power variability, and reduce dimensionality while retaining approximately 90% of the variance. This step is particularly important for high-dimensional data and is incorporated in both the DL and ML model pipelines to lighten redundancy and reduce computational complexity.

2.2.4 Sequence generation via sliding window

In the Transformer models, the normalized feature matrix is segmented into overlapping sliding windows (e.g., 50 timesteps per window). This transformation converts the static dataset into a sequence-based format suitable for models that capture temporal dependencies. In contrast, while the ML model primarily operates on tabular data, sliding window sequences may also be employed to frame forecasting as a regression problem.

2.3 Data augmentation

Data augmentation strategies are applied to enhance model robustness and reduce overfitting, particularly in scenarios with noisy or limited training data.

2.3.1 Noise injection

Gaussian noise, characterized by a zero mean and a variance of 0.01, was added to the training data. This perturbation simulates slight variations in the input features, thereby enabling the network to learn more robust representations. The addition of noise effectively emulates the inherent variability in real-world sensor data and enhances the resilience of the models to minor fluctuations.

2.3.2 Variational mode decomposition (VMD) for feature enhancement

Variational mode decomposition (VMD) enhances the input feature space by breaking down the wind speed signal into multiple intrinsic mode functions (IMFs), each representing different frequency components. This allows the model to capture both short-term variations and long-term patterns. By appending these IMFs to the original features, we provide richer temporal information, enabling the models to better learn complex, multi-scale dynamics in wind power data. Unique to the ML preprocessing pipeline, where VMD decomposes the wind speed feature into intrinsic mode functions, this helps in extracting multiple frequency components to capture hidden dynamic patterns [34]. The resulting VMD modes are then concatenated with the original features to form an enhanced feature set. This enriched representation enables the ML model to better capture the multi-scale behavior of the time series data. This significantly improves the forecasting performance, particularly for ensemble models like XGBoost and LightGBM.

2.4 Model implementation

2.4.1 Machine learning models

In this research, three ensemble machine learning models, Random Forest, XGBoost, and LightGBM in Fig. 3, are employed for short-term wind power forecasting using normalized inputs and VMD preprocessing technique [35]. Random Forest uses a bagging approach, averaging predictions from independent decision trees to reduce variance. XGBoost and LightGBM, based on gradient boosting frameworks, optimize a regularized objective function to capture complex nonlinear relationships in wind power data.

$$\hat{y} = \frac{1}{T} \sum_{i=1}^{T} h_i(x), \tag{1}$$

where $h_i(x)$ represents the prediction from the i-th tree. This aggregation reduces variance and enhances robustness. In contrast, XGBoost and LightGBM are based on gradient boosting frameworks that optimize a regularized objective function:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{2}$$

where $l(y_i, \hat{y}_i)$ denotes the loss function, and $\Omega(f_k)$ controls the complexity of the k-th tree to prevent overfitting. XGBoost employs second-order Taylor expansion by incorporating both the gradient (g_i) and Hessian (h_i) information to approximate and optimize the objective efficiently:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t), \tag{3}$$

where f_t is the model added at the t-th iteration. LightGBM improves training efficiency by utilizing histogram-based feature binning and growing trees leaf-wise rather

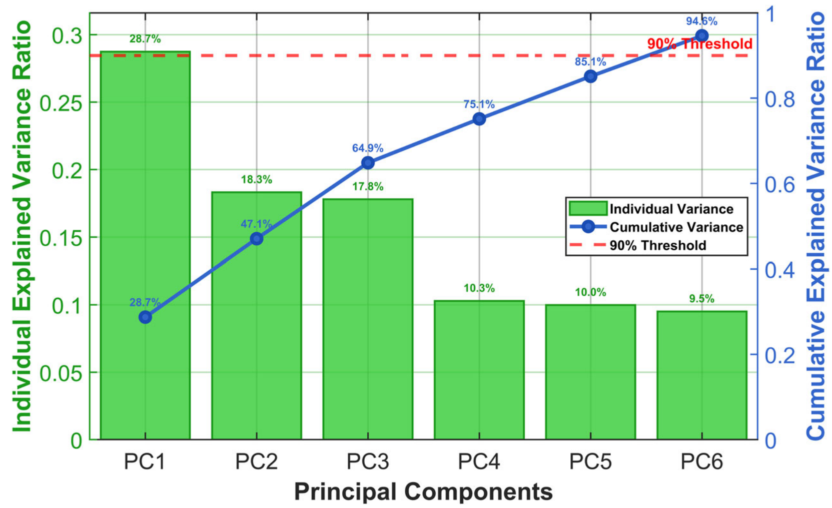


Fig. 2 Dimensionality reduction using PCA

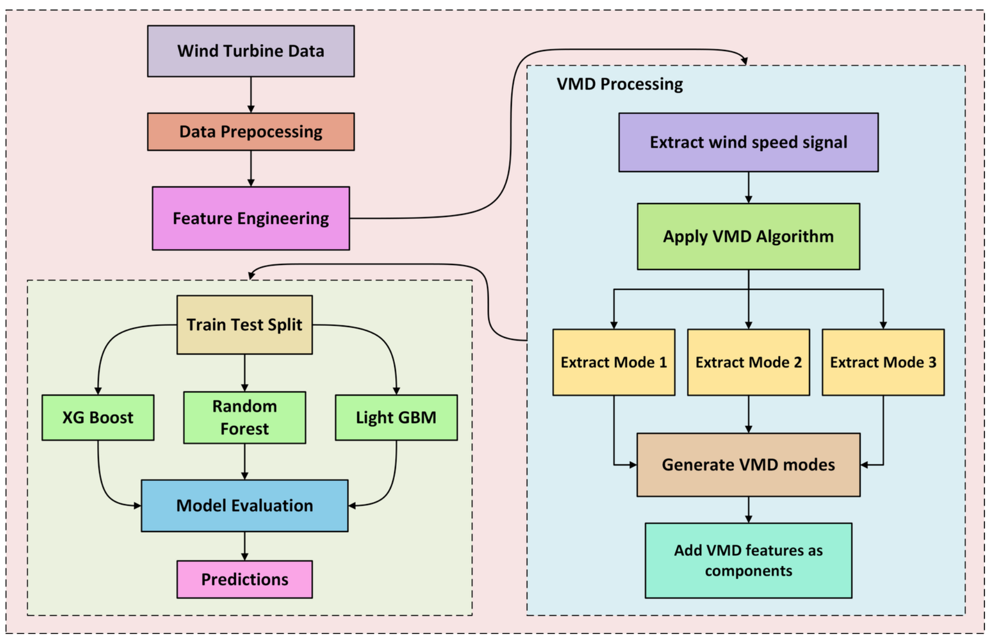


Fig. 3 Workflow of Wind Power Forecasting Using Ensemble ML Models and VMD Processing

than level-wise, selecting the leaf with the maximum loss reduction. Despite sharing a similar objective function with XGBoost, LightGBM offers enhanced speed and scalability, especially on large datasets. Experimental results revealed that XGBoost and LightGBM achieved superior forecasting accuracy and computational efficiency compared to Random Forest, demonstrating their effectiveness in modeling complex nonlinear relationships inherent in wind power generation data.

2.4.2 Deep learning transformer models

The DL-based pipelines, including the Vanilla Transformer, Informer, Autoformer, and Temporal Fusion Transformer (TFT) in Fig. 4, show the models' architecture. These share a common preprocessing strategy involving temporal alignment, cyclic feature encoding, lag feature incorporation, data normalization via StandardScaler and dimensionality reduction using PCA to improve learning efficiency.

The Vanilla Transformer model processes an input sequence $X = [x_1, x_2, \ldots, x_L]$ by first projecting each $x_i \in \mathbb{R}^{d_{\text{model}}}$ into an embedding space and adding a positional encoding

$$\text{PE}_{(pos,2i)} = \sin\left(pos/10000^{2i/d_{\text{model}}}\right), \tag{4}$$

$$\text{PE}_{(pos,2i+1)} = \cos\left(pos/10000^{2i/d_{\text{model}}}\right), \tag{5}$$

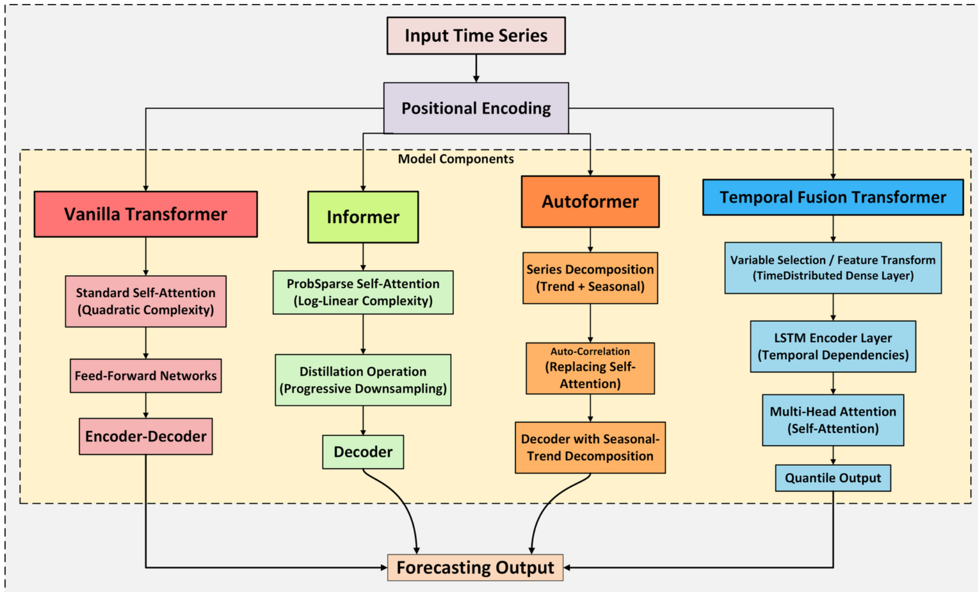


Fig. 4 Comparative Architecture of Transformer-Based Models for Wind Power Forecasting

to retain order information [36]. It then applies multi-head self-attention

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\big(QK^\top/\sqrt{d_k}\big)V \tag{6}$$

with h parallel heads (each of dimension $d_k = d_{\mathrm{model}}/h$), concatenating their outputs and projecting back via W^O. A position-wise feed-forward network

$$\mathrm{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{7}$$

follows, and residual connections with layer normalization ($\mathrm{LN}(x)$) are used around both sub-layers. Dropout (rate p) is applied to attention weights and FFN outputs. Hyperparameters such as $h = 4$, $d_{\mathrm{ff}} = 128$, and dropout $= 0.1$ are chosen empirically to balance learning capacity and regularization.

The Informer extends this backbone by adopting ProbSparse self-attention, which reduces complexity from $O(L^2)$ to $O(L \log L)$ by selecting only the top-u queries (where $u \approx L \ln L$) with highest Kullback–Leibler divergence from a uniform distribution. A convolutional distillation module then applies a 1D convolution (kernel size = 3, stride = 2) to compress the sequence length, removing redundancy before subsequent encoder layers. Finally, the generative decoder predicts all T' future steps in a single forward pass, avoiding error propagation inherent in auto-regressive decoding.

The Autoformer integrates a series decomposition block into each encoder-decoder layer, splitting the input X into trend $T = \mathrm{MA}(X)$ (via moving average) and seasonal $S = X - T$ components. Each component is processed separately through an auto-correlation attention mechanism that identifies periodic lags and captures seasonality explicitly. By stacking multiple decomposition + attention layers, the Autoformer effectively models both long-term trends and repeating patterns, improving interpretability and accuracy for data with strong periodic behaviors.

The Temporal Fusion Transformer (TFT) combines recurrent and attention mechanisms with dynamic feature selection. At each timestep t, its variable selection network computes a weight vector $\alpha_t = \mathrm{softmax}(W_{\mathrm{vs}}\, x_t + b_{\mathrm{vs}})$ to choose among static, known, and unknown inputs. A bidirectional LSTM encoder–decoder

$$h_t = \mathrm{LSTMCell}(x_t, h_{t-1}) \tag{8}$$

captures sequential dynamics, after which an interpretable multi-head attention layer produces

$$y_t = g_t \odot \mathrm{Attention}(Q, K, V) + (1 - g_t) \odot h_t, \tag{9}$$

$$g_t = \sigma(W_g\, h_t + b_g), \tag{10}$$

blending attention outputs with recurrent states via a gating mechanism. Gated residual connections and layer normalization throughout ensure stable training, while this hybrid design delivers both high forecasting accuracy and model interpretability.

2.5 Evaluation of the model

The evaluation of the proposed models was carried out using standard metrics that quantify the prediction error. The following metrics were computed based on the ground truth (y_i) and the predicted values ($\hat{y}_i$).

Coefficient of determination (R^2)

The coefficient of determination R^2 measures the proportion of variance in the dependent variable that is predictable from the independent variables. It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} \tag{11}$$

where $\bar{y}$ is the mean of the observed values and n is the number of observations. A value of R^2 closer to 1 indicates a better fit. The code implementations calculate R^2 using standard libraries, ensuring consistency across model evaluations.

Mean squared error (MSE)

It is widely adopted to capture the average squared difference between the actual and predicted values:

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

The use of MSE in the code provides a measure of the overall error magnitude, giving higher weight to larger errors due to the squaring term.

Root-mean-squared error (RMSE)

RMSE, the square root of the MSE, has the same units as the target variable and is given by:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2} \tag{13}$$

This metric provides a direct measure of the average error magnitude, making it easier to interpret in the context of the problem.

Mean absolute error (MAE)

MAE quantifies the average absolute difference between the predicted and actual values and is expressed as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left|y_i - \hat{y}_i\right| \tag{14}$$

Being less sensitive to outliers compared to MSE, MAE is particularly useful for obtaining a balanced measure of error magnitude across different scales of prediction.

3 Results and discussion

The anomalies in wind power data could indicate sensor malfunctions, such as recording incorrect wind speeds or power values; power system defects, like turbine failure or unexpected energy loss; or external conditions, like poor weather, maintenance activities, or sudden wind direction changes. These are identified to ensure dataset integrity.

3.1 Anomaly detection

The anomaly detection approach depicted in Fig. 5 employs an LSTM autoencoder, where the raw wind power data are initially normalized using a RobustScaler to reduce the impact of outliers. Time-window sequences are constructed to maintain the temporal structure of the data, and the LSTM autoencoder is trained on these sequences to learn a compact representation of typical operating conditions. It calculates the reconstruction error for each time frame by calculating the difference between the original input sequence and the sequence reconstructed by the autoencoder, typically using a measure such as mean squared error. The reconstruction error for each sequence is computed as the mean squared error (MSE):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \tag{7}$$

where x_i is the original input value and $\hat{x}_i$ is the reconstructed value at position i, with n being the sequence length.

A dynamic threshold is established to detect anomalies, set at the 95th percentile of the MSE values observed in the training data. Sequences with an MSE exceeding this threshold are classified as anomalous. This approach effectively captures significant deviations from normal operational patterns, ensuring the reliability of the dataset for subsequent forecasting tasks. This method effectively identifies significant deviations from normal behavior, leading to detection of anomalies in wind power data as shown in Fig. 6.

The physics-informed neural network (PINN) model uses physical constraints directly into its training process, complementing anomaly detection [37]. The PINN incorporates dropout and L2 regularization to mitigate overfitting, to improve generalization on unseen data. A custom physics loss is added into the network's objective function, which enforces Betz' law, a principle stating that the maximum extractable power from wind is defined by:

$$P_{\max} = \frac{16}{27} \cdot \frac{1}{2} \rho A v^3, \tag{15}$$

where ρ represents air density, A implies the rotor area, and v indicates the wind speed. This constraint ensures that

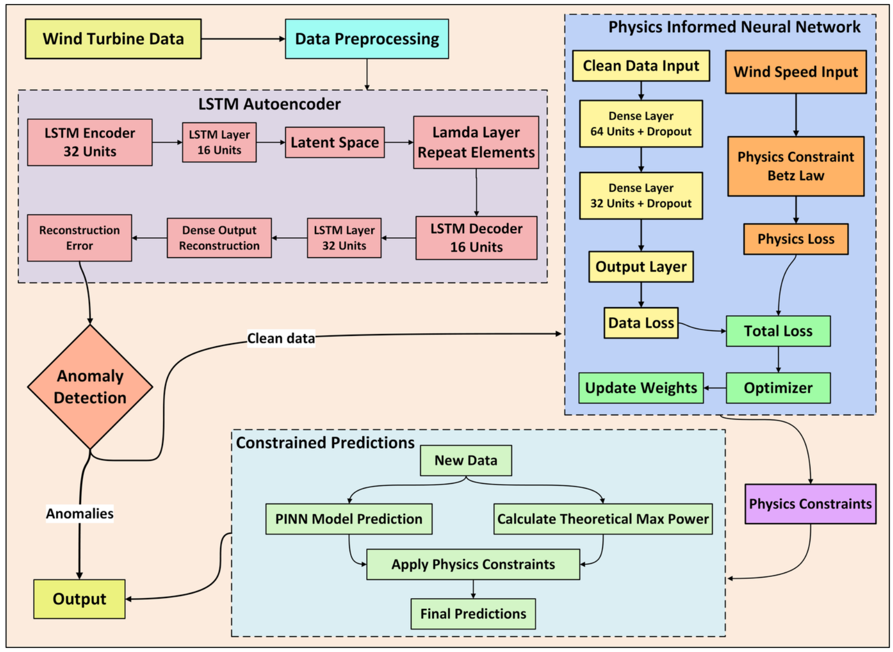


Fig. 5 Model for Detecting Anomalies in Wind Power Time Series

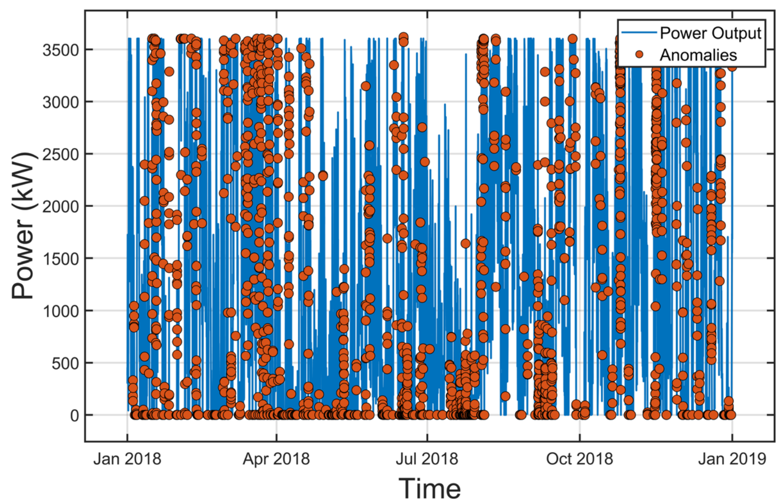


Fig. 6 Wind Power Output with Detected Anomalies

the forecasted wind power values remain within possible physical limits. To enforce Betz' law within the PINN framework, a physics-based loss term that regulates predictions exceeding the theoretical wind power limit was designed. Physics loss is mean squared error between predictions and theoretical power. The total loss function combines data fidelity and physical consistency:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda \, \mathcal{L}_{\text{physics}} \tag{16}$$

where

$$\mathcal{L}_{\text{physics}} = \frac{1}{N} \sum_{i=1}^{N} \max\bigl(0, \; P_{\text{pred},i} - P_{\max,i}\bigr)^2. \tag{17}$$

The weight λ (0.1) of the physics loss is tunable and can be adjusted to balance the influence of physical constraints. Early stopping is employed to optimize the training process, which prevents overfitting and ensures smooth convergence. A validation step follows prediction to ensure that the predicted values fall within the limits of Betz' law so that the model's reliability is improved.

LSTM autoencoder detected 2,533 anomalies (5% of data). The majority of the data points lie far below the theoretical power curve, indicating less-than-expected power production due to probable inefficiencies like environmental conditions (e.g., turbulence, air density fluctuations), operational restrictions (e.g., mechanical malfunction or suboptimal turbine settings), sensor miscalculations. The presence of anomalies suggests that pure data-driven models could be misleading without incorporating physical constraints (such

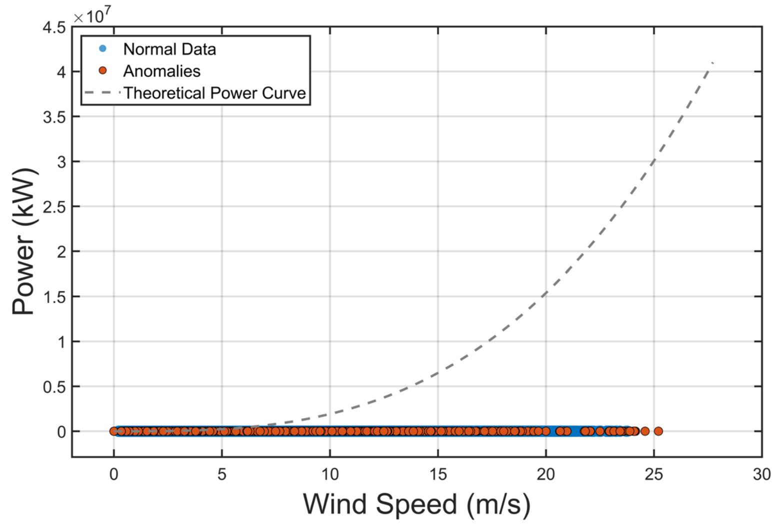


Fig. 7 Wind Speed vs Power Output with Anomalies

as Betz' law) into the learning process. The PINN is important in preserving the physically realistic predictions unlike the traditional machine learning models that may blindly fit observed trends without considering fundamental wind power physics. Figure 7 illustrates that combination of data-driven learning (like LSTMs) with physics-informed models (like PINN) is essential for making accurate and physically meaningful predictions in wind power forecasting. The results show that the threshold-based anomaly detection technique correctly identifies anomalous behavior and that the addition of physical constraints ensures that the prediction from the model stays within reasonable aerodynamic limits. Figure 8 helps to understand the underlying patterns, cyclical components, and irregularities (residuals) that might indicate anomalies or external effects.

The cleaned dataset, free of anomalies, is split into are evaluated using standard 80:20 train–test split and used to train the AI-driven forecasting model. This ensures the model learns from representative data, enhancing prediction accuracy for short-term wind power forecasts. After assessing the models that are listed earlier, the experimental details are discussed below. Historical wind power and meteorological data from a particular wind farm are used in the experiments. Techniques such as early stopping and learning rate scheduling are employed to mitigate overfitting and enhance generalization. Performance metrics, including mean squared error (MSE), root-mean-squared error (RMSE), mean absolute error (MAE), and R^2 score, are presented in the table below. These metrics as depicted in Table 1 provide a comprehensive evaluation of the models' predictive accuracy and robustness. The stable and consistent range of performance metrics are observed for a confidence interval of 95%, thereby ensuring the robust predictions by the developed models and not heavily influenced due to data variations.

3.2 Learning curve

The learning curves for each model provide insights into the training and validation process, highlighting the convergence behavior of the developed models and performance stability over epochs. The learning curves offer key information on their training dynamics and generalizability.

The Vanilla Transformer Fig. 9 exhibits a steep decline in training and validation losses during initial epochs, followed by stabilization. This indicates rapid learning and effective convergence. The close alignment of training and validation losses indicates that the model generalizes well to unseen data with strong generalization, minimal overfitting, aided by early stopping and learning rate scheduling.

The Informer model Fig. 10 shows a similar pattern of rapid initial learning and gradual stabilization. The slight gap between training and validation losses indicates a balanced trade-off between underfitting and overfitting, demonstrating robust performance on unseen data. The initial few epochs show significant improvements, highlighting the model's ability to quickly capture relevant features.

The Autoformer's learning curve Fig. 11 displays a sharp initial decrease in losses, followed by stabilization. The close proximity of training and validation losses throughout training suggests strong generalization ability. Noise augmentation and regularization techniques help prevent overfitting, ensuring reliable performance.

The TFT model's learning curve Fig. 12 shows rapid initial improvements and eventual stabilization, indicating effective learning and convergence. The alignment of training and validation losses suggests good generalization, with early stopping and learning rate scheduling prevents overfitting.

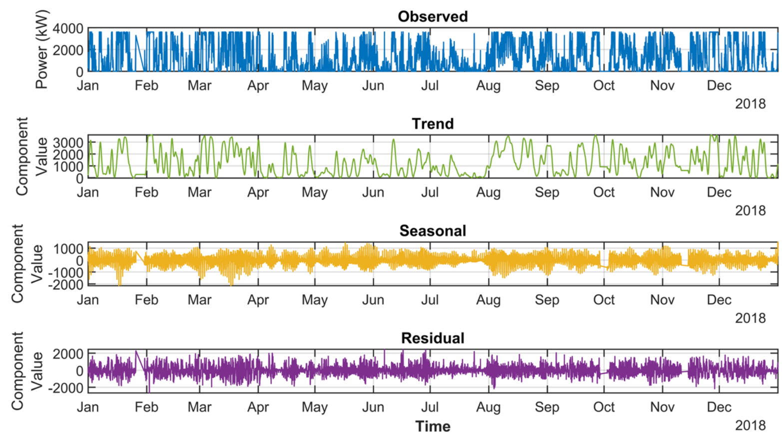


Fig. 8 Seasonal-Trend decomposition using Loess (STL) separating the time series into observed signal, trend, seasonal, and residual components

Table 1 Wind Power Forecasting

Model	R^2	MSE	RMSE	MAE
XGBoost	0.9890	0.0015	0.0385	0.0151
LightGBM	0.9799	0.0026	0.0514	0.0233
Temporal Fusion Transformer	0.9649	0.0349	0.1868	0.1109
Random Forest	0.9651	0.0045	0.0674	0.0281
Informer	0.9483	0.0067	0.0821	0.0519
Vanilla Transformer	0.9484	0.0078	0.0881	0.0640
Autoformer	0.9231	0.0666	0.2760	0.1657

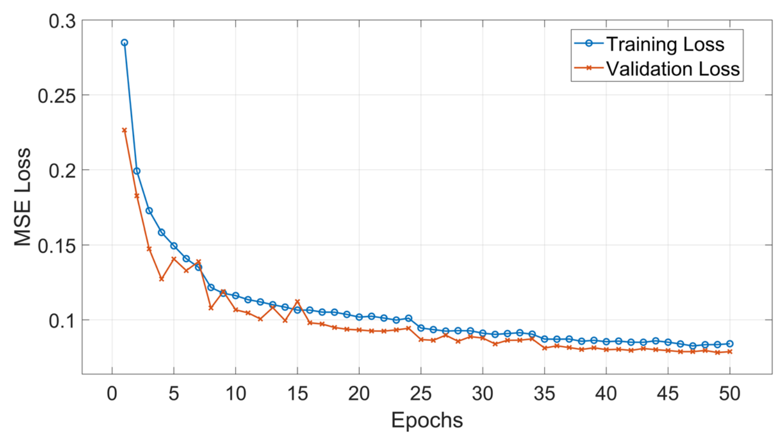


Fig. 9 Learning Curve for Vanilla Transformer Model

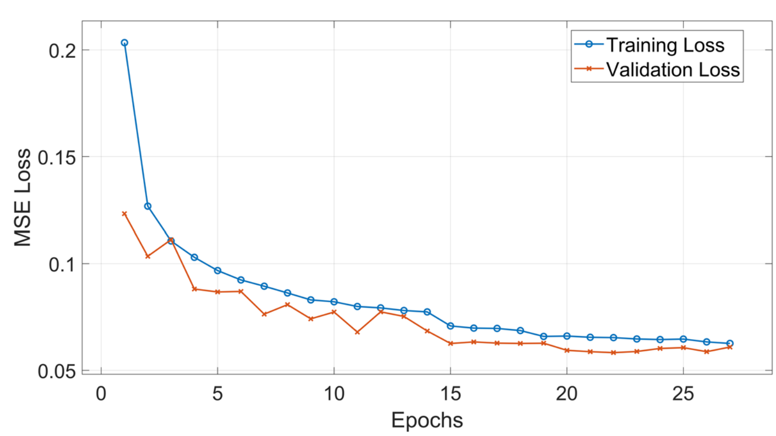


Fig. 10 Learning Curve for Informer Model

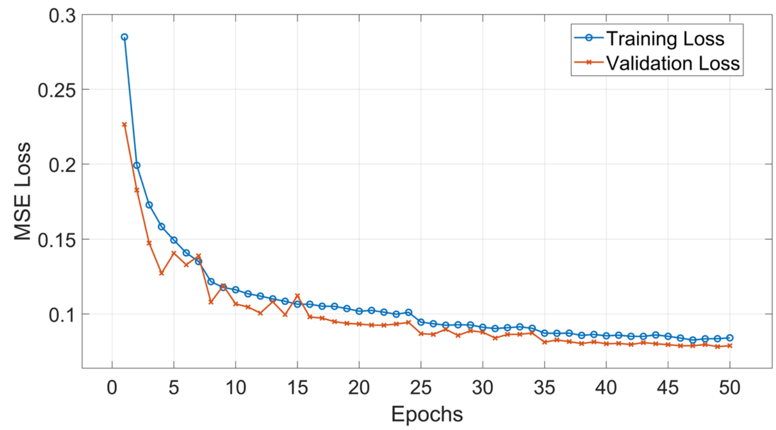


Fig. 11 Learning Curve for Autoformer Model

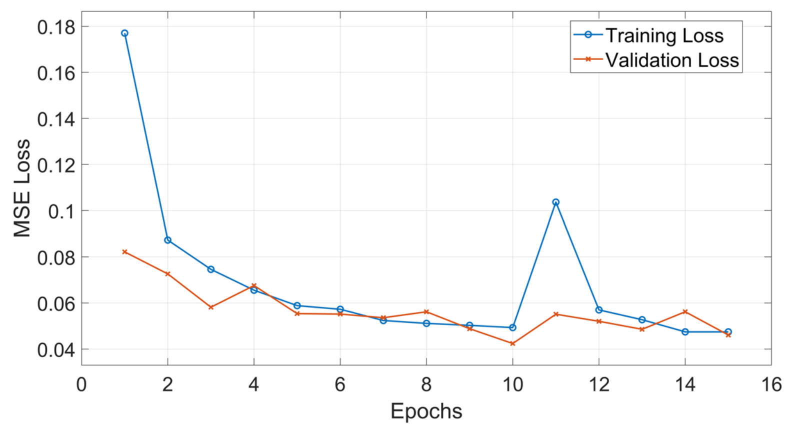


Fig. 12 Learning Curve for Temporal Fusion Transformer Model

3.3 Prediction vs true values

In this subsection, the predicted wind power values against the actual observed values are compared to evaluate the accuracy of each model. For a better visualization, a subset of the test dataset, around 1000 samples, are plotted for good visualization. Figure 13 to Fig. 19 show the prediction versus true values plots for each of the models.

From the LightGBM model Fig. 13, it generally tracks the actual wind power values with decent accuracy despite the spiky and discontinuous nature of the dataset.

Similarly, the Random Forest model Fig. 14 produces stable predictions that align well with actual values, although it appears to slightly underfit in highly variable regions due to its ensemble averaging nature.

On the other hand, the XGBoost model Fig. 15 demonstrates the most reliable prediction performance, following the actual wind power variations closely with fewer deviations. All three ML models show capability in capturing the wind power generation pattern, but their performance varies in precision and handling of rapid fluctuations. Among them, XGBoost stands out as the best-performing model, achieving the closest match between predicted and actual values across the entire sample set. LightGBM provides a strong balance of speed and accuracy but occasionally shows local mismatches. Random Forest offers consistent predictions with smooth transitions, making it a reliable choice but slightly less accurate in volatile segments.

The Vanilla Transformer model Fig. 16 demonstrates a high level of consistency between predicted and true values, as evident in the close alignment of the orange (predictions) and blue (true values) lines in the plot. Preprocessing steps like comprising cyclic encoding, lag feature generation, normalization, and PCA have enabled the model to capture essential temporal dependencies. The proximity of the predictions to the actual values indicates a high degree of accuracy and reliability. Despite the fluctuating nature of the true values, the model maintains good precision.

The Informer model Fig. 17 exhibits a strong alignment between predicted and true values, with the orange and blue lines closely following each other throughout the plot. The incorporation of convolutional distillation within the architecture, combined with robust preprocessing steps, enhances its ability to manage long-range dependencies. Consequently, the observed predictions exhibit minimal fluctuation relative to the true values, underscoring the Informer's aptitude to extract and retain critical temporal features with high precision.

The Autoformer model Fig. 18 shows a remarkable closeness between predicted and true values, as seen in the

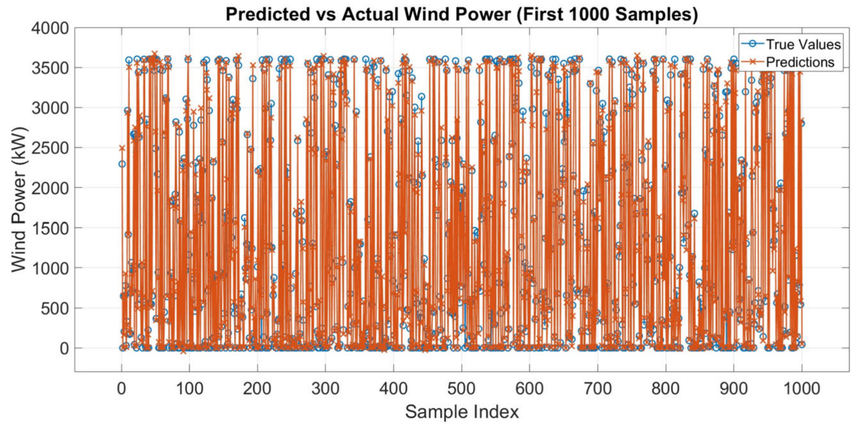


Fig. 13 Actual vs Predicted Values for Light GBM model

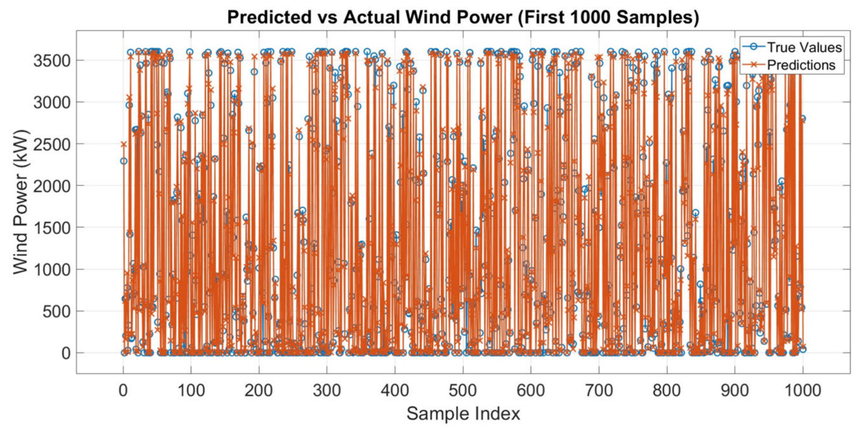


Fig. 14 Actual vs Predicted Values for Random Forest model

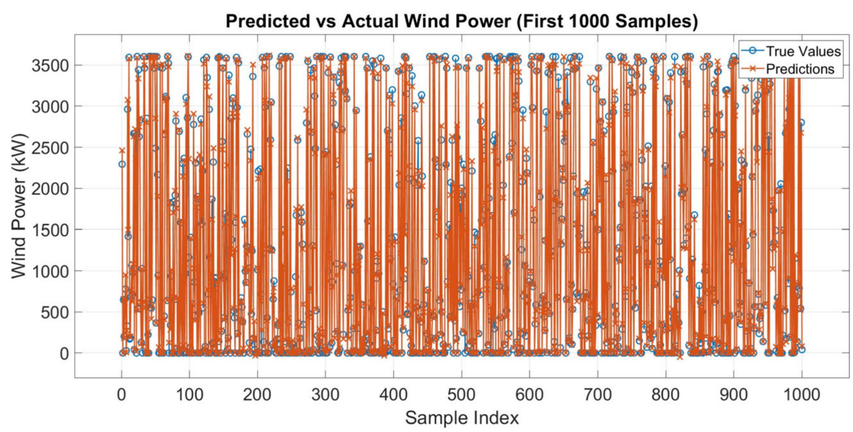


Fig. 15 Actual vs Predicted Values for XGBoost model

overlapping orange and blue lines in the plot. By leveraging series decomposition to separate trend and seasonal components, the Autoformer effectively isolates and models complex underlying patterns. This leads to a high correlation between predicted and actual values, even in the presence of data variability. The auto-regressive learning capability, coupled with effective preprocessing, results in forecasts that align closely with the real observations, confirming the model's reliability.

The TFT model Fig. 19 shows a high level of consistency between predicted and true values, indicating accurate and reliable forecasts. Normalization and noise augmentation enhance its learning capability. It is enhanced through variable selection and an LSTM-based encoder, and the TFT benefits from both dynamic feature extraction and attentive mechanisms. The preprocessing steps provide it with a well-calibrated temporal feature vector, which in turn results in highly precise predictions that mirror the actual time series data. Overall, the TFT demonstrates robust performance with minimal prediction error across varying temporal segments true values.

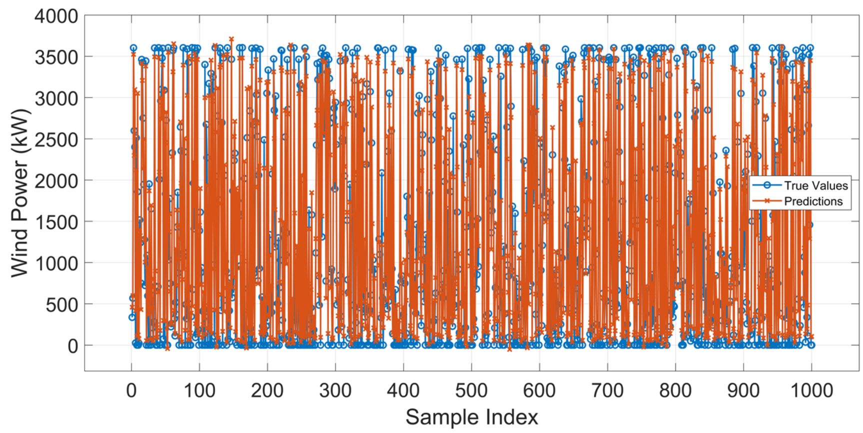


Fig. 16 Actual vs Predicted Values for Vanilla Transformer Model

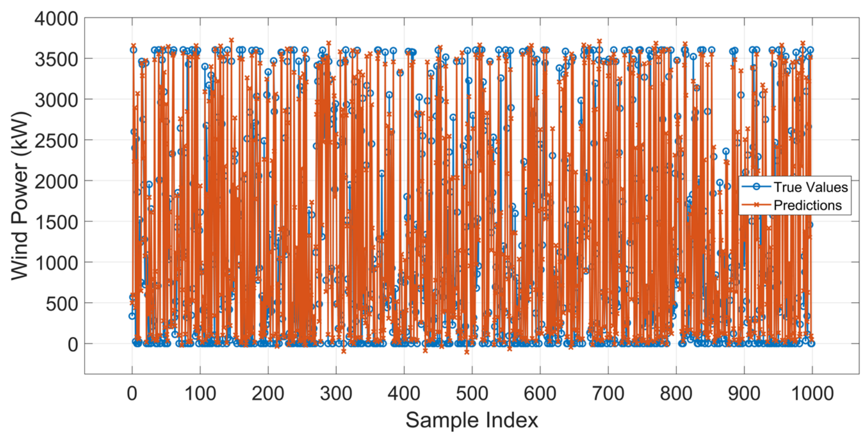


Fig. 17 Actual vs Predicted Values for Informer Model

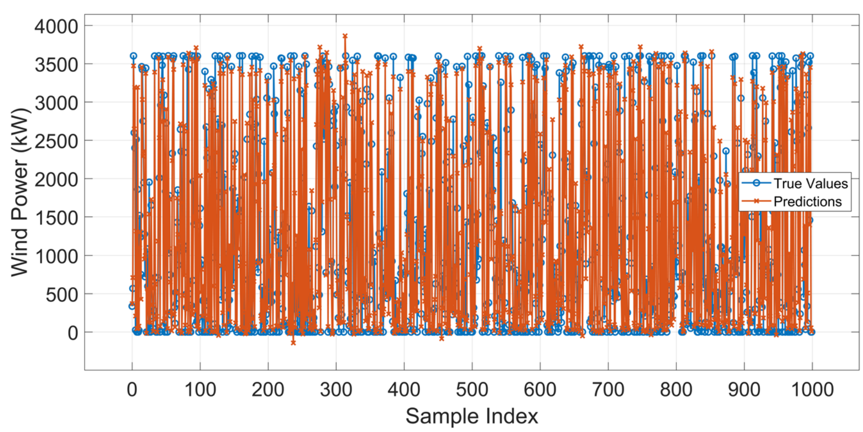


Fig. 18 Actual vs Predicted Values for Autoformer Model

4 Conclusion

This research introduced an AI-augmented methodology for short-term wind power forecasting, expanding upon previous foundational studies in the domain. This study captures the nonlinear temporal dynamics of wind power generation by using ensemble machine learning models (XGBoost, LightGBM, Random Forest) and deep learning architectures (Vanilla Transformer, Informer, Autoformer, TFT). The proposed strategy integrated preprocessing techniques, namely VMD, Gaussian noise augmentation, and PCA which significantly enhanced feature quality and model resilience. Moreover, anomaly detection with LSTM autoencoders and PINNs guaranteed data integrity and observation to physical limitations such as Betz' law. The learning curves provide significant information about DL model training and validation proves that the model works with on different datasets.

Among the models, XGBoost exhibited the highest performance, attaining a R^2 score of 0.9890, MSE of 0.0015, RMSE of 0.0385, MAE of 0.0151, illustrating its generalization capabilities on high-frequency SCADA data. The strong correlation between predicted and actual values further validates

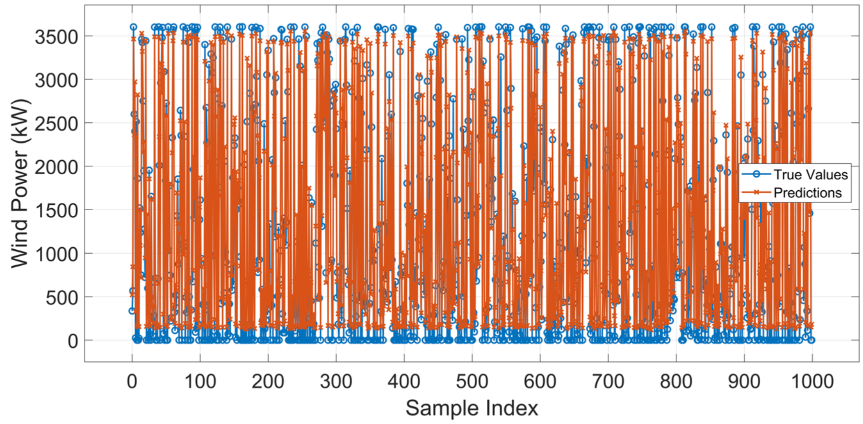


Fig. 19 Actual vs Predicted Values for Temporal Fusion Transformer Model

the reliability of the hybrid approach. Therefore, XGBoost is recommended for wind power forecasting tasks where higher precision is essential. This work confirms that combining ensemble methods and attention-based structures together with preprocessing and domain-specific constraints produces accurate and physically reasonable forecasts. Future directions involve integrating satellite data and geographical characteristics to improve predictions and increase the framework's scalability across various wind farm topographies. The integration of satellite photographs, atmospheric pressure measurements, and geographic data is essential for improving the precision of wind electricity forecasting models. Automatic function planning with statistics processing can enhance forecasting accuracy and speed up model education.

References

1. Vishnutheerth E, Vijay V, Satheesh R, Kolhe M (2024) A comprehensive approach to wind power forecasting using advanced hybrid neural networks. IEEE Access
2. Jency W, Judith J (2023) Performance analysis on wind power forecasting in machine learning and deep learning. In: 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), pp. 592–596. IEEE
3. Kiasari MM, Aly HH (2024) Enhancing wind power forecasting accuracy in canada using a solar data-enhanced hybrid machine learning model: Integrating ann, lstm, and svr. In: 2024 IEEE 12th International Conference on Smart Energy Grid Engineering (SEGE), pp. 189–194. IEEE
4. Khan M, Naeem MR, Al-Ammar EA, Ko W, Vettikalladi H, Ahmad I (2022) Power forecasting of regional wind farms via variational auto-encoder and deep hybrid transfer learning. Electronics 11(2):206
5. Hu M, Gao X, Duan J, Shu C, Sun B, Zhang J, Song D, Chen B (2024) Wind power forecasting based on sparrow search algorithm and kernel-based extreme learning machine. In: 2024 8th International Conference on Power Energy Systems and Applications (ICoPESA), pp. 575–579. IEEE
6. Kalele G (2023) Hybrid machine learning approach for wind power forecasting. In: 2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC), pp. 70–73. IEEE
7. Finamore AR, Calderaro V, Galdi V, Graber G, Ippolito L, Conio G (2023) Improving wind power generation forecasts: a hybrid ann-clustering-pso approach. Energies 16(22):7522
8. Bali V, Kumar A, Gangwar S (2019) Deep learning based wind speed forecasting-a review. In: 2019 9th International Conference on Cloud Computing, Data Science & Engineering (confluence), pp. 426–431. IEEE
9. Wang Q, Wang Y, Zhang K, Liu Y, Qiang W, Han Wen Q (2023) Artificial intelligent power forecasting for wind farm based on multi-source data fusion. Processes 11(5):1429
10. Karaman ÖA (2023) Prediction of wind power with machine learning models. Appl Sci 13(20):11455
11. Belletreche M, Bailek N, Abotaleb M, Bouchouicha K, Zerouali B, Guermoui M, Kuriqi A, Alharbi AH, Khafaga DS, El-Shimy M et al (2024) Hybrid attention-based deep neural networks for short-term wind power forecasting using meteorological data in desert regions. Sci Rep 14(1):21842
12. Sulaiman MH, Mustaffa Z (2024) Enhancing wind power forecasting accuracy with hybrid deep learning and teaching-learning-based optimization. Cleaner Energy Systems 9:100139
13. Ma Z, Mei G (2022) A hybrid attention-based deep learning approach for wind power prediction. Appl Energy 323:119608
14. Huang B, Liang Y, Qiu X (2021) Wind power forecasting using attention-based recurrent neural networks: a comparative study. IEEE Access 9:40432–40444
15. Chen J, Zeng G-Q, Zhou W, Du W, Lu K-D (2018) Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy Convers Manage 165:681–695
16. Chen X, Wang Y, Zhang H, Wang J (2024) A novel hybrid forecasting model with feature selection and deep learning for wind speed research. J Forecast 43(5):1682–1705
17. Wu Q, Guan F, Lv C, Huang Y (2021) Ultra-short-term multi-step wind power forecasting based on CNN-LSTM. IET Renew Power Gener 15(5):1019–1029
18. Zhang Y, Zhao Y, Gao S (2019) A novel hybrid model for wind speed prediction based on VMD and neural network considering atmospheric uncertainties. IEEE Access 7:60322–60332
19. Wu Z, Luo G, Yang Z, Guo Y, Li K, Xue Y (2022) A comprehensive review on deep learning approaches in wind forecasting applications. CAAI Trans Intell Technol 7(2):129–143
20. Tian Y, Wang D, Zhou G, Wang J, Zhao S, Ni Y (2023) An adaptive hybrid model for wind power prediction based on the ivmd-fe-ad-informer. Entropy 25(4):647
21. Ahmed U, Muhammad R, Abbas SS, Aziz I, Mahmood A (2024) Short-term wind power forecasting using integrated boosting approach. Front Energy Res 12:1401978

22. Yu G, Shen L, Dong Q, Cui G, Wang S, Xin D, Chen X, Lu W (2024) Ultra-short-term wind power forecasting techniques: comparative analysis and future trends. Front Energy Res 11:1345004
23. Zhu M, Li Z, Lin Q, Ding L (2025) Fast-powerformer: A memory-efficient transformer for accurate mid-term wind power forecasting. arXiv preprint arXiv:2504.10923
24. Wang H-K, Song K, Cheng Y (2022) A hybrid forecasting model based on CNN and informer for short-term wind power. Front Energy Res 9:788320
25. Perr-Sauer J, Tripp C, Optis M, King J (2020) Short-term wind forecasting using statistical models with a fully observable wind flow. In: Journal of Physics: Conference Series, vol. 1452, p. 012083. IOP Publishing
26. Zhang W, Yan H, Xiang L, Shao L (2025) Wind power generation prediction using lSTM model optimized by sparrow search algorithm and firefly algorithm. Energy Inf 8(1):35
27. Singh U, Rizwan M (2021) A systematic review on selected applications and approaches of wind energy forecasting and integration. J Institut Eng Series B 102(5):1061–1078
28. Badjan A, Rashed GI, Gony HAI, Haider H, Bahageel AO, Shaheen HI (2025) Improving short-term wind power forecasting in Senegal's flagship wind farm: a deep learning approach with attention mechanism. Electr Eng 107(3):3307–3321
29. Sireesha PV, Thotakura S (2024) Wind power prediction using optimized mlp-nn machine learning forecasting model. Electrical Engineering, 1–24
30. Jin F, Jiang C, Wen X (2025) Wind power scenario forecasting based on combination of generative adversarial network and long short-term memory network. Electrical Engineering, 1–11
31. Arati DC, Menon PS, Velayudhan J, Poornachandran P, Raj AK, OK S (2024) Enhancing wind power prediction through machine learning. In: 2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP), pp. 1–5. IEEE
32. İsen B (2018) Wind Turbine SCADA Dataset. Accessed: 2025-06-19. https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset
33. Tarek Z, Shams MY, Elshewey AM, El-kenawy E-SM, Ibrahim A, Abdelhamid AA, El-dosuky MA (2023) Wind power prediction based on machine learning and deep learning models. Computers, Materials & Continua **75**(1)
34. Zhao Z, Yun S, Jia L, Guo J, Meng Y, He N, Li X, Shi J, Yang L (2023) Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. Eng Appl Artif Intell 121:105982
35. Alkesaiberi A, Harrou F, Sun Y (2022) Efficient wind power prediction using machine learning methods: a comparative study. Energies 15(7):2327
36. Ayene SM, Yibre AM (2024) Wind power prediction based on deep learning models: The case of adama wind farm. Heliyon **10**(21)
37. Wang H, Yang Z, Kang W, Sun P, Konstantinou G, Chen Z (2024) Physics-informed learning based wind farm two-machine aggregation model for large-scale power system stability studies. IEEE Transactions on Power Systems