IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 July 2024, accepted 19 August 2024, date of publication 26 August 2024, date of current version 16 September 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3450096*

RESEARCH ARTICLE

# A Comprehensive Approach to Wind Power Forecasting Using Advanced Hybrid Neural Networks

**E. P. VISHNUTHEERTH[1], VIVEK VIJAY[1], RAHUL SATHEESH[1], (Member, IEEE), AND MOHAN LAL KOLHE[2]**

[1]Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India
[2]Department of Engineering Sciences, University of Agder, 4604 Kristiansand, Norway

Corresponding authors: Rahul Satheesh (s_rahul1@cb.amrita.edu) and Mohan Lal Kolhe (mohan.l.kolhe@uia.no)

This work was supported by the University of Agder, Norway.

**ABSTRACT** Wind power prediction is important in successfully integrating renewable energy sources into the grid. This study is focused on a sub-domain of wind power prediction and compares Bidirectional Long Short Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU) architectures. Additionally, these models are enhanced by advanced pre-processing techniques, including such methods as Discrete Wavelet Transform (DWT) and Fourier Synchrosqueezed Transform (FSST), as well as hybrid models involving Convolutional Neural Network (CNN) and Random Forest (RF) together with BiLSTM and BiGRU Models. It was found that the hybrid model consisting of CNN and BiGRU performed better than other hybrids by returning an $R^2$ score of 0.9093, RMSE of 0.1095, MSE of 0.0120 and MAE of 0.0466; this shows that our model had a much greater level of accuracy compared to others ones developed before. These model performance indices demonstrated its better trustworthiness and error level for further utilization in wind energy forecast applications required for efficiency improvements and reliability enhancement in wind energy management. The current study emphasizes the usefulness of combining deep learning approaches like BiLSTM and BiGRU for more accurate wind power predictions hence improving reliability and effectiveness in managing wind energy resources.

**INDEX TERMS** Bidirectional long short term memory (BiLSTM), bidirectional gated recurrent unit (BiGRU), convolutional neural networks (CNN), Discrete Wavelet Transform (DWT), hybrid deep learning, wind power.

## I. INTRODUCTION

Wind energy has been considered very essential in the renewable energy system owing to its capability of sparing the environment with little impact. However, achieving high values for wind power generation remains a difficult task because it is as unpredictable as the wind patterns. Hence, accurate wind power forecasting [1] is important to adequately balance power supply and demand in the power grid, optimize energy dispatch, and reduce reliance on power generated from fossil fuels. Consequently, constructing effective and reliable predictive models for energy supply becomes crucial. Traditional methods are inadequate in describing temporal patterns prevalent in wind power data [2]. Statistical models and physical-based methods, such as ARIMA (Auto-Regressive Integrated Moving Average), often fail to respond to sudden changes in wind patterns, leading to large prediction errors. To address this issue, deep learning algorithms such as LSTM (Long short Term Memory) and GRU (Gated Recurrent Unit) are employed, which can learn from sequential data. These models are designed to work best for time series data; therefore, they are suitable for wind power forecasting. By extending these architectures to bidirectional versions, it is possible to achieve higher prediction accuracy because BiLSTM and BiGRU take

The associate editor coordinating the review of this manuscript and approving it for publication was Jiaxing Lei.

into account the context from both preceding and subsequent time steps [3].

Studies on the use of blended deep learning models for wind power prediction are very encouraging. It has been observed that by incorporating features of several forms of preprocessing and/or modeling, these kinds of models can substantially deliver top performance. In this study, both BiLSTM and BiGRU are integrated with recent preprocessing procedures like DWT (Discrete Wavelet Transform) [4] and FSST (Fourier Synchrosqueezed Transform) [5], and also integrated with other deep learning and machine learning techniques such as CNN (Convolutional Neural Network) and RF (Random Forest) respectively. The preprocessing algorithms help to split the wind power time series into more manageable batches for deep learning models, increasing training efficiency. The ability of the model to capture long-term and short-term trends [6] can be aided by the use of DWT since it effectively breaks down the time series data into frequency components. The representation of non-stationary signals can be improved similarly with the use of FSST, and spatial characteristics of the data can be effectively learned using CNNs [7]. In the case of Random Forest with BiLSTM or BiGRU, through an ensemble learning approach, variance reduction is done along with addressing of the overfitting issue. This model can even recognize the minor features in the data as well. The main aim of this work is to analyze the performance of different models and discover the most effective model for wind power prediction. Experiments indicate that the proposed hybrid CNN with BiGRU model performs best among all the models, achieving the highest $R^2$ value of 0.9093, the lowest MSE of 0.0120, and a reduced MAE of 0.0466. These results highlight the effectiveness of integrating CNN with BiGRU as a hybrid model in analyzing the complex dynamics of wind power data. The useful features from the input data are extracted using CNN and the long term dependencies in the sequential data are learned by BiGRU, which enhances the model's performance.

Major research in this area showcases methods that make use of BiGRU-Attention [8] processes, which gives attention to useful features hence improving predictions of both high and low frequency components [9]; and Transformers [10], which is used to identify long-term relationships within the data. Apart from this, Time2Vec data and WDCNN (Wide and Deep Convolutional Neural Network) [11] is incorporated with BiLSTM for wind power and photovoltaic predictions. Time2Vec effectively learns time-related features, WDCNN [12] extracts features at multiple scales, and BiLSTM handles temporal relationships, enhancing the model's predictive power. Additionally, a data screening algorithm developed using the RANSAC (Random Sample Consensus) [13] method and for prediction using a Seq2Seq (Sequence-to-Sequence) [14] model with Attention and BiGRU [15]. The RANSAC technique eliminates noise and outliers, while the Seq2Seq-Attention BiGRU [16]model focuses on sequences to provide better forecasts [17]. Feature extraction and selection using optimized Variational Mode Decomposition (VMD) [18] and an improved Multi-Objective Coati Optimization Algorithm [19] predict wind power with high accuracy by selecting the best features from the dataset.

Advancements in the reliability of wind power forecasts are crucial for renewable energy [20] management, given the inherent weaknesses of wind power. Newer forecast prediction methods [21] ensure easier integration of wind energy [22] into the power system, reducing dependence on fossil fuel backup power and improving overall power system reliability. The techniques and knowledge gained can also be applied to other renewable energy sources, contributing to sustainable energy systems [23]. Studies demonstrate the necessity of integrating deep learning and hybrid models to solve renewable energy forecasting problems [24], paving the way for future innovative solutions. Recent approaches in wind power prediction have seen considerable improvements with data-driven techniques. Wang et al. reviewed these methods, particularly for Automatic Generation Control performance assessment of wind–thermal bundled power systems [25]. Xiong et al. proposed an enhanced deep learning model based on Extreme Learning Machine (ELM) [26] and BiLSTM, using an improved reptile search algorithm. Tian, Hou, and Yan introduced an adaptive hybrid model, IVMD-FE-Ad-Informer (Improved Variational Mode Decomposition with Feature Extractor), to minimize errors and enhance wind power prediction accuracy. Sun et al. used a modified Particle Swarm Optimization (PSO) [27] approach with Attention-Based LSTM for improved accuracy [28]. Various novel methodologies for wind power prediction have been explored, such as integrating ARMA [29], PSO-SVM, and clustering [30], and the EALSTM-QR (Error Attribution Long Short-Term Memory - Quantile Regression) model combining numerical weather prediction with deep learning interval forecasting [31]. New trends in prediction time are discussed by Sawant et al. [32] and trends in data-based prediction is mentioned by Liu et al. [33]. These findings portray the growth of methodologies for wind power prediction and the use of combined and sophisticated approaches.

To sum up, this paper provides a valuable comparative investigation of hybrid BiLSTM and BiGRU structures for wind power prediction. By implementing various preprocessing techniques like DWT and FSST with BiLSTM and BiGRU, along with implementing hybrid models like incorporating CNN and RF with BiLSTM and BiGRU; wind patterns' variability is addressed. The results show that the CNN + BiGRU hybrid model is effective, achieving the highest $R^2$ of 0.9093, the lowest RMSE of 0.1095, making it a useful and accurate means of wind power prediction. This work paves way for the usage of preprocessing techniques and other feasible hybrid models with bidirectional recurrent neural networks, which can contribute to the advancement of renewable energy forecasts and sustainable energy supply systems.

## II. DATASET DESCRIPTION

The dataset is a compilation of Wind Power data generated in the year 2018. It includes readings at various intervals for a particular day, which clearly shows the fluctuations in readings for an accurate wind power prediction. This dataset is publicly accessible [34]. The dataset contains 50,530 instances, each of which have four features and a label. The features included in the dataset are as follows:

Date/Time has the timestamps of each reading. It is crucial for the temporal analysis of the data. It also takes into account the fluctuations in the readings. Wind Speed, measured in meters per second (m/s), is another basic feature that aids in wind power prediction. It has a direct effect on wind power generated as wind power is directly proportional to the cube of wind speed, thus making it an important aspect. The Theoretical Power Curve Value, measured in kilowatt-hours (kWh), measures the power output of the wind turbine based on the wind speed according to the theory. It provides a benchmark for comparing the actual power output and helps in understanding the efficiency of the wind power generation process happening in the described data. Wind Direction, measured in degrees, can affect the performance of wind turbines. The wind blow direction can impact the efficiency of power generation, thus making it an important variable for consideration. The target variable or the label in this dataset is the LV Active Power, measured in kilowatts (kW). LV Active Power represents the actual power output value, also referred to as the ground truth value. It is measured at the wind turbines and is used to compare with the predicted values of the models.

The dataset used in this research contains high-resolution temporal data in addition to the geographically influenced variables like wind direction. This wide range of features incorporates all the main aspects behind wind power generation into consideration. This can lay a strong foundation for the importance of using this dataset for building the predictive model. Hence by taking this dataset into action, the main aim is to improve the accuracy of wind power predictions.

## III. METHODOLOGY

The procedure used here includes several crucial processes, such as data preprocessing, data augmentation, model implementation, hyperparameter tuning, and model evaluation. Under data preprocessing, missing values are handled and normalized in the values under the same range after which an 80:20 train-test split is performed. Preprocessing techniques [35] such as Discrete Wavelet Transform (DWT) and Fourier Synchrosqueezed Transform (FSST) are employed. This can capture multiple frequency components and time-frequency representations. After data preprocessing, data augmentation is performed for better generalization by the model. Techniques such as the addition of Gaussian noise and temporal shifting are employed for the same. The model implementation phase is divided into two categories: models with specific preprocessing techniques and hybrid models. The former includes combinations like DWT with BiLSTM, DWT with BiGRU, FSST with BiLSTM, and FSST + BiGRU. The hybrid ones comprises of CNN with BiLSTM, CNN with BiGRU, RF with BiLSTM, and RF with BiGRU. The grid search method is used to get the fine-tuned optimal parameters for the best performance of the model. The models are then evaluated using metrics such as $R^2$, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). This thorough approach aims to provide accurate and reliable interim wind forecasts, thereby improving the precision of wind power predictions [36] and contributing to the efficient management of renewable energy resources. Following are the major steps involved:

### A. DATA PREPROCESSING

Here, initially, the missing values are handled and then the dataset values are normalized into the same range, after which the 80:20 train-test split is done. Additionally, two preprocessing techniques are utilized which are Discrete Wavelet Transform (DWT) and Fourier Synchrosqueezed Transform (FSST).

The use of DWT and FSST has boosted the model's performance on non-stationary data. It is these methods that resolve the limitations of traditional ways like Fourier Transform that usually are not able to properly track time-localized frequency components especially in non-stationary signals where the frequency content changes over time.

#### 1) DWT (DISCRETE WAVELET TRANSFORM)

To decompose the time series data into frequency components, DWT technique is leveraged. This technique helps in capturing both short-term and long-term patterns in the data. By transforming the time series data into the frequency domain, it facilitates the identification of major patterns and features that are crucial for accurate wind power prediction that would have been missed by only employing time-invariant methods. DWT has made it possible to analyze data finely, thereby capturing details of variability typical for non-stationary signals. This enhances the ability of the model to learn from various frequency components, leading to improved prediction performance. The DWT of a signal [4] $f(t)$ is given by:

$$DWT(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \psi \left( \frac{t-b}{a} \right) dt \tag{1}$$

where $\psi\left(\frac{t-b}{a}\right)$ is the mother wavelet, $a$ is the scale parameter, and $b$ is the translation parameter.

#### 2) FSST (FOURIER SYNCHROSQUEEZED TRANSFORM)

To decompose the time series data into a time-frequency representation, the FSST technique is leveraged. Non-stationarity commonly found in wind power datasets are analyzed in detail using this technique. It creates a time-frequency matrix form where transient features and key patterns can be observed. It refines time frequency representation by re-distributing frequency components better. By preprocessing the data with FSST, the model's capability to understand and predict complex temporal dynamics, is enhanced. This technique is particularly useful for capturing sudden changes and variations in wind power generation. The FSST of a signal [5] $f(t)$ is defined as:

$$FSST(\tau, f) = \int_{-\infty}^{\infty} f(t) \frac{|f(t-\tau)|}{\sqrt{2\pi\sigma^2}} e^{-j2\pi ft} dt \tag{2}$$

where $\sigma$ is the standard deviation of the Gaussian window.

### B. DATA AUGMENTATION

One way is by adding noise at this stage, which helps the model generalize better and enhance its ability to learn. For example, a Gaussian noise can be added as part of the small random changes in numerical data, enabling the model to still perform effectively even when dealing with noisy real-world values. Moreover, transformation involves shifting over time in time series; through temporal shift, the time-series data is shifted slightly in one direction or the other, introducing a slight temporal relationship adjustment. This adjustment is intended to make the model learn time-varying relationships that are stable even when there are minor perturbations. Another technique that can help with scaling: letting the model learn from data points at varying wind power magnitudes in addition to these two, Salt and Pepper noise is introduced so that the model can detect sudden spikes or outliers.

### C. MODEL IMPLEMENTATION

Components for model implementation that have been used are BiLSTM and BiGRU. These models are particularly proven effective for sequence learning tasks because of this advantage of the architectures. It captures dependencies over the current time step as well as dependencies over the previous and/or future time steps. Both directions of data are processed through BiLSTM and it improves the variants showing how the original model is capable of learning long-term dependencies. BiGRU has similar capabilities, but fewer parameters compared to BiLSTM.

The forward pass $\overrightarrow{h_t}$ and backward pass $\overleftarrow{h_t}$ for BiLSTM are defined as:

$$\overrightarrow{h_t} = LSTM(x_t, \overrightarrow{h_{t-1}}) \tag{3}$$

$$\overleftarrow{h_t} = LSTM(x_t, \overleftarrow{h_{t+1}}) \tag{4}$$

The forward pass $\overrightarrow{h_t}$ and backward pass $\overleftarrow{h_t}$ for BiGRU are defined as:

$$\overrightarrow{h_t} = GRU(x_t, \overrightarrow{h_{t-1}}) \tag{5}$$

$$\overleftarrow{h_t} = GRU(x_t, \overleftarrow{h_{t+1}}) \tag{6}$$

The output is obtained by concatenating $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$.

#### 1) MODELS WITH SPECIFIC PREPROCESSING TECHNIQUES

a) When BiLSTM is incorporated with DWT preprocessing technique, better predictions are noticed as it makes use of the decomposed frequency components. BiLSTM processes the input data in both forward and backward directions, capturing dependencies from both past and future time steps. This bidirectional approach, enhanced by the features extracted through DWT, enables the model to better understand the temporal patterns and improve forecasting accuracy.

b) When DWT is combined with BiGRU, it makes use of the frequency components extracted using DWT and the efficient learning capacity of GRU. BiGRU processes the data in both directions, capturing contextual information from past and future time steps. DWT with BiGRU combination enhances the ability of the model to predict wind power by capturing temporal dependencies and patterns found in the frequency domain.

c) When FSST is combined with BiLSTM, it makes use of the high-resolution time-frequency matrix created by FSST. BiLSTM processes this enriched data bidirectionally, allowing the model to capture intricate temporal patterns and dependencies. This combination improves the predictive performance of the model by making use of the detailed time-frequency features extracted by FSST.

d) When FSST is combined with BiGRU, it makes use of the detailed time-frequency matrix created by FSST. The bidirectional nature of the BiGRU captures dependencies from both past and future time steps, enhancing the understanding of the model about the temporal dynamics in the data. This combination improves wind power prediction accuracy by effectively utilizing the time-frequency features extracted by FSST.

#### 2) HYBRID MODELS

In case of the implementation of the hybrid models, CNN is as a deep learning model used for feature extraction and RF is a machine learning model for pattern recognition.

The convolution operation for input $x$ and filter $w$ is given by:

$$s(t) = (x * w)(t) = \sum_{\tau} x(\tau)w(t-\tau) \tag{7}$$

where $x(\tau)$ is the input signal at time $\tau$, $w(t-\tau)$ is the filter (or kernel) applied at time $t-\tau$ and $s(t)$ is the output signal after convolution.

The Random Forest prediction for wind power $\hat{y}$ is given by:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \tag{8}$$

where $N$ is the total number of trees and $T_i(x)$ is the prediction of the $i$-th decision tree.

The CNN architecture used consists of a Conv1D layer with a kernel size 1 and 64 filters, which is then followed by multiple BiLSTM and BiGRU layers with the optimized number of layers and units, and an output Dense layer.

The RF architecture used consists of 100 decision trees with each one trained on subsets of the train split data, and the mean value of the outputs of these independent trees makes the final prediction.

a) CNN enables the extraction of spatial features when combined with BiLSTM, allowing the former to process the extracted features in both forward and backward directions empowering the latter to capture long-term dependencies thereby enhancing prediction accuracy. This hybrid model blends the strengths of CNN and BiLSTM together, ensuring that it provides better wind power prediction while utilizing each model's positive attributes.

b) CNN extracts the spatial features from the input data, when CNN is coupled with BiGRU. As the features are bi-directionally processed by BiGRU, it implies that contextual information is well captured in both the directions. The integration of these two models form a fusion of CNN's feature extraction capabilities and BiGRU's effective temporal processing, resultantly forming this hybrid model. An optimal performance in prediction tasks is thus achievable through this hybrid model that takes advantages from both ends: CNN and BiGRU.

c) The combination of RF with BiLSTM enables the RF to extract key features of wind power input data, which are further fed into BiLSTM to capture temporal dependencies. By this hybrid model, RF gains from robust feature extraction and powerful sequence modeling capabilities of BiLSTM which enhances wind power forecasting.

d) When RF is incorporated with BiGRU, it is the RF model which is being used mainly for deriving the relevant features out of the given input data. These features are processed bidirectionally by BiGRU. This approach results in preservation of meaningful features and some temporal dependencies, and, thus, provides better predictive performance of the created model. The Hybrid model can take advantage of RF's feature extraction capability as well as BiGRU's capability in learning the temporal dynamics of a given context.

### D. HYPERPARAMETER OPTIMIZATION

The Grid Search method is used to fine-tune and get the optimal hyperparameters for both BiLSTM and BiGRU models. This method involves systematically exploring a predefined set of hyperparameters to identify the best combination that yields the highest model performance. Parameters such as the number of layers, number of units per layer, learning rate, and batch size are optimized. Learning rates between 0.001 and 0.01 are tested in achieving a balance between stability, while dropouts were explored from 0.1 to 0.5 to avoid overfitting as well as stability with convergence speed. The reason for choosing this search space was the need to adapt the model to specific problem settings concerned with non-stationary data. The temporal and frequency components of variability called for such high dimensions because they needed to be able to change patterns adaptively and capture specifics. By evaluating the performance of the model on a validation set for each combination of hyperparameters, it is ensured that the selected parameters result in the best predictive accuracy. This fine-tuning of the models is crucial to achieve optimal performance in wind power prediction.

### E. EVALUATION OF THE MODELS

The developed models are evaluated using $R^2$ (Coefficient of Determination), MSE (Mean Squared Error), RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error).
$R^2$ (Coefficient of Determination) is given as:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{9}$$

MSE (Mean Squared Error) is given as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

RMSE (Root Mean Squared Error) is given as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{11}$$

MAE (Mean Absolute Error) is given as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{12}$$

Here, $y_i$ is the actual observed value, $\hat{y}_i$ is the model predicted value, and $\bar{y}$ is the mean of the actual observed values.

To assure the generalization by the model, k-fold cross validation is employed with k = 10 to keep a check on the potential overfitting scenario. The model is trained through these folds and the average of the performance metrics values across the folds is utilized.

The methodology followed is as shown in Figure 1.

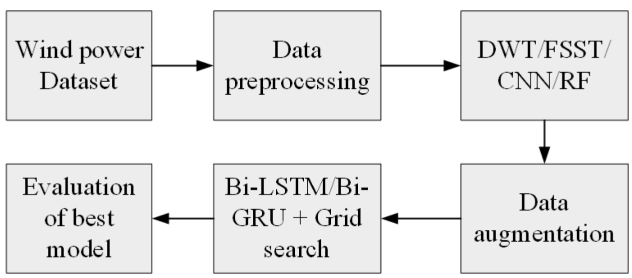



**FIGURE 1.** Methodology block diagram.

## IV. RESULTS AND DISCUSSIONS

After assessing the models that are listed earlier, the experimental details are discussed in this section. Historical wind power and meteorological data from a particular wind farm were used in the experiments. To guarantee a just comparison, the dataset is split into training and testing sets while keeping the same proportion for every model. The performance metrics used to evaluate the models are coefficient of determination ($R^2$), mean squared error (MSE), root mean square error (RMSE), and mean absolute error (MAE). These metrics, as depicted in Table 1 provide a complete evaluation of the models' ability to generalize to new data as well as their predictive accuracy.

**TABLE 1. Comparison between the models.**

| Model | $R^2$ | MSE | RMSE | MAE |
|---|---|---|---|---|
| DWT + BiLSTM | 0.9071 | 0.0123 | 0.1109 | 0.0521 |
| DWT + BiGRU | 0.9088 | 0.0121 | 0.1100 | 0.0516 |
| FSST + BiLSTM | 0.9084 | 0.0121 | 0.1100 | 0.0486 |
| FSST + BiGRU | 0.9087 | 0.0121 | 0.1100 | 0.0479 |
| CNN + BiLSTM | 0.9078 | 0.0122 | 0.1105 | 0.0465 |
| CNN + BiGRU | 0.9093 | 0.0120 | 0.1095 | 0.0466 |
| RF + BiLSTM | 0.8957 | 0.0138 | 0.1175 | 0.0506 |
| RF + BiGRU | 0.8968 | 0.0136 | 0.1166 | 0.0479 |

The stable and consistent range of performance metrics were observed for a confidence interval of 95%, thereby ensuring the robust predictions by the developed models and not heavily influenced due to data variations.

### A. LEARNING CURVES

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. Figures 2 to Figure 9 illustrate the learning curves for each of the models.

The learning curve of DWT with BiLSTM model in Figure 2 indicates that the training and validation accuracies have been improving steadily, which means that the model can learn effectively from the data. From this graph, it can be observed that as we progress into different epochs, the loss for training tends to diminish throughout while at a specific time, there are some changes in validation loss though it still reduces on average. The learning trend shows how well the model has learned from the training set and how it continues performing on the validation dataset.

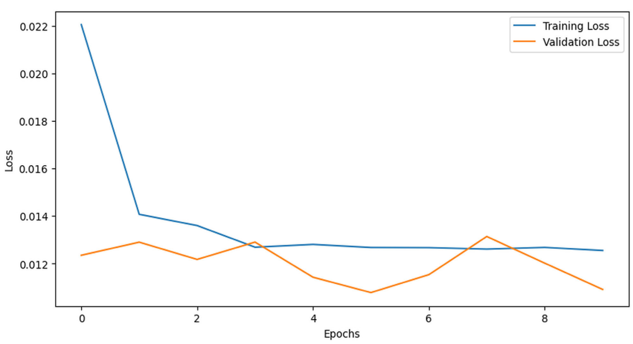



**FIGURE 2. Learning curve for DWT with BiLSTM model.**

A learning curve of the DWT with BiGRU model is shown in Figure 3 which indicates that it converges much faster and generalizes much better than the DWT with BiLSTM model. There is a rapid decline in training losses during the first few epochs, showing that learners learn quickly. Moreover, the development of validation losses has been steadily downward with slightly more oscillations, thus indicating efficient generalization and robustness to new observations. This comparison demonstrates that not only does the DWT with BiGRU model learn faster but also its performance on the validation set is better than that of the DWT with BiLSTM model.

In Figure 4, the learning curve shows that FSST with BiLSTM model can quickly learn patterns hidden in data and achieve high accuracy at the early stages of training. A sharp fall in training loss over the first epochs is displayed by the graph, which implies fast learning. Also, validation loss decreases while fluctuating slightly showing that the model generalizes well from training to novel data. It converges rapidly to low loss values suggesting that FSST with BiLSTM model is good at picking up key characteristics of a dataset.

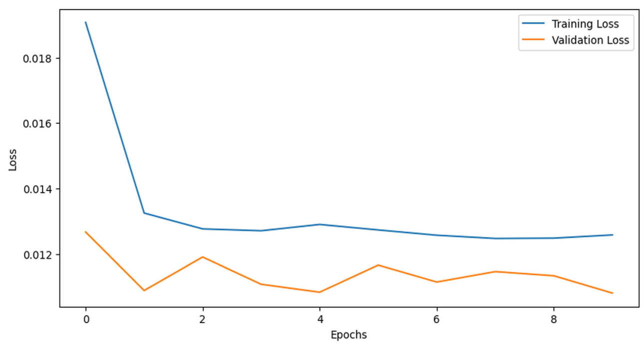



**FIGURE 3. Learning curve for DWT with BiGRU model.**

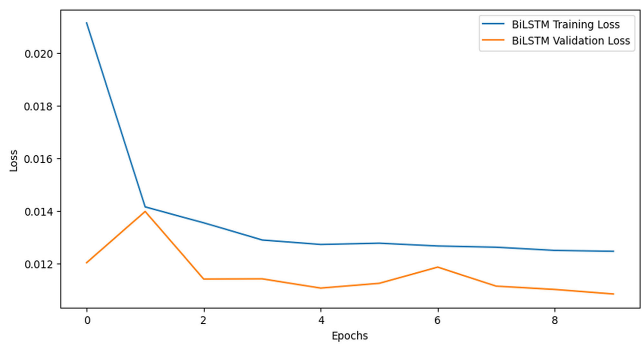



**FIGURE 4. Learning curve for FSST with BiLSTM model.**

As portrayed in Figure 5, the FSST with BiGRU model demonstrates steady learning with minimal overfitting, implying robust performance of the model. The training loss decreases monotonously indicating a constant type of learning. The validation loss generally moves downwards though not as smoothly as expected hence representing a good generalization ability of the model to unseen data. This indicates that FSST with BiGRU achieves a compromise between learning from training data and performing adequately on the validation set to avoid overfitting and keep its efficiency reliable.

Figure 6 demonstrates how CNN with BiLSTM models can extract relevant features leading to continuous improvement in both training and validation metrics. This consistent performance in both training and validation metrics demonstrates the effectiveness of combining CNN with BiLSTM in capturing important features and maintaining robust predictive accuracy. The CNN with BiGRU model learning curve graph in Figure 7 tells us that it is a better learner than the other models. The two lines on the graph are almost parallel and this indicates that their gap also remains minimal showing that there is increased generalization on the part of the model. As seen, loss decreases rapidly at first for training data, which implies efficient learning. Loss was also seen to drop to some extent before slightly rising again during validation but generally remaining close to training loss. These findings suggest that overfitting is not severe in this case and the latter effectively generalizes information for new inputs, thus showing robustness during both stages of training and validation.

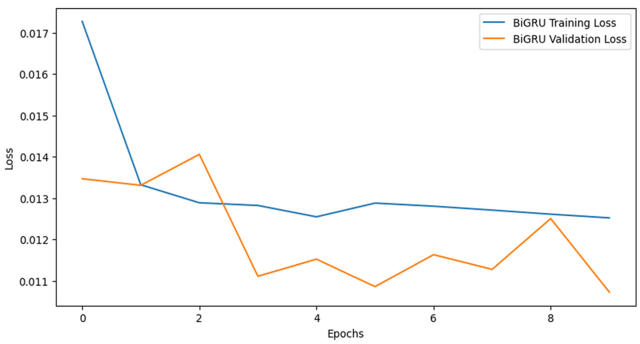



**FIGURE 5. Learning curve for FSST with BiGRU model.**

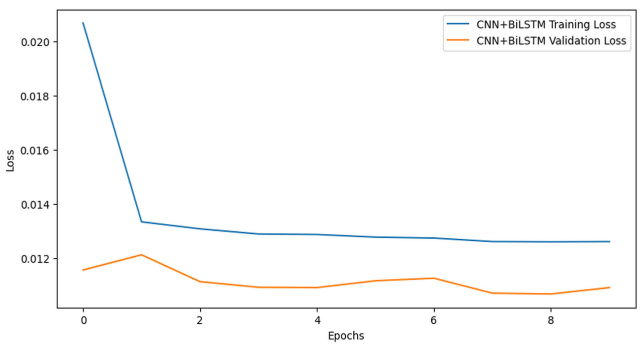



**FIGURE 6. Learning curve for CNN with BiLSTM model.**

Figure 8 is a learning curve of RF with the BiLSTM model which performs well consistently throughout the training process. Training loss decreases quickly in the first epoch and stabilizes, indicating efficient learning. The validation loss closely follows the training loss with both curves nearly overlapping, implying low overfitting and powerful generalization. These steady trends in both training and validation measures indicate that RF with BiLSTM model is robust enough to capture patterns effectively and generalize them in data. The learning curve of the RF with the BiGRU model in Figure 9 demonstrates accurate feature extraction and sequence learning, leading to high prediction accuracy.

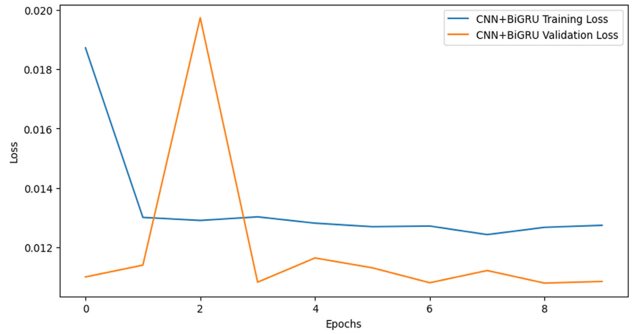



**FIGURE 7. Learning curve for CNN with BiGRU model.**

The initial few epochs show a steep drop in training loss implicating fast learning. The validation loss tightly follows the training loss and both seem to level off at extremely low values; this implies good generalization ability of the model for unseen examples. Additionally, the narrow difference between two curves indicates that overfitting is avoided by RF with the BiGRU model which provides excellent performance and liability during all stages of the training process.

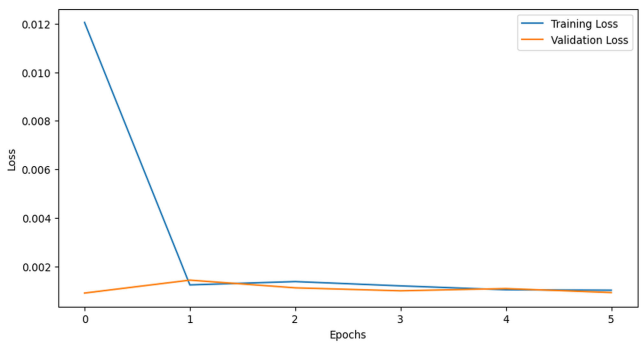



**FIGURE 8. Learning curve for RF with BiLSTM model.**

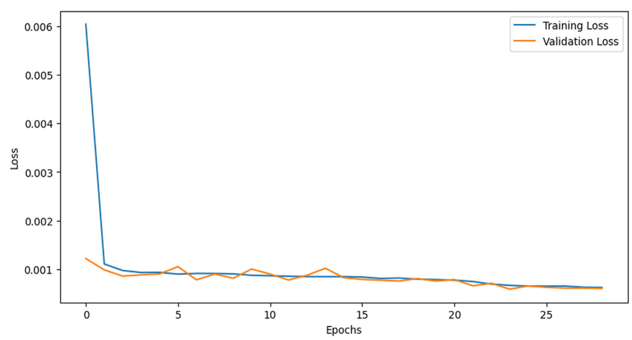



**FIGURE 9. Learning curve for RF with BiGRU model.**

The analysis of these learning curves provides valuable insights into the strengths and weaknesses of each model, guiding future improvements in wind power prediction methodologies.

### *B. PREDICTION VS TRUE VALUES*

In this subsection, the predicted wind power values against the actual observed values is compared to evaluate the accuracy of each model. Figure 10 to Figure 17 show the prediction versus true values plots for each of the models.

The closeness between the actual values and forecasts made using DWT with BiLSTM is well illustrated in Figure 10. It means that this model is very successful in recording all of the underlying patterns and trends from the observed data, hence making accurate predictions.

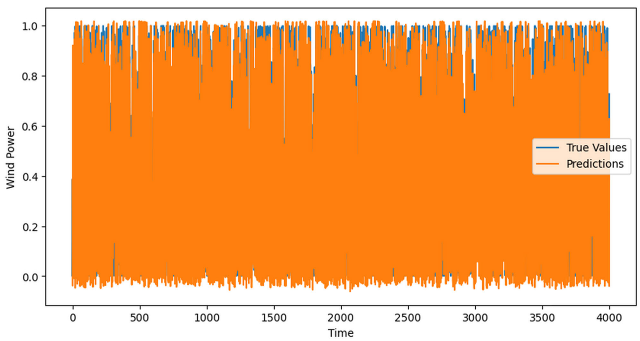



**FIGURE 10. Actual vs predicted values for DWT with BiLSTM model.**

Figure 11 shows that the combination of model DWT and BiGRU forecast is even closer to the actual wind power values. The enhancement of this closeness suggests that the model has better predictive accuracy as a result. On the graph, blue represents true values while orange represents predictions; the latter shadowing the former almost every step of the way along the time series line. Such proximity visibly solidifies how adept the model is at encapsulating and then echoing forthcoming trends of wind power. In turn, this only strengthens its worth and dependability by showing more evidence in support of its effectiveness and reliability.

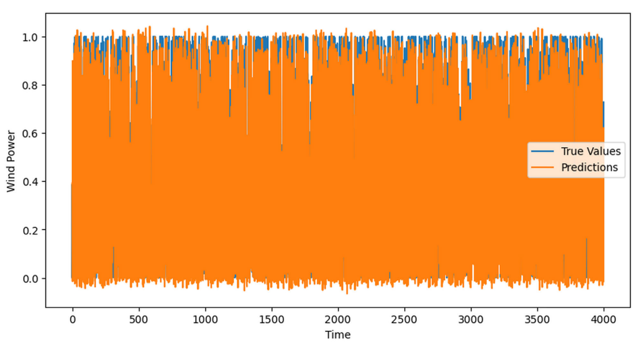



**FIGURE 11. Actual vs predicted values for DWT with BiGRU model.**

In Figure 12, the FSST through BiLSTM model is displayed to have a near perfect capability of predicting wind power. The actual values and predicted values almost overlay each other as depicted by the blue and orange lines respectively; this high level of consistency demonstrates the accuracy and reliability of this model in forecasting wind power trends. The ability of the model to forecast these fluctuating true values has good precision: these are shown by blue bars while the orange bars show predictions that are very close to them throughout the time series. It underscores the successful outcome in capturing wind power fluctuations— hence being able to depict real values with striking similarity, which confirms further credibility and robustness in performance. The FSST with BiGRU model as shown in Figure 13 shows excellent prediction accuracy, with minimal deviation from the actual values.

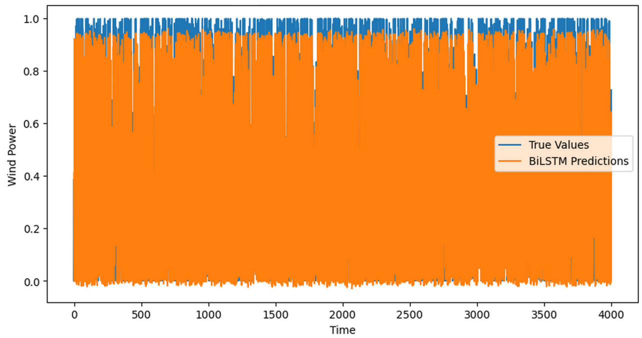



**FIGURE 12. Actual vs predicted values for FSST with BiLSTM model.**

The capacity of the CNN with BiLSTM model to grasp the real wind power values is illustrated in Figure 14. The orange lines denote the predictions made by this model, which are significantly parallel if not identical to the blue lines, representing actual wind power values. Throughout the entire time series, their consistency is conspicuous enough to underscore that the predictive accuracy of our model is quite high and it can learn well from the data while generalizing effectively. As we observe minimal disparities between these two sets at any point along our timeline, it strengthens the confidence in how reliable the CNN with BiLSTM model can be when used for forecasting wind power trends; such a demonstration highlights its worthiness in practice.

Figure 15 for CNN with BiGRU has the highest accuracy among all models and the model predictions come close to the actual values as shown in the graph below. In the graph, the blue line depicts the true values of wind power, and the predictions of the CNN with BiGRU model are shown in the orange line. It can seen clearly from the graph that the model predictions line is in line with the true values line. This slight difference between the predicted with the true value is because of many uncertainties in wind power generation and it is technically not possible to achieve perfect accuracy. Nevertheless, the accuracy achieved for this model is truly exceptional, and the model played a very important role in forecasting wind power for renewable energy management.

Figure 16 indicates that the RF with BiLSTM model provides reliable predictions with good alignment to the true values. The RF with BiGRU model in Figure 17 demonstrates its ability by closely matching the actual wind power values it predicted. The graph shows the actual wind power represented by the blue line and the model's predicted values with the orange line that easily correlate with one another. It shows that the RF with BiGRU model has learned a lot from the data that is being presented in the temporal feature vector and was able to scale the probability to any point in time and estimate the power production about the actual historical data providing us a reliable and precise data. The small gap between the real wind power and the model-estimated wind power indicates that this model is a good candidate to use for wind power prediction.

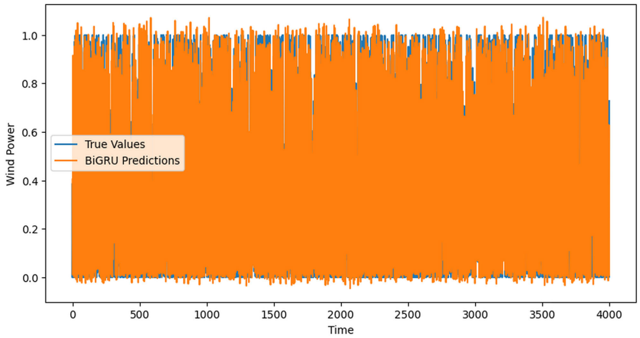



**FIGURE 13. Actual vs predicted values for FSST with BiGRU model.**

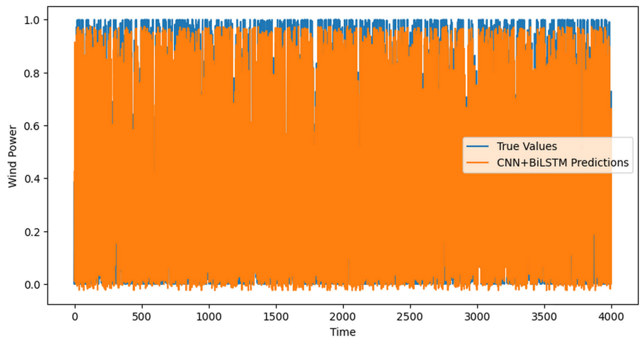



**FIGURE 14. Actual vs predicted values for CNN with BiLSTM model.**

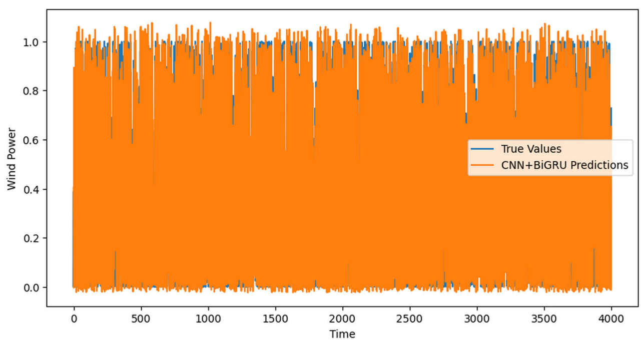



**FIGURE 15. Actual vs predicted values for CNN with BiGRU model.**

These plots show how well each model predicts, showing the effectiveness of combining different preprocessing techniques and using other models like CNN and RF with BiLSTM and BiGRU to forecast wind power. The superior performance of CNN with the BiGRU model is demonstrated by its high correlation between predicted and real values thus revealing the potentiality of such a hybrid approach for accurate and reliable wind power forecasting. The results showed that the CNN with BiGRU model performed

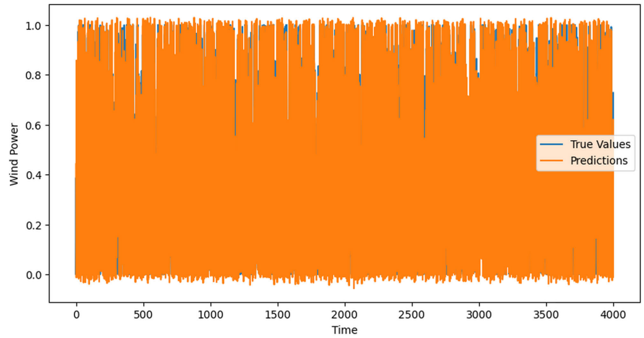



**FIGURE 16. Actual vs predicted values for RF with BiLSTM model.**

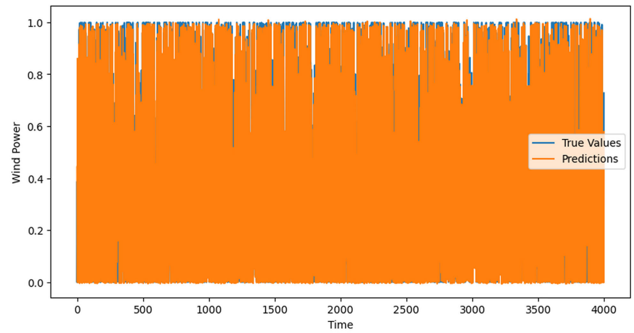



**FIGURE 17. Actual vs predicted values for RF with BiGRU model.**

better compared to other hybrid models by achieving the highest $R^2$ score of 0.9093 which represents the correlation between actual and predicted wind energy levels. This model has the lowest MSE (0.0120) and RMSE (0.1095), which are indicative of its ability to minimize forecast errors. In addition, the model's MAE of 0.0466 indicates consistent accuracy under different test conditions. These findings show that BiGRUs combined with CNNs for feature extraction significantly improved prediction performance by effectively capturing temporal relationships.

It is also found that combining the BiLSTM and BiGRU models with sophisticated preprocessing methods like Discrete Wavelet Transform (DWT) and Fourier Synchrosqueezed Transform (FSST) improved their prediction performance. The $R^2$, RMSE, MSE, and MAE of these models that made use of these preprocessing techniques were significantly higher than those of the baseline models. Deep learning models appear to be better at capturing complicated patterns and relationships in data by dividing the wind power time series into smaller, more manageable components. Overall, the experimental findings demonstrate the validity of the proposed method's ability to generate reliable and accurate wind power predictions, hence creating a more stable and effective renewable energy policy.

## V. CONCLUSION

More sophisticated techniques are employed such as BiGRU (Bi-directional Gated Recurrent Units), Discrete Wavelet Transform (DWT), etc. which have demonstrated significant enhancement of the accuracy and robustness in forecasting wind energy. The evaluation metrics for the superior CNN with BiGRU model are: $R^2$ equal to 0.9093, MSE equal to 0.0120, RMSE equal to 0.1095 and MAE equal to 0.0466. These range of values for a model is an indicator that corresponds to ideal metrics; Incorporating information from past and future data sources, this model provides a mechanism robust for stacked context while also capturing non-trivial temporal dependencies through wind energy statistics. Since big data in wind power forecasting is highly dynamic, the bidirectional analysis of the BiGRU model referring to both forward and reverse requires sets plays a crucial role where it outperforms significantly compared with most reflecting forecastable signals. Training and validation loss curves almost overlap giving the signature of overfitting but it is not, rather it proves that the model works with confidence on different datasets.

Satellite photographs, atmospheric pressure, and geographic information should be included in studies to improve the accuracy of wind electricity forecasts. Valuable information about wind variations and patterns can be offered by combining satellite information with advanced meteorological fashions. Exploring advanced device learning strategies inclusive of ensemble procedures and hybrid models can improve forecasting performance. strategies along with fusion mastering and switch learning can leverage a couple of data assets while preserving statistics privacy. New statistics are vital for continuous improvement and accuracy. New facts make programs more usable. Automatic function planning and strong statistics processing can enhance forecast accuracy and speed up model education.

## REFERENCES


[1] P. Xu, M. Zhang, Z. Chen, B. Wang, C. Cheng, and R. Liu, "A deep learning framework for day ahead wind power short-term prediction," *Appl. Sci.*, vol. 13, no. 6, p. 4042, Mar. 2023, doi: 10.3390/app13064042.

[2] D. El Bourakadi, A. Yahyaouy, and J. Boumhidi, "Improved extreme learning machine with AutoEncoder and particle swarm optimization for short-term wind power prediction," *Neural Comput. Appl.*, vol. 34, no. 6, pp. 4643–4659, Mar. 2022, doi: 10.1007/s00521-021-06619-x.

[3] W. Chen, W. Qi, Y. Li, J. Zhang, F. Zhu, D. Xie, W. Ru, G. Luo, M. Song, and F. Tang, "Ultra-short-term wind power prediction based on bidirectional gated recurrent unit and transfer learning," *Frontiers Energy Res.*, vol. 9, Dec. 2021, Art. no. 808116, doi: 10.3389/fenrg.2021.808116.

[4] A. Barjasteh, S. H. Ghafouri, and M. Hashemi, "A hybrid model based on discrete wavelet transform (DWT) and bidirectional recurrent neural networks for wind speed prediction," *Eng. Appl. Artif. Intell.*, vol. 127, Jan. 2024, Art. no. 107340, doi: 10.1016/j.engappai.2023.107340.

[5] L. Li, H. Cai, H. Han, Q. Jiang, and H. Ji, "Adaptive short-time Fourier transform and synchrosqueezing transform for non-stationary signal separation," *Signal Process.*, vol. 166, Jan. 2020, Art. no. 107231, doi: 10.1016/j.sigpro.2019.07.024.

[6] S. Guan, Y. Wang, L. Liu, J. Gao, Z. Xu, and S. Kan, "Ultra-short-term wind power prediction method combining financial technology feature engineering and XGBoost algorithm," *Heliyon*, vol. 9, no. 6, Jun. 2023, Art. no. e16938, doi: 10.1016/j.heliyon.2023.e16938.

[7] Y. Chen, H. Zhao, R. Zhou, P. Xu, K. Zhang, Y. Dai, H. Zhang, J. Zhang, and T. Gao, "CNN-BiLSTM short-term wind power forecasting method based on feature selection," *IEEE J. Radio Freq. Identificat.*, vol. 6, pp. 922–927, 2022, doi: 10.1109/JRFID.2022.3213753.

[8] Y. Meng, C. Chang, J. Huo, Y. Zhang, H. M. M. Al-Neshmi, J. Xu, and T. Xie, "Research on ultra-short-term prediction model of wind power based on attention mechanism and CNN-BiGRU combined," *Frontiers Energy Res.*, vol. 10, May 2022, Art. no. 920835, doi: 10.3389/fenrg.2022.920835.

[9] S. Wang, J. Shi, W. Yang, and Q. Yin, "High and low frequency wind power prediction based on transformer and BiGRU-attention," *Energy*, vol. 288, Feb. 2024, Art. no. 129753, doi: 10.1016/j.energy.2023.129753.

[10] M. Yu, D. Niu, T. Gao, K. Wang, L. Sun, M. Li, and X. Xu, "A novel framework for ultra-short-term interval wind power prediction based on RF-WOA-VMD and BiGRU optimized by the attention mechanism," *Energy*, vol. 269, Apr. 2023, Art. no. 126738, doi: 10.1016/j.energy.2023.126738.

[11] D. Geng, B. Wang, and Q. Gao, "A hybrid photovoltaic/wind power prediction model based on Time2Vec, WDCNN and BiLSTM," *Energy Convers. Manag.*, vol. 291, Sep. 2023, Art. no. 117342, doi: 10.1016/j.enconman.2023.117342.

[12] G. Zhou, G. Hu, D. Zhang, and Y. Zhang, "A novel algorithm system for wind power prediction based on RANSAC data screening and Seq2Seq-attention-BiGRU model," *Energy*, vol. 283, Nov. 2023, Art. no. 128986, doi: 10.1016/j.energy.2023.128986.

[13] C. Wang, H. Lin, H. Hu, M. Yang, and L. Ma, "A hybrid model with combined feature selection based on optimized VMD and improved multi-objective coati optimization algorithm for short-term wind power prediction," *Energy*, vol. 293, Apr. 2024, Art. no. 130684, doi: 10.1016/j.energy.2024.130684.

[14] Z. Zhao, H. Nan, Z. Liu, and Y. Yu, "Multi-step interval prediction of ultra-short-term wind power based on CEEMDAN-FIG and CNN-BiLSTM," *Environ. Sci. Pollut. Res.*, vol. 29, no. 38, pp. 58097–58109, Aug. 2022, doi: 10.1007/s11356-022-19885-6.

[15] H. He and M. Yuan, "A novel few-sample wind power prediction model based on generative adversarial network and quadratic mode decomposition," *Frontiers Energy Res.*, vol. 11, Jul. 2023, Art. no. 1211360, doi: 10.3389/fenrg.2023.1211360.

[16] H. Sun, Q. Cui, J. Wen, L. Kou, and W. Ke, "Short-term wind power prediction method based on CEEMDAN-GWO-Bi-LSTM," *Energy Rep.*, vol. 11, pp. 1487–1502, Jun. 2024, doi: 10.1016/j.egyr.2024.01.021.

[17] S. Wang, B. Li, G. Li, B. Li, H. Li, K. Jiao, and C. Wang, "A comprehensive review on the development of data-driven methods for wind power prediction and AGC performance evaluation in wind–thermal bundled power systems," *Energy AI*, vol. 16, May 2024, Art. no. 100336, doi: 10.1016/j.egyai.2024.100336.

[18] J. Xiong, T. Peng, Z. Tao, C. Zhang, S. Song, and M. S. Nazir, "A dual-scale deep learning model based on ELM-BiLSTM and improved reptile search algorithm for wind power prediction," *Energy*, vol. 266, Mar. 2023, Art. no. 126419, doi: 10.1016/j.energy.2022.126419.

[19] Y. Tian, D. Wang, G. Zhou, J. Wang, S. Zhao, and Y. Ni, "An adaptive hybrid model for wind power prediction based on the IVMD-FE-Ad-Informer," *Entropy*, vol. 25, no. 4, p. 647, Apr. 2023, doi: 10.3390/e25040647.

[20] Y. Sun, X. Wang, and J. Yang, "Modified particle swarm optimization with attention-based LSTM for wind power prediction," *Energies*, vol. 15, no. 12, p. 4334, Jun. 2022, doi: 10.3390/en15124334.

[21] I. Colak, S. Sagiroglu, and M. Yesilbudak, "Data mining and wind power prediction: A literature review," *Renew. Energy*, vol. 46, pp. 241–247, Oct. 2012, doi: 10.1016/j.renene.2012.02.015.

[22] N. Li, Y. Wang, W. Ma, Z. Xiao, and Z. An, "A wind power prediction method based on DE-BP neural network," *Frontiers Energy Res.*, vol. 10, Feb. 2022, Art. no. 844111, doi: 10.3389/fenrg.2022.844111.

[23] Y. Wang, D. Wang, and Y. Tang, "Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods," *IEEE Access*, vol. 8, pp. 17071–17079, 2020, doi: 10.1109/ACCESS.2020.2968390.

[24] X. Peng, H. Wang, J. Lang, W. Li, Q. Xu, Z. Zhang, T. Cai, S. Duan, F. Liu, and C. Li, "EALSTM-QR: Interval wind-power prediction model based on numerical weather prediction and deep learning," *Energy*, vol. 220, Apr. 2021, Art. no. 119692, doi: 10.1016/j.energy.2020.119692.

[25] M. Sawant, R. Patil, T. Shikhare, S. Nagle, S. Chavan, S. Negi, and N. D. Bokde, "A selective review on recent advancements in long, short and ultra-short-term wind power prediction," *Energies*, vol. 15, no. 21, p. 8107, Oct. 2022, doi: 10.3390/en15218107.

[26] D. Li, X. Yu, S. Liu, X. Dong, H. Zang, and R. Xu, "Wind power prediction based on PSO-Kalman," *Energy Rep.*, vol. 8, pp. 958–968, Jul. 2022, doi: 10.1016/j.egyr.2022.02.077.

[27] Y. Liu, Y. Wang, Q. Wang, K. Zhang, W. Qiang, and Q. H. Wen, "Recent advances in data-driven prediction for wind power," *Frontiers Energy Res.*, vol. 11, Jun. 2023, Art. no. 1204343, doi: 10.3389/fenrg.2023.1204343.

[28] J. Ren, Z. Yu, G. Gao, G. Yu, and J. Yu, "A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism," *Energy Rep.*, vol. 8, pp. 437–443, Aug. 2022, doi: 10.1016/j.egyr.2022.02.206.

[29] K. Yuan, K. Zhang, Y. Zheng, D. Li, Y. Wang, and Z. Yang, "Irregular distribution of wind power prediction," *J. Modern Power Syst. Clean Energy*, vol. 6, no. 6, pp. 1172–1180, Nov. 2018, doi: 10.1007/s40565-018-0446-9.

[30] Z. Li, R. Xu, X. Luo, X. Cao, and H. Sun, "Short-term wind power prediction based on modal reconstruction and CNN-BiLSTM," *Energy Rep.*, vol. 9, pp. 6449–6460, Dec. 2023, doi: 10.1016/j.egyr.2023.06.005.

[31] G. An, Z. Jiang, X. Cao, Y. Liang, Y. Zhao, Z. Li, W. Dong, and H. Sun, "Short-term wind power prediction based on particle swarm optimization-extreme learning machine model combined with AdaBoost algorithm," *IEEE Access*, vol. 9, pp. 94040–94052, 2021, doi: 10.1109/ACCESS.2021.3093646.

[32] Y. Sheng, H. Wang, J. Yan, Y. Liu, and S. Han, "Short-term wind power prediction method based on deep clustering-improved temporal convolutional network," *Energy Rep.*, vol. 9, pp. 2118–2129, Dec. 2023, doi: 10.1016/j.egyr.2023.01.015.

[33] H. Zheng, Z. Hu, X. Wang, J. Ni, and M. Cui, "VMD-CAT: A hybrid model for short-term wind power prediction," *Energy Rep.*, vol. 9, pp. 199–211, Jun. 2023, doi: 10.1016/j.egyr.2023.02.061.

[34] B. Erisen. *Wind Turbine Scada Dataset*. Accessed: Jun. 7, 2024. [Online]. Available: https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset

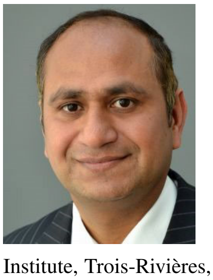
[35] V. Vanitha, D. Raphel, and R. Resmi, "Forecasting of wind power using variational mode Decomposition-Adaptive Neuro Fuzzy inference system," in *Proc. Innov. Power Adv. Comput. Technol. (i-PACT)*, vol. 1, Mar. 2019, pp. 1–4.

[36] S. Shivkumar, R. Padam Singh, and P. Kumar, "Wind power prediction using transformers: A federated learning approach," in *Proc. 9th Int. Conf. Signal Process. Commun. (ICSC)*, Dec. 2023, pp. 491–497, doi: 10.1109/ICSC60394.2023.10441057.


**E. P. VISHNUTHEERTH** received the B.Tech. degree in electronics and communication engineering from the Amrita School of Engineering, Amritapuri, India, in 2023. He is currently pursuing the M.Tech. degree in data science with the Amrita School of Artificial Intelligence, Coimbatore, India. His research interests include machine learning, deep learning, natural language processing, the IoT, and blockchain.

**VIVEK VIJAY** received the B.Tech. degree in electrical and electronics engineering from the Amrita School of Engineering, Amritapuri, India, in 2021. He is currently pursuing the M.Tech. degree in data science with the Amrita School of Artificial Intelligence, Coimbatore, India. After completing the B.Tech. studies, he joined Tata Consultancy Services (TCS), where he started his professional career in IT industry. His role involved working on various projects that honed his technical and analytical skills. His research interests include deep learning, natural language processing, and speech processing.

**RAHUL SATHEESH** (Member, IEEE) received the B.Tech. degree in electrical and electronics engineering from the University of Kerala, India, in 2012, the M.E. degree in power systems from Anna University, Chennai, India, in 2014, and the Ph.D. degree in power systems from the Department of Electrical Engineering, National Institute of Technology Calicut, Kerala, India. Following his doctoral studies, he was a Senior Engineer with Bosch Global Software Technologies, where he specialized in electrification and system engineering topics in electric vehicles (EVs). He is currently an Assistant Professor with the Amrita School of AI, Amrita Vishwa Vidyapeetham, Coimbatore. He has published several research articles in various journals and attended several conferences. His current research interests include computational and data-driven algorithms, wide-area monitoring systems, power oscillation studies, power grid resilience, smart grids, electric vehicle technology, and AI applications in power systems. He is an Active Member of the IEEE Power and Energy Society. He was honored with the 2019 Young Innovator Award by the Kerala State Government; the 2022 Outstanding Engineer Award from IEEE PES HQ; and the Outstanding Chapter Volunteer Award from IEEE Kerala Section, in 2023.



**MOHAN LAL KOLHE** is currently a Professor of smart grid and renewable energy with the Faculty of Engineering and Sciences, University of Agder, Norway. He is a leading renewable energy technologist with three decades of the academic experience at an international level. He has held various academic positions at prestigious universities, such as University College London; the University of Dundee, U.K.; the University of Jyvaskyla, Finland; and the Hydrogen Research Institute, Trois-Rivières, QC, Canada. He was also a Board Member of the Government of South Australia's Renewable Energy Board (2009–2011) as an Advisor on renewable energy policies. He is an Expert Evaluator of projects for funding at various international research councils, such as the European Commission: Erasmus Higher Education Capacity Building; the Royal Society London, U.K.; and the Engineering and Physical Sciences Research Council (EPSRC U.K.).