DTU Library

Wind Power Forecasting Using LSTM Incorporating Fourier Transformation based Denoising Technique

Kahraman, Aysegül; Yang, Guangya; Hou, Peng

Published in:
Proceedings of Wind Integration Workshop 2021

Publication date:
2021

Document Version
Peer reviewed version

Wind Power Forecasting Using LSTM Incorporating Fourier Transformation based Denoising Technique

Aysegul Kahraman
Department of Electrical Engineering[1]
Technical University of Denmark[1]
Sino-Danish Center for Education and Research[2]
University of Chinese Academy of Sciences[2]
Kgs. Lyngby, Denmark
ayska@elektro.dtu.dk

Guangya Yang
Department of Electrical Engineering[1]
Technical University of Denmark[1]
Kgs. Lyngby, Denmark
gyy@elektro.dtu.dk

Peng Hou
SEWPG European Innovation Center[3]
Aarhus, Denmark
houpeng@shanghai-electric.com

***Abstract*—Forecasting of wind power is necessary to remove the system operational uncertainties via ensuring more reliable inputs for power system scheduling and control. In order to achieve accurate forecast of wind power for up to 50 seconds (very short-term forecasting), this paper proposes the method Fourier Denoising combined with Long Short-Term Memory (FD-LSTM). The FD-LSTM cascades the output of the Fast Fourier Transform (FFT)-based denoising algorithm to the input of the Long Short-Term Memory (LSTM) forecaster. In this method, first, the inclusion of the FFT-based denoising algorithm ensures better and balanced performance for different forecasting horizons by removing the frequencies. Next, short-term forecasting by employing LSTM reduces the uncertainty and improves both the quality of the operation and the planning. In this paper, we first evaluate the performance of the proposed FD-LSTM method on increased forecasting horizons based on Mean Square Error (MSE), Mean Absolute Error (MAPE), and R-Squared, and compare the results against linear regression. Afterwards, the method is tested with data sets where false data is present. The results show that FD-LSTM outperforms the other forecasting methods under the presence of false data.**

Keywords—Fourier transform, long short-term memory, wind power, forecasting, false data

I. INTRODUCTION

With global concerns on climate change and the increasing energy consumption, deployment of renewable energy has become a necessity for the environment. This may cause an increased interest in microgrids and smart grids, which represent key technologies to integrate renewable energy sources. Despite this interest, the intermittent characteristics of renewable sources may cause operational and planning problems. In this context, wind energy stands out with its high potential [1]. However, as wind energy system is usually connected to the main power system, in order to ensure the safe, stable, and economical operation of the electric system, accurate prediction of the generated wind energy is essential.

There have been many studies on wind forecasting in the literature. In order to forecast wind power, most of the recent works apply deep learning techniques by employing wind and weather data sets. In general, the studies are focused on hourly forecasts. While according to the forecasting horizon, it is divided into three main categories as short, mid-, and long-term forecasts. Long-term forecasts usually focus on major strategic decisions, while mid-term forecasts deal with minor strategic decisions and operational issues that vary from month to year. Short-term forecasts, on the other hand, are completely related to short-term system planning and operational issues, and the horizon is varying from days to minutes [2].

With the focus on smart grid operations, ultra-short-term forecast also has become important. Statistical methods such as exponential smoothing, regression models, Box-Jenkins, and Auto-Regressive Integrated Moving Average (ARIMA), which were frequently used in the beginning, presented good results in the field [2]. However, in the last decades, Artificial Intelligence (AI) based methods mainly Neural Networks (NN) have started to attract great attention. Furthermore, hybrid models have been created by combining AI and statistical methods to achieve more accurate results. In [3], ARIMA, ANN, and the hybrid version of ARIMA and ANN were applied and the hybrid method performed significantly better than the others. In [4], the authors formulated the wind power generation problem as a regression problem using linear regression, Support Vector Regression (SVR) and k-Nearest Neighbor Regression, by hiring only historical power measurements and the highest accuracy was achieved with the SVR method. In general, features are selected as a combination of weather data and previous wind energy records.

In [5], by applying wavelet transform, high and low-frequency data were separated and then combined with NN, short-term wind power generation was estimated. Although hybrid methods are known to improve forecasting performance, this may change as the forecasting horizon changes. In [6], wind speed and power are forecasted for different horizons by ARIMA, ANN, SVM methods and their various combinations. Although ARIMA-ANN gives better results in general, it has been noted that individual methods also give better results at different horizon windows. Within the comparison of ARIMA-Kalman and ARIMA-ANN, it was reported that ARIMA-Kalman is slightly better than others with better tracking for wind speed estimation from one-step to three steps [7]. Miguel and Rui have compared various ANN-based methods and outperformed the benchmark for one-step ahead wind power forecast [8]. Overall, it is difficult to find a single architecture fit for all forecasting problems. The length and quality of the dataset, prediction horizon, feature selection, and the forecasting methods are correlated and jointly determines the overall performance.

With the improvement AI-based hybrid methods, preprocessing and post processing methods have also been thoroughly developed. In [9], a novel hybrid short-term power forecasting method was presented by introducing correction of Numerical Weather Prediction (NWP) and was applied with both NWP data and spatial correlations. Considering microgrid, forecasting performance was improved for load demand, solar, and wind power generation using the fuzzy prediction interval model [10]. Incorporating meteorological data for renewable energy forecast may also cause breakdown or malfunction if the data is not accurate. The study in [11] makes compared solar power forecasting by applying six different methods, where temperature forecast data was not used as a feature as it was found little impact on the forecast performance. Similarly, wind speed data as input reduced the overall forecast performance with Bi-LSTM [12].

If there is no additional available weather data with historical wind generation, Long Short-Term Memory (LSTM) which is a specific type of Recurrent Neural Networks (RNN) performs well for long-term dependencies. In addition, having measurements within shortened time intervals bring certain levels of noise and denoising techniques are applied to improve the forecasting performance. With these circumstances, it becomes necessary to investigate the preprocessing methods combined with LSTM to investigate multi-step forecasting performance in an increasing horizon with very short intervals. Related to this, low and high-frequency components of the PV power data have been decoupled via fast Fourier decomposition and a convolutional neural network was implemented for forecasting of each component separately where the result is reconstructed from the individual forecast. The results show that Frequency-Domain Decomposition (FDD) is helping to achieve better performance for ultra-short-term PV power prediction [13].

In this paper, we present an LSTM-based forecasting model for a univariate wind power generation dataset with serial Fast Fourier Transform and demonstrate multi-step forecasting capability under false data injection attacks.

II. METHODOLOGY

Historical data plays a vital role in forecasting. Accurate predictions are possible only if with the right forecasting model as well as quality dataset. However, there may be inevitably inaccurate or missing samples in the dataset. In such cases, data processing is widely used as it contributes to the creation of clean and valid data sets. Finding and removing outliers and anomalies from the dataset can be accomplished in different ways. However, potential problems with datasets are not limited to these. With the development of the smart grid, the operation relies more on measurement and communication infrastructure, which makes the system more vulnerable to cyber-attacks. One of the attacks is false and inaccurate data injection to mislead the operation and control decisions [14]. Since cyber-attacks can cause not only economic but also the quality and physical damage with the grid overload and instability, studies on these attacks have recently gained attention. Machine Learning techniques can recognize and catch such attacks and respond quickly without interrupting any operation [15].

In this study, false data injection attacks are generated during the training and testing of the forecasting algorithm, and forecasting is performed with the proposed FD-LSTM and compared with linear regression performance. The general flow of the paper is given in Fig. 1.

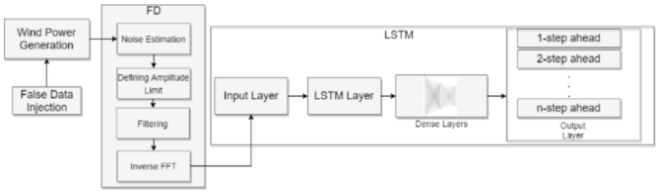

Fig. 1. The flow of the proposed model.

A. *Fast Fourier Transform*

Time series may have various frequencies and characteristics. Fourier Transform can decompose the function from time domain to the frequency domain. Fourier Transform is discretized to be used in practical applications and called Discrete Fourier Transform (DFT). The formula of DFT is given as follow [15]:

$$X_k = \sum_{n=1}^{N-1} x_n \left[cos(2\pi kn/N) - i.sin(2\pi kn/N)\right] \quad (1)$$

where, N is number of samples, n represents current sample, k is current frequency and x_n is the input sequence.

DFT performs successfully but also includes outrageous calculations based on its definitions. Fast Fourier Transformation (FFT) is computationally less time-consuming and fast when it is compared with DFT and FFT is one of the most used and powerful algorithms in time series data analysis [16]. Regardless of how complex or simple the function is, it helps to de-noise the data and keep only the essential and necessary components for the next steps. Frequency shows variation depending on application and preferences, higher frequency often introduces more noise, which makes FFT application an initial step for forecasting studies. FFT declares that each individual can be decomposed into a sum of different sinusoidal signals.

As wind power generation is uncertain and seasonal, studies have been focusing on identifying the seasonality to achieve more accuracy in forecast. In this study, uncertainty and trend are emphasized due to the high frequency time series data. In addition to capturing the trend, we also aim to eliminate undesired and false frequencies.

The FFT procedure starts with decomposing data with the Fourier transform, calculating power density while the data is in the frequency domain, setting a boundary to filter out the noise, and finally reconstructing the data by taking inverse FFT to return to the time-domain. This provides noise-free wind power that is ready to be used in forecasting.

B. *LSTM*

Long Short-Term Memory (LSTM) is a special type of Recurrent Neural Networks (RNNs) which are dynamic and more complex than feed-forward neural networks in terms of keeping the necessary historical information by using special hidden units. LSTM networks are obtained by replacing hidden units in the RNN with memory gates. The main responsibility of memory cells in the LSTM network is using the new input vector. LSTM has 3 special gates, forget (f_t), input (i_t), and output gates (o_t), which help to decide the irrelevant information, take the current state and valuable information from the past, and decide to what should be next hidden state, respectively. One of the intermediate states is called a cell state and it is calculated by Eq. 2. To update the state, the multiplication of old state with forget gate which is 0 or 1 and sum with the multiplication of new information $\widetilde{C_t}$ with input gate for detecting its importance.

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \tag{2}$$

One of the main equations of the LSTM which calculates the output of the network is Eq 3.

$$h_t = o_t \odot \tanh(C_t) \tag{3}$$

where, h_t is output and C_t is a state value. There is three main activation function in the literature: ReLU, sigmoid, and tanh. Rectifier Linear Unit (ReLU) is preferred in many neural networks, it works between 0 and z, z is a positive real number. While sigmoid returns a value between 0 and 1, tanh works between -1 and 1 in deep learning literature. When the data set is normalized as between 0-1, the sigmoid function becomes a convenient type for implementation. Although many deep learning studies employ ReLU function to manage the vanishing gradient problem by increasing the depth or number of parameters for the network, LSTM can handle this issue without being affected by the change of activation function. Based on the proposed technique, it seems possible to use both ReLU and sigmoid activation functions. However, since it is desired to observe the forecast performance of the LSTM network against false data insertion attacks, the data set is given to the network without normalization in order not to reduce the effect of false data and to preserve its discernibility by the network, thus, ReLU activation function is preferred.

C. Forecasting Performance Evaluation

The performance of regression problem can be evaluated by comparing the actual value and the forecasted value for the overall goal of evaluating how good and appropriate to this data set. Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE) and R-squared (R^2) are generally applied for performance evaluation. However, we criticized the network performance for various steps ahead, instead of having only an average error metric or one-step evaluation. It is possible to examine and compare the forecast performance separately for each step from one to n steps.

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_j - \widehat{y_j}| \tag{4}$$

$$MSE = \frac{1}{n}\sum_{j=1}^{n} \left(y_j - \widehat{y_j}\right)^2 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n} \left(y_j - \widehat{y_j}\right)^2} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{j=1}^{n}\left(y_j - \widehat{y_j}\right)^2}{\sum_{j=1}^{n}\left(y_j - \overline{y_j}\right)^2} \tag{7}$$

$$R_{adj}^2 = 1 - \frac{(1 - R^2)(n-1)}{n-k-1} \tag{8}$$

where y_j and $\widehat{y_j}$ are the actual and estimated values of the wind power generation. n and k represent row and column. While our model tries to decrease all the MAE, MSE, RMSE, MAPE, it tries to maximize R^2 and R_{adj}^2.

III. Case Study

The dataset used in this study belongs to daily wind generation power data for a day. The frequency of the data set is 5 seconds and the total number of samples is 17352.

Besides applying the proposed FD-LSTM to the wind energy dataset, another goal of the study is to evaluate the impact of cyber-attacks on smart grid operations related to false data insertion and the overall prediction performance of deep learning against these possible attacks. For this reason, random anomalies were created based on a uniform distribution in the data set that could occur during possible attacks, estimation methods were completed by FD-LSM and linear regression methods for both this data set and the data set that were not exposed to any attack, and the results are presented below.

In forecast applications, it is prevalent to apply different normalization techniques. They are fed into the input layer of an LSTM network, which is designed to more simply train the network and bring all features into a similar range, especially when you have more than one feature. However, in this study, we only have one feature, which is the previous electricity generation; besides, we aim to keep the effect of false data injection during both training and testing to show the forecasting performance against these kinds of attacks. For the stated reasons, the data set was used on its own scale without being normalized.

The procedure of FD-LSTM begins with the FFT application. With taking the FFT, data turns into more smooth and dispose of noise and inserted artificial data. The LSTM has one dense layer which follows the input layer with 64 and 32 neurons, respectively. The output layer presents more than one step ahead; here it is chosen as 10 steps; since it is also beneficial for the representation of forecasting horizons based on various horizons. The case study is investigated under these 2 circumstances; (1) data with the cyber-attacks and (2) data without cyber-attacks. The linear regression and FD-LSTM results are presented separately. For the first case, we assume that the attacks are corresponding 1% of the total data based on uniform distribution between defined lower and upper boundaries which are 50 and 250. The wind power data before and after the attacks are plotted in Fig. 2.

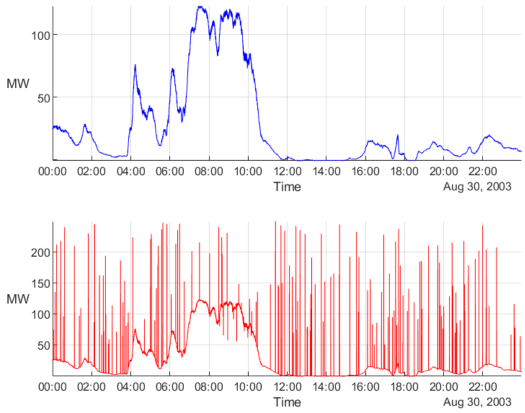

Fig. 2. The wind data set before and after the attacks.

The data set is divided into 80%, 10%, and 10% for train, validation, and test data set. The forecasting horizon is selected as 10, the error change for each step-ahead is observed individually. Since the data does not include any other meteorological data or different features, power generation is taken as the only feature and 10 previous steps are fed to the network to forecast the next 10 steps-ahead. It is possible to feed more lags; however, a high number of lags can cause less relevant inputs to be included in the network and reach complexity, which will degrade the performance of the network. Although wind power generation is quite uncertain and we do not have different features, it is possible to obtain high forecasting performance using only historical data, since this study focuses on demonstrating the ability of LSTM architecture against external attacks. The model is trained by using the data between the current time and N previous lags and is aimed to predict the next N steps. The forecasting of the next 10 steps can be expressed as a function of $f(X_{t+1}, X_{t+2}, X_{t+3}, \dots, X_{t+N}) =$

$$\{f(X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-N})\}.$$

Before the testing, training and validation data is shuffled to increase the prediction performance. The optimizer is Adam, the loss function is MSE, the batch size is 8 and the number of epochs is 15.

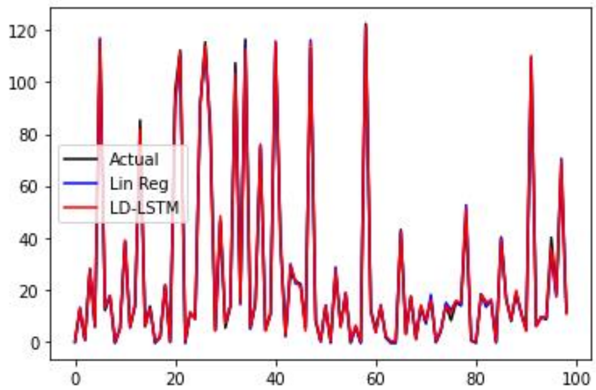

Fig. 3. Forecasting results of the first 100 data in the test set with the regression and LSTM methods and their actual values.

In the first case, a network is created according to the mentioned above, but it is assumed that there is no cyber-attack initially. The training is completed and forecast results that is for the next 10 steps ahead are obtained. The estimated and actual values for each step are compared for the proposed FD-LSTM and the regression method, and it is observed that the results are extremely close and have high accuracy. Fig. 3. illustrates the first 100 values of the results obtained for step 9. Fig. 4. represents the results of performance criteria based on the test set for each step separately.

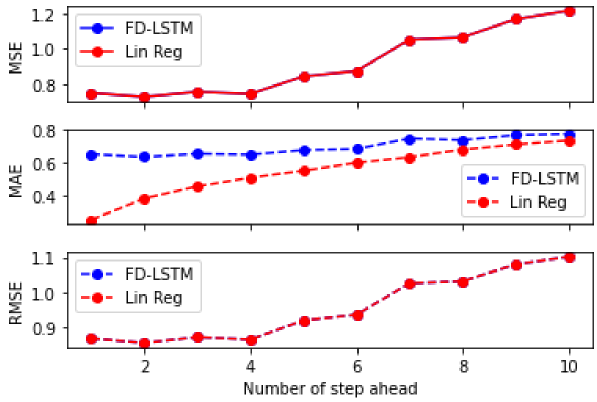

Fig. 4. Forecasting performance for the 10 steps-ahead with the regression and LSTM methods.

It should be underpinned that MAE increases rapidly for further steps compared to FD-LSTM for linear regression.

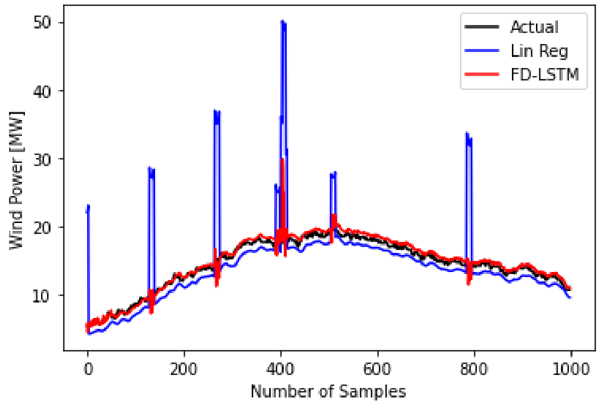

Fig. 5. Forecasting results of the first 1000 data in the test set with the regression and FD-LSTM methods under the attack and their actual values.

It can be realized that with the high-frequency time series data both the analytic and deep learning method works well, in this case, there might not be necessary to employ deep learning techniques since it is easier and faster to implement statistical methods. However, these methods cannot deal with damaged data, often data preprocessing helps to catch and clear it, but recently cyber-attacks for inserting false data have made it difficult to catch these anomalies and it has become a real problem. When the data is healthy, linear regression might be even a better option by having less complexity and computationally less expensive. False data attack is simulated and with its presence the forecasting results are compared with actual ones in Fig. 5 and Fig. 6. It shows that the proposed method is better at catching peaks that come with attacks, and although it cannot completely ignore these peaks, it gives more reasonable responses. In general, the result from FD-LSTM is closer to the actual data than from linear regression. Here, it is different than the first case because MAE is not presenting a major difference for further steps ahead.

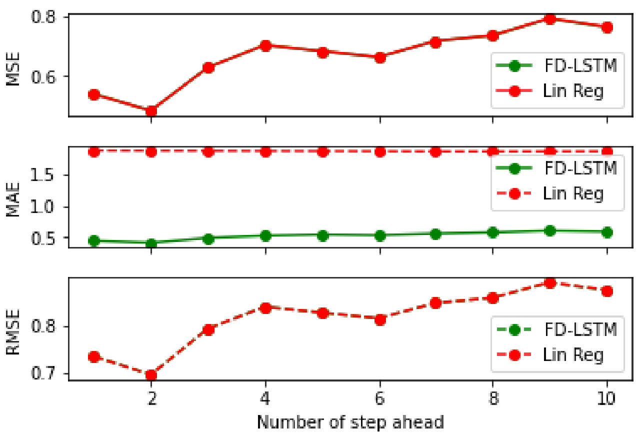

Fig. 6. Forecasting performance for the 10 steps-ahead with the regression and FD-LSTM methods with the false data insertion.

A similar conclusion can be extracted from Table 1 which gives the average of all steps for each case study.

TABLE 1. The change in performance evaluations for proposed FD-LSTM and linear regression under the attacks.

	Without Attacks		**With Attacks**	
	FD-LSTM	Linear Regression	FD-LSTM	Linear Regression
MAE	0.6963	0.5493	0.5296	1.871
RMSE	0.9597	0.9705	0.8192	3.839
R-squared	0.9992	0.9991	0.9572	0.06192
Adjusted R^2	0.9992	0.9991	0.9572	0.06125

Although linear regression performs slightly better in predicting wind energy without attacks, it turns out that this model may not be suitable when cyber-attacks are present. FD-LSTM gives more balanced results in both cases and better in dealing with attacks for its considerable good performance.

IV. Conclusion

Based on current developments in forecasting studies, FD-LSTM is proposed to achieve more accurate forecasting of wind power generation with false data injection attacks. The evaluation criteria illustrated that LSTM has a high potential for dealing with false data injection attacks, as it can analyze previous attacks and perform very well during the test phase. This study shows that (1) cyber-attacks jeopardize the traditional model performance while more advanced models such as LSTM are better to withstand them with good performance (2) FFT is helping the forecasting model to get better results under false data inserting attacks. (3) LSTM can be designed specifically for different application contexts. The proposed FD-LSTM method is verified using MAE, MSE, RMSE, R^2 and R^2_{adj} which provides significantly better results than the benchmark ones. Therefore, by reasonably integrating statistical methods with deep learning methods, forecasting performance can be improved against specific conditions and cyber-attacks.

References

[1] Z. Tasneem, et al., "An analytical review on the evaluation of wind resource and wind turbine for urban application: Prospect and challenges," Developments in the Built Environment, 100033, 2020.

[2] S. Thomassey, "Intelligent demand forecasting systems for fast fashion," in Information systems for the fashion and apparel industry (pp. 145-161), Woodhead Publishing, 2016.

[3] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, 50, 2003, pp. 159-175.

[4] N. A. Treiber, J. Heinermann, and O. Kramer, "Wind power prediction with machine learning," in Computational sustainability, Springer, Cham, 2016, pp. 13-29.

[5] J. P. D. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "Short-term wind power forecasting in Portugal by neural networks and wavelet transform," Renewable energy, 36(4), 2011, pp. 1245-1251.

[6] J. Shi, J. Guo, and S. Zheng, "Evaluation of hybrid forecasting approaches for wind speed and power generation time series," Renewable and Sustainable Energy Reviews, 16(5), 2012, pp. 3471-3480.

[7] H. Liu, H. Q. Tian, and Y. F. Li, "Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction," Applied Energy, 98, 2012, pp. 415-424.

[8] M. Godinho, and R. Castro, "Comparative performance of AI methods for wind power forecast in Portugal," Wind Energy, 24(1), 2021, pp. 39-53.

[9] S. Hu, et al., "Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction," Applied Energy, 293, 116951, 2021.

[10] D. Sáez, F., Ávila, D. Olivares, C. Cañizares, and L. Marín, "Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids," IEEE Transactions on Smart Grid, 6(2), 2014, pp. 548-556.

[11] L. Gigoni, et al., "Day-ahead hourly forecasting of power generation from photovoltaic plants," IEEE Transactions on Sustainable Energy, 9(2), 2017, pp. 831-842.

[12] M. Ko, et al., "Deep Concatenated Residual Network With Bidirectional LSTM for One-Hour-Ahead Wind Power Forecasting," IEEE Transactions on Sustainable Energy, 12(2), 2020, pp. 1321-1335.

[13] J. Yan, et al. "Frequency-Domain Decomposition and Deep Learning Based Solar PV Power Ultra-Short-Term Forecasting Model," IEEE Transactions on Industry Applications, 2021.

[14] E. Drayer, and T. Routtenberg, "Detection of false data injection attacks in smart grids based on graph signal processing," IEEE Systems Journal, 14(2), 2019, pp. 1886-1896.

[15] A. Sayghe, et al., "Survey of machine learning methods for detecting false data injection attacks in power systems," IET Smart Grid, 3(5), 2020, pp. 581-595.

[16] T. Fischer-Cripps, "Newnes interfacing companion: computers, transducers, instrumentation and signal processing," Elsevier, 2002.