NONLINEAR AUTOREGRESSIVE NEURAL NETWORK FOR WIND POWER GENERATION FORECASTING

[1]RAHINA MOMMY SA'ADU, [2]HAKEEM A. SULAIMON, [3]MUHAMMAD AMINU AHMAD, [4]MUSTAPHA LAWAL ABDULRAHMAN

[1,3]Department of Computer Science, Kaduna State UniversityKaduna State, Nigeria
[2]Department of Computer Science, Federal College of Education, ZariaKaduna State, Nigeri.
[4]National Center for Remote sensing Jos, Rizek Village, Plateau State

Corresponding Author:rahinasaadu@gmail.com

ABSTRACT

Wind energy is increasingly being utilized globally, in part as it is one of the renewable energy sources characterized by the lowest cost of electricity production and has experienced a significant expansion in installed capacity in recent years. Hence forecasting wind behavior, For example, wind speed is important for energy managers and electricity traders, to overcome the risk of unpredictability when using wind energy. One of the challenges in integrating wind power into the electrical grid is its intermittency. One approach to deal with wind intermittency is forecasting future values of wind power production. But these models rely only on historical data which is still not convincible enough. Moreover, coping with nonlinear time series data in forecasting medium to long term with high accuracy is still a challenging task. In this paper, Nonlinear Autoregressive Neural Network - Long-Short-Term Memory (NARX- LSTM) model for forecasting wind power has been proposed. This model is able to determine long-term dependencies in the context of wind power predictions. The proposed model has Root Mean Square Error of 0.16000 and MAPE of 0.23000. The prediction accuracy of the model is better than the existing models including BPNN with RMSE of 23.877 and MAPE of 22.032, RBFNN which has RMSE of 18.729 and MAPE of 16.735, and ARIMA-RBF which has RMSE of 3.00 and MAPE of 2.659. The paper concluded that in terms of correlation error R, the proposed model obtains a regression value close to 1, which clearly interprets that the model perfectly fits the data used in both the training, testing, and validation phase.

Keywords: Artificial Neural Network, Machine Learning, Deep Learning, Recurrent Neural Network, Nonlinear Autoregressive Neural Network

INTRODUCTION

Wind energy has been an important part of the electricity market in every country around the world because it offers many advantages, including clean green energy and low prices and it does not produce emissions that cause acid rain or greenhouse gases. Wind energy is a vital source of renewable energy with large reserves and wide distribution (Liu *et al*., 2019).The wind power produced by a wind farm critically depends on the stochastic nature of the wind speed, and unexpected variations in the wind power output increase the operational cost of the electricity system (Zhou *et al*., 2019). Thus, wind power or wind speed forecasting methods have been reported in the literature over the past few years to improve forecast accuracy.According to Mocanu *et al*., (2016) forecasting can be grouped into either one of these three groups, they include (i) short term forecast usually ranging from day-week (ii) medium-term forecast usually ranging from week-year, and (iii) long-term forecast usually ranging from a year and above. These methods as presented by Hong and Rioflorido, (2019), which are Physical methods that include Numerical Weather Prediction (NWP). NWP uses hydro- and thermo-dynamic models of the atmosphere to predict the weather, considering initial values and boundary conditions. In case the accuracy of NWP is poor, the wind power generation forecasting becomes inaccurate. The statistical methods include probability mass bias, probabilistic auto-regression, vector autoregressive model, and Bayesian framework. These methods concern the relationship between wind power generation and explanatory variables.

The traditional artificial neural networks (ANN) are used for predicting wind power; they include the multi-feed-forward neural network (MFNN), the Radial Basis Function Neural Network (RBFNN) (Chang *et al*., 2016), the wavelet neural network (WNN), the extreme learning machine (ELM) (Zhou, Yu, and Jin, 2018) and Elman recurrent neural network (Wang *et al*., 2018). The advantage of these methods is that they require no predefined mathematical model. Commonly used methods in the networks are auto-encoders, long-short-term memory (LSTM) (Fu, Hu, Tang, Yu, and Liu, 2018), the restricted Boltzmann machine (RBM) (Santhosh, Venkaiah, and Kumar, 2019), and the Convolutional Neural Network (CNN) (Hong and Rioflorido, 2019). Long-Short-Term Memory (LSTM) is an artificial recurrent neural network architecture used in the field of deep learning. It has feedback connections, and it can process the entire sequence of data such as images, speech, and video. It is also well-suited to classifying, processing, and making predictions based on time series data.

Nonlinear Autoregressive Neural Network (NARX) is a recurrent dynamic network, with feedback connections enclosing several layers of the network making it suitable for time series forecasting. Time series is a sequence of data, numerical values, or observations usually recorded at constant time intervals (Brockwell, Davis, and Calder, 2002). It is measured normally every second, minute, hour, day, and a week or in some cases year. Compared with the traditional ANNs, deep learning neural networks do not need extra unsupervised networks or data (signal) preprocessing, for example, decomposition. An artificial neural network is a computing intelligence that is made up of neurons. It is inspired by the biological neural network. Activation of each neuron is computed, and the response is sent over the output node of neurons. The traditional ANN is consisting of three main layers namely, the input, the hidden, and output layers. Each of these layers is comprised of multiple neurons with an activation function. There is no systematic framework for choosing the numbers of hidden layers although three categories of parameters are implored to define ANN. They include the connection style between different layers of neurons. Deep learning neural networks outperform the traditional neural network in renewable power forecasting problems.

On the other hand, ensemble models have also attained global attention in recent years. Nowadays, around 90% of the developed wind speed and power forecasting approaches are ensemble models. These hybrid models can be implemented by combining the superior features of the individual models. In general, hybrid models have proved to achieve high accuracy particularly short-term forecast, for example, an hour by Qing and Niu, (2018) to a week by Wafune *et al*. (2014). But there are few limited works on medium to long term prediction with previous work showing errors above 40-50% as regards medium to long term forecasting (Yun *et al*., (2012). This paper implements a hybrid neural network model that can forecast medium to long-term forecasting while focusing on improving the forecasting performance, Thus, this paper puts forward a deep learning approach to enhance the prediction performance using NARX and LSTM for the prediction of air pollution concentration in smart cities. The remainder of this paper is divided as follows, section 2 provides the review of related work, section 3 presents the method used in this research work, section 4 discuss the implementation and results and finally section 5 concludes the research findings and proffer future pathways

MATERIALS AND METHODS

The major concern of this work is to improve on the prediction accuracy of the existing system by developing a model using a NARX and LSTM for the forecasting of wind power generation. The overall strategy is implemented in two (2) stages. To implement the model, the proposed framework of the NARX- LSTM model is described as shown in figure 1 below.

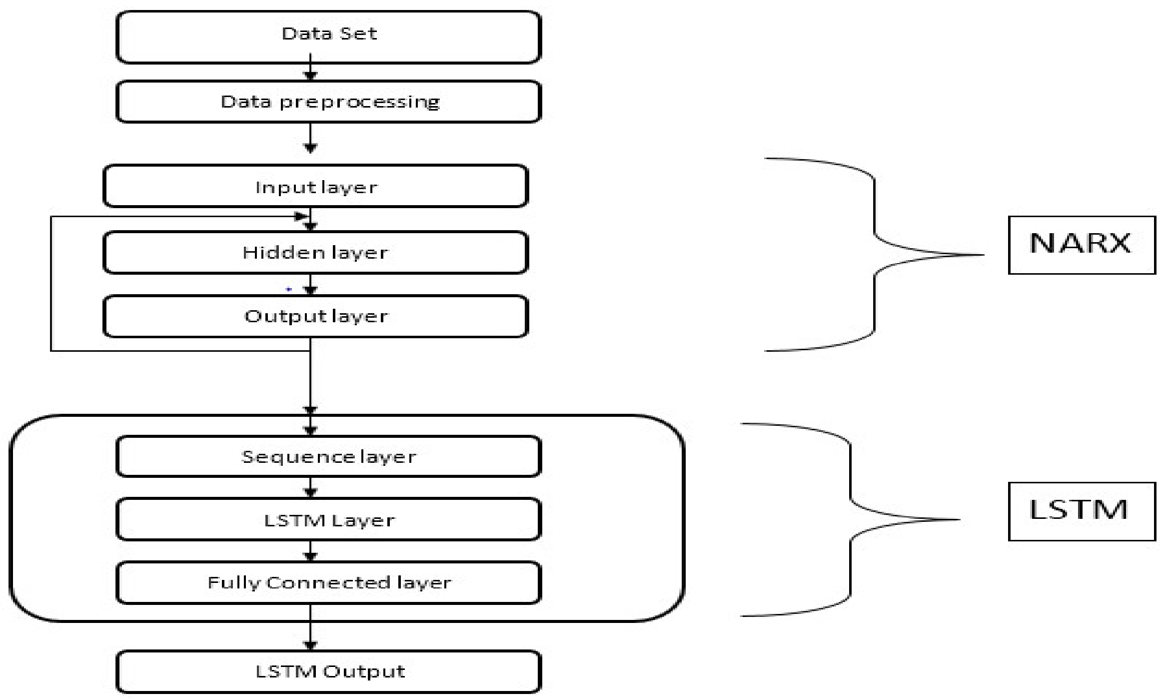

Figure 1: Proposed Frameworks of NARX-LSTM

The framework consists of three main stages which are further explained in the following sub-sections for a better understanding of how the system works to achieve the main objective.

Model Development

Stage 1: Data set Collection and Preprocessing Stage

The Data set collection stage involves the identification of predictor variables and Data collection stage. After identification of the relevant input variables, the Data set is collected from the official sources and saved in the appropriate data format (.xlsx). The Data set was pre-processed by carrying out Data cleaning/Scrubbing to remove typographical errors and inconsistencies in the data. We then saved the Data set in the format required by MATLAB through the Data formatting phase. We carry out a Data normalization scheme using the minmax (-1 to 1) to normalize the dataset for fast convergence.

Stage 2: Nonlinear Autoregressive Neural Network with Exogenous Inputs Stage

The NARX Stage is the stage built on the Sequence Layer of the LSTM Stage. In this experiment, the normalized data set is fed as input to the NARX input layer which is the topmost layer of the NARX network. A sequence input layer inputs sequence data to a network for processing and training to achieve the desired forecast accuracy. The NARX network is composed of three layers which are the input, hidden, and an output layer with feedback connections from the output to the hidden layer. The NARX stage is a dynamic recurrent neural network with feedback from the Output layer connected to the input of the hidden layer. The Input and feedback inputs are fed via input and feedback delays. The Figure 2 below shows the NARX Network used in this research.

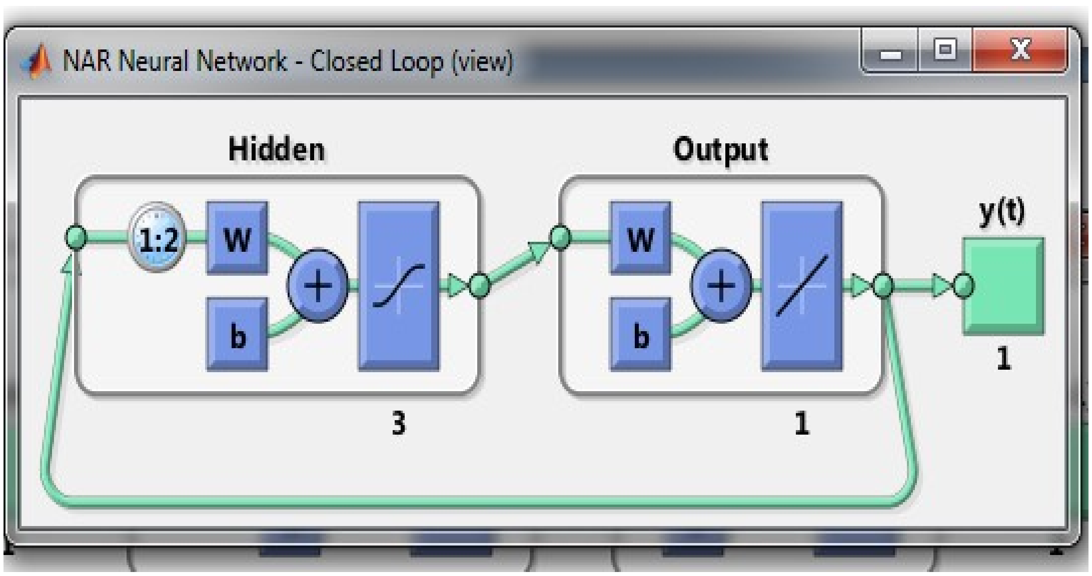

Figure 2: NAR Neural Network Closed Loop

Stage 3: LSTM Stage

The LSTM layer in our work learns long-term dependencies between time steps of sequence data. This layer also models the time series and sequence data in the network and performs additive interactions then configures the next layer to be a Fully Connected Layer that combines all the features learned by the sequence and LSTM layers to identify the larger patterns in the data set to make the forecast. Hence, the Output Size parameter in the last fully connected layer was equated to the number of classes in the target data. In this experiment, The LSTM Output Layer also known as the Regression Output layer holds the loss function used for training the network for regression, and the response purpose.

Data Collection

The measured wind speed and wind power data for training the model are taken from an actual wind farm in Taiwan Power Company at a 10-min resolution obtain from the Kaggle repository portal in excel file format. The data set contain wind power and weather forecast data for a period of 1st January 2012 to 31 December 2013. The columns contain the following: TIME STAMP: 10-min resolution. TARGETVAR: Hourly-mean wind power is normalized by the maximum output of the wind farm. U10: Forecast zonal wind velocity (m/s) at 10m above ground. V10: Forecast meridional wind velocity (m/s) at 10m above ground. U100: Forecast zonal wind velocity (m/s) at 100m above ground and V100: Forecast meridional wind velocity (m/s) at 100m above ground

Performance Metrics

In this experiment, a Standard Statistical forecasting metrics are used to evaluate the network performance in other to select the model with the best performance. The developed network is tested and evaluated by using the Mean Square Error and Root Mean Square Error in the MATLAB Neural Network (NN-tools) package. Also, Mean Average Percentage Error (MAPE) and correlation error (R) was used to ascertain the model's prediction accuracy.

The MSE, RMSE and MAPE is computed in the formula below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}} \quad (1)$$

Where,

Predicted: value predicted by the model, Actual: observed values and N: Total number of observations

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \hat{Y}_i|}{Y_i} \quad (2)$$

Where; y_i: value predicted by the model, $\acute{y}_i$: observed values and n: Total number of observations

RESULTS AND DISCUSSION

In this paper, the accuracy of the estimated forecasts of the proposed model will be compared with the other models to ascertain which model gives a more accurate forecast through the use of RMSE, MSE, and correlation error (R). The result will now be presented the result based on forecasting using all the algorithms and present the results by the performance standard of each of the model and discuss the findings. From the results of our simulation, we obtain the following plots which will demonstrate how perfectly fit is our model. To validate the developed network, the Error histogram was used. Figure 3 shows the error histogram for the fit set of the proposed model. It shows how the error sizes are slightly well distributed. When most errors are near zero, it has been observed a better-trained model. In this case, however, it is confirmed that the network also has errors near zero.

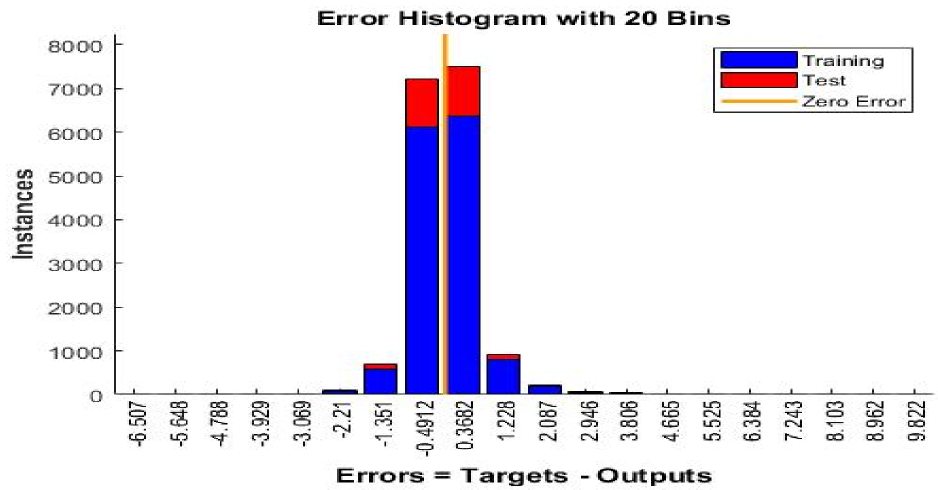

Figure 3: Error histogram for the fit set

The training algorithm used in this research work to train the proposed model is Bayesian regularization back propagation algorithm which converged after 290 epochs, and it showed stability (no increase after converging) and no overshoot (no increase before converging), as shown in Figure 4.

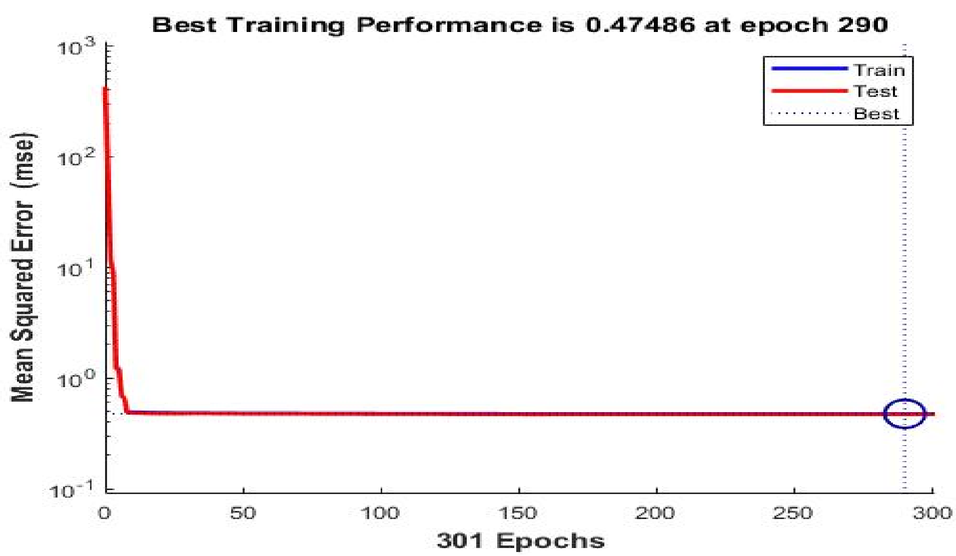

Figure 4: Training Performance for the proposed NARX-LSTM

Figure 5 shows the regression graph for the proposed model; this shows how accurately the proposed trained model fits the dataset. It was observed that the value of R is close to 1 (Good) which demonstrates that the model prediction is very close to the actual dataset. If it was close to zero (bad) then it shows that the model completely fails in making a correct prediction. In our own case, the model prediction was close to the actual test which demonstrates a good prediction by the proposed model as shown in table 3.

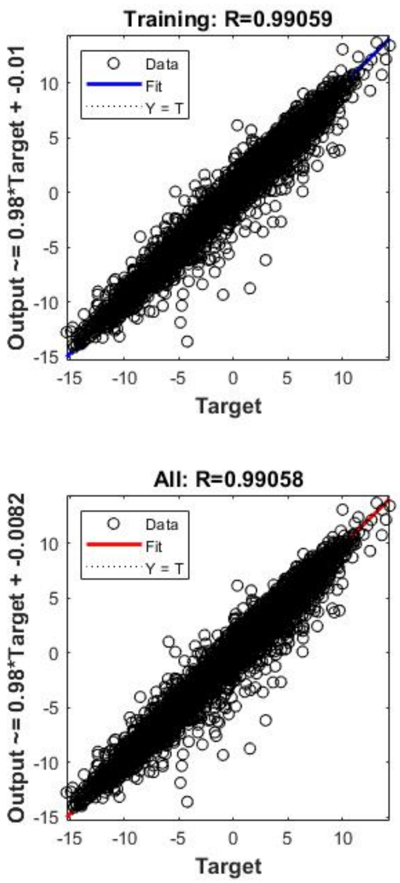

Figure 5: Error regression on forecast vs. actual

Figure 6 displays the error autocorrelation function. The figure shows how the prediction errors are related in time. Similarly, for a perfect prediction model, there should be only one nonzero value of the autocorrelation function. These nonzero values should occur at zero lag; this is also called mean square error. Such an autocorrelation function would imply a complete un-correlation of predicted errors with each other. Base on the graph as shown in Figure 6, the correlations, except for the one at zero lag, fall approximately within the 98% confidence limits around zero, thus the model appears to be suitable. If there were significant correlations in the prediction errors, then it's possible to improve and enhance the prediction accuracy maybe by changing the neural network structure or increasing the number of delays in the network.

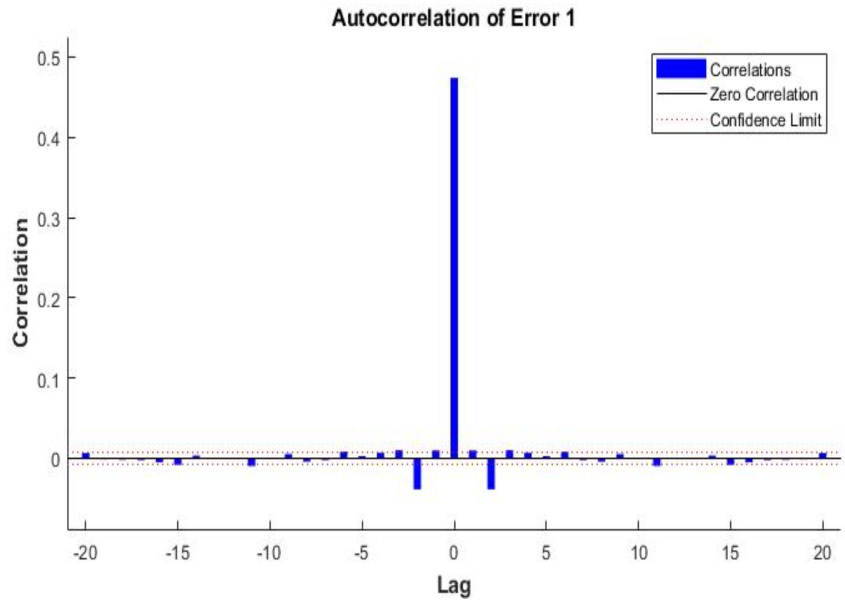

Figure 6: Autocorrelation error for the proposed model

Table 1 illustrates how the performances of the proposed hybrid approach compare with those of the existing model for the forecast of wind power generation. For the MSE, RMSE, and MAPE, the smaller the value, the better the performance.

Table 1: Illustrates how the performances of the proposed NARX- LSTM model when compared with those of the existing ARIMA-RBF, BPNN and RBFNN model in terms of training MSE, and validation MAPE and RMSE.

	NARX-LSTM	BPNN	RBFNN	ARIMA-RBF
MSE	0.47486	0.0510	1.6547	0.5843
MAPE	0.23000	22.032	16.735	2.659
RMSE	0.16000	23.877	18.729	3.133

Result Analysis in Terms of Training MSE

For the MSE, RMSE, and MAPE, the smaller the value, the better the performance, in this case, the training MSE of the proposed NARX-LSTM is 0.074205 while the training MSE of the existing BPNN is 0.051094, RBFNN is 1.65471 and the ARIMA-RBF is 0.5843. This means that the existing system achieved a better training performance against all the other models, although the training MSE of the proposed model was comparatively better than the other two models i.e RBFNN and ARIMA_RBF. Here are proposed system fails to outperform the existing BPNN in terms of training performance.

Result Analysis in Terms of Forecasting Accuracy (RMSE and MAPE)

Similarly, for the Validation RMSE and MAPE, the smaller the value the better the accuracy, in this case, the proposed NARX-LSTM model has RMSE of 0.16000 and MAPE of MAPE 0.23000 respectively. This is by far the lowest value when compared against the existing model that is. BPNN has RMSE of 23.877 and MAPE of 22.032, RBFNN which has RMSE of 18.729 and MAPE of 16.735, and ARIMA-RBF which has RMSE of 3.1330 and MAPE of 2.659 respectively. Thus, the proposed NARX-LSTM has clearly enhanced the forecasting accuracy as against all the existing models that were used for the evaluation by achieving the lowest values in terms of validation RMSE and MAPE. Thus, the proposed model demonstrates superiority in forecasting long term wind power generation model with high accuracy as against the state of the art.

CONCLUSION

The analysis of the proposed hybrid neural network model presented in this paper performs very well in forecasting wind power for medium to long-term time horizons.The proposed NARX-LSTM, in general, performs better in terms of validation RMSE, MAPE, and correlation error (R) than the existing model in the case of wind power forecasting. The proposed model was able to account for long-term dependencies in the context of wind power predictions. The proposed NARX-LSTM model has RMSE of 0.16000 and MAPE of MAPE 0.23000. This is by far the lowest value when compared against the existing model including BPNN which has RMSE of 23.877 and MAPE of 22.032, RBFNN which has RMSE of 18.729 and MAPE of 16.735, and ARIMA-RBF which has RMSE of 3.1330 and MAPE of 2.659. Thus, the proposed NARX-LSTM has clearly enhanced the forecasting accuracy against all the existing models that were used for the evaluation by achieving the lowest values in terms of RMSE and MAPE validation. Thus, the proposed model demonstrates superiority in forecasting long term wind power generation model with high accuracy as against the state of the art. Similarly, in terms of correlation error R, the proposed model obtains a regression value close to 1 which clearly interprets that the model is perfectly fit for the target data in both the training, testing, and validation phase.

REFERENCES

Brockwell, P. J., Davis, R. A., and Calder, M. V. (2002). *Introduction to time series and forecasting* (Vol. 2): Springer.

Cheng, J., and Li, Q. (2008). Reliability analysis of structures using artificial neural network based genetic algorithms. *Computer methods in applied mechanics and engineering, 197*(45-48), 3742-3750.

Fu, Y., Hu, W., Tang, M., Yu, R., and Liu, B. (2018). *Multi-step Ahead Wind Power Forecasting Based on Recurrent Neural Networks.* Paper presented at the 2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC).

Hong, Y.-Y., and Rioflorido, C. L. P. P. (2019). A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. *Applied Energy, 250*, 530-539.

Iwafune, Y., Yagita, Y., Ikegami, T., and Ogimoto, K. (2014). *Short-term forecasting of residential building load for distributed energy management.* Paper presented at the 2014 IEEE International Energy Conference (ENERGYCON).

Mocanu, E., Nguyen, P. H., Gibescu, M., and Kling, W. L. (2016). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks, 6*, 91-99.

Qing, X., and Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy, 148*, 461-468.

Santhosh, M., Venkaiah, C., and Kumar, D. V. (2019). Short-term wind speed forecasting approach using Ensemble Empirical Mode Decomposition and

Deep Boltzmann Machine. *Sustainable Energy, Grids and Networks*, 100242.

Wang, G., Jia, R., Liu, J., and Zhang, H. (2019). A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning. *Renewable Energy*.

Yun, K., Luck, R., Mago, P. J., and Cho, H. (2012). Building hourly thermal load prediction using an indexed ARX model. *Energy and Buildings, 54*, 225-233.

Zhou, Q., Wang, C., and Zhang, G. (2019). Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems. *Applied Energy, 250*, 1559-1580.