# Deep Learning Based Recurrent Neural Networks to Enhance the Performance of Wind Energy Forecasting: A Review

Senthil Kumar Paramasivan

School of Information Technology & Engineering, Vellore Institute of Technology, Vellore 632014, India

Corresponding Author Email: psenthilkumar@vit.ac.in

**ABSTRACT**

In the modern era, deep learning is a powerful technique in the field of wind energy forecasting. The deep neural network effectively handles the seasonal variation and uncertainty characteristics of wind speed by proper structural design, objective function optimization, and feature learning. The present paper focuses on the critical analysis of wind energy forecasting using deep learning based Recurrent neural networks (RNN) models. It explores RNN and its variants, such as simple RNN, Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional RNN models. The recurrent neural network processes the input time series data sequentially and captures well the temporal dependencies exist in the successive input data. This review investigates the RNN models of wind energy forecasting, the data sources utilized, and the performance achieved in terms of the error measures. The overall review shows that the deep learning based RNN improves the performance of wind energy forecasting compared to the conventional techniques.

## 1. INTRODUCTION

In the modern era, the wind energy is attracted by many companies for power generation. It is more competitive due to the economic and cost-effective manner compared to traditional power generation. Due to the clean, green, and naturally replenished characteristics of renewable energy, it acts as a promising alternative to fossil fuels such as natural gas, oil, and coal. The reliability and stability of the energy systems depend on the proper scheduling of the energy generation. However, the uncertain nature of the renewable energy imposes issues in reliability and stability of energy systems. The wind energy, biomass energy, solar energy, geothermal energy, and hydropower are the existing renewable energy sources in the world. Among the number of renewable energy sources available, the wind energy source plays an important role in producing power, and it is the rapidly growing wind energy farm. From the data published by World Wind Energy Association (WWEA), the total installed capacity of all wind turbines reached 650.8 Gigawatt globally by the end of 2019. Figure 1 shows the year-wise growth of wind energy in terms of total installed capacity.

The reserve capacity of the wind energy systems may increase due to its uncertain characteristics such as randomness, volatility, and intermittent. It is an essential requirement in electrical power and energy systems for proper planning, operation, and management [1]. The wind energy forecasting plays an important role in timely power generation through accurate forecasting. Based on the time horizon, it is mainly categorized into four types of wind energy forecasting, namely very-short term, short-term, medium-term, and long-term forecasting. The very short-term wind energy forecasting is utilized to control the wind turbine and monitoring load in time ranges from a few seconds to 30 minutes. The time

horizon of short term wind energy forecasting ranges from 30min to 6 hours and is utilized for load sharing. The medium-term wind energy forecasting is utilized for energy trading & management of power systems and the time horizon ranges from 6 hours to 24 hours. The long-term wind energy forecasting is utilized for scheduling the wind turbine maintenance and it ranges from 1 day to 7 days [2].
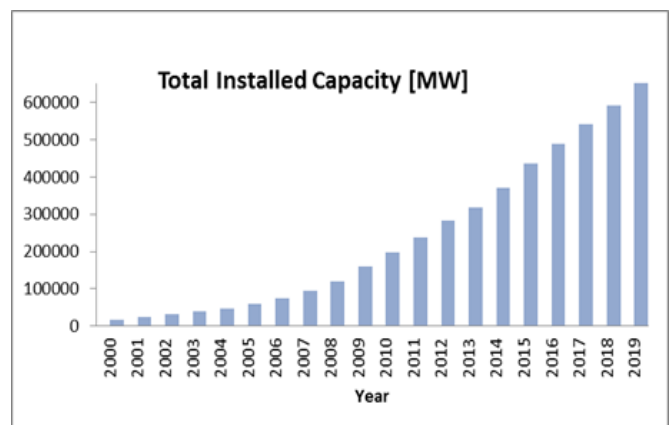


**Figure 1.** Growth of wind energy

The uncertain nature of the wind speed creates a big challenge for a few minutes to hours ahead of wind energy forecasting. In the literature, several models based on physical, statistical, and hybrid approaches were devoted to improving wind speed and wind power forecasting. The forecasting using physical approaches considers the parameters related to wind flow's physical characteristics inside and outside the wind farm, such as roughness, farm layout & obstacles, and weather forecast data such as humidity, temperature, and pressure. On

the other hand, in the statistical approach, the forecasting is performed by utilizing the historical measurement data and produces the forecast output by employing the statistical models. It does not consider the physical phenomena for the forecasting. The popular statistical approaches in use are the regression tree [3], Auto Regressive Moving Average (ARMA) [4], artificial neural networks [5], fuzzy logic, and support vector machine. The statistical approach guarantees good forecasting accuracy for the time series forecasting [6].

The hybrid models integrate two or more methods to avail of the advantages of them. The hybrid models guarantee better performance than the individual models in wind energy forecasting. The hybrid models include ensemble learning, optimization, feature selection, and decomposition techniques.

The ensemble learning based models construct different models and then integrates them to solve problems [7]. The heuristics optimization methods improve forecasting performance by optimizing the parameters of the model [8]. The feature selection and decomposition methods consider the series of historical wind speed and wind power data and improve the performance of forecasting by reducing the forecasting error. The decomposition based models belong to the category of hybrid model that decomposes the stationary series of data into multiple non-stationary subseries of data and then constructs the forecasting models for each subseries of data.

The wavelet transformation is the popular method utilized for the time series analysis to perform the transformation in time and also frequency domains. In wind energy forecasting, the discrete wavelet transform is applied to the discrete form of wind speed data. The Empirical Mode Decomposition (EMD) is another method for decomposing the time series data into a set of Intrinsic Mode Functions (IMFs) where for each IMF, different residue and frequency bands are assigned. The local properties of the time series data define frequency band and residue of the IMF. The empirical mode decomposition method has proved its efficiency in a variety of applications involving the nonlinear and non-stationary processes [9-13].

In general, the forecasting model's performance depends on the quality of the input data provided for the training process. The feature selection is an important pre-processing method for selecting the significant features related to the target feature from the list of input features. It identifies the significant features by measuring the correlation between the features. Hence, it tunes the input given to the forecasting model, which leads to improving the performance of the wind energy forecasting model [14-17]. The generalization and feature extraction capabilities of the artificial intelligence-based approaches make them outperform the physical and statistical approaches in forecasting wind speed and wind power.

This paper focusses on the critical analysis of the recent review on wind energy forecasting using RNN models. This organization of the paper consists of six sections: Section 1 discusses the challenges available in the wind energy forecasting and types of forecasting. Section 2 presents the significance of deep learning and a recurrent neural network. Section 3 explains the deep learning approaches such as Simple RNN, Bidirectional RNN, LSTM, and GRU. Section 4 discusses the performance measures utilized for the assessment of RNN methodologies in wind energy forecasting. Section 5, discusses the development of wind energy forecasting models using deep learning based RNN. Finally, the conclusion of this study is provided in section 6.

## 2. RECURRENT NEURAL NETWORKS FOR DEEP LEARNING

The Artificial Neural Network (ANN) is a popular method employed by various researchers for forecasting. It applied in many applications for the evaluation of nonlinear network structure, forecasting, pattern recognition, classification, clustering, and optimization techniques. The network is tuned to reduce the error by updating the bias and weight values during training. The performance of the network is improved with the number of samples. It consists of three dominant learning paradigms. They are supervised learning, unsupervised learning, and reinforcement learning. Supervised learning utilizes the training data comprised of the input vector along with the target vector. During learning, the difference between the actual target vector is compared against the forecast vector. The network is adjusted according to the difference until the forecast vector matches the actual vector. Unsupervised learning utilizes the training data with input vector only. During the training, the network learns by using the input patterns and forms the clusters. It can find the patterns, features, relations, categories, and regularities of input over the output. In reinforcement learning, the network receives some feedback from the environment. Many experts performed research using the ANN technique for improving the performance of wind energy prediction and exposed the importance of the selection procedure in achieving the goal [18, 19].

Li and Shi [20] investigated three different artificial neural networks namely radial basis function, adaptive linear element, and backpropagation for forecasting the wind speed. Dumitru and Gligor [21] proposed the feed-forward neural network based model for forecasting the daily average wind power. Liu at al. [22] introduced the probabilistic neural network and complex-valued recurrent neural networks to predict wind power. Wu et al. [23] proposed the neural network model for wind power forecasting, where the radial basis function is utilized. Monfared et al. [24] developed the fuzzy logic and ANN-based model for wind speed forecasting. However, these ANN algorithms need a feature extraction from the input data. Feature extraction is a difficult task, which requires expert knowledge to perform appropriately.

Features taken from each sample of data are fed into neural network algorithms. Such algorithms referred to as "shallow model" algorithms because they consist of very few composition layers. The shallow models have the neural network structure without hidden layers or with only one hidden layer. The learning process of shallow models requires more knowledge, skill, and challenging to analyze theoretically. Subsequently, shallow models can suffer from network instability, feature extraction process, weak generalization capability, and non-convergence parameters because of the uncertain and volatile nature of wind energy data. To avoid this difficulty from the shallow model, deep learning concepts were introduced. It consists of one or more hidden layers. The main aim is to automatically learn the feature hierarchy, avoid data overfitting problems, solve complex features, and transfer learning.

The wind energy forecasting with deep learning architecture was developed based on these characteristics [1, 25]. RNN models are popular approaches that are branches in the field of deep learning. Recurrent models follow the sequential approach to input data processing and the temporal dependency between successive data can be well captured.

Nowadays, deep learning has more attractiveness due to its dominant features such as feature engineering on its own, satisfactory results with unstructured data, strong generalization capability, handling the big-data & time-series data. It is most suitable for real-world applications. The neural network is built by arranging neurons in three layers. They are the input layer, hidden, and output layers. The network consists of only one input and output layer and one or more number of hidden layers in between input and output layers.

The neurons in subsequent layers are connected through weighted links. Each neuron is characterized by its weight, bias, and activation function and these are organized into three layers. The weight and bias of the neuron are updated based on the error value. The main task performed by each neuron in the network is to calculate the weighted sum of the input signal and then apply the activation function on it. There is one node corresponding to each input in the input vector. So, the number of neurons forming the input layer depends on the number of attributes or features that acts as an input to the neural network. The input layer passes the data to the first hidden layer. The hidden layers are well connected with the input layer, and they integrated with weight the input values to pass into the output layers. The output layer performs the summation of the weighted information received from the hidden layer neurons and produces the final classification or prediction outcome.

An essential feature of the artificial neural network is the activation functions. The information the neuron receives is relevant to the information given means activating the neuron. Otherwise, it should be ignored. The activation function like linear function, step function, sigmoid function, tanh function, Rectified Linear Unit (ReLU) function, Softmax function and Swish (A Self-Gated) function. The appropriate activation function for the fast convergence of the network is selected on the nature of the problem. Figure 2 shows the structure of deep learning architecture.
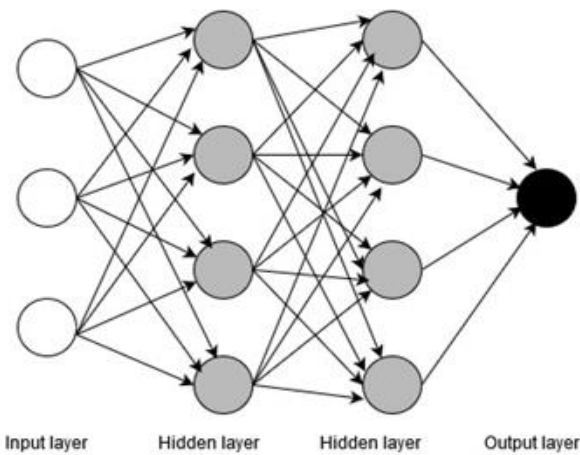


**Figure 2.** Deep learning architecture

The deep learning architecture greatly expands the neural network functionalities in terms of the number of problems and the type of problems it can address. The most popular deep learning architectures are Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN), Deep Belief Network (DBN), Deep Stacking Networks (DSN), Autoencoders, Generative Adversarial Networks (GANs), and Deep Residual Networks. In this study, the deep learning based recurrent neural network is focused on enhancing the performance of wind energy

forecasting. Figure 2 shows the deep learning architecture. The structure of RNN architecture differs from other artificial neural network architectures in representing the data in its input and output. The artificial neural network structure passes the data linearly in both feed-forward process and backpropagation process.

The RNN follows the recurrence relation during the forward pass and uses the backpropagation through time for learning. The sequence data has a time dependency among all its features. Many real-time applications like speech synthesis, natural language processing, music generation, and image captioning generate sequence data. The RNN is developed for handling these types of data. It handles the sequence data well by identifying the short-term and long-term sequence dependencies among different data points. From these dependencies, the RNN extracts the hidden pattern and utilizes this knowledge for the forecast. The RNN processes one input vector from the sequence of input vector at a time and retains that state information in the network itself. It loops the connection and produces the output by considering the previous state information and the current input [26].
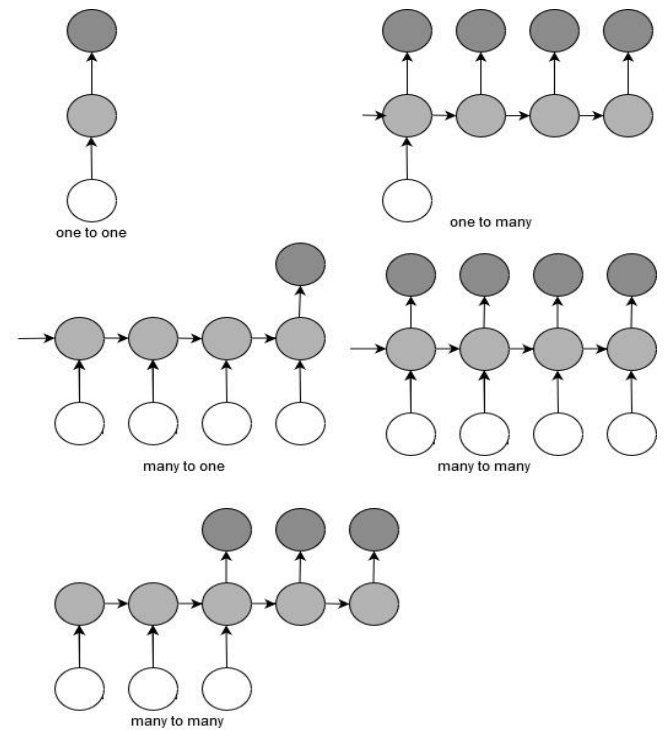


**Figure 3.** Types of RNN architectures

The primary reason for the RNN to be considered as an exciting is that they enable us to operate over long sequences of a vector. The predictive performance of the RNN is improved by designing its grid in both horizontally and vertically. The best approach is the number of elements, which are used as inputs and the expected sequence length as the output. The deep learning networks synchronize the RNN output to get the proper results. Based on the number of inputs given and outputs generated, the RNN is classified into four types, namely, one-to-one, one-to-many, many-to-one, and many-to-many. Figure 3 shows the different types of RNNs which differ in the number of inputs given and the number of outputs produced.

The one-to-one type of RNN takes only one fixed size of the input and produces only one fixed output. The one-to-many type of RNN utilizes only one fixed-sized input as the previous

case, but it produces a sequence of outputs. This model is used in generating a music and image processing area. Whereas, the many to one type of RNN architecture gets multiple sequences of inputs and produces a single output. It is mainly used for time series analysis, energy forecasting, sentimental analysis, and stock market prediction. Finally, the many-to-many type of RNN takes multiple inputs and produces multiple outputs. It is represented in two ways. The first type is fixed-size input and output sequence of data. Another type is input and output different size of a sequence of data. It mainly used for machine translation models.

## 3. METHODOLOGY

### 3.1 Simple recurrent neural networks

The RNN is a kind of neural network which can handle the large datasets easily by looping back the past information in each unit. For each time step, the recurrent neural network utilizes the number of activation function units. Each of these units contains the hidden state as an internal state of the unit. The hidden state represents the past information, which is processed earlier by the unit and holds at the specific time step. This state information is updated regularly for each time step to show the updated knowledge. In RNN, the hidden state is updated by using the recurrence relation. At the time 't', a single time step is provided as an input. Then, the current state is calculated by using the inputs provided to the network and the previous state value. Now, the calculated current state '$h_t$' will be utilized as a previous state value for the next time step at the time 't-1'. Thus the current state '$h_t$' at the time 't' becomes the previous state '$h_{t-1}$' at the time period 't-1'. The output is calculated for all the time steps. Once all the time steps are completed, the final current state is calculated. The final output of the recurrent network is calculated from the final current state [27]. After that, the error value is calculated by comparing the calculated output with the actual output. Then, this error is backpropagated to the network and the weights are updated.

Let '$x_t$' be the present input, '$h_t$' be the new hidden state, '$h_{t-1}$' be the hidden state at 't-1', and '$f_w$' be the fixed function with tangible weight. The activation function utilized for updating the hidden state is as follows,

$$h_t = f_w(x_t, h_{t-1}) \tag{1}$$

$$h_t = \tanh(w_{xh}x_t + w_{hh}h_{t-1}) \tag{2}$$

where, '$w_{hh}$' represents the weight at recurrence relation and, '$w_{xh}$' represents the weight at the input. The output of the recurrent network '$y_t$' is calculated as follows

$$y_t = w_{hy}h_t \tag{3}$$

where '$w_{hy}$' represents the weight value at the output layer of the recurrent network. The general architecture of the RNN is shown in Figure 4. Even though the primary recurrent neural network works effectively, it has some limitations due to the backpropagation of error in an extensive network. The backpropagation leads to two major problems in the recurrent network, namely vanishing gradients and exploding gradients.

The vanishing gradient and exploding gradient problems are generated in the network when the backpropagated error is too

small, nearly zero and when the error becomes too large respectively. As a solution to this problem the threshold can be set on the gradients when it is passed back in time. But this may introduce some efficiency issues in the network. So, to provide an optimal solution for overcoming these problems, the two variants of RNN are developed, namely LSTM and GRU.
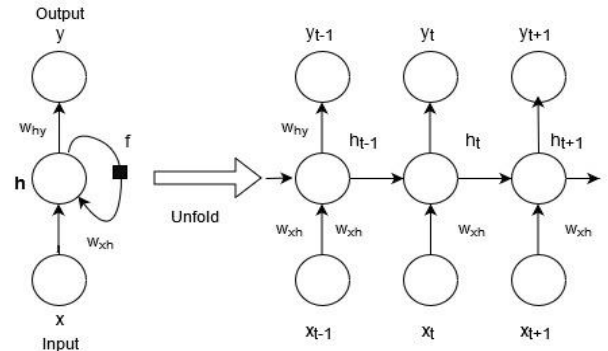


**Figure 4.** Recurrent Neural Network architecture

### 3.2 Long Short Term Memory (LSTM)

The LSTM is an extension of the recurrent neural network that remembers the hidden state information for a more extended period using the vector of internal cell state. The short term memory of the simple RNN may create a barrier for achieving good accuracy. But the LSTM solves short term memory issue by introducing the long term memory. It keeps all the required information from past learning and removes that information irrelevant from past learning. Figure 5 shows the general structure of long short term memory.
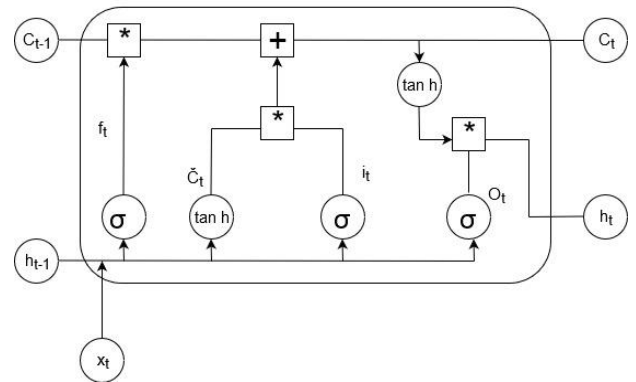


**Figure 5.** Structure of Long Short Term Memory

It achieves this filtering function with the help of gates. There are three different kinds of gates utilized by LSTM cell for different purposes: input, forget, and output gates. The input gate identifies the information that is required for the next process and should be kept in internal cell state whereas, the forget gate finds the information that should be removed and should not be kept in the internal cell state from the past learning and the output gate finds what information should be generated as an output from the internal cell state and will be utilized as the next hidden state. The following section discusses the working of the long short term memory network.

First, the sigmoid layer identifies the information that should be thrown away from the internal cell state. It decides to keep the internal state information for the next cell state

from the two inputs; one is the previous state information '$h_{t-1}$' and another one is the input at the current state '$x_t$'. It generates 0 or 1 as an output for every information in the cell state '$C_{t-1}$'. The output 1 shows that the particular information should be kept in the cell state whereas the output 0 shows that the information needs to be removed from the cell state.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{4}$$

Second, the new information to be stored in the internal cell state is identified by using two layers, namely sigmoid and tanh layer. The sigmoid layer called as input gate layer that identifies what information must be updated, followed by the tanh layer. It generates the new candidate vector '$C_t$' and add it to the internal cell state. Consequently, these two are combined to produce the updation in the cell state.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{5}$$

$$\widetilde{C}_t = tanh(W_C.[h_{t-1}, x_t] + b_C) \tag{6}$$

Finally, the cell state is updated from '$C_{t-1}$' to '$C_t$'.

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{7}$$

Third, the output '$O_t$' of the RNN will be generated by using two consecutive layers, sigmoid layer followed by output layer as follows

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{8}$$

$$h_t = O_t * tanh(C_t) \tag{9}$$

Followed by LSTM other networks such as Depth Gated LSTM [28], Multiplicative LSTMs (mLSTMs) [29] and Bidirectional LSTMs [30] are also developed for overcoming the limitation of the simple RNN.

**3.3 Gated Recurrent Unit (GRU)**

The extended version of RNN, which is also an alternative network model to LSTM for handling the vanishing gradient problem of the basic recurrent neural networks, is the gated recurrent unit. It also utilizes three gates, namely update gate, current memory gate and reset gate. The update gate behaves similarly to the output gate and decides what information should be passed to the future. The reset gate acts similarly to the combined version of input and forget gate of LSTM and helps to decide the information to be forgotten. It does not maintain any internal state. Instead of this, it incorporates the internal state information of the LSTM into the hidden state of the GRU. Finally, the collection of this information is passed into the next GRU. The current memory gate is incorporated into the reset gate and made as a subpart of the input gate. It adds some non-linearity with the input and makes the input to be a zero-mean. It minimizes the effect of previous knowledge on the current information by making the current memory gate as a subpart of the reset gate. GRU combines the two gates, namely input and forget gate into the update gate and makes changes in the combined information of cell state and hidden state. Figure 6 shows the general architecture of GRU.

First, GRU takes the current input and the previous hidden state as input vectors. Then, it performs the multiplication on element basis and calculates the parameterized current input

and past hidden state vectors for each gate [31]. The appropriate activation function is applied on each gate as follows,

$$z_t = \sigma(W_z.[h_{t-1}, x_t]) \tag{10}$$

$$r_t = \sigma(W_r.[h_{t-1}, x_t]) \tag{11}$$

The current memory gate calculates different from others in which it performs the Hadmard product of the reset gate with the previously hidden state information. After that, this information is parameterized and added to the current input vector.

$$\widetilde{h}_t = tanh(W.[r_t, * h_{t-1}, x_t]) \tag{12}$$

The current hidden state information is calculated as follows:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{13}$$
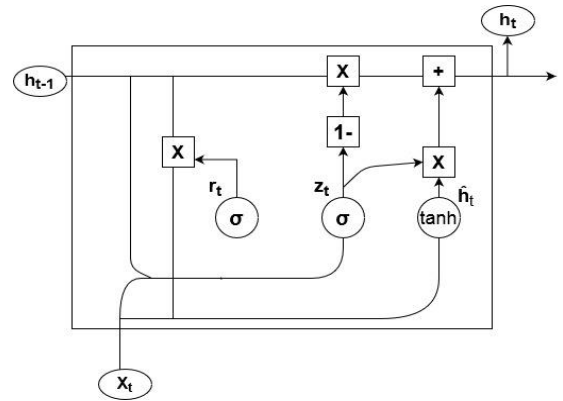


**Figure 6.** Structure of Gated Recurrent Unit

**3.4 Bidirectional recurrent neural networks**

The bidirectional recurrent neural network is formed by combining two independent RNNs in which the information is fed into one network in one direction and for another network in the reverse direction. At each time step, the output of these two networks is combined. So, at any time, the network has forward and backward sequence information [32]. The general structure of the Bidirectional RNN is shown in Figure 7. The bidirectional RNN is suitable for various applications such as prediction of the energy forecasting, protein structure, machine translation, speech recognition, and handwritten recognition. It provides good accuracy compared to the simple RNN.
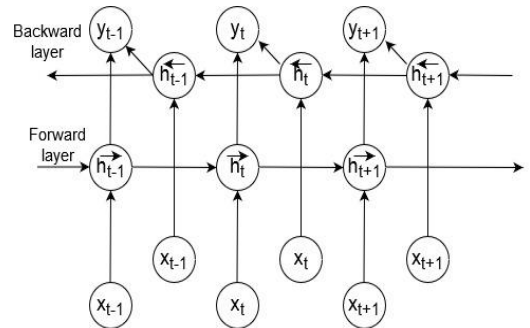


**Figure 7.** Structure of bidirectional Recurrent Neural Network

## 4. PERFORMANCE ASSESSMENT OF WIND ENERGY FORECASTING

The performance of the forecasting methods can be evaluated by using the measures such as Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean absolute Percentage Error (MAPE), and Mean Bias Error (MBE) etc. In general, the error generated by the forecasting methods is measured by calculating the difference between the actual and forecast value of the target feature as the base [33-35]. The mean absolute error is calculated as the average absolute difference between the actual and forecast value and is calculated as follows.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Y_{forecast} - X_{actual}| \qquad (14)$$

The mean square error measures the how the estimator is qualified. It is always positive. The forecasting method with the MSE value closer to zero is better. It is measured as average of the squared difference between the actual and forecast value. It is calculated as follows

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_{forecast} - X_{actual})^2 \qquad (15)$$

The root mean square error is calculated as square root of MSE, which is directly proportional to the square error. It forecasting models with a smaller error are better. It is measured as follows

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{forecast} - X_{actual})^2} \qquad (16)$$

The mean absolute percentage error is an important measure utilized by various researchers in the literature for comparing the accuracy of prediction methods. In machine learning, it is utilized as a loss function, especially for regression problems. It is calculated as follows

$$MAPE = \left(\frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_{forecast} - X_{actual}}{X_{actual}}\right|\right) * 100 \qquad (17)$$

The mean bias error is the primary measure for capturing the average bias in the forecasting method. It is calculated as follows

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(Y_{forecast} - X_{actual}) \qquad (18)$$

## 5. DEVELOPMENT OF DEEP LEARNING BASED RNN IN WIND ENERGY FORECASTING

Shao et al. [36] proposed the deep learning approach for predicting short-term wind speed by utilizing the combination of recurrent neural network and the infinite feature selection (Inf-FS). First, the essential features that contribute to improving the accuracy of the wind power forecast are identified by using the Inf-FS. Consequently, the non-

stationary components are reduced by using the wavelet decomposition. Finally, the deep learning based recurrent neural network performs the non-linear mapping and forecast short-term wind power. The proposed RNN with Inf-FS model outgunned other methodologies. Gangwar et al. [37] presented a deep learning based LSTM for forecasting the wind speed. The accuracy of the proposed model is tested against the Support Vector Machine (SVM). The result shows that the proposed LSTM based recurrent neural network provides better accuracy than SVM.

Shi et al. [38] developed the deep learning model for wind speed forecasting in which the spatial temporal correlation theory is utilized as an essential technique. First, the correlation of the adjacent wind turbines with target wind turbine are identified as an important factor for the forecasting by using continuous wavelet transforms. Then, the Wavelet Coherence Transformation analysis is introduced for analyzing the wind turbines and the time lag characteristics. Finally, the LSTM recurrent neural network is trained and the parameters are tuned. The result shows that the LSTM based deep learning model produces an improved accuracy than other traditional models.

Sun et al. [39] introduced an RNN based model for monitoring the health of the wind turbine. Due to the generation of individual faults, Supervisory Control and Data Acquisition (SCADA) variables of the wind turbine may change continuously. The variance level of each SCADA variable is combined with LSTM recurrent neural network and monitors the health status of the wind turbine. As a result, the proposed model outperformed other models. Cali and Sharma [40] proposed a deep learning based wind power forecasting. The sensitivity analysis is performed on the Numerical Weather Prediction (NWP) data and the relevant features (feature selection) are identified. Then, the LSTM based RNN is utilized for forecasting a 24 hrs ahead wind power. The weather data and wind power have a sequential dependency. The short term temporal dependency between these data is modelled effectively by using LSTM. The LSTM based RNN achieves better accuracy compared to other models.

Zu and Song [41] introduced the short term wind power forecasting model in which a wavelet packet decomposition is applied for decomposing the time series wind power sequences into a number of sub sequences. An improved GRU method with Scaled Exponential Linear Unit (SELU) activation function is utilized for predicting the wind power for each subsequence. The forecast wind power output is reconstructed from each subsequence to obtain complete wind power. The proposed model produces an improved accuracy of the short term wind power forecasting.

Liu et al. [42] proposed the short-term wind power forecasting model using Discrete Wavelet Transform (DWT) and LSTM. First, the DWT is utilized for decomposing the wind power into signals. Then, the LSTM is applied to each sub-signal for predicting the wind power. Consequently, the final prediction result is formed by combining the predicted results from each sub-signal. The proposed method produces an improved accuracy compared to other methodologies.

Pradhan and Subudhi [43] developed the Recurrent Wavelet Neural Network (RWNN) based wind speed forecasting model. The model employs the wavelet technique for decomposing the wind speed and then utilizes the recurrent wavelet neural network on the decomposed data to forecast wind speed. As a result, the proposed RWNN achieves better performance compared to conventional RNN.

**Table 1.** Summary of deep learning based RNN in wind energy forecasting

| Sl.No | Authors | Forecast variable | Energy Data set | RNN based Forecasting Methodology | Comparison Forecasting methodlogy | Performance measures | Remarks |
|---|---|---|---|---|---|---|---|
| 1 | Shao et al. [36] | Wind power | NREL | Infinite feature selection with RNN | MLP | RMSE and RSD | Accuracy of wind power forecast increased in spring, summer, autumn and winter. |
| 2 | Gangwar et al. [37] | Wind speed | Kaggle | LSTM | SVM | RMSE | LSTM has more significance results. |
| 3 | Shi et al. [38] | Wind speed | Buckley City wind farm, USA | Spatial temporal correlation with LSTM (SC-LSTM) | BP, Elman, ELM and SVM | RMSE, MAE and MAPE | SC-LSTM are produced better results. |
| 4 | Sun and Sun [39] | SCADA variable | Wind farm in Hebei province | Analysis of variance level in each SCADA variable with LSTM | ELM and ELMAN | RMSE, MAE BIAS and SDE | The performance of the SCADA variable with LSTM is acceptable. It is utilized for assessing the health status of wind turbine. |
| 5 | Cali and Sharma [40] | Wind power | Sotavento, in Spain | Sensitive analysis and LSTM | - | nRMSE and nMAE | The LSTM with sensitive analysis improves the performance by the positive (temperature) and negative (surface pressure) effect of the features in the forecast. |
| 6 | Zu and Song [41] | Wind power | Belgian electric power operator Elia | WPD-GRU-SELU (Wavelet Packet Decomposition-GRU-Scaled Exponential Linear Units) | GRU-SELU, WD-GRU-SELU, and GRU | MAPE and MAE | WPD-GRU-SELU hybrid model is superior than other models especially for the wind power with large fluctuations. |
| 7 | Liu et al. [42] | Wind power | Wind farms in Mongolia, Netherlands, China | DWT and LSTM | BP, RNN, LSTM, DWT-BP and DWT-RNN | MAE, MAPE and RMSE | DWT-LSTM method provides better prediction results compared to other models. |
| 8 | Pradhan and Subudhi [43] | Wind speed | NREL | Maximum overlap discrete wavelet transform and recurrent wavelet neural network (RWNN) | Conventional RNN | MAE | The RWNN has better and fast learning ability compared to RNN |
| 9. | Liu et al. [44] | Wind Speed | Wind farm, China | VMD-SSA-LSTM-ELM | ARIMA,LSTM, ELM,VMD-ELM, VMD-LSTM-ELM, EMD-SSA-LSTM-ELM and WPD-LSTM-ELM | MAE, MAPE and RMSE | The proposed multistep model (VMD-SSA-LSTM-ELM) extracts the trend information effectively and perform well in forecasting the wind speed. |
| 10 | Zhu et al. [45] | Wind speed | National Wind Energy Technology Center | The Hybrid model of Top-down relevant feature search (TDRG) with Gaussian Process Regression (GPR) and LSTM (TGPLSTM) | MLP and GLM | RMSE, RSE, MAE and $SS_{CRPS}$ | The TGPLSTM hybrid method provides good accuracy for the interval and point forecasting. |
| 11. | Fu et al. [46] | Wind power | North China | LSTM/GRU with wind speed correction process | ARIMA and SVM | MAE, SMAPE, RMSE and $P_{SMAPE}$ | LSTM/GRU forecasting model with the input correction process produces better performance. |
| 12. | Liu et al. [47] | Wind Speed | Neimenggu, Northwest China | Mutual Information (MI) +Stacked denoising auto-encoder (SDAE)+ LSTM | MLP and LSTM | RMSE, MAPE and MAE | This model outperforms other two forecasting models. |
| 13 | Yu et al. [48] | Wind Speed | - | Wavelet-RNN-SVM, Wavelet –LSTM-SVM, | ELM, BPN, SVM | MAE, MAPE and RMSE | WT-LSTM-SVM model and WT-GRU-SVM |

| No. | Author | Type | Location | Model | Compared with | Metrics | Findings |
|---|---|---|---|---|---|---|---|
| | | | | Wavelet-GRU-SVM | RNN, LSTM, GRU, WT_SVM, WT_LSTM and WT-GRU | | models produced recommended performance. |
| 14 | Ding et al. [49] | Wind power | Sichuan Province, China | Bidirectional GRU | SVM and ANN | RMSE and MAE | The bidirectional GRU shows better forecasting results compared to other models. |

Liu et al. [44] presented a multistep forecasting model for wind speed by using decomposition techniques followed by LSTM and ELM. First, it decomposes the original wind speed data into a sequence of sublayers by using variational mode decomposition. After that, the trend information of all sub-layers are extracted by using singular spectrum analysis. The ELM and LSTM are utilized to forecast the high and low-frequency sublayers obtained from VMD-SSA respectively. The multistep model effectively extracts the trend information from the historical wind speed data. Table 1 summarize the deep learning based RNN in wind Energy forecasting. Now days an effective renewable energy forecasting can be attained by analysing large volume of meteorological data. The main objective of the big data analytics is to assist the predictive modelers, analytics professionals and data scientists in taking the right business decisions by analysing the large volume of transactional and other forms of data. It is utilized in various areas such as energy [50], finance [51], healthcare [52], text mining [53] and telecommunication [54], load forecasting [55]. Hence, the big data analytics adds much power to wind energy forecasting. The forecasting of wind power can also be performed by using the big data based prediction framework [56].

## 6. CONCLUSION

The recurrent neural network has a multiple variants of network structure. It loops back the previous state information to predict the current state along with the current input. The RNN has the short term memory, whereas, the variants of RNN has the capability of holding the long sequences of information by employing different network structure. In this paper, the necessity of deep learning in energy forecasting and the research efforts employed by the deep neural networks such as simple RNN, LSTM, GRU, and bidirectional RNN are discussed. The review shows that the accurate prediction of wind energy is possible with the deep learning based RNN methodologies. The findings from the literature show that the RNN providing an improved performance compared to other conventional methods in wind energy forecasting. The finding from the review specified in this paper would help the researchers to choose the right method for satisfying their desired tasks and requirements in the wind energy. In future the decomposition techniques, ensemble learning techniques and the feature selection concepts can be combined with RNN and its varients to enhance the performance of the wind energy forecasting.

## REFERENCES

[1] Wang, H., Lei, Z., Zhang, X., Zhou, B., Peng, J. (2019). A review of deep learning for renewable energy forecasting. Energy Conversion and Management, 198: 111799. https://doi.org/10.1016/j.enconman.2019.111799

[2] Jung, J., Broadwater, R.P. (2014). Current status and future advances for wind speed and power forecasting. Renewable and Sustainable Energy Reviews, 31: 762-777. https://doi.org/10.1016/j.rser.2013.12.054

[3] Troncoso, A., Salcedo-Sanz, S., Casanova-Mateo, C., Riquelme, J.C., Prieto, L. (2015). Local models-based regression trees for very short-term wind speed prediction. Renewable Energy, 81: 589-598. https://doi.org/10.1016/j.renene.2015.03.071

[4] Milligan, M., Schwartz, M., Wan, Y.H. (2003). Statistical wind power forecasting models: Results for US wind farms (No. NREL/CP-500-33956). National Renewable Energy Lab. (NREL), Golden, CO (United States).

[5] Gomes, P., Castro, R. (2012). Wind speed and wind power forecasting using statistical models: autoregressive moving average (ARMA) and artificial neural networks (ANN). International Journal of Sustainable Energy Development, 1(1/2).

[6] Liu, H., Chen, C., Lv, X., Wu, X., Liu, M. (2019). Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods. Energy Conversion and Management, 195: 328-345. https://doi.org/10.1016/j.enconman.2019.05.020

[7] Heinermann, J., Kramer, O. (2016). Machine learning ensembles for wind power prediction. Renewable Energy, 89: 671-679. https://doi.org/10.1016/j.renene.2015.11.073

[8] de Andrade, C.F., dos Santos, L.F., Macedo, M.V.S., Rocha, P.A.C., Gomes, F.F. (2019). Four heuristic optimization algorithms applied to wind energy: determination of Weibull curve parameters for three Brazilian sites. International Journal of Energy and Environmental Engineering, 10(1): 1-12. https://doi.org/10.1007/s40095-018-0285-5

[9] Qian, Z., Pei, Y., Zareipour, H., Chen, N. (2019). A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. Applied Energy, 235: 939-953. https://doi.org/10.1016/j.apenergy.2018.10.080

[10] Abhinav, R., Pindoriya, N.M., Wu, J., Long, C. (2017). Short-term wind power forecasting using wavelet-based neural network. Energy Procedia, 142: 455-460. https://doi.org/10.1016/j.egypro.2017.12.071

[11] Berrezzek, F., Khelil, K., Bouadjila, T. (2019). Efficient wind speed forecasting using discrete wavelet transform and artificial neural networks. Revue d'Intelligence Artificielle, 33(6): 453-460. https://doi.org/10.18280/ria.330607

[12] Zheng, Z.W., Chen, Y.Y., Zhou, X.W., Huo, M.M., Zhao, B., Guo, M. (2013). Short-term wind power forecasting using empirical mode decomposition and RBFNN. International Journal of Smart Grid and Clean Energy, 2(2): 192-99. https://doi.org/ 10.12720/sgce.2.2.192-199

[13] Zhang, Y., Zhang, C., Sun, J., Guo, J. (2018). Improved

wind speed prediction using empirical mode decomposition. Advances in Electrical and Computer Engineering, 18(2): 3-11. https://doi.org/10.4316/AECE.2018.02001

[14] Senthil Kumar, P., Lopez, D. (2015). Feature selection used for wind speed forecasting with data driven approaches. Journal of Engineering Science and Technology Review, 8(5): 124-127.

[15] Senthil Kumar, P., Lopez, D. (2016). Forecasting of wind speed using feature selection and neural networks. International Journal of Renewable Energy Research, 6: 833-837.

[16] Kumar, P.S., Lopez, D. (2016). A review on feature selection methods for high dimensional data. International Journal of Engineering and Technology, 8(2): 669-672.

[17] Jensen, R., Shen, Q. (2008). Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches, 8: John Wiley & Sons.

[18] Sivanandam, S.N., Deepa, S.N. (2007). Principles of Soft Computing. John Wiley & Sons.

[19] Ahmad, T., Zhang, H., Yan, B. (2020). A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustainable Cities and Society, 102052. https://doi.org/10.1016/j.scs.2020.102052

[20] Li, G., Shi, J. (2010). On comparing three artificial neural networks for wind speed forecasting. Applied Energy, 87(7): 2313-2320. https://doi.org/10.1016/j.apenergy.2009.12.013

[21] Dumitru, C.D., Gligor, A. (2017). Daily average wind energy forecasting using artificial neural networks. Procedia Engineering, 181: 829-836. https://doi.org/10.1016/j.proeng.2017.02.474

[22] Liu, Z., Gao, W., Wan, Y.H., Muljadi, E. (2012). Wind power plant prediction by using neural networks. In 2012 IEEE Energy Conversion Congress and Exposition (ECCE), pp. 3154-3160. https://doi.org/10.1109/ECCE.2012.6342351

[23] Wu, X., Hong, B., Peng, X., Wen, F., Huang, J. (2011). Radial basis function neural network based short-term wind power forecasting with Grubbs test. In 2011 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), pp. 1879-1882. https://doi.org/10.1109/DRPT.2011.5994206

[24] Monfared, M., Rastegar, H., Kojabadi, H.M. (2009). A new strategy for wind speed forecasting using artificial intelligent methods. Renewable Energy, 34(3): 845-848. https://doi.org/10.1016/j.renene.2008.04.017

[25] Pasupa, K., Sunhem, W. (2016). A comparison between shallow and deep architecture classifiers on small dataset. In 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), pp. 1-6. https://doi.org/10.1109/ICITEED.2016.7863293

[26] Patterson, J., Gibson, A. (2017). Deep learning: A practitioner's approach. O'Reilly Media, Inc.

[27] Gers, F.A., Schmidhuber, J. (2000). Recurrent nets that time and count. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, 3: 189-194. https://doi.org/10.1109/IJCNN.2000.861302

[28] Yao, K., Cohn, T., Vylomova, K., Duh, K., Dyer, C. (2015). Depth-gated LSTM. arXiv preprint arXiv:1508.03790.

[29] Krause, B., Lu, L., Murray, I., Renals, S. (2016). Multiplicative LSTM for sequence modelling. arXiv preprint arXiv:1609.07959.

[30] Graves, A., Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5-6): 602-610. https://doi.org/10.1016/j.neunet.2005.06.042

[31] Dey, R., Salemt, F.M. (2017). Gate-variants of gated recurrent unit (GRU) neural networks. In 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 1597-1600. https://doi.org/10.1109/MWSCAS.2017.8053243

[32] Schuster, M., Paliwal, K.K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11): 2673-2681. https://doi.org/10.1109/78.650093

[33] Sobri, S., Koohi-Kamali, S., Rahim, N.A. (2018). Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management, 156: 459-497. http://dx.doi.org/10.1016/j.enconman.2017.11.019

[34] Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H.A., Nielsen, T.S. (2005). Standardizing the performance evaluation of short-term wind power prediction models. Wind Engineering, 29(6): 475-489. https://doi.org/10.1260/030952405776234599

[35] Subbiah, S.S., Chinnappan, J. (2020). A review of short term load forecasting using deep learning. International Journal on Emerging Technologies, 11(2): 378-384.

[36] Shao, H., Deng, X., Jiang, Y. (2018). A novel deep learning approach for short-term wind power forecasting based on infinite feature selection and recurrent neural network. Journal of Renewable and Sustainable Energy, 10(4): 043303. https://doi.org/10.1063/1.5024297

[37] Gangwar, S., Bali, V., Kumar, A. (2020). Comparative Analysis of Wind Speed Forecasting Using LSTM and SVM. EAI Endorsed Transactions on Scalable Information Systems, 7(25): e1 http://dx.doi.org/10.4108/eai.13-7-2018.159407

[38] Shi, X., Huang, S., Huang, Q., Lei, X., Li, J., Li, P., Yang, M. (2019). Deep-learning-based wind speed forecasting considering spatial–temporal correlations with adjacent wind turbines. Journal of Coastal Research, 93(sp1): 623-632. https://doi.org/10.2112/SI93-084.1

[39] Sun, Z., Sun, H. (2018). Health status assessment for wind turbine with recurrent neural networks. Mathematical Problems in Engineering, 2018. https://doi.org/10.1155/2018/6972481

[40] Cali, U., Sharma, V. (2019). Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection. International Journal of Smart Grid and Clean Energy, 8(2): 103-110. http://dx.doi.org/10.12720/sgce.8.2.103-110

[41] Zu, X.R., Song, R.X. (2018). Short-term wind power prediction method based on wavelet packet decomposition and improved GRU. In Journal of Physics: Conference Series, 1087(2): 022034. http://doi :10.1088/1742-6596/1087/2/022034

[42] Liu, Y., Guan, L., Hou, C., Han, H., Liu, Z., Sun, Y., Zheng, M. (2019). Wind power short-term prediction based on LSTM and discrete wavelet transform. Applied Sciences, 9(6): 1108.

https://doi.org/10.3390/app9061108

[43] Pradhan, P.P., Subudhi, B. (2020). Wind speed forecasting based on wavelet transformation and recurrent neural network. International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, 33(1): e2670. https://doi.org/10.1002/jnm.2670

[44] Liu, H., Mi, X., Li, Y. (2018). Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Conversion and Management, 159: 54-64. https://doi.org/10.1016/j.enconman.2018.01.010

[45] Zhu, S., Yuan, X., Xu, Z., Luo, X., Zhang, H. (2019). Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast. Energy Conversion and Management, 198: 111772. https://doi.org/10.1016/j.enconman.2019.06.083

[46] Fu, Y., Hu, W., Tang, M., Yu, R., Liu, B. (2018). Multi-step ahead wind power forecasting based on recurrent neural networks. 2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 217-222. https://doi.org/10.1109/APPEEC.2018.8566471

[47] Liu, X., Zhang, H., Kong, X., Lee, K.Y. (2020). Wind speed forecasting using deep neural network with feature selection. Neurocomputing, 397: 393-403. https://doi.org/10.1016/j.neucom.2019.08.108

[48] Yu, C., Li, Y., Bao, Y., Tang, H., Zhai, G. (2018). A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. Energy Conversion and Management, 178: 137-145. https://doi.org/10.1016/j.enconman.2018.10.008

[49] Ding, M., Zhou, H., Xie, H., Wu, M., Nakanishi, Y., Yokoyama, R. (2019). A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. Neurocomputing, 365: 54-61. https://doi.org/10.1016/j.neucom.2019.07.058

[50] Grolinger, K., L'Heureux, A., Capretz, M.A., Seewald, L. (2016). Energy forecasting for event venues: Big data and prediction accuracy. Energy and Buildings, 112: 222-233. https://doi.org/10.1016/j.enbuild.2015.12.010

[51] Bennett, M. (2013). The financial industry business ontology: Best practice for big data. Journal of Banking Regulation, 14(3-4): 255-268.

[52] Murdoch, T.B., Detsky, A.S. (2013). The inevitable application of big data to health care. Jama, 309(13): 1351-1352. https://doi.org/10.1001/jama.2013.393

[53] Sivasankari, S., Baggiya Lakshmi, T. (2016). Operational analysis of various text mining tools in bigdata. International Journal of Pharmacy & Technology (IJPT), 8(2): 4087-4091.

[54] Ruiz, M., Germán, M., Contreras, L.M., Velasco, L. (2016). Big data-backed video distribution in the telecom cloud. Computer Communications, 84: 1-11. https://doi.org/10.1016/j.comcom.2016.03.026

[55] Subbiah, S.S., Chinnappan, J. (2020). An improved short term load forecasting with ranker based feature selection technique. Journal of Intelligent & Fuzzy Systems, 39(5): 6783-6800. https://doi.org/10.3233/JIFS-191568

[56] Yin, X., Zhao, X. (2019). Big data driven multi-objective predictions for offshore wind farm based on machine learning algorithms. Energy, 186: 115704. https://doi.org/10.1016/j.energy.2019.07.034

**NOMENCLATURE**

| | |
|---|---|
| ANN | Artificial Neural Networks |
| ARIMA | Auto Regressive Integrated Moving Average |
| BP | Back Propagation |
| DWT | Discrete Wavelet Transform |
| ELM | Extreme Learning Machine |
| EMD | Empirical Mode Decomposition |
| GLM | Generalized Linear Mode |
| GRU | Gated Recurrent Unit |
| LSTM | Long Short Term Memory |
| MLP | Multilayer Perceptron |
| MAE | Mean Absolute Error |
| MSE | Mean Square Error |
| RMSE | Root Mean Square Error |
| MAPE | Mean Absolute Percentage Error |
| MBE | Mean Bias Error |
| nMAE | Normalized Mean Absolute Error |
| nRMSE | Normalized Root Mean Square Error |
| RNN | Recurrent Neural Network |
| RSD | Relative Standard Deviation |
| SCADA | Supervisory Control and Data Acquisition |
| SSA | Singular Spectrum Analysis |
| $SS_{CRPS}$ | Forecast Skill Score Based on Continuous Ranked Probability Score |
| SVM | Support Vector Machine |
| NREL | National Renewable Energy Laboratory |
| VMD | Variational Mode Decomposition |
| tanh | Hyperbolic Tangent Function |
| ReLU | Rectified Linear Unit |
| WPD | Wavelet Packet Decomposition |
| WWEA | World Wind Energy Association |