Deep Learning Ensemble Based New Approach for Very Short-Term Wind Power Forecasting

Dan A. Rosa de Jesús, *Student Member, IEEE*, Paras Mandal, *Senior Member, IEEE*, Yuan-Kang Wu, *Senior Member, IEEE*, and Tomonobu Senjyu, *Senior Member, IEEE*

***Abstract*—This paper presents a new prediction approach based on deep learning ensemble for very short-term (10-minute-ahead) wind power forecasting for a look–ahead period of 1h, 3h, and 6h. The proposed deep learning ensemble approach combines several individual and hybrid deep learning models, such as Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), Hybrid Deep Neural Network (HDNN), with the formation of four different ensembles, in particular HDNN+CNN, HDNN+LSTM, CNN+LSTM, and HDNN+CNN+LSTM. The proposed approach considers the historical data of wind speed as major input through ensemble averaging in order to produce the final wind power prediction. The major advantage of the proposed ensemble learning is that they make the best use of predictions from multiple deep learning models and their capability to effectively "cancel out" the individual errors, which in turn help enhance the final prediction accuracy. The simulation on actual data, acquired from the real wind farm in Texas, demonstrates the effectiveness of the presented approach to produce a higher degree of very short-term wind power forecast accuracy for multiple seasons of the year in comparison to other soft computing as well as to individual deep learning models.**

***Index Terms*—Convolutional Neural Networks, Deep Learning, Ensemble Learning, Hybrid Deep Neural Network, Long Short-Term Memory Networks, Wind Power Forecasting.**

I. INTRODUCTION

The integration of wind energy into the power grid has been rapidly increasing due to several advantages that it offers, such as, but not limited to, clean-generation technology, grid decarbonization, and improvement in grid efficiency and reliability. However, wind power is highly dependent on weather conditions [1], and its output power variability brings challenges to the power system operators while integrating this environmental-friendly renewable resource into the grid. Several wind power forecasting techniques are available in order to facilitate the integration of wind power into the grid. However, there is still a great need to develop a more accurate forecasting technique. This calls for a new and better forecasting approach, such as deep learning algorithms, to enhance its efficient integration to the grid [2].

D. A. Rosa de Jesús and P. Mandal are with Power and Renewable Energy Systems (PRES) Lab. within the Department of Electrical and Computer Engineering, The University of Texas at El Paso, TX 79968, USA (e-mail: danrosa@miners.utep.edu; pmanda@utep.edu).

Y-K Wu is with the Department of Electrical Engineering, National Chung Cheng University, Chiayi 62102, Taiwan (email: allenwu@ccu.edu.tw).

T. Senjyu is with the Department of Electrical and Electronics Engineering, University of the Ryukyus, Nishihara, Okinawa 9030213, Japan (email: b985542@tec.u-ryukyu.ac.jp).

Recently, deep learning models have been used for wind power forecasting. In [3], the data set is partitioned prior to training and preprocessed using a state-space reconstruction technique with delay embedding, and they achieved accuracy close to 97% using an auto-encoder Elman Recurrent Neural Network (RNN) considering several wind power time intervals. Furthermore, as it is reported in [4], a Long Short-Term Memory (LSTM) network obtained a mean absolute error of 0.5432 considering previous values of wind speed, temperature, and pressure, and they utilized a fuzzy-rough set theory to reduce the dimension of the input data in order to speed up the training process by eliminating noisy and redundant data samples. Additionally, Chen *et al.* [5] proposed an application of Convolutional Neural Network (CNN) and achieved a low absolute average wind power forecast error. In [5], the model hyperparameters are tuned using a genetic algorithm. However, the data preprocessing and hyperparameter tuning are found to be computationally expensive and time-consuming.

Although deep learning models have achieved state-of-the-art results across many applications, they are sensitive to specific training data features and hyperparameter values [6]. Additionally, deep learning algorithms overfit when the data set is not extensive, or the number of features is limited [7]. One way of alleviating the aforementioned issues is through ensemble learning, which is a deep learning technique that combines several base deep learning models to produce optimal predictive models [8]. This parts from the idea that individual models perform predictions with mutually exclusive errors. Therefore, making the best use of predictions from multiple models because their individual errors "cancel out." Moreover, ensembles provide extra degrees of freedom, allowing for solutions that would be difficult or impossible to obtain by individual deep learning models.

There have been several ensemble models proposed in the literature including boosting [9], AdaBoost [10], and random forest [11] showing promising results across several domains. In [12], two deep learning ensemble-based classification models, including a CNN ensemble and deep residual network ensemble, perform hyperspectral image classification with accuracy of more than 90% where only the noisy bands were removed from the data set before training. A deep ensemble machine for video classification achieved classification accuracy up to 91.6% on five different action recognition data set without data preprocessing [13]. Furthermore, in [14], an ensemble of CNNs reduce false-positive detection of lung nodules while increasing the sensitivity. However, to the

978-1-7281-5508-1/20/$31.00 ©2020 IEEE

best of our knowledge, the application of ensemble models has not been thoroughly studied in power system forecasting problems, e.g., wind power forecasting. The work described in this paper contributes to developing new ensemble models combining several deep learning models through ensemble averaging [15], [16] to significantly improve the accuracy of very short-term (10-minute-head) wind power forecasting for different forecasting horizons, e.g., 1h, 3h, and 6h. This paper presents four types of ensemble models, such as (i) combination of Hybrid Deep Neural Network (HDNN) and CNN, i.e., HDNN+CNN, (ii) HDNN+LSTM, (iii) CNN+LSTM, and (iv) HDNN+CNN+LSTM. The advantages of the proposed ensemble models over existing deep learning and soft computing techniques reported in the literature are: (1) decreased architecture complexity as the individual models are trained before they are combined, (2) significant reduction in training time, (3) lower degree of risk of overfitting, and (4) avoiding the need of feature engineering or hyperparameter tuning.

This paper is organized as follows. Section II describes the data set. Section III introduces the proposed ensemble models for wind power forecasting. In Section IV, the experimental results are presented and compared to other models. Finally, the conclusion and future work are discussed in Section V.

II. Data Description

The data set of wind power for training, validation, and testing the proposed ensemble models is acquired from a wind farm located close to Midland, Texas, including 10-minute interval samples of the year 2006. Fig. 1 shows the variable output power characteristic of the wind farm for the selected month. The features of the collected data include the wind speed in m/s and wind power in MW. The capacity of the selected wind farm is 450 MW. The first step in preparing the data set involves dividing the samples into seasons, i.e., winter, spring, summer, and fall. The last step involves normalizing the feature vectors within the unit range $[0, 1]$ using the min-max function defined in [17].

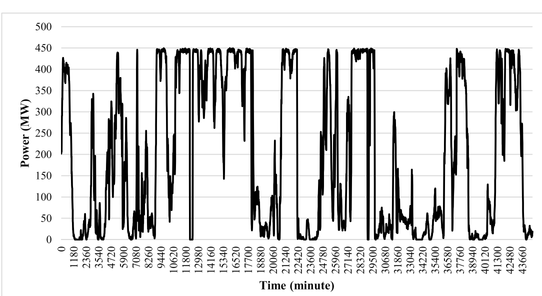

Fig. 1: Variable output power of a wind farm during March 2006.

III. Proposed Deep Learning Ensemble Models

The proposed ensemble models consist of the input layer followed by a set of deep learning models including the CNN, LSTM, Multi-Layer Perceptron Network (MLPN), and HDNN. The combination of these models through ensemble averaging allows the proposed models to perform improved feature extraction from historical trends in the data.

A. Input Layer

The input layer distributes the data feature vectors to each of the deep learning models. Each data sample corresponds to a vector with one feature, including the historical wind speed. The output dimension of this layer is $(1, 1)$.

B. Soft Computing Models

1) Convolutional Neural Network: CNNs is a deep learning algorithm that emulates the organization of the visual cortex in the animal brain. In each convolutional layer, a set of kernels perform feature extraction through convolution operations [18]. Our CNN model contains three convolutional layers with 512, 256, and 128 kernels, respectively. Each convolutional layer is followed by a pooling layer that replaces the output vector of the preceding layer with a summary statistic of the closest outputs [19].

2) Long Short-Term Memory: The LSTM is a type of RNN with cells containing gates that manage the storage and flow of information [19]. Our LSTM model contains three LSTM of 256, 256, and 64 cells, respectively.

3) Multi-Layer Perceptron Network: The MLPN is a type of neural network composed of multiple layers of perceptrons that map weighted inputs to the output of each unit through activations. This network has three layers with 512, 1024, and 256 units, respectively.

4) Hybrid Deep Neural Network: The HDNN [20] model consists of the combination of a convolutional, LSTM, and perceptron units at the layer level (See Fig. 2).

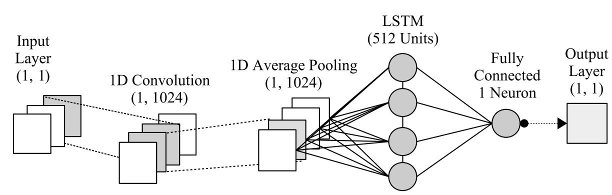

Fig. 2: The HDNN model, including the convolutional and LSTM layers of 1024 units and 512 cells, respectively.

C. Ensemble Averaging and Proposed Forecasting Procedure

Ensemble averaging [15], [16] is a type of committee machine that takes advantage of the "divide and conquer" strategy by combining the predictions of multiple deep learning models. Ensembling deep learning models involves three stages (see Fig. 3):

Stage 1. Building $N \geq 2$ deep learning models with initial weight values,

Stage 2. Training the models to obtain the set of weight values that yield the best forecasts, and

Stage 3. Combining the models through ensemble averaging to obtain improved final predictions.

The deep learning models are built using their corresponding hyperparameters. During the training stage, individual

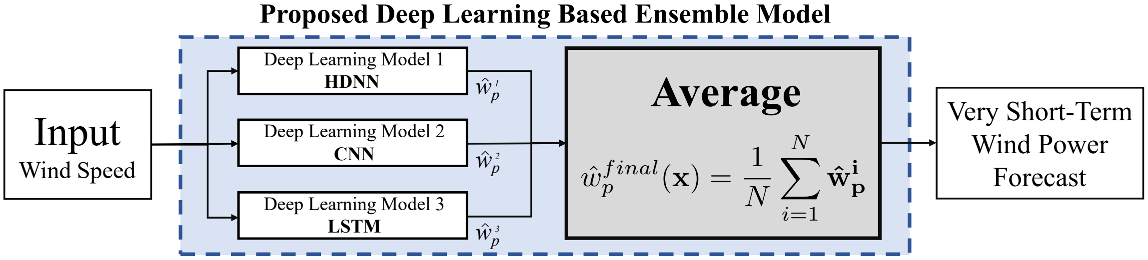

Fig. 3: Flow of forecast process: An example of the high-level architecture of one of the four ensemble-models, i.e., ensemble-4 model (ENS4) presented inside the dashed lines, for wind power forecasting that considers ensembling of three different deep learning models (HDNN, CNN, and LSTM). In this figure, the individual model first produces a separate wind power forecast ($\hat{w}_p^1$ by HDNN, $\hat{w}_p^2$ by CNN, and $\hat{w}_p^3$ by LSTM) that are averaged to obtain the final wind power forecasts. Other three ensemble-models 1–3 (not shown here) are ENS1 (HDNN+CNN), ENS2 (HDNN+LSTM), and ENS3 (CNN+LSTM) follow the similar forecast process.

wind speed data samples are forward passed through each model to obtain the wind power forecast values. With the associated wind power actual and forecasted values, an error is computed and backpropagated through the layers of the models to adjust the weights accordingly. This process is repeated several times until the models achieve convergence. During the testing stage, the deep learning models are loaded using the weights obtained during training. Some of the models are connected to the average layer, depending on the specific ensemble model implementation. Then, the wind speed data samples are forward passed through the models to obtain their associated wind power forecast values. The result is the average of the individual deep learning model's wind power forecasts $\hat{\mathbf{w}}_{\mathbf{p}}^{\mathbf{i}}$ that contribute to the final forecasting $\hat{w}_p^{final}$, which is given by:

$$\hat{w}_p^{final}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{w}}_{\mathbf{p}}^{\mathbf{i}}, \tag{1}$$

where $\mathbf{x}$ is the input and N is the number of deep learning models for ensembling purpose. The benefits of this approach include decreased architecture complexity as the individual models are trained before they are combined and decreased training time and risk of overfitting as there are fewer weight values to set. Also, no feature engineering is required as the individual models consider different aspects of the data improving the final prediction. In this paper, four ensembles (ENS1—ENS4) are developed for wind power forecasting.

D. Activation Functions, Loss Function, and Optimizer

All the models implement the linear activation function for the hidden layers and sigmoid for the output layer [19]:

$$S(x) = \frac{e^x}{e^x + 1}, \tag{2}$$

where $S(x)$ is the sigmoid activation function and x the weighted sum of the layer's input and the associated weight values. The Huber loss function [21] is used for all the deep learning models in this work because it is less sensitive to outliers in comparison to the mean squared error. The Nesterov-accelerated Adaptive Moment Estimation (Nadam) optimizer is used in all the models and is similar to the Root Mean Square Propagation (RMSprop) but with momentum [19]. A detailed description of these model architectures and their hyperparameters can be found in [20].

IV. Simulation Results and Discussion

We compare the performance of the proposed ensemble models (ENS1—ENS4) with other deep learning models: CNN, LSTM, MLPN, and HDNN using the following accuracy measures: Mean Absolute Percentage Error (MAPE), Normalized Mean Absolute Error (NMAE), and Normalized Root Mean Squared Error (NRMSE) [1]:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|W_t^a - \widehat{W}_t^p|}{\overline{W_t}^{a,N}} \times 100\%, \tag{3}$$

where N is the number samples with the actual wind power data (W_t^a) at time t, vector containing the predicted wind power ($\widehat{W}_t^p$) at time t, and average of the actual wind power:

$$\overline{W_t}^{a,N} = \frac{1}{N}\sum_{t=1}^{N} W_t^a, \tag{4}$$

$$NMAE = \frac{1}{N}\sum_{t=1}^{N}\frac{|W_t^a - \widehat{W}_t^p|}{W_N} \times 100\%, \tag{5}$$

where W_N is the capacity of the wind farm system (450 MW).

$$NRMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\frac{W_t^a - \widehat{W}_t^p}{W_N}\right)^2} \times 100\%. \tag{6}$$

Table I presents the performance metric values obtained from the considered models applied over the one-hour look-ahead period. The MAPE values range from 0.782% to 89.845% through the seasons, being the LSTM and CNN the models with the lowest values among the deep learning models. The proposed ensemble models show superior performance with MAPE values between 0.691% and 3.288% for the spring and summer seasons, respectively. This demonstrates the effectiveness of combining several deep learning models to improve wind power forecasts.

TABLE I: Comparison of 10-minute-ahead wind power forecasting performance of the proposed ensemble models with other deep learning models over the horizon of 1 hour.

Model	Metric (%)	Spring	Summer	Fall	Winter
CNN	**MAPE**	1.214	3.433	2.818	1.812
	NMAE	0.012	0.004	0.026	0.012
	NRMSE	1.380	0.483	3.300	1.460
RNN	**MAPE**	1.099	3.682	2.637	1.842
	NMAE	0.011	0.004	0.024	0.012
	NRMSE	1.256	0.490	3.351	1.499
LSTM	**MAPE**	0.782	4.981	2.745	2.102
	NMAE	0.008	0.006	0.026	0.013
	NRMSE	0.914	0.685	3.664	1.661
MLPN	**MAPE**	1.214	89.845	2.762	1.803
	NMAE	0.012	0.102	0.026	0.011
	NRMSE	1.380	10.250	3.246	1.449
HDNN	**MAPE**	2.056	5.925	2.734	1.647
	NMAE	0.020	0.007	0.026	0.011
	NRMSE	2.230	0.774	3.517	1.146
ENS1 **HDNN+CNN**	**MAPE**	0.850	3.595	2.776	**1.630**
	NMAE	0.008	0.004	0.026	**0.010**
	NRMSE	1.052	0.512	3.384	**1.130**
ENS2 **HDNN+LSTM**	**MAPE**	1.029	5.453	**2.277**	1.816
	NMAE	0.010	0.006	**0.021**	0.012
	NRMSE	1.151	0.729	**3.214**	1.340
ENS3 **CNN+LSTM**	**MAPE**	0.988	**3.288**	2.311	1.638
	NMAE	0.010	**0.004**	0.021	0.011
	NRMSE	1.131	**0.452**	3.497	1.172
ENS4 **HDNN+CNN+LSTM**	**MAPE**	**0.691**	3.835	2.403	1.641
	NMAE	**0.007**	0.004	0.022	0.011
	NRMSE	**0.890**	0.561	3.383	1.134

TABLE II: Comparison of 10-minute-ahead wind power forecasting performance of the proposed ensemble models with other deep learning models over the horizon of 3 hours.

Model	Metric (%)	Spring	Summer	Fall	Winter
CNN	**MAPE**	3.004	6.002	4.034	2.576
	NMAE	0.026	0.014	0.010	0.017
	NRMSE	3.142	1.696	1.277	2.121
RNN	**MAPE**	2.901	5.979	4.823	2.563
	NMAE	0.025	0.014	0.013	0.017
	NRMSE	3.055	1.711	1.515	2.124
LSTM	**MAPE**	3.191	4.924	3.064	2.547
	NMAE	0.026	0.012	0.008	0.017
	NRMSE	3.352	1.352	0.925	1.918
MLPN	**MAPE**	2.994	124.032	4.430	2.588
	NMAE	0.026	0.304	0.011	0.017
	NRMSE	3.134	31.527	1.400	2.125
HDNN	**MAPE**	2.508	4.918	3.035	2.484
	NMAE	0.021	0.011	0.008	0.016
	NRMSE	2.720	1.343	0.929	1.888
ENS1 **HDNN+CNN**	**MAPE**	2.416	5.452	3.510	2.483
	NMAE	0.021	0.013	0.009	0.016
	NRMSE	2.623	1.486	1.063	1.952
ENS2 **HDNN+LSTM**	**MAPE**	**2.179**	**4.878**	**3.029**	2.473
	NMAE	**0.018**	**0.011**	**0.008**	0.016
	NRMSE	**2.429**	**1.341**	**0.923**	1.845
ENS3 **CNN+LSTM**	**MAPE**	2.508	5.385	3.529	**2.386**
	NMAE	0.021	0.013	0.009	**0.016**
	NRMSE	2.720	1.469	1.057	**1.802**
ENS4 **HDNN+CNN+LSTM**	**MAPE**	2.356	5.224	3.356	2.418
	NMAE	0.020	0.012	0.008	0.016
	NRMSE	2.558	1.421	1.003	1.829

TABLE III: Comparison of 10-minute-ahead wind power forecasting performance of the proposed ensemble models with other deep learning models over the horizon of 6 hours.

Model	Metric (%)	Spring	Summer	Fall	Winter
CNN	**MAPE**	6.597	7.334	4.516	5.082
	NMAE	0.014	0.009	0.012	0.008
	NRMSE	1.710	1.132	1.571	1.089
RNN	**MAPE**	7.415	7.816	4.876	5.204
	NMAE	0.016	0.010	0.013	0.008
	NRMSE	1.876	1.154	1.665	1.099
LSTM	**MAPE**	4.328	7.243	4.539	4.399
	NMAE	0.009	0.009	0.012	0.007
	NRMSE	1.305	1.260	1.442	0.884
MLPN	**MAPE**	6.544	115.862	4.618	4.990
	NMAE	0.014	0.144	0.012	0.008
	NRMSE	1.699	15.394	1.593	1.068
HDNN	**MAPE**	4.303	7.172	4.306	4.768
	NMAE	0.009	0.009	0.011	0.008
	NRMSE	1.292	1.231	1.381	0.929
ENS1 **HDNN+CNN**	**MAPE**	5.160	**6.752**	**4.154**	4.726
	NMAE	0.011	**0.008**	**0.011**	0.008
	NRMSE	1.435	**1.115**	**1.383**	0.973
ENS2 **HDNN+LSTM**	**MAPE**	**4.273**	7.199	4.412	4.488
	NMAE	**0.009**	0.009	0.012	0.007
	NRMSE	**1.289**	1.244	1.408	0.887
ENS3 **CNN+LSTM**	**MAPE**	4.824	6.803	4.252	**4.328**
	NMAE	0.011	0.008	0.011	**0.007**
	NRMSE	1.382	1.133	1.407	**0.872**
ENS4 **HDNN+CNN+LSTM**	**MAPE**	4.592	6.768	4.226	4.503
	NMAE	0.010	0.008	0.011	0.007
	NRMSE	1.342	1.151	1.386	0.905

To further study the wind power forecasting capability of the proposed ensemble models, we performed 10-minute-ahead wind power forecasting for the look-ahead period of the next 3 hours and 6 hours (see Tables II and III). The HDNN shows better performance than the other deep learning models with MAPE values ranging from 2.484% and 4.918% for the 3-hour horizon scenario. However, in the same case, the proposed ensemble models showed improved performance with the MAPE values between 2.179% and 4.878%, being the HDNN+CNN ensemble (ENS2) the best model on three out of four seasons (from spring to fall). Furthermore, in the 6-hour time horizon case, the HDNN obtained MAPE values between 4.303% and 7.172%, which outperform the other deep learning models. In the same scenario, the proposed ensemble models produced the MAPE values between 4.154% and 6.752%, which are superior to the values obtained by the other deep learning models.

From the results presented in Tables I–III, it can be observed that the ENS2 (HDNN+LSTM) performed better than the other ensembles, winning two out of three scenarios in spring (3-hour and 6-hour time horizons) and fall (1-hour and 3-hour time horizons). The ENS3 obtained the winning accuracy on the winter data set for 3-hour and 6-hour time horizon cases. These ensembles can extract useful information from historical trends in the data taking advantage of their individual deep learning models to enhance the forecasting accuracy.

Furthermore, ENS3, ENS2, and ENS1 displayed their best performance during the summer data set for 1-hour, 3-hour, and 6-hour forecasting horizon cases, respectively. ENS1, which combines the HDNN and CNN deep learning models, performed better feature extraction on the longest forecasting horizon scenario as both models include convolutional units. The rest of the accuracy measures and associated values are presented in Tables I-III.

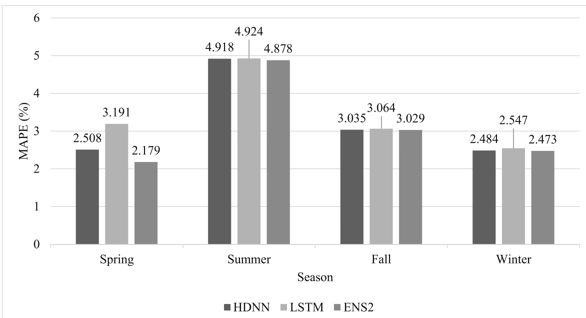

Fig. 4: Seasonal forecasting accuracy improvement as the data passes through different models in the HDNN+LSTM ensemble (ENS2).

Figure 4 illustrates the MAPE as the data flows from the input to the output layers of the ENS2 (HDNN+LSTM) in different seasons for 10-minute-ahead wind power forecasting over the period of 3 hours. This figure further demonstrates the capability of the proposed ensemble learning to improve wind power prediction combining the strengths of the individual forecasting models.

V. Conclusion

This paper developed ensemble models combining various deep learning models through ensemble averaging to predict very short-term wind power. It is found that wind speed is the most influencing factor in predicting wind power output. We used historical data of wind speed and fed it to the proposed ensemble and other deep learning models to perform 10-minute-ahead wind power forecasting over the horizon of 1, 3, and 6 hours in multiple seasons of the year. The test results demonstrated that the proposed deep learning-based ensemble approach show superior performance over other five individual deep learning models, i.e., a CNN, LSTM, MLPN, and HDNN. Therefore, our proposed ensemble models are suitable for renewable electric power forecasting application as they yield higher forecasting accuracy regardless of the season of the year. The proposed ensemble models can also be applied to other power system forecasting and classification problems. Future work includes (i) performing wind power forecasting considering more meteorological features and (2) integrating the proposed ensemble models into power systems scheduling problems such as unit commitment and economic dispatch.

References

[1] R. M. Peri, P. Mandal, A. U. Haque, and B. Tseng, "Very short-term prediction of wind farm power: An advanced hybrid intelligent approach," in *2015 IEEE Industry Applications Society Annual Meeting*, 2015, pp. 1–8.

[2] X. Luo, J. Sun, L. Wang, W. Wang, W. Zhao, J. Wu, J. Wang, and Z. Zhang, "Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4963–4971, 2018.

[3] Z. Shi, H. Liang, and V. Dinavahi, "Direct interval forecast of uncertain wind power based on recurrent neural networks," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1177–1187, 2018.

[4] W. Yao, P. Huang, and Z. Jia, "Multidimensional lstm networks to predict wind speed," in *2018 37th Chinese Control Conference (CCC)*, 2018, pp. 7493–7497.

[5] G. Chen, J. Shan, D. Y. Li, C. Wang, C. Li, Z. Zhou, X. Wang, Z. Li, and J. J. Hao, "Research on wind power prediction method based on convolutional neural network and genetic algorithm," in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2019, pp. 3573–3578.

[6] M. P. Ranjit, G. Ganapathy, K. Sridhar, and V. Arumugham, "Efficient deep learning hyperparameter tuning using cloud infrastructure: Intelligent distributed hyperparameter tuning with bayesian optimization in the cloud," in *2019 IEEE 12th International Conference on Cloud Computing (CLOUD)*, 2019, pp. 520–522.

[7] X. Chen, J. Weng, W. Lu, J. Xu, and J. Weng, "Deep manifold learning combined with convolutional neural networks for action recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 3938–3952, 2018.

[8] H. Li, Y. Cui, Y. Liu, W. Li, Y. Shi, C. Fang, H. Li, T. Gao, L. Hu, and Y. Lu, "Ensemble learning for overall power conversion efficiency of the all-organic dye-sensitized solar cells," *IEEE Access*, vol. 6, pp. 34 118–34 126, 2018.

[9] V. Soto, S. García-Moratilla, G. Martínez-Muñoz, D. Hernández-Lobato, and A. Suárez, "A double pruning scheme for boosting ensembles," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2682–2695, 2014.

[10] J. Cui, W. Li, C. Fang, S. Su, J. Luan, T. Gao, L. Hu, Y. Lu, and G. Chen, "Adaboost ensemble correction models for tddft calculated absorption energies," *IEEE Access*, vol. 7, pp. 38 397–38 406, 2019.

[11] J. Xia, W. Liao, J. Chanussot, P. Du, G. Song, and W. Philips, "Improving random forest with ensemble of features and semisupervised feature extraction," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1471–1475, 2015.

[12] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep learning ensemble for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 6, pp. 1882–1897, 2019.

[13] J. Zheng, X. Cao, B. Zhang, X. Zhen, and X. Su, "Deep ensemble machine for video classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 2, pp. 553–565, 2019.

[14] C. Li, G. Zhu, X. Wu, and Y. Wang, "False-positive reduction on lung nodules detection in chest radiographs by ensemble of convolutional neural networks," *IEEE Access*, vol. 6, pp. 16 060–16 067, 2018.

[15] X. Zhang and D. Wang, "A deep ensemble learning method for monaural speech separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 5, pp. 967–977, 2016.

[16] N. Yu, L. Qian, Y. Huang, and Y. Wu, "Ensemble learning for facial age estimation within non-ideal facial imagery," *IEEE Access*, vol. 7, pp. 97 938–97 948, 2019.

[17] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.

[18] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, ISSN: 0018-9219.

[19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[20] D. R. de Jesús, P. Mandal, S. Chakraborty, and T. Senjyu, "Solar pv power prediction using a new approach based on hybrid deep neural network," in *2019 IEEE Power Energy Society General Meeting*, 2019.

[21] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, Mar. 1964.