Improved Stacked Ensemble based Model For Very Short-Term Wind Power Forecasting

Monsef Tahir, *IEEE Memeber*, Ramadan El-Shatshat, *IEEE Senior memeber*, and M.M.A Salama, *IEEE Fellow*

***Abstract*-- the high intermittency and uncertainty associated with wind energy have brought serious challenges to the power system operators, such as increasing the operating costs by increasing the requirements on primary reserves. Hence, the need to apply more robust and accurate forecasting model has become a persistent need to meet these challenges. This paper proposes a generic two-layer stacked ensemble based model for a very-short-term wind power forecasting. The proposed approach combines a heterogeneous machine learning algorithms composed of three well-established models SVR, RBFNN, and RF heuristically with a parameter tuned SVR on the upper layer. First, the base learners are trained and tested on the original data, and then their outputs are used to train the upper layer learner to predict the final output. In this paper, a criterion for selecting the most efficient base predictors is used based on a reasonable decision for the number and type of the combined learners that takes into consideration: *(i)* the reliability of predictor(s) for forecasting the wind power data set; and *(ii)* the number of predictors that produce high performance. Many experiments had been conducted and the results prove that the proposed model is consistently better than the state-of-the-art models. The proposed model can assist network operators to make the right decision to regulate and schedule their resources optimally.**

Keywords—short-term forecasting, regression, artificial intelligence, ensemble, stacking, time series, wind power, SVR, ANN, RBFNN.

I. Introduction

Due to the growing demand for clean and renewable energy, wind power farm installations have increased considerably in recent years. Wind power capacity has been expanded and reached 369.553 GW, fulfilling 4% of the total worldwide power demand by the end of 2014[1]. However, the stochastic and the intermittent wind power generation bring serious challenges to the power system at the large-scale penetration. The uncertainty associated with the generated power may lead to unreliable decisions when it comes to real-time grid operation and regulation actions[2]. Therefore, to safely integrate wind power into the electricity grid, a reliable very-short-term wind power forecasting technique is required.

Several prediction techniques applied successfully to forecast wind power for 15 min, 30 min, and 1 hour ahead are reported in the literature. In [2][3][4][5][6], a comprehensive overview has been presented on current and new developments in wind power forecasting. A classification of wind power forecasting according to time–scale ranging from very-short-term (few minutes) to long-term (weeks or years) is reported in [6]. Forecasting models are mainly classified into three categories: statistical models, artificial intelligence based models, and hybrid/ensemble models. Statistical models such as the autoregressive (AR), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), Bayesian approach, and gray predictions are used for short time forecasting, but the prediction error increases when the variations of the entered data is high [6]. Artificial intelligence based models such Neural networks (ANN), support vector regression (SVR), fuzzy logic (FL), and heuristic techniques are widely applied in the wind and power forecasting because of their ability to efficiently handle the non-linear relationship between the variables [7]. Recently, probabilistic wind power forecasting using radial basis function neural network (RBFNN) has been proposed in [8]. The developed RBFNNs are trained with the ordinary orthogonal least square algorithm, and their parameters are tuned optimally using particle swarm optimization. In other attempts, the forecasting accuracy of RBFNN showed some improvement when the RBFNN parameters are automatically selected using genetic algorithm (GA) as in [9]. Investigation on the performance of the SVR algorithm against that of a multilayer perceptron (MLP) proved the effectiveness of SVRs for wind speed prediction [10]. For improving the prediction accuracy of SVR, optimization algorithms are usually used to tune its parameters, which are the regularization parameter C, ε-insensitive loss function, and the bandwidth of Kernel function. Parameter tuning of SVR using GA to improve the forecasting accuracy by minimizing the prediction error is presented in [11]. Ensemble forecasting models for wind power forecasting, such as bagging and boosting ensemble algorithms have been reported in [12]. In [13], a review of ensemble methods for wind and solar power forecasting is presented, in which the author classified the ensemble models into competitive and cooperative models. Random forest (RF) ensemble algorithm for combining SVRs from the base learners is presented in [14], where the author used an intelligent weighting approach that takes into account weather situations and the past forecast of the models. The authors in [15] investigated some methods for predicting the probability density function of generated wind power from wind farm locations. They constructed probability density forecasts from weather ensemble predictions, which is a new type of weather forecast generated from atmospheric models. It has been observed that the perdition error of the most common forecasting techniques has a tendency to increase, specifically when the prediction horizon increases; and sometimes they are unable to generalize from the trend.

This paper proposes an efficient stacked ensemble forecasting model that combines a family of learning machines heuristically based on a reasonable decision for the number and type of the combined learners. The model combines three diverse and widely applied time series forecasting techniques (SVR, RBFNN, and RF) via a parameter tuned SVR to form the final output. The criterion for constructing an efficient ensemble predictor is based on; (i) reliability of the base predictor(s) for forecasting the wind power data set, and (ii) the number of the predictors that can achieve high performance. The proposed stacked ensemble model has been trained to predict the future wind power up to 2 hours ahead. Many experiments had been conducted and the results show that the model can accurately forecast the future wind power output. Also, the results show that the proposed forecasting model has superior performance compared with the classical forecasting models such as Linear regression (LR), Random tree (RT), and Multilayer perceptron (MLP).

This paper is structured as follows: section I introduces different forecasting algorithms with a focus on SVR, RF, RBFNN, and Ensemble models. The proposed stacked ensemble model is presented in section II. Data processing and evaluation, and experiment results and discussion are presented in section III and section IV, respectively. The conclusion is in section V.

II. STACKED ENSEMBLE MODEL

Ensemble learning generally improves the accuracy of the machine learning by combining several models, which were originally developed to reduce the variability in classification/regression decisions and thereby increase generalization performance [16]. Unlike Bagging and Boosting algorithms, which originally combine base learners of the same type, Stacking algorithm is usually applied to base learners that built by different learning algorithms. ***Stacking*** algorithm is an ensemble learning technique, in which the predictions of a group of individual learners (base learners) are given as inputs to a second-level learning algorithm (meta-learner). This second-level algorithm combines the model predictions optimally to form a final set of predictions [17].

The proposed stacked ensemble model used in this paper is illustrated in Figure. 1. The process starts with preprocessing the time series wind power data, as discussed in section III, and then use it for training and testing the base learners located in the first layer. The base learners used in the first layer are composed of RF, RBFNN, and SVR. The outputs of those base learners are then used to train the second layer forecasting models, which is SVR. The second layer SVR model is then used as a combiner for the outputs of base forecasting models. The structure of the model and the way of processing the data improves the prediction accuracy by automatically combining diverse learners heuristically as discussed in section IV. The following is a brief description of each learning algorithm used in this model.

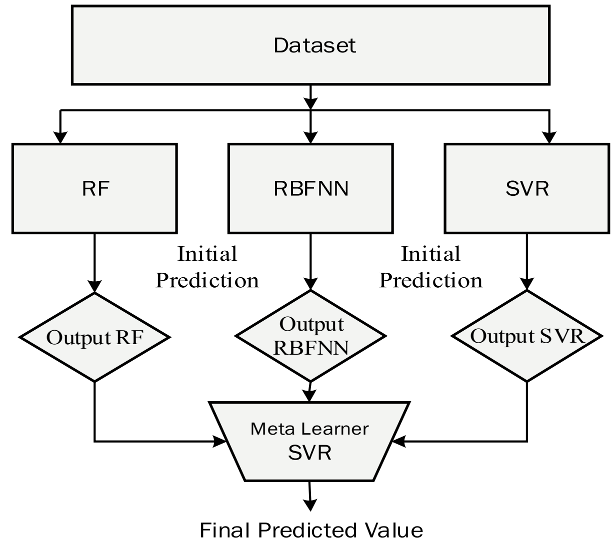

Figure. 1. Proposed forecasting model

1. Support Vector Regression (SVR)

The support vector regression technique, which is based on Vapnik's concept of support vectors [18], is introduced firstly in [19]. The basic idea of SVR is based on the computation of a linear regression function in a high dimensional feature space where the input data are mapped via a nonlinear function. In addition, one of the main characteristics of SVR is that instead of minimizing the observed training error, it attempts to minimize the generalized error bound to achieve generalized performance. This generalization error bound is the combination of the training error and a regularization term that controls the complexity of the hypothesis space [20].

Giving a set of training samples:

$$\{(x_1, y_1), \dots \dots (x_n, y_n)\} \; with \; x_i \in R^d \quad and \quad y_i \in R$$

The linear regression model can be expressed as

$$f(x) = w_1 x_1 + w_2 x_2 + \cdots . w_n x_n + b = \langle w, x \rangle + b \qquad (1)$$

Where $w = [w_1 \, w_2 \, \dots . \, w_3]^T$ represents the vector of coefficients and b represents the bias. The regression problem can be addressed by solving the following constrained optimization problem[21].

$$\underbrace{Min}_{w,b,\xi,\xi^*_,} \; \frac{1}{2} \, w^T w + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \qquad (2)$$

In order to avoid underfitting and overfitting of the training data set, the meta-learner SVR minimizes the training error and the regularization parameter of the terms

$$\left(\frac{1}{2}\right) w^T w \;\; and \; C \sum\nolimits_{i=1}^{n} (\xi_i + \xi_i^*).$$

Subject to the following constraints

$$y_i - (w^T \emptyset(x_i) + b) \leq \varepsilon + \xi_i$$
$$(w^T \emptyset(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \quad (3)$$
$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \ldots\ldots, n$$

Where x_i is mapped to a higher dimensional space by the function $\emptyset$. ξ_i^* And ξ are slack variables subject to ε - insensitive tube. The parameters, which control the prediction accuracy, are the regularization parameter C , the width of the tube ε and the mapping function $\emptyset$.

By applying the Lagrange multiplier method and fulfilling Karnsh-Kuhn-Tucker (KKT) conditions, the dual form is transformed into an optimization function as in [22].

$$Min_{\alpha,\alpha^*} \ \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*) \quad (4)$$

S.T

$$\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha^{(*)} \leq C$$

Where $Q_{ij} = \emptyset(x_i)^T \emptyset(x_i)$. The inner product can be replaced by Kernel function from the type of RBF, which is a commonly used in similar applications.

2. Radial Basis Function Neural Networks (RBFNN)

Radial Basis Function Neural Network is a particular kind of neural network widely used in short-term predictions. It is a three-layer feed-forward network consisting of an input layer, hidden layer, and an output layer. RBFNN generally uses a linear transfer function for the output units and a nonlinear transfer function for the hidden units. Its input layer simply consists of the source nodes connected by weighted connections to the hidden layer and the net input to a hidden unit is a distance measure between the input presented at the input layer and the point represented by the hidden unit. [23].

3. Random Forest (RF)

Random forest is a non-parametric ensemble base learning technique used for classification and regression suggested by L. Breiman [24]. RFs for regression are formed by growing trees depending on a random vector such that the tree predictor takes on numerical values an opposed to class labels. The output values are numerical and we assume that the training set is independently taken from the distribution of the random vector. The random forest estimator is formed by taking the average over a number of trees similar to the classification scenario [24].

The pseudo code of the proposed stacking ensemble algorithm is shown in the squared area [25]. The code shows the sequence followed to construct the model starting from the input data to the base models in the first layer and ends with the output data from the SVR at the second layer.

```
Input: Data set      D={(x_1, y_1), (x_2, y_2), ……… (x_n, y_n)}
  First-level learning algorithm  L_1, …… L_T
  Second-level learning algorithm  L
Process
For  t = 1 ,2, …, T :
  h_t = L_t(D)   % Train a 1st level individual learner h_t
        %Train an algorithm L_t to the original data set D
End;
D  = ∅     % Generate a new data set
For t = 1,2, …. , m
        For t = 1,2, …. , T
z_it = h_t(x_i) % h_t to predicit the training example x_i
End;
              D' = D'  ∪ {((z_i1, z_i2, …… z_ij), y_i )}
End;
h' = L(D').     % Train the 2nd level learner
 L              %learn algorithm L to the original D
Output:   H(x) = h'(h_1(x), h_2 (x), …… . , h_T(x))
```

III. DATA PROCESSING AND EVALUATION

1. Data Processing

Wind power was measured chronologically from January 1, 2010, to Dec 15, 2011, at a site in Ontario, Canada, with readings taken every 15 min. The measured data represent a very short term interval recorded from a wind power farm. First, data is normalized between 0 and 1 after deleting outside range values and filling in missing values using imputation. The data is then divided into training and testing sets representing 80% and 20% of the total dataset respectively.

Table 1, gives the descriptive statistics of the training and testing data.

Table 1. Characteristics of training/testing data

	Training Data	Test Data
Data Size	52424	13106
Max/Min Values	1/0	0.9979/0
Mean	0.4191	0.4392
Standard deviation	0.3162	0.3152

2. Model Evaluation

To assess the prediction performance of the base models and the proposed stacked model, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are employed to evaluate the accuracy of the model. These indices are defined as:

$$MAE = \frac{1}{T} \sum_{t=}^{T} |f_t - y_t| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (f_t - y_t)^2} \quad (6)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{f_t - y_t}{y_t} \right| \quad (7)$$

Where f_t and y_t are the forecasted and the desired values respectively while T is the total number of samples.

IV. RESULTS AND DISCUSSION

In order to achieve precise and generalized forecasting process, the parameters of the model should be selected carefully. From the literature, some researchers tuned the parameters of the SVR and RBFNN by applying some optimization algorithms such as GA as discussed in [9][11] using the training data set. However, the tuning process using heuristic search techniques adds more computational burden to the forecasting process without remarkably improving the prediction accuracy. Therefore, tuning the base learners' parameters and the meta-learner parameters SVR is conducted based on trial and error offline until the appropriate values are determined. Fig. 2, shows the impact of changing the C parameter of the SVR on the RMSE and MAE values. It can be seen that changing the regularization parameter C from 1 to 100 does not change the prediction accuracy significantly. However, having a value of more than 1000 deteriorates the accuracy of the forecaster.

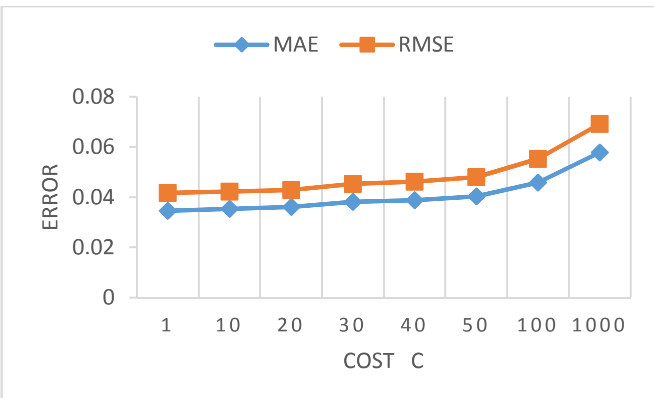

Fig. 2. Impact C parameter of SVR on MAE and RMSE

In this paper, the parameters of the SVR model have been chosen after several trials. The regularization parameter C is selected to be 1. The RBF Kernel used in SVR is shown in equation (8), Where σ = 0.001 and the term $\|x_n - x_i\|^2$ is the square Euclidean distance between the two feature vectors.

$$K(x, y) = \exp\left(-\frac{\|x_n - x_i\|^2}{2\sigma^2}\right) \quad (8)$$

When the appropriate parameters of the meta-learner (SVR) are selected, the ensemble model is built, trained, tested, and used to predict future wind power up to 2 hours ahead.

Figure. 3, shows the performance of the base models SVR, RBFNN, and RF on a sample of the test data set for four consecutive days, from July 7, 2011, to July 10, 2011, with a prediction step of 15 min. Each model has been trained and tested on the same data set individually. It can be seen from the figure that the SVR algorithms slightly better than RF and RBFNN specially when comparing the RMSE as shown in Table 2. The predictions of the three models individually are acceptable, but the accuracy deteriorates especially when the variation of the data is high as shown in Fig. 3.

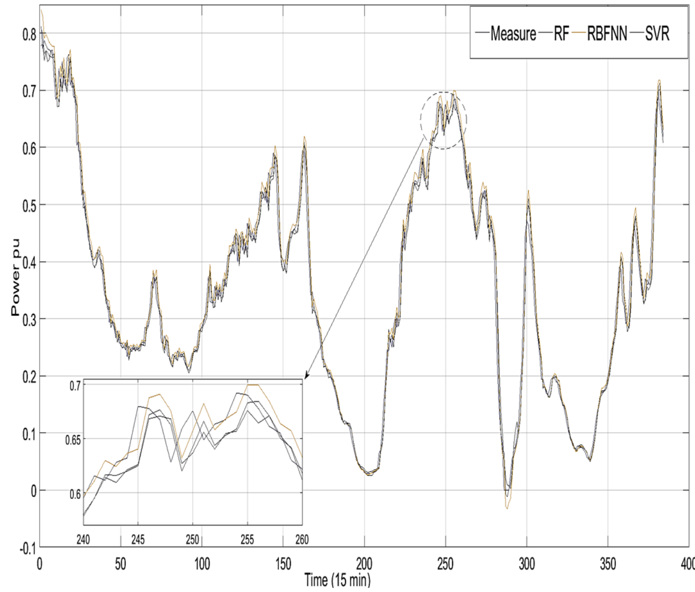

Fig. 3. Predictions of the base models (July 7- July 11)

Fig. 4. Shows the performance of the stacked ensemble model applied to the same subset of the data. The measured data and the forecasted data are plotted for the same consecutive days, from July 7, 2011 to July 10, 2011 with a prediction step of 15 min. It can seen that the proposed model can fit the test data more precisely compared with individual base models, and yields better forecasting accuracy. By combining the three base models using the proposed stacked ensemble model, the prediction accuracy increases even at the point where the variation is high as seen in Fig. 4.

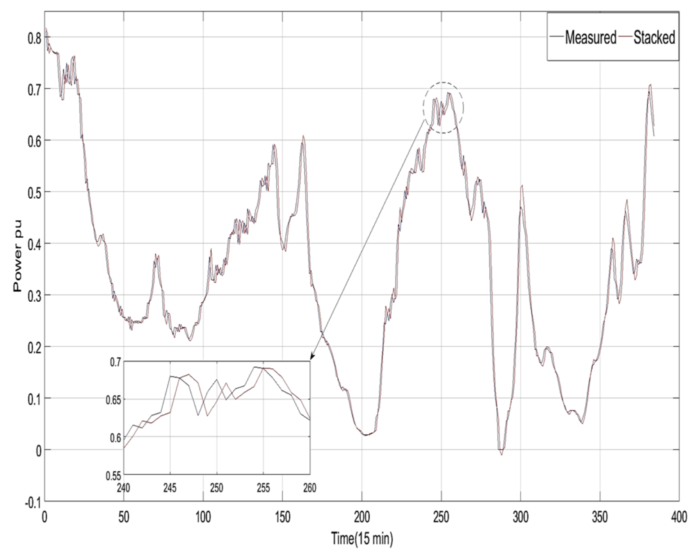

Fig. 4. Prediction of the stacked model (July 7- July 11)

Furthermore, the proposed staked ensemble model is evaluated and compared against classical algorithms, such as linear regression (LR), multilayer perceptron (MLP), and random tree

(RD) for a very-short-term wind power forecasting. Fig. 5. Shows the performance of LR, MLP, and RD trained and tested on the same data set. It is clear from Table 2 and Fig.5 that the classical models are not adequate when dealing with highly non-linear data set. Therefore, they are good when predicting for one or two points ahead or the prediction error increases.

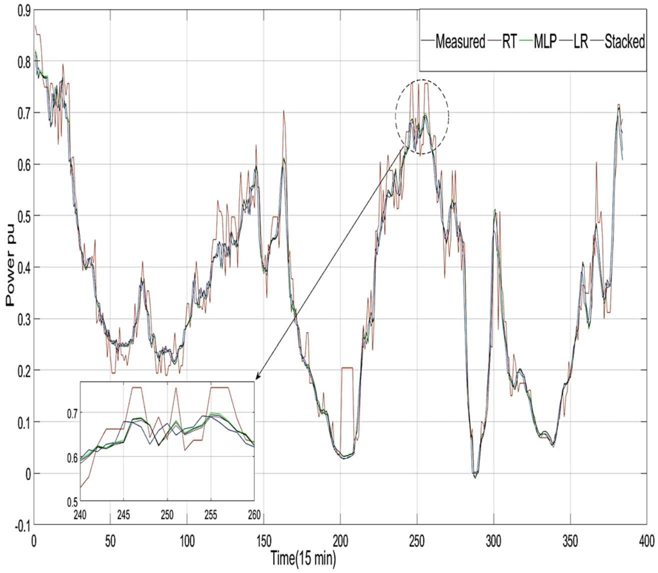

Fig. 5. Predictions of the classical models RT, MLP, and LR

Table 2, shows the performance metrics of each model after the evaluation of the test data set. The forecasting process has been conducted on each model individually at first then the proposed stacked model is used for comparison purposes. Noticeably, the proposed model outperformed the individual models in terms of prediction error minimization specifically for the MAPE. Inherently, the strength of the model comes from its ability to learn from three heterogeneous models learned by different algorithms.

Table 2. Comparison between the proposed and traditional models

Model	MAPE%	RMSE	MAE
SVR	8.555	0.0254	0.0171
RBFNN	7.5484	0.0256	0.0171
RF	7.4264	0.026	0.0173
STACKED	6.84	0.0251	0.0165
MLP	15.572	0.0267	0.0194
RT	44.6885	0.0594	0.0401
LR	9.6854	0.0259	0.0177

One advantage of the proposed model is that the forecasting accuracy does not deteriorate noticeably when predicting for a few hours ahead. As can be seen from Fig. 6, the proposed model has less forecasting error comparing with the individual models when forecasting up to 2 hours ahead beyond the test data set.

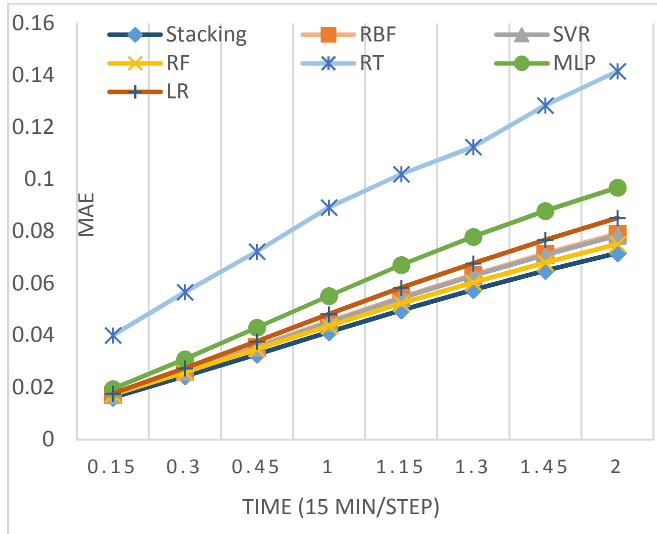

Fig. 6. Prediction of all learners over the horizon of 2 hours

V. CONCLUSION

Precise forecasting of wind power plays a significant role in improving the efficiency and reliability of power systems. In this paper, a stacked ensemble-forecasting model has been established and applied to time series wind power data for very-short-term future power forecasting. The prediction models SVR, RBFNN, and RF are combined stochastically by SVR to build the ensemble model at the first input layer that is connected to the input samples. The three base models are then combined by the SVR model, which is trained by the outputs of the base models to predict the final output. The evaluation of the stacked ensemble model using MAE, RMSE, and MAPE showed that the proposed model outperforms the individual base models in addition to other classical models, such as MLP, LR, and RT. Our results show that combining the base predictors via parameter tuned SVR improves the stacked ensemble model prediction accuracy, which makes it a reliable technique for the power network operators.

I. REFERENCES

[1] M. S. Aziz, S. Ahmed, U. Saleem, and G. M. Mufti, “Wind-hybrid Power Generation Systems Using Renewable Energy Sources-A Review,” *Int. J. Renew. ENERGY Res.*, vol. 7, no. 1, 2017.

[2] A. Tascikaraoglu and M. Uzunoglu, “A review of combined approaches for prediction of short-term wind speed and power,” *Renew. Sustain. Energy Rev.*, vol. 34, pp. 243–254, 2014.

[3] M. S. El Moursi, A. Al Hinai, E. B. Ssekulima, and M. B. Anwar, “Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review,” *IET Renew. Power Gener.*, vol. 10, no. 7,

pp. 885–989, 2016.
[4] J. Melorose, R. Perroy, and S. Careas, “an Overview on Wind Power Forecasting Methods,” *Statew. Agric. L. Use Baseline 2015*, vol. 1, 2015.
[5] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, “Current methods and advances in forecasting of wind power generation,” *Renew. Energy*, vol. 37, no. 1, pp. 1–8, 2012.
[6] W.-Y. Chang, “A Literature Review of Wind Forecasting Methods,” *J. Power Energy Eng.*, vol. 2, no. 4, pp. 161–168, 2014.
[7] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, “A review on the forecasting of wind speed and generated power,” *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, 2009.
[8] G. Sideratos and N. D. Hatziargyriou, “Probabilistic wind power forecasting using radial basis function neural networks,” *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1788–1796, 2012.
[9] A. F. Sheta and K. De Jong, “Time-series forecasting using GA-tuned radial basis functions,” *Inf. Sci. (Ny).*, vol. 133, no. 3–4, pp. 221–228, 2001.
[10] M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, “Support vector machines for wind speed prediction,” *Renew. Energy*, vol. 29, no. 6, pp. 939–947, 2004.
[11] G. Santamaría-Bonfil, A. Reyes-Ballesteros, and C. Gershenson, “Wind speed forecasting for wind farms: A method based on support vector regression,” *Renew. Energy*, vol. 85, pp. 790–809, 2016.
[12] H. Allende and C. Valle, “Ensemble Methods for Time Series Forecasting,” in *A Passion for Multi-Valued Logic and Soft Computing*, vol. 349, springer, 2017, pp. 217–282.
[13] Y. Ren, P. N. Suganthan, and N. Srikanth, “Ensemble methods for wind and solar power forecasting — A state-of-the-art review,” *Renew. Sustain. Energy Rev.*, vol. 50, pp. 82–91, 2015.
[14] M. Abuella and B. Chowdhury, “Random Forest Ensemble of Support Vector Regression Models for Solar Power Forecasting,” *arXiv1705.00033 [cs]*, pp. 2–6, 2017.
[15] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. De Sousa, *Ensemble approaches for regression*, vol. 45, no. 1. 2012.
[16] E. Zhang, Cha, and Yunqian Ma, *Ensemble Machine Learning*, vol. 8, no. 2. Boston, MA: Springer US, 2012.
[17] J. Sill, G. Takacs, L. Mackey, and D. Lin, “Feature-Weighted Linear Stacking,” 2009. [Online]. Available: http://arxiv.org/abs/0911.0460.
[18] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[19] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, “Support vector regression machines,” *Adv. Neural Inf. Process. Dystems*, vol. 1, pp. 155–161, 1997.
[20] D. Basak, S. Pal, and D. C. Patranabis, “Support Vector Regression,” *Neuronal Inf. Process. - Lett. Rev.*, vol. 11, no. 10, pp. 203–224, 2007.
[21] K. Chen and J. Yu, “Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach,” *Appl. Energy*, vol. 113, pp. 690–705, 2014.
[22] Y. Ren, P. N. Suganthan, and N. Srikanth, “A Novel Empirical Mode Decomposition With Support Vector Regression for Wind Speed Forecasting,” *IEEE Trans Neural Netw Learn Syst*, vol. 27, no. 8, pp. 1793–1798, 2014.
[23] H. Du and N. Zhang, “Time series prediction using evolving radial basis function networks with new encoding scheme,” *Neurocomputing*, vol. 71, no. 7–9, pp. 1388–1400, 2008.
[24] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[25] G. Wang, J. Hao, J. Ma, and H. Jiang, “A comparative assessment of ensemble learning for credit scoring,” *Expert Syst. Appl.*, vol. 38, no. 1, pp. 223–230, 2011.