

Performance comparison of artificial intelligence techniques in short term current forecasting for photovoltaic system

Muhammad Murtadha Othman, Mohammad Fazrul Ashraf Mohd Fazil,
Mohd Hafez Hilmi Harun, Ismail Musirin, Shahril Irwan Sulaiman
Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia

Article Info

Article history:

Received Nov 27, 2018

Revised Jan 24, 2019

Accepted Mar 13, 2019

Keywords:

Artificial neural network

Multiple time lags

Random forest

Short term photovoltaic current forecasting

Wavelet denoising

ABSTRACT

This paper presents artificial intelligence approach of artificial neural network (ANN) and random forest (RF) that used to perform short-term photovoltaic (PV) output current forecasting (STPCF) for the next 24-hours. The input data for ANN and RF is consists of multiple time lags of hourly solar irradiance, temperature, hour, power and current to determine the movement pattern of data that have been denoised by using wavelet decomposition. The Levenberg-Marquardt optimization technique is used as a back-propagation algorithm for ANN and the bagging based bootstrapping technique is used in the RF to improve the results of forecasting. The information of PV output current is obtained from Green Energy Research (GERC) University Technology Mara Shah Alam, Malaysia and is used as the case study in estimation of PV output current for the next 24-hours. The results have shown that both proposed techniques are able to perform forecasting of future hourly PV output current with less error.

*Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Muhammad Murtadha Othman,
Faculty of Electrical Engineering,
Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia.
Email: mamat505my@yahoo.com

1. INTRODUCTION

The past few years have shown a remarkable growth in the use of solar energy for residential, commercial, and industrial sectors. The growing capacity for global solar PV sector already reached 178GW in 2014, and estimated to reach 540GW in 2019 [1, 2]. In recent years, solar PV system has been developed drastically and the reason behind it is because of the nature usage of PV that is maintenance free, long lasting used, and environmentally friendly [3-8]. However, PV system is operating in a non-stationary random process caused by the variability of solar irradiance and other environmental factors that may affect the output current.

In general, artificial neural network (ANN), support vector machine (SVM), and fuzzy logic have been used as forecasting methods due to several advantages [9, 10]. These methods have been used for solar irradiance forecasting due to its increasing demand in producing accurate output. Other than any basic AI technique, there is also a combination of several AI techniques that can produce accurate forecasting result in the future of solar irradiance. A time series with ANN, fuzzy logic with ANN and wavelet based ANN are examples of famous combination for ANN.

In particular, ANN is an alternative model that capable of handling uncertainty matters of solar irradiance [11, 12]. The main advantages of using ANN can be seen in its stability to solve complex modelling especially a non-linear model [9]. Random forest (RF) is another advance AI used for forecasting. RF uses ensemble machine learning that consists of many decision tree models for classification and regression [13]. The construction of tree does not depend on the previous tree since the trees are created independently by using bootstrap aggregation technique as well as the bagging [14, 15]. RF has the advantage in terms of non-

overfitting the output results, the run time process is fast and efficient when handling a large dataset thus gives it superior predictive performance.

This paper presents the ANN and RF methods that used to perform short-term PV output current forecasting for the next 24-hours. There is not much research that has been done regarding to the PV output current forecasting using RF. The input data used for this method is current, irradiance, hours and temperature will pass through the filtration process using the wavelet decomposition to eliminate the noises in each data. Then, the multiple time lags is used due to its capabilities to identify the pattern and behavior of filtered data while improving it for accurate estimation in the next 24 hours of PV output current forecasting [16-20]. The case study uses PV output current, temperature, irradiance and hours in 2015 with the total of 7460 hourly data obtained from the Green Energy Research Center (GERC), University Technology Mara Shah Alam, Malaysia. The robustness of both models in forecasting are compared by referring to the mean square error (MSE), mean absolute percentage error (MAPE) and regression between the forecasted and actual (targeted) values.

2. RESEARCH METHOD

This segment will explain the concept of feature extraction or data preparation for the ANN and RF models used in the PV output current forecasting [21-25]. The structure used for this process is shown in Figure 1 wherein the STPCF process begins from the original data selection. In this case, the hourly PV output current, temperature, irradiance and hours are selected as the input data. After data selection, the data preparation for the input and target data is performed by using the wavelet decomposition and multiple time lags technique. Subsequently, the forecasting models for ANN and RF are designed. Finally, the training and testing procedure is performed to obtain the forecasting outcome from ANN and RF.

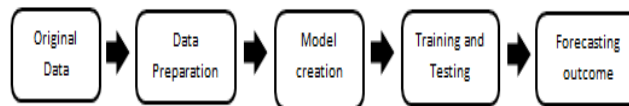


Figure 1. Block diagram of PV output current forecasting modeling

2.1. Input data of chronological parameter

The information of data used in forecasting is acquired from the Green Energy Research Centre (GERC) of UiTM Shah Alam, Malaysia. The data is obtained in the form of MATLAB software. The data is obtained consisting with five parameters in 2015. In the GERC laboratory, this information is collected by data logger for every 5 minutes. This data will be analyzed and the hour, irradiance, temperature, power and current with maximum parameter value will be used in forecasting as shown in Table 1.

Table 1. Information for each parameter

Parameter	Maximum Value	Unit
Hour	24	Hour
Irradiance	1320	W/m ²
Temperature	47.9	°C
Power	26847.61	Watt
Current	9.98	Ampere

The data preparation for ANN and RF models is shown in Figure 2 and its procedure is elaborated as follows.



Figure 2. Block diagram for data preparation

- Collect the raw data consisting with 7460 hourly information of hour, irradiance, temperature, power and current.
- Perform the filtering process of input data by using the wavelet decomposition to reduce any noise inside it.
- Normalize all of the filtered data by dividing with its maximum value in order to reduce data redundancy within the range of 0 and 1.
- Perform the multiple time lags to improve the data to determine the movement pattern of every input data required by the ANN and RF techniques.
- Use the multiple time lags to estimate the future variable and the lagged (past period) variable that will evolved in the future [9, 9, 9, 9, 9]. The input data improved by the multiple time lagsable to determine the movement pattern of data in the neural network based on (1). The total number of time interval lagging is K=24 hours.

$$Lagk = Z_t - Z_{t-k} \tag{1}$$

where,
t : time interval.
k : time interval lags 1, 2, 3...K.
 K : total number of time interval lagging.

In (1), the total number of time interval lagging used is K=24 where the value of K is stated to be equivalent to the time interval in the forecasted variable. The value of K is fixed to 24 for forecasting the next 24 hours of PV output current. The input data is in the form of k-by-t matrix where each column will be used to forecast PV output current for the next 24-hours. The first column of training data, *Lagk* is used to forecast the target data of *X48*. The input data arrangement for training and target data is shown in Figure 3.

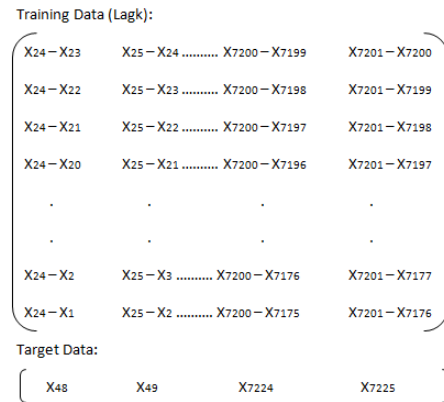


Figure 3. The input data arrangement for training data and target data for ANN and RF in Lag 24

Two sets of data which is training data and target data is created after all data have been converted into multiple time lags. The arrangement of training data for each line is a combination of hour, temperature, irradiance and current. The last part of current data will be used as target data. The training and target data formed will be used for forecasting using ANN and RF methods. The chronological arrangement of input data is shown in Table 2.

Table 2. Chronological arrangement of input data

Input data	
Parameter for each line after Lag 24	
1 to 23	hour
24 to 46	temperature
47 to 69	irradiance
70 to 92	current
Target Data	

2.2. Artificial neural network (ANN)

The Artificial Neural Network (ANN) is an alternative method that has been efficiently carried out in this paper as it is also suited to tackle solar energy uncertainty issues. The ANN is using the Levenberg-Marquardt technique to forecast the PV output current for the next 24 hours [4].

In this case study, the ANN model for forecasting the PV output current is consisting of one input layer, two hidden layers, and one output layer. The output layer of ANN is consisting of one neuron which will provides the predicted PV output current for the next 24 hours. The Levenberg-Marquardt technique is used in the ANN model as back propagation algorithm for optimization of the data during the training process. This technique is commonly used in forecasting the training set of ANN due to its algorithm that compromise between the accuracy and stability of prediction to achieve the steepest method for measuring minimal errors. The ANN model for forecasting PV output current is shown in Figure 4 and the ANN procedure is explained below.

- Divide input data into three sets of training, testing and validating for the multiple time lags of $K=24$ hours.
- In the training process, the synapses minimize the error between the actual output and the targeted output by regulating the learning rate and momentum.
- Select the number of hidden layers is based on the fact of one hidden layer is sufficient to estimate any function. Therefore, two hidden layers is used in this ANN models that will provide more precise results with minimum RMS error in forecasting the next 24 hours of PV output current.
- Repeat the error minimization process until the optimization process in forecasting is converged yielding to the smallest error in its output. Then, the training procedure is terminated once the minimum error becomes plateau for several iterations of optimization process involved in the ANN.
- Identify the strength of ANN in producing the correct STPCF results that can be proven by conducting the testing then validation processes by using different set of input data.



Figure 4. Block diagram for ANN model

2.3. Random forest (RF)

Random Forest is a model comprising with two significant components of tree bagging and random decision trees [6]. The TreeBagger defined as B is containing with the number of trees ($NTrees$) with the X and Y as the ensemble function that been used for creating a decision tree. The decision tree uses the input function X to predict the target response Y . The procedure of Random Forest is explained below.

- Perform bootstrap samples, N randomly drawn from the training data of RF model, to create a regression trees for each sample. The bootstrap sample is having the same size as the original training data.
- Perform the bagging technique that divides the bootstrap sample into two sets of data which is two-third is for the In-Bag while the remaining data is for the Out-Of-Bag (OOB).
- Use the InBag to create a forest wherein the tree growth technique will produce the best leaves. . The OOB data is used to run the unbiased prediction error as trees are added into the forest during tree growth phase using the InBag data. The primary role of OOB data in tree growth technique is to compare its estimation with the predicted values obtained from the InBag to find the best leaves with minimal error rate from every tree.
- Halt the growth of the tree once the final node of best leaf in every tree is obtained. Upon finishing the final nodes, the prediction value from the final node of best leaf is collected from every tree and the average prediction is calculated from the final node leaf of all trees. Figure 5 shows the structure of RF.

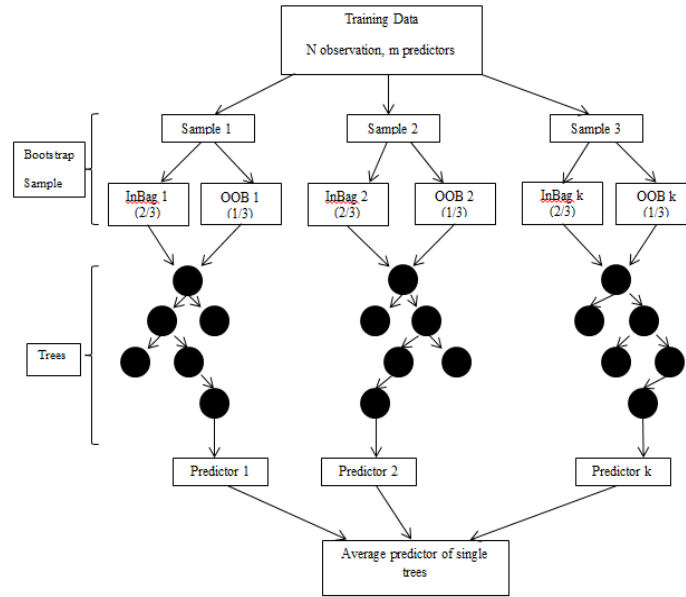


Figure 5. Structure of RF algorithm

3. RESULTS AND ANALYSIS

This section discussed on the STPCF results determined by using the ANN and RF models. The data of hourly solar irradiance, temperature, hour, and current obtained from Green Energy Research (GERC) University Technology Mara Shah Alam, Malaysia is used for the case study of STPCF. The input data undergoes the wavelet decomposition to eliminate the noise inside the data and then the multiple time lags of K=24-hours is applied to the filtered data. The data size used in ANN and RF procedure is 17520 columns. The data is divided into three sets wherein the data size for training is 5785 columns, testing data is 720 columns and validation data is 720 columns.

3.1. Artificial neural network (ANN)

The input data of ANN is the combination of multiple time lags of hour, temperature, irradiance and current. Training and testing procedures of ANN are performed where the input data having the multiple time lags of K = 24 hours. In the ANN model, the number of neurons for the first hidden layer is 20 and second hidden layer is 10. While, learning rate and momentum is 0.3 respectively. The numbers for first hidden layer, second hidden layer, learning rate and momentum are selected by performing sensitivity analysis where the selected values of learning rate and is referring to the minimum RMSE value of output.

Table 3. Results of ANN considering all the best parameters

ANN Output for K=24	
Training sets	5785
Testing sets	720
Number of neuron in 1 st hidden layer	20
Number of neuron in 2 nd hidden layer	10
Number of output	1
Learning rate	0.3
Momentum	0.3
Training function	Levenberg- Marquardt
Training RMSE	0.5270
Testing RMSE	0.5301
Regression	0.99342
Minimum MAPE	0.2832
Maximum MAPE	13.2377
Mean MAPE	4.4217

Figures 6(a) and 6(b) represent the result of forecasted PV output current versus actual targeted values, and the regression of output results obtained during the testing procedure of ANN, respectively. In Figure 6(a), the actual targeted output is in blue colour and the forecasted PV output current is in red colour. The forecasted pattern of hourly PV output current is almost the same with the pattern of actual targeted values at certain hours. However, there is inconsistency with several large error in the variation between the forecasted and targeted PV output current for the next 24 hours.

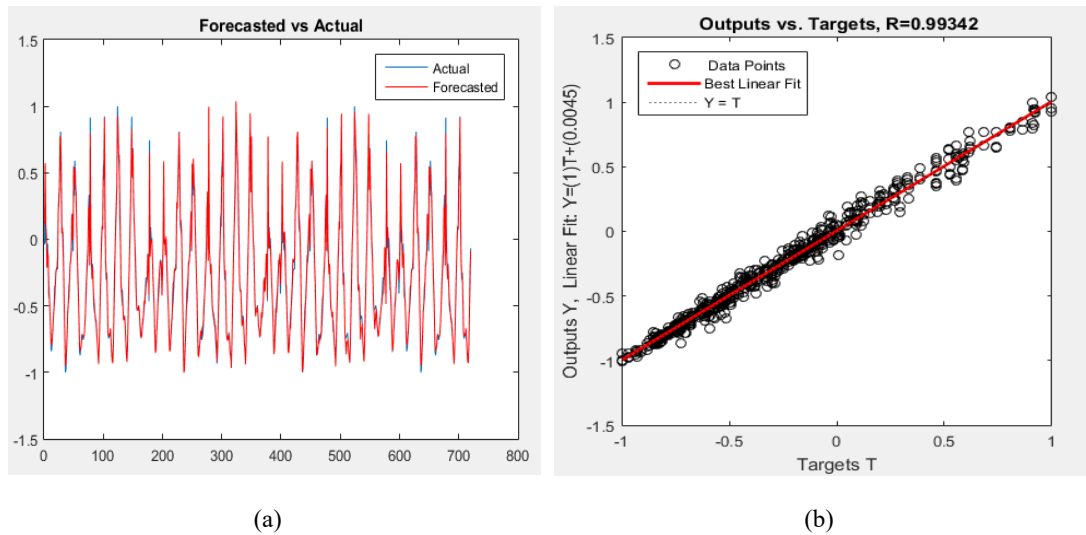


Figure 6. STPCF for the next 24-hours using ANN for the (a) forecasted PV output current versus actual targeted values, (b) regression of forecasted versus actual PV output current

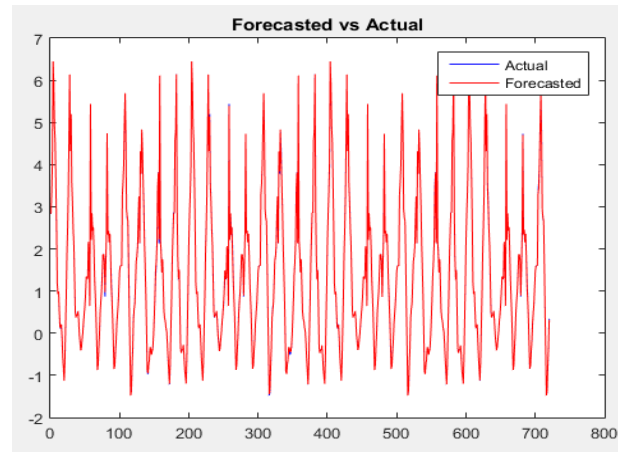
3.2. Random forest (RF)

The multiple time lags of hour, temperature, irradiance and current are used as the input data of RF. The training, testing and validating processes of RF are conducted using the input data with multiple time lags of $K=24$ hours. The RF is conducted at three different cases of 1, 5 and 10 number of trees (*Ntrees*) and every tree consisting of 5 leaves. The selection for the number of trees is based on the fact that single tree is sufficient to estimate any function. Therefore, two trees will provide more precision in determining the minimum error. The mean square error (MSE) is obtained from the training procedure. However, in the testing procedure, the MSE is automatically compared its output with the targeted data at each leaf in every tree. These comparisons are performed until the finest trees expansion is achieved giving the minimum average of RMS error for the final node leaf of all trees. The number of trees chosen for the sensitivity analysis to determine the best prediction of PV output current with minimum RMSE values RF is shown in Table 4.

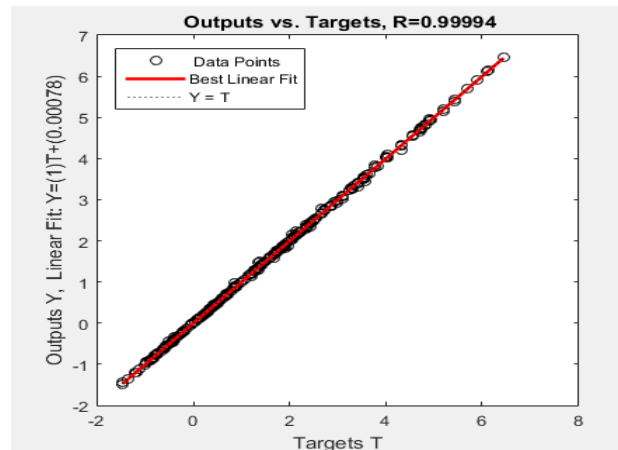
Table 4. Results of RF considering all the best parameters

RF Output for $K=24$			
Training sets	5785		
Testing sets	720		
Validation sets	720		
Training function	Bootstrapping		
Output Function	Regression		
Number of trees	1	5	10
Number of leaves	5	5	5
Training RMSE	0.8725	0.8753	0.8756
Testing RMSE	0.0476	0.0089	0.0078
Regression	0.99967	0.99993	0.99994
Minimum MAPE	0.0061	0.0016	0.0111
Maximum MAPE	2.0735	0.6109	0.7953
Mean MAPE	0.2731	0.2286	0.2000

Figures 7(a) and 7(b) represent the comparative results and regression of forecasted PV output current versus actual targeted values obtained from the testing procedure of ANN, respectively. It can be observed that the forecasted hourly PV output current provide a very similar variation with minimum error as compared to the actual targeted values pattern.



(a)



(b)

Figure 7. STPCF for the next 24-hours using RF for the (a) forecasted PV output current versus actual targeted values, (b) regression of forecasted versus actual PV output current

3.3. MAPE comparison between the performance of ANN and RF

Tables 3 and 4 have shown the results of ANN and RF, respectively. The output result is obtained by considering to multiple time lags $K = 24$ hours during testing procedure for both techniques. The testing procedure for both techniques with multiple time lags is further investigated by comparing the MAPE results of PV output current.

By referring to the MAPE results of ANN and RF in Table 5, it can be observed that the ANN model produces a higher MAPE value of 5.0836%, in contrast with the MAPE of 0.0579% determined by the RF in day two. It is perspicuous in Table 5 that the RF provides most accurate prediction with the minimum average MAPE results in forecasting the PV output current as compared to the ANN. It is obvious that bagging technique improves the training and testing processes of RF in obtaining the best results with minimum error in forecasting. This implies that the ANN is far more complicated than the RF in terms of interpreting and understanding the weight, easy to over-fit the model and unpredicted in its performance.

Table 5. MAPE of forecasted PV output current for the next 24-hours obtained from the ANN and RF

Day	ANN	RF
	MAPE (%)	MAPE (%)
1	5.9144	0.1184
2	5.0836	0.0579
3	3.0745	0.3008
4	2.9441	0.1622
5	2.4173	0.1184
6	6.0031	0.1219
7	3.3752	0.7556
8	4.7268	0.1420
9	2.5261	0.0864
10	6.9916	0.2823
11	8.9894	0.7953
12	2.2936	0.0190
13	6.2087	0.0574
14	1.7835	0.2749
15	4.5219	0.3175
16	1.0799	0.0118
17	6.7759	0.2154
18	13.2377	0.1462
19	11.4929	0.0111
20	0.3173	0.0243
21	0.9236	0.2045
22	0.2832	0.0320
23	3.3776	0.0203
24	4.4665	0.4610
25	4.4561	0.0198
26	5.9024	0.1180
27	5.0569	0.3915
28	3.0791	0.3008
29	2.9465	0.2508
30	2.4028	0.1835
Average MAPE	4.4217	0.2000

4. CONCLUSION

The application of artificial neural network (ANN) and random forest (RF) with wavelet denoising and multiple time lags $K=24$ in performing short term photovoltaic current forecasting (STPCF) has been discussed elaborately in this paper. The results shown proved that the models proposed for the case study have the benefit of providing accurate result of STPCF. However, the RF method shown the important of choosing the accurate number of tree and leaf to be used as it will affect the performance of RF. The result shown that the RF method able to forecast the PV output current for the next 24 hours and provide more accurate results of STPCF with minimum error compared to ANN.

ACKNOWLEDGEMENT

This research was supported by the Long-Term Research Grant (LRGS), Ministry of Education Malaysia for the program titled "Decarbonisation of Grid with an Optimal Controller and Energy Management for Energy Storage System in Microgrid Applications" with project code 600-IRMI/LRGS 5/3 (001/2019). The authors would also like to acknowledge The Institute of Research Management & Innovation (IRMI), Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia for the facilities provided to support on this research.

REFERENCES

- [1] M. Hosenuzzaman, N. A. Rahim, J. Selvaraj, M. Hasanuzzaman, A. B. M. A. Malek, and A. Nahar, "Global prospects, progress, policies, and environmental impact of solar photovoltaic power generation," *Renew. Sustain. Energy Rev.*, vol. 41, pp. 284–297, 2015.
- [2] J. F. M. Pessanha and N. Leon, "Forecasting long-term electricity demand in the residential sector," *Procedia Comput. Sci.*, vol. 55, no. Itqm, pp. 529–538, 2015.
- [3] G. R. T. Esteves, B. Q. Bastos, F. L. Cyrino, R. F. Calili, and R. C. Souza, "Long term electricity forecast: A systematic review," *Procedia Comput. Sci.*, vol. 55, no. Itqm, pp. 549–558, 2015.
- [4] M. M. Othman, and I. Musirin, "Optimal sizing and operational strategy of hybrid renewable energy system using

Performance comparison of artificial intelligence techniques in short term... (Muhammad Murtadha Othman)

- homer." In *IEEE 2010 4th International Power Engineering and Optimization Conference (PEOCO)*, 2010.
- [5] A. Qais, M. M. Othman, N. Khamis, and I. Musirin, "Optimal sizing and operational strategy of PV and micro-hydro," In *2013 IEEE 7th International Power Engineering and Optimization Conference (PEOCO)*, pp. 714-717, 2013.
- [6] M. Ghazvini, A. Abbaspour-Tehrani-Fard, M. Fotuhi-Firuzabad, and M. M. Othman, "Optimizing size and operation of hybrid energy systems," In *2013 IEEE 7th International Power Engineering and Optimization Conference (PEOCO)*, pp. 489-494, 2013.
- [7] S. M. H. W. Dawi, M. M. Othman, I. Musirin, A. A. M. Kamaruzaman, A. M. Arriffin, and N. A. Salim, "Gamma Stirling Engine for a Small Design of Renewable Resource Model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 2, pp. 350-359, 2017.
- [8] F. Zahari, M. M. Othman, I. Musirin, A. A. M. Kamaruzaman, N. A. Salim, and B. N. Sheikh Rahimullah, "Design of a small renewable resource model based on the stirling engine with alpha and beta configurations," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 2, pp. 360-367, 2017.
- [9] F. Wang, Z. Mi, S. Su, and H. Zhao, "Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters," *Energies*, vol. 5, no. 5, pp. 1355-1370, 2012.
- [10] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol.105, pp. 569-582, 2017.
- [11] M. Ding, L. Wang, and R. Bi, "An ANN-based approach for forecasting the power output of photovoltaic system," *Procedia Environ. Sci.*, vol. 11, no. PART C, pp. 1308-1315, 2011.
- [12] P. Mandal, S. T. S. Madhira, A. Ul haque, J. Meng, and R. L. Pineda, "Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques," *Procedia Comput. Sci.*, vol. 12, no. 915, pp. 332-337, 2012.
- [13] I. A. Ibrahim, T. Khatib, A. Mohamed, and W. Elmenreich, "Modeling of the output current of a photovoltaic grid-connected system using random forests technique," *Energy Explor. Exploit.*, vol. 36, no. 1, pp. 132-148, 2018.
- [14] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, 2001.
- [15] M. Kayri, I. Kayri, and M. T. Gencoglu, "The performance comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by using photovoltaic and atmospheric data," *2017 14th Int. Conf. Eng. Mod. Electr. Syst. EMES 2017*, pp. 1-4, 2017.
- [16] M. H. H. Harun, M. M. Othman, and I. Musirin, "Short term load forecasting (STLF) using artificial neural network based multiple lags of time series," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5507 LNCS, no. PART 2, pp. 445-446, 2009.
- [17] M. H. H. Harun, M. M. Othman, and I. Musirin, "Short term load forecasting (STLF) using artificial neural network based multiple lags and stationary time series," In *2010 4th International Power Engineering and Optimization Conference (PEOCO)*, pp. 363-370, 2010.
- [18] M. M. Othman, M. H. H. Harun, and I. Musirin, "Forecasting short term electric load based on stationary output of artificial neural network considering sequential process of feature extraction methods," In *2012 IEEE International Power Engineering and Optimization Conference Melaka, Malaysia*, pp. 485-489, 2012.
- [19] M. M. Othman, M. H. H. Harun, N. A. Salim, and M. L. Othman, "Sequential process of feature extraction methods for artificial neural network in short term load forecasting," *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 19, pp. 8830-8838, 2015.
- [20] M. M. Othman, M. H. H. Harun, and I. Musirin, "Short term load forecasting using artificial neural network with feature extraction method and stationary output," In *2012 IEEE International Power Engineering and Optimization Conference Melaka, Malaysia*, pp. 480-484, 2012.
- [21] J. S. Armstrong, "Illusions in Regression Analysis @ upenn.academia.edu," vol. 2012, no. 3, pp. 961-967, 2012.
- [22] P.-H. Chiang, S. P. V. Chiluvuri, S. Dey, and T. Q. Nguyen, "Forecasting of Solar Photovoltaic System Power Generation Using Wavelet Decomposition and Bias-Compensated Random Forest," *2017 Ninth Annu. IEEE Green Technol. Conf.*, pp. 260-266, 2017.
- [23] V. Lo Brano, G. Ciulla, and M. Di Falco, "Artificial Neural Networks to Predict the Power Output of a PV Panel," *Int. J. Photoenergy*, vol. 2014, p. 12, 2014.
- [24] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker J. D. Renné, and T. E. Hoff, "Validation of Short and Medium Term Operational Solar Radiation Forecasts in the U.S.," no. 2010, 2009.
- [25] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Sol. Energy*, vol. 83, no. 10, pp. 1772-1783, 2009.