Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MAYFLY OPTIMIZED EXTREME GRADIENT BOOSTING FOR SINGLE DECOMPOSITION BASED SHORT TERM SOLAR POWER PREDICTION

RAJ KUMAR PARIDA [1], MONIDEEPA ROY [2], *AJAYA KUMAR PARIDA [3], ASIF UDDIN KHAN [4], SUDAN JHA [5]

[1234] School of Computer Engineering, KIIT University, India [5] Department of Computer Science and Engineering, Kathmandu University, Nepal

Email: [1] 2081024@kiit.ac.in, [2] monideepafcs@kiit.ac.in, [3] ajaya.paridafcs@kiit.ac.in, [4] asif.khanfcs@kiit.ac.in, [5] sudan.jha@ku.edu.np

ABSTRACT

Solar power is one of the cleanest form of renewable energy, which can be mostly used for grid interactive mode without much difficulty. PV power prediction plays an important role when grid connected mode is considered. At present scenario proper prediction is greatly valued as it is directly related to various environmental conditions. Accurate prediction helps in proper maintenance planning. This paper develops a hybrid model which employs mayfly optimization and extreme gradient boosting technique (XGB). The Meta heuristic optimization technique is used for obtaining the optimized hyper parameters. As per the previous literature study XGB is observed to give more accurate result as compared to previously implemented techniques like Support Vector Machine, Extreme Learning Machine etc., but the main drawback of XGB is observed when external noise is inserted i.e it faces the over fitting problem. The proposed model with the help of optimization technique optimizes the learning rate and maximum depth thus providing more accurate result. The proposed method is verified under different weather condition and different geographical location. The experimental result supports the fact that the proposed hybrid model performs better than the Random Biased Functional Network technique, Support Vector Machine and Extreme Learning Machine technique. The performance accuracy is supported by different statistical tests like Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Correlation coefficient (R2). The MAPE for RBFN is obtained to be 4.87 % which is greater than the proposed hybrid model by 1.19 %.

Keywords: *Solar Power Prediction, Meta heuristic optimization, mayfly optimization and extreme gradient boosting technique*

1. INTRODUCTION

Because of the serious problem of pollution and different environmental hazards the modern world is opting for different renewable energy resources implementation, like wind energy, solar energy, tidal energy etc. Among all other resources solar energy became the most effective supplement for the conventional energy resources. Till date solar has been obtaining more attention from scientists and industrial applications. The main challenges are observed in the field where penetration is required. Accurate prediction makes a better and successful integration of solar PV into the power grid.

Prediction plays an important role while managing the increment and decrement in loads, matching of the peak load, decision making purpose. The solar power stability is majorly affected by different atmospheric conditions like solar irradiance, different climatic conditions (i.e variation in the atmospheric condition due to change in seasons.) It also acts a prime factor for the operation of energy management system. Table.1 shows different forecasting horizons and there corresponding importance. Accurate solar power prediction is essential for various applications [1]. A great progress has been achieved by the researchers over the past years in the field of solar power prediction. Various machine learning techniques have been have been used for prediction of solar power. Machine learning techniques includes SVM, NN, ELM, CNN etc. [2-7]. over the years. In the year 2016 two researchers named Guestrin and Chen proposed an ensemble tree based method known as extreme gradient boosting (XGB). This method possessed a higher computational power and remedy for over-fitting problem [9]. Different machine learning techniques used in the previous years for prediction includes different statistical, AI based methods and hybrid methods. ARMA, ARIMA, ARMAX has been implemented for prediction of solar power and irradiation [10-13]. To remove the disadvantage of the statistical methods different machine learning process like SVM, NN, RBFLN, ELM etc. has been used for various prediction horizons. More recently different hybrid models and ensemble models are used for more accurate prediction. In the previous literature survey it has been observed that the XGB model when combined with deep learning technique

showed better prediction output when compared with SVM. It is also observed that the XGB model is by far not used for the prediction of solar power. Hence as a new contribution to the previous literature survey in this paper XGB is used for predicting the solar power where the hyper parameters of the XGB technique is optimized using the mayfly optimization process. To make the result more accurate even without combining with the deep learning technique, the nonlinear solar power is decomposed using the VMD technique [14]. When compared with SVM, ELM, XGB-DNN showed better prediction accuracy. The main concern in solar power prediction is sudden change in atmospheric condition, for example in case of sudden cloud covering or foggy weather the uninterrupted power supply must be continuing. Hence a more accurate prediction technique is required. The basic XGB model is observed to be more efficient based on the previous literature study but it is observed that the optimization process makes it more accurate in case of noisy

condition by optimizing the maximizing depth and learning rate.

The paper is organized as follows: Section 2 describes the different techniques used for prediction purpose, section 3 describes the data collection and pre-processing of the data. Section 4

shows the experimental verification of the proposed technique and lastly section 5 concludes the paper and also discus about the future study

Table.1. application of solar power prediction at different prediction horizons

Prediction horizon	Time horizon	Application
Very Short Term	Up to few minutes	Unit commitment
Short Term	1hour-1day	Decision making
Medium Term	1 day to -1 month	Maintenance
Long Term	More than a month	Planning

2. METHODOLOGY

This section gives an overview of different models used in the paper for designing of the proposed model and predicting the solar power. Different techniques include the decomposition process using the Variational Mode Decomposition technique followed by extreme gradient boosting process where the parameters are optimized using the mayfly optimization process. Figure.1 shows the schematic representation of the proposed model

2.1. Extreme Gradient Boosting method

This method was originally proposed as a university research project at Washington by Tiang Chen, It has very little contribution in the field of solar power prediction. [15]

Let us consider a specific number of samples (n) such that $D = \{x, y\}$ m, where D is the input vector with m features in it, and y being its corresponding output. This is a form of binary tree where the splitting rule is used to decide whether the sample will belong to the left branch or right branch depending upon the single input. Individual leaf is given a value that is the predicted output. The splitting rule makes each leaf a subset of the input. Like other boosting techniques the XGB makes a set of regression trees in a sequential manner and combines them by simple addition for predicting the output.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

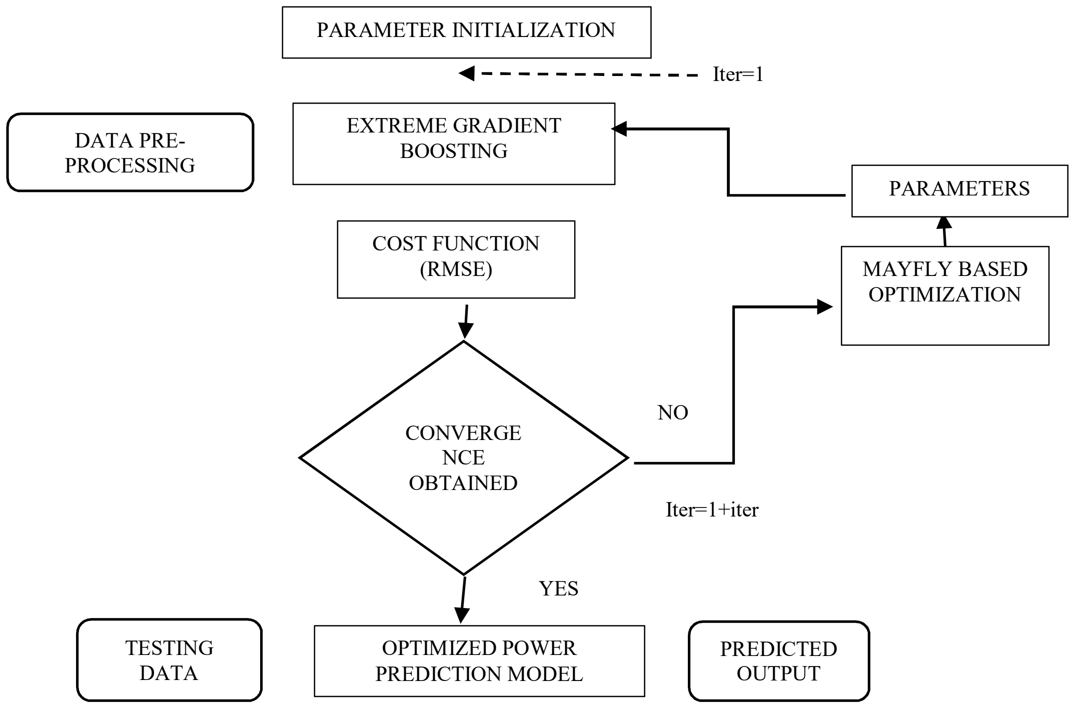

Figure.1. structure of proposed model

$y_i(t-1)$ Being the predicted output (yi) at the instant *(t-1)* and *i.* hence the XGB forms a tree of f_t (prediction function) for minimizing f_t (objective functions) as given below

$$L_t = \sum_{i=1}^{n} l_f\left(y_i, y_i(t)\right) + \delta\left(f(t)\right)$$

$$= \sum_{i=l}^{n} l_f\left(y_i, y_i(t-1) + f_t(x_i)\right) + \delta(f_t)$$

(1)

l_f denotes the loss function, i.e. it determines the prediction accuracy. Mean Square error is considered to be the loss function. δ (Regularization) is used to reduce the problem of over fitting.

$$\delta(f) = \lambda T + \frac{1}{2}\gamma \sum_{j=1}^{T} w^2{}_j \qquad (2)$$

Where λ represents the minimum reduction, γ being the regularization, Taylor approximation of L_t is used to obtain f_t.

$$L_t \approx \sum_{i=1}^{n}\left[l\left(y_i, y_i(t-1) + g_1 f_1(x_1)\right) + \frac{1}{2} h_1 f_i^2(x_i)\right] + \delta(f_t)$$

(3)

$y_i(t-1)$ is obtained at step t, therefore removing the constant from eq (3) and hence minimizing the objective function

$$L_t = \sum_{i=1}^{n}[g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i)] + \delta(f_t)$$

$$= \sum_{i=1}^{T}\left[\left(\sum_{i\in I_j} g_i\right) w_j + \frac{1}{2}\left(\sum_{i\in I_j} h_i + \lambda\right) w_j^2\right] + \delta T$$

(4)

Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Ij being the subset of input set associated with leaf j,

i.e. Optimal w_j and L_t can be obtained as follows:

$$w_j^* = \frac{-\sum_{i\in Ij} g_i}{\sum_{i\in I_j} h_i + \zeta}' L_t(q)$$

$$= -\frac{1}{2}\sum_j^T \frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i + \zeta} + \delta T$$

(5)

Eq (5) is used to get the optimal weight when q (structure of the tree) is already determined. In this method feature values are used to sort the input set to construct the tree with depth zero. T every level of iteration a new tree is obtained by splitting the present value. Splitting now maximizes the loss reduction which is obtained as

$$L_s = \frac{1}{2}\left[\frac{\left(\sum_{i\in Ij} g_i\right)^2}{\sum_{i\in L_R g} h_i + \delta} + \frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_R g} h_i + \delta} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in Ift} h_i + \delta}\right]$$

(6)

I_{Rg} and I_{ft} is a set associated with left and right node respectively. Before performing the XGB certain parameters should be decided like the learning rate, maximum depth, sub sample and learning rate. Table.2 shows the algorithm of XGB technique.

Table.2 algorithm for XGB technique

Initialise the parameters: step size (η), regularization parameter number of iteration and minimum reduction parameter (δ).
Using the values of the features set the input values.
Calculate g_i and h_i for every iteration
Obtain the split value by examining the current tree
Compute the weight of the new tree
Compute the new predicted data
Return to the trained XGB model

2.2. Mayfly Optimization Technique

It is a technique where the concept of May flies are included for obtaining the optimal value. The name is so given as the insects are found in the month of May in United Kingdom. They stay in the form of aquatic nymphs until they are matured enough to move towards the surface of the water. The MMFs (matured male Mayflies) joins in swarms above the water surface for the purpose of alluring its counter female Mayflies. In this process they create a format. The copulating takes place when the female moves towards these swarms and resulting in release of eggs in the water and this continuous process continues. Therefore executing the required alterations required for finer execution of the process. The location of the MMFs is updated using the equation (7)

$$x_i^{t+1} = x_i^t + v_i^{t+1} \qquad (7)$$

Where x_i^t represents the current location, and the position x_i^{t+1} is obtained along with the velocity factor. The MMFs maintains the position some meters over the surface of the water thus establishing an increase in speed. Thus the velocity of the MMFs can be determined as

$$V_{mn}^{t+1} = g^* V_{mn}^t + a_{C1}^* e^{-Br^2} * \left(Pb_{mn} - x_{mn}^t\right) + a_{C2} * e^{-br^2} * (Pb - x_{mn}^t)$$

(8)

$$Pb_{mn} = \left\{ i_{fitness} \left(x_{kj}^{t+1}\right)^{x_m^{t+1}} \left\langle fitness(Pb_m) \right. \right.$$

(9)

Where, V is the velocity, *mn* defines the velocity of *m* mayfly in the six n

kj being the velocity of an individual *k* mayfly in the size j at the time t. Where ac1 and ac2 determines the positive attraction constant, which is useful for calculating the cognitive and social element. g gravitational coefficient where as *B* gives the visibility coefficient. Pb_{mn} is the ideal position travelled by an insect m. the Pb value gets updated at every step having the finest MMF position. In a similar manner to obtain the FMF. This is followed by producing a crossover where

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

two offspring's are obtained and then mutation is conducted.

3. DATA COLLECTION

In this section the nonlinear data collected from the nrel site [16] is observed to be highly nonlinear. In order to make the model more accurate the raw data is subjected to normalization and decomposition. In ref [14] a through knowledge about the VMD based technique is discussed. Where it is observed from the experimental setup that the decomposed data performed better than the individual model. The entire data is normalized using the equation (10)

$$A_{norm} = \frac{A_{act} - A_{\min}}{A_{\max} - A_{\min}} \quad (10)$$

Here A_{norm} is the normalized data sample of a precise sample, A_{act} is the actual value, whereas A_{min} and A_{max} are the minimum and maximum values obtained in the data set. Table 3. Shows different parameters of the XGB model without applying optimization process. The hybrid model of MMF- XGB process is performed by initialization of four different hyper-parameters within the range of their search value. The learning rate is randomly selected between [0.01 0.5], the depth of the tree ranges between 3 to 10, regularization factor is initiated between the range of 100 and 200. The boosting number has been set till 200 subjected to the reduction in cost function. The swam size for MMF is set to 100 and the iter-max is selected till 200.

4. EXPERIMENTAL RESULT ANALYSIS

The experimental performance is performed with Intel Core i3 (2.0 GHz.) processor and 8GB RAM, the proposed model is developed in MATLAB. The performance of the proposed model is evaluated using different matrices namely Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) There are various other prediction measurement techniques but in this paper only four are considered for obtaining the prediction accuracy. The prediction matrices helps in comparative study between different other prediction models. Table.3 shows the value of different optimized parameters obtained by using MMF and the minimization function of MAE.

$$MAPE = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\left| T_t(x_i) - O_p(x_i) \right|}{T_t(x_i)} \right) *100 \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left| T_t(x_i) - O_p(x_i) \right|^2} \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| T_t(x_i) - O_p(x_i) \right| \quad (13)$$

To have a better evaluation of the proposed model (MMF-XGB), the present study is compared with other methods used mostly for prediction of solar power that includes ELM, SVM and RBFN. The data set consist of 2520 samples, the data has been collected with an interval of 15min from 6am to 6pm. The negative impact of the data is removed using equation (10).

Table.3. optimized values of different parameters

Parameters	Optimised Values
Learning rate	0.32
Maximum Depth	10
Regularization	100
Ratio of samples per input	0.95

Two different seasons has been considered for predicting the solar power, summer season and rainy season has been taken into consideration for performing the result analysis. Out of the entire season, data of one single month is consider for calculating the models accuracy.

For summer season the month of March is taken into consideration whereas for rainy season the month of June is considered. Figure.3. shows the original sunny weather of Alabama, Figure.4 shows a comparative study of different methods considered for obtaining the performance accuracy of the proposed method for summer season.

Table 4 gives the corresponding values of the comparative analysis. Although the proposed model is a better functioning prediction technique when noisy data is considered but still now sudden change in weather condition is not measured in terms of error calculation. In our work we have not focused on partial shading which also plays a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

major role in describing the performance of the proposed model.

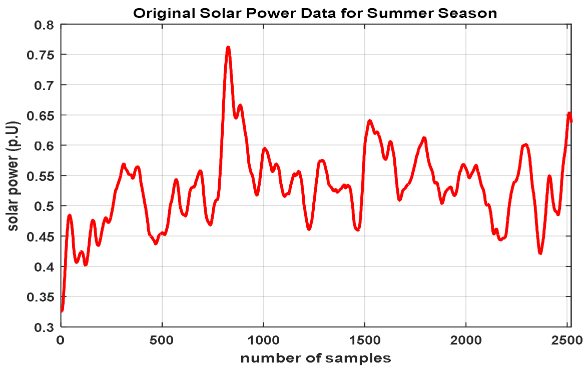

Figure.3. Original Solar Power for Summer Season

Table.5 gives a comparison of previously implemented techniques for prediction of solar irradiation and power from which we it is clear that the best performing model is XGB which is further modified in the proposed study to obtain more accurate prediction. It is clear that the proposed technique performs better even when subjected to different weather condition.

5. CONCLUSION

It is observed from experimental verification that the proposed technique out performs different other models shown in this study that has been well used by previous researchers for prediction of solar powers. The model's accuracy is obtained based on different calculations as given by various performance indices calculation. Table.4. gives a better overview of the outperformance of the proposed technique as compared to other techniques. It is observed that the model performs better for seasonal variations also. The decomposition method is used to obtain the prediction accuracy up to a maximum limit. The main contribution of the proposed technique is that it gives better predicted output even when subjected to different atmospheric condition. The main objective of any prediction model is to predict the desired output with less computational time and this is fulfilled by the proposed method.

In future study we can be use this prediction solar power technique with different input parameters (weather fluctuations) and also be tested for sudden variation in the input data and thus utilization in green building can also be opted.

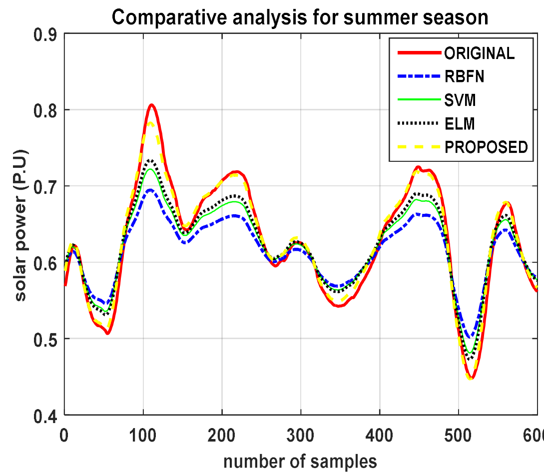

Figure.4. Comparative analysis for Summer Season

Table.4. Comparative analysis of different prediction model for solar power prediction

Season	Method	MAPE	MAE	RMSE
summer	RBFN	4.871	0.031	0.039
	SVM	3.44	0.022	0.027
	ELM	2.89	0.018	0.023
	Proposed	1.196	0.007	0.0094
Rainy	RBFN	4.935	0.046	0.051
	SVM	3.850	0.025	0.038
	ELM	3.013	0.023	0.031
	Proposed	1.242	0.024	0.013

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table.5.Summary of existing solar irradiance forecasting approaches

Prediction Model	Dataset	Feature selected	Best performing model	ref
MARS, GLM, GLMNET, ICR, KNN, LR, SVR, NH,GP, GB, GBst, MLP,	National Solar Radiation Database (NSRDB) of 7 different Stations in USA	-	SVR,RF, MLP	17
RF, ANN, XGBoost	Hawaii meteorological data of USA	RF, PCA	XGBOOST	18
MLP, Physics Based Deterministic Model	Numerical Weather Prediction Data of USA		MLP	19
SVR, GBR, RFR, Hybrid (ALL)	Numerical Weather Prediction data of 7 different locations in spain		SVR hybrid	20

REFERENCEs

[1] Sharadga, Hussein, Shima Hajimirza, and Robert S. Balog. "Time series forecasting of solar power generation for large-scale photovoltaic plants." *Renewable Energy* 150 (2020): 797-807.

[2] Chen, Changsong, et al. "Online 24-h solar power forecasting based on weather type classification using artificial neural network." *Solar energy* 85.11 (2011): 2856-2870.

[3] Wan, Can, et al. "Photovoltaic and solar power forecasting for smart grid energy management." *CSEE Journal of Power and Energy Systems* 1.4 (2015): 38-46.

[4] Zeng, Jianwu, and Wei Qiao. "Short-term solar power prediction using a support vector machine." *Renewable energy* 52 (2013): 118-127.

[5] Jang, Han Seung, et al. "Solar power prediction based on satellite images and support vector machine." *IEEE Transactions on Sustainable Energy* 7.3 (2016): 1255-1263.

[6] Majumder, Irani, Manoja Kumar Behera, and Niranjan Nayak. "Solar power forecasting using a hybrid EMD-ELM method." *2017 international conference on circuit, power and computing technologies (ICCPCT)*. IEEE, 2017.

[7] Izgi, Ercan, et al. "Short–mid-term solar power prediction by using artificial neural networks." *Solar Energy* 86.2 (2012): 725-733.

[8] Ahmed, Razin, et al. "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization." *Renewable and Sustainable Energy Reviews* 124 (2020): 109792.

[9] Torres-Barrán, Alberto, Álvaro Alonso, and José R. Dorronsoro. "Regression tree ensembles for wind energy and solar radiation prediction." *Neurocomputing* 326 (2019): 151-160.

[10] Wan, Can, et al. "Photovoltaic and solar power forecasting for smart grid energy management." *CSEE Journal of Power and Energy Systems* 1.4 (2015): 38-46.

[11] Huang, Rui, et al. "Solar generation prediction using the ARMA model in a laboratory-level micro-grid." *2012 IEEE third international conference on smart grid communications (SmartGridComm)*. IEEE, 2012.

[12] Ren, Ye, P. N. Suganthan, and N. Srikanth. "Ensemble methods for wind and solar power forecasting—A state-of-the-art review." *Renewable and Sustainable Energy Reviews* 50 (2015): 82-91.

[13] Wan, Can, et al. "Photovoltaic and solar power forecasting for smart grid energy management." *CSEE Journal of Power and Energy Systems* 1.4 (2015): 38-46.

[14] Majumder, Irani, P. K. Dash, and Ranjeeta Bisoi. "Variational mode decomposition based low rank robust kernel extreme

learning machine for solar irradiation forecasting." *Energy conversion and management* 171 (2018): 787-806.

[15] Ahmed, Razin, et al. "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization." *Renewable and Sustainable Energy Reviews* 124 (2020): 109792.

[16] https://www.nrel.gov/solar/ [last operated on 2nd jan 2023]

[17] Yagli, Gokhan Mert, Dazhi Yang, and Dipti Srinivasan. "Automatic hourly solar forecasting using machine learning models." *Renewable and Sustainable Energy Reviews* 105 (2019): 487-498.

[18] Munawar, Usman, and Zhanle Wang. "A framework of using machine learning approaches for short-term solar power forecasting." *Journal of Electrical Engineering & Technology* 15 (2020): 561-569.

[19] Pierro, Marco, et al. "Deterministic and stochastic approaches for day-ahead solar power forecasting." *Journal of Solar Energy Engineering* 139.2 (2017).

[20] Gala, Yvonne, et al. "Hybrid machine learning forecasting of solar radiation values." *Neurocomputing* 176 (2016): 48-59.