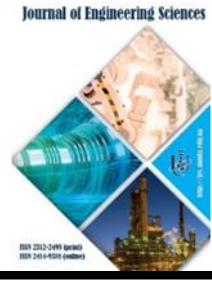
JOURNAL OF ENGINEERING SCIENCES

Volume 9, Issue 2 (2022)

Mbah O. M., Madueke C. I., Umunakwe R., Maurison N. A. (2022). Extreme gradient boosting: A machine learning technique for daily global solar radiation forecasting on tilted surfaces, Vol. 9(2), pp. E1-E6, doi: 10.21272/jes.2022.9(2).e1



Extreme Gradient Boosting: A Machine Learning Technique for Daily Global Solar Radiation Forecasting on Tilted Surfaces

Mbah O. M.[1*], Madueke C. I.[2], Umunakwe R.[2], Maurison N. A.[3]

[1] Department of Mechanical Engineering, Federal University Oye-Ekiti, Ikole, City, 370105, Ekiti-State, Nigeria;
[2] Department of Material and Metallurgical Engineering, Federal University Oye-Ekiti, Street, Ikole, 370105, Ekiti -State, Nigeria;
[3] Department of System Engineering, University of Lagos, Street, Akoka, 100213, Lagos State, Nigeria

Article info:
Submitted: August 13, 2022
Accepted for publication: October 27, 2022
Available online: November 2, 2022

***Corresponding email:**
oguejiofor.mbah@fuoye.edu.ng

Abstract. Enhancing solar irradiance and accurate forecasting is required for improved performance of photovoltaic and solar thermal systems. In this study, Extreme Gradient Boosting (XGBoost) model was developed using three input parameters (time, day number, and horizontal solar radiation) and was utilized to forecast daily global solar radiation on tilted surfaces. The proposed model was built using XGBRegressor with five generations, 100 n estimators, and a learning rate of 0.1. Three statistical metrics, such as the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), were used to compare the model's results to observed solar radiation data from the Nation Centre for Energy, Research and Development, University of Nigeria, Nsukka. The results showed improved prediction accuracy and XGBoost capability to estimate daily global solar radiation on tilted surfaces. In the training section, the proposed model had a statistical performance of $R^2 = 0.9977$, RMSE = 1.6988, and MAE = 1.081, and in the testing section, $R^2 = 0.9934$, RMSE = 2.8558, and MAE = 2.033. XGBoost model demonstrated a better performance when compared with other models in the literature. As a result, the proposed model provides an effective approach for estimating solar radiation.

Keywords: machine learning model, extreme gradient boosting, solar radiation prediction.

1 Introduction

Electromagnetic radiation from the sun is referred to as solar radiation. Captured solar radiation can be transformed into electricity or heat energy, which can be utilized to heat thermal systems and power homes. When designing, installing, and forecasting a solar conversion system's energy output, knowing a particular location's solar irradiance is extremely important for maximum performance. Solar systems are installed on tilted surfaces to maximize their exposure to direct sunlight, which increases their efficiency. Most solar radiation measuring instruments take measurements on horizontal surfaces rather than tilted surfaces. As a result, many techniques for estimating global solar radiation on tilted surfaces, such as empirical models [1, 2], satellite-derived models [3] and stochastic models [4] were developed. Empirical models are dependent on sunshine duration, the amount of cloud cover, and the maximum and minimum temperatures [5, 6]. However, problems arise when there is a dearth of or limited sunshine data and maximum and minimum temperatures. Besides that, both the satellite and stochastic models have drawbacks in terms of cost and behavioral characteristics. As a result, these techniques are unstable when used for solar irradiance forecasting and produce approximations. They are also ineffective when the database contains null or missing data. However, the advent of machine learning appears promising in overcoming these obstacles.

2 Literature Review

In Nigeria, empirical models have been used in predicting solar radiation [7, 8]. Chabane et al. [9] developed a model for daily solar radiation estimation in Biskra, Algeria, based on wavelength optical depth. Similarly, a mathematical model was developed by Herath et al. [10] for predicting solar radiation using multiple linear regression analysis. Sunshine hours, humidity, pressure, temperature, radiation, sunset time, date, wind direction, and time served as the model's input parameters. An Artificial Neural Network (ANN) was used to evaluate the model's performance, and the correlation coefficient was 0.5973. Shourehdeli et al. [11] assessed injector fuel efficiency in basic mode utilizing four conventional models of isentropic coefficients.

Machine learning approaches have been broadly utilized in estimating global solar radiation in many regions. For example, Olatomiwa et al. [12] utilized an adaptive network-based fuzzy inference system (ANFIS) to forecast solar irradiance in Nigeria. The ANFIS model produced a correlation coefficient of 0.6567. Similarly, Hacioğlu and Rifat [13] devised a model for estimating solar radiation based on humidity, pressure, temperature, and wind speed. Both linear and gaussian regression models were used to assess the model's performance. Also, the Support Vector Machine Regression (SVM-R) algorithm has been employed in predicting global solar radiation [14, 15, 16, 17]. Similarly, the ANN technique has been utilized in predicting daily global solar radiation and other observed data [18, 19, 20, 21]. Marzo et al. [22] created an ANN model for estimating daily solar radiation utilizing three input parameters: extraterrestrial radiation, daily maximum, and minimum temperature. Data on solar radiation was used to validate the model. The average relative root mean square derivation (RRMSD) produced by ANN was 0.13, and its correlation coefficient was 0.8.

Furthermore, various researchers have used hybrid machine-learning models to predict global solar radiation. Torabi et al. [23] developed a Cluster-Based Approach that uses support vector machine and artificial neural networks (CBA-SVM-ANN) to estimate the horizontal surface's daily solar radiation. The findings showed that the independent models outperformed the hybrid model (CBA-SVM-ANN) based on the mean absolute percentage inaccuracy. Likewise, three hybrid models (gradient-boosted regression, random forest, and support vector machine) were used by Gala et al. [24] to forecast global solar radiation. The outcomes demonstrate how well the hybrid model works. Achour et al. [25] used the hybrid model to estimate monthly mean global solar radiation in Southern Algeria. Comparative performance of three machine learning models (SVM, ANN, and ANFIS) was done by Quej et al. [26] in estimating daily horizontal surface solar radiation in Mexico. From the results, the support vector machine performed better. A power coefficient equation was proposed by Augbulut et al. [27] in Turkey for predicting system power output and validated the proposed model using four machine learning approaches (Deep Learning, SVM, Kernel and Nearest-Neighbor, and ANN).

Furthermore, four Turkish districts used four machine-learning algorithms to forecast daily solar radiation using six input parameters [28]. The results showed that ANN outperformed the other models. Rabehi et al. [29] created a hybrid model for predicting global solar radiation in Southern Algeria that combines a multilinear-layer perceptron, a boosted decision tree, and linear regression (MLP-BDT-LR). The model's performance was evaluated using the Applied Research Unit for Renewable Energies data. Mbah et al. [30] compared two machine learning techniques (K-nearest neighbor and extreme gradient boosting) to four sky models (Tian, Koronakis, Badescu, Liu, and Jordan) for estimating daily global solar radiation on a tilted surface. According to the results, extreme gradient boosting outperformed other models. Similarly, Feng et al. [31] compared four machine-learning models and four empirical models for predicting daily global solar radiation in China. The MEA-ANN model generally outperformed the other four machine learning models with a correlative coefficient of 0.779, while the fan model outperformed the other four empirical models with a correlative value of 0.733.

This research aims to investigate the applicability of Extreme Gradient Boosting (XGBboost) in forecasting global solar radiation in Nigeria. This inquiry was inspired by the importance of reliable solar radiation data in Engineering, Agriculture, Ecology, and Hydrological investigations.

3 Research Methodology

3.1 Site and measurement

The proposed supervised machine learning model was developed using a daily global solar radiation dataset measured at the Energy Centre from 2016 to 2017, located inside the University of Nigeria, Nsukka. The model's three input parameters were time, day number, and horizontal solar radiation. Solar radiation data was measured using two solarimeters. The first solarimeter measures hourly horizontal solar radiation, while the second measures solar radiation at 15° inclinations toward the south. Time was measured in ten-minute incremental steps using a smartwatch. The proposed model was developed in Python using XGBRegressor with five generations, 100 estimators, and a learning rate of 0.1. The measured solar radiation dataset was split into two sets for training and testing the model. 70% of the dataset was used for sample training, with the remaining 30% used for model testing.

3.2 Machine learning models

Machine learning models are generated by algorithms that try to find the relationship between feature and target variables to detect patterns and make decisions based on previously unseen datasets. The proposed machine learning model is described below.

The Extreme Gradient Boosting algorithm is a supervised machine learning algorithm that forecasts a target output using train data with various features. It applies to both classification and regression problems. The XGBoost model's regularized objective is given by

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, y_i^{\wedge(t-1)} + f_t(X_i)\right) + \Omega(f_t) \quad (1)$$

where Ω – a regularized term used to avoid overfitting; l – loss function between the actual and estimated value; f_t – tree output; X_i – training dataset; Y_i – predicted dataset.

The second-order Taylor expansion of the loss function is expressed as follows:

$$Obj^{(t)} = \sum_{i=1}^{n} l\,(y_i, y_i^{\wedge(t-1)}) + g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i) + \Omega(f_t) \quad (2)$$

where g_i, h_i – the loss function's first and second derivatives, respectively.

3.3 Evaluation of model performance

The XGBoost model's performance was evaluated using the statistical indicators listed below.

The coefficient of determination is a measure that assesses the ability of a model to predict an outcome in linear regression. The coefficient of determination is given as

$$R^2 = \frac{\sum_{i=1}^{n}[(H_{mi}-\overline{H_{mi}}).(H_{pi}-\overline{H_{pi}})]^2}{\sum_{i=1}^{n}(H_{mi}-\overline{H_{mi}})^2 . \sum_{i=1}^{n}(H_{pi}-\overline{H_{pi}})^2}. \quad (3)$$

Also, MAE is determined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{H_{pi}-H_{mi}}{n}\right). \quad (4)$$

The mean absolute error is the sum of all differences between the observed and measured values. It is employed in the evaluation of regression models.

Root mean square error, which is frequently used for regression analysis, informs how concentrated the data points are around the best fit line:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(H_{pi} - H_{mi})^2}, \quad (5)$$

where n – sample size; H_{mi} – actual value; H_{pi} – estimated value.

4 Results and Discussion

4.1 Statistical parameters of the dataset

The XGBoost model was developed in this study using three input parameters: time, day number, and horizontal surface hourly solar irradiance. Table 1 lists the statistical parameters for the dataset used in this study.

Table 1 – Statistical parameters of the dataset

Process	R^2	RMSE	MAE
Training	0.9977	1.6988	1.081
Testing	0.9934	2.8558	2.033

4.2 XGboost model analysis

Figure 1 depicts the solar radiation training dataset and predicted values using the XGBoost model, while Figure 2 portrays the testing dataset and the predicted values.

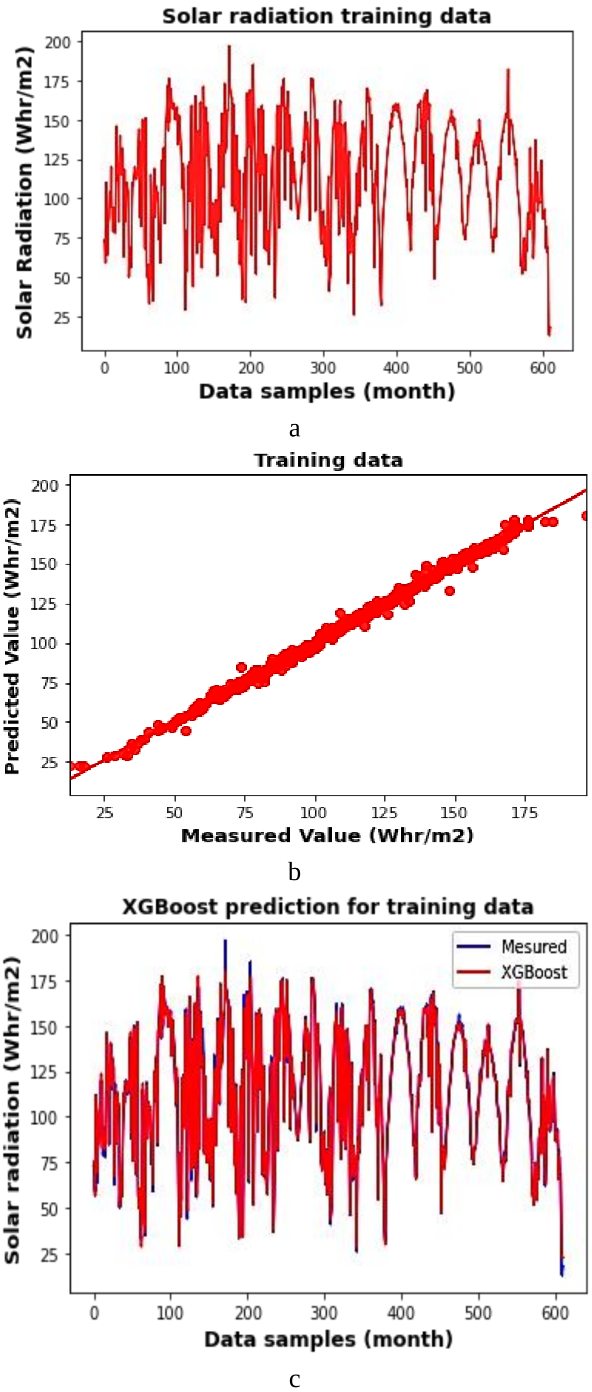
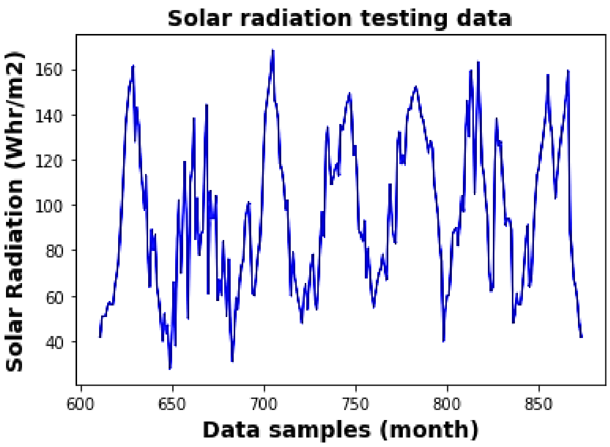
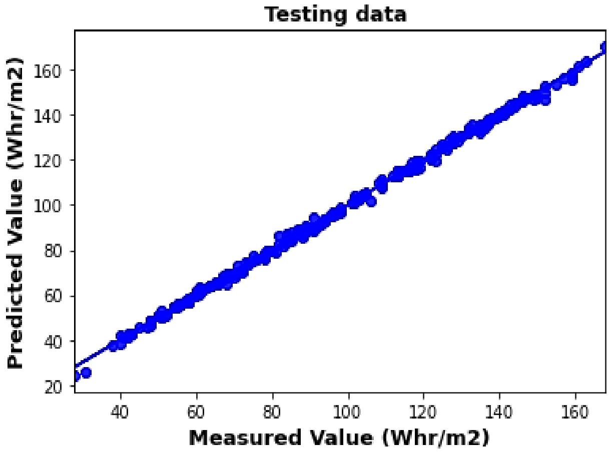


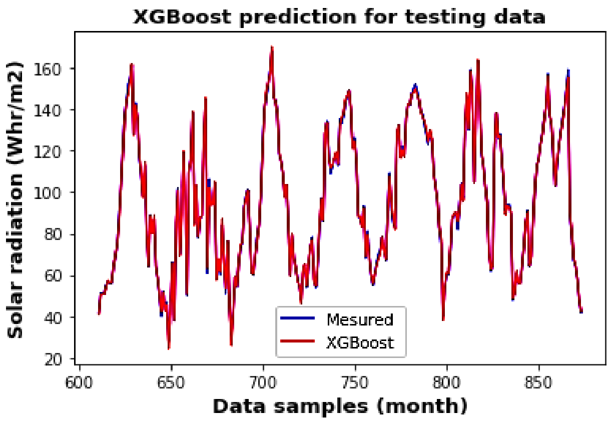
Figure 1 –Samples of training data (a), scatter plots of training data and predicted values using XGBoost model (b), and forecasting solar radiation using XGBoost model (c)



a



b



c

Figure 2 – Samples of testing data (a), scatter plots of testing data and predicted values using XGBoost model (b), and forecasting solar radiation using XGBoost model (c)

Figures 1b and 2b depict the relationship between the developed model and observed solar radiation data on the tilted surface during sample training and testing with scattered plots.

While also providing a more tightly compacted set of data points between the proposed model and the observed solar radiation data, both figures demonstrated a very high similar positive linear correlation. As a result, the XGBoost model has a high coefficient of determination for both training and testing data.

The XGBoost model's performance in estimating global solar radiation on a tilted surface was assessed using statistical measures such as coefficient of determination, root mean square error, and mean absolute error.

Table 2 summarizes the XGBoost model's prediction accuracy.

Table 2 – Statistical performance of the XGBoost model

Variable	Min	Max	Mean	Standard deviation
Time	9.16	15.90	12.55	2.01
Day number	20	411	359.14	29.18
Solar radiation on a horizontal surface	12	160	94.80	30.28
Solar radiation on a tilted surface	13	197	108.71	35.43

4.3 Model validation

A correlation between the proposed model and nine other solar radiation prediction models previously proposed by various authors [9, 10, 12, 14, 21, 22, 29, 31] validates XGBoost precision and accuracy in forecasting solar radiation.

As shown in Table 3, a statistical indicator of coefficient of determination was used to compare the proposed model prediction accuracy with other developed models in literature in order to evaluate its performance.

Table 3 – Comparison between the XGBoost model and existing models from the literature

Reference	Model type	Inputs	Country	R^2
Chabane et al. [9]	Empirical	3	Algeria	0.757
Feng et al. [30]	Empirical	1	China	0.733
Herath et al. [10]	Empirical	10	Sri Lanka	0.597
Guermoui et al [14]	SVM-R	1	Algeria	0.974
Olatomiwa et al. [12]	ANFIS	3	Nigeria	0.657
Rahimikhoob [21]	ANN	3	Iran	0.889
Marzo et al. [22]	ANN	3	Chile, Israel, Australia	0.800
Rabehi et al. [29]	Hybrid	4	Algeria	0.977
Feng et al. [30]	Hybrid	1	China	0.779
Present study	XGBoost	3	Nigeria	0.993

Table 3 shows that the XGBoost model has the highest coefficient of determination of the nine models tested. As a result, the proposed model outperforms the benchmark models (empirical, other machine learning, and hybrid models).

5. Conclusion

The importance of accurate solar radiation data in engineering, geography, crop production, and ecological studies cannot be overstated. This study developed an Extreme gradient boosting (XGBoost) model for estimating global solar radiation on tilted surfaces using three input parameters (time, day number, and horizontal solar radiation). The following findings are based on the proposed model's performance evaluation.

The analysis revealed that the XGBoost model predicts global solar radiation on tilted surfaces with high precision and accuracy, as evidenced by the statistical performance metrics obtained during the model testing; R^2, RMSE, and MAE values of 0.9934, 2.8558, and 2.033, respectively.

The proposed model shows significant prediction improvement compared to the literature's reference models (empirical, other machine learning, and hybrid models).

The proposed model can be deployed into a web application browser to predict solar irradiance on tilted surfaces because it is highly efficient and capable of handling various input parameters.

References

1. Halawa, E., Ghaffarian Hoseini, A., Li, D. H. W. (2014). Empirical correlations as a means for estimating monthly average daily global radiation: a critical overview. *Renewable Energy*, Vol. 72, pp. 149-153.
2. Besharat, F., Dehghan, A. A., Faghih, A. R. (2013). Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, Vol. 21, pp. 798-821.
3. Pinker, R. T., Frouin, R., Li, Z. (1995). A review of satellite methods to derive surface shortwave irradiance. *Remote Sensing of Environment*, Vol. 51, pp. 108-124.
4. Hansen, J. W. (1999). Stochastic daily solar irradiance for biological modeling applications. *Agricultural and Forest Meteorology*, Vol. 94, pp. 53-63.
5. Chen, J.-L., Li, G.-S. (2013). Estimation of monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China. *International Journal of Climatology*, Vol. 33, pp. 487-498.
6. Wu, G., Liu, Y., Wang, T. (2007). Methods and strategy for modeling daily global solar radiation with measured meteorological data – A case study in Nanchang station, China. *Energy Conversion and Management*, Vol. 48, pp. 2447-2452.
7. Mbah, O. M., Mgbemene, C. A., Enibe, S. O., Ozor, P. A., Mbohwa, C. (2018). Comparison of experimental data and isotropic sky models for global solar radiation estimation in Eastern Nigeria. *World Congr. Eng*,, Vol. 2, pp. 4-8.
8. Mbah, O. M., Ozor, P., Mgbemene, C., Enibe, S. O., Mbohwa, C. (2018). Comparative analysis of anisotropic sky models and experimental data in estimating solar radiation on tilted surface in Sub-Saharan African climate. *IEOM Conference. IEOM 2018*, 2018.
9. Chabane, F., Arif, A., Moummi, N., Brima, A. (2020). Prediction of Solar Radiation According to Aerosol Optical Depth. *Iranian (Iranica) Journal of Energy & Environment*, Vol. 11, pp. 271-276.
10. Herath, H., Ariyathunge, S., Karunasena, G. (2021). Development of a Mathematical Model to Forecast Solar Radiation and Validating Results Using Machine Learning Technique. *European PMC, Research Square*, https://doi.org/ 10.21203/rs.3.rs-669429/v1
11. Shourehdeli, S. A., Mobini, K., Asakereh, A. (2022). Modeling of Isentropic Coefficients Used in One Dimensional Model to Predict Ejector Performance at Critical Mode. *Iranian (Iranica) Journal of Energy & Environment*, Vol. 13, pp. 111-123.
12. Olatomiwa, L., Mekhilef, S., Shamshirband, S., Petković, D. (2015). Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria. *Renewable and Sustainable Energy Reviews*, Vol. 51, pp. 1784-1791.
13. Hacioğlu, R. (2017). Prediction of solar radiation based on machine learning methods. *The Journal of Cognitive Systems*, Vol. 2, pp. 16-20.
14. Guermoui, M., Rabehi, A., Gairaa K., Benkaciali, S. (2018). Support vector regression methodology for estimating global solar radiation in Algeria. *The European Physical Journal Plus*, Vol. 133, pp. 1-9.
15. Chen, J.-L., Li, G.-S. (2014). Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables. *Theoretical and Applied Climatology*, Vol. 115, pp. 627-638.
16. Chen, J.-L., Li, G.-S., Wu, S.-J. (2013). Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. *Energy Conversion and Management*, Vol. 75, pp. 311-318, 2013.
17. Chen, J.-L., Liu, H.-B., Wu, W., Xie, D.-T. (2011). Estimation of monthly solar radiation from measured temperatures using support vector machines – A case study. *Renewable Energy*, Vol. 36, pp. 413-420.
18. Benmouiza, K., Cheknane, A. (2013). Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Conversion and Management*, Vol. 75, pp. 561-569.
19. Motameni, H. (2020). Determining the composition functions of Persian non-standard sentences in terminology using a deep learning fuzzy neural network model. *International Journal of Engineering*, Vol. 33, pp. 2471-2481.

20. Mahdavi Jafari, M., Khayati, G. R., Hosseini, M., Danesh-Manesh, H. (2017). Modeling and optimization of roll-bonding parameters for bond strength of Ti/Cu/Ti clad composites by artificial neural networks and genetic algorithm. *International Journal of Engineering*, Vol. 30, pp. 1885-1893.
21. Rahimikhoob, A. (2010). Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. *Renewable Energy*, Vol. 35, pp. 2131-2135.
22. Marzo, A., Trigo-Gonzalez, M., Alonso-Montesinos, J., Martínez-Durbán, M., López, G., Ferrada, P., Fuentealba, E., Cortés, M., Batlles, F. J. (2017). Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. *Renewable Energy*, Vol. 113, pp. 303-311.
23. Torabi, M., Mosavi, A., Ozturk, P., Varkonyi-Koczy, A., Istvan, V. (2018). A hybrid machine learning approach for daily prediction of solar radiation. *International Conference on Global Research and Education*, 2018.
24. Gala, Y., Fernández, Á., Díaz, J., Dorronsoro, J. R. (2016). Hybrid machine learning forecasting of solar radiation values. *Neurocomputing*, Vol. 176, pp. 48-59.
25. Achour, L., Bouharkat, M., Assas, O., Behar, O. (2017). Hybrid model for estimating monthly global solar radiation for the Southern of Algeria : (Case study: Tamanrasset, Algeria). *Energy*, Vol. 135, pp. 526-539.
26. Quej, V. H., Almorox, J., Arnaldo, J. A., Saito, L. (2017). ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *Journal of Atmospheric and Solar-Terrestrial Physics*, Vol. 155, pp. 62-70.
27. Ağbulut, Ü., Gürel, A. E., Ergün, A., Ceylan, İ. (2020). Performance assessment of a V-Trough photovoltaic system and prediction of power output with different machine learning algorithms. *Journal of Cleaner Production*, Vol. 268, 122269.
28. Ağbulut, Ü., Gürel, A. E., Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews*, Vol. 135, 110114.
29. Rabehi, A., Guermoui, M., Lalmi, D. (2020). Hybrid models for global solar radiation prediction: A case study. *International Journal of Ambient Energy*, Vol. 41, pp. 31-40.
30. Mbah, O. M., Madueke, C. I., Umunakwe, R., Okofor, C. O. (2022). Machine learning approach for solar irradiance estimation on tilted surfaces in comparison with sky models prediction. *Journal of Engineering Sciences*, Vol. 9(2), G1-G6, https://doi.org/10.21272/jes.2022.9(2).e1
31. Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L., Cui, N. (2019). Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Conversion and Management*, Vol. 198, 111780.