

## Hybrid Ensemble-Learning Approach for Renewable Energy Resources Evaluation in Algeria

El-Sayed M. El-Kenawy<sup>1,2</sup>, Abdelhameed Ibrahim<sup>3</sup>, Nadjem Bailek<sup>4,\*</sup>, Kada Bouchouicha<sup>5</sup>,  
Muhammed A. Hassan<sup>6</sup>, Basharat Jamil<sup>7</sup> and Nadhir Al-Ansari<sup>8</sup>

<sup>1</sup>Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, Egypt

<sup>2</sup>Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 35712, Egypt

<sup>3</sup>Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt

<sup>4</sup>Energies and Materials Research Laboratory, Faculty of Sciences and Technology, University of Tamanghasset, 10034, Tamanghasset, Algeria

<sup>5</sup>URERMS, Centre de Développement des Energies Renouvelables (CDER), 01000, Adrar, Algeria

<sup>6</sup>Mechanical Power Engineering Department, Faculty of Engineering, Cairo University, Giza, 12613, Giza, Egypt

<sup>7</sup>Department Computer Sciences, Universidad Rey Juan Carlos, Móstoles, 28933, Madrid, Spain

<sup>8</sup>Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 97187, Lulea, Sweden

\*Corresponding Author: Nadjem Bailek. Email: prbailek@gmail.com

Received: 01 September 2021; Accepted: 09 October 2021

**Abstract:** In order to achieve a highly accurate estimation of solar energy resource potential, a novel hybrid ensemble-learning approach, hybridizing Advanced Squirrel-Search Optimization Algorithm (ASSOA) and support vector regression, is utilized to estimate the hourly tilted solar irradiation for selected arid regions in Algeria. Long-term measured meteorological data, including mean-air temperature, relative humidity, wind speed, alongside global horizontal irradiation and extra-terrestrial horizontal irradiance, were obtained for the two cities of Tamanrasset-and-Adrar for two years. Five computational algorithms were considered and analyzed for the suitability of estimation. Further two new algorithms, namely Average Ensemble and Ensemble using support vector regression were developed using the hybridization approach. The accuracy of the developed models was analyzed in terms of five statistical error metrics, as well as the Wilcoxon rank-sum and ANOVA test. Among the previously selected algorithms, K Neighbors Regressor and support vector regression exhibited good performances. However, the newly proposed ensemble algorithms exhibited even better performance. The proposed model showed relative root mean square errors lower than 1.448% and correlation coefficients higher than 0.999. This was further verified by benchmarking the new ensemble against several popular swarm intelligence algorithms. It is concluded that the proposed algorithms are far superior to the commonly adopted ones.

**Keywords:** Renewable energy resources; hybrid modeling; tilted solar irradiation; arid region



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1 Introduction

The performance of photovoltaic or thermal energy conversion systems depends on the orientation and the inclination angle of their collection fields compared to the horizon. Usually (for technical and economic considerations), these systems are installed in a fixed tilted position according to the considered site location for maximum solar energy collection. On the other hand, the inclination angle of photovoltaic arrays or solar thermal collectors could also be adjusted few times a year according to optimum tilt angles defined for specific periods or seasons [1]. The energy gain of these systems is dependent on the availability of solar irradiation on the inclined surface. As a result, determining the generated power requires knowledge of solar irradiance on a photovoltaic (PV) panel in the plane-of-Array (POA). Even though global radiation on a horizontal plane is measured at many meteorological stations around the world, direct measurements of solar radiation on inclined surfaces are very scarce [2]. Over the years, various mathematical models have been developed to calculate solar radiation on tilted surfaces from measured data on the horizontal surface. Evaluation of some of the most widely used such models can be found in [3]. Demain et al. [4] compared 14 models for transposing horizontal radiation to inclined surfaces. They utilized the data for eight months (April to November 2011) from the Royal Meteorological Institute of Belgium in Uccle. The performance of the models was validated against the measured solar radiation data on an inclined surface held at 50.79°. It was reported that the discrepancies appearing in the models' outputs could be attributed to intermediate sky conditions, which were found to be very small in case of clear or overcast situations. It was further reported that the isotropic sky consideration yields better results against the anisotropic sky assumption. Notton et al. [5] used ANNs to estimate the hourly solar radiation on the inclined surface from the global horizontal solar radiation using three models for the Mediterranean site of Ajaccio, France. They used five years of measured data (2006–2010). Optimization of the ANN structure was performed for the number of layers, the number of neurons per layer, and the input data size. They also performed a comparison of empirical models with the newly developed ANN model. It was finally concluded that the ANN method produces more reliable results than the conventional models and that the modeling and optimization of PV equipment can rely on the performance estimation of ANN techniques. Shaddel et al. [6] analyzed the performance of the ANN technique to estimate global solar irradiance on tilted surfaces for application in PV and photothermal applications in the city of Mashhad (Iran). Several input parameters were considered, such as the global solar irradiation, extraterrestrial horizontal irradiation, number of days, collector angle, solar altitude angle, and latitude of the location. The ANN was analyzed and optimized for performance for various tilts 0°, 45° and 60° at an interval during the day length of 6:00 to 17:00 h for the year 2013. The performance of the ANN was verified against the ground measured data for inclined surfaces. It was reported that ANN could effectively be used as a reliable tool for estimating solar radiation on inclined surfaces. Takilalte et al. [7] proposed a new methodology to estimate global irradiation on inclined surfaces at five-minute intervals from the data of global horizontal irradiation. They combined the traditional Liu and Jordan model with Perrin Brichambau with an ANN technique to optimize the model and introduced sky-related conditions to transform the isotropic model into an anisotropic model. They compared the newly developed model with the similar form available in the literature previously. It was reported that the new model, tested under complex sky conditions, has smaller errors and better estimation accuracy due to the effective combination of the ANN model with the conventional ones. Dahmani et al. [8] transposed the 5-min interval measured solar radiation data (obtained for two years) for an inclined surface using the ANN technique for the region of Bouzareah, Algeria. The optimization of the number and types of input data was accomplished using sensitivity analysis. One hidden layer was deemed to be sufficient for estimating the 5-min interval data on the inclined surface. Notton et al. [9]

presented the neural network approach to estimate solar radiation on a tilted plane from horizontal ones at a 10-min interval using three models. The ANN was based on the Levenberg-Marquardt algorithm (LM). It was developed and optimized using five years of solar data obtained from the “Sciences for Environment” laboratory at the University of Corsica Pascal Paoli with a meteorological station installed at Ajaccio. The accuracy of the optimal configuration was reported to be around 9% in terms of root mean square error (RMSE) and around 5.5% for the relative mean absolute error (RMAE), and it was concluded that these errors are indeed lower in comparison for those obtained from empirical correlations available in the literature and used for the estimation of hourly data.

Hybrid-learning methods have attracted increasing attention amongst researchers. In recent years, various hybrid-learning techniques have been proposed and used to evaluate solar energy resources. Alrashidi et al. [10] proposed a framework by hybridizing Support Vector Regression (SVR) with the Grasshopper Optimization Algorithm (GOA) and the Boruta-based feature selection algorithm (BA), abbreviated as (VR-GOA-BAK), for forecasting global solar radiation values at different locations of Saudi Arabia. It was reported that the proposed algorithm has a lower mean absolute percentage error (MAPE) and outperforms the classical SVR models by 32.15–39.69% for the selected locations of the study. Demircan et al. [11] improved the performance of the empirical Angstrom-Prescott model for solar radiation estimation using the Artificial Bee Colony (ABC) algorithm for the city of Muğla, Turkey. Both annual and semi-annual models were developed, and the performance was enhanced using the computational algorithm. It was also reported that the models solely relying on sunshine duration do not exhibit reliable performances, and therefore, sunset-sunrise hour angle should also be included in the models to enhance the performance. Zhou et al. [12] presented a review of machine learning models based on 232 previously reported research articles. They focused on the type of input parameters included, the kind of feature selection method, and the model development procedure. Seven classes of machine-learning models were defined based on the pre-processing data algorithms, output ensemble methods, and the purpose of the models. It was suggested that the quality control of data used for developing the model should be performed to remove the solar radiation measurement errors and account for instrument failure. Further, it was recommended that novel and combined machine-learning models be the focus of studies in the future. Pang et al. [13] investigated the use of deep learning techniques for solar radiation estimation for the region of Alabama. An ANN model and a recurrent neural network (RNN) were analyzed with different sampling frequencies and moving window algorithms to verify accuracy and efficiency. It was reported that RNN outperforms the ANN algorithm and has a 26% higher accuracy. The performance of RNN was further improved from 0.9% normalized mean bias error (NMBE) to 0.2% using a moving window algorithm. Bellido-Jimenez et al. [14] developed and compared different machine learning models to estimate solar radiation at various locations of southern Spain and the USA. Intrahourly inputs were used to improve the performance of the models. It was suggested that machine learning models outperform the self-calibrated empirical models. It was reported that the multi-layer perceptron (MLP) algorithm outperforms the other algorithms and exhibits the best performance using the new proposed variables for the locations considered with medium aridity values. Further, SVR and random forest (RF) models were suggested to be better for the aridest and most humid sites.

In most aforementioned papers, the climate variables are used to estimate solar radiation on a horizontal plane, or the horizontal solar radiation is used as an input variable to estimate radiation on a tiled surface [15–17]. In this paper, we estimated hourly tilted solar irradiation using climate variables. Moreover, irradiance is estimated by hybrid ensemble-learning approach, hybridizing Advanced Squirrel Search Optimization Algorithm (ASSOA) and Support Vector Regression methods, which to the best of our knowledge, has not been used for solar radiation modeling. In addition, the results

are compared to those of several other popular swarm intelligence algorithms, including the genetic algorithm, particle swarm optimization [18], and Grey Wolf Optimizer [19].

## 2 Study Area and Datasets

Hourly solar radiation on tilted surfaces measured at the ground is processed in this study. Data were collected from two radiometric stations located in the southern region of Algeria. Tab. 1 presents the station list, including the geographical coordinates and the measurement period. According to Köppen's climate classification, Tamanrasset (TAM) and Adrar (ADR) have a hot desert (BWh) climate with distinct wet and dry seasons. A statistical summary of all measured variables is given in Tab. 2. In this table, GHI, Tmed, Hum, and WS respectively represent the global horizontal irradiation, mean air temperature, relative humidity, and wind speed.

**Table 1:** Geographical information and weather characteristics of the study sites

Station ID	Lat. [°N]	Long. [°E]	Elev. [m]	Data periods	Tilt Angle [°]
ADR	22.785	5.522	1378	2002–2006	23
TAM	27.874	−0.293	257	2009–2012	28

**Table 2:** Statistical summary of all measured variables

#Station	Statistic	GHI [Wh/m <sup>2</sup> ]	Tmed [°C]	Hum [%]	WS [m/s]
ADR	Max.	32.40	41.40	100.00	5.50
	Min.	0.00	0.00	7.50	0.00
	Mean	15.04	21.22	60.04	1.69
	SD	6.69	7.71	16.15	0.58
TAM	Max.	32.30	43.30	100.00	6.20
	Min.	0.00	0.00	16.20	0.00
	Mean	15.11	22.13	61.95	1.37
	SD	7.91	8.92	20.76	0.86

## 3 Materials and Methods

### 3.1 The ASSOA Regression Algorithm

The Advanced Squirrel Search Optimization Algorithm (ASSOA) was first proposed in [20] for the classification purpose based on two stages to classify chest X-ray images. In this paper, the ASSOA algorithm is applied for the regression problem for the evaluation of renewable energy resources in the Saharan Algeria region. The ASSOA algorithm considers that the agents/individuals are changing their positions in the search space between three kinds of places named-normal, oak, and hickory trees. The optimal solution (nuts food sources) are represented as oak-and-hickory trees. In a mathematical form, the ASSOA algorithm assumes the searching agents are moving to search for the best solution (hickory tree) and next best solutions (three oak trees) for  $n$  agents ( $FS$ ). The agents' locations,  $FS_{i,j}$  for  $i^{th}$  agents in  $j^{th}$  dimension in which  $i \in 1, 2, 3, \dots, n$  and  $j \in 1, 2, 3, \dots, d$ , and velocities,  $V_{i,j}$  for  $i^{th}$

agents in the  $j^{th}$  dimension, are represented as follows.

$$FS = \begin{bmatrix} FS_{1,1} & FS_{1,2} & FS_{1,3} & \dots & FS_{1,d} \\ FS_{2,1} & FS_{2,2} & FS_{2,3} & \dots & FS_{2,d} \\ FS_{3,1} & FS_{3,2} & FS_{3,3} & \dots & FS_{3,d} \\ \dots & \dots & \dots & \dots & \dots \\ FS_{n,1} & FS_{n,2} & FS_{n,3} & \dots & FS_{n,d} \end{bmatrix} \tag{1}$$

$$V = \begin{bmatrix} V_{1,1} & V_{1,2} & V_{1,3} & \dots & V_{1,d} \\ V_{2,1} & V_{2,2} & V_{2,3} & \dots & V_{2,d} \\ V_{3,1} & V_{3,2} & V_{3,3} & \dots & V_{3,d} \\ \dots & \dots & \dots & \dots & \dots \\ V_{n,1} & V_{n,2} & V_{n,3} & \dots & V_{n,d} \end{bmatrix} \tag{2}$$

The initial locations of  $FS_{ij}$  are assumed to have uniform distribution within the bounds. The objective function values,  $f = f_1, f_2, f_3, \dots, f_n$ , can be calculated for each agent as

$$f = \begin{bmatrix} f_1(FS_{1,1}, FS_{1,2}, FS_{1,3}, \dots, FS_{1,d}) \\ f_2(FS_{2,1}, FS_{2,2}, FS_{2,3}, \dots, FS_{2,d}) \\ f_3(FS_{3,1}, FS_{3,2}, FS_{3,3}, \dots, FS_{3,d}) \\ \dots \\ f_n(FS_{n,1}, FS_{n,2}, FS_{n,3}, \dots, FS_{n,d}) \end{bmatrix} \tag{3}$$

The best solution value of the objective function indicates a hickory tree. The calculated values are sorted in ascending order. The first best solution is  $FS_{ht}$  (agent is on the hickory tree), whereas the following three best solutions are indicated as  $FS_{at}$  (agents are on acorn trees). Other solutions are indicated as  $FS_{nt}$  (agents are on normal trees). The locations are updated in each iteration during the algorithm as in the following different cases for a random value  $p \geq 0.5$ .

**Case 1:** Location of  $FS_{at}$  and moving to hickory tree:

$$FS_{at}^{t+1} = \begin{cases} FS_{at}^t + d_g \times G_c(FS_{ht}^t - FS_{at}^t) & \text{if } R_1 \geq P_{dp} \\ \text{Random location} & \text{otherwise} \end{cases} \tag{4}$$

**Case 2:** Location of  $FS_{nt}$  and moving to acorn trees:

$$FS_{nt}^{t+1} = \begin{cases} FS_{nt}^t + d_g \times G_c(FS_{at}^t - FS_{nt}^t) & \text{if } R_2 \geq P_{dp} \\ \text{Random location} & \text{otherwise} \end{cases} \tag{5}$$

**Case 3:** Location of  $FS_{nt}$  and moving to hickory tree:

$$FS_{nt}^{t+1} = \begin{cases} FS_{nt}^t + d_g \times G_c(FS_{ht}^t - FS_{nt}^t) & \text{if } R_3 \geq P_{dp} \\ \text{Random location} & \text{otherwise} \end{cases} \tag{6}$$

where  $R_1, R_2$ , and  $R_3$  are random numbers  $\in [0, 1]$ . The  $d_g$  is a random distance parameter and  $t$  represents the current iteration.  $G_c$  is a constant, equal to 1.9, and it is used to achieve the exploitation and exploration balance. The  $P_{dp}$  probability value is equal to 0.1 for the cases (Case 1, Case 2, and Case 3). For a random value  $p < 0.5$ . The following cases, Case 4, Case 5, and Case 6, will be applied.

**Case 4:** Location of  $FS_{nt}$  and moving diagonally:

$$FS_{nt}^{t+1} = \begin{cases} FS_{nt}^t + V_{nt}^t + c_1 r (FS_{ht}^t - FS_{nt}^t) + c_2 r (FS_{at}^t - FS_{nt}^t) & \text{if } P_a < a \\ \text{Random } FS_{rand}^t \in FS_{nt}^t & \text{otherwise} \end{cases} \quad (7)$$

where  $c_1$ ,  $c_2$ ,  $r$ ,  $P_a$ , and  $a$  are random numbers  $\in [0, 1]$ . For a random agent  $FS_{rand}^t$  from normal agents  $FS_{nt}^t$ , the objective functions as  $F_n(FS_{rand}^t)$  and  $F_n(FS_{nt}^t)$  are calculated to determine between horizontal or vertical movement. If  $F_n(FS_{rand}^t) < F_n(FS_{nt}^t)$ , the movement will be in the vertical direction. Otherwise, the movement will be in the horizontal direction, as in the following case.

**Case 5:** Location of  $FS_{nt}$  and moving horizontally or vertically based on the  $F_n(FS_{rand}^t)$  value:

$$FS_{nt}^{t+1} = \begin{cases} FS_{nt}^t + V_{nt}^t + c_3 r (FS_{rand}^t - FS_{nt}^t) & \text{if } F_n(FS_{rand}^t) < F_n(FS_{nt}^t) \\ FS_{nt}^t + V_{nt}^t + c_1 r (FS_{ht}^t - FS_{nt}^t) & \text{otherwise} \end{cases} \quad (8)$$

where  $c_3$  is a random number  $\in [0, 1]$ . If the horizontal and vertical movement condition is not achieved, the following case will be applied.

**Case 6:** Location of  $FS_{nt}$  and moving will be exponentially:

$$FS_{nt}^{t+1} = FS_{nt}^t + |(FS_{rand}^t - FS_{nt}^t)| \exp(bt) \cos(2\pi t) \quad (9)$$

where  $b$  is a random number  $\in [0, 1]$ .

The constant of seasonal,  $S_c$ , and the seasonal constant minimal value,  $S_{min}$ , are calculated as in the following equations to check the monitoring condition ( $S_c' < S_{min}$ ) for  $t$  as the current iteration and  $t_m$  indicates the maximum value of iterations.

$$S_c' = \sqrt{\sum_{k=1}^d (FS_{at,k}^t - FS_{ht,k}^t)^2}, t = 1, 2, 3. \quad (10)$$

$$S_{min} = \frac{10E^{-6}}{(365)^{t/(t_m/2.5)}} \quad (11)$$

The value of  $S_{min}$  can affect the exploration and exploitation capabilities of the ASSOA algorithm during iterations. If a specific condition is met, such an agent's relocation is calculated by Eq. (12).

$$FS_{nt}^{new} = FS_{ht}^t + 2r \left( (FS_{ht}^t - FS_{nt}^t) \left( 1 - \left( \frac{FS_{ht}^t + FS_{nt}^t}{FS_{nt}^t} \right)^2 \right) \right) \quad (12)$$

The ASSOA algorithm is explained step-by-step in Algorithm 1. Steps 1 and 2 initialize the algorithm parameters. The objective function is then calculated for each agent to sort them and get the first, second, and third best agents and normal agents in Steps 3 and 4. From Step 5 to Step 44, the algorithm is working to update the agents' positions. Steps from 45 to 50 calculate and update positions based on the seasonal constant to avoid local minima. The best solution is obtained by the end of the algorithm at Step 53.

---

**Algorithm 1:** The ASSOA algorithm [20].

---

1. **Initialize** ASSOA population positions  $FS_i (i = 1, 2, \dots, n)$  with size  $n$  using Eq. (1), velocities  $V_i (i = 1, 2, \dots, n)$  using Eq. (2), maximum iterations  $t_m$ , and objective function  $f_n$ .
  2. **Initialize** ASSOA parameters  $R_1, R_2, R_3, n_1, n_2, n_3, P_{dp}, G_c, c_1, c_2, c_3, r, b, P_a, P_d, a, d, p, t = 1$
  3. **Calculate** objective function  $f_n$  for each  $FS_i$  using Eq. (3) and **Sort** agents locations in ascending order
  4. **Find** first best agent  $FS_{ht}$ , next three best agents  $FS_{at}$ , and normal agents  $FS_{nt}$
  5. **while**  $t \leq t_m$  **do**
  6.     **if**  $(p \geq 0.5)$  **then**
  7.         **for**  $(t = 1:n_1)$  **do**
  8.             **if**  $(R_1 \geq P_{dp})$  **then**
  9.                  $FS_{at}^{t+1} = FS_{at}^t + d_g \times G_c (FS_{ht}^t - FS_{at}^t)$
  10.             **else**
  11.                  $FS_{at}^{t+1} = \text{Random location}$
  12.             **end if**
  13.         **end for**
  14.         **for**  $(t = 1:n_2)$  **do**
  15.             **if**  $(R_2 \geq P_{dp})$  **then**
  16.                  $FS_{nt}^{t+1} = FS_{nt}^t + d_g \times G_c (FS_{at}^t - FS_{nt}^t)$
  17.             **else**
  18.                  $FS_{nt}^{t+1} = \text{Random location}$
  19.             **end if**
  20.         **end for**
  21.         **for**  $(t = 1:n_3)$  **do**
  22.             **if**  $(R_3 \geq P_{dp})$  **then**
  23.                  $FS_{nt}^{t+1} = FS_{nt}^t + d_g \times G_c (FS_{ht}^t - FS_{nt}^t)$
  24.             **else**
  25.                  $FS_{nt}^{t+1} = \text{Random location}$
  26.             **end if**
  27.         **end for**
  28.     **else**
  29.         **if**  $(P_a < a)$
  30.              $FS_{nt}^{t+1} = FS_{nt}^t + V_{nt}^t + c_1 r (FS_{ht}^t - FS_{nt}^t) + c_2 r (FS_{at}^t - FS_{nt}^t)$
  31.         **else**
  32.             **Choose** random agent  $FS_{rand}^t$  from normal agents  $FS_{nt}^t$
  33.             **if**  $(P_d < d)$  **then**
  34.                 **Calculate** fitness function  $F_n (FS_{rand}^t)$  for  $FS_{rand}^t$
  35.                 **if**  $(F_n (FS_{rand}^t) < F_n (FS_{nt}^t))$  **then**
  36.                      $FS_{nt}^{t+1} = FS_{nt}^t + V_{nt}^t + c_3 r (FS_{rand}^t - FS_{nt}^t)$
  37.                 **else**
  38.                      $FS_{nt}^{t+1} = FS_{nt}^t + V_{nt}^t + c_1 r (FS_{ht}^t - FS_{nt}^t)$
  39.                 **end if**
  40.         **else**
- 

(Continued)

**Algorithm 1:** Continued

---

```

41.            $FS_{nt}^{t+1} = FS_{nt}^t + |(FS_{rand}^t - FS_{nt}^t)|\exp(bt)\cos(2\pi t)$ 
42.           end if
43.       end if
44.   end if
45.   Calculate seasonal constant ( $S_c^t$ ) using Eq. (10)
46.   Calculate the minimum value of seasonal constant ( $S_{min}$ ) using Eq. (11)
47.   if ( $S_c^t < S_{min}$ ) then
48.        $FS_{nt}^{new} = FS_{ht}^t + 2r \left( (FS_{ht}^t - FS_{nt}^t) \left( 1 - \left( \frac{FS_{ht}^t + FS_{nt}^t}{FS_{nt}^t} \right)^2 \right) \right)$ 
49.   end if
50.   Update  $S_{min}$  using Eq. (11)
51.   Set  $t = t + 1$ .
52. end while
53. Return optimal solution  $FS_{ht}$ 

```

---

### 3.2 Performance Metrics

The performance metrics used for the classification measurements are RRMSE, MBE, RMSE, MAE, and R. The mean absolute error (MAE) determines the average of absolute errors. The mean bias error (MBE) indicates whether the tested model is under-or over-predicting the actual measurements. The correlation coefficient (R) indicates the strength of the correlation between actual and estimated values. The root mean square error (RMSE) and the relative RMSE (RRMSE) provide estimates of the absolute and relative random error components [21–23].

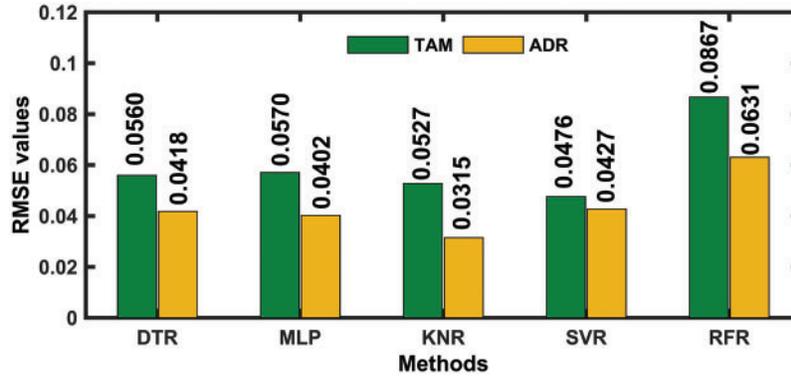
## 4 Experimental Results

### 4.1 Ensemble Selection Scenario

The experiments are divided into two parts. First, the effectiveness of the five models is evaluated for technique selection based on the long-term measured meteo-solar datasets including mean air temperature, relative humidity, wind speed, and global horizontal irradiation, alongside extra-terrestrial horizontal irradiance, for the two Algerian stations. Second, an ensemble-learning method is selected to be used for the estimation of solar radiation. Each investigated dataset is divided randomly into three subsets for training, validation, and testing (60%, 20%, and 20%, respectively). The training set is used to train a k-nearest neighbors (KNN) classifier during the learning phase, the validation set is used when calculating the fitness function for a specific solution, and the testing set is used to evaluate the efficiency of the used techniques. For the KNN classifier, the number of k-neighbors is 5, and the k-fold cross-validation value is set to 10. The selected methods are Decision Tree Regressor (DTR), MLP Regressor (MLP), K-Neighbors Regressor (KNR), Support Vector Regression (SVR), and Random Forest Regressor (RFR). In addition, Average Ensemble (AVE) and Ensemble using SVR (SVE) are two new proposed ensembles for developing solar radiation estimators.

The evaluation of the accuracy of the five selected methods is provided based on the testing dataset, i.e., to determine their capability of handling input data. Fig. 1 shows a comparison between the RMSE values for the different models in each station based on the testing data. It is observed that

the KNR approach presents the lowest RMSE values, followed by the MLP approach at ADR and the SVR approach at TAM.



**Figure 1:** RMSE for five selected methods and each station

Tab. 3 presents the main error metrics of the five selected machine-learning techniques used to select the new ensemble-learning approach. In general, MAE ranges between 0.0260 and 0.0867 for the two sites. The obtained RMSE values also imply a good performance, with high positive correlations, as expected. Although each statistical parameter can evaluate the performance of the model from a specific point of view, it is very difficult for a single statistical parameter to comprehensively evaluate the performance of models. We, therefore, propose to use a supplement statistical score, namely the Global Performance Index (GPI), which has been extensively employed in recent years because of its comprehensiveness and wide applicability. GPI considers different statistical parameters to evaluate the performance of models [24,25] The GPI calculation method is given by Eq. (13).

$$GPI_i = \sum_j \alpha_j (\bar{y}_j - \bar{y}_{ij}) \tag{13}$$

**Table 3:** Results of single classification for the TAM and ADR station

Model	TAM					ADR						
	MAE	RMSE	RRMSE %	R	GPI	Rank	MAE	RMSE	RRMSE %	R	GPI	Rank
DTR	0.0401	0.056	11.410	0.973	-0.286	6	0.0301	0.0418	8.272	0.980	-0.620	6
MLP	0.0468	0.057	11.632	0.980	-0.326	7	0.0336	0.0402	7.957	0.987	-0.295	7
KNR	0.040	0.053	10.756	0.982	0.000	4	0.0240	0.031	6.231	0.996	0.945	3
SVR	0.039	0.048	9.715	0.984	0.213	3	0.0370	0.043	8.465	0.991	-0.383	4
RFR	0.0748	0.0867	17.679	0.938	-2.617	5	0.0501	0.0631	12.487	0.978	-2.591	5
AVE	0.0343	0.0414	8.439	0.989	0.606	2	0.0238	0.0297	5.882	0.995	1.003	2
SVE	0.0224	0.0278	5.650	0.994	1.383	1	0.0227	0.0257	5.080	0.998	1.409	1

Here,  $\alpha_j$  represents the weight of each error metric ( $j$ ).  $\alpha_j$  is taken as  $-1.0$  for  $R$  and  $1.0$  for other error metrics. On the other hand,  $\bar{y}_j$  represents the median of a scaled error metric ( $j$ ), and  $\bar{y}_{ij}$  represent the  $j^{\text{th}}$  scaled error metric's value for the model ( $i$ ). It can be seen from Tab. 3 that the highest GPI values (1.383 and 1.409) at the two sites correspond to the SVE ensemble-learning models.

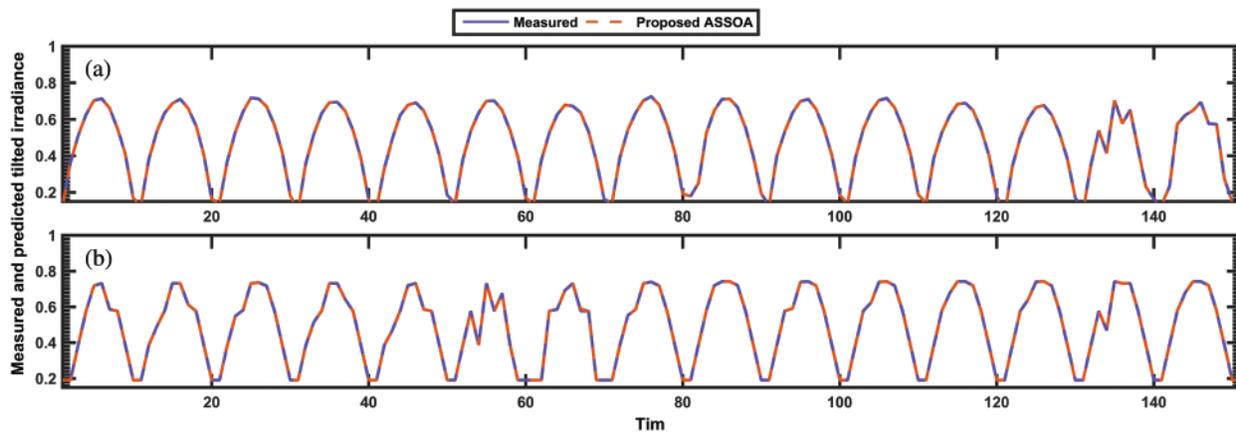
#### 4.2 Proposed Hybrid Model Results

The ability of the proposed SVR ensemble-based ASSOA algorithm for improving the accuracy is investigated vs. other algorithms. The parameters of  $\alpha$  and  $\beta$  in the fitness function are set as 0.99 and 0.01, respectively. The detailed structural information of the proposed model is listed in [Tab. 4](#).

**Table 4:** The ASSOA algorithm configuration

Parameter	Value	Parameter	Value
# Iterations	50	$C_1, C_2, C_3$	[0, 1]
# Agents	10	$P_a$	[0, 1]
$R_1, R_2, R_3$	[0, 1]	$P_d$	[0, 1]
$G_c$	1.9	a, r, b, p	[0, 1]
$P_{dp}$	0.1		

[Fig. 2](#) presents a graph of the measured and predicted hourly tilted irradiance for the combined ASSOA approach. The figure shows a high degree of agreement between actual values and the estimates, which suggests that the proposed model in the study is more reliable than the available approaches. ANOVA statistical tests are performed to ensure the quality of the proposed ASSOA algorithm. The results are shown in [Tabs. 5](#) and [6](#). On the other hand, [Fig. 3](#) depicts the distributions of estimated residuals at the two locations.



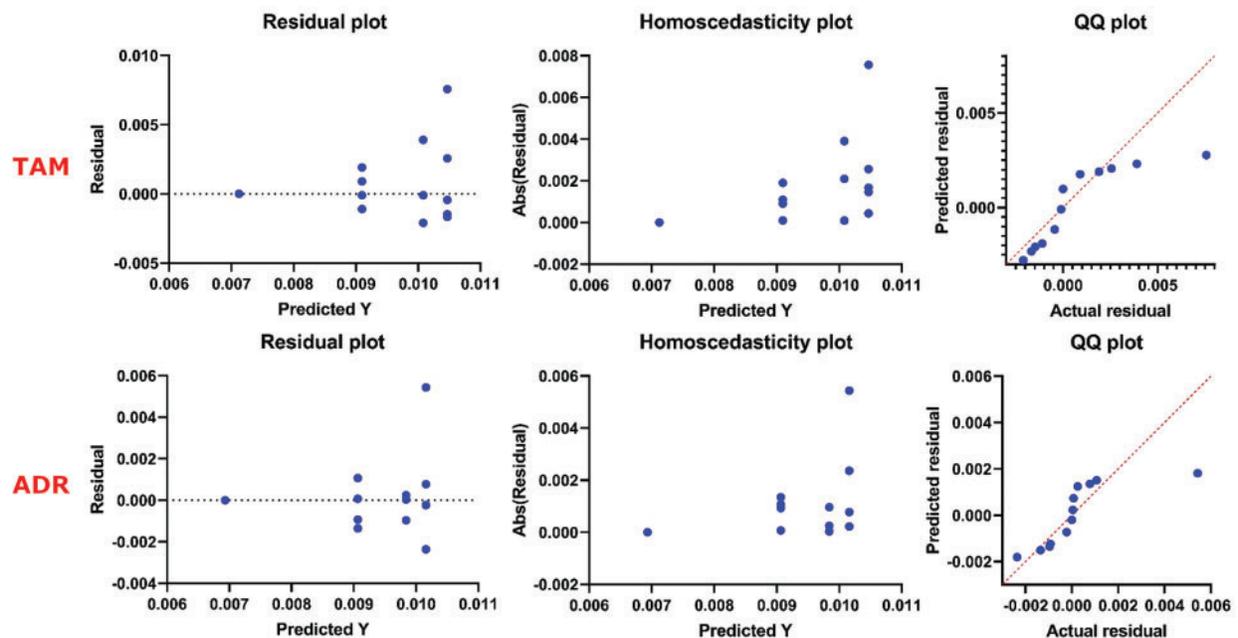
**Figure 2:** Optimized weights predicted values by the proposed model during 15 consecutive days at (a) ADR, and (b) TAM stations

**Table 5:** Results of ANOVA test for the optimization ensemble based on the TAM station data

	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.000134	3	4.48E-05	F (3, 76) = 34.95	P < 0.0001
Residual (within columns)	9.74E-05	76	1.28E-06	-	-
Total	0.000232	79	-	-	-

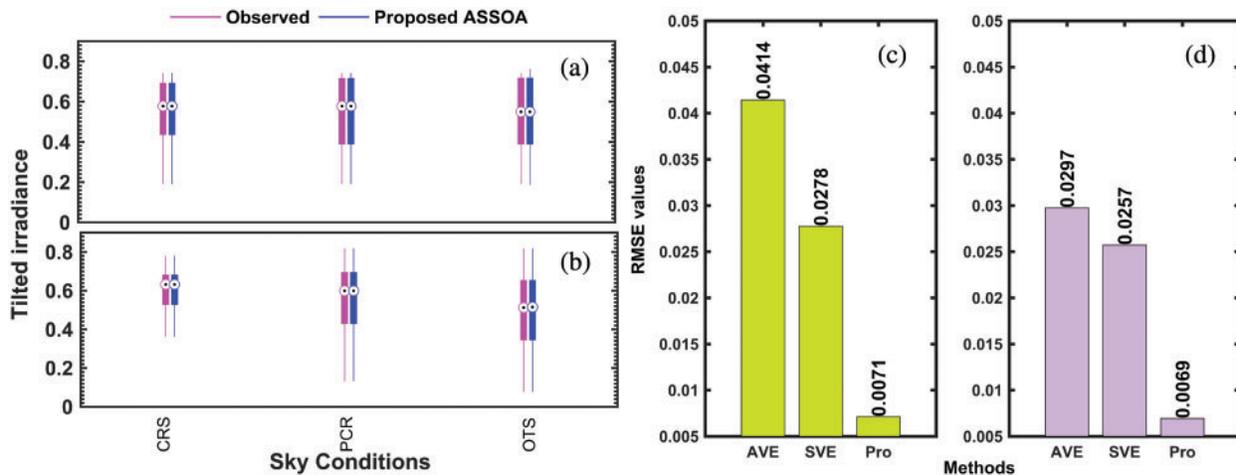
**Table 6:** Results of ANOVA test for the optimization ensemble based on the ADR station data

	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.000126	3	4.21E-05	F (3, 76) = 77.12	P < 0.0001
Residual (within columns)	4.15E-05	76	5.46E-07	-	-
Total	0.000168	79	-	-	-



**Figure 3:** Residual plots of the proposed model at the two stations

To check the sensitivity of the proposed model to the weather conditions, different sky conditions (cloudiness) were considered. This time, the database was divided into three sub-databases according to the cloudiness index value, namely clear sky (CRS), partly cloudy sky (PCR), and complete overcast sky (OTS). The Box plots given in Fig. 4 represent the corresponding results of testing in each class of sky conditions. Without exception, the proposed model performs better under CRS conditions. The model achieves a quite similar accuracy for the PCR and OTS conditions at ADR.



**Figure 4:** Boxplot for the proposed model test for different sky conditions at (a) ADR and (b) TAM, and comparison between the proposed SVR ensemble-based ASSOA algorithm and two selected ensembles models based on overall RMSE-score at (c) ADR and (d) TAM

### 4.3 Proposed Optimization Ensemble vs. Other Algorithms

This experimental test is used to investigate how the ASSOA optimization algorithm helps to improve the accuracy of solar radiation prediction models. To verify the effectiveness of the proposed prediction model, the following groups of comparative experiments are carried out using previously described ensembles models (AVE and SVE). The verification experiments are carried out based on measurement at ADR and TAM. The corresponding RMSEs are depicted in Fig. 4. It is found that the RMSE values for the proposed model are extremely less than the corresponding values for the two-best ensemble models from the previous investigation.

Next, to verify the superiority of the proposed method, several other popular swarm intelligence algorithms including the genetic algorithm (GA), particle swarm optimization (PSO), and Grey Wolf Optimizer (GWO) are used as the benchmark algorithms. The configurations of the GA, PSO, and GWO algorithms, including the number of iterations (generations), population size, and other parameters, are shown in Tab. 7.

The testing results of the selected swarm intelligence algorithms in this study are presented in Tab. 8 in terms of the performance indicators described in Section 3.2. Among these algorithms, GA's estimates are characterized by the highest uncertainty (roughly the RRMSE ranges between 1.96% and 2.04%) while PSO achieves the smallest uncertainty (RRMSEs range between 1.37% and 1.45%). In

fact, PSO takes the first position in three swarm intelligence algorithms. GWO performs comparable or slightly better than GA in terms of RMSE and REMSE.

**Table 7:** Configurations of compared algorithms

Algorithm	Parameter (s)	Value (s)
GA	#Generations	50
	#Population size	10
	Ratio of mutation	0.1
	Crossover	0.9
	Mechanism of selection	Roulette wheel
PSO	Inertia $W_{\max}$ , $W_{\min}$	[0.9, 0.6]
	$C_1$ , $C_2$	[2, 2]
	#Iterations	50
GWO	#Wolves	10
	a	2 to 0

**Table 8:** Results for each station optimized by different algorithms

Optimizer	TAM					ADR				
	MAE	MBE	RMSE	RRMSE	R	MAE	MBE	RMSE	RRMSE	R
GWO	0.00446	-0.00028	0.00712	1.448	0.999	0.00438	-0.00026	0.00693	1.369	0.999
GA	0.00500	-0.00213	0.01003	2.040	0.999	0.00510	-0.00267	0.00993	1.962	0.997
PSO	0.00510	-0.00029	0.00998	2.031	0.999	0.00499	-0.00030	0.00914	1.806	0.999

The variability of the RMSE indicators of the proposed approach, compared to the benchmark algorithms, is shown in Fig. 5. By integrating the proposed hybrid ensemble-learning approach using ASSOA, the model estimates better the hourly solar radiation on tilted surfaces than using PSO, GWO, or GA algorithms (RMSE inferior to 0.008). The Histograms of RMSEs obtained by employing different swarm intelligence algorithms are shown in Fig. 6. It can be noticed that the proposed algorithm using ASSOA can achieve a maximum frequency of error roughly around 0.007 for the two studied sites, which signifies that the ASSOA outperforms the other algorithms. The proposed ASSOA Algorithm's output results in comparison to other algorithms using Wilcoxon's Rank-Sum Based on Average Error Metric are listed in Tabs. 9 and 10, which show that ASSOA is the optimal solution in solving optimization problems in the solar radiation estimation.

Another evaluation is conducted for further assessment of the proposed model against the state-of-the-art models. From Tab. 11, we can see that the proposed hybridization approach yields plausible scores with respect to several prior contributions. The novel adopted approach hybridizing ASSOA and SVR in this study shows high performance in the estimation of hourly solar radiation on tilted surfaces.

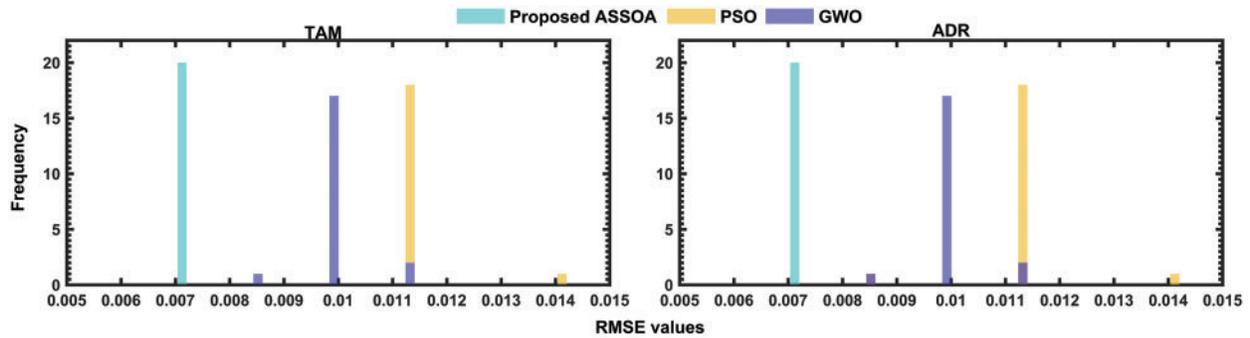


Figure 5: Histogram of RMSE for different algorithms for the TAM and ADR stations

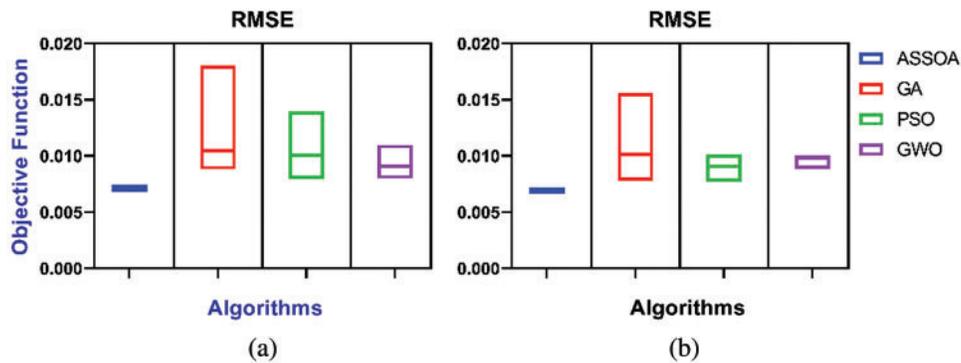


Figure 6: Algorithms, performance vs. the objective function for the TAM and ADR stations. (a) TAM station (b) ADR station

Table 9: Wilcoxon Signed Rank Test results of the optimization ensemble for TAM station

	ASSOA	GA	PSO	GWO
Theoretical median	0	0	0	0
Actual median	0.007122	0.01003	0.009982	0.008998
Number of values	20	20	20	20
Wilcoxon Signed Rank Test				
Sum of signed ranks (W)	210	210	210	210
Sum of positive ranks	210	210	210	210
Sum of negative ranks	0	0	0	0
P-value (two tailed)	<0.0001	<0.0001	<0.0001	<0.0001
Exact or estimate?	Exact	Exact	Exact	Exact
P-value summary	****	****	****	****
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes

(Continued)

**Table 9:** Continued

	ASSOA	GA	PSO	GWO
How big is the discrepancy?				
Discrepancy	0.007122	0.01003	0.009982	0.008998

**Table 10:** Wilcoxon Signed Rank Test results of the optimization ensemble for ADR station

	ASSOA	GA	PSO	GWO
Theoretical median	0	0	0	0
Actual median	0.006935	0.009934	0.009136	0.00988
Number of values	20	20	20	20
Wilcoxon Signed Rank Test				
Sum of signed ranks (W)	210	210	210	210
Sum of positive ranks	210	210	210	210
Sum of negative ranks	0	0	0	0
P-value (two tailed)	<0.0001	<0.0001	<0.0001	<0.0001
Exact or estimate?	Exact	Exact	Exact	Exact
P-value summary	****	****	****	****
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes
How big is the discrepancy?				
Discrepancy	0.006935	0.009934	0.009136	0.00988

**Table 11:** A comparative study between developed models and different techniques in previous studies

Model type	Ref.	Site of study	RMSE	RRMSE %
Present study	-	TAM	0.00712	1.44817
Present study	-	ADR	0.00693	1.36881
Olmo et al.	[26]	Granada	-	27%
Notton et al.	[3]	Ajaccio	-	8.11–10.71%
Padovan et al.	[27]	Padova	39.43	6.4%–8.7%
Notton et al.	[5]	Ajaccio	18.23	5.28%

## 5 Conclusions

In this paper, a novel combined approach using the ASSOA optimization algorithm is proposed. For this purpose, measurements of hourly solar irradiation on inclined surfaces covering 2002–2006 in Tamanrasset and 2009–2012 in Adrar, Algeria, were used. The proposed model accuracy was assessed based on MBE, MAE, RMSE, RRMSE, and R.

Different experiments are performed using the mentioned data; firstly, five models of technique selection were considered (DTR, MLP, KNR, SVR, and RFR) in developing solar radiation estimating models. The estimation is based on the long-term measured meteorological datasets, including relative humidity, mean air temperature, maximum air temperature, minimum air temperature, and daily temperature range. The performance is evaluated and compared to two new proposed ensembles known as Average Ensemble (AVE) and Ensemble using SVR (SVE). The identification of model performance rankings is then conducted based on the Global Performance Index (GPI). The results of the comparative analysis show the superiority of the AVE and the SVE models.

In the second experimentation, the ASSOA optimization algorithm effect on improving the accuracy results, in the estimation of the solar, was investigated. Comparative analyses were carried out using previously described ensembles models (AVE and SVE). It was concluded that the proposed is more successful than the two better ensemble-based models, AVE and SVE.

Thirdly, to check the proposed model's sensitivity to the climate condition, the proposed evaluation at different sky conditions is conducted. The results show that the ASSOA perform better in CRS condition than in PCR end OTS. The model achieves a quite similar accuracy for the PCR end OTS condition in the ADR site.

The proposed ASSOA algorithm was also compared to GA, PSO, and GWO optimizers for feature selection to validate its efficiency. Compared to the other swarm intelligence algorithms, the proposed ASSOA shows superior performance. When the proposed model in this study is compared against the state-of-the-art models, it can be said that the model offers high performance in the estimation of estimate the hourly solar radiation on tilted surfaces classified as arid climate.

The present work demonstrated the potential of estimating solar radiation on tilted surfaces while also considering the influential meteorological parameters. Thus, it will help in decision-making. Combining the proposed optimization algorithm with more accurate estimation models can be further improved in the future. Moreover, regional models can be established. This study can contribute to enhancing solar design and the facilitation of implementing solar technologies for remote areas.

**Funding Statement:** The authors received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] N. Bailek, K. Bouchouicha, N. Aoun, M. EL-Shimy, B. Jamil *et al.*, "Optimized fixed tilt for incident solar energy maximization on flat surfaces located in the algerian big south," *Sustainable Energy Technologies and Assessments*, vol. 28, pp. 96–102, 2018.
- [2] C. K. Pandey and A. K. Katiyar, "Hourly solar radiation on inclined surfaces," *Sustainable Energy Technologies and Assessments*, vol. 6, pp. 86–92, 2014.

- [3] G. Notton, P. Poggi and C. Cristofari, "Predicting hourly solar irradiations on inclined surfaces based on the horizontal measurements: Performances of the association of well-known mathematical models," *Energy Conversion and Management*, vol. 47, no. 13–14, pp. 1816–1829, 2006.
- [4] C. Demain, M. Journée and C. Bertrand, "Evaluation of different models to estimate the global solar radiation on inclined surfaces," *Renewable Energy*, vol. 50, pp. 710–721, 2013.
- [5] G. Notton, C. Paoli, S. Vasileva, M. L. Nivet, J. L. Canaletti *et al.*, "Estimation of hourly global solar irradiation on tilted planes from horizontal one using artificial neural networks," *Energy*, vol. 39, no. 1, pp. 166–179, 2012.
- [6] M. Shaddel, D. S. Javan and P. Baghernia, "Estimation of hourly global solar irradiation on tilted absorbers from horizontal one using artificial neural network for case study of mashhad," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 59–67, 2016.
- [7] A. Takilalte, S. Harrouni, M. R. Yaiche and L. Mora-López, "New approach to estimate 5-min global solar irradiation data on tilted planes from horizontal measurement," *Renewable Energy*, vol. 145, pp. 2477–2488, 2020.
- [8] K. Dahmani, R. Dizene, G. Notton, C. Paoli, C. Voyant *et al.*, "Estimation of 5-min time-step data of tilted solar global irradiation using ANN (Artificial Neural Network) model," *Energy*, vol. 70, pp. 374–381, 2014.
- [9] G. Notton, C. Paoli, L. Ivanova, S. Vasileva and M. L. Nivet, "Neural network approach to estimate 10-min solar global irradiation values on tilted planes," *Renewable Energy*, vol. 50, pp. 576–584, 2013.
- [10] M. Alrashidi, M. Alrashidi and S. Rahman, "Global solar radiation prediction: Application of novel hybrid data-driven model," *Applied Soft Computing*, vol. 112, pp. 107768, 2021.
- [11] C. Demircan, H. C. Bayrakçı and A. Keçebaş, "Machine learning-based improvement of empiric models for an accurate estimating process of global solar radiation," *Sustainable Energy Technologies and Assessments*, vol. 37, pp. 100574, 2020.
- [12] Y. Zhou, Y. Liu, D. Wang, X. Liu and Y. Wang, "A review on global solar radiation prediction with machine learning models in a comprehensive perspective," *Energy Conversion and Management*, vol. 235, pp. 113960, 2021.
- [13] Z. Pang, F. Niu and Z. O'Neill, "Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons," *Renewable Energy*, vol. 156, pp. 279–289, 2020.
- [14] J. A. Bellido-Jiménez, J. Estévez Gualda and A. P. García-Marín, "Assessing new intra-daily temperature-based machine learning models to outperform solar radiation predictions in different conditions," *Applied Energy*, vol. 298, pp. 117211, 2021.
- [15] E. -S. M. El-Kenawy, "Solar radiation machine learning production depend on training neural networks with ant colony optimization algorithms," *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, vol. 7, no. 5, pp. 1–4, 2018.
- [16] J. Almorox, J. A. Arnaldo, N. Bailek and P. Martí, "Adjustment of the Angstrom-PreScott equation from Campbell-Stokes and Kipp-Zonen sunshine measures at different timescales in Spain," *Renewable Energy*, vol. 154, pp. 337–350, 2020.
- [17] K. Bouchouicha, M. A. Hassan, N. Bailek and N. Aoun, "Estimating the global solar irradiation and optimizing the error estimates under algerian desert climate," *Renewable Energy*, vol. 139, pp. 844–858, 2019.
- [18] A. Ibrahim, M. Noshay, H. A. Ali and M. Badawy, "PAPSO: A power-aware VM placement technique based on particle swarm optimization," *IEEE Access*, vol. 8, pp. 81747–81764, 2020.
- [19] E. -S. M. El-Kenawy, M. M. Eid, M. Saber and A. Ibrahim, "MbGWO-SFS: Modified binary grey wolf optimizer based on stochastic fractal search for feature selection," *IEEE Access*, vol. 8, pp. 107635–107649, 2020.
- [20] E. -S. M. El-Kenawy, S. Mirjalili, A. Ibrahim, M. Alrahmawy, M. El-Said *et al.*, "Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient COVID-19 x-ray chest image classification," *IEEE Access*, vol. 9, pp. 36019–36037, 2021.

- [21] K. Bouchouicha, N. Bailek, M. E. -S. Mahmoud, J. A. Alonso, A. Slimani *et al.*, “Estimation of monthly average daily global solar radiation using meteorological-based models in adrar, algeria,” *Applied Solar Energy*, vol. 54, no. 6, pp. 448–455, 2018.
- [22] B. Jamil and A. T. Siddiqui, “Estimation of monthly mean diffuse solar radiation over India: Performance of two variable models under different climatic zones,” *Renewable Energy*, vol. 25, pp. 161–180, 2018.
- [23] F. J. Olmo, J. Vida, I. Foyo, Y. Castro-Diez and L. Alados-Arboledas, “Prediction of global irradiance on inclined surfaces from horizontal global irradiance,” *Energy*, vol. 24, pp. 689–704, 1999.
- [24] N. Bailek, K. Bouchouicha, Z. Al-Mostafa, M. El-Shimy, N. Aoun *et al.*, “A new empirical model for forecasting the diffuse solar radiation over sahara in the Algerian Big south,” *Renewable Energy*, vol. 117, pp. 530–537, 2018.
- [25] N. Bailek, K. Bouchouicha, M. A. Hassan, A. Slimani and B. Jamil, “Implicit regression-based correlations to predict the back temperature of PV modules in the arid region of south algeria,” *Renewable Energy*, vol. 156, pp. 57–67, 2020.
- [26] F. J. Olmo, J. Vida, I. Foyo, Y. Castro-Diez and L. Alados-Arboledas, “Prediction of global irradiance on inclined surfaces from horizontal global irradiance,” *Energy*, vol. 24, pp. 689–704, 1999.
- [27] A. Padovan and D. Del Col, “Measurement and modeling of solar irradiance components on horizontal and tilted planes,” *Solar Energy*, vol. 84, no. 12, pp. 2068–2084, 2010.