AIMS Energy

AIMS Energy, 12(1): 62–83.
DOI: 10.3934/energy.2024004
Received: 30 October 2023
Revised: 06 December 2023
Accepted: 11 December 2023
Published: 27 December 2023

http://www.aimspress.com/journal/energy

Research article

Solar radiation forecasting based on ANN, SVM and a novel hybrid FFA-ANN model: A case study of six cities south of Algeria

Halima Djeldjli[1,*], Djelloul Benatiallah[1], Camel Tanougast[2] and Ali Benatiallah[3]

[1] Material Sciences Department, Faculty of Material Sciences, Mathematics and Computer Science Laboratory of Sustainable Development and Computer Science (LSDCS), University Ahmed Draia, Adrar, 01000, Algeria
[2] Laboratory of Systems Design, Optimization and Modeling (LCOMS), University of Lorraine, Metz 57070, France
[3] Laboratory of Energy Environment and Information System (LEEIS), University Ahmed Draia, Adrar, 01000, Algeria

* **Correspondence:** Email: djeldjlihalima@univ-adrar.edu.dz; Tel: +213696821352.

Abstract: This study was conducted for six cities in southern Algeria, where the accuracy of three models—support vector machines (SVM), artificial neural networks (ANN) and a novel hybrid firefly algorithm-based model (FFA-ANN)—were investigated when estimating global solar irradiation throughout an eleven-year period, utilizing nine input parameters as input data. The goal of our novel suggested a hybrid FFA-ANN model, where we relied on the optimization Firefly algorithm to enhance the ANN model created. Despite the fact that the ANN and SVM models produced promising results, our suggested FFA-ANN hybrid model outperformed the stand-alone ANN-based model using three statistical factors—correlation coefficient, relative root mean squared error and mean absolute percent error—with the best values of (R = 0.9321, rRMSE = 9.35% and MAPE = 6.29%). The findings demonstrated that FFA-ANN was preferable to the optimized SVM and ANN models when forecasting daily global solar irradiation in all zones. Furthermore, after comparing the combinations, the study's findings showed that the ANN model depended on: Extraterrestrial solar irradiation (H_0), declination and average temperature (T_{avg}) together with relative humidity (RH) as inputs in order to estimate daily sun radiation. Thus, the findings of this study suggest that in regions with dry climates and other places with comparable climates, the created model may be used to estimate daily global solar radiation whenever data is accessible.

Keywords: hybrid model; artificial neural network (ANN); firefly algorithm (FFA); solar radiation; optimization; energy systems; support vector machine (SVM)

Abbreviations: DGSR: Global solar irradiation per day (W/m^2); H_0: Extraterrestrial solar radiation(W/m^2); DE: Declination (°); T_{avg}: Average temperature (C°); RH: Relative humidity (%); HA: Hour angle (Degree°); BP: Atmospheric pressure (hPa); WS: Wind speed (m/s)

1. Introduction

Ensuring a precise and dependable forecast of solar radiation is crucial for the best possible layout and operation of thermal and solar photovoltaic systems. For the generation of clean and renewable energy, this is crucial. Because they yield good results with a limited number of accessible parameters as inputs and they have different architectures, artificial neural networks (ANNs) are extensively employed in solar radiation forecasting [1].

Readers can see from the broad overview that ANN has been considered in several kinds of research, such as the work of Gairaa et al. who proposed a novel combination approach to predict the daily global solar irradiation in Algeria by combining the nonlinear ANN model with the linear autoregressive moving average (ARMA) model. The suggested model (ARMA-ANN) guarantees a sufficient level of accuracy. In terms of mean absolute error (MAE), the combined model outperformed the ARMA model by a significant margin (over 18%), whereas it outperformed the ANN model by less than 3% [2]. Also in Algeria, Benatiallah et al. created a multilayer ANN model, to test different scenarios. It has been determined that the logistic Sigmoid function with a hidden layer of 15 neurons might be preferable to estimate the hourly sun irradiation [3].

Multilinear regression techniques (MLR), (ANN) and empirical equations were utilized in Greece to predict solar radiation [4]. While the findings of the ANN model when compared to MLR and the Hargreaves method using the same dataset are consistent between them, the accuracy of the results are increased by using the square root of the daily temperature differential and extraterrestrial radiation when using MLR and ANN. Nonetheless, ANN was considered to be more complex. Amiri et al. proposed a new method based on a multitask hybrid evolutionary neural network. After comparing the potential of three distinct basis functions in the hidden layer, it was determined that Sigmoidal units produced superior outcomes. Compared to single-task models, the suggested multitask alternative was simpler, far easier to use, more computationally efficient and may even perform slightly better [5].

Using ANNs, [6] estimated the monthly average of the daily global solar radiation across Italy using data collected from 45 different locations and 13 different input parameters. The most significant input for the correct monthly average daily global solar irradiation prediction was identified using the automatic relevance determination technique (ARD). The optimum ANN design only comprises a few parameters: Top of the Atmosphere radiation, day length, number of rainy days, average rainfall, latitude and altitude. The model exhibits excellent handling efficiency, geographic variety and accuracy. Model complexity and computational resource requirements are some of the factors to be taken into account. Also, in the work of [7], multiple ANN models were presented to estimate daily global solar radiation (GSR) on a horizontal area using meteorological data, with the findings revealing that relative humidity is an extremely important parameter determining prediction performance.

In addition, [8] provided an application of ANNs to anticipate daily solar radiation, analyze the impact of external meteorological data using a multivariate approach as a time series for the optimized

Multilayer Perceptron (MLP) and compare it with different prediction methods. The utilization of exogenous data is captivating in winter; however, endogenous data as inputs on a preprocessed ANN appears sufficient in summer. In Algeria's south western region, an ANN trained with the backpropagation algorithm was used by [9] to estimate global solar radiation based on air temperature and relative humidity data. The outcome demonstrates a very good accuracy. For various cities in India, the accuracy of three models involving different combinations of input parameters is evaluated by [10]. The work focused on forecasting monthly mean daily solar irradiation using the most influential inputs (month, maximum temperature, latitude and bright sunshine hours) determined by Waikato Environment for Knowledge Analysis (WEKA) software and relative humidity as the least affecting input ANN model is outperformed by the support vector machine (SVM) model, which uses the most influential inputs.

The SVM model has been applied in various applications, including solar radiation prediction and classification. In fact, the SVM model is successful in solving nonlinear regression problems, and being able to minimize prediction error and maximize prediction generalization makes it one of the best machine learning techniques available [1]. To anticipate monthly mean daily horizontal solar irradiation, [11] created two hybrid methods that combined SVM with the firefly algorithm (SVM-FFA) and the wavelet transform algorithm (SVM-WT). The findings suggest that the two hybrid techniques outperform the single SVM in terms of performance. According to the results, the SVM-WT appears to be more effective than the SVM-FFA approach for the particular case study. Additionally, using only sunshine ratio for input, [12] demonstrated the potential of creating an easy-to-use model based on support vector regression (SVM-R) that could be utilized for estimating daily global solar irradiation (DGSR) on the horizontal surface in Algeria. The results demonstrate the SVM-R's strong qualification for DGSR estimation utilizing only the sunshine ratio. On the other hand, in Spain, to evaluate the performance of three types of neural computation approaches in a problem of solar radiation prediction, the structure sigmoid unit-product unit with evolutionary training has been shown to be the best model, as well as alternative machine learning [13].

Furthermore, in Algeria, [14] presented the application of an SVM for the purpose of predicting global sun radiation on a horizontal surface. For the future forecast (daily or monthly), various combinations of calculated sunshine duration, observed ambient temperature and extraterrestrial solar radiation were taken into consideration. The solar radiation values that were anticipated and observed agreed rather well, according to the findings. The most important advantage is the fact that just a few basic parameters are needed for the suggested SVM models to achieve reliable precision. In Iran (Kerman city), by combining the WT algorithm with SVM, a new model was created by [15] for the aim of predicting daily horizontal diffuse solar radiation. Daily observed global and diffuse solar radiation datasets are used to evaluate the validity of the hybrid SVM-WT approach. The cloudiness index is associated with the clearness index to be the only input data for the proposed SVM-WT model. The ANN model, an empirical model and a radial basis function SVM-RBF were used to assess the applicability of SVM-WT. The outcomes show that SVM-WT is a more precise and effective method compared to other models.

In another study, [16] utilized SVMs for predicting GSR for Sharurha, southwest of Saudi Arabia. The SVM model was trained using measured relative humidity and temperatures. The GSR values were predicted using four combinations of datasets. The obtained results show that the SVM method is capable of predicting GSR from measured values of temperature and relative humidity. In paper [17], they used fuzzy regression functions (FRF) and the SVM technique to estimate the yearly average global horizontal sun radiation. They carried out an empirical analysis using a dataset gathered in Turkey and implemented the FRF-SVM technique with several kernel functions to show the efficacy of the approach. They found that in an area with complicated meteorological and spatial features, the employment of hybrid SVM and fuzzy function techniques were useful in long-term forecasting. In paper [18], three methods—ANN, SVM and adaptive neuro fuzzy inference system (ANFIS)—were evaluated for their efficacy and reliability in forecasting daily global sun radiation using meteorological data that was observed in Mexico. Root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) were among the statistical metrics used to evaluate the performance of the model. According to the outcomes of the evaluation, the SVM technique outperforms the other methods and shows promise as a replacement for conventional methods for solar radiation forecast. Its input data included rainfall, extraterrestrial solar radiation and the daily lowest and highest temperature.

In addition, according to [19], horizontal GSR is estimated using support vector regression. The effectiveness of the SVM-R model was assessed using various combinations of metrological features. It was discovered that the SVM-R model produced better outcomes than other models, such as ANN. It was determined that temperature is the most crucial factor, followed by wind speed, relative humidity, day number and atmospheric pressure. On the other hand, [20] suggested three novel hybrid SVM models for the prediction of daily diffuse solar radiation Rd in areas with air pollution: SVM-PSO (particle swarm optimization algorithm), SVM-BAT (bat algorithm) and SVM-WOA (whale optimization algorithm). As a result, in comparison to SVM, the suggested hybrid models further enhanced the forecasting accuracy as well as the convergence rate in Rd modeling. The findings demonstrated how crucial it is to take air pollution into account for a more precise estimation of daily Rd in areas where air pollution is in existence. Additionally, to boost the effectiveness of stand-alone machine learning models, heuristic algorithms—such as BAT—are strongly advised. In paper [21], for daily diffuse radiation estimates in China, three types of models were developed: Copula-base nonlinear quantile regression (CNQR), empirical models and the SVM-FFA algorithm. While empirical models fared somewhat better than the comparable CNQR, SVM-FFA surpassed the corresponding models. Consequently, SVM-FFA's total computing costs were much more than CNQR's. Given the trade-off between computing costs and accuracy, CNQR was strongly advised for the Rd estimate.

A combined SVM-FFA approach was employed in Iran by [22] in order to forecast the monthly average horizontal GSR. Models are checked for accuracy, taking different groups of captured climate-related datasets into consideration and using each technique and long-term horizontal GSR observations. Using a variety of datasets, the model performed remarkably well in all scenarios. Furthermore, extraterrestrial solar radiation has been included as a key component for precise GSR estimation. It was discovered that the model offers a definite benefit over empirical models with comparably identical input parameters. A blend machine learning method's accuracy in forecasting sun radiation from a subset of meteorological data has been studied in Nigeria [23]. In order to do that, a novel technique known as SVM-FFA has been developed, which predicts the monthly median horizontal global sun radiation based on three meteorological factors. The created SVM-FFA model outperforms the ANN model in terms of prediction accuracy, according to the collected findings. It is possible to identify the created SVM-FFA model as an effective machine learning method for precise horizontal GSR forecasting. The outcomes demonstrated the model's feasibility using temperature measurements and sunshine duration.

The primary goal of this investigation is to experimentally simulate our proposed hybrid firefly algorithm-based (FFA-ANN) model as a hybrid approach to machine learning for modeling and predicting solar radiation and compare it to two stand-alone models using meteorological data from nine selected study regions that are found in Algeria.

2. Models and materials

2.1. Areas of research and datasets

We have chosen the southwestern part of Algeria as the study area. The provinces of Adrar, Eloued, Ouargla, Tamanrasset, Timimoun and Bechar were selected as study areas by considering their climatic characteristics. This region has a lot of solar radiation potential. The greatest insolation time across this area exceeds 3,900 hours [24]. On horizontal surfaces, the mean solar energy collected is 5 kWh/m^2 within the vast majority of Algeria (Figure 1), or roughly 2,263 kWh/m^2/year in Algeria's south [25].

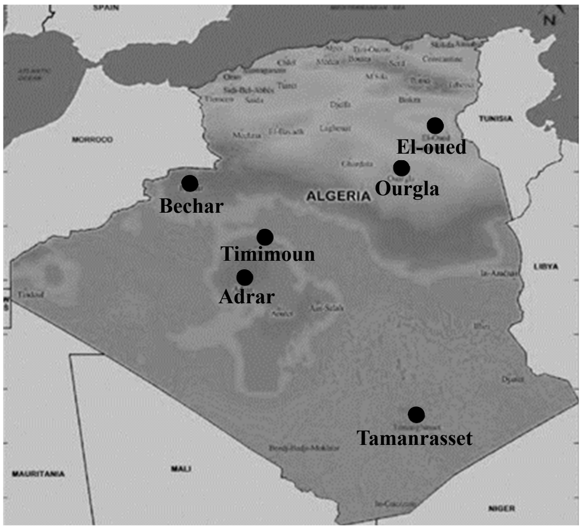

Figure 1. Displays the places of each of the sites under consideration on an Algerian map.

Given in Table 1, the provinces that were chosen and their respective locations.

Table 1. Location's coordinates.

Provinces	Latitude (°)	Longitude (°)	Altitude (m)
Adrar	26.489	−1.358	286
El-oued	33.512	6.783	63
Ouargla	30.998	6.766	178
Tamanrasset	24.335	4.455	810
Timimoun	30.024	0.849	431
Bechar	31.386	−2.012	785

We gather a vast quantity of data, enough to establish a representative dataset. The datasets applied are made up of nine parameters that are represented in Table 2 and are related to the above-mentioned study areas, over an eleven-year period (from the first of January 2010 to the last day of the year 2021). The SODA (Simple Ocean Data Assimilation) database supplied the data [26].

Table 2. The Parameters used.

Parameters	Abbreviation	Unit	Category	Type
The year	Y	/	Meteorological	Numerical
Day of the year	D	/		
Wind speed	WS	m/s		
Atmospheric pressure	BP	hPa		
Average temperature	T_{avg}	°C		
Relative humidity	RH	%		
Declination	DE	Degree (°)	Astronomical	
Hour angle	HA	Degree (°)		
Extraterrestrial solar irradiation	H_0	Wh/m^2		

We used astronomical and meteorological parameters to forecast DGSR. The datasets are separated into two subgroups: One for sample training the model (80%), and the remaining (20%) for testing and validating.

2.2. Methods and models

2.2.1. ANN

An ANN is an abstract computational approach that follows the behavior of the human brain [27]. For each of the i^{th} hidden layers, each neuron calculates weight by the Wij sum of the input signal yi, which is then applied to a nonlinear activation function to generate the output signal uj. This function's structure is:

$$uj = \sum_{i=0}^{n} Wij\, yi \tag{1}$$

The multilayer feed-forward neural network (MLF) using the back-propagation algorithm (BP) has been most frequently employed for predicting solar radiation [28]. This approach is used to represent nonlinearly separable problems. The input layer, a number of hidden layers, as well as a final output layer compose MLF. Weights Wjk with Wij link each layer, where each neuron adds threshold term or a bias to the total before nonlinearity shifts the sum to create an output. The node's activation function is a term given to this nonlinear transition. In MLFs, the hidden and output layers typically utilize logistic sigmoid (Eq 2) and linear functions (Eq 3), respectively [29]:

$$f(w) = \frac{1}{1+e^{-w}} \tag{2}$$

$$f(x) = x \tag{3}$$

The input to the output layer is represented by x, while the weighted sum of the input is denoted by w.

When an error is calculated at the output layer and propagated backward to the input layer, it is referred to as BP [30].

2.2.2. SVM

Vapnik came up with SVM, a supervised learning method that can be used for regression. SVMs fall within the machine learning subfield and they are based on the theory of structural risk reduction, which aims to reduce the learning machine's empirical risk as well as its confidence interval. This helps the machine acquire strong adaptation capabilities.

Statistical learning theory provides SVMs with a strong mathematical foundation. SVM techniques were first created to solve classification difficulties, but they may also be effectively used to solve regression problems (support vector regression). SVM can be used to estimate a regression for the dataset $\{(x, y)\} = N/i = 1$, at which yi is the output result, xi is an input vector and N is the overall number of input data. This is achieved by using the nonlinear function $\varphi(x)$ to convert x into a parameter space. Afterward determining a regression function in the manner is described below:

$$f\ (x) = \omega . \varphi\ (x) + b \tag{4}$$

With its variables, weight vector ω and bias value b, this function can most closely approach, with ε (error tolerance), the real outcome y. The regularized risk function that follows is minimized to determine the coefficients b and ω:

$$R\ (C) = C \sum_i^N L_\varepsilon\ (f\ (x_i), y_i) = \frac{1}{2} \mid\mid \omega \mid\mid^2 \tag{5}$$

The degree of complexity for the model is regulated by the term $\frac{1}{2} \mid\mid \omega \mid\mid^2$, which enhances the SVM's generalization. The amount of empirical inaccuracy in the user-selected optimization problem is determined by C (positive trade-off factor).

The primary distinction between this method and traditional regression is the utilization of an innovative loss function (ε). Vapnik's linear loss function with ε-insensitivity zone is:

$$L_\varepsilon\ (f(x_i), y_i) = 0 \text{ for } |f\ (x_i) - y_i| \leq \varepsilon, \text{otherwise } |f\ (x_i) - y_i| - \varepsilon \tag{6}$$

Therefore, if there is less than ε distinction between the observed and anticipated values, then the loss appears equivalent to zero. The loss error equals zero if the predicted value is within the tube. The gap between the anticipated value alongside the tube's radius ε represents the extent of the loss for all other expected sites outside the tube. The gap widths ξ "above" and ξ* "below" the tube ε in order to prevent outliers. The following reduces the risk:

minimize

$$R\ (\xi, \xi^*, b, \omega) = \frac{1}{2} \mid\mid \omega \mid\mid^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \tag{7}$$

subjected to

$$y_i - \omega\phi\ (x_i) - b_i \leq \varepsilon + \xi_i \tag{8}$$

$$b_i + \omega\phi\ (x_i) - y_i \leq \varepsilon + \xi_i^* \tag{9}$$

$$\xi_i \qquad \xi_i^* \geq 0 \tag{10}$$

Empirical risk severity is managed by $C \sum_{i=1}^N (\xi_i + \xi_i^*)$. α and α*, Lagrange multipliers, have been added within the constraint equations to address the optimization problem. The equation may be expressed using dual form as follows:

$$R\,(\alpha\,,\alpha^*) = \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K\,(x_i, x_j) \tag{11}$$

with limitations: $\sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) = 0\,, 0 \le \alpha_i \le C\;,\; 0 \le \alpha_i^* \le C_i$ =1,2,.....,N, where $K\,(x_i, x_j) = \phi\,(x_i).\phi\,(x_j)$ is referred to as a kernel function. This function gives SVM the capacity to simulate complex separating hyperplanes, enabling them to generate nonlinear boundaries. Following the computation of Lagrange multipliers, the regression hyperplane's ideal desired weights vector is discovered as follows:

$$\omega = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\phi\,(x_i) \tag{12}$$

and (Eq 4) is recast in the following way:

$$f(x, \alpha_i,\, \alpha_i^*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\, K\,(x_i, x_j) + b \tag{13}$$

where b is the computed bias based on training samples and N is the total number of support vectors. In general, the fundamental mathematical role of the SVM statistical process of learning is

$$y = f\,(\,x) = \sum_{i=1}^{M} \alpha_i.\phi\,(x_i) = w\,.\phi\,(x) \tag{14}$$

$\varphi(x)$ performs nonlinear transformation, yielding the linearly weighted sum of M. SVM's decision-making function is

$$y = f\,(x) = (\sum_{i=1}^{N} \alpha_i.K\,(x_i, x_j)) - b \tag{15}$$

where b and α_i are parameters, xi are vectors employed during the training phase. xj indicates an independent vector and N is the total quantity of training data.

Maximizing the objective functions of $\alpha_i$and b yields their respective parameters. Even if the data was not separable in the initial input space, it can be separated in the space of features with the right kernel selection (see Table 3) [31].

Table 3. Different kernel functions.

Kernel function	Equation	
Radial basis function (RB-F)	$K\,(x_i, x_j) = exp\;(-\gamma\, x_i\,.x_j\|\|\,(x_i - x_j\|\|^2 + r)$	(16)
Polynomial K	$K\,(x_i, x_j) = \;(\gamma x_i\,.x_j + r)^d$	(17)
Linear	$K\,(x_i, x_j) = \;x_i\,.x_j$	(18)
Sigmoid	$K\,(x_i, x_j) = \;tanh\;(\gamma x_i\,.x_j + r)$	(19)

Where: d, r and γ are kernel parameters.

2.2.3. FFA

The FFA is an evolutionary population based metaheuristic algorithm. FFA was first proposed by Xin-She Yang in late 2007 and 2008 [32,33]. FFA has been inspired from the behavior of the swarm such as bird folks, insects and fish schooling in nature. Numerous recent research has shown that FFA is a favorable algorithm that performs better than the performance of other metaheuristic algorithms such as genetic algorithms [32–35]. Inspired by real fireflies, FFA possesses three flashing features and idealized rules that may be summed up as follows:

1. Because they are all unisex, fireflies will mate with another firefly of any gender or sex.
2. A firefly's attraction is inversely correlated with brightness and diminishes with increasing distance from another firefly. The brighter firefly approaches a less brighter one. In the absence of a brighter firefly, the firefly travels randomly.
3. The brightness of the firefly depends on the value of the objective function.

When using the FFA, because light is absorbed by the surroundings, a firefly's attractiveness I diminishes with distance r from its source (another firefly). The variations in brightness $I(r)$ is:

$$I(r) = \frac{I_0}{r^2} \tag{20}$$

I_0 is the source's brightness. With the fixed light absorption coefficient γ and to prevent singularity at r = 0, it is possible to approximate the combined impact of the absorption with (Eq 20) to a Gaussian form; that is:

$$I(r) = I_0 e^{-\gamma r^2} \tag{21}$$

Since a firefly's attractiveness is proportional to the light intensity, the attractiveness function of the firefly can be defined as

$$B(r) = B_0 e^{-\gamma r^2} \tag{22}$$

B_0 at $r = 0$ represents the initial attractiveness.

The distance r_{ij} between two fireflies positions, x_{i} and x_{j}, may be described as either a Cartesian or Euclidian distance, as shown in [32,34,36]; that is,

$$r_{\mathrm{ij}} = ||x_{\mathrm{i}} - x_{\mathrm{j}}|| = \sqrt{\sum_{d=1}^{D} (x_{\mathrm{i,d}} - x_{\mathrm{j,d}})^2} \tag{23}$$

The total number of dimensions equals D, and k corresponds to the dth component for the spatial coordinates x_{i} concerning i th firefly.

When firefly j is more luminous than firefly i, the firefly i's motion toward firefly j may be described as

$$x_i = x_i + \beta_0 * exp(-\gamma r_{ij}^2)\,(x_{\mathrm{i}} - x_{\mathrm{j}}) + \alpha * (rand - 0.5) \tag{24}$$

The very first term refers to the firefly's present position; the following is how attractive it is to neighboring fireflies based on brightness; the final term is the firefly's erratic movement within the absence of a brighter firefly. A random number, rand, and a randomization parameter, α, both equal ∈[0,1].

Absorption coefficient γ gives rise to two specific scenarios in the FFA:

In the scenario when $\gamma = \infty$, the fireflies are almost invisible to one another and have very little attraction to brightness. The FFA functions similarly to a random walk technique as a result.

However, when $\gamma = 0$, the attraction coefficient remains constant ($\beta = \beta_0$) and the brightness does not diminish as the distance (r) grows. The PSO technique and the FFA are consistent in this scenario.

2.3. Models' architecture

2.3.1. ANN architecture

In general, the database (input and output) affects the ANN multilayer neural network's MLP design and structure. In this model, we have used five hidden layers for daily solar radiation estimation. Features of the developed ANN are given in Table 4 and Figure 2.

Table 4. The ANN features.

Activation function		Traning algorithm
Hidden layer: Tansig	Output layer: Pureline	Trainlm

For our output layer, we've only utilized one neuron (see Figure 2). In each of the six cities we selected, we used a script file generated by the MATLAB software for all models.

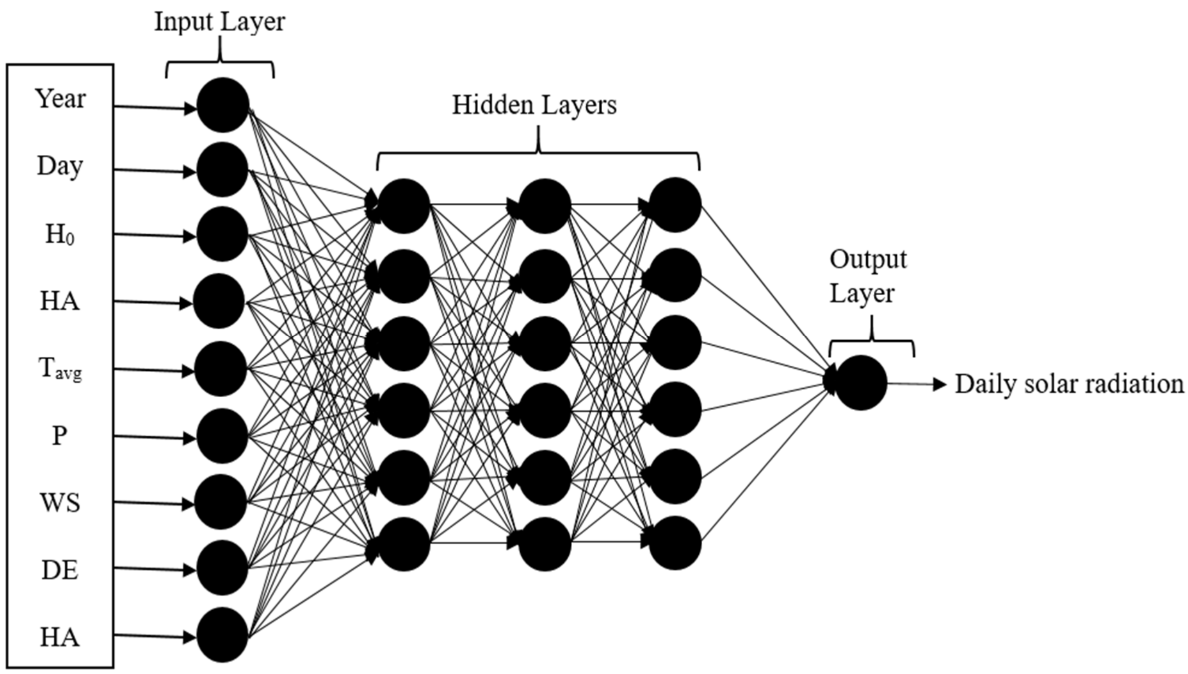

Figure 2. ANN based model.

2.3.2. Hybrid FFA-ANN based model

The following are the basic steps of the FFA:

- Step 1: The FFA starts with the initial settings of the most crucial parameters (β_0, γ, α), in addition to population size P and maximum generation number (MGN), which serves as the FFA's termination criterion (see Table 5).
- Step 2: The primary population, x_i, i = {1,...,P}, is created at random, and each solution $f(x_i)$ within the population is assessed for fitness by computing its associated objective function.
- Step 3: Until the required number of MGN iterations is reached, the following procedures are repeated in order to satisfy the termination condition.

It is challenging to choose the optimal subset of input variables. Testing each potential subset's fit methodically is the obvious way to choose this subset. Even though this thorough search ensures the best conclusion for a particular ANN model, it necessitates the exploration of (2d-1) combinations and, thus, entails a significant processing load. The ANN's capacity to accurately anticipate solar radiation is reliant on the selection of its inputs. In this work, the ANN model was generated and trained using various input combinations (see Figure 3).

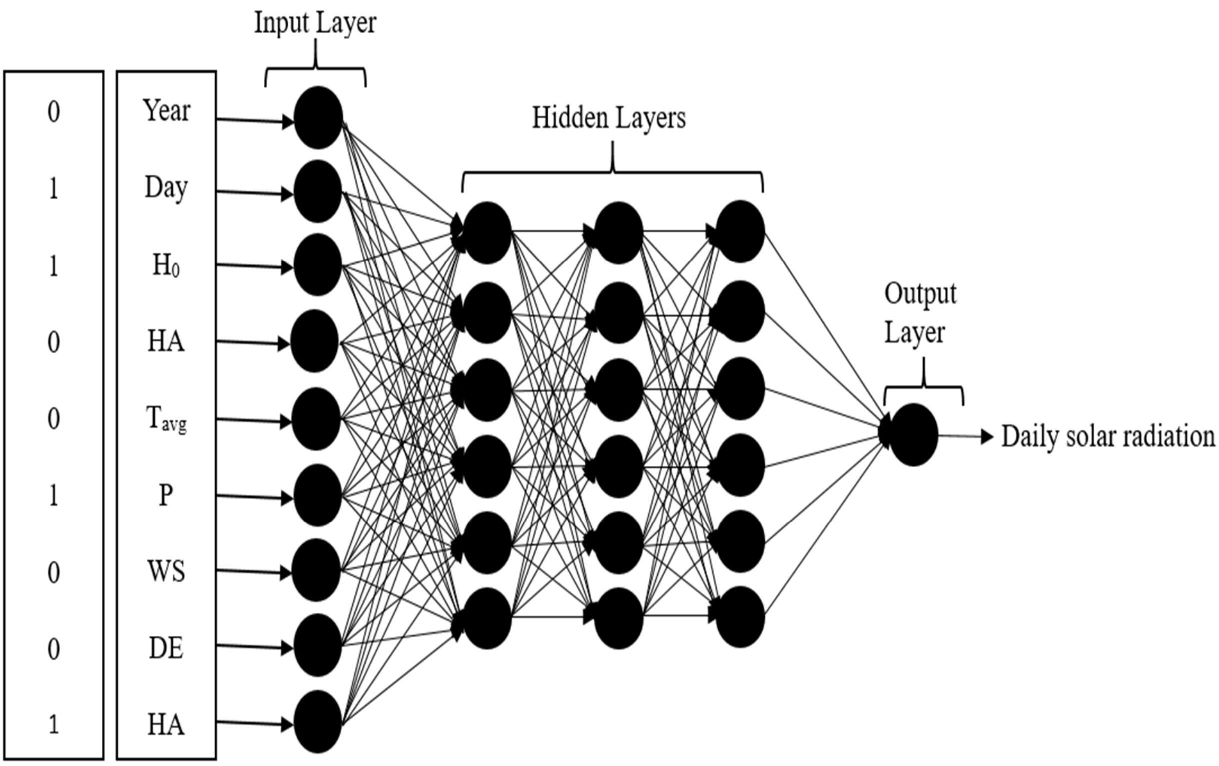

Figure 3. FFA-ANN based model.

In our proposed model, we adopt a hybrid machine learning algorithm based on the FFA-ANN to favor selecting the small and most relevant dataset of inputs with strong relationships to solar irradiation to train and validate the ANN model. The most relevant inputs are elected automatically using an evolution paradigm based on the FFA.

Table 5. Firefly algorithm parameters.

Description	Population size	Maximum generation	A	γ	Delta	β_0
Value	10	50	1	0.6	0.8	1

2.3.3. The SVM model

The SVMs were optimized and trained using the leaner app regression, using Matlab 2021a. In this study, the cubic and linear kernel following a series of trial and error tests were selected as the best. SVM performance depends on the slack parameter (ε), according to SVM hypothesis. In this study, this parameter (ε) was optimized by using the Bayesian algorithm with 30 iterations, adjusting the ε values from ε = 1.7529 to ε = 259485.0259. By adding up a couple of query points as well as their function value, an overall solution may be obtained by repeating the following two steps: (i) Simulating a surrogate function; (ii) figuring out where to query next by maximizing an acquisition function.

Only when the overall optimizer of the acquisition function is located and chosen as the subsequent query point during each round do convergence guarantees hold true. Local optimizers for the acquisition function are also employed in practice, though, as locating an overall optimizer is sometimes a difficult and time-consuming undertaking. The optimized values of the parameter ε for the SVM models are presented in Table 6.

Table 6. SVM optimized parameters.

Station	Kernel function	Optimum values Ɛ
Adrar	Cubic	2.0255
El-oued	Cubic	24.6855
Ouargla	Linear	2.682
Tamanrasset	Cubic	16.3562
Timimoun	Cubic	7.6419
Bechar	Linear	5.4471

2.4. Assessment indicators

Table 7. Statistical assessment indices.

Indice	Optimal value	Equation	
$\boldsymbol{rRMSE}$ %	0	$= \frac{1}{\bar{G}_{Act}}\sqrt{\frac{\sum_{i=1}^{N}\left(G_{Sim,i}-G_{Act,i}\right)^2}{N}}$	(25)
$\boldsymbol{MAPE}$ %	0	$= \frac{100}{N}\sum_{i=1}^{N}\left\lvert\frac{G_{Sim,i}-G_{Act,i}}{X_{Act,i}}\right\rvert$	(26)
$\boldsymbol{R}$	1	$= \frac{\sum_{i=1}^{N}\left(G_{Sim,i}-\bar{G}_{Sim}\right)\left(G_{Act,i}-\bar{G}_{Act}\right)}{\sqrt{\sum_{i=1}^{N}\left(G_{Sim,i}-\bar{G}_{Sim}\right)^2}\times\sqrt{\sum_{i=1}^{N}\left(G_{Act,i}-\bar{G}_{Ac}\right.}}$	(27)

Measures that are often used in assessment scores were used to examine the models' efficiency [37]. The definitions of these indices are shown in Table 7 below, where N is the overall number of observations, G_{Act} is the actual value and G_{Sim} are the predicted values.

3. Results and discussion

The ANN, FFA-ANN and SVM models were developed to forecast for daily solar radiation in six locations in Algeria. Table 8 displays the statistical indicators of performance for every model. Based on the findings, it is readily apparent that the suggested hybrid FFA-ANN approach performed very well and gave a very high accuracy for: Adrar city with (MAPE = 6.62; rRMSE = 10.10; R = 0.911); Ouargla city with (MAPE = 7.9; rRMSE = 11.47; R = 0.923); Tamanrasset city with (MAPE = 6.29; rRMSE = 9.35; R = 0.905); Bechar city with (MAE = 8.23; rRMSE = 11.76; R = 0.920); EL-Oued city with (MAPE = 8.26; rRMSE = 11.96; R = 0.9321); Timimoun city with (MAPE = 6.88; rRMSE = 10.48; R = 0.928).

Table 8. Statistical indicators.

		MAPE%	rRMSE%	R
	ANN	7.37	10.81	0.895
Adrar	FFA-ANN	6.62	10.01	0.911
	SVM	6.09	10.54	0.908
	ANN	9.49	13.17	0.917
El oued	FFA-ANN	8.26	11.96	0.9321
	SVM	7.41	12.43	0.9320
	ANN	8.02	11.84	0.918
Ouargla	FFA-ANN	7.9	11.47	0.923
	SVM	8.96	13.37	0.898
	ANN	6.91	10.01	0.890
Tamanrasset	FFA-ANN	6.29	9.35	0.905
	SVM	5.79	9.77	0.903
	ANN	7.87	11.59	0.911
Timimoun	FFA-ANN	6.88	10.48	0.928
	SVM	6.21	11.05	0.926
	ANN	8.38	12.06	0.915
Bechar	FFA-ANN	8.23	11.76	0.920
	SVM	9.73	13.99	0.888

We can clearly notice, as indicated by Table 8 and Figure 4, that the FFA-ANN hybrid approach outperformed the stand-alone ANN. Table 9 shows the optimal combination for each site when using the ANN technique.

Table 9. FFA-ANN best combination.

Location	The best combination [Y, D, H_0, DE, T_{avg}, RH, BP, HA, WS]
Adrar	[0.0.1.1.1.1.1.0.1]
El-oued	[0.0.1.1.1.1.1.0.1]
Ouargla	[0.0.1.1.1.1.1.1.0]
Tamanrasset	[0.0.1.1.1.1.1.1.1]
Timimoun	[0.0.1.1.1.1.1.0.1]
Bechar	[0.0.1.1.1.1.0.0.0]

We can conclude from the results shown in Table 9 that each city has its own best input data combination. However, we notice the importance of some inputs for the ANN model when we see their participation in each combination for all cities. These inputs are: H_0, DE, T_{avg} and RH.

However, when trying the same inputs for the SVM model, the accuracy of the model was not necessarily enhanced and, therefore the optimal combinations for the ANN model might not necessarily be the optimal combinations for the SVM model. Aspects of features unique to a given model and the complexity of the modeling approaches can all contribute to variations in the ideal feature set.

For this research, a hybrid FFA-ANN approach was developed in order to estimate global solar radiation. For comparing and evaluating the performances of our proposed model, the two MAPE and rRMSE main statistical output indicators were used. The findings of the overall average statistical of rRMSE and MAPE for both models involved are reported in Table 8 and Figure 4.

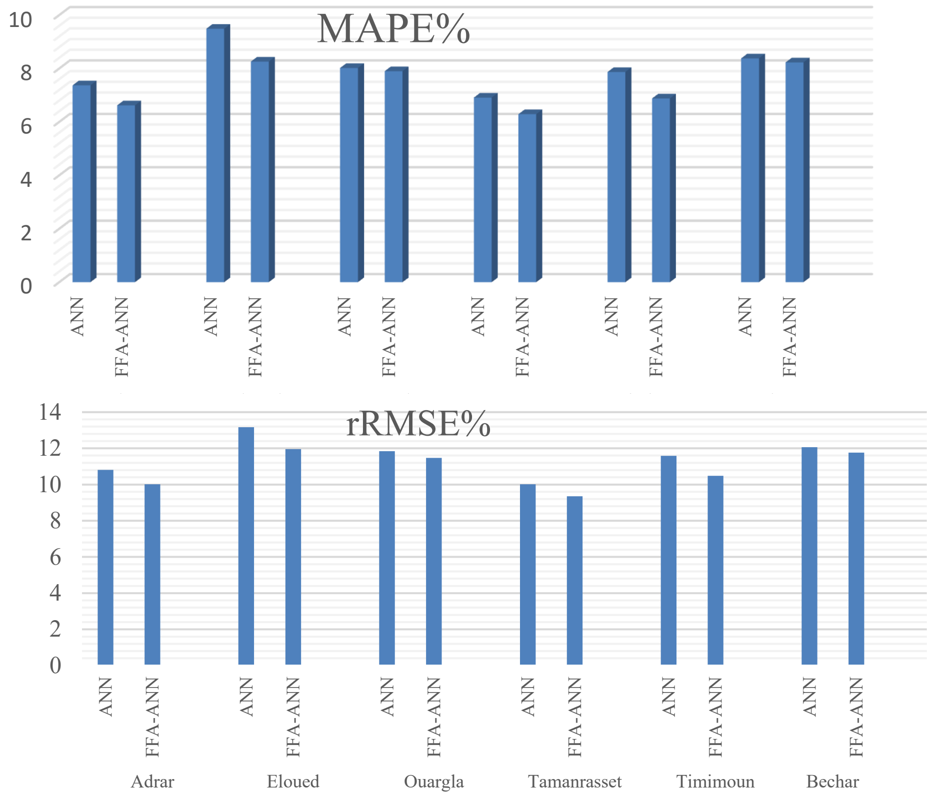

Figure 4. MAPE (%) and rRMSE (%) For the ANN and FFA-ANN.

Both figures demonstrate that when using the proposed firefly algorithm to choose the best combination ANN model, it performed better, leading us to the fact that the FFA-ANN hybrid model outperformed the stand-alone ANN model.

The actual and anticipated values of the most accurate model, FFA-ANN, over the research period are shown in Figure 5. This model has high accuracy, which makes it recommended for solar radiation regression and ANN model optimization. It has the advantage of giving the best combination for each location, which would be less time-consuming in the future and easier when collecting only needed data.

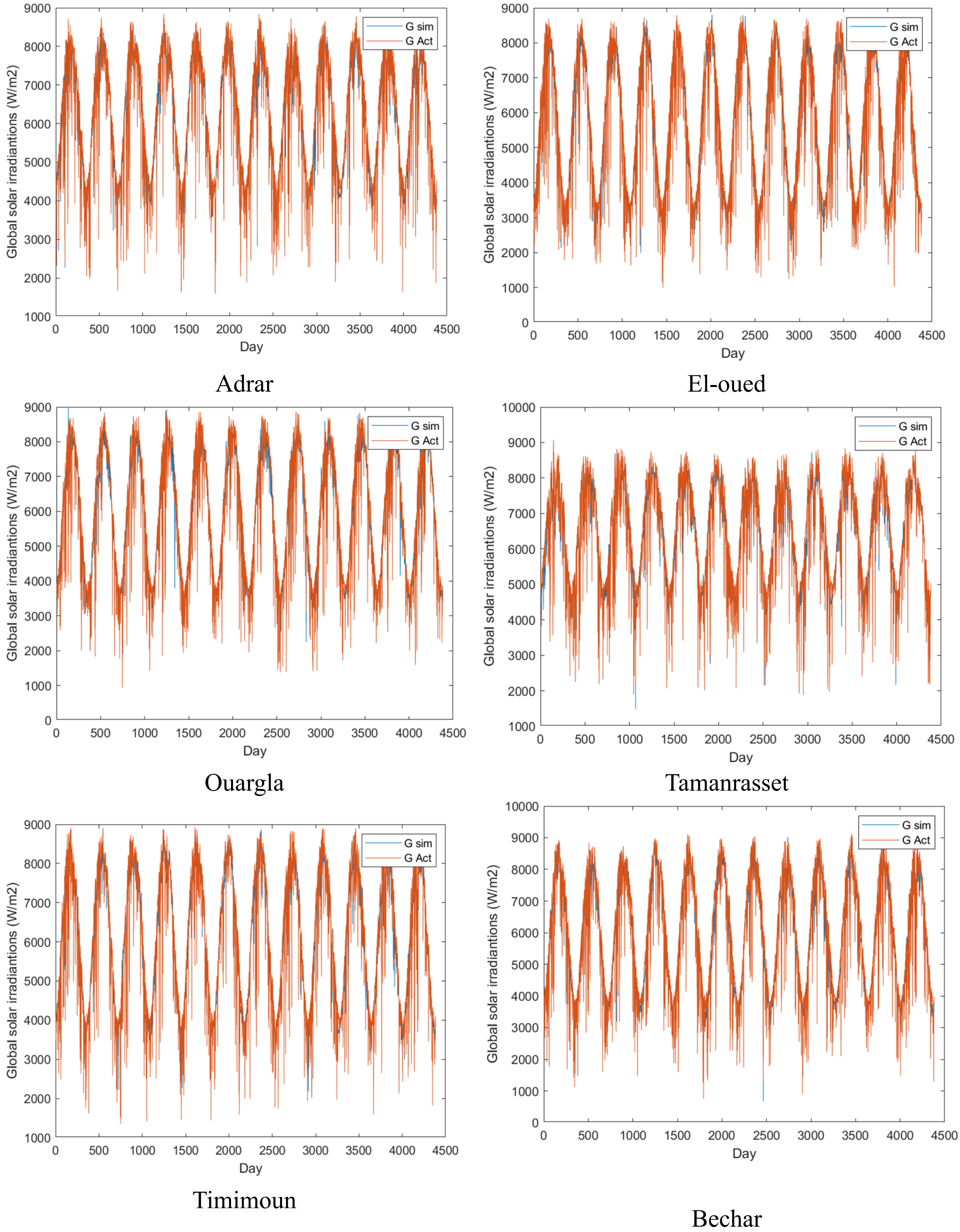

Figure 5. Predicted and collected values of the FFA-ANN models during an eleven-year period.

Similarly, scatter plots (see Figure 6) of the FFA-ANN models predicted DGSR and showed the forecasted data distributed as a series of points close to the linear (red) ideal fit, showing the relationship between the anticipated and actual values across the six selected cities throughout the course of the study.

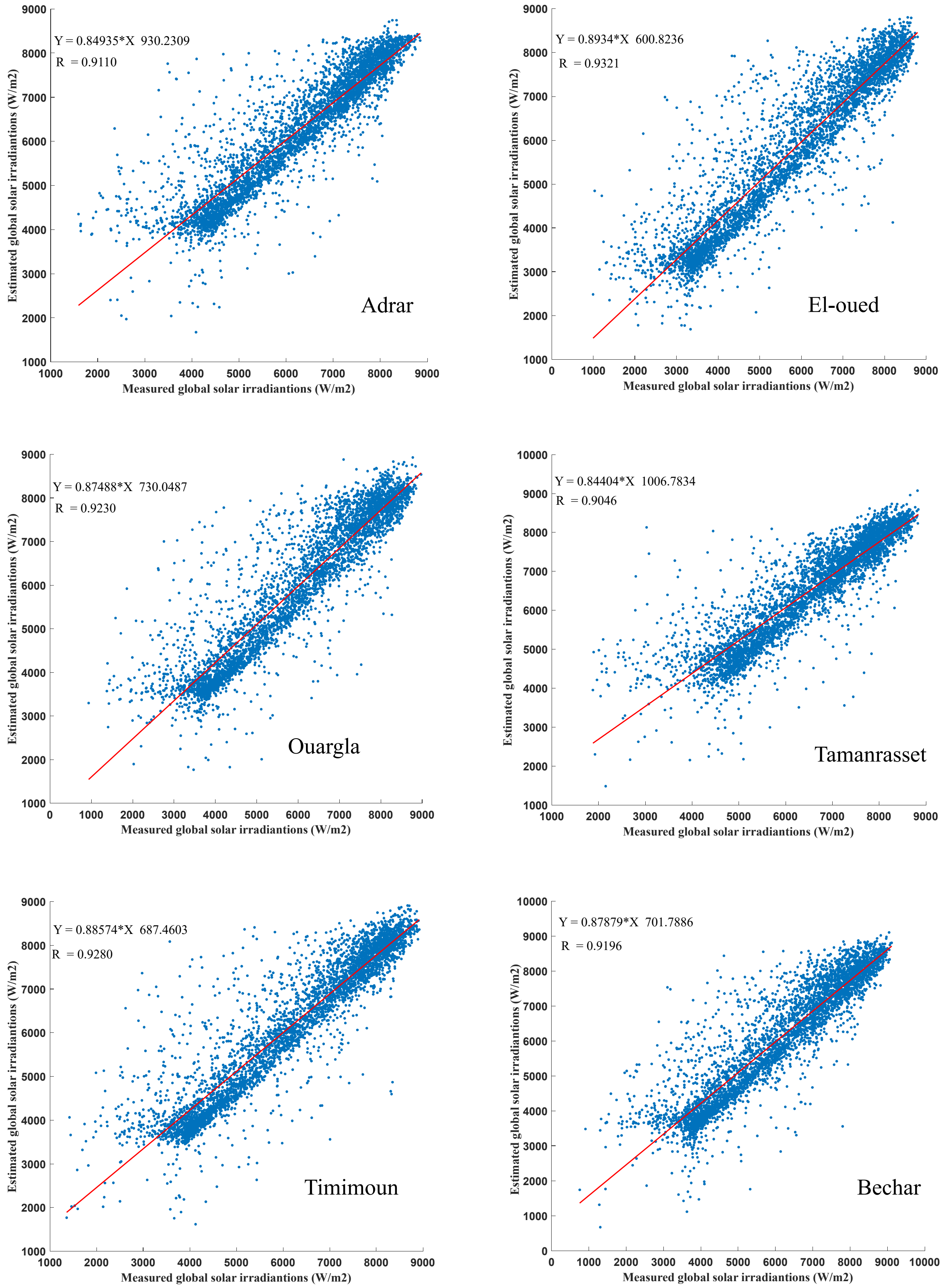

Figure 6. Scatter plot for FFA-ANN models.

Unlike traditional machine learning (ML) models with hyperparameter optimization such as stand-alone ANN and SVM, FFA-ANN has a higher computational cost due to the need for multiple runs of the ANN model, which leads the running time to be noticeably longer. However, when it comes to the effectiveness of simulation, FFA-ANN excels in capturing complex relationships, which leads to the highest accuracy, while traditional ML models perform well.

A variety of models created to forecast sun irradiation are shown in Table 10. Neural networks and empirical methods are compared with the suggested hybrid FFA-ANN model. Three statistical index errors—R, rRMSE and MAPE—are used to examine the reliability of predictions of the mentioned and present techniques. When compared to the other techniques, the findings of the proposed investigation demonstrate high and compatible effectiveness in regards to predictability and better forecasting capabilities. The results are consistent between them.

Table 10. statistics index errors across many models.

Reference	Model	Time step	Location	R	rRMSE%	MAPE%
[38]	ANN	Daily	Nevs¸ehir city, Turkey	0.965	14.10	15.92
[38]	k-NN	Daily	Karaman city, Turkey	0.941	25.19	30.24
[38]	SVM	Daily	Tokat city, Turkey	0.938	18.43	23.37
[39]	Empirical	Daily	Tamanrasset city, Algeria	0.891	-	-
[40]	ANN-FFA	Hourly	Klang Valley, Malaysia	-	23.47	13.00
The presented study	FFA-ANN	Daily	ELoued city, Algeria	0.9321	11.96	8.26

4. Conclusions

This study was conducted for the following six south Algerian cities: Adrar, EL-oued, Ouargla, Timimoun, Tamanrasset and Bechar, where three models (ANN, FFA-ANN and SVM) were examined for accuracy to estimate the DGSR during an eleven-year period (2010 to 2021), using nine input parameters as input data. Our proposed hybrid FFA-ANN model was developed with the aim of optimizing the ANN model built by our team. Despite the fact that the ANN and SVM models provided promising results, based on the statistical indicators (R, rRMSE and MAPE), the findings showed that our suggested FFA-ANN hybrid method performed higher than the stand-alone ANN-based model. Undeniably, the findings showed that FFA-ANN was better suited for the forecast process in all sites and even outperformed the SVM model. Moreover, comparing the combinations revealed the ANN model's reliance on H_0, DE, T_{avg} and RH as inputs when forecasting DGSR.

Thus, as a consequence, the findings of the investigation suggest that whenever data is accessible, the created model may be employed to forecast DGSR in locations with dry climates as well as for other locations with comparable climates. It may also be helpful for choosing the installation of solar energy systems.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are grateful to the laboratory of Sustainable Development and Computer Science (LSDCS) for supporting this research.

References

1. Kabouche N, Chellali F, Recioui A (2021) A review on solar radiation assessment and forecasting in Algeria: (Part 2; Solar radiation forecasting). *Algerian J Sign Syst* 6: 130–146. https://doi.org/10.51485/ajss.v6i3.141
2. Gairaa K, Khellaf A, Messlem Y, et al. (2016) Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach. *Renewable Sustainable Energy Rev* 57: 238–249. https://doi.org/10.1016/j.rser.2015.12.111
3. Benatiallah D, Benatiallah A, Bouchouicha K, et al. (2020) Prediction of hourly solar radiation using artificial neural networks. *Algerian J Env Sci Technol* 6: 1236–1245. Available from: https://www.aljest.net/index.php/aljest/article/view/272.
4. Antonopoulos VZ, Papamichail DM, Aschonitis VG, et al. (2019) Solar radiation estimation methods using ANN and empirical models. *Comp Electron Agric* 160: 160–167. https://doi.org/10.1016/j.compag.2019.03.022
5. Amiri B, Gómez-Orellana AM, Gutiérrez PA, et al. (2020) A novel approach for global solar irradiation forecasting on tilted plane using hybrid evolutionary neural networks. *J Clean Produc* 287: 125577. https://doi.org/10.1016/j.jclepro.2020.125577
6. Alsina EF, Bortolini M, Gamberi M, et al. (2016) Artificial neural network optimisation for monthly average daily global solar radiation prediction. *Energy Conv Manag* 120: 320–329. https://doi.org/10.1016/j.enconman.2016.04.101
7. Assas O, Bouzgou H, Fetah S, et al. (2014) Use of the artificial neural network and meteorological data for predicting daily global solar radiation in Djelfa, Algeria. *2014 Int Conf on Composite Materials Renewable Energy Applications (ICCMREA)*, *Sousse, Tunisia,* 1–5. https://doi.org/10.1109/ICCMREA.2014.6843807
8. Voyant C, Muselli M, Paoli C, et al. (2011) Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. *Energy* 36: 348–359 https://doi.org/10.1016/j.energy.2010.10.032
9. Hasni A, Sehli A, Draoui B, et al. (2012) Estimating global solar radiation using artificial neural network and climate data in the south-western region of Algeria. *Energy Proc* 18: 531–537. https://doi.org/10.1016/j.egypro.2012.05.064

10. Meenal R, Immanuel Selvakumar A (2018) Assessment of SVM, Empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renewable Energy* 121: 324–343. https://doi.org/10.1016/j.renene.2017.12.005
11. Mohammadi K, Shamshirband S, Danesh AS, et al. (2020) Retracted article: Horizontal global solar radiation estimation using hybrid SVM-firefly and SVM-wavelet algorithms: A case study. *Natural Hazards* 102: 1613–1614. https://doi.org/10.1007/s11069-015-2047-5
12. Guermoui M, Rabehi A, Gairaa K, et al. (2018) Support vector regression methodology for estimating global solar radiation in Algeria. *European Phys J Plus* 133: 22. https://doi.org/10.1140/epjp/i2018-11845-y
13. Guijo-Rubio D, Durán-Rosal AM, Gutiérrez PA, et al. (2020) Evolutionary artificial neural networks for accurate solar radiation prediction. *Energy* 210: 118374. https://doi.org/10.1016/j.energy.2020.118374
14. Belaid S, Mellit A (2016) Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. *Energy Conver Manage* 118: 105–118. https://doi.org/10.1016/j.enconman.2016.03.082
15. Shamshirband S, Mohammadi K, Khorasanizadeh H, et al. (2016) Estimating the diffuse solar radiation using a coupled support vector machine-wavelet transform model. *Renewable Sustainable Energy Rev* 56: 428–435. https://doi.org/10.1016/j.rser.2015.11.055
16. Bakhashwain JM (2016) Prediction of global solar radiation using support vector machines. *Int J Green Energy* 13: 1467–1472. https://doi.org/10.1080/15435075.2014.896256
17. Baser F, Demirhan H (2017) A fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation. *Energy* 123: 229–240 https://doi.org/10.1016/j.energy.2017.02.008
18. Queja VH, Almoroxa J, Arnaldob JA, et al. (2017) ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *J Atmos Sol-Terrestrial Physic* 155: 62–70. https://doi.org/10.1016/j.jastp.2017.02.002
19. Bhola P, Bhardwaj S (2019) Estimation of solar radiation using support vector regression. *J Inf Optim Sci* 40: 339–350. https://doi.org/10.1080/02522667.2019.1578093
20. Fan J, Wu L, Ma X, et al. (2019) Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in airpolluted regions. *Renewable Energy* 145: 2034–2045. https://doi.org/10.1016/j.renene.2019.07.104
21. Liu Y, Zhou Y, Chen Y, et al. (2020) Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: A case study in China. *Renewable Energy* 146: 1101–1112. https://doi.org/10.1016/j.renene.2019.07.053
22. Shamshirband S, Mohammadi K, Tong CW, et al. (2016) Retracted article: A hybrid SVM-FFA method for prediction of monthly mean global solar radiation. *Theor Appl Climatol* 125: 53–65. https://doi.org/10.1007/s00704-015-1482-2
23. Olatomiwa L, Mekhilef S, Shamshirband S, et al. (2015) A support vector machine–firefly algorithm-based model for global solar radiation prediction. *Sol Energy* 115: 632–644. https://doi.org/10.1016/j.solener.2015.03.015
24. Benatiallah D, Bouchouicha K, Benatiallah A, et al. (2019) Forecasting of solar radiation using an empirical model. *Algerian J Renewable Energy Sustainable Dev* 1: 212–219. https://doi.org/10.46657/ajresd.2019.1.2.11

25. Benatiallah D, Benatiallah A, Harouz A, et al. (2016) Development and modeling of a geographic information system solar flux in Adrar, Algeria. *Int J Syst Mod Simul* 1: 15–19. Available from: https://www.aljest.net/index.php/aljest/article/download/524/500.
26. SODA data. Available from: www.soda-pro.com/web-services#meteodata.
27. Haykin S, Lippmann R (1994) Neural networks, a comprehensive foundation. *Int J Neural Syst* 5: 363–364. https://doi.org/10.1142/S0129065794000372
28. Yadav AK, Chandel SS (2014) Solar radiation prediction using artificial neural network techniques: A review. *Renewable Sustainable Energy Rev* 33: 772–781. http://doi.org/10.1016/j.rser.2013.08.055
29. Ata R (2015) Artificial neural networks applications in wind energy systems: A review. *Renewable Sustainable Energy Rev* 49: 534–562. https://doi.org/10.1016/j.rser.2015.04.166
30. Esmaeelzadeh SR, Adib A, Alahdin S (2015) Long-term streamflow forecasts by adaptive neuro-fuzzy inference system using satellite images and K-fold crossvalidation (case study: Dez, Iran). *KSCE J Civ Eng* 19: 2298–2306. https://doi.org/10.1007/s12205-014-0105-2
31. Vapnik V (2013) The nature of statistical learning theory. *Springer Sc & Bus Media, Berlin, Heidelberg, Germany* 267–287. Available from: https://statisticalsupportandresearch.files.wordpress.com/2017/05/vladimir-vapnik-the-nature-of-statistical-learning-springer-2010.pdf.
32. Yang XS (2008) Nature-inspired metaheuristic algorithms. Luniver Press, UK, 81–89. Available from: https://www.academia.edu/457296/Nature_inspired_metaheuristic_algorithms.
33. Chantasut N, Charoenjit C, Tanprasert C (2004) Predictive mining of rainfall predictions using artificial neural networks for chao phraya river. *4th Inter Conf of The Asian Feder of Info Tech in Agriculture and The 2nd World Cong on Comp in Agriculture and Natural Res*, *Bangkok, Thailand,* 117–122. Available from: https://www.thaiscience.info/Journals/Article/NETC/10438493.pdf.
34. Lukasik S, Zak S (2009) Firefly algorithm for continuous constrained optimization tasks. *In Proceedings of the Inter Conf on Comp and Comput Intelligence (ICCCI 09). Springer, Wroclaw, Poland* 5796: 97–106. Available from: https://link.springer.com/chapter/10.1007/978-3-642-04441-0_8.
35. Yang XS (2010) Firefly algorithm, stochastic test functions and design optimization. *Int J Bio-Inspired Comp* 2: 78–84. https://doi.org/10.1504/IJBIC.2010.032124
36. Yang XS (2009) Firefly algorithms for multimodal optimization. In: Watanabe, O., Zeugmann, T., *Stochastic Algorithms: Foundations and Applications.* SAGA 2009. Lecture Notes in Computer Science, Berlin, Heidelberg. 5792: 169–178. https://doi.org/10.1007/978-3-642-04944-6_14
37. Stone RJ (1993) Improved statistical procedure for the evaluation of solar radiation estimation models. *Sol Energy* 89: 51–91. https://doi.org/10.1016/0038-092X(93)90124-7
38. Agbulut Ü, Etem Gürel A, Biçen Y (2021) Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable Sustainable Energy Rev* 135: 110114. https://doi.org/10.1016/j.rser.2020.110114
39. Ihaddadene N, El Hacen Ould Ahmedou M, Jed B, et al. (2019) Daily global solar radiation estimation based on air temperature: Case of study south of Algeria. *E3S Web Conf* 80: 01002. https://doi.org/10.1051/e3sconf/20198001002

40. Anwar Ibrahim I, Khatib T (2017) A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. *Energy Conv Manage* 138: 413–425. https://doi.org/10.1016/j.enconman.2017.02.006