Smart Grids and Sustainable Energy (2024) 9:3
https://doi.org/10.1007/s40866-023-00179-w

ORIGINAL PAPER

A Novel Non-Linear Model Based on Bootstrapped Aggregated Support Vector Machine for the Prediction of Hourly Global Solar Radiation

Abdennasser Dahmani[1,2] · Yamina Ammi[2] · Salah Hanini[2]

Received: 9 September 2022 / Accepted: 3 November 2023

Abstract
The prediction of global solar radiation (GSR) for various regions is of great importance as it provides guidance for the design, modeling, and operation of solar energy conversion systems, the selection of suitable regions, and even informs future investment policies for decision-makers. This paper presents the methods for predicting hourly mean solar radiation using support vector machines (SVM), which is a machine learning algorithm based on statistical learning theory. The primary objective of this paper was to investigate the use of a support vector machine (SVM) based on quantitative structure–activity relationships, specifically a single support vector machine (SSVM) and a bootstrap aggregated support vector machine (BASVM), to predict hourly global solar radiation in Bouzareah city. A dataset consisting of 3603 data points was employed to develop both the SSVM and BASVM models. Bootstrap aggregation of SVM is utilized to enhance the accuracy and robustness of SVM models constructed from limited training datasets. The training dataset is resampled using bootstrap resampling with replacement to create an ensemble of SVM models, each trained on a different sample from the training set. A support vector machine model is developed, and individual Support Vector Machines (SVMs) are then combined to form a Bootstrap Aggregated Support Vector Machine (BASVM). Experimental data for global solar radiation (GSR) were compared to the calculated GSR, and excellent correlation coefficients (R) were found (0.9913) during the testing phase. This novel BASVM model could be utilized by researchers and scientists to design high-efficiency solar devices.

Keywords Bootstrap · Support vector machine · Prediction · Solar radiation

Introduction

The demand for renewable energy has been rapidly increasing in recent years due to the negative effects of fossil fuel-based energy sources on our environment and their contribution to climate change. Consequently, there has been a growing interest in clean energy resources, such as solar energy [1]. As a result, predicting solar radiation reaching the Earth's surface has paramount importance for various applications, including engineering designs, heating and cooling systems, building energy systems, medical studies, agriculture, climatological research, evapotranspiration studies, solar collector efficiency, and seawater desalination [2, 3]. Reliable solar radiation statistics are essential for achieving successful outcomes in any of these fields of research.

Algeria is strategically positioned in the Sunbelt, offering a significant advantage in terms of its solar energy potential. Throughout the national territory, annual sunlight duration exceeds 3,000 h, and in the high plateaus and the Sahara, it can reach up to approximately 3,900 h [4]. Regrettably, obtaining accurate solar irradiation measurements in various regions of Algeria remains a challenge, primarily due to the high costs associated with measurement equipment such as solarimeters and pyranometers, as well as the expense, maintenance, and calibration requirements of the systems involved. While Algeria hosts numerous meteorological stations in different parts of the country, the availability of solar irradiation measurements is not always guaranteed. This is

✉ Abdennasser Dahmani
dahmani.abdennasser@univ-relizane.dz

[1] Department of Mechanical Engineering, Faculty of Science and Technology, GIDD Industrial Engineering and Sustainable Development Laboratory, University of Relizane, Bourmadia, 48 000 Relizane, Algeria

[2] Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Medea, Urban Pole, 26000 Medea, Algeria

Published online: 30 November 2023

often due to issues related to recording problems caused by frequent power outages, especially during the summer months, or limitations in the number of variables that can be recorded. As a result, it becomes significantly more important to employ sophisticated procedures for accurately estimating solar radiation using readily available meteorological data [5]. In light of the absence of solar radiation measurements in all regions of the Earth, various models have been devised to estimate solar radiation in areas lacking monitoring stations. The progress in developing global solar radiation models is continuous; however, it's essential to recognize that these models may produce differing outcomes across various regions. Consequently, it holds significance to construct location-specific models whenever possible. As part of this initiative, a study was conducted to formulate solar radiation models customized explicitly for India [6].

Over time, several artificial intelligence (AI) methods for predicting global solar radiation on a horizontal surface have been developed. These methods include the support vector machine approach for estimating global solar radiation while accounting for the influence of fog and haze [7], as well as the least squares-support vector machine (LS-SVM) [8]. Hansen and Salamon [9] introduced the concept of an aggregated or stacked neural network, which enhances a model's generalization by training multiple neural networks and fusing their outputs. This highly effective approach has found wide application [10]. Research has demonstrated that stacked neural networks outperform individual ones in terms of generalization capability [11]. Literature studies have shown that artificial neural networks (ANN) are superior to traditional empirical models in predicting solar radiation (SR) [12]. Support vector machines (SVM), developed by Vapnik [13], have recently found wide application in computer science, bioinformatics, and environmental science [14, 15]. Previous studies have proven that SVMs perform better than neural networks and other statistical models [14]. However, there is limited literature on the application of SVMs in predicting SR. Thus, the goal of this research work is to develop a support vector machine-based bootstrap aggregated support vector machine (BASVM) model to predict hourly global solar radiation received on the horizontal plane over one year in the Bouzareah region of Algeria. This prediction will be based on nine meteorological and climatological parameters: Month, Day, Time (h), Average Temperature (K), Relative Humidity (%), Atmospheric Pressure (mbar), Wind Speed (m/s), Wind Direction (°), and global solar radiation (Wh/m^2).

To the best of our knowledge, no studies using a bootstrap-based support vector machine for predicting solar radiation or in any other domain have been described in the literature. This will be the first study to predict global solar radiation using the BASVM Model. We will compare the individual support vector machine (ISVM) and a single support vector machine (SSVM) to the BASVM. The paper is structured as follows: Section 2 presents the materials and methods, Section 3 introduces the evaluation criteria, Section 4 covers the results and discussion, and Section 5 summarizes the conclusions of our research.

Literature Review

The modeling of solar irradiation has been the focus of several research and studies, with the most significant conducted in the last two decades. Quej et al. [16] employed three machine learning algorithms, specifically SVM, ANN, and ANFIS, to predict daily global solar radiation data for six stations in Mexico. The algorithms were trained using extraterrestrial solar radiation, rainfall, minimum temperature, and temperature data. The comparative analysis revealed that SVM outperformed the other models, achieving the best results with a root mean square error (RMSE) of 2.578, mean absolute error (MAE) of 1.97, and coefficient of determination (R^2) of 0.689. Dos Santos et al. [17] evaluated hourly and daily direct solar radiation using two methods, ANN and SVM, with 13 years of data. The results demonstrated the positive performance of both methods. Lima et al. [18] applied three predictive models, namely multilayer perceptron back propagation neural network (MLPBP-NN), the Radial Basis Function network (RBF), and the support vector machine (SVM), to predict daily solar radiation. The input variables used for these models were solar irradiance and temperature. The results showed that the proposed model exhibited high efficiency in forecasting daily solar radiation, especially in areas located close to the Equator line.

Takilate et al. [19] developed a novel approach to estimate inclined irradiation in 5-min intervals across three distinct climatic regions: Algiers and Ghardaïa in Algeria, and Malaga in Spain. The model is a combination of two conventional models, namely the Perrin Brichambaut and Liu and Jordan models. The results demonstrated that the normalized root mean square error (nRMSE) ranged from 4.7% to 6.41%, indicating the model's accuracy in predicting inclined irradiation in the specified areas. Gao et al. [20] propose a hybrid hourly irradiance forecasting method that combines CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) with Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM). They conclude that the proposed hybrid approach yields better results than a large number of alternative methods. Peng et al. [21] introduced a novel hybridization methodology, primarily relying on the utilization of the recent CEEMDAN algorithm as a pre-processing technique, in combination with the sine cosine search algorithm (SCA) for feature selection, and Bidirectional Long

Short-Term Memory (BiLSTM) as the core prediction model. The proposed CEEMDAN-SCA-Bi-LSTM model demonstrated superior forecasting accuracy compared to seven reference models. Keshtegar et al. [22] Conducted a study that evaluated the effectiveness of four empirical regression methods – namely Kriging, MARS, RSM, and M5 Tree – in accurately assessing solar energy using diverse input data from Adana and Antakaya in Turkey. The findings revealed that the Kriging model exhibited superior performance when compared to the cyclic MARS, RSM, and M5 Tree models. This investigation took place in the context of West Africa. Nwokolo et al. [22] Provided a quantitative evaluation of the global solar energy literature. Utilizing a range of models such as sunlight-based, temperature-based, precipitation-based, cloud-core, comparative humidity-based, and hybrid parameter-based models, they amassed a collection of 356 empirical models and 68 functional forms. These studies collectively showcase the evolution of solar irradiation modeling, with a growing emphasis on the integration of advanced algorithms and hybrid methodologies to achieve more accurate and reliable predictions, In the context of this present research work, we have developed a non-linear model based on bootstrapped aggregated support vector machine (BASVM) for predicting hourly global solar radiation received on the horizontal plane over one year in the region of Bouzareah (Algeria).

Material and Methods

In this research study, we utilized parameters widely used in the literature [23–26] we used hourly data for one year, We collected hourly data for one year (2015) from the radiometric station 'Shems,' which is part of the Centre for Renewable Energy Development (CDER) located in Bouzareah, Algiers, at a latitude of 36.8° and a longitude of 3.17°. These data were applied to predict hourly global solar radiation using both the single support vector machine (SSVM) and the bootstrapped aggregated support vector machine (BASVM) with nine different parameter configurations. Figure 1 illustrates the measurement instruments at the Bouzareah station in Algeria.

This database (DB) comprises 3603 data points and has been utilized to optimize the parameters of the bootstrapped aggregated support vector machine (BASVM). In the database, values less than 120 W/m^2 (from 5 a.m. to 5 p.m.) have been excluded, as defined by the World Meteorological Organization (WMO), which establishes sunshine duration when global solar radiation values exceed 120 W/m^2 [28]. The statistical analysis of the input and output data was performed in terms of the minimum (min), the average (mean), the maximum (max), the sum (sum), the sample variance

Fig. 1 Photo of the measuring station at CDER [27]

(Var), and the standard deviation (STD), all of which are detailed in Table 1.

Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised learning method that has become exceedingly popular for predicting meteorological data such as temperature [29], wind speed [30], and global solar radiation [7] in the past few years. Due to its simplicity and flexibility, it can handle a range of classification and regression difficulties in different fields, for example, mechanical engineering [31], energy [32], finance [33], and other fields. SVMs distinctively afford balanced predictive performance, even in studies where sample sizes may be limited.

The regression function can use the nonlinear relationship between the input and output in a support vector machine model. The output of the SVM model is obtained by the following equation [34]:

$$f\left(x_i\right) = \omega^T \varnothing\left(x_i\right) + b,\ i = 1, 2, \ldots,\ n \tag{1}$$

$f\left(x_i\right)$: the predicted data of the SVM model.

$\varnothing\left(x_i\right)$: the implicitly constructed nonlinear function that transforms input finite-dimensional space into higher-dimensional space.

ω^T: This is the weight vector, which corresponds to the coefficients associated with the feature vector in the high-dimensional feature space. It helps determine the importance of each feature in the regression process.

b: the bias of the SVM model.

i = 1,2,…,n: This indicates that the regression function is calculated for each input sample in the dataset, where n is the total number of samples.

The dataset has a D-dimensional input vector $x_i \in R^D$ and a scalar output $y_i \in R$.

The following equations provide the SVM optimization model (for the training set):

Table 1 Numerical analysis of inputs and output

		Min	Mean	Max	Std	Var	Sum
Inputs	Month	1.0000	6.4838	12.0000	3.1671	10.0305	23,361.0000
	Day	1.0000	15.7144	31.0000	8.8143	77.6910	56,619.0000
	Time (h)	5.0000	11.3369	17.0000	3.0199	9.1196	40,847.0000
	Temperature (k)	280.2600	294.6823	308.5400	6.2204	38.6928	1.0617e + 06
	Relative Humidity (%)	22.8800	58.6105	95.0000	11.4007	129.9762	2.1117e + 05
	Pressure (mbar)	972.59	996.4820	102.0190	6.2419	38.9617	3.5903e + 06
	Wind speed(m/s)	0.0800	4.2924	16.5900	2.2269	4.9589	1.5465e + 04
	Wind direction(°)	0.0200	172.0608	359.7800	118.3522	1.4007e + 04	6.1993e + 05
Output	Global solar radiation (Wh/m^2)	120.0109	515.5762	1030.7230	247.8800	6.1444e + 04	1.8576e + 06

he total number of variables, including Month, Day, Time (h), Average Temperature (K), Relative Humidity (%), Atmospheric Pressure (mbar), Wind Speed (m/s), and Wind Direction (°), was reduced using a correlation matrix. This matrix is available as Supplementary Data (Table A1)

$$\begin{cases} \min\ R(w,\xi,\xi^*,\varepsilon) = \frac{1}{2}\,\|\,w\,\|^2 + C\left[v\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)\right] \\ \text{subjective to}: y_i - w^T\varphi\left(x_i\right) - b \le \varepsilon + \xi_i \\ w^T\varphi\left(x_i\right) + b - y_i \le \varepsilon + \xi_i \\ \xi^*, \varepsilon \ge 0 \end{cases} \tag{2}$$

$\frac{1}{2}\|w\|^2$: represents the regularization term or the norm of the weight vector.

C: the factor that balances model complexity with empirical risk $\|w\|^2$

ξ_i^*: the slack variable to denote the distance of the ith sample outside of the ε-tube.

As a standard nonlinear constrained optimization problem, the above problem can be resolved by constructing the dual optimization problem based on the Lagrange multipliers techniques:

$$\begin{cases} \max\ R\left(a_i, a_i^*\right) = \sum_{i=1}^{N} y_i\left(a_i, a_i^*\right) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(a_i, a_i^*\right)\left(a_j, a_j^*\right)K(x_i, x_j) \\ \text{subjective to}: \sum_{i=1}^{N} y_i\left(a_i, a_i^*\right) = 0 \\ 0 \le a_i, a_i^* \le C/N \\ \sum_{i=1}^{N}\left(a_i + a_i^*\right) \le C.v \end{cases} \tag{3}$$

$K(x_i, x_j)$: the kernel function satisfying the Mercer's condition;

a_i and a_i^*: the nonnegative Lagrange multipliers.

$$\hat{y} = f\left(x_i\right) = \sum_{i=1}^{N}\left(a_i - a_i^*\right)K\left(x - x_i\right) + b, i = 1, 2, \ldots, n \tag{4}$$

Bootstrap Aggregated Support Vector Machine (BASVM)

One crucial strategy for enhancing the robustness and performance of prediction models involves improving a collection of prediction models, such as Support Vector Machines (SVMs), and subsequently combining them. The development of the Bootstrap Aggregated Support Vector Machine model (BASVM) entails the process of sampling the training datasets using a MATLAB function. Figure 2 provides a visual representation of the Bootstrap Aggregated SVM (BASVM), wherein multiple individual SVM models (ISVM) are created to model the same underlying relationship. This approach enables a more robust and accurate prediction through the combination of these individual models.

The process, focused on designing and optimizing the architecture of both ISVM and BASVM, is depicted in Fig. 3. It begins with resampling the training dataset using a bootstrap technique to generate a set of n different training datasets, where n takes values of 10, 15, 20, 25, and 30. Subsequently, for each of these training datasets, an ISVM model is constructed and evaluated using the testing dataset. The ISVM models that are developed are then combined by taking the average using the following equation:

$$y = \frac{\sum_{i=1}^{n} y_i}{n} \tag{5}$$

where: y_i is the output of the individual SVM "ISVM", y represent the output of the BASVM, and n is the number of ISVM models. The output of BASVM is the mean of the outputs of ISVM.

Modeling the Support Vector Machine

This research study introduces a novel approach aimed at enhancing and refining the architecture of the support vector machine. The method, as illustrated in Fig. 3, encompasses the creation of three distinct support vector machine models: the single support vector machine (SSVM), an individual support vector machine (ISVM), and a bootstrapped aggregated support vector machine (BASVM, which stacks

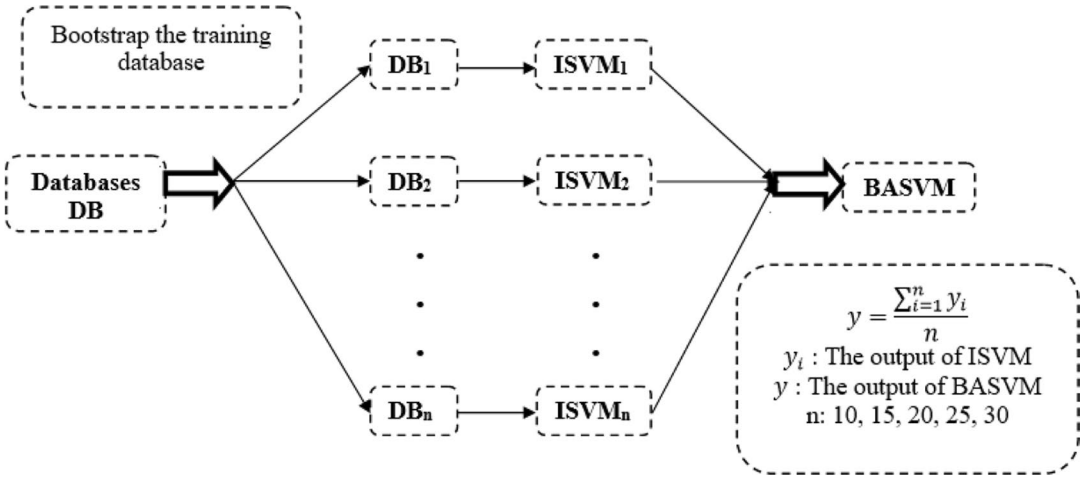


Figure 2 Bootstrap aggregated support vector machine

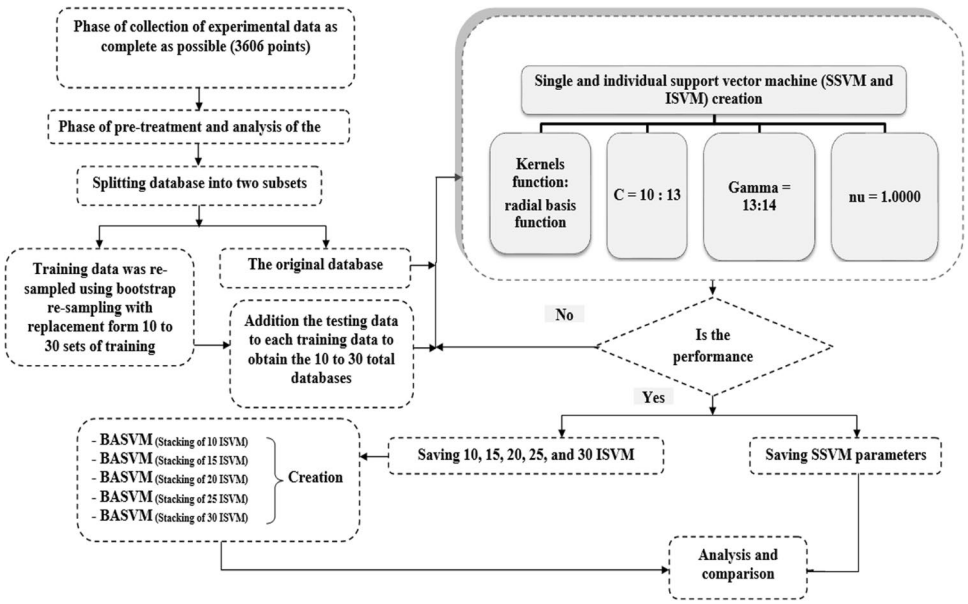


Fig. 3 Flow diagram for support vector machine development (SSVM, ISVM, and BASVM (Stacking of 10, 15, 20, 25, and 30 ISVM))

10, 15, 20, 25, and 30 ISVMs). The bootstrap technique is employed to calculate the average of the outputs from these individual support vector machines. To predict hourly global solar radiation, the study leveraged SVM modeling and executed the analysis using both MATLAB and STATISTICA software.

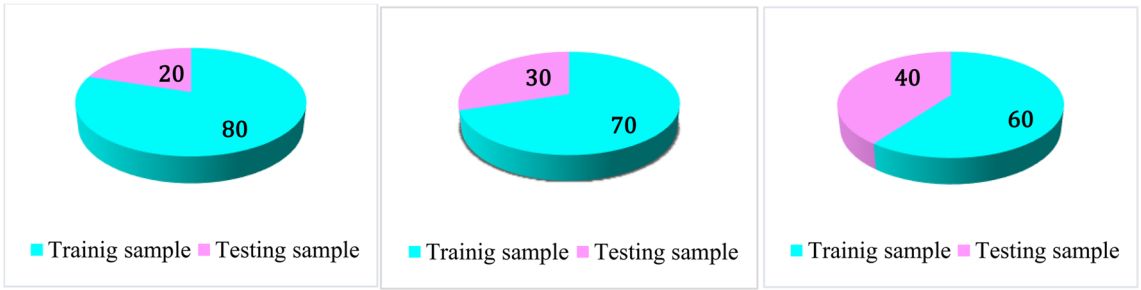
Evaluation Criteria

In our current study, we employed a variety of error measures to assess the effectiveness of our prediction models. These measures included the Correlation Coefficient (R), the Mean Absolute Error (MAE) along with its normalized counterpart (nMAE), the Model Predictive Error (MPE), the Root Mean Squared Error (RMSE) and its normalized counterpart (nRMSE), as well as the Standard Error of Prediction (SEP) [35, 36]:

$$\overline{y} = \sum_{i=1}^{N} y_{i,cal}/N \tag{6}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_{i,exp} - y_{i,cal}\right| \; ; \; nMAE = MAE/\overline{y} \tag{7}$$

$$MPE(\%) = \frac{100}{N}\sum_{i=1}^{n}\left|\frac{(y_{i,exp} - y_{i,cal}}{y_{i,exp}}\right| \tag{8}$$



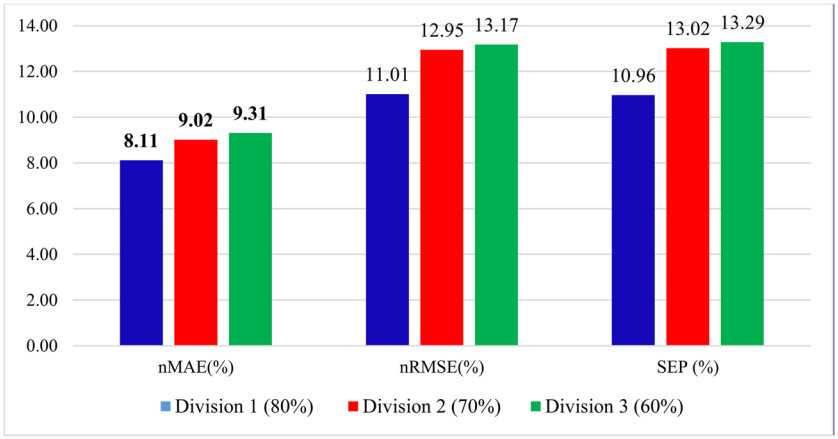
(a) (b) (c)

Fig. 4 The division of the whole database: **a** "Division 1", **b** "Division 2", **c** "Division 3"

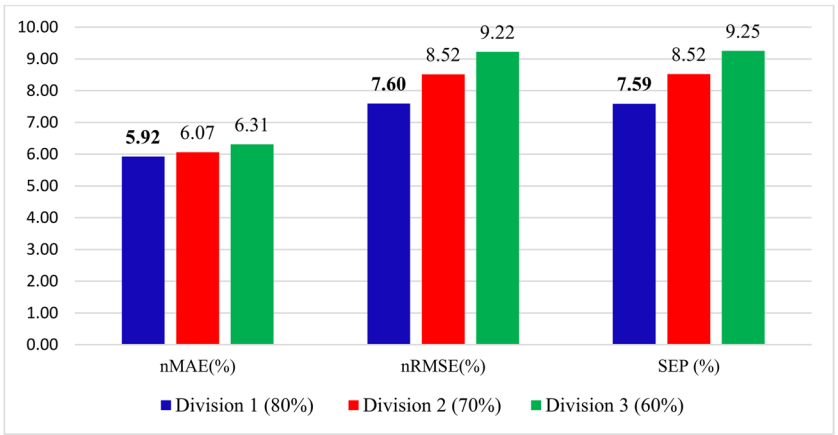
$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(y_{i,cal} - y_{i,exp}\right)^2}{N}} \; ; \; \mathrm{nRMSE} = \mathrm{RMSE}/\overline{y} \tag{9}$$

According to Despotovic et al. [37] the model accuracy is considered excellent if nRMSE < 10%, good if 10% < nRMSE < 20%, fair if 20% < nRMSE < 30% and low if nRMSE > 30%.

$$\mathrm{SEP}(\%) = \frac{\mathrm{Rmse}}{Y_{I,exp}} \times 100 \tag{10}$$



(a)



(b)

Fig. 5 Effect of the division of database: **a** test phase, **b** total phase

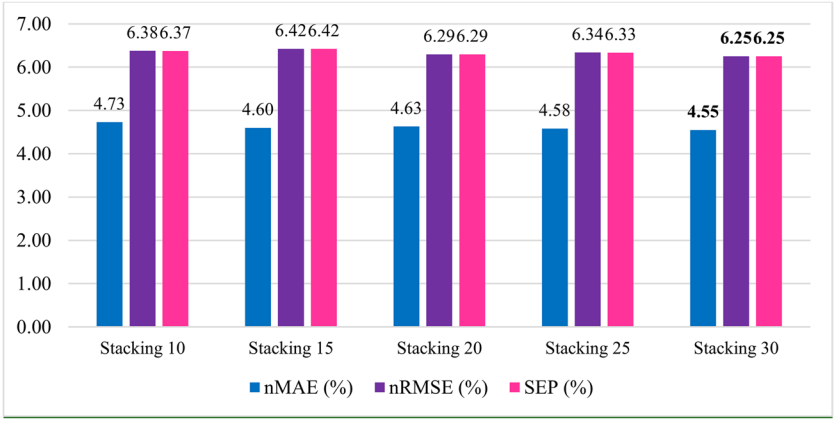


Figure 6 nMAE, nRMSE, and SEP for the different stacking support vector machine models for testing sample

where n is the total number of data; $Y_{i,exp}$ and $Y_{i,cal}$ are the experimental and calculated data point of global solar radiation, respectively.

Results and Discussion

Effect of the Division of Database

In our current study, we segregated the entire database into two distinct samples: the training dataset, which constitutes the bulk of the database, and a testing dataset, which we employed to gauge the performance of the Support Vector Machine (SVM) in practical scenarios and assess its predictive prowess. A visual representation of this database partition is illustrated in Fig. 4.

Figure 5a and b showcase the results pertaining to the relative Mean Absolute Error (nMAE), relative Root Mean Squared Error (nRMSE), and the Standard Error of Prediction (SEP) in the context of predicting hourly global solar radiation when considering different database partitioning schemes. Notably, the initial partition of the dataset, denoted as the first sample, yielded the most favorable outcomes during the testing phase. Consequently, the individual support vector machine (ISVM) was built based on this initial database partition.

Comparison between Different Stacking "BASVM" Models

In order to compare different stacking models of BASVM, five distinct stacking models were implemented: stacking 10 ISVM, stacking 15 ISVM, stacking 20 ISVM, stacking 25 ISVM, and stacking 30 ISVM. The process involved resampling the training data using bootstrap resampling with replacement [38] to create different sets of training data, resulting in 10 datasets for stacking 10 ISVM, 15 datasets for stacking 15 ISVM, 20 datasets for stacking 20 ISVM, 25 datasets for stacking 25 ISVM, and 30 datasets for stacking 30 ISVM.

For each of these stacking models, an Individual Support Vector Machine (ISVM) was created for each training dataset. Each ISVM was configured with eight parameters in the input layer and one unit responsible for generating the predicted values of global solar radiation in the output layer. The radial basis function kernel was consistently used for each model (SSVM, ISVM), while the values of C and gamma were varied within the ranges of 10 to 13 and 13 to

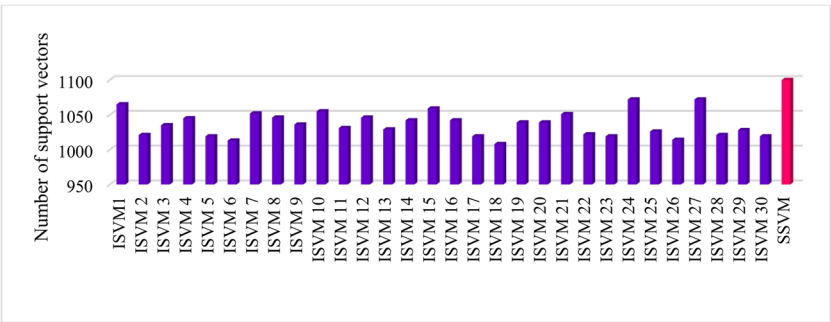


Fig. 7 Optimize the structure of ISVM and SSVM models

14, respectively, with Nu set to 1. The optimal correlation coefficient (R) for each ISVM was selected through a trial and error method.

In each stacking model, the output of the Bootstrap Aggregated Support Vector Machine (BASVM) was computed as the mean of the ISVM outputs, following Eq. 5. The performance of these developed stacking models was evaluated using key metrics such as the correlation coefficient (R), relative Mean Absolute Error (nMAE), relative Root Mean Squared Error (nRMSE), and the Standard Error of Prediction (SEP) across different phases, including training, testing, and the total dataset [50].

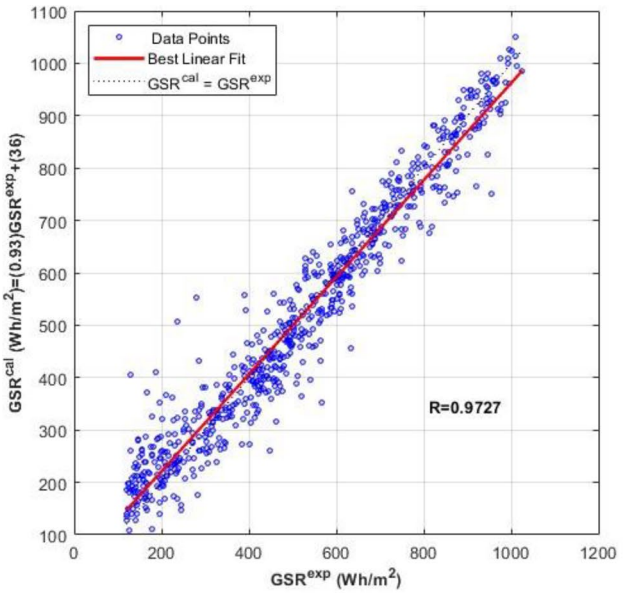
A comparison of nMAE, nRMSE, and SEP for the various BASVM stacking models is depicted in Fig. 6. It becomes evident that the BASVM stacking 30 ISVM model demonstrates superior robustness compared to other stacking models, with nMAE at 4.5487%, nRMSE at 6.2509%, and SEP at 6.2493%. Consequently, this paper places further emphasis on the BASVM (stacking 30 ISVM) model due to its remarkable performance.

Performance SVM Models

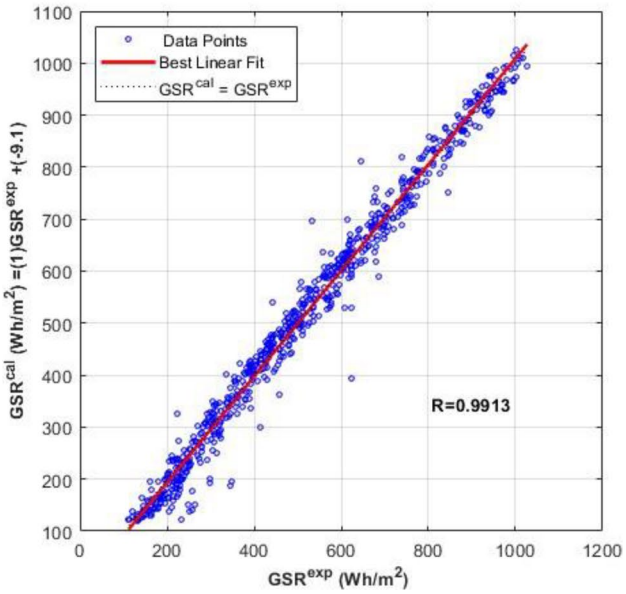
The structures of the individual support vector machine model labeled as "ISVM" and the single support vector machine model denoted as "SSVM" can be found in Fig. 7. An evident observation is that these support vector machines, "ISVM" and "SSVM," exhibit dissimilar structures and do not exhibit a harmonious relationship in their design. In particular, when comparing the thirty individual support vector machines in "ISVM" to the single support vector machine in "SSVM," it is notable that "ISVM" had a lower count of support vectors. Additionally, each individual support vector machine in "ISVM" achieved a higher correlation coefficient "R" in contrast to "SSVM".

Based on the preceding discussion, two support vector machine (SVM) models were developed, namely SSVM and BASVM (stacking 30 ISVM models), with the primary objective of predicting global solar radiation. The plots and the parameters of linear regression are distinctly discernible. In Fig. 8a and b, a comparison is presented between the experimental and calculated global solar radiation, where agreement vectors closely approach the ideal values [i.e., a = 1 (slope), b = 0 (intercept), R = 1 (regression coefficient)] during the adjustment of the support vector machine profiles.

For the SSVM model in the test phase, the parameter values were [a, b, R] = [0.9264, 35.9732, 0.9727], while for the BASVM model (stacking 30 ISVM models) during the testing phase, the parameters were [a, b, R] = [1.0173, -9.1196, 0.9913]. Notably, the slope in both SVM models is very close to 1 during the testing phase, indicating a strong correlation with the ideal value. Furthermore, the intercept (b) is in proximity to 0 for the testing phase in both SVM models, which is indicative of minimal bias in the predictions. The regression coefficients (R) fall within the generally accepted excellent range ($0.90 \leq R \leq 1.00$) for SVM models (both SSVM and BASVM with 30 networks). This



(a)



(b)

Fig. 8 Evaluating global solar radiation through experimental and calculated data: a SSVM ''Testing phase'', b BASVM (Stacking of 30 ISVM) ''Testing phase''

Table 2 Errors of ISVM, and SSVM models

	Phases	R	nMAE (%)	MPE (%)	nRMSE (%)	SEP (%)
$ISVM_1$	Training	0.9931	4.7654	6.5035	5.6983	5.6854
	Testing	0.9806	6.9267	10.9263	9.8331	9.8487
	Total	0.9904	5.2012	7.3886	6.7469	6.7368
$ISVM_2$	Training	0.9932	4.6404	6.1995	5.5737	5.5724
	Testing	0.9807	6.6643	10.2862	9.3045	9.3501
	Total	0.9906	5.0568	7.0173	6.5432	6.5486
$ISVM_3$	Training	0.9933	4.6649	6.2717	5.5424	5.5482
	Testing	0.9796	6.8791	10.2169	9.4757	9.5658
	Total	0.9905	5.1164	7.0612	6.5571	6.5752
$ISVM_4$	Training	0.9933	4.6141	6.1275	5.5363	5.5268
	Testing	0.9796	6.7184	10.1131	9.4341	9.5141
	Total	0.9905	5.0438	6.9251	6.5431	6.5454
$ISVM_5$	Training	0.9926	4.8869	6.6201	5.8414	5.8324
	Testing	0.9822	7.0191	10.9305	9.3293	9.4221
	Total	0.9904	5.3196	7.4827	6.7093	6.7144
$ISVM_6$	Training	0.9923	4.9180	6.5730	5.8980	5.8963
	Testing	0.9766	7.1615	11.1847	10.0572	10.1109
	Total	0.9892	5.3705	7.4959	6.9485	6.9543
$ISVM_7$	Training	0.9934	4.6053	6.2856	5.4700	5.4788
	Testing	0.9820	6.8777	10.5707	9.3041	9.3900
	Total	0.9910	5.0578	7.1431	6.4144	6.4343
$ISVM_8$	Training	0.9927	4.9153	6.8204	5.8015	5.7861
	Testing	0.9794	6.8947	10.1820	9.6698	9.7341
	Total	0.9901	5.3117	7.4931	6.7567	6.7513
$ISVM_9$	Training	24.2544	4.6980	6.4852	5.6087	5.6171
	Testing	0.9812	6.8174	10.1184	9.2824	9.3555
	Total	0.9908	5.1260	7.2122	6.5272	6.5453
$ISVM_{10}$	Training	0.9931	4.7874	6.5926	5.6690	5.6693
	Testing	0.9755	7.2996	11.3561	10.5330	10.5806
	Total	0.9896	5.2942	7.5459	6.9400	6.9466
$ISVM_{11}$	Training	0.9932	4.7305	6.5414	5.6588	5.6640
	Testing	0.9777	7.2785	11.1248	10.3682	10.4114
	Total	0.9900	5.2385	7.4586	6.8569	6.8676
$ISVM_{12}$	Training	0.9937	4.5021	6.1812	5.3555	5.3616
	Testing	0.9811	6.7021	10.1289	9.3771	9.4318
	Total	0.9911	4.9443	6.9712	6.3750	6.3883
$ISVM_{13}$	Training	0.9927	4.8811	6.5966	5.8059	5.7887
	Testing	0.9803	7.0024	10.5813	9.5373	9.5516
	Total	0.9901	5.3087	7.3939	6.7329	6.7190
$ISVM_{14}$	Training	0.9924	5.0023	5.9276	5.9208	5.9276
	Testing	0.9808	7.0129	10.8850	9.5139	9.6065
	Total	0.9901	5.4028	7.6999	6.7861	6.8055
$ISVM_{15}$	Training	0.9930	4.7916	6.6742	5.7322	5.7360
	Testing	0.9811	6.9812	10.9010	9.4025	9.4952
	Total	0.9905	5.2380	7.5201	6.6632	6.6801
$ISVM_{16}$	Training	0.9930	4.8015	6.4883	5.7250	5.7127
	Testing	0.9774	7.5494	10.9757	10.5271	10.4988
	Total	0.9898	5.3473	7.3863	6.9401	6.9245
$ISVM_{17}$	Training	0.9925	4.8898	6.5936	5.8179	5.8198
	Testing	0.9803	7.0289	11.0913	9.5016	9.5839
	Total	0.9900	5.3216	7.4936	6.7335	6.7469

Table 2 (continued)

	Phases	R	nMAE (%)	MPE (%)	nRMSE (%)	SEP (%)
$ISVM_{18}$	Training	0.9929	4.7839	6.5666	5.7197	5.7184
	Testing	0.9811	6.9604	10.8713	9.5591	9.6101
	Total	0.9905	5.2167	7.4280	6.6561	6.6619
$ISVM_{19}$	Training	0.9932	4.7334	6.5626	5.6524	5.6532
	Testing	0.9796	6.6982	9.9275	9.5491	9.5882
	Total	0.9906	5.1281	7.2359	6.6253	6.6315
$ISVM_{20}$	Training	0.9927	4.7798	6.5142	5.6766	5.6766
	Testing	0.9823	6.6641	9.6245	9.1624	9.1403
	Total	0.9905	5.1537	7.1366	6.5116	6.5085
$ISVM_{21}$	Training	0.9934	4.6677	6.3626	5.5816	5.5796
	Testing	0.9796	6.9427	10.3063	9.6632	9.6662
	Total	0.9907	5.1264	7.1518	6.6177	6.6163
$ISVM_{22}$	Training	0.9936	4.5688	6.3240	5.4859	5.4794
	Testing	0.9789	6.7458	10.1733	9.6794	9.6884
	Total	0.9907	5.0179	7.0943	6.6049	6.6000
$ISVM_{23}$	Training	0.9922	5.0677	6.9436	6.0934	6.0896
	Testing	0.9811	6.9684	10.6575	9.6216	9.6873
	Total	0.9899	5.4480	7.6868	6.9443	6.9502
$ISVM_{24}$	Training	0.9930	4.7834	6.4804	5.7430	5.7392
	Testing	0.9767	7.3088	11.2501	10.6020	10.6313
	Total	0.9896	5.2920	7.4348	7.0058	7.0060
$ISVM_{25}$	Training	0.9939	4.4100	5.8450	5.2693	5.2685
	Testing	0.9785	7.0501	10.8734	10.1471	10.2103
	Total	0.9906	4.9447	6.8512	6.5711	6.5786
$ISVM_{26}$	Training	0.9933	4.5847	6.2938	5.4534	5.4509
	Testing	0.9795	7.1209	11.1833	9.9258	9.9827
	Total	0.9903	5.0954	7.2722	6.6091	6.6143
$ISVM_{27}$	Training	0.9926	4.8508	6.6048	5.7877	5.7924
	Testing	0.9787	7.3291	11.5283	10.1224	10.1709
	Total	0.9897	5.3419	7.5900	6.8580	6.8690
$ISVM_{28}$	Training	0.9931	4.7525	6.4114	5.6037	5.6071
	Testing	0.9746	7.4873	11.4298	10.6745	10.7876
	Total	0.9894	5.3086	7.4156	6.9612	6.9795
$ISVM_{29}$	Training	0.9930	4.8087	6.5069	5.6871	5.6785
	Testing	0.9759	7.3125	11.0113	10.7678	10.8097
	Total	0.9895	5.3040	7.4083	6.9779	6.9748
$ISVM_{30}$	Training	0.9929	4.6767	6.2786	5.6110	5.6169
	Testing	0.9776	7.4268	11.6311	10.3973	10.5047
	Total	0.9897	5.2205	7.3497	6.8158	6.8354
SSVM	Training	0.9911	5.3752	7.3309	6.4595	6.4593
	Testing	0.9727	8.1104	12.1255	11.0112	10.9646
	Total	0.9875	5.9235	8.2904	7.5957	7.5891
BASVM	Testing	0.9913	4.5487	5.8823	6.2509	6.2493

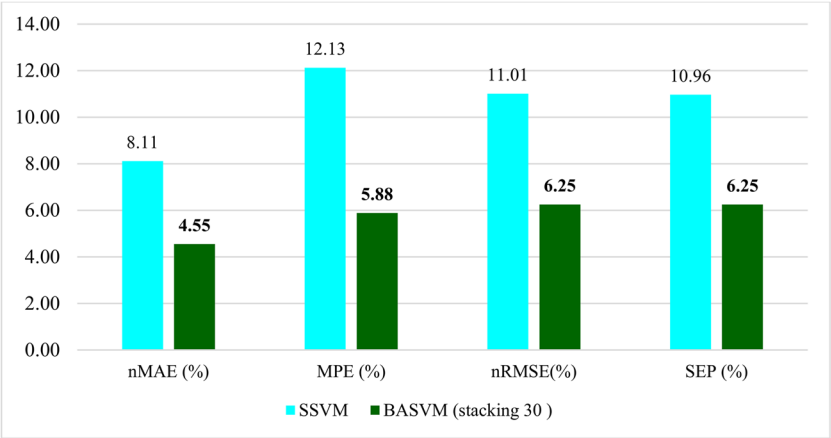


Figure 9 nMAE, MPE, nRMSE, and SPE of bootstrap aggregated support vector machine "BASVM" $_{(\text{stacking 30 ISVM})}$ and signal support vector machine "SSVM" for testing test set

attests to the robustness of the established SVM models and their capability to reliably predict global solar radiation.

Comparison Between ISVM, BASVM, and SSVM

Table 2 provides a comprehensive overview of the performance metrics for thirty individual support vector machine (ISVM) models, the bootstrap aggregated support vector machine (BASVM), and the single support vector machine (SSVM) across different datasets, including training data, testing data, and the combined total datasets for ISVM and SSVM. Specifically, the metrics considered for comparison include the relative Mean Absolute Error (nMAE), Model Predictive Error (MPE), relative Root Mean Squared Error (nRMSE), and the Standard Error of Prediction (SEP).

This comparative analysis serves to underscore that the BASVM models represent a credible alternative to the SSVM models. It is worth noting that the performance of these support vector machines (SSVM, ISVM, and BASVM) can vary across the training, testing, and total datasets. In some cases, a support vector machine that exhibits minimal errors in the training dataset might exhibit more substantial errors when applied to the test dataset.

For instance, ISVM 20 demonstrates a lower relative Root Mean Squared Error (nRMSE) of 5.6766% on the training set, 9.1624% on the testing set, and 6.5116% on the combined total datasets. Notably, the BASVM outperforms in terms of nRMSE, achieving a lower value of 6.2509% on the testing set. This highlights a significant enhancement in accuracy achieved by the collaborative approach of combining multiple models within the BASVM.

Figure 9 illustrates a comparison a comparison between the BASVM (Stacking 30 ISVM) and SSVM models. This comparison of testing outcomes between the BASVM (Stacking 30 ISVM) and SSVM models clearly underscores the advantages of the bootstrap aggregated support vector machine model over the Single Support Vector Machine model (SSVM). It underscores the enhanced performance of the BASVM model in terms of precision, illustrating its superior capability to provide more accurate predictions of global solar radiation when compared to the single support vector machine model.

Table 3 Overview of various models for predicting global solar radiation

Reference	Location	Type of model	Evaluation Index
This study	Algeria	Bootstrap aggregated support vector machine BASVM	R = 0.9913
Shamshirband et al. [39]	Iran	Support vector machine –wavelet transform model (SVM-WT)	R = 0.9631
Linares-Rodríguez et al.[40]	Spain	Multilayered perception (ANN-MLP)	R = 0.9400
Dahmani et al. [11]	Algeria	Bootstrap aggregated neural networks (BANN)	R = 0.9680
Yao et al. [7]	China	Support vector machine(SVM)	R = 0.8760
Guermoui et al. [41]	Algeria	Support vector machine(SVM)	R = 93.06
Mehdi Lotfinejad et al. [42]	Iran	Bat neural network (BNN)	R = 0.9810
Fadare [43]	Nigeria	Multilayered perception, feedforward, back-propagation (MLP-FBP)	R = 0.9560

Comparison with Other Models

To assess the significance of our findings, we conducted a comparative analysis with other studies carried out by different researchers, with a particular focus on models that used similar inputs to ours. These models were all aimed at predicting solar radiation. The results of this assessment provide strong evidence of the effectiveness and accuracy of the developed $BASVM_{(Stacking\ of\ 30\ ISVM)}$ model for predicting global solar radiation. Table 3 displays the outcomes obtained from these aforementioned models alongside the results from our study.

Conclusion

The primary objective of this current research study is to enhance the predictive capabilities of two robust support vector machine models, namely SSVM and BASVM (Stacking of 30 ISVM), by leveraging an accessible structure–activity relationship. These models are designed for the precise prediction of hourly global solar radiation. The comparative analysis between SSVM and BASVM (Stacking of 30 ISVM) serves to highlight the robustness, reliability, and effectiveness of support vector machine models when applied to meteorological input parameters, including variables such as month, day, time, average temperature, relative humidity, atmospheric pressure, wind speed, and wind direction.

The results of this study exhibit a noteworthy performance difference between the two models. During the testing phase, BASVM (Stacking of 30 ISVM) achieves remarkable consistency between the calculated and experimental data, boasting a relative Root Mean Squared Error (nRMSE) of 6.2509%. In contrast, SSVM records an nRMSE of 11.0112%. This novel model, BASVM (Stacking of 30 ISVM), proves to be a valuable tool for predicting solar radiation, especially in locations without access to measurement equipment such as solarimeters or pyranometers and associated systems. It particularly shines when dealing with scenarios marked by limited available data, the presence of outliers necessitating exclusion, or instances of missing data.

Additionally, this model can play a pivotal role in supporting the installation of solar-energy systems and evaluating thermal conditions in building studies, particularly in regions like Algeria or those with similar climatic characteristics. Its ability to deliver accurate solar radiation predictions makes it a valuable asset for both energy planning and building design in such areas.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40866-023-00179-w.

Acknowledgements We would also like to extend our thanks to the centre for development of renewable energies in Bouzareah, Algeria.

Funding None.

Data Availability The data used in the study is owned by the "Centre de Développement des Énergies Renouvelables (CDER)" and cannot be provided for auditing or scientific use by others.

References

1. Gielen D, Boshell F, Saygin D, Bazilian MD, Wagner N, Gorini R (2019) The role of renewable energy in the global energy transformation. Energy Strateg Rev 24:38–50
2. Kurt E, Demirci M, Şahin HM (2022) Numerical analyses of the concentrated solar receiver pipes with superheated steam. Proc Inst Mech Eng Part A J Power Energy 236:893–910
3. Joshi U, Shrestha PM, Maharjan S, Bhattarai A, Bhattarai N, Chapagain NP et al (2022) Estimation of Solar Insolation and Angstrom-Prescott Coefficients Using Sunshine Hours over Nepal. Adv Meteorol 2022:15. https://doi.org/10.1155/2022/3593922
4. Stambouli AB, Khiat Z, Flazi S, Kitamura Y (2012) A review on the renewable energy development in Algeria: Current perspective, energy scenario and sustainability issues. Renew Sustain Energy Rev 16:4445–4460
5. Laidi M, Hanini S, Rezrazi A, Yaiche MR, El Hadj AA, Chellali F (2017) Supervised artificial neural network-based method for conversion of solar radiation data (case study: Algeria). Theor Appl Climatol 128:439–451. https://doi.org/10.1007/s00704-015-1720-7
6. Makade RG, Chakrabarti S, Jamil B (2021) Development of global solar radiation models: A comprehensive review and statistical analysis for Indian regions. J Clean Prod 293:126208
7. Yao W, Zhang C, Hao H, Wang X, Li X (2018) A support vector machine approach to estimate global solar radiation with the influence of fog and haze. Renew Energy 128:155–162. https://doi.org/10.1016/j.renene.2018.05.069
8. Guermoui M, Gairaa K, Boland J, Arrif T (2021) A novel hybrid model for solar radiation forecasting using support vector machine and bee colony optimization algorithm: review and case study. J Sol Energy Eng 143:20801
9. Hansen LK, Salamon P (1990) Neural network ensembles. IEEE Trans Pattern Anal Mach Intell 12:993–1001
10. Parasuraman K, Elshorbagy A, Si BC (2006) Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles. Soil Sci Soc Am J 70:1851–1859
11. Dahmani A, Ammi Y, Hanini S, RedhaYaiche M, Zentou H (2023) Prediction of hourly global solar radiation: comparison of neural networks/bootstrap aggregating. Kem u Ind Časopis kemičara i Kem inženjera Hrvat 72:201–213
12. Rezrazi A, Hanini S, Laidi M (2015) An optimisation methodology of artificial neural network models for predicting solar radiation: A case study. Theor Appl Climatol 123:769–783. https://doi.org/10.1007/s00704-015-1398-x
13. Vapnik VN, Lerner AY (1963) Recognition of patterns with help of generalized portraits. Avtomat i Telemekh 24:774–780
14. Shah K, Patel H, Sanghvi D, Shah M (2020) A comparative analysis of logistic regression, random forest and KNN models for the text classification. Augment Hum Res 5. https://doi.org/10.1007/s41133-020-00032-0
15. Ahir K, Govani K, Gajera R, Shah M (2019) Application on virtual reality for enhanced education learning, military training and sports. Augment Hum Res 5. https://doi.org/10.1007/s41133-019-0025-2

16. Quej VH, Almorox J, Arnaldo JA, Saito L (2017) ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J Atmos Solar-Terrestrial Phys 155:62–70
17. dos Santos CM, Escobedo JF, Teramoto ET, da Silva SHMG (2016) Assessment of ANN and SVM models for estimating normal direct irradiation (Hb). Energy Convers Manag 126:826–836
18. Lima M, Carvalho PCM, Braga A, Ramírez LM, Leite JR (2018) MLP back propagation artificial neural network for solar resource forecasting in equatorial areas. Renew Energy Power Qual J 1:175–180
19. Takilalte A, Harrouni S, Yaiche MR, Mora-López L (2020) New approach to estimate 5-min global solar irradiation data on tilted planes from horizontal measurement. Renew Energy 145:2477–2488. https://doi.org/10.1016/j.renene.2019.07.165
20. Gao B, Huang X, Shi J et al (2020) Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. Renew Energy 162:1665–1683
21. Peng T, Zhang C, Zhou J, Nazir MS (2021) An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting. Energy 221:119887
22. Keshtegar B, Mert C, Kisi O (2018) Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree. Renew Sustain Energy Rev 81:330–341
23. Azadeh A, Maghsoudi A, Sohrabkhani S (2009) An integrated artificial neural networks approach for predicting global radiation. Energy Convers Manag 50:1497–1505. https://doi.org/10.1016/j.enconman.2009.02.019
24. Siham CM, Salah H, Maamar L, Latifa K (2017) Artificial neural networks based prediction of hourly horizontal solar radiation data: Case study. Int J Appl Decis Sci 10:156–174. https://doi.org/10.1504/IJADS.2017.084312
25. Rezrazi A, Hanini S, Laidi M (2016) An optimisation methodology of artificial neural network models for predicting solar radiation: a case study. Theor Appl Climatol 123:769–783. https://doi.org/10.1007/s00704-015-1398-x
26. Pang Z, Niu F, O'Neill Z (2020) Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. Renew Energy 156:279–289. https://doi.org/10.1016/j.renene.2020.04.042
27. Guermoui M, Bouchouicha K, Benkaciali S, Gairaa K, Bailek N (2022) New soft computing model for multi-hours forecasting of global solar radiation. Eur Phys J Plus 137(1):162. https://doi.org/10.1140/epjp/s13360-021-02263-5
28. Kalogirou SA, Mathioulakis E, Belessiotis V (2014) Artificial neural networks for the performance prediction of large solar systems. Renew Energy 63:90–97
29. Radhika Y, Shashi M (2009) Atmospheric Temperature Prediction using Support Vector Machines. Int J Comput Theory Eng 55–58. https://doi.org/10.7763/ijcte.2009.v1.9
30. Mohandes MA, Halawani TO, Rehman S, Hussain AA (2004) Support vector machines for wind speed prediction. Renew Energy 29:939–947. https://doi.org/10.1016/j.renene.2003.11.009
31. Du M, Zhao Y, Liu C, Zhu Z (2021) Lifecycle cost forecast of 110 kV power transformers based on support vector regression and gray wolf optimization. Alex Eng J 60:5393–5399
32. Hu G, Xu Z, Wang G, Zeng B, Liu Y, Lei Y (2021) Forecasting energy consumption of long-distance oil products pipeline based on improved fruit fly optimization algorithm and support vector regression. Energy 224:120153
33. Liu M, Luo K, Zhang J, Chen S (2021) A stock selection algorithm hybridizing grey wolf optimizer and support vector regression. Expert Syst Appl 179:115078
34. García-Alba J, Bárcena JF, Ugarteburu C, García A (2019) Artificial neural networks as emulators of process-based models to analyse bathing water quality in estuaries. Water Res 150:283–295. https://doi.org/10.1016/j.watres.2018.11.063
35. Bailek N, Bouchouicha K, Al-Mostafa Z, El-Shimy M, Aoun N, Slimani A et al (2018) A new empirical model for forecasting the diffuse solar radiation over Sahara in the Algerian Big South. Renew Energy 117:530–537. https://doi.org/10.1016/j.renene.2017.10.081
36. Ammi Y, Khaouane L, Hanini S (2021) Stacked neural networks for predicting the membranes performance by treating the pharmaceutical active compounds. Neural Comput Appl 33:12429–12444. https://doi.org/10.1007/s00521-021-05876-0
37. Despotovic M, Nedic V, Despotovic D, Cvetanovic S (2015) Review and statistical analysis of different global solar radiation sunshine models. Renew Sustain Energy Rev 52:1869–1880
38. Tibshirani RJ, Efron B (1993) An introduction to the bootstrap. Monogr Stat Appl Probab 57:1–436
39. Shamshirband S, Mohammadi K, Khorasanizadeh H, Yee L, Lee M, Petković D et al (2016) Estimating the diffuse solar radiation using a coupled support vector machine–wavelet transform model. Renew Sustain Energy Rev 56:428–435. https://doi.org/10.1016/j.rser.2015.11.055
40. Linares-Rodriguez A, Ruiz-Arias JA, Pozo-Vazquez D, Tovar-Pescador J (2013) An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. Energy 61:636–645
41. Guermoui M, Rabehi A (2020) Soft computing for solar radiation potential assessment in Algeria. Int J Ambient Energy 41:1524–1533
42. Lotfinejad MM, Hafezi R, Khanali M, Hosseini SS, Mehrpooya M, Shamshirband S (2018) A comparative assessment of predicting daily solar radiation using Bat Neural Network (BNN), Generalized Regression Neural Network (GRNN), and Neuro-Fuzzy (NF) system: A case study. Energies 11:1–15. https://doi.org/10.3390/en11051188
43. Fadare DA (2009) Modelling of solar energy potential in Nigeria using an artificial neural network model. Appl Energy 86:1410–1422. https://doi.org/10.1016/j.apenergy.2008.12.005