










Estimating of Solar Radiation Based on Machine Learning Approaches under Algerian Desert Climate

Djelloul Benatallah^{1*}, Ayoub Zahouani¹, Abdelmalek Yahiaoui¹, Halima Djeldji², Bahous Nasri²,
Ali Benatallah², Bouchra Benabdelkrim²

¹ Department of Mathematics and Computer Science, LSDCS Laboratory, Ahmed DRAIA University Adrar, Adrar 01000, Algeria

² Department of Material Sciences, Ahmed DRAIA University Adrar, Adrar 01000, Algeria

Corresponding Author Email: dje.benatallah@univ-adrar.edu.dz

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/i2m.230504>

ABSTRACT

Received: 16 July 2024

Revised: 23 September 2024

Accepted: 29 September 2024

Available online: 25 October 2024

Keywords:

estimation, machine-learning, linear regression, Random Forests, solar radiation, XGBoost

This study focuses on developing machine learning models to accurately estimate global solar radiation in southern Algeria, specifically in Illizi, Tamanrasset, and Timimoun. Traditional methods for measuring solar radiation are costly, hindering effective assessment and utilization of solar energy potential. By employing various techniques such as linear regression, XGBoost, Random Forest, and a hybrid model, the study aims to enhance the sustainability and efficiency of solar energy utilization. Results indicate that Random Forest performed exceptionally well in Illizi, achieving an R value of 0.9255. In Timimoun, the hybrid model combining Random Forest and XGBoost demonstrated outstanding performance, yielding an R value of 0.9640. In Tamanrasset, the same hybrid model showed good performance with an R value of 0.8868. These findings underscore the efficacy of machine learning approaches in accurately estimating solar radiation; thus, facilitating better utilization of solar energy resources in the region; consequently, it is advised to utilize the models in south Algeria as well as any other place with comparable climatic conditions.

1. INTRODUCTION

The depletion of conventional energy resources such as fossil fuels, oil, and gas, coupled with increasing pollution and the effects of climate change, underscores the urgent need for energy diversification. Integrating renewable energies into our energy mix offers a sustainable solution to these pressing challenges. By harnessing renewable sources such as solar, wind, hydro, and geothermal power, we can reduce our reliance on finite fossil fuels, mitigate environmental degradation, and combat climate change [1]. Embracing renewable energy technologies not only fosters energy security and independence but also promotes economic growth, job creation, and innovation in the clean energy sector. It's imperative that we accelerate the transition towards renewable energies to build a more sustainable and resilient energy future for generations to come [2]. Any country's ability to prosper economically and socially depends heavily on its energy supply. Solar radiation estimation is a crucial aspect in various fields, including environmental science, agriculture, renewable energy, and climate studies [3]. Accurate measurement and estimation of solar radiation are vital for understanding and predicting weather patterns, optimizing the performance of solar energy systems, and managing agricultural practices effectively [4]. By estimating solar radiation, scientists and engineers can design and deploy solar panels in locations that maximize energy production, thereby

enhancing the efficiency and sustainability of solar power generation. Additionally, farmers can use solar radiation data to make informed decisions about crop planting and irrigation, ultimately leading to improved agricultural yields [5]. Furthermore, in climate research, understanding solar radiation helps in modeling and predicting climate changes, which is essential for developing strategies to mitigate the impacts of global warming [6]. Therefore, the importance of solar radiation estimation cannot be overstated, as it plays a fundamental role in advancing technology, improving agricultural productivity, and addressing environmental challenges [7].

Artificial intelligence (AI) and machine learning (ML) are revolutionizing the estimation of solar radiation, enhancing accuracy and efficiency [8]. AI refers to the development of computer systems that can perform tasks typically requiring human intelligence, while ML, a subset of AI, involves training algorithms to recognize patterns and make predictions based on data [9]. In the context of solar radiation estimation, ML models analyze vast amounts of meteorological and environmental data to predict solar energy potential with high precision [10]. These advanced techniques help optimize solar energy systems, improve energy forecasts, and support the integration of renewable energy into the grid, ultimately contributing to more reliable and sustainable energy solutions [11].

In the field of solar radiation prediction and photovoltaic

energy production, various types of models exist, including physical models relying on mathematical equations and AI-based models [12]. Machine learning is particularly significant in modeling solar radiation, using statistical techniques to identify patterns and relationships within data. Several important studies have been conducted, contributing to the advancement of this field [5]. For instance, Van der Meer et al. [13] conducted a comprehensive review on probabilistic forecasting of photovoltaic energy production and electricity consumption, emphasizing the importance of using probabilistic models to improve forecast accuracy. On the other hand, Lorenz et al. [14] focused on solar radiation prediction for forecasting energy production from grid connected photovoltaic systems, demonstrating that meteorological models can significantly enhance short-term forecast accuracy. Moreover, Alessandrini et al. [15] utilized a set of short-term models for probabilistic solar energy forecasting, proving that an integrated approach can reduce uncertainty in forecasts and provide more accurate predictions. Antonanzas et al. [7] provided a review of photovoltaic energy production forecasts, highlighting the importance of historical data analysis and using machine learning models to improve future forecast accuracy. Additionally, Sagade et al. [16] introduced an ARIMA model for assessing various photovoltaic energy techniques' performance and found that using these models can improve the models' ability to predict photovoltaic system performance under changing climatic conditions [7]. This study aimed to enhance the estimation of solar radiation using three models: ANN, GA-ANN, and ANFIS in southern Algeria. The findings revealed that GA-ANN and ANFIS were more accurate than ANN in predictions, making them useful for determining the size and installation of solar energy systems [9]. The study explored the efficiency of solar energy systems by analyzing solar radiation availability. Solar radiation measurements are sparse across Africa, making accurate sizing and optimization of solar energy projects challenging. To address this, the study aimed to predict global hourly solar irradiation using a neural model based on solar geometry and astronomical data from a city in North Africa. Nine models and three activation functions were tested using data collected by the Saharan Renewable Energy Research Unit from 2013 to 2018. Eighty percent of the data was used for training, and the remainder for validation. Various combinations of input data were explored, yielding different levels of precision. The logistic Sigmoid function with 15 neurons in the hidden layer achieved a correlation coefficient of 98.25% between measured and estimated global solar irradiation. This model is recommended for estimating solar radiation intensities at the study site and other locations with similar climatic conditions [17]. This study developed an ANN model to predict daily solar irradiance in Bechar. Among four models tested, Model1, using MLF with the Levenberg-Marquardt back-propagation algorithm, showed the best accuracy ($R = 0.9198$, $MAPE = 7.57$). This model is valuable for solar system design in Algeria [18]. The previous study developed four models to estimate daily solar radiation in Reggane, Algeria, using artificial neural networks. These models utilized meteorological and astronomical data over six years and indicated that Model 4 performed the best, with a correlation coefficient (R) exceeding 0.91 and a mean absolute error below 7%. This suggests its suitability for designing solar energy systems in desert regions [19]. The hybrid solar radiation forecast model proposed in this study combines an Artificial Neural Network (ANN) model with a Numerical

Weather forecast (NWP) simulation. Using a six-hourly interval $1^\circ \times 1^\circ$ NCEP FNL analysis data, meteorological and solar radiation parameters were simulated for the entire year of 2016 using the Weather Research and Forecasting (WRF) model. The annual results of the WRF model were fed into the ANN model to estimate solar radiation and temperature for a solar farm in Algeria's Adrar Province. The outcomes showed that the ANN model significantly increased the solar and temperature predictions' accuracy. Evaluated using RMSE, MAE, and correlation metrics, the hybrid WRF-ANN model outperformed the WRF simulation alone in hourly solar radiation predictions [20]. The study aimed to forecast daily global irradiation on a horizontal plane using the empirical Campbell model and compare it with measured results. Results indicated a high accuracy level, with a mean absolute percentage error (MAPE) of less than 7%, mean bias error below 3% in absolute value, relative root mean square error (RMSE) under 7%, and a correlation coefficient exceeding 0.99 for annual global radiation. These findings suggested that the Campbell model could effectively predict global solar radiation for this site and other locations with similar climatic conditions [21]. Hybrid models combining ANFIS and ANN, enhanced with GA and PSO, were developed to accurately forecast daily solar radiation. These models demonstrated superior performance over traditional models like ARIMA, SVR, and Random Forest Regression, reducing the root mean square error by over 10% to less than 6%. Cross-validation confirmed their reliability and ability to predict daily solar radiation accurately, highlighting their potential in advancing solar energy as a sustainable and reliable source [22]. Overall, previous studies consistently highlight the superiority of machine learning methods over traditional approaches in estimating solar radiation, particularly in regions with complex climates. Machine learning models, due to their flexibility and ability to model non-linear relationships, can handle diverse datasets and provide more accurate and robust estimates. However, the success of machine learning methods depends on the availability of high-quality data, model selection, and proper tuning, making them more resource-intensive compared to traditional empirical models. Hybrid approaches combining both traditional and ML methods offer a promising avenue for further improving GSR estimation accuracy.

Integrating machine learning-based solar radiation estimation techniques into standardized measurement practices and calibration procedures offers a promising advancement in metrology. These models, trained on vast datasets, can enhance the accuracy and precision of solar radiation instruments by compensating for environmental factors that traditional methods may overlook, such as cloud cover, atmospheric composition, and temperature fluctuations, particularly in extreme climates like the Algerian desert. By providing continuous, data-driven predictions, machine learning approaches can be employed for real-time calibration and performance validation of solar sensors, ensuring more reliable and standardized solar radiation measurements across diverse geographic regions. This integration supports the refinement of measurement standards, improving consistency and accuracy in renewable energy assessments and climate studies.

This study aims to develop machine learning models to accurately estimate global solar radiation in southern Algeria, with a particular focus on the regions of Illizi, Tamanrasset and Timimoun, and to design an application for accurate

predictions based on regional characteristics. In our project, we used machine learning techniques, including linear regression, Random Forests and XGBoost, along with a hybrid model. We also developed an application called ZAHYA_SOLAR using these models to easily and efficiently estimate solar radiation levels.

2. MATERIAL AND MODEL

2.1 Study areas and data collection

The research regions were determined by taking into account the meteorological features of the provinces of Illizi, Tamanrasset, and Timimoun. Table 1 shows that the provinces that were chosen are in the southern region of Algeria. They have a great potential for solar power since they get an average of 2263 kWh/m² of solar energy annually, with a maximum of 3900 hours of sunshine per year [23]. The locations of the suggested sites are shown on an Algerian map in Figure 1.

Table 1. The coordinates of the sites

Sites	Illizi	Timimoun	Tamanrasset
Latitude (°)	26.50	30.02	24.37
Longitude (°)	8.47	0.84	4.32
Altitude (m)	568	431	810

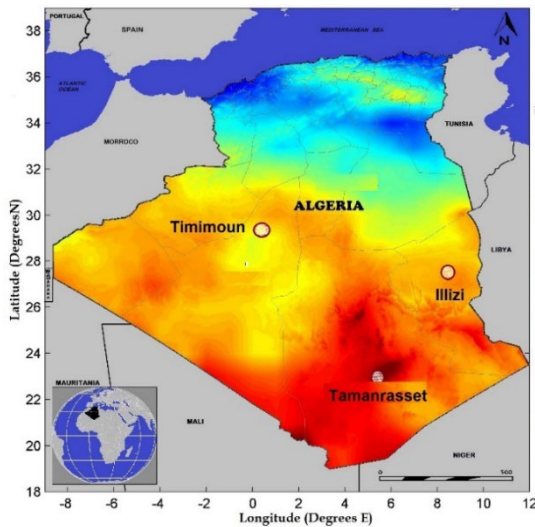


Figure 1. Algeria's solar potential and the locations of three sites

The geographical location of the selected sites is given in Table 1.

Data of global solar radiation were collected in specific regions of Algeria, specifically in Illizi and Tamanrasset, from 2015 to 2020, with 2192 each, and in Timimoun from 2016 to 2020, with 1827 obtained from the Nasa [24]. Global collected of solar radiation are available daily, enabling us to estimate medium solar energy every day, and 10% of the data was used for testing and the rest was set to the training stage. In order to predict the output of daily global solar radiation, we worked with the meteorological and astronomical parameters (see Table 2) and used the following as inputs: average temperature, wind speed, relative humidity, and atmospheric pressure. We then calculated the parameters using Eqs. (1)-(2) for extraterrestrial solar irradiation and sunshine duration of the day, respectively.

$$E_x = \frac{24}{\pi} g_0 \left(\cos(\theta) \cos(\delta) \sin(w) + \sin(\theta) \sin(\delta) \frac{\pi}{180} w \right) \quad (1)$$

$$S_d = \left(\frac{2}{15} \right) w \quad (2)$$

Using θ (°) for the city's latitude, δ (°) for the sun's declination, w (deg.) for the angle of solar height (°), and g_0 (dimensionless) for the coefficient of solar radiation from extraterrestrial sources.

Table 2. The parameters used

Parameters	Unit	Category
Year (Y)	2015-2020	
Day of the year (Jday)	1-365/366	
Temperature (T_{avg})	Degree	Meteorological
Relative Humidity (Rh)	%	
Pressure (Pr)	hPa	
Wind speed (Ws)	m/s	
Extraterrestrial solar irradiation (E_x)	Wh/m ²	
Sunshine duration (Sd)	h	

In this study, basic quality control methods were implemented to enhance the accuracy of the collected data, which initially contained gaps or had some entries removed during quality checks. The global solar irradiance was evaluated using the horizontal surface values of extraterrestrial irradiance as a physical threshold [25]. Based on the quality control results, the missing data accounted for approximately 5%, 2%, and 3% of the total dataset for the first, second, and third cities, respectively.

However, in practical applications, solar radiation is commonly measured using instruments such as pyranometers, which are designed to measure global solar radiation. Pyranometers are typically calibrated according to World Meteorological Organization (WMO) standards, ensuring high accuracy and reliability. High-quality pyranometers have an uncertainty of less than 5% under most conditions, with precision ranging between ± 1 to 5 W/m². Other meteorological parameters, such as temperature, humidity, and wind speed, are often measured using sensors like thermometers, hygrometers, and anemometers, which also undergo regular calibration to maintain accuracy. Instrument calibration is typically performed against standardized reference instruments to ensure measurement precision and reduce systematic errors. Data quality checks and periodic recalibration of these instruments are essential to ensure that the measurements used for machine learning model training remain accurate and representative of the actual environmental conditions, which directly affects the reliability of the model's predictions [26].

2.2 Model 1: Linear regression

One of the simplest and most widely used machine learning methods is linear regression. It's a statistical technique for forecasting analysis. Linear regression produces predictions for numerical or continuous variables such as age, salary, sales, and product pricing, among others.

A process that shows a linear relationship between one or more independent (x) variables and a dependent (y) variable is called "linear regression" (see Figure 2). The way that the value of the dependent variable changes in response to the

value of the independent variable is determined by linear regression since it shows a linear relationship [27]:

A linear equation is used in linear regression to model the relationship between the input features (X) and the target variable (Y) [28]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_n X_n + \varepsilon \quad (3)$$

where,

Y is the global solar radiation.

β_0 is the intercept.

$\beta_1, \beta_2, \dots, \beta_n$ are the coefficients for the input features X_1, X_2, \dots, X_n .

X_2, \dots, X_n .

ε is the error term.

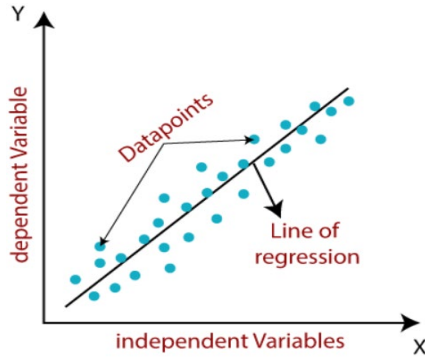


Figure 2. An example of how to use linear regression

2.3 Model 2: Random Forest

The output of several decision trees is combined by the potent machine-learning algorithm Random Forest to produce a single outcome. An ensemble learning technique called Random Forest constructs several decision trees (DTEs) and averages the predictions made by each (see Figure 3). A subset of the features and data are used to construct each decision tree [29, 30].

$$Y = \frac{1}{N} \sum_{i=1}^N T_i(X) \quad (4)$$

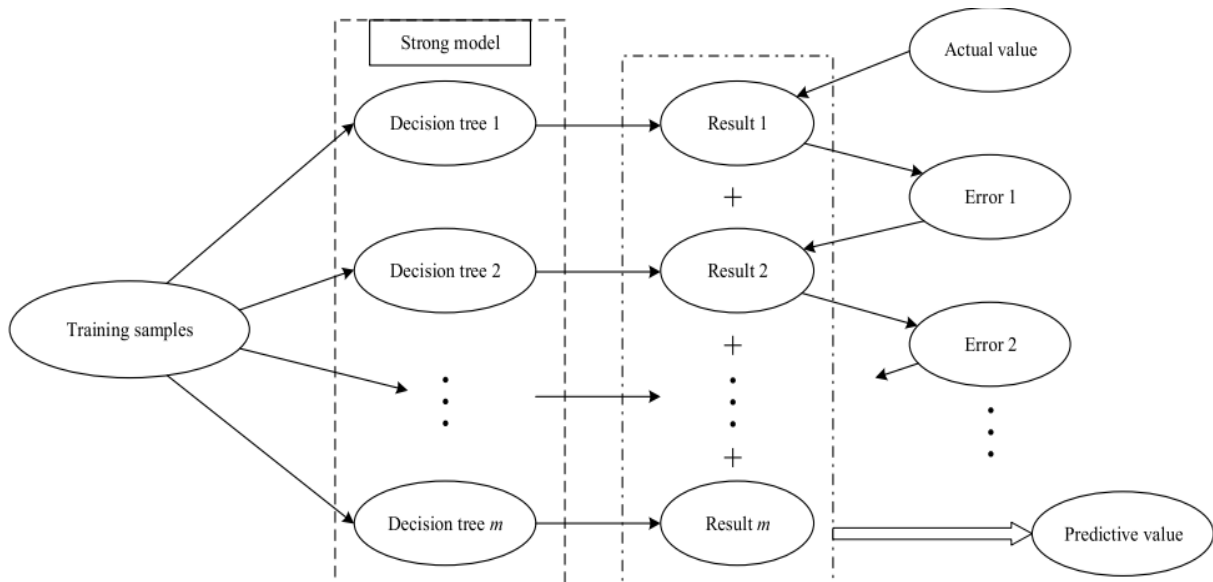


Figure 4. XGBoost lifting principle

where,

Y is the global solar radiation.

N is the number of trees.

$T_i(X)$ is the prediction of the i -th tree for the input features X .

The number of trees in forest are 50, 100, and 150.

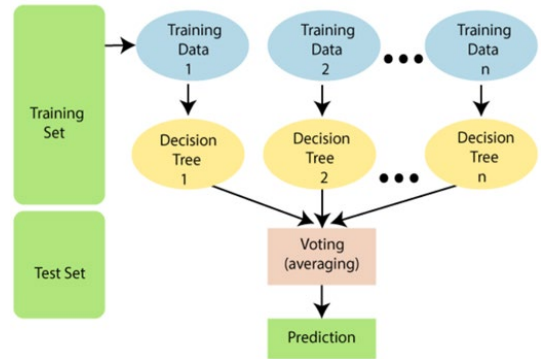


Figure 3. An illustration shows how the Random Forest algorithm operates

2.4 Model 3: XGBoost

The Gradient Boosting Decision Tree (GBDT) is the foundation of the classification and regression method XGBoost, which creates a sequence of weak learners and then integrates them to create a powerful prediction model [31]. XGBoost trains these weak learners by constructing a sequence of weak models, primarily regression or classification trees. After training, it merges the weak models to produce the final model (see Figure 4). With an eye on lowering the total model error, a new model is introduced based on the residual error from the prior model [32].

Next, the sample's prediction function is stated as follows:

$$Y = \sum_{k=1}^K f_k(X), f_k \in F \quad (5)$$

where,

Y is the global solar radiation.

K is the number of trees.

$f_k(x)$ is the prediction of the k -th tree.

Learning rate, tested values are 0.01, 0.1, 0.3.

Table 3. Statistical score equations

Abbreviation	Equations
R	$R = \frac{\sum_{i=1}^N (G_{for,i} - \bar{G}_{for})(G_{obs,i} - \bar{G}_{obs})}{\sqrt{\sum_{i=1}^N (G_{for,i} - \bar{G}_{for})^2} \times \sqrt{\sum_{i=1}^N (G_{obs,i} - \bar{G}_{obs})^2}} \quad (7)$
MAE	$MAE = \sqrt{\frac{1}{N} \sum_{i=1}^n G_{for,i} - G_{obs,i} } \quad (8)$
MBE	$MBE = \frac{1}{N} \sum_{i=1}^N (G_{for,i} - G_{obs,i}) \quad (9)$
$RMSE$	$RMSE = \sqrt{\frac{\sum_{i=1}^N (G_{for,i} - G_{obs,i})^2}{N}} \quad (10)$

2.5 Model 4: Hybrid model

The hybrid model that combines XGBoost and Random Forest makes use of each model's advantages to enhance prediction accuracy (see Figure 5). This method uses both Random Forest and XGBoost at the same time, combining their predictions to provide a final forecast [33]. To capitalize on their differences, a hybrid model integrates the predictions from many models, including XGBoost and Random Forest.

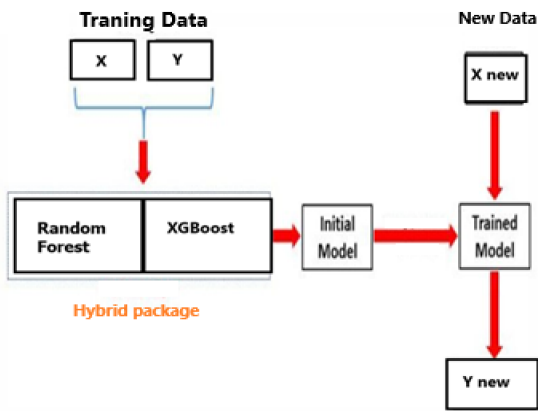


Figure 5. Hybrid model principle

Projections from each model are gathered and divided by the total number of models to get the final forecast in the hybrid model, which is computed using the constant ratio of mediation of various models' projections. Every model type is guaranteed to be assigned the proper weight in the final projection because of this constant ratio.

A weighted average or another combination approach can be used as the final prediction [34]:

$$Y = \omega_1 Y_{RF} + \omega_2 Y_{XGB} \quad (6)$$

where,

Y is the global solar radiation.

Y_{RF} is the prediction from the Random Forest model.

Y_{XGB} is the prediction from the XGBoost model.

ω_1 and ω_2 are the weights assigned to each model is prediction.

2.6 Statistical evaluation indices

A number of commonly used assessment score metrics, including Root Mean Square Error (RMSE), Mean Bias Error

(MBE), Mean Absolute Error (MAE), and coefficient of correlation value (R) [35], were employed to evaluate the performance of the models under study. The use of statistical metrics like R, MAE, RMSE, and MBE is crucial for evaluating the accuracy and reliability of machine learning models in solar radiation estimation. Each of these metrics offers a different perspective on the model's performance and its physical significance in the context of solar radiation measurements. Together, these metrics provide a comprehensive understanding of model performance, balancing accuracy, error sensitivity, and bias—key factors in ensuring reliable solar radiation estimates [36]. Table 3 provides definitions for these metrics, where N denotes the total number of data points, and G_{for} and G_{obs} stand for the forecast values of solar radiation and observed values, respectively.

3. RESULTS AND DISCUSSIONS

The current study's primary objective was to predict daily global solar radiation using four machine-learning models and compare the simulated and collected results. As a result, appropriate computer programs for all sky solar radiation condition estimation in the three locations were constructed using Python software.

Table 4 presents the performance of different machine learning models in the estimation of solar radiation across three cities in southern Algeria: Illizi, Timimoun, and Tamanaraset. The table includes the correlation coefficient values (R), the Mean absolute error (MAE), and the Root Mean Square Error (RMSE) per model in each city.

In the city of Illizi, Model 2 outstrips other models with a high R value and a low MAE and RMSE value; this model shows excellent performance with an R value of 0.9255 and the lowest MAE values of 429.49 and RMSE of 612.95 among the four models. Second place Model 4 also shows a very good performance with R values of 0.9225, MAE values of 445.51 and RMSE 625.58.

In the city of Timimoun, model 4 shows a distinct performance of R of 0.9640 and the lowest MAE values of 352.58 and RMSE of 471.22 among the four models. Model 2 ranks second with a very good performance with an R value of 0.9638.

In the city of Tamanaraset, model 4 shows the best performance again with a high R value of 0.8868 and the lowest MAE values of 483.99 and RMSE 681.62. Model 2 comes in second with good performance with an R value of

0.8817. The superior model in each city is highlighted in bold.

Table 4. Results of the model’s overall performance

City	Model	RMSE (wh.m ² /day)	MAE (wh.m ² /day)	R
Illizi	Model 1	953.58	730.57	0.8076
	Model 2	612.95	429.49	0.9255
	Model 3	688.92	472.51	0.906
	Model 4	625.58	445.51	0.9225
Timimoun	Model 1	814.47	642.51	0.8857
	Model 2	472.42	356.75	0.9638
	Model 3	519.94	379.35	0.9562
	Model 4	471.22	352.58	0.964
Tamanaraset	Model 1	836.92	626.33	0.8232
	Model 2	694.58	486.7	0.8817
	Model 3	752.16	531.79	0.8614
	Model 4	681.62	483.99	0.8868

Figure 6 represents a comparison of rRMSE values to assess the performance of different models in the three cities: Illizi, Timimoun, and Tamanaraset.

In Illizi City, we note that Model 2 shows the lowest value for rRMSE (8.99%), which means it is the most accurate model in forecasts and significantly reduces relative errors. Model 4 comes in second in the value of rRMSE (9.17%), which is close to model 2 but slightly less precise. Model 3 is ranked third with rRMSE (10.10%), superior to model 1 but less performing than models 2 and 4. Model 1 records the

highest rRMSE value (13.99%), indicating that it is less accurate in forecasts.

As we can see in Timimoun, Model 4 records the lowest value for rRMSE (7.01%), making it the most accurate model in forecasts for this city. Model 2 is very close to model 4 for rRMSE (7.03%), indicating excellent performance. Model 3 ranks third with rRMSE (7.73%), which is a good performance but is lower than models 2 and 4. Model 1 again, the worst performance of rRMSE is 12.12%.

We also observe in Tamanaraset. Model 4 records the lowest value for rRMSE (10.23%), making it the most accurate model in forecasts in this city. Model 2 ranks second with rRMSE (10.43%), reflecting very good accuracy. Model 3 is ranked third with rRMSE (11.29%), superior to model 1 but with less performance than models 2 and 4. Model 1 records the highest rRMSE value (12.57%), indicating that it is less accurate in forecasts.

Figure 7 represents a comparison of rMAE values to assess the performance of different models in the three cities: Illizi, Timimoun, and Tamanaraset.

We note in Illizi that Model 2 shows the lowest value for rMAE (6%), which means it is the most accurate model in forecasts and significantly reduces absolute errors. Model 4 second in the value of rMAE (6.53%), which is very close to model 2 but slightly less precise. Model 3 ranked third in rMAE (6.93%), superior to model 1 but with less performance than models 2 and 4. Model 1 records the highest value of rMAE (10.71%), indicating that it is less accurate in forecasts.

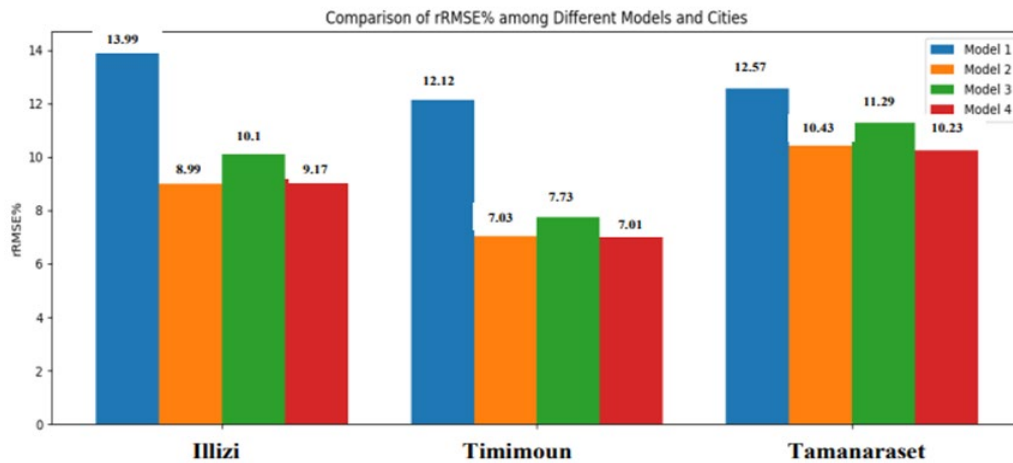


Figure 6. Comparison of rRMSE

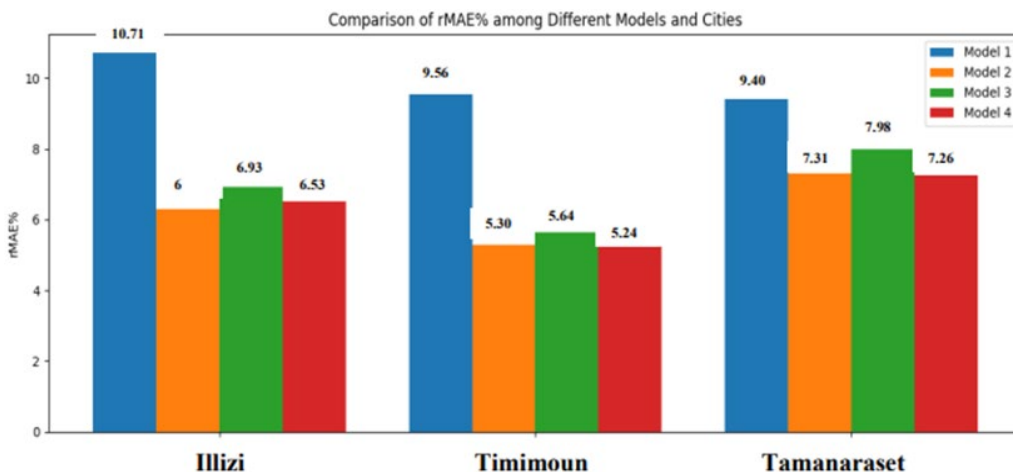


Figure 7. Comparison of rMAE

As we can see in Timimoun, model 4 records the lowest value for rMAE (5.24%), making it the most accurate model in forecasts for this city. Model 2 is very close to Model 4 for rMAE (5.30%), indicating excellent performance. Model 3 ranks third with rMAE (5.64%), which is a good performance but is lower than models 2 and 4. Model 1 again, the worst performance of rMAE (9.56%).

We also note in Tamanaraset Model 4 records the lowest value for rMAE (7.26%), making it the most accurate model in forecasts in this city. Model 2 comes second in the value of rMAE (7.31%), reflecting very good accuracy. Model 3 ranks third with rMAE (7.98%), superior to model 1 but with less performance than models 2 and 4. Model 1 records the highest value of rMAE (9.40%), indicating that it is the least accurate forecast.

Charts in Figure 8, we note that in model 2 (with R=0.9255),

the points are very close to the red line and the points are distributed systematically and less dispersed around the red line than in the other three models, meaning that the difference between actual and predictive values is small, and therefore Model 2 has fewer predictive errors. This indicates that Model 2 is the most accurate and consistent predictor of actual data values in the city of Illizi.

Based on the charts presented in Figure 9, it can be observed that Model 4 (with R=0.9640) has fewer predictive errors than the other three models because its points are distributed systematically and are closer to the red line than those in the other models. This suggests that there is little variation between the estimated and actual values. Based on this, it can be concluded that Timimoun is actual data values are most reliably and accurately predicted by Model 4.

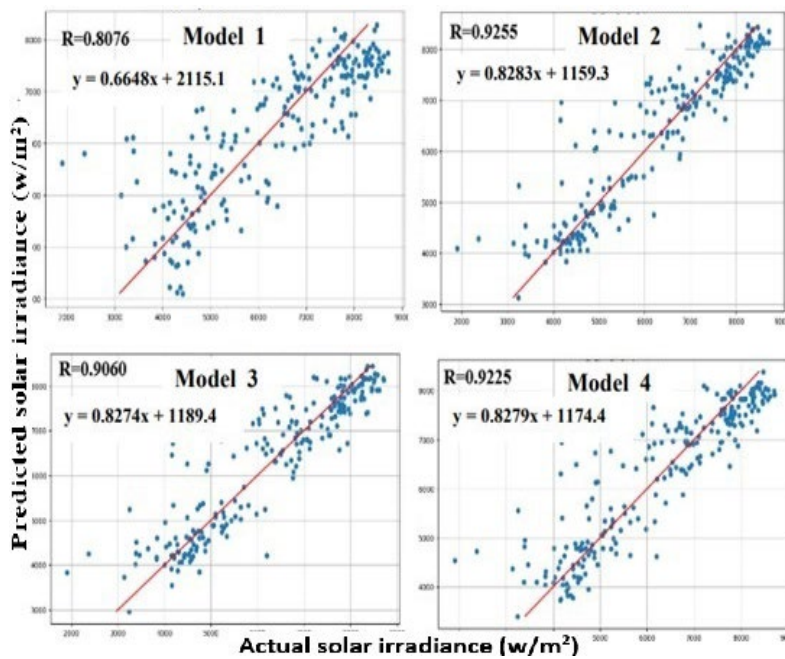


Figure 8. Scatter plot of Illizi

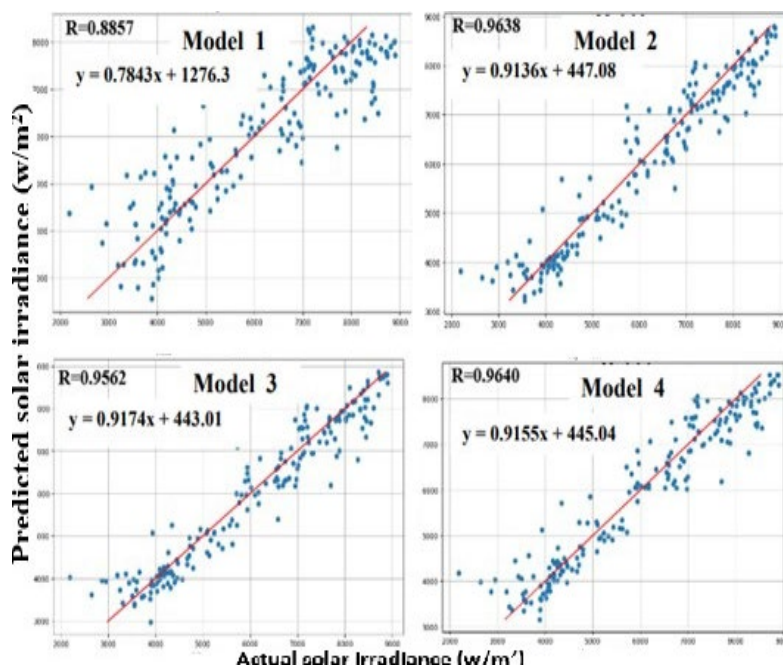


Figure 9. Scatter plot of Timimoun

According to the charts in Figure 10, Model 4 (with $R=0.8868$) has fewer predictive errors than the other three models because the points are distributed more systematically and less randomly around the red line than in the other three models. This indicates that there is less of a difference between the actual and predicted values. This suggests that, for real data values in the city of Tamanaraset, Model 4 is the most reliable and consistent predictor.

Figures 11, 12 and 13 present predicted and actual daily global solar radiation values following the acquisition of the prediction outcomes. These figures allow for the comparison of forecasted values with actual values, enabling evaluation of variations between them.

The blueprints created in these figures shape the expected results, facilitating the easy identification of differences and

discrepancies between predicted and collected values.

Figure 11 presents the comparison of actual and forecasted values in Illizi during six (6) years of study period in all sky conditions. The analysis of Model 2 is considered the best model to note that red dots are very close to the blue dots, indicating high accuracy in prediction.

Figure 12 illustrates a comparison between actual and predicted values in Tamanaraset city during six years of study period in all sky conditions using Model 4, which is considered the best model. Red dots match very close to blue dots, indicating high accuracy in prediction.

Figure 13 illustrates a comparison between actual and predicted values in Timimoun city during this study using Model 4, which is considered the best model. Red dots match very close to blue dots, indicating high accuracy in prediction.

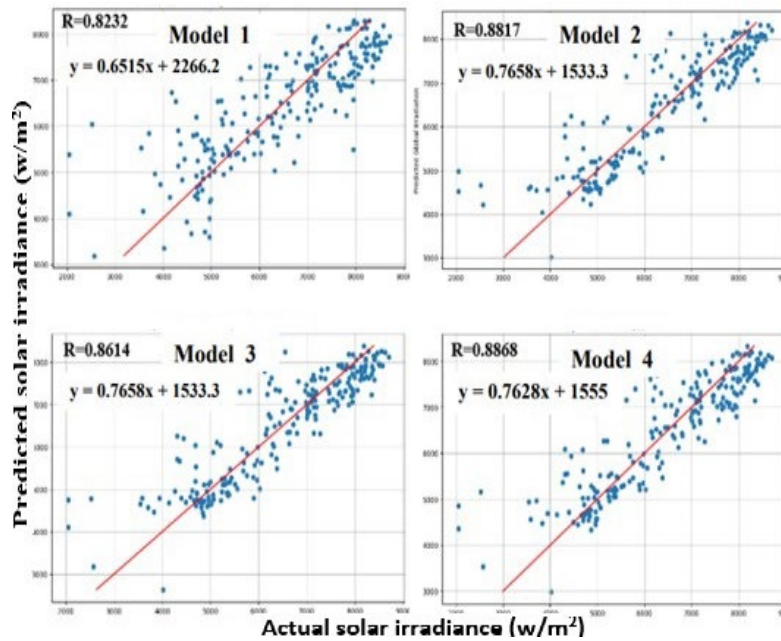


Figure 10. Scatter plot of Tamanaraset

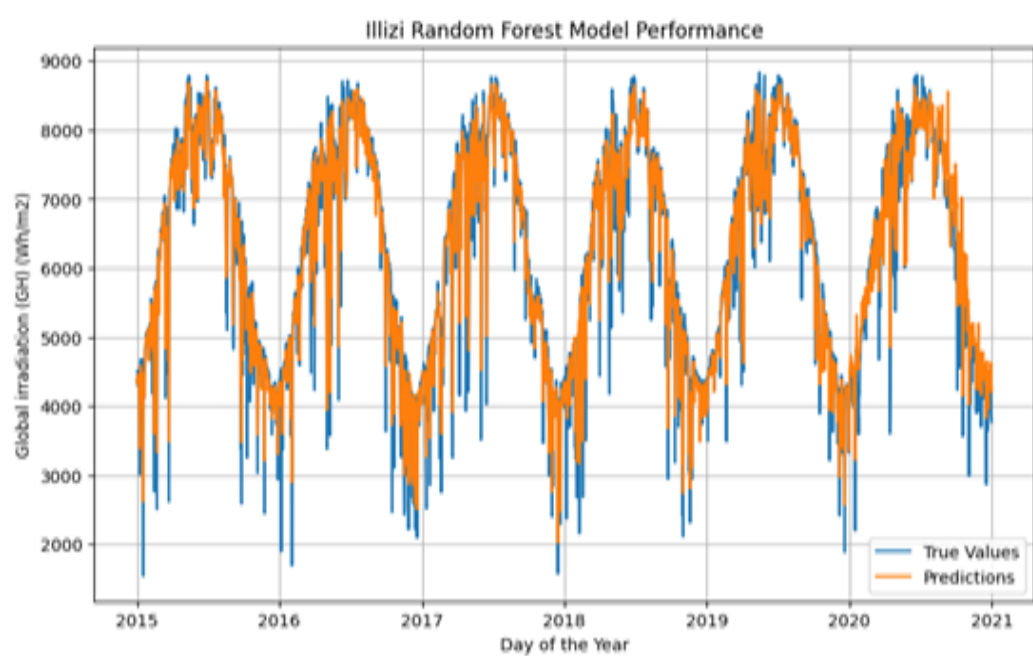


Figure 11. Predicted and actual values of the best model (Model 2) for Illizi

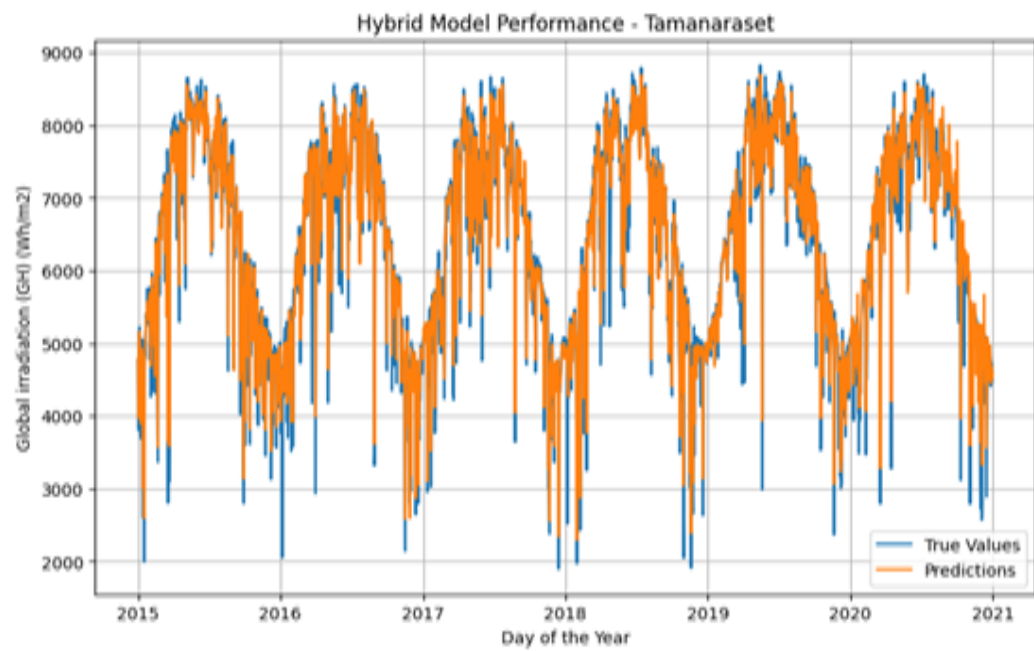


Figure 12. Predicted and actual values of Tamanaraset (Model 4)

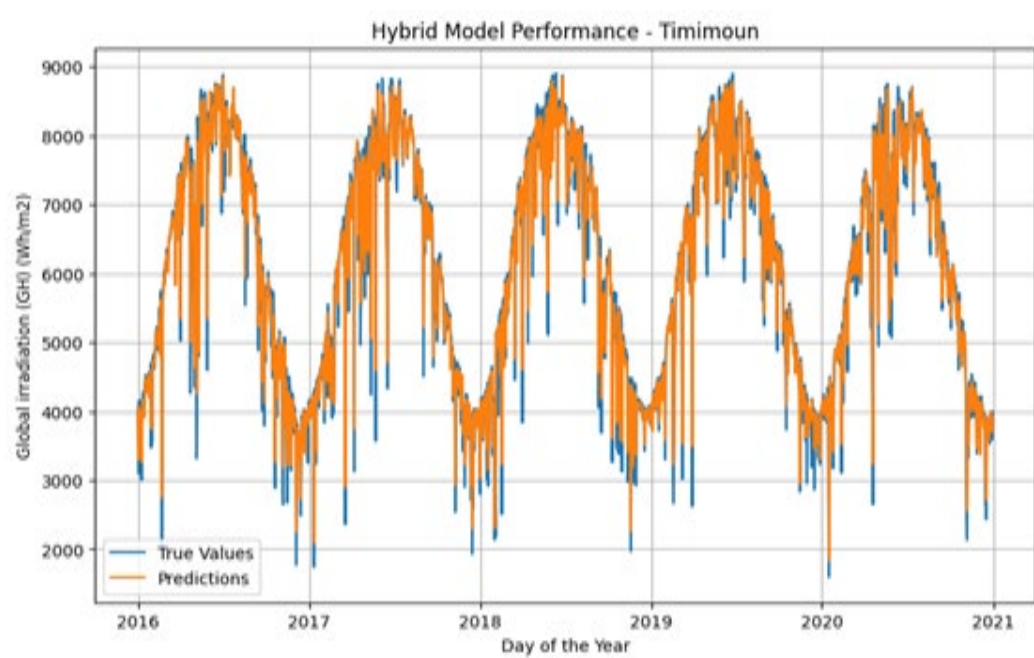


Figure 13. Predicted and actual values of Timimoun (Model 4)

Table 5. Statistics assess errors across the different previous works

Ref.	Model	Location	R	rRMSE%	RMSE	MAE	rMAE%
[37]	RF	Tamil Nadu, a state in southern India	0.4490	38.58	933.32	569.59	23.13
[37]	MLR	Tamil Nadu, a state in southern India	0.4520	38.14	923.44	554.07	22.44
[38]	RF	El Oued- City in Algeria	/	/	/	317.32	/
[39]	Linear Regression	Perlis, Northern Malaysia	0.7780	/	/	/	/
This study	RF	Illizi -City in Algeria	0.9255	8.99	612.95	429.49	6
This study	Hybrid	Timimoun -City in Algeria	0.9640	7.01	471.22	352.58	5.24

The hybrid model combining Random Forest and XGBoost exhibits precision in forecasting solar radiation in both Tamanaraset, and Timimoun regions due to several key factors:

-Ensemble Learning: Both Random Forest and XGBoost are ensemble learning methods that combine multiple weak

learners to create a stronger model. This approach helps in reducing overfitting and improving generalization.

-Handling Non-linearity: Both algorithms are capable of capturing complex, non-linear relationships in the data. This is particularly useful in modeling solar radiation, which can be influenced by various non-linear factors such as weather

patterns and geographical features.

-Feature Importance: Random Forest and XGBoost can provide insights into feature importance, helping to identify which variables contribute the most to the model's predictions. This can be crucial in understanding the factors affecting solar radiation.

-Robustness to outliers: Both models are relatively robust to outliers and noise in the data, which can be common in environmental datasets.

Combining these strengths, the hybrid model leverages the advantages of both Random Forest and XGBoost, resulting in high precision in forecasting global solar radiation in the two cities.

The results of the research imply that the XGBoost machine-learning model and the Random Forest-XGBoost hybrid model are equally successful in estimating Algeria's radiation levels. While the Random Forest -XGBoost model provides results in the Timimoun and Tamanrasset regions, the accuracy of the Random Forest model is particularly beneficial to the Illizi region. These models provide data that can be used to build, design, and optimize the energy systems in these areas. In summary, Linear Regression performs well in stable, predictable climates, while XGBoost and Random Forest excel in regions with complex or highly variable climatic and geographical conditions. Each model's performance is closely tied to its capacity to capture the underlying relationships in local weather and terrain.

Table 5 displays several models used to predict solar irradiation. The results of the proposed models show high and consistent effectiveness in terms of predictability and improved predictability when compared to other previous models. The results are better. Based on the data presented in the table, it appears that the RF and MLR models were applied in Tamil Nadu, a state in southern India, where their results were almost similar in terms of their ability to predict solar irradiation. The results for R, rRMSE, and rMAE are divergent and not close in terms of solar radiation estimates compared to the results of our study at the bottom of the table.

However, the RF model applied in Illizi, a city in Algeria, seems to have performed better compared to the other models, as it had a higher correlation value (R) and lower error percentage (% rRMSE), as well as lower Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Finally, the Hybrid RF-XGBoost model applied in Timimoun, a city in Algeria, appears to have achieved the best performance among all models. It had the highest correlation value (R) and the lowest percentage of both Root Mean Square Error (% rRMSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), indicating that this model performed the best in predicting solar irradiation at the specified location, and can be generalized to similar regions with the same input parameters.

4. SOFTWARE FOR ESTIMATING SOLAR RADIATION

Our software called ZAHYA_SOLAR (see Figure 14) is designed using Python. It utilizes data libraries such as Pandas for data loading and analysis, and Scikit-learn for building and training machine learning models such as Linear Regression, Random Forest, and XGBoost. Users can select the city and choose the appropriate model, then easily view the results and charts within the interface. The application is intended to

provide a graphical interface for users to evaluate and predict solar radiation at any location in all sky conditions using machine learning models. It allows users to load test data, preprocess it, and apply one of four prediction models. After obtaining the results of solar radiation estimation at the specified location, it can be used in the design and installation of any renewable energy project.

The application, ZAHYA_SOLAR, is designed to seamlessly integrate with the machine learning models developed for estimating solar radiation. The application serves as a user-friendly interface that allows users to input relevant meteorological data (such as temperature, humidity, and solar angles), which is then fed into the models. These models, built using techniques like regression, neural networks, or other machine learning algorithms, process the input data and generate solar radiation estimates for the user in real-time or over specified periods. Future practical applications could have practical potential in the field of solar measurement such as: solar power plant design and optimization, energy production prediction, site evaluation, performance monitoring, ...etc.

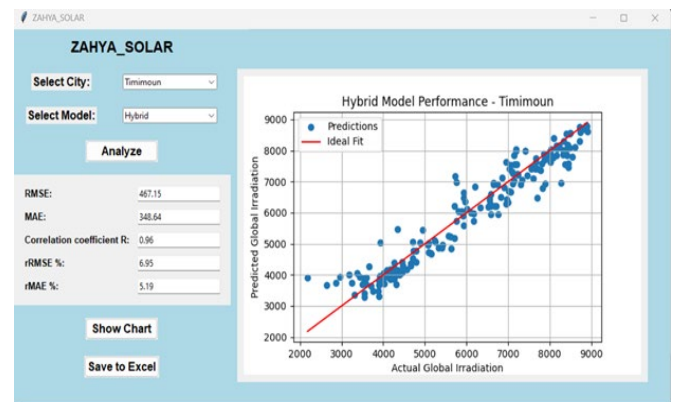


Figure 14. ZAHYA_SOLAR software

5. CONCLUSIONS

The study underscores the significant potential of artificial intelligence techniques in predicting global solar radiation, particularly in regions abundant in solar resources in southern Algeria. By employing diverse machine learning models, including the innovative hybrid model combining XGBoost and Random Forest, precise and reliable results were achieved, enhancing our understanding of optimizing solar energy utilization. The results obtained from the successful analysis of hybrid models, incorporating Random Forest and XGBoost, demonstrate high quality in solar radiation prediction. In the city of Illizi, Random Forest exhibited excellent performance with an R value of 0.9255. In Timimoun, the hybrid model (RF-XGBoost) displayed outstanding performance with an R value of 0.9640, while in Tamanrasset, the hybrid model showed good performance with an R value of 0.8868. These results provide additional support for the effectiveness of hybrid techniques in solar radiation prediction and enhancing the sustainability of solar energy projects. Based on these accurate solar radiation forecasts, significant improvements can be made in the efficiency and sustainability of solar energy projects. They enable operators to better plan for harnessing available solar resources, thereby promoting sustainable development and reducing reliance on traditional energy

sources. When data is available for solar energy system design, heating, and cooling in Saharan climate zones, this hybrid model can be applied. Future research could focus on applying machine learning (ML) models for solar radiation estimation in real-world solar energy projects, particularly in optimizing solar power generation and grid integration. This includes enhancing the accuracy of radiation forecasts to improve energy management, reducing operational costs, and integrating renewable energy sources into smart grid systems. Here are some promising research directions for improving solar radiation estimation and expanding the future application's capabilities:

- Incorporating Real-Time Data Streams: Satellite and Sensor Integration, IoT and Smart Sensors
- Exploring Advanced Machine Learning Techniques: Deep Learning and Neural Networks, Hybrid Models, Ensemble Learning
- Incorporating Climate Change Projections:
- Expanding Geographic Scope: Transfer Learning for Regional Adaptation, Global Weather Data Models.

REFERENCES

- [1] Un, C. (2023). A sustainable approach to the conversion of waste into energy: Landfill gas-to-fuel technology. *Sustainability*, 15(20): 14782. <https://doi.org/10.3390/su152014782>
- [2] Günen, M.A. (2021). A comprehensive framework based on GIS-AHP for the installation of solar PV farms in Kahramanmaraş, Turkey. *Renewable Energy*, 178: 212-225. <https://doi.org/10.1016/j.renene.2021.06.078>
- [3] Voyant, C., Notton, G., Kalogirou, S., Nivet, M.L., Paoli, C., Motte, F., Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105: 569-582. <https://doi.org/10.1016/j.renene.2016.12.095>
- [4] Djeldjli, H., Benatiallah, D., Tanougast, C., Benatiallah, A. (2024). Solar radiation forecasting based on ANN, SVM and a novel hybrid FFA-ANN model: A case study of six cities south of Algeria. *AIMS Energy*, 12(1): 62-83. <https://doi.org/10.3934/energy.2024004>
- [5] Ağbulut, Ü., Gürel, A.E., Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews*, 135: 110114. <https://doi.org/10.1016/j.rser.2020.110114>
- [6] Kumar, R. (2023). Solar radiation analysis for predicting climate change using deep learning techniques. In *Effective AI, Blockchain, and E-Governance Applications for Knowledge Discovery and Management*. IGI Global, pp. 58-68. <https://doi.org/10.4018/978-1-6684-9151-5.ch004>
- [7] Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J., Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. *Solar Energy*, 136: 78-111. <https://doi.org/10.1016/j.solener.2016.06.069>
- [8] Dalapati, G.K., Ghosh, S., Sherin, T., Ramasubramanian, B., Samanta, A., Rathour, A., Wong, T.K.S., Chakraborty, S., Ramakrishna, S., Kumar, A. (2023). Maximizing solar energy production in ASEAN region: Opportunity and challenges. *Results in Engineering*, 20: 101525. <https://doi.org/10.1016/j.rineng.2023.101525>
- [9] Halima, D., Djelloul, B., Mehdi, G., Camel, T., Ali, B., Bouchra, B. (2024). Solar radiation estimation based on a new combined approach of artificial neural networks (ANN) and genetic algorithms (GA) in South Algeria. *Computers, Materials & Continua*, 79(3). <http://doi.org/10.32604/cmc.2024.051002>
- [10] Gligorea, I., Cioca, M., Oancea, R., Gorski, A.T., Gorski, H., Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, 13(12): 1216. <https://doi.org/10.3390/educsci13121216>
- [11] Benatiallah, D., Bouchouicha, K., Benatiallah, A., Harouz, A., Nasri, B. (2021). Artificial neural network based solar radiation estimation of Algeria Southwest Cities. In *4th International Conference on Artificial Intelligence in Renewable Energetic Systems*, Tipaza, Algeria, pp. 573-583. https://doi.org/10.1007/978-3-030-63846-7_54
- [12] Gupta, S., Singh, A.K., Mishra, S., Vishnuram, P., Dharavat, N., Rajamanickam, N., Kalyan, C.N.S., AboRas, K.M., Sharma, N.K., Bajaj, M. (2023). Estimation of solar radiation with consideration of terrestrial losses at a selected location—A review. *Sustainability*, 15(13): 9962. <https://doi.org/10.3390/su15139962>
- [13] Van der Meer, D.W., Widén, J., Munkhammar, J. (2018). Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renewable and Sustainable Energy Reviews*, 81: 1484-1512. <https://doi.org/10.1016/j.rser.2017.05.212>
- [14] Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G. (2009). Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1): 2-10. <https://doi.org/10.1109/JSTARS.2009.2020300>
- [15] Alessandrini, S., Delle Monache, L., Sperati, S., Cervone, G. (2015). An analog ensemble for short-term probabilistic solar power forecast. *Applied Energy*, 157: 95-110. <https://doi.org/10.1016/j.apenergy.2015.07.059>
- [16] Sagade, A.A., Samdarshi, S.K., Panja, P.S. (2018). Experimental determination of effective concentration ratio for solar box cookers using thermal tests. *Solar Energy*, 159: 984-991. <https://doi.org/10.1016/j.solener.2017.11.021>
- [17] Benatiallah, D., Benatiallah, A., Bouchouicha, K., Nasri, B. (2020). Prediction du rayonnement solaire horaire En utilisant les reseaux de neurone artificiel. *Algerian Journal of Environmental Science and Technology*, 6(1): 1236-1245.
- [18] Djeldjli, H., Benatiallah, D., Bouchouicha, K., Benatiallah, A. (2022). Solar radiation forecasting based on artificial neural network: A case study of Bechar City, Southwest Algeria. In *International Conference on Artificial Intelligence in Renewable Energetic Systems (IC-AIRES 2022)*, Tipasa, Algeria, pp. 3-12. https://doi.org/10.1007/978-3-031-21216-1_1
- [19] Benatiallah, D., Bouchouicha, K., Benatiallah, A., Harouz, A., Nasri, B. (2022). Assessment of global solar energy under all-sky condition using artificial neural network. In *International Conference on Artificial Intelligence in Renewable Energetic Systems (IC-AIRES 2021)*, Tipasa, Algeria, pp. 167-174. https://doi.org/10.1007/978-3-030-92038-8_16

- [20] Bouchouicha, K., Bailek, N., Bellaoui, M., Oulimar, B., Benatiallah, D. (2021). ANN-based correction model of radiation and temperature for solar energy application in South of Algeria. In 4th International Conference on Artificial Intelligence in Renewable Energetic Systems (ICAIREs 2020), Tipaza, Algeria, pp. 584-591. https://doi.org/10.1007/978-3-030-63846-7_55
- [21] Benatiallah, D., Bouchouicha, K., Benatiallah, A., Harrouz, A., Nasri, B. (2019). Forecasting of solar radiation using an empirical model. *Algerian Journal of Renewable Energy and Sustainable Development*, 1(2): 212-219. <https://doi.org/10.46657/ajresd.2019.1.2.11>
- [22] Senouci, A., Maouedj, R., Benatiallah, A., Benatiallah, D., Bouchouicha, K. (2023). Enhanced daily global solar radiation prediction through hybrid artificial neural network and adaptive neuro-fuzzy inference system with meta-heuristic algorithm integration. *Instrumentation Mesure Métrologie*, 22(6): 241-251. <https://doi.org/10.18280/i2m.220603>
- [23] Benatiallah, D., Bouchouicha, K., Benatiallah, A., Harrouz, A., Nasri, B. (2021). Estimation of global solar radiation using adaptive neuro-fuzzy inference system. *Algerian Journal of Environmental Science & Technology*, 7(4): 2097-2106.
- [24] Nasa data. Available at: <http://gmao.gsfc.nasa.gov/reanalysis/MERRA-2>
- [25] Journée, M., Bertrand, C. (2011). Quality control of solar radiation data within the RMIB solar measurements network. *Solar Energy*, 85(1), 72-86. <https://doi.org/10.1016/j.solener.2010.10.021>
- [26] World Meteorological Organization. (2018). Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8). World Meteorological Organization.
- [27] Ibrahim, S., Daut, I., Irwan, Y.M., Irwanto, M., Gomesh, N., Farhana, Z. (2012). Linear regression model in estimating solar radiation in Perlis. *Energy Procedia*, 18: 1402-1412. <https://doi.org/10.1016/j.egypro.2012.05.156>
- [28] Montgomery, D.C., Peck, E.A., Vining, G.G. (2021). Introduction to Linear Regression Analysis. John Wiley & Sons.
- [29] Soufiane, G., Ouafia, F., Ahmed, A. (2022). Solar radiation time-series prediction using random forest algorithm-based feature selection approach. In International Conference on Digital Technologies and Applications (ICDTA 2022), Benguerir, Morocco, pp. 659-668. https://doi.org/10.1007/978-3-031-02447-4_68
- [30] Villegas-Mier, C.G., Rodriguez-Resendiz, J., Álvarez-Alvarado, J.M., Jiménez-Hernández, H., Odry, Á. (2022). Optimized random forest for solar radiation prediction using sunshine hours. *Micromachines*, 13(9): 1406. <https://doi.org/10.3390/mi13091406>
- [31] Li, X.L., Ma, L.F., Chen, P., Xu, H., Xing, Q.J., Yan, J.H., Lu, S.Y., Fan, H.H., Yang, L., Cheng, Y.Q. (2022). Probabilistic solar irradiance forecasting based on XGBoost. *Energy Reports*, 8: 1087-1095. <https://doi.org/10.1016/j.egypr.2022.02.251>
- [32] Phan, Q.T., Wu, Y.K., Phan, Q.D. (2021). Short-term solar power forecasting using XGBoost with numerical weather prediction. In 2021 IEEE International Future Energy Electronics Conference (IFEEEC), Taipei, Taiwan, pp. 1-6. <https://doi.org/10.1109/IFEEEC53238.2021.9661874>
- [33] Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., Zeng, Z. (2021). Solar radiation prediction using different machine learning algorithms and implications for extreme climate events. *Frontiers in Earth Science*, 9: 596860. <https://doi.org/10.3389/feart.2021.596860>
- [34] Venu, K., Prakash, K.I., Jayaram, S., Karan, N.S., Raja, M.M., Renu, K. (2023). Solar radiation prediction using machine learning model. In 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, pp. 52-57. <https://doi.org/10.1109/ICSCDS56580.2023.10104904>
- [35] Stone, R.J. (1993). Improved statistical procedure for the evaluation of solar radiation estimation models. *Solar Energy*, 51(4): 289-291. [https://doi.org/10.1016/0038-092X\(93\)90124-7](https://doi.org/10.1016/0038-092X(93)90124-7)
- [36] Yang, D., Kleissl, J. (2016). Preprocessing and parameter selection for machine learning of solar radiation. *Renewable Energy*, 89: 144-153. <https://doi.org/10.1016/j.renene.2015.11.062>
- [37] Obiora, C.N., Ali, A., Hasan, A.N. (2021). Implementing extreme gradient boosting (XGBoost) algorithm in predicting solar irradiance. In 2021 IEEE PES/IAS PowerAfrica, Nairobi, Kenya, pp. 1-5. <https://doi.org/10.1109/PowerAfrica52236.2021.9543159>
- [38] Krithika, K.M., Maheswari, N., Sivagami, M. (2022). Models for feature selection and efficient crop yield prediction in the groundnut production. *Research in Agricultural Engineering*, 68(3): 131-141. <https://doi.org/10.17221/15/2021-RAE>
- [39] Kourt, M., Arbaoui, M. (2023). A new artificial intelligence method for estimating solar radiation. Master's Thesis in Computer Science, Adrar University.

NOMENCLATURE

AI	Artificial Intelligence
LR	Linear Regression
MAE	Mean absolute error
ML	Machine Learning
R	Correlation coefficient
RF	Random Forest
RMSE	Root Mean Square Error
XGBoost	Extreme gradient boosting