International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Harnessing Madagascar's Solar Potential: Global Horizontal Irradiance Forecasting in Mahajanga using Satellite Data Analysis

Rakotomalala Lovasoa Feno Fanantenana[1], Totozafiny Theodore[1], Randrianirina Jean Marc Fabien Sitraka[2], Liva Graffin Rakotoarimanana[2], Zely Arivelo Randriamanantany[2]
[1]ISSTM and FSTE department, Mahajanga University, Mahajanga, Madagascar
[2]Physics department/ IME, Antananarivo University, LTTC Laboratory, Antananarivo, Madagascar
Email address: ralovas@gmail.com, theodore.totozafiny@gmail.com, randrianirinajeanfabien@gmail.com, zelyran@yahoo.fr

Abstract— *Global Horizontal Irradiance (GHI) is an essential factor in determining solar energy potential and photovoltaic (PV) system efficiency. Accurate estimation of GHI is important for enhancing solar power generation. Madagascar is considered as a high potentiality in this range of energy resource. This paper presents an approach for predicting GHI in Mahajanga city of Madagascar using its own satellite data combined with assemble of learning techniques. The study incorporates multiple data analyses steps including data collection and data pre-processing. To develop the analysis of the features, several methods were performed such as Correlation Matrix to select the element which combine the real data of GHI for its finest forecast. Thus, several machine learning models, such as Linear Regression, Artificial Neural Network (ANN), Decision Tree, K-Nearest Neighbours (KNN), Random Forest, Long Short Term Memory (LSTM), Support Vector Machine (SVM), DBSCAN and Light Gradient Boosting Machine (Light GBM with XGBoost /Feature engineering) were used to verify the compatibility of the data, highlighted the Random Mean Squad Error (RMSE %) of each model. Then, with the minimum RMSE results, four (04) models from them are chosen such as Light GBM (3.85 %), Random Forest (3.92 %), ANN (4.13 %) and Decision Tree (5.23%) to achieve the prediction of the GHI with features BHI, BNI, TOA along with Clear sky BHI as well as Clear sky GHI. (R^2 score: 0.9992, Mean Absolute Error (MAE): 0.84 and Root Mean Squared Error (RMSE): 2.46)*

Keywords— *Global horizontal irradiance; machine learning; correlation, prediction, data analysis.*

I. Introduction

The increasing global demand for sustainable energy underscores the importance of solar power as a key contributor to future energy systems. In high-potential regions such as Madagascar where solar resources remain largely underexploited, the ability to accurately forecast solar radiation is critical for optimizing photovoltaic (PV) performance and enhancing grid integration. Global Horizontal Irradiance (GHI), in particular, serves as a fundamental parameter in determining solar energy potential and system efficiency, [1].

This study aims to develop robust forecasting models for GHI in Mahajanga, precisely in Madagascar, using satellite-derived data in combination with advanced machine learning techniques. By improving the precision of GHI predictions, this work aims to support the efficient deployment of solar technologies and inform about energy planning in tropical regions.

The research addresses several key areas: the collection and preprocessing of high-resolution satellite data, the identification of features most strongly correlated with GHI then the evaluation of multiple machine learning models to determine their predictive performance.

The remainder of this paper is organized as follows: Section 1 reviews the current state of research in solar irradiance forecasting. Section 2 describes the data processing and modeling methodology. Section 3 presents the experimental results and comparative analysis of forecasting models. Finally, Section 4 deals with the implications of the findings and so will suggest directions for future work.

II. Literature Review

Liu et al. demonstrated that the combined use of the WRF meteorological model and the Kalman filter significantly reduces the root mean square error (RMSE) of global horizontal irradiance forecasts, from 102.9 W/m² to 79.1 W/m² for intraday predictions, [2].

In 2021, a study developed and compared several deep neural network architectures (LSTM, Bi-LSTM, GRU, Bi-GRU, CNN, CNN-LSTM, CNN-BiLSTM) for day-ahead forecasting of global horizontal irradiation (GHI) in Hail (correlation coefficient of 96% and a mean absolute error below 1 kWh/m²/day), Saudi Arabia, demonstrating that deep learning is highly effective in capturing seasonal variations in solar radiation using only historical data.

In 2021 [3], deep neural network architectures such as LSTM, Bi-LSTM, GRU, Bi-GRU, CNN, CNN-LSTM, and CNN-BiLSTM were developed and evaluated for day-ahead forecasting of global horizontal irradiation (GHI) in Hail, Saudi Arabia. These models were trained using only historical GHI data, without incorporating additional meteorological variables. The Bi-LSTM model achieved the highest performance, with a correlation coefficient of 96% and a mean absolute error below 1 kWh/m²/day. These results confirm that deep learning approaches are highly effective in capturing seasonal patterns in solar radiation, even in arid climates with limited data sources.

In 2022 [4], a study conducted in Errachidia, Morocco, explored the effect of feature selection on the accuracy of Direct Normal Irradiance (DNI) forecasting using assemble of machine learning models. The research applied four tree-based regressors

Rakotomalala Lovasoa Feno Fanantenana, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Liva Graffin Rakotoarimanana, and Zely Arivelo Randriamanantany, "Harnessing Madagascar's Solar potential: Global Horizontal Irradiance forecasting in Mahajanga using Satellite Data Analysis," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 17-23, 2025.

Random Forest, XGBoost, CatBoost, and LightGBM to identify the most influential meteorological and solar parameters from a dataset covering 2017 to 2019. Among the evaluated models, the Random Forest regressor achieved the best performance, reaching an accuracy of 99.57% with 11 selected features and a training time of 3 minutes and 23 seconds. These findings highlight that carefully selecting input features, especially GHI, DHI, and solar zenith angle, significantly improves DNI prediction without requiring the full assemble of parameters.

In 2022 [5], a multichannel deep learning framework combining wavelet transform, convolutional neural networks (CNN), and bidirectional LSTM (MC-WT-CBiLSTM) was proposed to enhance the accuracy of short-term GHI forecasting using temperature and irradiance data. Compared to classical and hybrid models, the proposed MC-WT-CBiLSTM achieved superior results across all time intervals (10, 30, and 60 minutes), with an R^2 value reaching 0.998 and significantly lower MAE and RMSE values, demonstrating its robustness for capturing solar irradiance fluctuations in real-world conditions.

In 2018 [6], experimental results demonstrated that the horizon-group approach achieved the best performance, with a global nRMSE of 0.227, outperforming all individual forecasting models.

In 2023 [7], the results showed that artificial neural network (ANN) models achieved the highest accuracy, with a coefficient of determination $R^2 = 0.999$, a root mean square error (RMSE) of 0.234 MJ/m²/day, and a mean absolute percentage error (MAPE) below 5% in several reviewed studies.

In 2025 [8], the study found that the Linear Regression model achieved the best performance, with an RMSE of 139.71 and a Mean Absolute Error of 98.32, significantly outperforming Support Vector Regression (RMSE: 487.20) and Decision Tree (RMSE: 504.67), making it the most effective approach for GHI prediction in Saudi Arabia.

III. METHOD DESCRIPTIONS

The whole approach steps are listed as follows:

- Data analysis tasks of all datasets used in this research are carried out ahead containing data collection and data preprocessing to prepare the data for several steps before results.
- The choice of the models to perform the forecasting.

II.1 Data analysis

A. Data Collection

In this work a potential solar area is taken into consideration for the collection of the data and the simulation: Mahajanga is localized in the west coast of Madagascar, characterized by a tropical climate which is considered as the warmest region of the island with a temperature around 27°C to 35°C. The average annual solar radiation is 200 Wh/m^2 while the geographical coordinates are 15°43'S and 46°19'E (fig.1). To build the proposed forecasting model proposed in this paper, solar datasets are collected from Solar Energy Copernicus satellite data. The data is recorded every fifteen (15) minutes range from the website within the location of the site. The observation period is one (01) recent year from 2024-01-01 to 2025-01-01. Figure 1 shows the variation of the features data deployed, thus, the unity of measurement is Wh/m^2 (ordinate axis). The following data are depicted: Irradiation on horizontal plane at the top of atmosphere(TOA), Sky global irradiation on horizontal plane at ground level clear (Clear sky GHI), Clear sky beam irradiation on horizontal plane at ground level (Clear sky BHI), Clear sky diffuse irradiation on horizontal plane at ground level (Clear sky DHI), Clear sky beam irradiation on mobile plane following the sun at normal incidence (Clear sky BNI), Global irradiation on horizontal plane at ground level (GHI), Beam irradiation on horizontal plane at ground level(BHI), Diffuse irradiation on horizontal plane at ground level(DHI), Beam irradiation on mobile plane following the sun at normal incidence(BNI). One has the Proportion of reliable data in the summarization (Reliability). Reliability is a precision factor of the data, without unity but varies from 0 to 1.

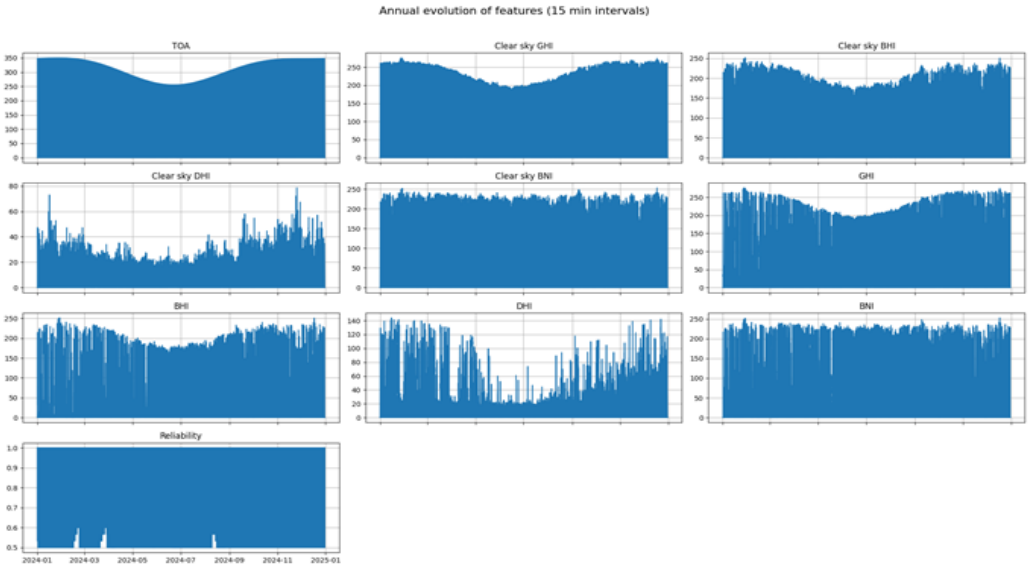

Figure 1: Annual data collected features

Rakotomalala Lovasoa Feno Fanantenana, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Liva Graffin Rakotoarimanana, and Zely Arivelo Randriamanantany, "Harnessing Madagascar's Solar potential: Global Horizontal Irradiance forecasting in Mahajanga using Satellite Data Analysis," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 17-23, 2025.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

B. Data Preprocessing

An inspection of the used datasets is involved, we spot that the data are worst presented, alphanumerical and string are mixed on. For that, data separation and cleaning steps are applied. These steps are very important to have reliable datasets because unclean data can decline the efficiency of model accuracies.

1- Organizing dataset

It is a method of transforming the data received with separate column and rows for enhancing understanding, visualizing and dealing with the data.

2- Input data standardization

Input data scaling, also known as normalization and standardization, is a very useful implementation when applying machine learning. The importance of standardization is mainly to avoid the possible incompatibility of numeric values with different units. Additional significant feature is to overcome numerical difficulties during computation processes and to minimize the RMSE for an efficient model. Each data attribute is linearly standardized to the range: 0 to 1, using equation:

$$X^{(i,j)} = \frac{x^{(i,j)} - \mu_x}{\sigma_x}, i = 1, \dots, n \; et \; j = 1, \dots, n \quad (1)$$

Where $X\ (i,j)$: Standardized value of a variable, $x\ (i,j)$ is the value of an explanatory variable (j=3), μ_x is the mean of the variable $x\ (i)$ et σx: is the standard deviation of the variable $x\ (i)$.

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x^{(i,j)} \; et \; \sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x^{(i,j)} - \mu_x)^2} \quad (2)$$

3- Visualizing dataset

A significant quantity of data is used for the model (around 40000 rows), a detailed analysis of each data is very difficult to achieve consequently a recent technique is adopted to visualize the assemble of each column data by Boxplot and Frequency distribution plot (Fig. 2). Using redundant data to train and test the algorithms would make the model inefficient with important score and error as well as increase the time of executing the algorithms. The data are shown to be homogeneous and compact in the boxplot. The uniformity of the data is shown by its frequency, if an outliers appeared on, the frequency plot is able to illustrate it. Fig. 2 shows a sample Boxplot and Distribution Frequency by GHI.

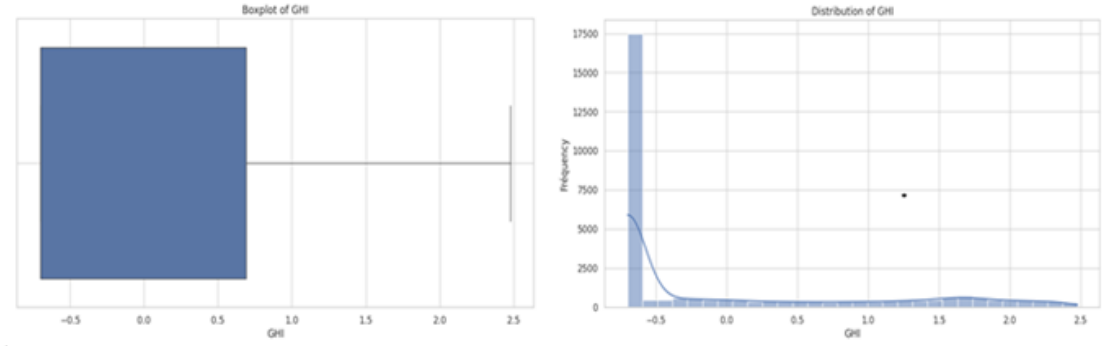

Figure 2: Boxplot and Frequency distribution of GHI dataset

4- Features construction/selection

The study's resolution is to forecast the GHI, many features with GHI were evaluated to determine which one generates the best results. As seen in figure 3, an experimental approach architecture was created to identify elements that improve the accuracy of solar radiation forecasting. Our architecture comprises two main steps as follows: First, the feature importance for our dataset was computed. Second, the effect of the number of features was investigated. In this phase, each set of classifier takes different assemble of features. These matrix correlations (fig.3) demonstrate how each feature relationship level takes effect on the target which is GHI. By definition, a Correlation Coefficient (CC) evaluates the relation between every single feature data. The CC evaluation gave a range of value: [-1;1]. The features with high correlation have the maximum value: 1, in fact, worst correlation is negative or null. In our case of study, we use the attribute which has high correlation with GHI (sixth columns on the matrix). The CC choice here is from 0,91 to 0,99. Then Clear sky GHI, Clear sky BHI, Clear sky BNI, TOA, BHI and BNI. These six (06) attributes are selected for the prediction of GHI with the enumerated models below.

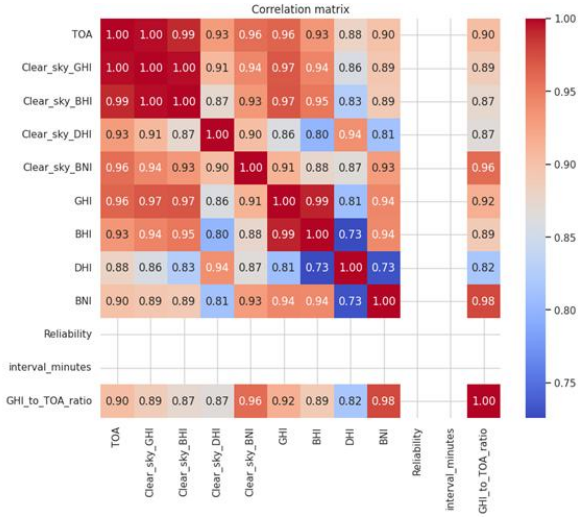

Figure 3: Parameters correlation matrix

Rakotomalala Lovasoa Feno Fanantenana, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Liva Graffin Rakotoarimanana, and Zely Arivelo Randriamanantany, "Harnessing Madagascar's Solar potential: Global Horizontal Irradiance forecasting in Mahajanga using Satellite Data Analysis," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 17-23, 2025.

5- Separating data into training and testing

Modelling was performed on Google Colab using Jupyter Notebook and Python 3.8. The training assemble was 80%, while 20% for the testing assemble. To analyze the performance of each parameter importance method, the results of each model have been expressed in terms of precision by the RMSE and the score R^2. And then, the experimental results were inspected in follow.

II-2 Machine Learning Models

1- Artificial neural networks

Artificial neural networks are generalizations of mathematical models [9]. As illustrated in Fig. 4, they have a parallel and distributed processing structure [10]. Signals are transmitted between neurons through these connections, with each link applying a weight to the transmitted input. Each neuron uses an activation function often non-linear to transform its net input into an output signal.

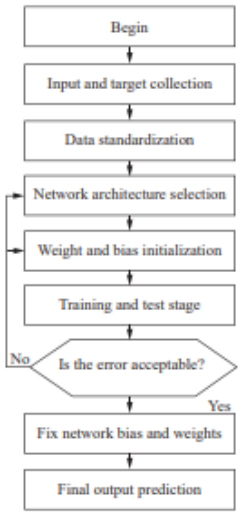

Figure 4: Training Process of an Artificial Neural Network

This activation function two (2) in the hidden layer enables the model outputs to fall within a range between 0 and 1, which is particularly relevant for binary classification problems as well as for activating neurons within each layer of the network.

$$P(X) = P_{max} \times \frac{1}{1-e^{-k(X-X_c)}} \quad (3)$$

Where X is the input values (TOA, Clear sky GHI, Clear sky BHI, BHI, BNI), P_{max} is the maximum standardized value observed in the data, X_c is the point at which the standardized power reaches approximately 50% of P_{max} ,and k controls the rate of increase.

2- Decision tree

A decision tree is a hierarchical structure that partitions data based on information criteria. It is built recursively by selecting the best feature at each node to maximize class separation [11]. Let X be a discrete random variable taking n values $x_1, \dots, x_2$with corresponding probabilities $p_1, \dots, p_n$. The entropy of X, usually denoted $H_b(S)$, is defined as the following quantity:

$$H_b(S) = -\sum_{i=1}^{n} p_i log_i(p_i) \quad (4)$$

Where b denotes the base of the logarithm (most often, $b = 2$).

The tree learns to predict Y (GHI) by splitting the data based on the following variables: TOA, Clear sky GHI, Clear sky BHI, BHI and BNI.

3- Random Forest

Random Forest is an ensemble of decision trees built from data subsamples. Each tree votes for a final prediction, which reduces the risk of overfitting and improves the model's robustness [12].

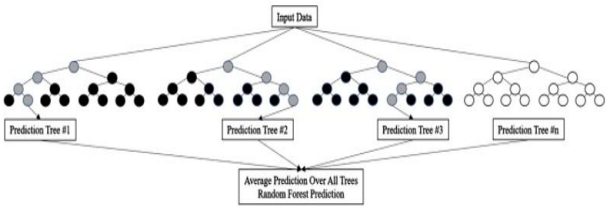

Figure 5: Random Forest Logic

Random Forest combines tree bagging and feature sampling.

$$\hat{P}_{bag}(X) = \frac{1}{B}\sum_{b=1}^{B} \hat{P}^{(b)}(\mathrm{X}) \quad (5)$$

Where X is the input values (TOA, Clear sky GHI, Clear sky BHI, BHI, BNI), B is the number of decision trees in the random forest, $\hat{P}^{(b)}(X)$ is the active power prediction made by the $b-th$ tree, and $\hat{P}_{bag}(X)$ is the final prediction obtained by averaging the predictions from all trees.

4- LightGBM (Light Gradient Boosting Machine)

LightGBM is a gradient boosting algorithm optimized for speed and performance. It builds decision trees in a leaf-wise rather than a depth-wise manner, which improves convergence and accuracy while maintaining low memory usage [13].

$$\hat{F} = \underset{F}{\operatorname{argmin}}\, \mathbb{E}_{x,y}[L(y, F(x)] \quad (6)$$

$\hat{F}$ is the optimal function that best approximates the relationship between inputs x and output y. $L(y, F(x)$ is the loss function that quantifies the difference between the prediction F(x) and the actual value of y.

$$\hat{P}(X) = \sum_{m=1}^{M} \omega_m \cdot h_m(X) \quad (7)$$

Where X is the input values (TOA, Clear sky GHI, Clear sky BHI, BHI, BNI), $\hat{P}(X)$is the predicted GHI, M is the total number of trees in the model, ω_m is the weight assigned to tree m, and $h_m(X)$ is the prediction made by tree m.

II-3 Ensemble learning Models comparison and choice

In this step, after data analysis, variety of models and tree-based classifiers including Linear Regression, Artificial Neural Network(ANN), Decision Tree, K-Nearest Neighbours (KNN), Random Forest, Long Short Term Memory(LSTM) with XGBoost /Feature engineering, Support Vector Machine(SVM), DBSCAN, and Light Gradient Boosting Machine were used to test the model performance with the six (06) influent parameters selected with the target, GHI. Secondly, the effect of the number of features was investigated. In this phase, each parameter obtained in the preceding step integrates all the models in order to obtain the highest accuracy. Then the results by fig.4

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

demonstrates the performance of each model tested with the data parameters within the prediction of GHI. The comparison of the model tested shows the RMSE (%). These values obtained describes in how the model familiarizes the data to achieve the finest forecasting by the real data of GHI. Our choice is to practice the models with RMSE under 5.5 %. Then, ANN, Decision Tree, Random Forest and Light Gradient Boosting Machine are the highest accuracy model to perform for an efficient results of forecasting. Figure 4 depicts this detailed information.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}} \quad (8)$$

Where y_i represents the observed values, $\hat{y}_i$ the predicted values, and n the number of data points.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (9)$$

Where R^2 is the coefficient of determination, $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the sum of squared residuals, and $\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the total sum of squares.

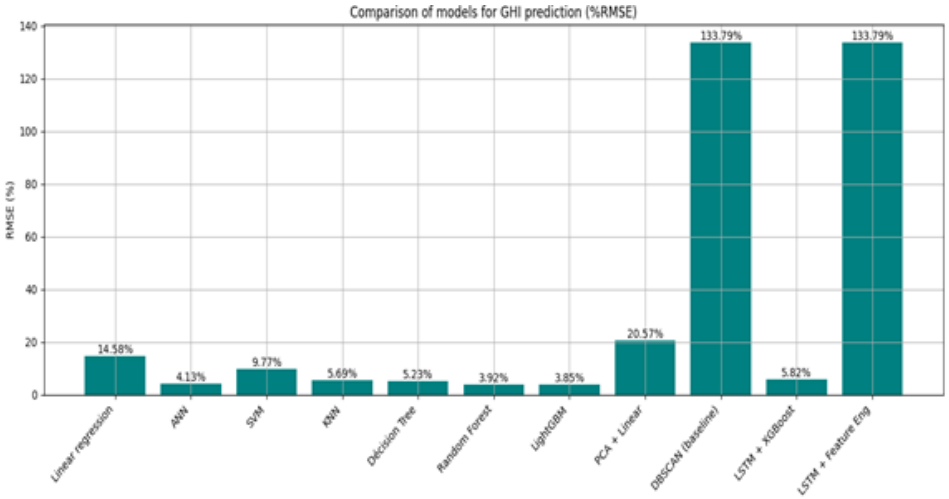

Figure 6 : RMSE (%) comparison for the ensemble learning models

The following table 1 presents the performance metrics of four regression models using R^2 and RMSE as evaluation criteria. LightGBM and Random Forest achieve the highest R^2 values, indicating excellent predictive accuracy. However, LightGBM shows the lowest RMSE, making it the most precise model overall.

TABLE 1: The performance metrics of four regression models

Models	R^2	***RMSE***
LightGBM	0.9992	2.46
Random Forest	0.9992	2.49
Decision Tree	0.9991	2.58
ANN	0.9984	3.42

IV. Results and Discussions

The evaluation highlights that four models ANN, Decision Tree, Random Forest and LightGBM, and consistently achieve the lowest RMSE values, indicating their superior accuracy and reliability in predicting Global Horizontal Irradiance (GHI). These models outperform others by maintaining RMSE values well below 6%, reflecting their capacity to capture complex data patterns and minimize prediction errors. In contrast to models like DBSCAN and LSTM + Feature Engineering, which yield extremely high RMSE values (over 133%), these four models demonstrate robustness and consistency, making them the most effective for precise GHI forecasting.

1. Artificial Neuronal Network (ANN)

The fig. 7 shows the alignment between red (predicted) and blue (real) points across all samples shows that the ANN captures the variability of GHI with high precision. The three-layer ANN (64 → 32 → 1) achieves reliable predictions while avoiding overfitting, demonstrating its suitability for nonlinear meteorological data. The consistency of predictions over 10,000 test samples indicates strong generalization capacity and statistical reliability. The red markers densely overlap the blue ones throughout the graph, reinforcing that the ANN maintains prediction accuracy across diverse solar conditions. The quality of prediction supports its application in real-time GHI estimation for solar energy systems and grid planning.

2. Decision Tree

The red points (predicted) follow the general spread of the blue points (real) (fig. 8), showing that the model captures key patterns of solar irradiance. The predictions cover the full range of GHI values, suggesting that the model does not collapse into biased outputs. Across 10,000 samples, the overlap between predicted and real values appears dense, indicating that the model remains stable on large datasets. While many red points align closely with blue ones, the overall spread shows local deviations, which is typical of decision tree behavior with high-variance data. Unlike ANN, the decision tree provides more explainable decision boundaries, but may struggle with the continuous, nonlinear fluctuations in GHI.

3. Random Forest

Rakotomalala Lovasoa Feno Fanantenana, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Liva Graffin Rakotoarimanana, and Zely Arivelo Randriamanantany, "Harnessing Madagascar's Solar potential: Global Horizontal Irradiance forecasting in Mahajanga using Satellite Data Analysis," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 17-23, 2025.

The red points (predicted) show a closer fit to the blue points (real) compared to the single decision tree (fig. 9), indicating enhanced accuracy. The aggregation of multiple trees allows the model to generalize better and smooth out local fluctuations in the data. Over 10,000 test samples, the predicted values maintain strong coherence with actual measurements, confirming model robustness. The clustering of red points over blue ones across the full GHI range highlights the model's ability to follow complex solar irradiance patterns. While more complex than a single tree, the Random Forest still provides a level of interpretability and avoids overfitting, making it a practical choice for real-world energy forecasting.

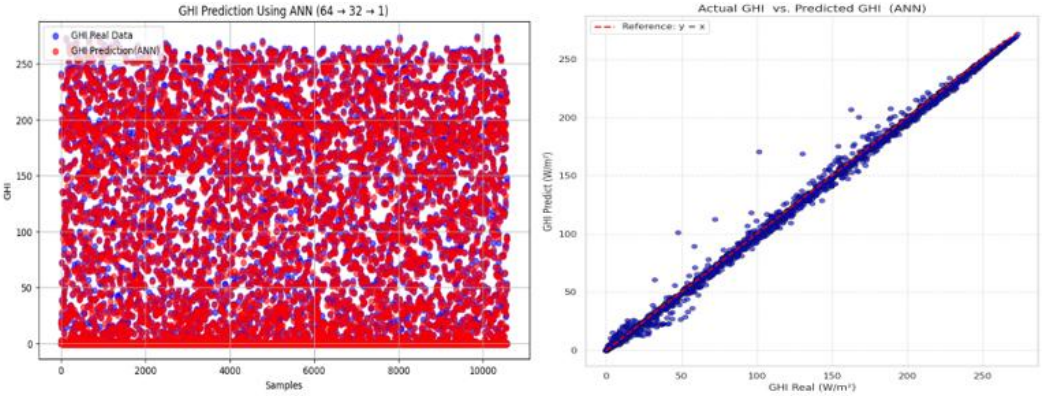

Figure 7: Models 001 GHI Prediction using ANN

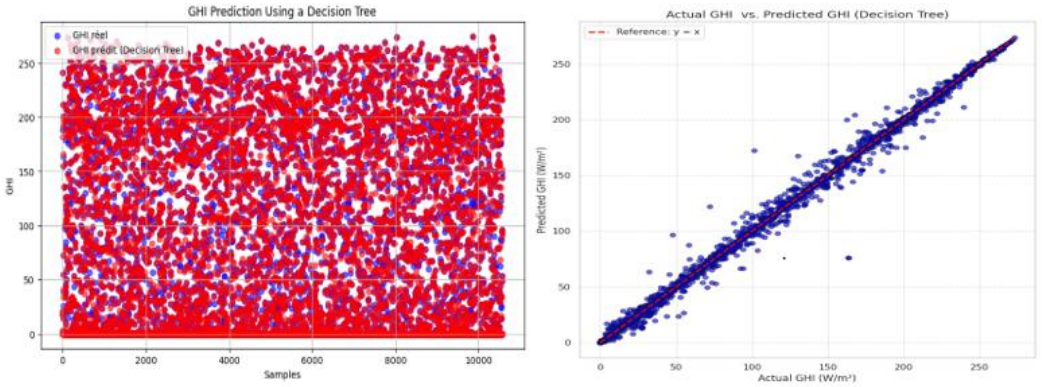

Figure 8: Model 002 GHI prediction using a Decision Tree

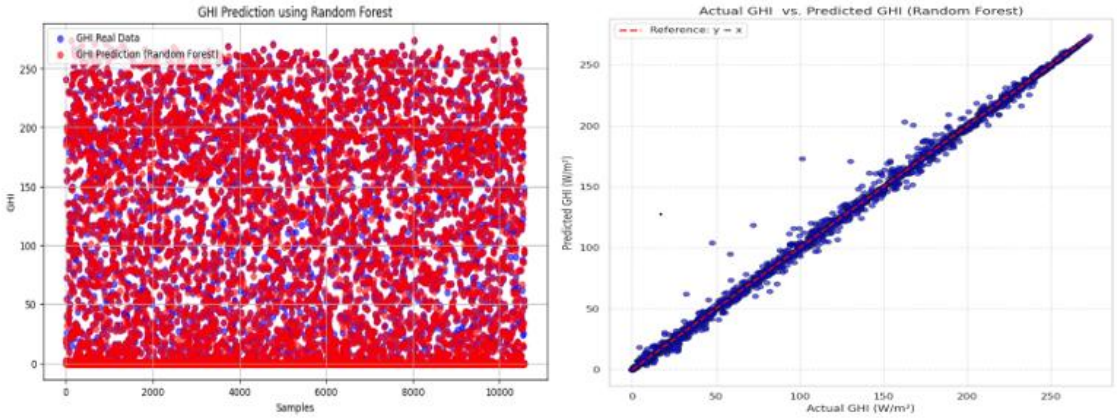

Figure 9: Model 003 GHI Prediction using Random Forest

4. LightGBM

The fig. 10 illustrates a comparison between the actual GHI (Global Horizontal Irradiance) values and those predicted by the LightGBM model across a dataset of over 10,000 samples. The blue dots represent the real data, while the red dots show the model's predictions. The strong overlap of the red dots on top of the blue ones along the x-axis (samples) indicates a high predictive accuracy of the LightGBM model. The visual alignment between both assemble of points shows that the model successfully captures the trend and variability of the

Rakotomalala Lovasoa Feno Fanantenana, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Liva Graffin Rakotoarimanana, and Zely Arivelo Randriamanantany, "Harnessing Madagascar's Solar potential: Global Horizontal Irradiance forecasting in Mahajanga using Satellite Data Analysis," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 17-23, 2025.

observed data, even under conditions of high dispersion. Furthermore, the consistent distribution of prediction errors across the entire sample range suggests that the model is robust and generalizes well, with no apparent overfitting. This behavior highlights the effectiveness of the LightGBM model for the regression task of GHI prediction based on the given input variables, and supports its use in large-scale solar forecasting scenarios.

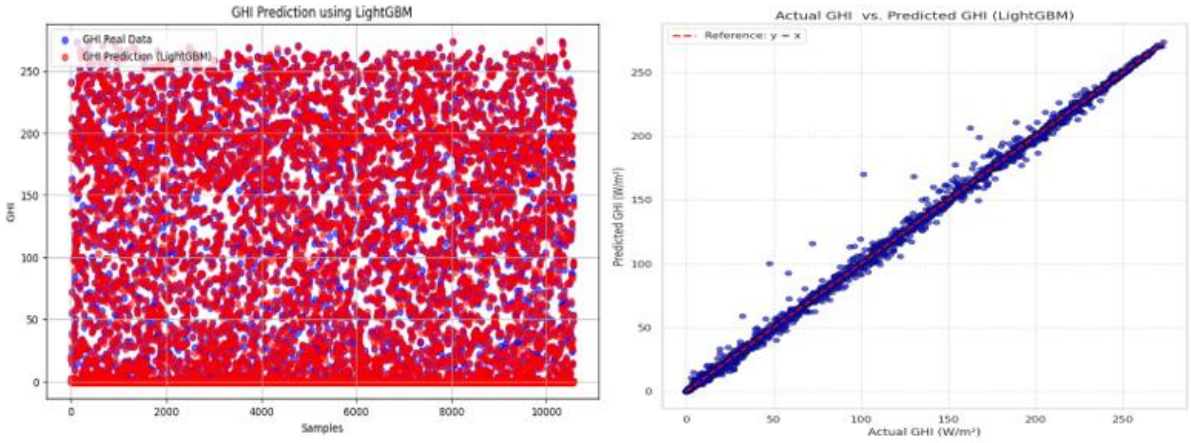

Figure 10: Models 004 GHI Prediction using LightGBM

V. Conclusion

This study effectively demonstrates the substantial solar energy potential of the Mahajanga region by leveraging satellite-derived data in combination with advanced machine learning techniques to produce accurate forecasts of Global Horizontal Irradiance (GHI). A rigorous selection of six highly correlated features-including TOA, BHI, BNI, and clear-sky GHI-enabled the integration of multiple algorithms to assess predictive performance under realistic conditions.

Among the nine machine learning models evaluated, four exhibited notably high accuracy, as measured by Root Mean Square Error (RMSE): LightGBM: 3.85% (highest accuracy achieved) Random Forest: 3.92% Artificial Neural Network (ANN): 4.13% Decision Tree: 5.23%.

These results underscore the superior performance of optimized ensemble methods, particularly LightGBM and Random Forest, in terms of robustness, generalization capacity, and their ability to model nonlinear meteorological data. The consistent accuracy of LightGBM confirms its relevance for real-time GHI forecasting and positions it as a powerful tool for intelligent and sustainable solar energy planning.

The proposed methodology provides a strategic framework for optimizing photovoltaic system design, supports the integration of renewable energy into smart grid infrastructures, and contributes to broader energy transition initiatives in Madagascar and other tropical regions with high solar potential.

References

[1] Quansah D. A., Addy J. M., Adaramola M. S., *Solar irradiance forecasting models: State-of-the-art and future prospects*, Renewable and Sustainable Energy Reviews, 152, 111707, 2021.

[2] Y. Liu, "*Solar GHI Ensemble Prediction Based on a Meteorological Model and Method Kalman Filter*", China, 2022, *Advances in Meteorology*, pp. 1–17.

[3] S. Boubaker, M. Benghanem, A. Mellit, A. Lefza, O. Kahouli, L. Kolsi, "*Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia*", Saudi Arabia, 2021, *IEEE Access*, pp. 36719–36729.

[4] M.K. Boutahir, Y. Farhaoui, M. Azrour, I. Zeroual, A. El Allaoui, "*Effect of Feature Selection on the Prediction of Direct Normal Irradiance*", Morocco, 2022, *Big Data Mining and Analytics*, vol. 5, no. 4, pp. 309–317.

[5] M. Pi, N. Jin, D. Chen, B. Lou, "*Short-Term Solar Irradiance Prediction Based on Multichannel LSTM Neural Networks Using Edge-Based IoT System*", China, 2022, *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2372748, 11 pages.

[6] J. Huertas-Tato, R. Aler, F.J. Rodríguez-Benítez, C. Arbizu-Barrena, D. Pozo-Vázquez, I.M. Galván, "*Predicting Global Irradiance Combining Forecasting Models Through Machine Learning*", Spain, 2018, HAIS: Hybrid Artificial Intelligent Systems, pp. 622–633.

[7] F. Nawab, A.S. Abd Hamid, A. Ibrahim, K. Sopian, A. Fazlizan, M.F. Fauzan, "*Solar irradiation prediction using empirical and artificial intelligence methods: A comparative review*", Malaysia, 2023, Heliyon 9, e17038.

[8] D. Albahdal, M. Almousa, W. Aljebreen, A.A. Almutairi, "*Sunrise in the Desert: Leveraging Big Data Analytics for Predictive Solar Energy Production in Saudi Arabia*", Saudi Arabia, 2025, IEEE Access, pp. 54585–54598

[9] Haykin S., *Neural Networks and Learning Machines*, 3rd ed., Pearson, 2009.

[10] Buturache A.-N., Stancu S., « *Wind Energy Prediction Using Machine Learning* », Low Carbon Economy, 12(1), 1-21, 2021.

[11] Haykin S., *Neural Networks and Learning Machines*, 3rd ed., Pearson, 2009.

[12] T. K. Ho, « *The Random Subspace Method for Constructing Decision Forests* », IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8), 832-844, 1998.

[13] Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y., « *LightGBM: A Highly Efficient Gradient Boosting Decision Tree* », Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 3149-3157, 2017.