



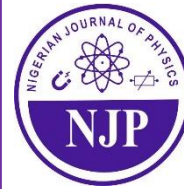
Nigerian Journal of Physics (NJP)

ISSN online: 3027-0936

ISSN print: 1595-0611

DOI: <https://doi.org/10.62292/njp.v35i2.2026.503>

Volume 35(2), June 2026



## Day-Ahead Hourly Forecasting of Solar Radiation using a Physics-based Hybrid Machine Learning Model in Some Selected Locations in Nigeria

\*Taofeek A. Otunla and Emmanuel O. Ayegboyin

Atmospheric Research Group, Department of Physics, University of Ibadan, Oyo State, Nigeria.

\*Corresponding Author's Email: [biodunotunla@gmail.com](mailto:biodunotunla@gmail.com)



### ABSTRACT

Day-ahead hourly forecasting of Solar Radiation (SR) is crucial for optimizing renewable energy generation, grid stability, and storage management, especially for locations with high solar potential but low utilization. This study develops a stacked hybrid forecasting model that combines statistical and machine learning approaches to forecast day-ahead hourly SR across eight Nigerian locations. Relevant historical meteorological data spanning one to two years. The datasets were divided into five temporal scales, consisting of annual, pre-rainy(MAM), rainy(JJA), post-rainy(SON), and dry(DJF) seasons. Three baseline models—Seasonal Auto-regressive Integrated Moving Average with exogenous variables (SARIMAX), Long Short-Term Memory (LSTM) networks, and Extreme Gradient Boosting (XGBoost) were developed by using 70:20 of the datasets in each temporal scale for the purpose of training and testing respectively and the remaining 10% for final evaluation of the hybridized linear regression stacked model. The performances of the models were evaluated using Root Mean Square Error (RMSE), Mean Bias Error (MBE), and Coefficient of Determination ( $R^2$ ). All the six possible combinations of the models were subjected to Diebold-Mariano(DM) test of significance. The RMSE together with DM were also carefully combined to rank the models. The results showed that while SARIMAX( $R^2$ : 0.60 to 0.88) captured seasonal patterns, it consistently underperformed against the machine learning models. LSTM excelled ( $R^2$ : 0.76 to 0.92) for datasets where clearer temporal patterns exist, and XGBoost was more effective( $R^2$ : 0.75 to 0.90) under irregular or noisy conditions. The hybrid model demonstrated the most robust and reliable performance overall( $R^2$ : 0.68 to 0.92) with potential to reduce bias (MBE:-7.24 to 17.21 W/m<sup>2</sup>). The performance of the model were at their best in the DJF ( $R^2$ : 0.90 to 0.92) for most of the locations, while MAM and SON were over- and under predicted respectively for most locations due to higher cloud-cover fluctuations. XGBoost and Hybrid models delivered the best performance with overall rank-score of two and four respectively.

### Keywords:

Stacked hybrid model,  
SARIMAX,  
LSTM,  
XGBoost.

### INTRODUCTION

Solar radiation (SR) is a primary component of renewable energy systems, influencing photovoltaic (PV) power generation, solar thermal applications, and climate studies. Accurate prediction of SR is crucial for optimizing energy production, grid control, and energy storage systems (Qing & Niu, 2018). Solar radiation is also crucial in agriculture by influencing the assimilation of nutrients by plants, as well as the growth of the plants. The growth, leaf size, and the development of flowers and fruits in plants are influenced by both the amount and

quality of sunlight they receive (He *et al.*, 2019). In addition, the moisture content of the soil and temperature can also be affected by SR, which in turn affects plant growth and other agricultural practices like irrigation, sowing, harvest times, etc. (Onwuka, 2018). However, SR is highly variable due to atmospheric conditions, geographical location, and seasonal changes, making its prediction a complex challenge (Besharat *et al.*, 2013) Nigeria receives a substantial amount of SR daily. On average, about 4.4 to 7.0 kWh/m<sup>2</sup>/day of SR, with the highest levels in the northern regions (e.g., Sokoto,

Maiduguri) and slightly lower but still substantial radiation in the southern regions (e.g., Lagos, Port Harcourt) (Oyedepo, 2012). These values far exceeded the electricity demand of the country. However, less than 1% of this solar potential has been tapped through the national grid (International Energy Agency, 2022), possibly due to spatio-temporal variations in solar radiation and other myriad environmental factors (Julius & Balogun, 2022). Nigeria is located in a tropical climate where solar irradiance can fluctuate significantly within an hour due to high cloudiness, turbidity, and atmospheric dynamics, making daily or monthly forecasts inadequate for real-time decisions.

Day-ahead hourly forecasting of SR is especially essential in some applications. For instance, it enables grid operators and solar farm managers to: one, schedule electricity generation more efficiently, two, make informed use of energy storage systems, and three, maintain a stable balance between electricity supply and demand. Besides, in rural locations where grid electricity is either unreliable or completely absent, day-ahead hourly solar forecasting can guide the development of cost-effective mini-grids, providing a sustainable alternative to conventional power. At the national level, day-ahead hourly forecasting of SR is also important, especially where the national grid is fragile and chronically under-powered. Hence, the ability to anticipate fluctuations caused by environmental factors can help stabilize supply. This will not only support higher integration of renewable energy but will also reduce the significant costs associated with balancing demand and generation (Ohijeagbon *et al.*, 2025; NERC, 2022; Agbo *et al.*, 2021). Knowledge of day-ahead SR is also important in irrigation schedules.

Solar radiation modelling is a very active area of research in Nigeria, and several approaches have been used to predict SR. These approaches include: empirical, parametric, statistical, and machine learning methods. The empirical method that leverages the correlation between SR and extraterrestrial radiation, sunshine duration, cloud-cover, and air temperature has been widely used, and the method works well for the daily and monthly time scale (Besharat *et al.*, 2013; Otunla & Koleaje, 2015; Mujabar & Chintaginjala, 2021; Ogbaka *et al.*, 2020; Davou & Umar, 2024). For instance, Otunla (2020) improved on angstrom-based modelling of daily SR by partitioning solar radiation into clear and all-sky radiation. The clear-sky transmissivity was modelled using a parametric approach, while cloud transmissivity was modelled using an empirical approach. This approach was an improvement over widely accepted empirical models for the locations used in the study. Auto-regressive Integrated Moving Average (ARIMA) for linearly correlated time series data has been used in a statistical approach of SR modelling (Alsharif *et al.*, 2019; Benitez *et al.*, 2023). Seasonal Auto-regressive

Integrated Moving Average (SARIMA) and its variant, SARIMAX, which include exogenous variables, were indicated to be superior to ARIMA when time series data are characterised by trends and seasonality (Aggarwal & Saini, 2014; Benitez *et al.*, 2023). Machine learning techniques emerged as powerful tools for improving the accuracy of SR forecasts over empirical and statistical methods by leveraging on historical meteorological data to develop predictive algorithms capable of capturing complex patterns in SR variability that usually characterised high resolution time series data (Mukhtar *et al.*, 2022; Belaid & Mellit, 2016; Isma'il & Aliyu, 2023; Ali *et al.*, 2023). Lamidi *et al.* (2024) reviewed 72 articles on machine learning techniques for SR prediction and identified Long Short-Term Memory (LSTM), Support Vector Machines (SVMs), and Random Forest as the top-performing algorithms. In a study by Alabi & Ojo (2024), meteorological data of 31 years from six Nigerian cities: Sokoto, Maiduguri, Ilorin, Ikeja, Enugu, and Port Harcourt were used to develop a hybrid model from Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures to forecast SR. The results indicated that the hybrid model outperforms each of the standalone models, with a high degree of alignment and modest prediction errors. Kassem *et al.* (2023) applied a linear-nonlinear hybrid model to forecast SR using meteorological data in 85 cities across Africa, including 13 cities from Nigeria. The results indicated that the suggested linear-nonlinear hybrid models outperformed all other models in terms of prediction accuracy.

Most Nigerian studies were based on daily and monthly time scales, possibly due to the scarcity of relevant meteorological data, with very limited attention given to hourly forecasting (Ohijeagbon *et al.*, 2025). Recent studies, such as the Ibadan irradiance study by Urhuerhi *et al.* (2025) and the deep learning models applied by Tajjour *et al.* (2023), highlighted the potential of short-term forecasting. This present study aimed to address this gap by developing a hybrid model that is both robust and accurate in forecasting SR for some selected locations in Nigeria. The stable and better interpretability of linear stacking method were explored over that of weighted averaging and meta-learner stacking methods in the study.

## MATERIALS AND METHODS

### Study Locations and Datasets

The study locations are Yola, Enugu, Nsukka, Bauchi, Lagos, Akure, Anyigba, and Abuja. fall within different climatic zones of Nigeria, reflecting the country's diverse weather patterns. Yola lies in the Sudan savanna zone, which is marked by semi-arid conditions and longer hours of sunshine. Enugu and Nsukka are both situated in the tropical rainforest zone, where heavy rainfall and frequent cloud cover influence solar availability. Bauchi represents the Guinea savanna zone with a transitional

climate with moderate rainfall and abundant sunshine. Lagos is located in the mangrove–rainforest transition zone along the coast, where a humid maritime climate and persistent cloudiness prevail. Akure also lies in the tropical rainforest belt, with high rainfall and distinct seasonal variation. Anyigba and Abuja are part of the Guinea savanna zone in north-central Nigeria, characterized by a balance of wet and dry seasons. The meteorological data used in this study are from the Center for Atmospheric Research of the National Space Research and Development Agency (CAR- NASRDA), Federal Ministry of Science and Technology, Anyigba,

Nigeria. One year of air temperature, relative humidity, and solar radiation data were obtained for each of the eight locations across Nigeria for the years ranging between 2009 and 2014. The original data were recorded at 5-minute intervals and resampled to hourly data to reduce noise and computational complexity.

The data set sequences were split into training, testing, and validation sets using a 70:20:10 ratio, respectively, for the linear regression model that hybridizes the base models and 70:20 of training and testing, respectively, for the base models.

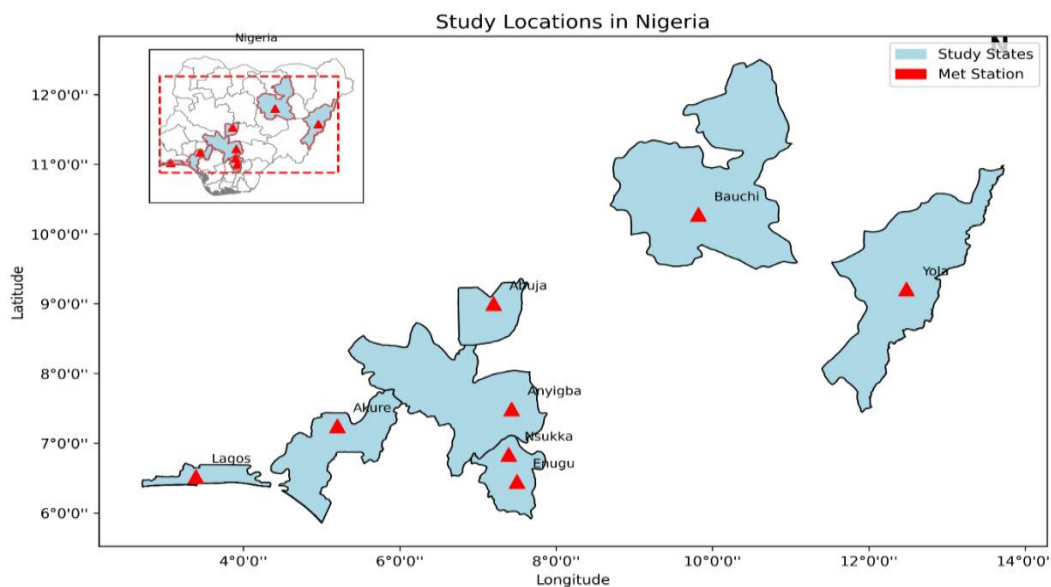


Figure 1: Map of Selected Locations in Nigeria

Across different geographical locations and climates, for instance, China (Zhu *et al.*,2019); Taiwan (Zhong &Wu, 2020 ); Germany & Australia (Santos *et al.*,2022), and India(Sharma *et al.*, 2022), SARIMAX, LSTM, and XGBoost models were indicated to be superior to other machine learning and statistical models. Hence, a linear regression combination of these baseline models was developed, and comparisons were made with the contributing baseline models using: one, one year of hourly data for each location, and two, three months of hourly data for each season, labelled as: pre-rainy(March-April-May (MAM)); rainy (June-July-August(JJA)); post-rainy (September-October-November(SON)), and dry season (December-January-February(DJF)).

**Features Engineering**

Hourly extraterrestrial radiation,  $I_{0h}$ , relative air mass,  $m_r$ , and absolute air mass,  $m_a$  were added as part of the features in the model engineering processes. Hourly extraterrestrial radiation is computed using Duffie and Beckman (2013) as:

$$I_{0h} = \frac{12 \cdot 3600}{\pi} G_{sc} d_r [\cos\varphi \cos\delta (\sin\omega_2 - \sin\omega_1) + (\omega_2 - \omega_1) \sin\varphi \sin\delta].$$

$$G_{sc} = 1367 \text{ W/m}^2;$$

$$\omega = 15(\text{local solar time} - 12);$$

$$d_r = 1 + 0.033 \cos\left(\frac{360N}{365}\right);$$

$$\delta = 23.45 \sin\left(\frac{360(284+N)}{365}\right).$$

Where  $G_{sc}$  is Solar constant, N is day of the year,  $\omega_1, \omega_2$  are hour angles in the unit of degree at the beginning and end of the hour,  $d_r$  is Earth-Sun’s distance correction factor,  $\varphi$  is latitude in radians;  $\delta$  is declination angle in degree,

Relative air mass,  $m_r$ , is engineered using solar zenith angle according to Kasten and Young (1990) as:

$$m_r = \frac{1}{\cos\theta_z + 0.50572(96.07995 - \theta_z)^{-1.6364}},$$

where  $\theta_z$  is solar zenith angle.

Absolute air  $m_a$  includes the effect of altitude through air pressure  $P$  as:  $m_a = m_r \left(\frac{P}{P_0}\right)$ , where  $P_0=101.325 \text{ kPa}$

**Normalization**

Feature variables were rescaled into the range [0,1] where necessary through the application of Min-Max normalization. This is necessary to ensure that all features' contributions are proportionate during model training.

**SARIMAX**

SARIMAX is a Seasonal variation of the Auto-regressive Integrated Moving Average (ARIMA) model with exogenous variables related to the response variable in the regression model. The SARIMAX time series model has a general multiplicative shape of the form: SARIMA(p,d,q)x(P, D, Q,s) (Hipel & McLeod, 1994). Where p,d, and q are the non-seasonal auto-regressive, differencing, and moving average orders, respectively. P, D, and Q are the seasonal auto-regressive, differencing, and moving average orders. The values of p,d,q, and P,

D, Q parameters were selected from a set of values using the pmdarima. arima.auto\_arima library in Python and they are listed in Table 1. In order to avoid poor forecasting accuracy in SARIMAX model, the stationarity of the datasets was tested using combination of the Augmented Dickey–Fuller (ADF), Phillips–Perron (PP), and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. The results of the ADF and PP tests indicated that the data were stationary, while the results from the KPSS test were a contradiction (non-stationary). This disagreement between KPSS and the other two tests indicates the presence of trend and seasonality in the datasets. The exogenous variables that feature in the model consist of both meteorological and engineering variables, which are: air temperature, relative humidity, air pressure, hourly extraterrestrial radiation, absolute air mass, and relative air mass.

**Table 1: Summary of Hyperparameters in SARIMAX Model in all the Study Locations**

Hyper-parameters	Abuja	Akure	Anyigba	Enugu	Lagos	Nsukka	Yola	Bauchi
P	1	1	1	1	1	1	0	0
D	0	0	0	0	0	0	0	0
Q	1	1	1	1	3	1	1	1
P	1	1	1	1	1	1	2	2
D	0	0	0	0	0	0	0	0
Q	2	2	2	2	2	2	2	2

**Extreme Gradient Boosting (XGBoost)**

XGBoost is a classification and regression algorithm based on the Gradient Boosting Decision Tree (Li *et al.*, 2022). XGBoost is designed to combine many simple models, usually decision trees, into a stronger predictive model. Each new tree will be trained in order to correct the mistakes made by the previous ones, which helps the overall model to forecast accurately. XGBoost uses a regularization process while optimizing its predictions, which prevents over-fitting and improves performance. It is also designed to run very efficiently, even on large

datasets, and often outperforms traditional machine learning methods, making it well-suited for time series forecasting. The XGBoost model features include all the variables that were used in SARIMAX, together with lag and rolling means of meteorological variables. The variables were chosen to capture the non-linear and non-seasonal features in the hourly solar radiation. Random search algorithm with early stopping were used to generate the hyperparameters (Table 2) that were used to develop the model.

**Table 2: Summary of Hyperparameters in Extreme Gradient Boosting Model(XGBoost) for the Locations in Pre-rainy(MAM), Rainy(JJA), Post-rainy(SON) and Dry(DJF) Seasons**

Hyperparameters	Abuja	Akure	Anyigba	Enugu	Lagos	Nsukka	Yola	Bauchi
Pre-rainy								
subsample	0.8	0.8	0.8	0.8	0.8	0.8	0.6	0.6
n_estimators	600	600	600	400	600	400	600	600
max_depth	4	4	4	4	4	4	3	4
learning_rate	0.1	0.1	0.1	0.01	0.1	0.01	0.01	0.1
colsample_bytree	1	1	1	1	1	1	0.6	0.8
Rainy								
subsample	0.6	0.8	1	0.6	0.6	0.6	1	0.6
n_estimators	600	400	400	600	200	600	600	600
max_depth	4	4	3	5	4	5	5	3
learning_rate	0.1	0.01	0.05	0.01	0.01	0.01	0.05	0.01
colsample_bytree	0.8	1	0.8	0.6	0.6	0.6	1	0.6

Hyperparameters	Abuja	Akure	Anyigba	Enugu	Lagos	Nsukka	Yola	Bauchi
Post rainy								
subsample	0.6	1	0.6	0.8	0.6	0.6	0.6	0.6
n_estimators	200	400	600	400	600	600	600	600
max_depth	4	3	3	4	4	4	4	4
learning_rate	0.01	0.05	0.01	0.01	0.1	0.1	0.1	0.1
colsample_bytree	0.6	0.8	0.6	1	0.8	0.8	0.8	0.8
Dry								
subsample	1	1	1	1	0.6	1	0.8	0.8
n_estimators	600	600	600	600	600	400	400	600
max_depth	5	5	5	4	3	3	4	4
learning_rate	0.05	0.05	0.05	0.1	0.1	0.05	0.01	0.1
colsample_bytree	1	1	1	0.8	0.6	0.8	1	1

### Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network (RNN) designed to process sequence and time series data. It can learn long-term dependencies, which enables it to solve the problem of vanishing gradients, thereby improving its forecasting accuracy (Ebtehaj & Bonakdari, 2024). LSTM is made up of three components, which are the input gate, forget gate, and output gate. The input gate takes in the information into the memory cell, the forget gate controls the removal of irrelevant information in the memory cell, and the output gate brings out the information in the memory cell. The mathematical formulation of LSTM is given in Gers *et al.* (2000) and Hochreiter & Schmidhuber (1997). The variables that were featured in the model include all the variables that were used in the XGBoost model, together with cyclic encoding of hour and daily sine and cosine angles. The hourly and daily cyclic features were encoded in the model to give a better representation of the diurnal and daily seasonal natures of the solar radiation data.

The following hyperparameters were implemented for all four seasons in all the stations: “epochs”=50; “batch\_size”=64; “Dense”=1; “verbose”=1; “optimizer”= “Adam”; “loss”= “MSE”. Early Stopping was also included.

### Hybrid Model

A hybrid model was developed to combine the strengths of each model to generate accurate and robust hourly forecasts. The predictions of the hybrid model were stacked using linear regression of the form:  $y_i = \beta_1 SARIMAX_i + \beta_2 XGBoost_i + \beta_3 LSTM_i + \beta_0$ . Where  $y_i$  is the forecast hourly solar radiation at the  $i$ -th observation;  $\beta_0$  is the Intercept of the regression;  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  are coefficients of SARIMAX, XGBOOST, and LSTM predictions, respectively, to the final stacked forecast.  $SARIMAX_i$ ,  $XGBoost_i$  and  $LSTM_i$  are the predicted values from the SARIMAX, XGBOOST, and LSTM models  $i$ th predictions, respectively.

### Evaluation Metrics

Three evaluation metrics were used in this study to evaluate the performance of the proposed model. They are: Root Mean Square Error (RMSE), Mean Bias Error (MBE), and Coefficient of Determination ( $R^2$ ). The smaller the value of RMSE, the better the forecast. The negative and positive signs of MBE indicate underestimation and overestimation, respectively. The higher the value of  $R^2$ , the better the fit between the measured and forecast values. The three performance metrics are given as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum (y - y_i)^2}$$

$$MBE = \frac{1}{n} \sum (y - y_i)$$

$$R^2 = 1 - \frac{\sigma_y^2}{\sigma_{y_i}^2}$$

Where  $y$  and  $y_i$  are the measured and model forecast values, respectively, and  $n$  is the number of data points.

Diebold-Mariano (DM) test of significance was used to adjudicate the better model between six possible pairs of the four models. It generates statistic that follows normal distribution which is given as:  $DM = \frac{d}{\sqrt{\frac{2\pi f_d(0)}{T}}}$ . Where  $d$

is mean of loss difference ( $d_t$ ) at time  $t$ .  $T$  is the total number of forecasts.  $f_d(0)$  is the spectral density of  $d_t$  at zero frequency that accounts for autocorrelation.  $d_t = e_{1,t}^2 - e_{2,t}^2$ , where  $e_{1,t}^2$  and  $e_{2,t}^2$  are the squared errors of the model 1 and model 2 forecasts respectively.

A combination of RMSE and DM for all seasons and locations of study was used to generate a comprehensive assessment of model accuracy and reliability.

### RESULTS AND DISCUSSION

Table 3 gives the results of the evaluation of hourly SR forecasting using SARIMAX, LSTM, XGBoost, and a stacked hybrid model across eight locations in Nigeria, based on one year of hourly data. The performance of the models is given by the error metrics of RMSE, MBE, and  $R^2$ . The results indicated that while SARIMAX captured the seasonal pattern of solar radiation reasonably well ( $R^2$ : 0.6 to 0.88), it underestimated in five locations

(MBE: -7.04 to -46.81 W/m<sup>2</sup>) and overestimated in three (MBE: 10.29 to 45.68 W/m<sup>2</sup>). This was as a consequence of SARIMAX's weakness in handling the non-linear and irregular variations that often occur in high-resolution hourly solar radiation data (Haider *et al.*, 2022). The coefficient of determination in both LSTM (R<sup>2</sup>: 0.76 to 0.92) and XGBoost (R<sup>2</sup>: 0.75 to 0.90) indicated better co-variations compared with SARIMAX. The two of them also have much lower error metrics (Table 4). Furthermore, LSTM gave the best results of the three baseline models in locations where clearer temporal patterns were present, such as Abuja, Bauchi, and Yola 2012 (Haider *et al.*, 2022; Lim *et al.*, 2022). By contrast, XGBoost performed better in locations where the data appeared more irregular or noisy, such as Anyigba, Lagos, and Enugu (Li *et al.*, 2022). These differences highlight how each model has distinct strengths: LSTM is better in learning sequential dependencies while XGBoost captures complex nonlinear relationships

better. The stacked hybrid model (R<sup>2</sup>: 0.68 to 0.92; RMSE: 54.13 to 108.62 W/m<sup>2</sup>) that linearly regressed the three baseline models (Table 5) offered the most robust performance by assigning greater weight to the best baseline model and, at the same time, factoring in the peculiarities of other models and by so doing avoiding overfitting. The stacked hybrid model generally showed the most reliable performance overall. In some locations, such as Nsukka, the hybrid model delivered clear improvements over the best single models, reducing both errors and bias. In some other locations, such as Anyigba and Yola, its performance is comparable with the best baseline model, indicating that it could remain competitive without being weaker. Importantly, the hybrid model reduced the bias in these locations, demonstrating its ability to balance out the over-prediction and under-prediction tendencies of the baseline models.

**Table 3: Metrics of SARIMAX, LSTM, XGBoost and Stacked Hybrid Models for all the Locations using One Year of Data**

Location	Year	Metrics	SARIMAX	XGBOOST	LSTM	Hybrid
Lagos	2009	RMSE	103.73	79.34	81.57	92.43
		MBE	-46.81	-1.54	9.96	-12.23
		R2	0.6	0.76	0.75	0.68
Akure	2012	RMSE	107.68	85.8	103.35	96.57
		MBE	-34.92	27.04	23.24	40.28
		R2	0.79	0.87	0.81	0.83
Nsukka	2009	RMSE	72.87	60.33	59.55	54.13
		MBE	10.29	-8.96	-0.01	-7.24
		R2	0.83	0.88	0.89	0.91
Enugu	2013	RMSE	209.52	91.8	75.13	82.37
		MBE	-26.01	7.79	6.8	6.88
		R2	0.26	0.86	0.9	0.89
Anyigba	2011	RMSE	96.77	78.28	89.35	78.87
		MBE	16.84	5.35	9.43	5.57
		R2	0.81	0.87	0.84	0.87
Abuja	2009	RMSE	128.42	99.4	94.96	99.8
		MBE	45.68	9.66	-13.73	10.36
		R2	0.77	0.86	0.88	0.86
Bauchi	2014	RMSE	155.68	116.02	82.46	108.62
		MBE	-11.58	13.87	-7.77	-17.71
		R2	0.64	0.8	0.9	0.82
Yola	2011	RMSE	94.67	76.57	98.39	76.56
		MBE	-7.04	7.22	15.94	6.1
		R2	0.88	0.92	0.87	0.92

The results of this one-year analysis indicated that integrating statistical and machine learning approaches improves the robustness of hourly solar radiation forecasts. While LSTM and XGBoost frequently outperformed SARIMAX on their own, the hybrid model produced more consistent accuracy across the locations. This suggests that the ensemble approach of stacking baseline models could be more suitable for hourly solar

radiation forecasting when the environment is diverse in seasonality and climatic conditions.

Five to six days of hourly plots of forecasted and measured hourly solar radiation were compared for three locations: Lagos, Anyigba, and Yola, each chosen to represent Coastal, Guinea and Sahel climates respectively during pre-rainy, rainy, post-rainy and dry seasons (Figures 2-5). Both measured and modelled

hourly solar radiation values were indicated to increase from the coastal to Sahel locations across all seasons. The effect of clouds, as indicated by fluctuations in solar radiation values, was captured by the measurements and the model. It reduces from the coastal to the Sahel locations, and also from the dry season (DJF) to the peak of the rainy season (JJA). The stacked Hybrid model, therefore, captured all essential features expected of both the climate location and season.

The performance metrics as indicated by RMSE and MBE between measured and forecasted hourly solar radiation were well below were well below 100 and 20 W/m<sup>2</sup> respectively for 25 out 32 combinations of location and season with exception of (Abuja, rainy), (Bauchi, rainy), (Bauchi, dry),(Enugu, rainy), (Enugu, dry), (Lagos, rainy) and (Yola, pre-rainy)(Figures 6 and 7 ) while R<sup>2</sup> (Figure 8) values were only below 0.7 at (Abuja, rainy), (Enugu, dry) and (Bauchi, dry).

The values of all the coefficients of the linear Hybrid model in all the locations used in the study are listed in

Table 4. The values of the coefficients ranged between -0.25603 and -0.01002 for SARIMAX; 0.923217 and 1.347482 for XGBoost; 0.451652 and -0.00013 of LSTM. The overall score for XGBoost is two and it is closely followed by the Hybrid model with overall score of four (Table 5). Thus, the XGBoost model mostly dominate the ensemble, accounting for the non-linearity and complexity of the hourly SR data. The values of the coefficients of SARIMAX and LSTM, which are mostly negative, together with the intercept values in each locations seemed to correct for biases and overfitting in the XGBoost model.

Thus, these results, which showed the robustness of the stacked hybrid model in characterising variations of hourly solar radiation across different climatic locations and seasons, together with the low error metrics, make the stacked hybrid model a model of choice for hourly forecasting of solar radiation.

**Table 4: Coefficients of SARIMAX( $\beta_1$ ), XGBoost ( $\beta_2$ ), LSTM( $\beta_3$ ) and Intercept ( $\beta_0$ ) in the Stacked Hybrid Model for Pre-rainy, Rainy, Post-rainy, and Dry seasons for all Eight Locations used in the Study**

Station	Season	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_0$
Abuja	PRE	-0.01105	1.011423	0.003096	-1.40379
	RAINY	-0.00439	1.004644	-0.00088	0.014386
	POST	-0.05705	1.323042	-0.19475	-17.2407
	DRY	-0.01285	1.041157	-0.02904	0.124209
Akure	PRE	-0.00586	1.029125	-0.02386	0.25664
	RAINY	0.059291	1.144286	-0.22163	-6.12008
	POST	-0.13257	1.171362	-0.06093	2.384663
Anyigba	PRE	-0.09957	1.098868	-0.00139	-3.73125
	RAINY	-0.01444	1.018309	-0.00373	-0.09751
	POST	-0.13518	1.108657	0.045267	-2.5995
Bauchi	PRE	-0.20559	1.347482	-0.1464	1.877502
	RAINY	-0.04126	1.078532	-0.04034	-0.9875
	POST	-0.01022	1.019192	-0.0077	-0.22245
Enugu	PRE	-0.01022	1.019192	-0.0077	-0.22245
	RAINY	-0.22134	1.154332	0.103422	-4.25754
	POST	-0.00428	1.013482	-0.00872	0.174132
Lagos	PRE	-0.00739	1.00837	-0.0003	-0.48016
	RAINY	-0.24064	0.923217	0.451652	-21.1203
	POST	-0.20924	1.236542	0.033532	-5.70142
Nsukka	PRE	-0.03814	1.094624	0.02077	-20.8257
	RAINY	-0.01002	1.008231	-0.00013	-0.10091
	POST	-0.00842	1.022837	-0.01252	-0.37257
Yola	PRE	-0.00034	1.261377	-0.10263	-11.5078
	RAINY	-0.00716	1.01394	-0.01036	0.433673
	POST	-0.2733	1.299657	0.033165	-6.9722
Yola	PRE	-0.05354	1.295941	-0.17832	-6.5275
	RAINY	-0.02073	1.104917	-0.08424	-2.06397
	POST	-0.01487	1.019481	-0.00544	-0.09105
Yola	PRE	-0.12773	1.20191	-0.07794	-0.82454
	RAINY	-0.24097	1.246102	0.020877	-3.46228
	POST	-0.01313	1.038241	-0.02637	0.265111
Yola	PRE	-0.01731	1.012285	0.001686	1.214264
	POST	-0.25603	1.290477	-0.03147	-1.6344

**Table 5: Mean RMSE(Mean\_RMSE), Number of times the model is significantly better(DM\_Wins), RMSE and DM Ranks, and overall score for all models using forecasts from all seasons and locations**

Model	Mean_RMSE(W/m <sup>2</sup> )	DM_Wins	RMSE_Rank	DM_Rank	Overall_Score
XGBoost	96.63	54	1	1	2
Hybrid	97.08	51	2	2	4
SARIMAX	132.9	19	3	3	6
LSTM	167.04	7	4	4	8

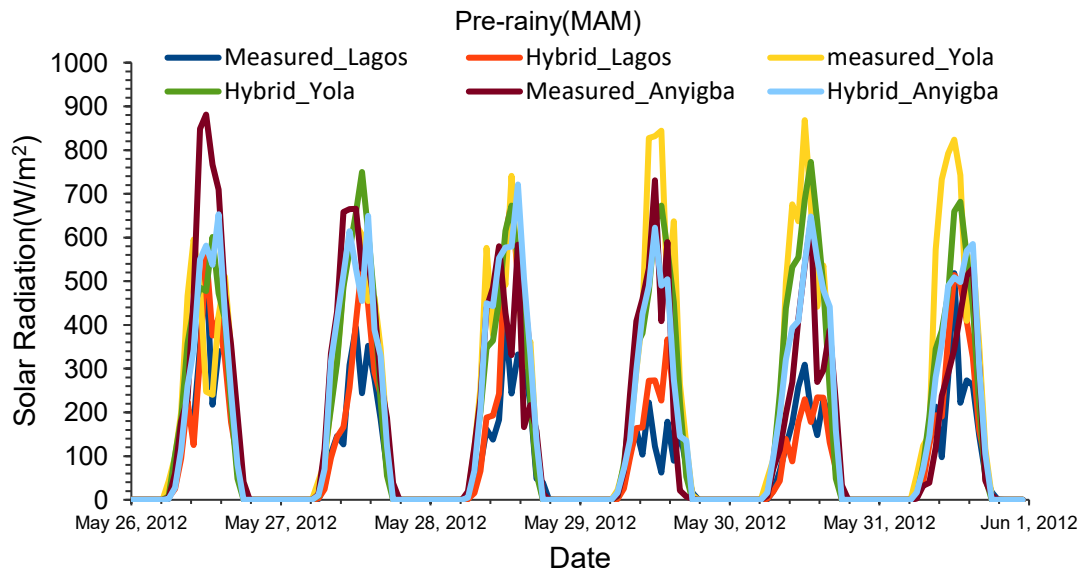


Figure 2: Hourly Variations of Measured and Solar Radiation Forecast for Different Climatic Locations: Lagos (Coastal), Anyigba (Guinea), and Yola (Sahel) in the Pre-rainy Season ( March-April-May)

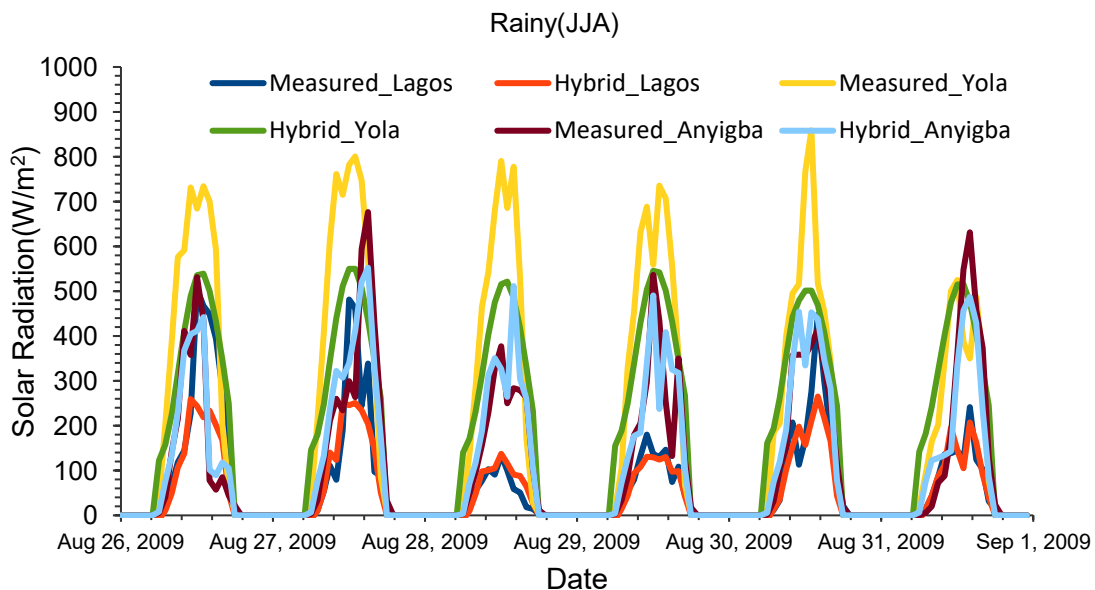


Figure 3: Hourly Variations of Measured and Solar Radiation for Different Climatic Locations: Lagos (Coastal), Anyigba (Guinea), and Yola (Sahel) in the Rainy Season (June-July-August)

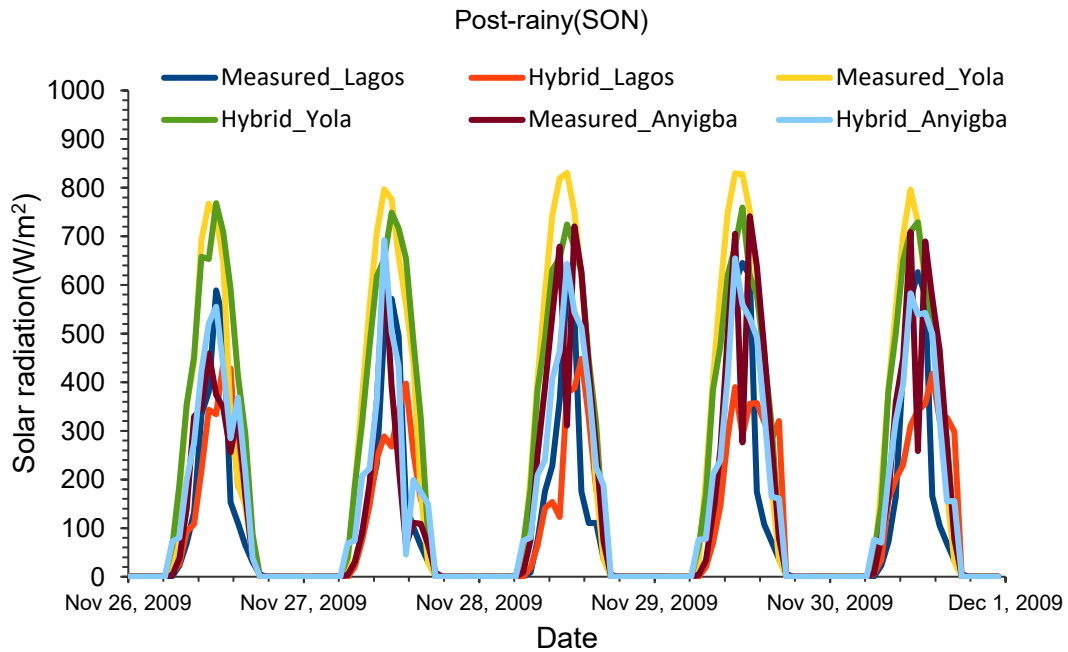


Figure 4: Hourly Variations of Measured and Solar Radiation Forecast for Different Climatic Locations: Lagos (Coastal), Anyigba (Guinea), and Yola (Sahel) in post-Rainy Season(September-October-November)

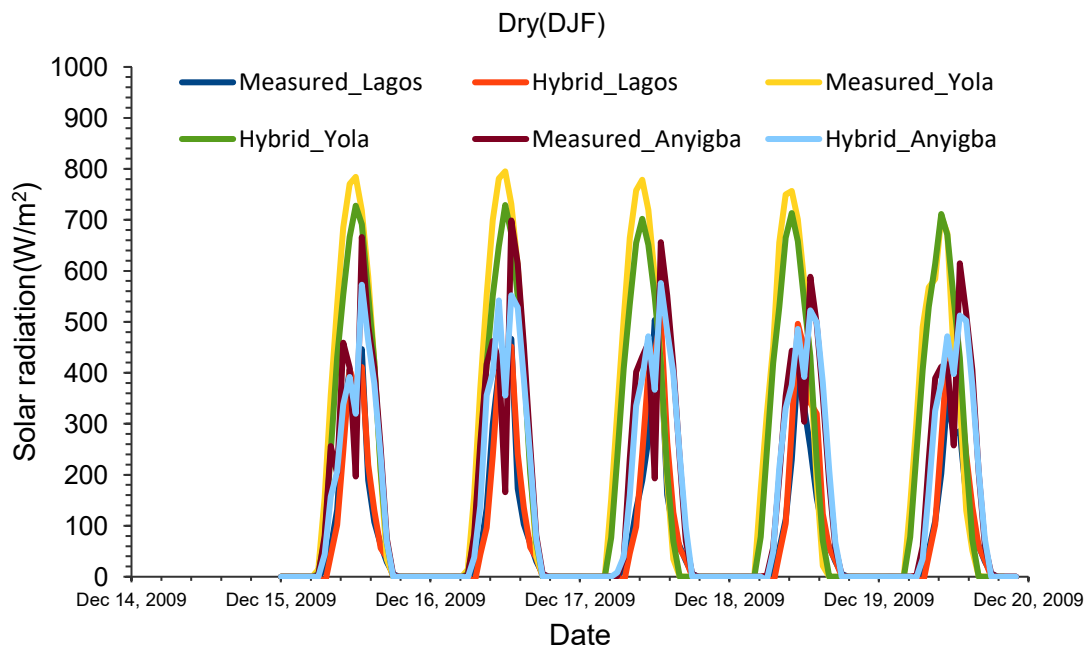


Figure 5: Hourly Variations of Measured and Solar Radiation Forecast for Different Climatic Locations: Lagos (Coastal), Anyigba(Guinea) and Yola (Sahel) in Dry Season (December-January-February)

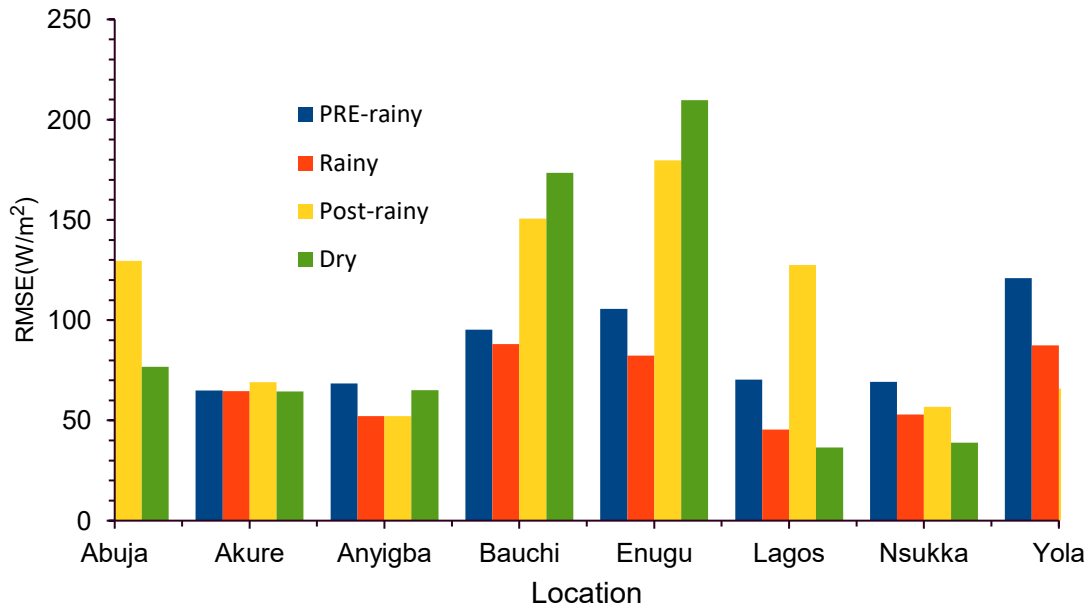


Figure 6: Seasonal Performance Metrics using Root mean Squared Error Between Measured and Stacked Hybrid Forecast of Hourly Solar Radiation for all the Eight Locations used in the Study

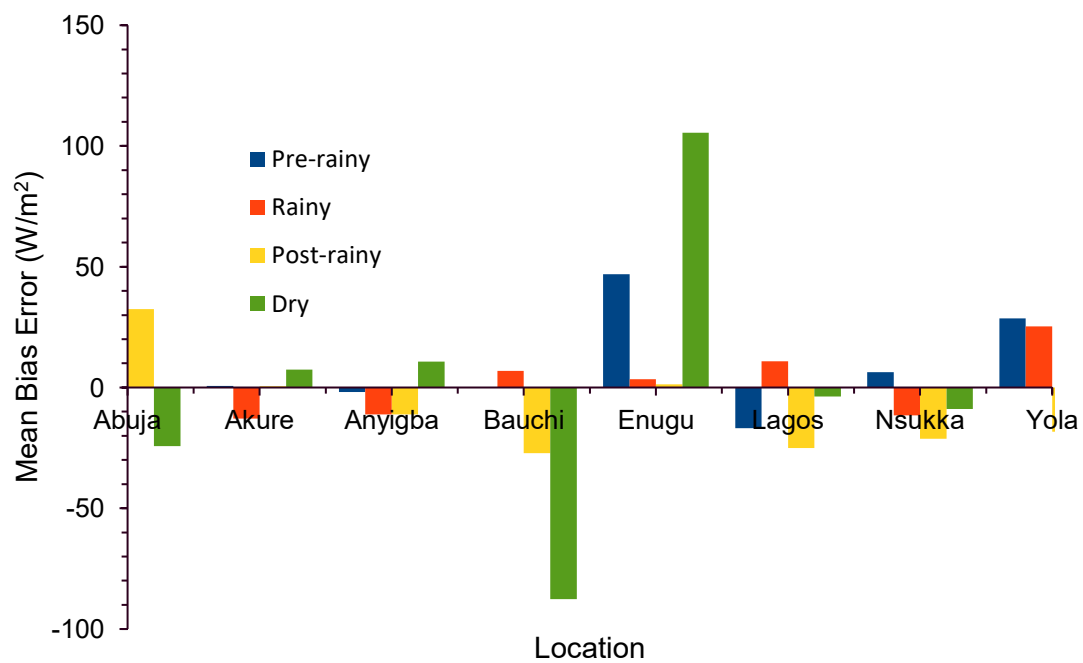


Figure 7: Seasonal Performance METRICS using Mean biased Error Between Measured and Stacked Hybrid Forecast of Hourly solar Radiation for all the Eight Locations used in the Study

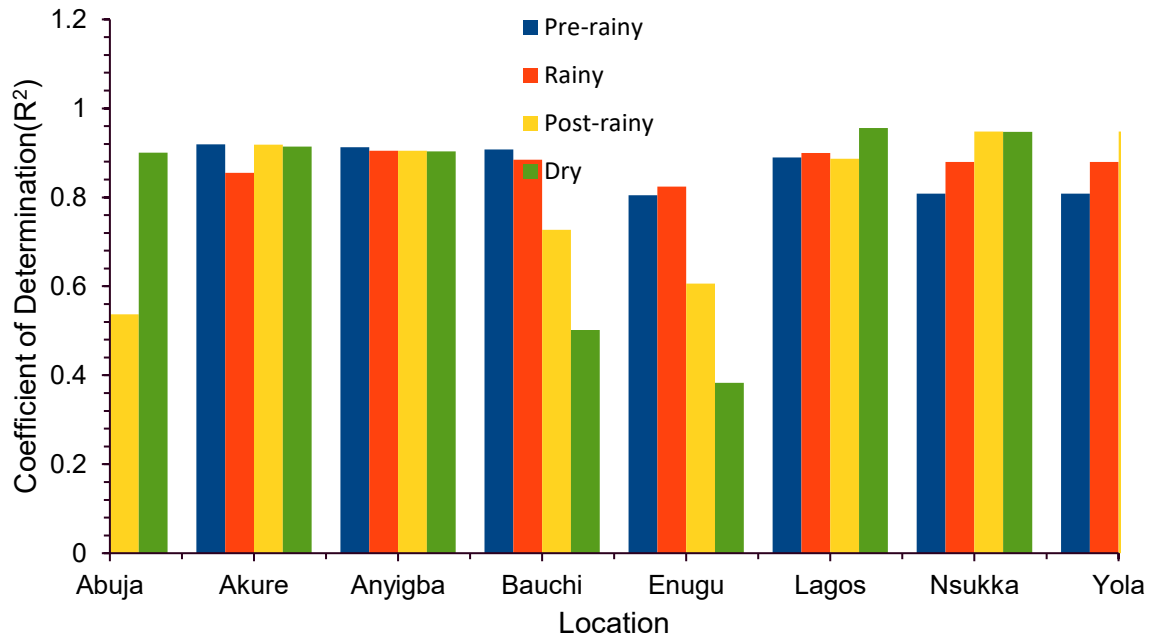


Figure 8: Seasonal Performance Metrics using Coefficient of Determination Between Measured and Stacked Hybrid Forecast of Hourly Solar Radiation for all the Eight Locations used in the Study

## CONCLUSION

Day-ahead hourly solar radiation is of crucial importance in many areas, especially in the optimization of photovoltaic energy generation and storage. A stacked hybrid machine learning model was developed from baseline statistical and machine learning models (i.e., SARIMAX, LSTM, and XGBoost) to forecast day-ahead hourly solar radiation. The performance of the models across the eight locations in Nigeria is given as follows: SARIMAX was able to capture the seasonal trends in the solar radiation sequential data with comparable higher errors to other models, highlighting its limitations in handling hourly data that exhibits irregular variations. LSTM and XGBoost consistently outperformed SARIMAX, with the LSTM model giving the best performance when clearer temporal patterns were present in the data, and XGBoost gave better performance when the data exhibited irregularities and non-linear fluctuations.

The stacked hybrid model proved the most robust overall as it consistently gave greater weight to the best-performing baseline model, and at the same time, assimilated the contributions of other models in the ensemble. Thereby creating a good balance between the overfitting tendency of some of the baseline models and, at the same time, preventing poor performance.

The stacked hybrid model also gave the characteristic features of greater fluctuations in SR values that are expected of MAM and SON seasons, and higher SR values at high latitudes at latitudes compare with low latitudes for the seasons that are comparable.

Thus the integration of the stacked hybrid model developed in this study into PV forecasting at hourly resolution has great potential to enhance renewable energy generation.

## REFERENCES

- Agbo, E. P., Edet, C. O., Magu, T. O., Njok, A. O., Ekpo, C. M., & Louis, H. (2021). Solar energy: A panacea for the electricity generation crisis in Nigeria. *Heliyon*, 7(5). <https://doi.org/10.1016/j.heliyon.2021.e07016>
- Aggarwal, S. K., & Saini, L. M. (2014). Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction Contest. *Energy*, 78, 247–256. <https://doi.org/10.1016/j.energy.2014.10.012>
- Alabi, N. O., & Ojo, G. O. (2024). Dihybrid Recurrent Neural Networks for Solar Radiation Prediction: A Comparative Study in Nigeria. *International Journal of Women in Technical Education and Employment*, 5(2), 62–71. <https://fpiwitedjournal.federalpolyilaro.edu.ng/administrator/docs/6101001.pdf>
- Ali, M. A., Elsayed, A., Elkabani, I., Akrami, M., Youssef, M. E., & Hassan, G. E. (2023). Optimizing Artificial Neural Networks for the Accurate Prediction of Global Solar Radiation: A Performance Comparison with Conventional Methods. *Energies*, 16(17). <https://doi.org/10.3390/en16176165>

- Alsharif, M. H., Younes, M. K., & Kim, J. (2019). Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea. *Symmetry*, 11(2), 240. <https://doi.org/10.3390/sym11020240>
- Belaid, S., & Mellit, A. (2016). Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. *Energy Conversion and Management*, 118, 105–118. <https://doi.org/10.1016/j.enconman.2016.03.082>
- Benitez, I. B., Ibañez, J. A., Lumabad, C. I. D., Cañete, J. M., & Principe, J. A. (2023). Day-Ahead Hourly Solar Photovoltaic Output Forecasting Using SARIMAX, Long Short-Term Memory, and Extreme Gradient Boosting: Case of the Philippines. *Energies*, 16(23), 7823. <https://doi.org/10.3390/en16237823>
- Besharat, F., Dehghan, A. A., & Faghih, A. R. (2013). Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, 21, 798–821. <https://doi.org/10.1016/j.rser.2012.12.043>
- Davou, N., & Umar, I. (2024). Evaluating the potential of large-scale photovoltaic power generation for rural electrification in Makurdi Town, Benue State.: Vol. 2(1), 8–19. *FNAS Journal of Applied and Physical Sciences*. [https://fnasjournals.com/index.php/FNAS-JAPS/article/view/439?utm\\_source=chatgpt.com](https://fnasjournals.com/index.php/FNAS-JAPS/article/view/439?utm_source=chatgpt.com)
- Duffie, J. A., & Beckman, W. A. (2013). *Solar Engineering of Thermal Processes* (1st ed.). Wiley. <https://doi.org/10.1002/9781118671603>
- Ebtehaj, I., & Bonakdari, H. (2024). CNN vs. LSTM: A Comparative Study of Hourly Precipitation Intensity Prediction as a Key Factor in Flood Forecasting Frameworks. *Atmosphere*, 15(9), 1082. <https://doi.org/10.3390/atmos15091082>
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, 12(10), 2451–2471. <https://doi.org/10.1162/089976600300015015>
- Haider, S. A., Hussain, A., Kim, C. H., & Shahzad, M. (2022). Deep learning and statistical methods for short- and long-term solar irradiance forecasting. *Renewable and Sustainable Energy Reviews*, 161, 112388. <https://doi.org/10.1016/j.rser.2022.112388>
- He, Y., Chen, G., Potter, C., & Meentemeyer, R. K. (2019). Integrating multi-sensor remote sensing and species distribution modeling to map the spread of emerging forest disease and tree mortality. *Remote Sensing of Environment*, 231, 111238. <https://doi.org/10.1016/j.rse.2019.111238>
- Hipel, K. W., & McLeod, A. Ian. (1994). *Time series modelling of water resources and environmental systems*. Amsterdam; New York : Elsevier. <https://lib.ugent.be/catalog/ebk01:1000000000555384>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- International Energy Agency, I. E. A. (2022). Nigeria; Energy policy review. <https://www.iea.org/reports/nigeria-energy-policy-review>
- Isma'il, M., & Aliyu, S. (2023). Daily Solar Radiation Forecasting for Northwest Nigeria Using Long Short-Term Memory. *International Journal of Science for Global Sustainability*, 9(1), 8. <https://doi.org/10.57233/ijsgs.v9i1.407>
- Julius, K. A., & Balogun, R. A. (2022). Characteristics and Distribution of Some Radiation Parameters over Nigeria. *European Journal of Environment and Earth Sciences*, 3(4), 32–40. <https://doi.org/10.24018/ejgeo.2022.3.4.255>
- Kassem, Y., Camur, H., Adamu, M. T., Chikowero, T., & Apreala, T. (2023). Prediction of Solar Irradiation in Africa using Linear-Nonlinear Hybrid Models. *Engineering, Technology & Applied Science Research*, 13(4), 11472–11483. <https://doi.org/10.48084/etasr.6131>
- Lamidi, L. O., Oyelakin, A. M., & Akinbi, M. B. (2024). A Survey on Machine Learning Techniques for The Prediction of Solar Power Production. *Indonesian Journal of Data and Science*, 5(2). <https://doi.org/10.56705/ijodas.v5i2.130>
- Li, X., Ma, L., Chen, P., Xu, H., Xing, Q., Yan, J., Lu, S., Fan, H., Yang, L., & Cheng, Y. (2022). Probabilistic solar irradiance forecasting based on XGBoost. *Energy Reports*, 8, 1087–1095. <https://doi.org/10.1016/j.egy.2022.02.251>
- Lim, S. C., Choi, J. H., & Park, J. Y. (2022). Solar power forecasting using CNN–LSTM hybrid model. *Energies*, 15(21), 8233. <https://doi.org/10.3390/en15218233>
- Mujabar, S., & Chintaginjala Venkateswara, R. (2021). Empirical models for estimating the global solar radiation of Jubail Industrial City, the Kingdom of Saudi

- Arabia. SN Applied Sciences, 3(1), 95. <https://doi.org/10.1007/s42452-020-04043-9>
- Mukhtar, M., Oluwasanmi, A., Yimen, N., Qinxiu, Z., Ukwuoma, C. C., Ezurike, B., & Bamisile, O. (2022). Development and Comparison of Two Novel Hybrid Neural Network Models for Hourly Solar Radiation Prediction. Applied Sciences, 12(3), 1435. <https://doi.org/10.3390/app12031435>
- NERC, N. (2022). Renewable energy policy and regulation in Nigeria. Abuja: Nigerian Electricity Regulatory Commission. <https://nerc.gov.ng>
- Ogbaka D.T, O., A.H, B., & V.Z, J. (2020). Angstrom-PreScott Model for Predicting Global Solar Radiation in Mubi, Nigeria. Asian Journal of Basic Science & Research, 02(02), 86–92. <https://doi.org/10.38177/AJBSR.2020.2209>
- Ohijeagbon, O., Waheed, A., & Ajayi, O. (2025). Machine Learning-Based Solar Radiation Forecasting Models for Enhancing Renewable Energy Integration in Nigeria. SSRN. <https://doi.org/10.2139/ssrn.5200160>
- Ohunakin, O. S., Adaramola, M. S., Oyewola, Olanrewaju, M., & Fagbenle, R. O. (2014). Solar energy applications and development in Nigeria: Drivers and barriers. Renewable and Sustainable Energy Reviews, 32, 294–301. <https://doi.org/10.1016/j.rser.2014.01.014>
- Onwuka, B. (2018). Effects of Soil Temperature on Some Soil Properties and Plant Growth. Advances in Plants & Agriculture Research, 8(1). <https://doi.org/10.15406/apar.2018.08.00288>
- Otunla, T. A. (2020). Estimation of daily solar radiation at equatorial region of West Africa using a more generalized Ångström-based broadband hybrid model. Meteorology and Atmospheric Physics, 132(3), 341–351. <https://doi.org/10.1007/s00703-019-00691-8>
- Otunla, T.A. & Kolebaje O.T.(2015). Assessing the performance of global solar radiation empirical models at a Sahelian Site, Sokoto, Nigeria. Journal of the Association of Mathematical Physics, 30, 489-496.
- Oyedepo, S. O. (2012). Energy and sustainable development in Nigeria: The way forward. Energy, Sustainability and Society, 2(1), 15. <https://doi.org/10.1186/2192-0567-2-15>
- Qing, X., & Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy, 148, 461–468. <https://doi.org/10.1016/j.energy.2018.01.177>
- Santos, M.L.; García-Santiago, X.; Camarero, F.E.; Gil, G.B.; Ortega, P.C. Application of Temporal Fusion Transformer for Day-Ahead PV Power Forecasting. Energies 2022, 15, 5232.
- Sharma, J.; Soni, S.; Paliwal, P.; Saboor, S.; Chaurasiya, P.K.; Sharifpur, M.; Khalilpoor, N.; Afzal, A. A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: A case study of India. Energy Sci. Eng. 2022, 10, 2909–2929.
- Tajjour, S., Chandel, S. S., Alotaibi, M. A., Malik, H., García Márquez, F. P., & Afthanorhan, A. (2023). Short-Term Solar Irradiance Forecasting Using Deep Learning Techniques: A Comprehensive Case Study. IEEE Access, 11, 119851–119861. <https://doi.org/10.1109/ACCESS.2023.3325292>
- Urhuetheri, O. P., Udomboso, C., & Sigauke, C. (2025). Forecasting solar power output in Ibadan: A machine learning approach leveraging weather data and system specifications. arXiv E-Prints, arXiv:2508.07462. <https://doi.org/10.48550/arXiv.2508.07462>
- Zhang, Z., Zhang, T., Zhang, R., Zhu, X., Wu, X., Tan, S., & Jian, Z. (2024). Predicting colorectal cancer risk: A novel approach using anemia and blood test markers. Frontiers in Oncology, 14, 1347058. <https://doi.org/10.3389/fonc.2024.1347058>
- Zhu, R.; Guo, W.; Gong, X. Short-Term Photovoltaic Power Output Prediction Based on k-Fold Cross-Validation and an Ensemble Model. Energies 2019, 12, 1220.
- Zhong, Y.-J.; Wu, Y.-K. Short-Term Solar Power Forecasts Considering Various Weather Variables. In Proceedings of the 2020 International Symposium on Computer, Consumer and Control (IS3C), Taichung City, Taiwan, 13–16 November 2020; pp. 432–435.