

A Performance Evaluation of Machine Learning Models for Solar PV Power Forecasting in Bamenda, Cameroon

Noel Nkwa Awangum¹, Derek Ajesam Asoh^{2,3*}, Jerome Ndam Mungwe³, Jean-De-Dieu Nguimfack⁴, Therese Nkwantoh⁵, Reeves Meli Fokeng⁶, Adelaide Nicole Kengnou Telem⁷, Patience Tifuh Taah¹, Carine Tanwie⁵, Daniel Agoons⁸

¹Department of Electrical/Electronic Engineering, NAHPI, University of Bamenda, Bamenda, Cameroon

²Electrical Engineering, Mechatronics, and Signal Processing Laboratory, ENSPY, UY-I, Yaounde, Cameroon

³Department of Renewable Energy Engineering, HTTTC, University of Bamenda, Bamenda, Cameroon

⁴Department of Electric Power Engineering, HTTTC, University of Bamenda, Bamenda, Cameroon

⁵Department of Forestry and Wildlife Technology, COLTECH, University of Bamenda, Bamenda, Cameroon

⁶Department of Geography and Planning, Faculty of Arts, University of Bamenda, Bamenda, Cameroon

⁷Department of Electrical and Electronic Engineering, COLTECH, University of Buea, Buea, Cameroon

⁸Agoons M & E Consultants, Simbock, Yaounde, Cameroon

Email: *derekasoh@gmail.com

How to cite this paper: Awangum, N.N., Asoh, D.A., Mungwe, J.N., Nguimfack, J.-D.-D., Nkwantoh, T., Fokeng, R.M., Telem, A.N.K., Taah, P.T., Tanwie, C. and Agoons, D. (2025) A Performance Evaluation of Machine Learning Models for Solar PV Power Forecasting in Bamenda, Cameroon. *Journal of Power and Energy Engineering*, 13, 1-20.

<https://doi.org/10.4236/jpee.2025.138001>

Received: July 4, 2025

Accepted: August 12, 2025

Published: August 15, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Facing increased energy demand which surpasses national grid supply capacity due to rapid population growth, urbanization, and economic activities, developing countries such as Cameroon are deploying solar photovoltaic power (SPVP) systems to supplement their energy needs; with these systems heralded for sustainability and environmental friendliness. However, the inherent intermittency of SPVP is a major concern since it cannot reliably fill the supply-demand gap with its associated risk of non-availability. Tackling this issue requires adequate forecasting of SPVP to guarantee better management of the energy shortfall. This study evaluates the performance of twenty-four machine learning models (MLMs) in forecasting SPVP in Bamenda, Cameroon. The study uses data from Photovoltaic Geographical Information System with six input features (direct beam irradiance, diffuse irradiance, reflected irradiance, sun height, ambient temperature, and wind speed) and training-testing split of 80% - 20% to forecast SPVP as output feature. Employing hold-out and re-substitution validation techniques, MLMs performance was evaluated using Coefficient of Determination (R^2) and Root Mean Squared (RMSE) metrics. Results reveal wide neural network model as the overall best performer with

*Corresponding author.

R^2 of 0.999 and RMSE of 9.377, compared to the other models with same or lower R^2 and higher RMSE ranging from 9.4522 to 458.97. This model was used to perform short-term SPVP forecast in Bamenda and may be used in the forecast of SPVP in geographically similar areas of Cameroon. This study underscores the role and importance of MLM performance evaluation to identify the best-yield model for SPVP to reliably fill supply-demand gaps.

Keywords

Renewable Energy, Solar PV Power Forecasting, Machine Learning Model, Performance Evaluation

1. Introduction

The emergence and development of a nation is dependent on its energy generation and use. Energy exists in several forms, one of which is electrical energy. Electrical energy is considered the most essential form of energy because it can be easily converted into other forms of energy and is widely used for domestic, industrial, commercial, and agricultural applications [1].

Fossil fuels constitute the major primary source of energy worldwide. However, fossil fuels are detrimental because they are the primary cause of climate change [2]. The adverse effects of energy generation through the use of fossil fuels drive the development and exploitation of renewable energy sources which are considered environmentally friendly. Renewable energy is energy from a naturally occurring source that has replenishment potential after use, for example, solar, wind, hydropower, biofuels, tidal, geothermal energy, etc. [3]

As a result of its availability, sustainability, and minimal negative environmental impact, solar energy has gained popularity in worldwide electricity generation. Solar photovoltaic (PV) technology has been developed for converting solar energy to electricity [2] [4] [5] and with improved and cheaper technology over the years, solar PV is more affordable technology [6]. Solar PV power generation and use play a positive role in fighting climate change, which adversely affects the environment and economies, given its emission free operation [7].

The positive impact, resulting from the use of solar PV power in fighting climate change is indirect and can be understood from the viewpoint that generating and using solar PV power reduces the use of fossil fuels, which in turn leads to a decrease in the emission of greenhouse gases (GHG), pollution of the environment, and ultimately less adverse environmental condition and climate change. It has been demonstrated that a 10 kW Solar PV system could generate, on the average, 40 kWh per day, thereby reducing GHG emission by 37 kg per day and 13,505 kg per year [8].

Although solar PV power can be heralded for its positive role in mitigating adverse climate change, it is however, an intermittent source of energy due to its

dependence on many factors including irradiance [9], temperature, wind speed, humidity, and dust [4] [10].

The maximum positive impact of solar PV power on climate change is attainable by encouraging massive use in tandem with its availability. Such massive use requires forecasting PV power availability efficiently as well as using cost-effective approaches.

Although solar PV power forecasting has gained significant attention in recent years as a result of increase in the demand for renewable energy [11], the process is complex, given that solar energy depends on several weather conditions such as solar irradiance, temperature, humidity, and wind speed which are uncontrollable [12]. In the quest for solutions, Machine Learning Techniques (MLT) have recently emerged as important tools for forecasting solar PV power generations [13]. The availability of large volume meteorological data and recent advancements in computing power have led to the development and use of Machine Learning Models (MLM) for solar PV power forecasting in many localities [14]. This work is a contribution to fill the gap in machine learning-based solar PV power forecasting research in Cameroon, where there is high solar PV power potential [15] [16] but little or no research has been done in solar PV power forecasting using MLTs.

2. Methodology

The procedure used in this study included: data acquisition, data preparation, training, validation, and testing of different MLMs, and evaluation and identification of the most accurate and better-yield MLM suitable for Solar PV power forecasting. **Figure 1** presents the flowchart of the methodological process.

2.1. The Study Area

The study area considered in this study is Bamenda in the North West Region of Cameroon, as shown in **Figure 2**. Bamenda is the main city of the North West Region, and is located at latitude 5.961°N and longitude 10.152°E. The study area was chosen because it is one of the localities in the North West Region targeted in the Renewable Energy Master Plan (REMP) of Cameroon for the deployment of solar PV generation stations [17].

2.2. Data Acquisition

The dataset for the study incorporates six meteorological variables as input (direct beam irradiance, diffuse irradiance, reflected irradiance, sun height, ambient temperature, and wind speed) and one output variable (PV power). Given the absence of datasets with ground measurements of these input and output variables for the study area, a dataset with a four-year time range (2017-2020) and hourly resolution was downloaded from Photovoltaic Geographical Information Systems (PVGIS); a platform which provides information on solar radiation and solar PV power generation potential for all locations in Europe, Africa, and a large part of

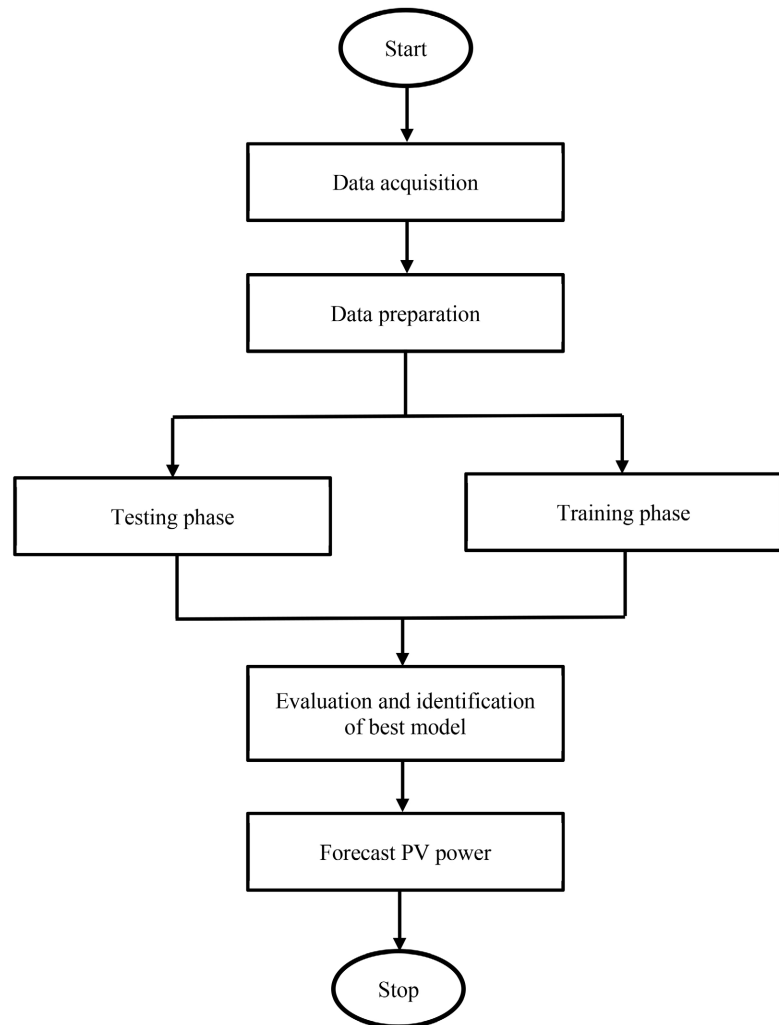


Figure 1. Methodological process.

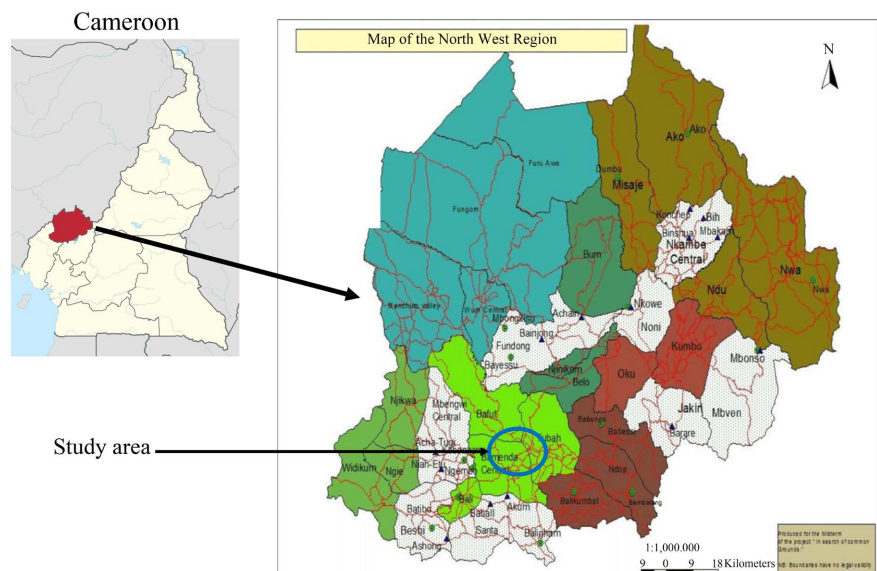


Figure 2. Political map of Cameroon [18].

Asia and America. The PVGIS data is obtained from satellite images and very high-quality *European Center for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (ERA5)* data, which has hourly estimates of a large number of atmospheric, land, and oceanic variables from 1940 up to date [19]. The raw data used in this study was made up of 35,064 records with the above-mentioned input and output variables; and enabled the forecast of the solar PV power generation potential of a 20 kW solar PV system in Bamenda, Cameroon.

2.3. Data Preparation

The first step in data preparation was feature selection. The six meteorological inputs and PV power output in the downloaded dataset were retained given that they align with the features in previous research reported in [20]. Although other features such as humidity, and precipitation affect solar PV power, they do not show any dominant relationship with solar PV power generation [21].

The second step in data preparation involved preprocessing. Data preprocessing is essential in Machine Learning (ML) because raw data usually contains duplicates, outliers, and missing values. Preprocessing was used to properly clean and transform the data into a format suitable for ML. The data obtained for this study was preprocessed using Microsoft Excel and MATLAB software. While Microsoft Excel was used to search for and remove duplicates from the dataset, MATLAB was used to remove outliers and missing values. The third and final step in the data preparation was splitting the preprocessed data into training and testing sets. In line with previous research, during the splitting, 80% of the data was retained for training and 20% for testing [20].

2.4. Variables in the Study

The preprocessed data comprised of six input variables or features (direct beam irradiance, diffuse irradiance, reflected irradiance, sun height, ambient temperature, and wind speed) and one output feature (solar PV power). The dataset used in this study contained more features as compared to two and four features in [21] and [22], respectively.

Solar PV power generation is dependent on meteorological conditions, implying that different meteorological parameters affect solar PV power generation differently [23]. Therefore, meteorological parameters were considered independent variables, used to forecast generated solar PV power as the dependent variable.

2.5. Machine Learning Techniques and Models

Machine learning techniques (MLTs) also referred to as machine learning algorithms (MLAs) often comprise two or more machine learning models (MLMs). Twenty-four MLMs stemming from six MLTs were trained, validated, and tested. The most accurate and better-yield model was later identified and selected for forecasting solar PV power. The MLMs considered in this study are presented below and summarized in **Table 1**.

Table 1. Summary of Machine Learning Techniques and corresponding models used in the study.

Machine learning technique	Machine learning model
Linear regression (LR)	Linear
	Interactions Linear
	Robust Linear
	Stepwise Linear
Regression trees (RT)	Fine Tree
	Medium Tree
	Coarse Tree
Support vector machines (SVM)	Linear SVM
	Quadratic SVM
	Cubic SVM
	Fine Gaussian SVM
	Medium Gaussian SVM
Ensembles of regression trees (ERT)	Coarse Gaussian SVM
	Boosted Trees
	Bagged Trees
Gaussian process regression (GPR)	Squared Exponential GPR
	Matern 5/2 GPR
	Exponential GPR
	Rational Quadratic GPR
Neural network (NN)	Narrow Neural Network
	Medium Neural Network
	Wide Neural Network
	Bilayered Neural Network
	Trilayered Neural Network

2.5.1. Linear Regression

Linear Regression models are supervised MLMs that use a polynomial slope-intercept form to connect the independent and dependent variables. In this case, the expected output is a continuous number. Linear Regression models learn the best fit of the line between the inputs (independent variable) and the outputs (dependent variable). This is achieved by minimizing the approximated error from the forecasted and actual responses [24]. The relationship between the input/independent variable, X , and the output/dependent variable, Y , is given by Equation (1) [11].

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

where:

- Y = the dependent variable,
- X = the independent variable,
- β_0 = the intersection point with the y-axis,
- β_1 = regression coefficient, and
- ε = error term.

Equation (1) is often referred to as a simple linear regression, where the dependent variable is modeled with one independent variable only. However, if the dependent variable is modeled with two or more independent variables, the relationship is called multivariate linear regression and is given by Equation (2) [11]. In this study, a multivariate regression was used, because the dataset employed, contained six independent variables.

$$Y = \beta_0 \sum_{i=1}^n \beta_i X_i + \varepsilon \quad (2)$$

where,

- X_i = independent variables,
- β_0 = a constant
- β_i = regression coefficients
- ε = error term

2.5.2. Regression Trees

Regression trees are MLMs developed using the fast divide and conquer greedy algorithm that recursively partitions the given training data into smaller subsets. Regression trees are well known for their efficiencies. However, Regression trees can also lead to poor decisions in lower levels of the trees, as a result of the unreliability of estimates, based on small samples of cases. Regression trees are also known for their instability. A small change in the training dataset may lead to a different choice when building a node, which may represent a dramatic change in the tree, especially if the change occurs at the top-level nodes. A simple regression tree is essentially the same as a decision tree, used in the context of response prediction. The branches of the tree represent splits on the value of a regressor, creating partitions in the independent data where the dependent data is more homogenous, like a form of clustering. The terminal nodes contain the prediction model for that branch of the tree. In this application, prediction models are constants; specifically, the medians of observations within the partition defined by that branch of the tree [25] [26].

2.5.3. Support Vector Machines

Support vector machines (SVM) are supervised MLMs, developed to solve problems of classification and regression [2] [24]. The key of support vector machines is to determine a function having n deviations from the actual vectors for training data. For regression problems, the main objective structures of support vector machines are to minimize training errors, and to reduce generalized errors to achieve generalized performance. Support vector machines are used extensively to achieve good performance in regression applications. In such applications, support vector

machines use a mathematical function referred to as the cost function for the measurement of the empirical risk to minimize error [24].

2.5.4. Gaussian Process Regression

Gaussian process regression is a non-parametric technique used frequently for solving extreme non-linear problems. Gaussian process regression is comprised of random variables. In a Gaussian process regression, all input and output data follow a Gaussian distribution profile. Gaussian process regression provides a distribution of all potential functions that are reliable with the training dataset. Gaussian process regression is then defined as a Gaussian process of the mean function $n(x)$ and kernel function $t(x, x')$. This can be represented mathematically as shown in Equation (3) [2].

$$F(x) = GP(n(x), t(x, x')) \quad (3)$$

where $n(x)$ = central tendency of F , x = test input.

2.5.5. Ensembles of Regression Trees

Ensembles algorithms are used to develop ensembles of regression trees used for reducing fluctuations in individual trees. This approach as applicable to machine learning, uses the bootstrap aggregation or bagging algorithm for creating similar datasets, tested from the identical source dataset. Ensembles of regression trees are machine learning models developed to overcome overfitting, a common problem with regression trees [2].

2.5.6. Neural Networks

Neural networks are MLM inspired by the way the human brain functions. These networks have interconnected layers of processing units, known as neurons [27] between which information is transmitted through interconnections during operation. Weights characterize each of these connections; and an activation function limits the amplitude of the output neurons. It is worth noting that neural networks are not developed for particular machine learning applications, but are trained to learn patterns from specific datasets. Once trained, neural networks can be used to make predictions or classifications. These networks can learn to recognize patterns in data from physical systems, computer programs, or other source [28].

2.6. Training and Validation of Machine Learning Models

Model training is a process that seeks to enable a MLM to learn the trends and relationships in the dataset. Model training involves feeding the MLM with input data [29]. MLM training in this study was performed using a MATLAB built-in application referred to as the Regression Learner. The Regression Learner is a specialized MATLAB application for training regression-based MLM. This application specializes in training MLM to make predictions on data. This application enables users to explore data, select features, specify validation techniques, train models and evaluate performance. With the Regression Learner, users can perform supervised machine learning using labeled data.

To train the MLM in this research, the training dataset of 26,400 records, which represented 80% of the preprocessed data of 33,000 records, was imported into the Regression Learner app in MATLAB, followed by selecting a validation technique. Validation is the process of accessing the ability of a trained MLM to perform on a new and unseen dataset. During validation, the MLM made predictions using the validation dataset. Then the model's performance was evaluated using different performance evaluation metrics such as RMSE, R^2 , MAE, and MAPE. Validation is crucial in identifying whether or not the MLM is overfitting the training dataset. The validation of all the 24 MLM in this study was performed using the Regression Learner in MATLAB in terms of the RMSE and R^2 . The RMSE and R^2 are the most widely used performance evaluation metrics [13]. Two validation techniques, namely, hold out and re-substitution validation were employed. These validation techniques are widely used with large datasets [30]. The use of the two validation techniques was to ensure that the most accurate MLM is rightly identified and selected for Solar PV power forecasting applications. With the Hold-out validation technique, 25% of the training dataset was held out (reserved) for accessing the performance of the MLM after training. With the re-substitution validation technique, the entire training dataset was used for assessing the performance of the different MLMs after training.

Validation of MLM was also performed with response plots, which are visual representations of the relationship between actual values and forecasted values.

2.7. Testing of Machine Learning Models

Testing of MLMs after training and validation is crucial. Just as validation, testing of MLMs involves the evaluation of the model's performance. Unlike validation which seeks to evaluate the performance of MLMs on new data, testing of MLMs is to evaluate their performance in real-world scenarios. The testing of MLMs was performed using the test data. Testing of all the 24 MLMs in this study was performed with the Regression Learner in MATLAB with a test dataset of 6600 records representing 20% of the preprocessed dataset of 33,000 records. During the testing the test dataset was uploaded into the Regression Learner, followed by running the 24 MLM on the test dataset. Thereafter, the RMSE, and R^2 values, and actual versus predicted plots generated by the models were observed and recorded.

2.8. Performance Evaluation and Identification/Selection of the Most Accurate MLM

The 24 MLMs were evaluated through the computation of the coefficient of determination (R^2) and Root Mean Squared (RMSE) values of each model, as shown in Equations (4) and (5) respectively [25] [31] [32].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5)$$

Root Mean Squared Error (RMSE) is a performance evaluation metric widely used in machine learning applications. The RMSE is a measure of the average difference between the forecasted values and the actual values and provides an estimation of how well the trained machine learning model can forecast the target value. Best machine model accuracy is demonstrated by lower RMSE values. The coefficient of determination (R^2), is another performance evaluation metric widely used in regression-based machine learning applications. R^2 is defined as the measure of the variance in the response variable that can be predicted using the predictor variable. R^2 values lie between 0 and 1, where values close to 1 indicate good model accuracy, while values close to 0 indicate poor model accuracy.

In order to select the most accurate MLM for forecasting solar PV power generation in Bamenda, Cameroon, the RMSE and R^2 values of all the trained and tested models were compared. Following this evaluation, the model with the lowest RMSE and highest R^2 value was selected as the most accurate MLM for forecasting solar PV power in the Bamenda, Cameroon.

Additionally, the most accurate MLM was identified visually, with the aid of response plots. By this technique, the model whose data points lie closest to the diagonal line compared to the other models was identified as the most accurate model and consequently selected as the best MLM for forecasting solar PV power in Bamenda, Cameroon.

2.9. Forecasting of Solar PV Power Generation for Study Area

The identified accurate MLM was used to perform Solar PV power forecasting in Bamenda, Cameroon based on short-term forecasting horizon. Firstly, the selected MLM was exported to the MATLAB workspace. Then, new data was fed into the model in order to perform forecasting. The MATLAB command shown in Equation (6) was used to feed data into the model:

$$\text{yfit} = \text{trainedModel.predictFcn}(T) \quad (6)$$

where:

- yfit contains the forecasted values for each data point,
- trained Model = the exported machine learning model,
- T = the test dataset in a tabular format.

3. Results and Discussions

3.1. Data Acquisition and Preparation

A dataset made up of both electrical and meteorological data was obtained and used for the training, validation, and testing of MLMs. The raw data acquired was made up of 35,064 records. After preprocessing the acquired raw data was reduced to 33,000 records, representing 94% of the original raw data. Solar PV power generation which was used as the output varied from 0 kW to 18 kW.

Both electrical and meteorological data sets were used for the training, validation, and testing of MLMs. The raw data acquired had 35,064 records. After pre-processing the acquired raw data was reduced to 33,000 records, representing 94% of the original raw data. Solar PV power generation used as the output varied from 0 kW to 18 kW as shown in **Figure 3**.

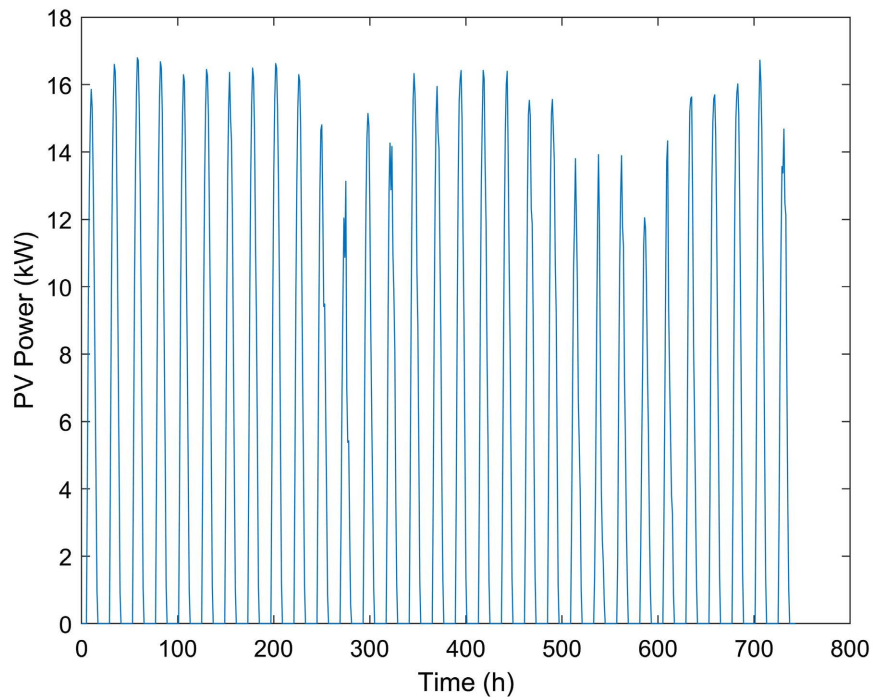


Figure 3. Output Range of Solar PV power generation.

The data acquired revealed that the study area considered has a high solar PV power potential, whose exploitation and use can go a long way to provide households and businesses with sustainable electricity supply.

3.2. Relationship between Meteorological Parameters and Solar PV Power Generation

Direct beam irradiance and ambient temperature influence Solar PV power generation more than other meteorological parameters [23]. Thus, after the data was acquired and preprocessed, the relationship between the direct beam irradiance and ambient temperature and solar PV power generation were studied. The results showed that solar PV power generation is directly proportional to the direct beam irradiance [23], while the optimum Solar PV power generation of 17 kW occurred when the ambient temperature was somewhere between 24°C and 28°C.

3.3. Training and Validation of the MLM

As mentioned in Section 2.5, a total of 24 MLMs were trained and validated. The MLMs were first validated using the hold out validation technique, and then using the re-substitution validation technique. The results obtained in terms of RMSE

and R^2 with both validation techniques are shown in **Table 2**. The hold-out validation yielded RMSE and R^2 values of the 24 MLMs ranging from 8.8227 to 477.19 and 0.994 to 0.999 respectively. The re-substitution validation yielded RMSE and R^2 values ranging from 2.4620 to 352.07 and 0.994 to 0.999 respectively. From **Table 2**, the Fine Gaussian SVM MLM was the least accurate with a RMSE value of 477.19 and an R^2 value of 0.990 after the hold-out validation and also the least accurate with an RMSE value of 352.07 and an R^2 value of 0.994 after the Re-substitution validation.

Table 2. Validation of the 24 MLM in terms of RMSE and R^2 .

S/N	ML model	RMSE		R^2	
		Hold out validation	Re-substitution validation	Hold out validation	Re-substitution validation
1	Wide Neural Network	8.8227	9.299	0.999	0.999
2	Matern 5/2 GPR	9.404	8.858	0.999	0.999
3	Rational Quadratic GPR	11.891	11.17	0.999	0.999
4	Medium Neural Network	13.276	16.32	0.999	0.999
5	Squared Exponential GPR	14.061	13.638	0.999	0.999
6	Trilayered Neural Network	14.322	16.748	0.999	0.999
7	Exponential GPR	14.45	2.462	0.999	0.999
8	Bilayered Neural Network	16.894	19.489	0.999	0.999
9	Narrow Neural Network	20.893	33.122	0.999	0.999
10	Stepwise Linear	86.806	88.904	0.999	0.999
11	Interactions Linear	86.807	88.899	0.999	0.999
12	Bagged Trees	144.06	86.84	0.999	0.999
13	Linear	146.2	146.74	0.999	0.999
14	Linear SVM	165.92	165.85	0.998	0.998
15	Medium Gaussian SVM	167.94	158.08	0.998	0.998
16	Coarse Gaussian SVM	170.83	171.79	0.998	0.998
17	Fine Tree	175.6	100.74	0.998	0.999
18	Cubic SVM	187.85	199.39	0.998	0.998
19	Medium Tree	189.11	130.25	0.998	0.999
20	Robust Linear	202.42	188.77	0.998	0.998
21	Coarse Tree	237.52	186.79	0.998	0.998
22	Quadratic SVM	242.32	234.22	0.997	0.997
23	Boosted Trees	347.19	335.93	0.994	0.995
24	Fine Gaussian SVM	477.19	352.07	0.99	0.994

After the Hold-out validation, the Wide Neural Network was the most accurate MLM with a RMSE values of 8.8227 and R^2 value of 0.999 while the Exponential GPR MLM had a RMSE value of 14.450 and R^2 of 0.999. After the re-substitution validation, Exponential GPR was the most accurate MLM with a RMSE value of 2.4620 and R^2 value of 0.999 while the Wide Neural Network had a RMSE of 9.2997 and a R^2 value of 0.999. Also, after the hold-out validation, the Matern 5/2 GPR MLM had a RMSE value of 9.4042 and a R^2 value of 0.999; and a RMSE value of 8.8580 and an R^2 value of 0.999 after the re-substitution validation. Therefore, the Wide Neural Network and Matern 5/2 GPR MLMs were potential candidates for the most accurate MLM for forecasting of Solar PV power generation in Bamenda, Cameroon, after training.

3.4. Test of the Machine Learning Models

Table 3 presents the outcome of the test of the 24 MLMs in terms of RMSE and R^2 . From **Table 3**, the Fine Gaussian SVM MLM was the least accurate MLM with a RMSE value of 458.97 and an R^2 value of 0.990 after the hold-out validation and also the least accurate with a RMSE value of 443.50 and an R^2 value of 0.991 after the Re-substitution validation.

Table 3. Test of the 24 MLMs in terms of the RMSE and R^2 .

S/N	Machine learning model	RMSE		R^2	
		Hold out validation	Re-substitution validation	Hold out validation	Re-substitution validation
1	Wide Neural Network	9.377	9.5279	0.999	0.999
2	Matern 5/2 GPR	9.452	10.104	0.999	0.999
3	Exponential GPR	10.299	14.861	0.999	0.999
4	Rational Quadratic GPR	10.922	12.248	0.999	0.999
5	Squared Exponential GPR	13.824	14.689	0.999	0.999
6	Trilayered Neural Network	15.21	17.071	0.999	0.999
7	Medium Neural Network	16.348	17.07	0.999	0.999
8	Bilayered Neural Network	19.63	20.034	0.999	0.999
9	Narrow Neural Network	29.851	34.105	0.999	0.999
10	Interactions Linear	90.252	88.225	0.999	0.999
11	Stepwise Linear	90.258	88.233	0.999	0.999
12	Bagged Trees	127.37	126.79	0.999	0.999

Continued

13	Linear	147.98	146.93	0.999	0.999
14	Medium Gaussian SVM	155.03	160.24	0.998	0.998
15	Fine Tree	163.96	168.41	0.998	0.998
16	Linear SVM	167.28	167.18	0.998	0.998
17	Medium Tree	174.04	176.32	0.998	0.998
18	Coarse Gaussian SVM	174.78	172.84	0.998	0.998
19	Robust Linear	191.3	186.03	0.998	0.998
20	Cubic SVM	192.94	199.56	0.998	0.998
21	Coarse Tree	210.16	214.86	0.998	0.998
22	Quadratic SVM	236.37	234.96	0.997	0.997
23	Boosted Trees	338.89	343.29	0.995	0.994
24	Fine Gaussian SVM	458.97	443.5	0.99	0.991

After the Hold-out validation, the Wide Neural Network was the most accurate MLM with a RMSE value of 9.377 and an R^2 value of 0.999, the Matern 5/2 GPR had a RMSE value of 9.4522 and R^2 value of 0.999 while the Exponential GPR MLM had a RMSE value of 10.299 and R^2 value of 0.999. After the re-substitution validation, the wide neural network was the most accurate MLM with a RMSE value of 9.5279 and R^2 value of 0.999, the Matern 5/2 GPR had a RMSE value 10.104 and a R^2 value of 0.999, while the Exponential GPR had a RMSE value of 14.861 and a R^2 value of 0.999.

Therefore, after testing of the MLMs, the Wide Neural Network and Matern 5/2 GPR MLMs were potential candidates for an accurate MLM for forecasting of Solar PV power generation in Bamenda, Cameroon. Additionally, 62% of the MLMs yielded more accurate results with the hold-out validation technique.

3.5. Performance Evaluation and Selection of the Accurate MLM

Performance evaluation of the 24 MLM was done by comparing the RMSE and R^2 values of the models achieved during the testing phase, as well as observing and comparing the actual versus forecasted plots of Solar PV power generation of the models. In the plots of actual versus forecasted Solar PV power generation, the spread of the data points along the diagonal is an indication of the accuracy of the MLM. The more spread the data points, the least accurate the MLM, the less spread the data point, the more accurate the MLM. The most performant MLM is one with the lowest RMSE and highest R^2 values and less spread of the data points on the plot of the actual vs forecasted Solar PV power generation, while the least performant model is one with highest RMSE and lowest R^2 values and the more spread the data points on the plot of actual vs forecasted Solar PV power generation. A MLM with lowest RMSE and highest R^2 values will be the most accurate model for Solar PV power forecasting in Bamenda, Cameroon. From the testing

of the 24 MLMs, the Fine Gaussian SVM yielded the least accuracy, while the wide neural network yielded the highest accuracy and was therefore identified as the most accurate model for forecasting Solar PV power generation in Bamenda, Cameroon. **Figure 4** is a plot of the actual vs forecasted Solar PV power generation for Fine Gaussian SVM (the least accurate MLM) while **Figure 5** is the actual vs predicted plot of Wide Neural Network (the most accurate MLM).

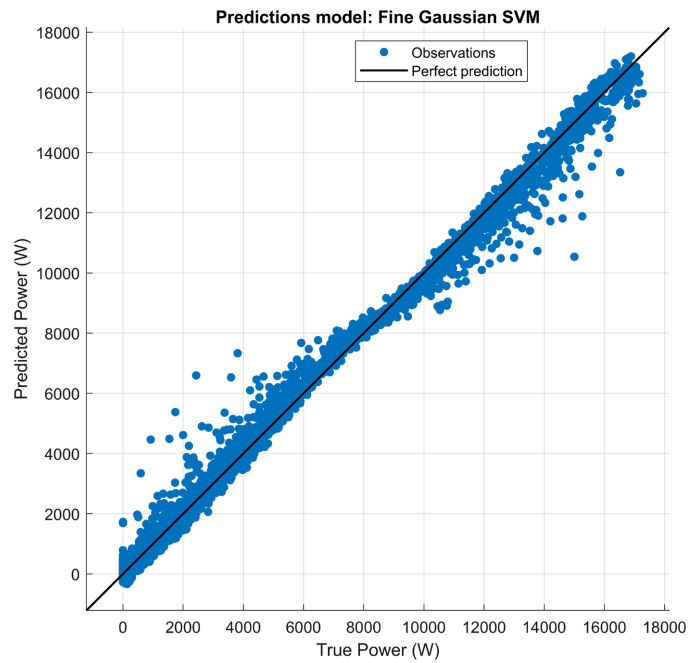


Figure 4. Actual vs forecasted solar PV power generation using Fine Gaussian SVM.

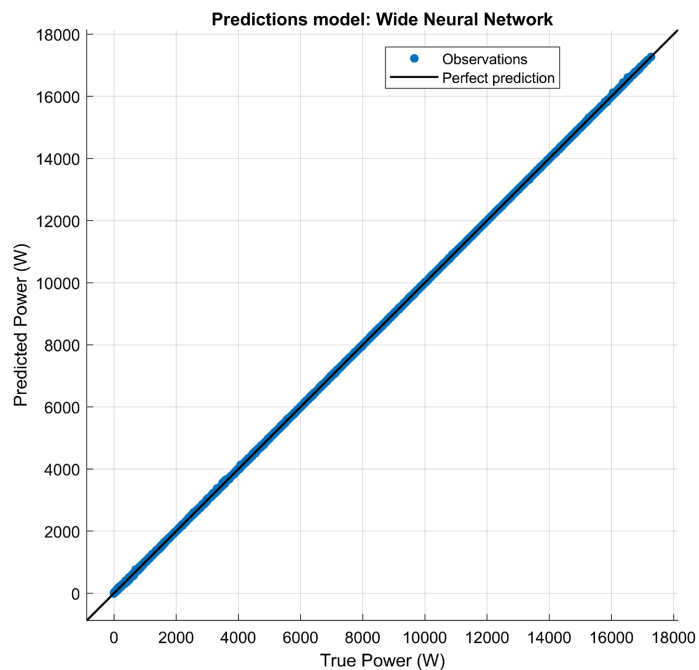


Figure 5. Forecasted vs actual solar PV power generation using the wide neural network.

Therefore, the wide neural network MLM, with a RMSE value of 9.377 and R^2 value of 0.999 from the hold-out validation during the testing, was identified to be the most accurate MLM for forecasting Solar PV power generation in the Bamenda, Cameroon. In terms of R^2 , the wide neural network, identified as the most accurate MLM in this study, yielded higher values, compared to the values obtained by [31]-[35].

Therefore, the wide neural network MLM, with a RMSE value of 9.377 and R^2 value of 0.999 from the hold-out validation during the testing, was identified to be the most accurate MLM for forecasting Solar PV power generation in Bamenda, Cameroon. Key hyper-parameters of the wide neural network MLM are shown in **Table 4**. In terms of R^2 , the wide neural network, identified as the most accurate MLM in this study, yielded higher values, compared to the values obtained in previous research as reported in [31]-[35].

Table 4. Hyper-parameters of wide neural network.

Number of hidden layers	1
First layer size	100
Activation function	ReLU
Iteration limit	1000
Optimizer	Not applicable

3.6. Forecasting of Solar PV Power Generation in Bamenda, Cameroon

The Wide Neural Network MLM was used to forecast the Solar PV power generation in Bamenda, North West Region following the procedure described in Section 2.9. The short-term forecasting results are shown graphically in **Figure 6**.

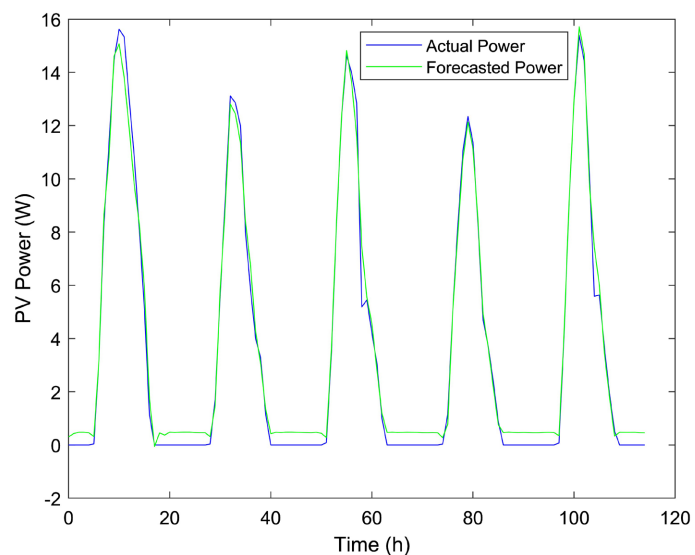


Figure 6. Wide neural network forecast of solar PV power generation in Bamenda, Cameroon.

The forecasting results revealed a small variation between the actual values of solar PV power generation and forecasted values of solar PV power generation.

4. Conclusions

The objective of this study was to identify an accurate MLM for forecasting of Solar PV power generation in Bamenda, North West Region of Cameroon. Raw data with 35,064 records and 7 features (six input and one output) was obtained from Photovoltaic Geographic Information System (PVGIS), hosted by the European Commission. The data provided the Solar PV power generation potential of a 20-kW solar PV system for Bamenda, Cameroon. The raw data was preprocessed using Microsoft Excel and MATLAB software. After preprocessing of the 35,064 data records, 33,000 records (94%) were retained for the training and testing of the MLMs. The preprocessed data with 33,000 records was split into training dataset of 26,400 records which represented 80% of the preprocessed data and testing dataset of 6600 records which represented 20% of the preprocessed dataset.

Using the training dataset, the MLMs were trained and validated using two validation techniques, namely: hold-out validation and re-substitution validation, providing valuable insights into MLM evaluation methodologies. Hold-out validation technique produced more accurate results than re-substitution validation technique, thereby underscoring the need for choosing an appropriate validation technique in training MLMs for solar PV power forecasting. Training and validation of the MLMs were followed by the testing of the MLMs using the test dataset. Wide Neural Network emerged as the most accurate MLM, with a RMSE of 9.377 W and R^2 of 0.999. Being the most accurate and better-yield MLM, Wide Neural Network was then used to forecast solar PV power generation in Bamenda, Cameroon, with results showing a small variation between actual solar PV power and forecasted solar PV power.

This study demonstrates significant advancements in employing machine learning for energy forecasting, a key step towards sustainable energy management. Overall, this study implies that adopting advanced MLMs like Wide Neural Networks, alongside rigorous validation techniques could be instrumental in optimizing solar PV power production. This approach not only enhances forecasting accuracy but also supports energy planning strategies, fostering sustainable development in the North West Region of Cameroon and potentially serving as a model for other regions with similar energy needs.

In this study, Wide Neural Network selected as the most accurate MLM is limited to solar PV power forecasting within the MATLAB environment. The deployment of this MLM to the cloud to enable real-time solar PV power forecasting is recommended.

Although this study demonstrates the utility of machine learning application in energy forecasting and the performance effectiveness of various MLMs, a limitation of the study is that the MLMs in the study were trained on a single location and the results cannot be generalized. Furthermore, performance checks of the

MLM on external sites were not performed. The direction of future research could address these limitations.

Acknowledgements

The authors are grateful to the Responsible Artificial Intelligence Network for Climate Action in Africa (RAINCA) Consortium, made up of the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM), the West African Science Centre on Climate Change and Adapted Land Use (WASCAL), and AKADEMIYA2063 for providing funding for this research with the support of IDRC (Grant #: 109705-001/002).

The Authors acknowledge the institutional support of The University of Bamenda, Cameroon.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Rahman, M.M., Islam, M.A., Karim, A.H.M.Z. and Ronee, A.H. (2012) Effects of Natural Dust on the Performance of PV Panels in Bangladesh. *International Journal of Modern Education and Computer Science*, **4**, 26-32. <https://doi.org/10.5815/ijmecs.2012.10.04>
- [2] Zazoum, B. (2022) Solar Photovoltaic Power Prediction Using Different Machine Learning Methods. *Energy Reports*, **8**, 19-25. <https://doi.org/10.1016/j.egy.2021.11.183>
- [3] Raghuvanshi, S.P., Raghav, A.K. and Chandra, A. (2008) Renewable Energy Resources for Climate Change Mitigation. *Applied Ecology and Environmental Research*, **6**, 15-27. https://doi.org/10.15666/aeer/0604_015027
- [4] Asoh, D.A. and Awangum, N.N. (2022) Low-Cost Automated PV Panel Dust Cleaning System for Rural Communities. *Smart Grid and Renewable Energy*, **13**, 173-199. <https://doi.org/10.4236/sgre.2022.138011>
- [5] Gaviria, J.F., Narváez, G., Guillen, C., Giraldo, L.F. and Bressan, M. (2022) Machine Learning in Photovoltaic Systems: A Review. *Renewable Energy*, **196**, 298-318. <https://doi.org/10.1016/j.renene.2022.06.105>
- [6] Mbinkar, E.N., Asoh, A.D., Tchuidjan, R. and Baldeh, A. (2021) Design of a Photovoltaic Mini-Grid System for Rural Electrification in Sub-Saharan Africa. *Energy and Power Engineering*, **13**, 91-110. <https://doi.org/10.4236/epe.2021.133007>
- [7] Narvaez, G., Bressan, M., Pantoja, A. and Giraldo, L.F. (2023) Climate Change Impact on Photovoltaic Power Potential in South America. *Environmental Research Communications*, **5**, Article ID: 081004. <https://doi.org/10.1088/2515-7620/acf02e>
- [8] Panahian, M., Ghosh, S. and Ding, G. (2017) Assessing Potential for Reduction in Carbon Emissions in a Multi-Unit of Residential Development in Sydney. *Procedia Engineering*, **180**, 591-600. <https://doi.org/10.1016/j.proeng.2017.04.218>
- [9] Woldegiyorgis, T.A., Benti, N.E., Chaka, M.D., Semie, A.G., Habtemicheal, B.A., Assamnew, A.D., *et al.* (2023) Harnessing Solar Power: Predicting Photovoltaic Potential in Fiche, Oromia, Ethiopia with Artificial Neural Networks. *Scientific African*, **21**, e01884. <https://doi.org/10.1016/j.sciaf.2023.e01884>

- [10] Panagea, I.S., Tsanis, I.K., Koutroulis, A.G. and Grillakis, M.G. (2014) Climate Change Impact on Photovoltaic Energy Output: The Case of Greece. *Advances in Meteorology*, **2014**, Article ID: 264506. <https://doi.org/10.1155/2014/264506>
- [11] Erten, M.Y. and Aydilek, H. (2022) Solar Power Prediction Using Regression Models. *International Journal of Engineering Research and Development*, **14**, s333-s342. <https://doi.org/10.29137/umagd.1100957>
- [12] Van Tai, D. (2019) Solar Photovoltaic Power Output Forecasting Using Machine Learning Technique. *Journal of Physics: Conference Series*, **1327**, Article ID: 012051. <https://doi.org/10.1088/1742-6596/1327/1/012051>
- [13] Alcañiz, A., Grzebyk, D., Ziar, H. and Isabella, O. (2023) Trends and Gaps in Photovoltaic Power Forecasting with Machine Learning. *Energy Reports*, **9**, 447-471. <https://doi.org/10.1016/j.egy.2022.11.208>
- [14] Pavithra, C.V., Divya, R., Bavithra, K. and Jeyashree, A. (2022) Machine Learning for Solar Power Forecasting. *Mathematical Statistician and Engineering Applications*, **71**, 1574-1591.
- [15] Nfah, E.M., Ngundam, J.M. and Tchinda, R. (2007) Modelling of Solar/Diesel/Battery Hybrid Power Systems for Far-North Cameroon. *Renewable Energy*, **32**, 832-844. <https://doi.org/10.1016/j.renene.2006.03.010>
- [16] Nfah, E.M. and Ngundam, J.M. (2009) Feasibility of Pico-Hydro and Photovoltaic Hybrid Power Systems for Remote Villages in Cameroon. *Renewable Energy*, **34**, 1445-1450. <https://doi.org/10.1016/j.renene.2008.10.019>
- [17] MINEE & KEEI (2017) A Study for Establishment of the Master Plan of Renewable Energy in Cameroon. Korea Energy Economics Institute.
- [18] Maps of India (2023) Political Map of Cameroon. <https://www.mapsofindia.com/world-map/cameroon/>
- [19] EU Science Hub (2023) Photovoltaic Geographical Information System (PVGIS). https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis_en
- [20] Anuradha, K., Erlapally, D., Karuna, G., Srilakshmi, V. and Adilakshmi, K. (2021) Analysis of Solar Power Generation Forecasting Using Machine Learning Techniques. *E3S Web of Conferences*, **309**, Article ID: 01163. <https://doi.org/10.1051/e3sconf/202130901163>
- [21] Sharkawy, A., Ali, M., Mousa, H., Ali, A. and Abdel-Jaber, G. (2022) Machine Learning Method for Solar PV Output Power Prediction. *SVU-International Journal of Engineering Sciences and Applications*, **3**, 123-130. <https://doi.org/10.21608/svusrc.2022.157039.1066>
- [22] Jacques, M.R.J., Raoul, D.N.S., Wira, P., Wulfran, F.M. and Saatong, K.T. (2023) Solar Irradiance Forecasting Based on Deep Learning for Sustainable Electrical Energy in Cameroon. *International Journal of Smart Grid*, **7**, 61-68.
- [23] Mahmud, K., Azam, S., Karim, A., Zobaed, S., Shanmugam, B. and Mathur, D. (2021) Machine Learning Based PV Power Generation Forecasting in Alice Springs. *IEEE Access*, **9**, 46117-46128. <https://doi.org/10.1109/access.2021.3066494>
- [24] Alhmoud, L., Al-Zoubi, A.M. and Aljarah, I. (2022) Solar PV Power Forecasting at Yarmouk University Using Machine Learning Techniques. *Open Engineering*, **12**, 1078-1088. <https://doi.org/10.1515/eng-2022-0386>
- [25] Gaboitaolelwe, J., Zungeru, A.M., Yahya, A., Lebekwe, C.K., Vinod, D.N. and Salau, A.O. (2023) Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison. *IEEE Access*, **11**, 40820-40845.

- <https://doi.org/10.1109/access.2023.3270041>
- [26] Essam, Y., Ahmed, A.N., Ramli, R., Chau, K., Idris Ibrahim, M.S., Sherif, M., *et al.* (2022) Investigating Photovoltaic Solar Power Output Forecasting Using Machine Learning Algorithms. *Engineering Applications of Computational Fluid Mechanics*, **16**, 2002-2034. <https://doi.org/10.1080/19942060.2022.2126528>
- [27] Ojo, O.S. and Ogunjo, S.T. (2022) Machine Learning Models for Prediction of Rainfall over Nigeria. *Scientific African*, **16**, e01246. <https://doi.org/10.1016/j.sciaf.2022.e01246>
- [28] Belu, R. (2013) Artificial Intelligence Techniques for Solar Energy and Photovoltaic Applications. In: *Handbook of Research on Solar Energy Systems and Technologies*, IGI Global, 376-436. <https://doi.org/10.4018/978-1-4666-1996-8.ch015>
- [29] MathWorks (2024) Train Regression Models in Regression Learner App. <https://www.mathworks.com/help/stats/train-regression-models-in-regression-learner-app.html>
- [30] MathWorks (2023) Select Data for Regression or Open Saved App Session. <https://www.mathworks.com/help/stats/select-data-and-validation-for-regression-problem.html>
- [31] Isma'il, M. and Aliyu, S. (2023) Daily Solar Radiation Forecasting for Northwest Nigeria Using Long Short-Term Memory. *International Journal of Science for Global Sustainability*, **9**, 142-149. <https://doi.org/10.57233/ijsgs.v9i1.407>
- [32] Olatomiwa, L., Mekhilef, S., Shamshirband, S., Mohammadi, K., Petković, D. and Sudheer, C. (2015) A Support Vector Machine-Firefly Algorithm-Based Model for Global Solar Radiation Prediction. *Solar Energy*, **115**, 632-644. <https://doi.org/10.1016/j.solener.2015.03.015>
- [33] Mohan, A. and Ngwira, M. (2019) Monthly Average Irradiation Forecasting for Malawi's Solar Resources. *International Journal of Innovative Technology and Exploring Engineering*, **8**, 1049-1060.
- [34] Allam, G.H., Elnaghi, B.E., Abdelwahab, M.N. and Mohammed, R.H. (2021) Using Machine Learning to Forecast Solar Power in Ismailia. *International Journal of Scientific and Research Publications*, **11**, 238-244. <https://doi.org/10.29322/ijsrp.11.12.2021.p12033>
- [35] Olatunde, O.O., Samuel, O.O., Israel, E. and Babatunde, A. (2021) Autoregressive Neural Network Models for Solar Power Forecasting over Nigeria. *Journal of Solar Energy Research*, **7**, 983-996.