*Original Research Article*

Comparative Analysis of Hybrid Machine Learning Models for Photovoltaic Energy Output Prediction: A Case Study from Nigeria's South-South Region

.

ABSTRACT

|  |
| --- |
| This study evaluated machine learning models for predicting photovoltaic (PV) energy output using daily meteorological data (temperature, relative humidity, wind speed, and solar radiation) from Nigeria’s South-South region over a decade (2012–2022). Data from Uyo, sourced from the University of Uyo’s Geography Department, were aggregated with PV output metrics and processed into a time-series dataset. Models including support vector machine (SVM), multiple linear regression (MLR), artificial neural network (ANN), deep neural network (DNN), gradient boosted decision tree (GBT), and hybrid architectures—convolutional neural network with long short-term memory (CNN-LSTM), reinforcement learning with LSTM (RL-LSTM), and their attention-enhanced variants (CNN-A-LSTM and Reinforcement Learning with Attention-enhanced LSTM (RL-A-LSTM)—were assessed using error metrics (MSE, MAE, RMSE, MAPE) and success rate (SR). The RL-A-LSTM model outperformed others, achieving the lowest MSE (0.004), MAPE (2.2%), MAE (0.004), and RMSE (0.063), with the highest SR (0.97), demonstrating exceptional accuracy in daily PV energy forecasting. CNN-A-LSTM followed closely (MSE: 0.005, SR: 0.96), while traditional models like SVM (MSE: 0.015) and MLR (MSE: 0.025) lagged due to limited nonlinear adaptability. The results accentuate the superiority of reinforcement learning-based hybrid models in capturing complex spatiotemporal dependencies, enabling precise predictions even under fluctuating humidity and irradiance conditions. Practically, these models enhance grid stability and operational efficiency by reducing forecast errors by up to 83% compared to conventional methods, enabling proactive energy management and cost savings. This study advocates integrating RL-A-LSTM into real-world PV systems to optimize renewable energy utilization, aligning with global decarbonization goals. The findings highlight the transformative potential of attention mechanisms and reinforcement learning in advancing sustainable energy systems. |

*Keywords: Machine learning, photovoltaic, meteorological data, reinforcement learning, forecasting*

1. INTRODUCTION

The global energy landscape is profoundly transforming as the world shifts towards more sustainable and renewable energy sources [1]. As stated by [2], renewable energy sources are the leading alternative solution to the increasing use of fossil fuels. Photovoltaic (PV) systems, which convert sunlight directly into electricity, are at the forefront of the transformation. With their capacity to provide clean, reliable, and decentralised energy, [3] noted that PV systems have become critical in reducing greenhouse gas emissions and combating climate change.

However, [4] stated that the performance of PV systems is influenced by several factors, including the geographic location, the angle of installation (tilt angle of PV module), and, most notably, the prevailing meteorological conditions. According to [5] Vidyanandan (2017), solar irradiance (the primary source of energy for PV systems), temperature (which affects PV cell efficiency and energy output), humidity (influences energy losses due to heat dissipation), wind speed (impacts cooling and energy losses) and shading (reduces energy output due to obstructed sunlight) are among the vital meteorological parameters that significantly impact the efficiency and output of PV systems. Accurate performance evaluation of these systems requires a good understanding of these meteorological factors and their complex interactions.

Although conventional performance evaluation methods have proven useful, they often fail to capture the intricate relationships between meteorological variables and photovoltaic (PV) system performance [6]. This shortcoming has led to the exploration of more advanced techniques, such as machine learning (ML), which can effectively model and predict PV system performance using historical and real-time data [4]. As a subset of artificial intelligence, [7] highlighted that machine learning has transformed numerous fields through its ability to learn from data, identify patterns, and make highly accurate predictions. In relation to PV systems, ML algorithms possess the capacity to be trained to analyse vast datasets of meteorological and performance data, revealing hidden insights and providing precise performance forecasts [8]. This sophisticated approach not only deepens the understanding of PV system behaviour under varying environmental conditions but also aids in optimising system design and operation.

This study aims to characterise the performance of a PV system based on the meteorological data of the installation location using a machine learning algorithm. The primary goal is to provide valuable insights for PV system optimisation and performance improvement.

2. Review of related LITERATURE

Studies have shown that meteorological data can significantly impact the efficiency of the energy output of a PV system [9]. Ref [10] reported that high temperatures reduce the efficiency of PV modules, while solar irradiance directly affects power output. Ref [11] analyzed the effect of temperature and solar irradiance on PV efficiency in various climates. Ref [12] developed a model to predict PV power output based on meteorological data for Vietnam. However, there is a need for more location-specific studies, as meteorological conditions vary significantly across different regions.

Ref [13] experimentally analyzed how fluctuating solar irradiation impacts photovoltaic (PV) efficiency and power output over 14 days. Using Microsoft Office Excel, they developed a novel colour contour method to categorize irradiance patterns via spectral analysis, enhancing predictive accuracy for energy generation. Although the study qualitatively linked irradiance variability to performance changes, it highlighted a need for quantitative models to precisely correlate irradiance levels with power outputs.

Ref [14] compared PV panels with and without cooling systems, finding a 15% efficiency for cooled panels versus 13% for uncooled ones over five days of temperature and radiation measurements. Their results emphasized temperature control’s role in mitigating efficiency losses. However, the study lacked quantitative analysis of temperature-power relationships, which highlights the need for advanced predictive models to enhance PV performance under thermal variations.

Ref [15] identified strong correlations between solar panel efficiency and five meteorological parameters (temperature, humidity, wind speed, solar intensity, dew point), determining a 26° tilt angle for optimal output. However, their correlation model showed absolute errors of 0.08–1.20% compared to experimental data, revealing a gap in quantitative precision.

Ref [16] demonstrated that Uyo’s lower relative humidity yielded higher solar generation efficiency than Port Harcourt, linking humidity inversely to PV performance. While the study emphasized regional climatic impacts, it lacked quantitative analysis of humidity-power output relationships.

Ref [17] simulated shading impacts on polycrystalline PV panels using MATLAB, showing a 33% power loss in fully shaded cells and validating bypass diodes’ mitigating role. However, the study omitted quantitative analysis of shading patterns’ specific effects on degradation, warranting research to model shading-induced losses systematically.

Ref [18] quantified how temperature (1.85–20.22% power loss) and humidity (up to 32.24% decline) reduce PV output, while wind mitigates losses via cooling and dust removal. Despite these insights, the correlations between environmental factors and performance lacked precision, necessitating advanced models to improve predictive accuracy for diverse climatic conditions.

3. Materials and Methods

The study requires a computer with a minimum of 8 GB RAM, 64 GB storage, and an Intel Core i3 processor (31XX series or newer). Compatible operating systems include Linux (Ubuntu 16.08 or later), macOS 10.10 (Yosemite or later), or Microsoft Windows 7 or newer. Essential software tools comprise a word processor (such as Microsoft Word), a spreadsheet application (such as MS Excel), and a photovoltaic (PV) system simulator for modelling and analysis. These specifications ensure seamless data processing, simulation, and documentation of experimental results.

**3.1 Methods**

**3.1.1 Predictive model development**

The preliminary procedure for the ML-based prediction model is presented in the section. From the 10-year meteorological data collected, solar radiation (S), temperature (T), humidity (H), wind speed (W) and PV energy output (E) were extracted. To combine meteorological data with weather data for creating a PV energy prediction model, the following steps as illustrated in Fig. 1 are applicable.

* 1. **Data Collection**: Data was collected from Geography department of the University of Uyo. The data collected included temperature, solar radiation, wind speed and relative humidity data. PV output data was sourced from an existing system in Uyo, Akwa Ibom State at BET Integrated Services, a solar company.
	2. **Data Pre-processing**: The raw meteorological and PV output data (2012–2022), sourced from the University of Uyo, underwent rigorous pre-processing to ensure robustness. Outliers were identified and removed using Interquartile Range (IQR) for domain-specific thresholds (such as solar radiation >1,400 W/m², humidity >100%) and Z-score ($\left|Z\right|>3$) for statistical anomalies, reducing the dataset by 8.2% but enhancing prediction stability (12% RMSE reduction). Missing values were addressed through linear interpolation for gaps $\leq 3$ days, while prolonged missing intervals (>50%) were discarded. Features were normalized to [0,1] using Min-Max scaling, and temporal variables (7-day solar radiation averages, cyclical day/month encoding) were processed to capture seasonal patterns. The dataset was partitioned temporally into training (2012–2019), validation (2020–2021), and test (2022) sets, preserving chronological dependencies.

 **Fig. 1. PV energy output prediction methodology.**

* 1. **Feature Selection**: Identification and selection of the most relevant features that impact PV energy output was done with the application of variance inflation factor (VIF).
	2. **Model Development**: Framework for the development of a predictive model using machine learning algorithm (RL) are highlighted in the succeeding sub-sections.
	3. **Model Training and Validation**: Model training using historical data to validate its performance using metrics like root mean square error (RMSE), mean absolute error (MAE), and R-squared (R²) values will be concluded after the pre-processing of the PV energy output data. Fine-tuning of the model parameters to improve prediction accuracy was integrated to enhance model accuracy.

**3.1.2 Fundamental relationship**

The energy output $\left(E\right)$ of a PV system can be generally modelled as a function of solar radiation $\left(S\right)$, temperature $\left(T\right)$, humidity $\left(H\right)$, and wind speed $\left(W\right)$, as expressed in Equation (1) [19]:

 $E=f(S, T, H, W)$ (1)

Each prediction model fashions a relationship along elements of Equation 3.1 to first obtain the global horizontal index (GHI) that further give an idea of the output energy generated.

**3.1.3 Linear and multiple linear regression (MLR)**

A simple linear regression model can be represented as given in Equation (2) [20]:

$E=β\_{O}+β\_{1}S+β\_{2}T+β\_{3}H+β\_{4}W+ϵ$ (2)

where $β\_{O}$ is the intercept, $β\_{1}, β\_{2}, β\_{2}, β\_{3}, β\_{4}$ are coefficients of each variable, and $ϵ$ is the error term.

According to [21], multiple linear regression (MLR) takes a different approach in that the main target is the correlation between the input features and response variables. MLR builds upon the foundation of linear regression (LR), which traditionally models the relationship between a single predictor and a dependent variable. MLR extends this by incorporating multiple independent variables simultaneously to better capture complex relationships. This approach is widely employed in forecasting applications due to its ease of implementation, interpretable results, and capability to detect anomalies or outliers in the predictor dataset. The mathematical representation of the MLR model is provided in Equation (3).

$E=b\_{0}+\sum\_{i=1}^{j}b\_{i}X\_{i}+ε$ (3)

where the model’s outcome and input features are denoted by $E$ and $X$, respectively, $b\_{0}$ is a constant parameter, and bi represents a regression coefficient for each $i-th$ independent variable, $i = 1, . . . , j$. The model’s error term (residual) is denoted by $ε$.

**3.1.4 Polynomial regression model**

If the relationship is non-linear, a polynomial regression can be used as exemplified in Equation (4) [20]:

$E=β\_{O}+β\_{1}S+β\_{2}T+β\_{3}H+β\_{4}W+β\_{5}S^{2}+β\_{6}T^{2}+β\_{7}H^{2}+β\_{4}W^{2}+ϵ$ (4)

**3.1.5 Multivariable regression model**

Considering interaction terms:

$$E=β\_{O}+β\_{1}S+β\_{2}T+β\_{3}H+β\_{4}W+β\_{5}S.T+β\_{6}S.H+β\_{7}S.W+β\_{8}T.H+β\_{9}T.W+β\_{10}H.W$$

 …(5)

**3.1.6 Predictive model using machine learning**

For more accurate predictions, Neural Networks (NN) will be employed. The basic structure for a machine learning model is expressed in Equation (6):

 $E=Model(S, T, H, W)$ (6)

The machine learning model is to be trained using collected meteorological data for the case study to find the best approximation of the function $f(S, T, H, W)$. (W/m2).

**3.1.7 Long short-term memory (LSTM) for time-series modelling**

In this study, in order to capture long-term dependencies in the sequential data obtained, a type of Recurrent Neural Network (RNN) that uses less memory known as long short-term memory (LSTM) will be employed. At time $t$, given input $X\left(t\right)$ (meteorological data) LSTM predicts $\hat{y}\left(t\right)$, the PV energy output.

**Mathematical Formulation of LSTM:** The LSTM cell consists of Forget Gate $\left(f\_{t}\right)$ which decides the part of the previous state $C\_{t-1}$ to discard, Input Gate $\left(i\_{t}\right)$ that decides what information to add to the cell state, Cell State $\left(C\_{t}\right)$ which is the internal memory of the cell and Output Gate $\left(o\_{t}\right)$ which decides the output of the cell. These culminate in an LSTM cell relationship as expressed in Equations (7) to (12).

 $f\_{t}=σ\left(W\_{f} . \left[h\_{t-1},X\_{t}\right]+b\_{f}\right)$ (7)

 $i\_{t}=σ\left(W\_{i} . \left[h\_{t-1},X\_{t}\right]+b\_{i}\right)$ (8)

 $\tilde{C}\_{t}=\tanh(\left(W\_{C} . \left[h\_{t-1},X\_{t}\right]+b\_{C}\right))$ (9)

 $C\_{t}=f\_{t}⊙C\_{t-1}+i\_{t}⊙\tilde{C}\_{t}$ (10)

 $o\_{t}=σ\left(W\_{o} . \left[h\_{t-1},X\_{t}\right]+b\_{o}\right)$ (11)

 $h\_{t}=o\_{t}⊙\tanh(\left(C\_{t}\right))$ (12)

where $h\_{t}$ represents the hidden state (output) of the LSTM at time $t$, $W\_{f}$, $W\_{i}$, $W\_{C}$, $W\_{o}$ are weight metrices for gates, $b\_{f}$, $b\_{i}$, $b\_{C}$, $b\_{o}$ are bias terms, $σ$ is a sigmoid activation function, $\tanh(\left(\right))$ is the hyperbolic tangent activation and $⊙$ represent element-wise multiplication.

**3.1.8 Reinforcement learning (RL) integration**

Reinforcement learning involves learning an optimal policy $π$ that maximises a reward signal $R$. The RL agent (predictor) interacts with the environment (meteorological data and PV energy outputs) to minimize the error between predicted and actual energy outputs.

**States, Actions, and Rewards**

**State** $\left(S\_{t}\right)$: The state at time $t$ consists of past meteorological data and previous PV energy output predictions:

 $S\_{t}=\left\{X\left(t-k\right), …,X\left(t-1\right),\hat{y}\left(t-1\right)\right\}$ (13)

where $k$ is the look-back window.

**Action** $\left(α\_{t}\right)$: The predicted PV energy output $\hat{y}\left(t\right)$.

**Reward** $\left(R\_{t}\right)$: The reward is a function of the prediction error as expressed in Equation 3.14:

 $R\_{t}=-\left|y\left(t\right)-\hat{y}\left(t\right)\right|$ (14)

where $y\left(t\right)$ is the actual PV output at $t$, and $\hat{y}\left(t\right)$ is the predicted output.

**Policy and Value Functions:** The RL agent optimizes a policy $π\left(a\_{t}|S\_{t}\right)$ to maximise the expected cumulative reward $G\_{t}$, where:

 $G\_{t}=\sum\_{k=0}^{\infty }γ^{k}R\_{t+k}$ (15)

where $γ$ is the discount factor $\left(0<γ\leq 1\right)$.

Using Q-learning (value-based RL), the agent learns a Q-value function $Q\left(S\_{t},a\_{t}\right)$ that represents the expected reward for taking action $a\_{t}$ in state $S\_{t}$, as portrayed in the Equation (16):

 $Q\left(S\_{t},a\_{t}\right)=E\left[R\_{t}+γ\max\_{a^{'}}Q\left(S\_{t+1},a^{'}\right)|S\_{t},a\_{t}\right]$ (16)

**3.1.9 Combined RL-LSTM framework development**

In the combined model perform specific role outline as follows:

1. LSTM predicts the next PV output $\hat{y}\left(t\right)$given the state $S\_{t}$.
2. The RL agent adjusts the prediction $\hat{y}\left(t\right)$ to minimize the prediction error by updating the policy 𝜋 π or Q-values.

**Procedures**

1. Initialize:
	1. LSTM network parameters $\left(W, b\right)$
	2. RL Q-value $Q\left(S, a\right)$ or policy $π$
2. At each time step $t$:
	1. Input the state $S\_{t}$ to the LSTM model to get an initial prediction $\hat{y}\_{L}\left(t\right)$.
	2. The RL agent refines the prediction by selecting an action $a\_{t}$ (adjustment) as given in Equation (17):

$\hat{y}\left(t\right)=\hat{y}\_{L}\left(t\right)+a\_{t}$ (17)

* 1. Observe the reward $R\_{t}=-\left|y\left(t\right)-\hat{y}\left(t\right)\right|$.
1. Update RL Q-values using the expression in Equation 3.18:

$Q\left(S\_{t},a\_{t}\right)⟵Q\left(S\_{t},a\_{t}\right)+α\left[R\_{t}+γ\max\_{a^{'}}Q\left(S\_{t+1},a^{'}\right)-Q\left(S\_{t},a\_{t}\right)\right]$ (18)

* 1. where $α$ is the learning rate.
1. Train LSTM:
	1. Use backpropagation through time (BPTT) to minimize the LSTM loss function, as given in Equation 3.19:

$L=\frac{1}{T}\sum\_{t=1}^{T}\left(y\left(t\right)-\hat{y}\left(t\right)\right)^{2}$ (19)

1. Repeat steps i-iv for multiple episodes until convergence.

The combined optimization objective is obtained by subtracting the product of the balancing hyperparameter $\left(λ\right)$ and RL cumulative reward function from the minimum of the loss prediction error as a function of LSTM parameters $\left(θ\right)$, as expressed in Equation (20):

 $\min\_{θ,π}L\left(θ\right)-λE\_{π}\left[G\_{t}\right]$ (20)

**3.1.10 Attention mechanism implementation**

The attention mechanism in the RL-A-LSTM model employs additive (Bahdanau-style) attention, where context vectors dynamically weight hidden states of the LSTM to prioritize critical temporal features. For each timestep $t$, attention scores $e\_{t}$ are computed as in (21):

 $e\_{t}=v^{T}\tanh(\left(W\_{h}h\_{t}+W\_{s}s\_{t-1}+b\right))$ (21)

Where $h\_{t}$ is the LSTM hidden state, $s\_{t-1}$ is the previous decoder state, and $v$, $W\_{h}$, $W\_{s}$, $b$ are learnable parameters. These scores were normalized through softmax to produce attention weights $α\_{t}$, which scale hidden states before final prediction. This allows the mode to focus on salient meteorological patterns (such as abrupt solar radiation drops) while suppressing noise.

**3.1.11 Hyperparameter selection process**

Hyperparameters (learning rate, LSTM units, attention dimension, batch size) were optimized using Bayesian optimization with a 5-fold time-series cross-validation on the 2012–2019 dataset. The search space included learning rates $\in \left[1^{-4},1^{-2}\right]$, LSTM units $\in \left[32, 256\right]$, and attention dimensions $\in \left[16, 64\right]$. The optimal configuration (learning rate: 0.001, LSTM units: 128, attention dimensions: 32) minimized validation MSE, avoiding overfitting in the process, as validated by loss curve plateaus and dropout regularization $\left(p=0.2\right)$

**3.2 Methods**

Using the data collected, the following outline were adopted for the combined model training:

1. Split the data into training and testing datasets
2. Train the Model
3. Use training set to fit the model

As outlined by [22], the performance of the proposed prediction model was assessed using a variety of performance indicators. The following indices expressed mathematically in Equations (21) to (25) were employed to evaluate the prediction model:

1. Mean Squared Error (MSE):

 $MSE=\frac{1}{n}\sum\_{i=1}^{n}\left(E\_{i}-\hat{E}\_{i}\right)^{2}$ (22)

1. Root Mean Squared Error (RMSE):

 $RMSE=\sqrt{\frac{1}{n}\sum\_{i=1}^{n}\left(E\_{i}-\hat{E}\_{i}\right)^{2}}$ (23)

1. Mean Absolute Error (MAE):

$MAE=\frac{1}{n}\sum\_{i=1}^{n}\left|E\_{i}-\hat{E}\_{i}\right|^{2}$ (24)

1. R-squared $\left(R^{2}\right)$:

 $\left(R^{2}\right)=1-\frac{1}{n}\sum\_{i=1}^{n}\left(E\_{i}-\hat{E}\_{i}\right)^{2}$ (25)

1. Mean Absolute Percentage Error (MAPE):

 $MAPE=\frac{1}{n}\sum\_{i=1}^{n}\frac{E\_{i}-\hat{E}\_{i}}{E\_{i}}×100$ (26)

where $E\_{i}$ is the actual energy output, $\hat{E}$ is the predicted energy output, $E$ is the mean of the actual energy output and $n$ is the number of observations.

This framework highlighted in the steps combined statistical methods and machine learning techniques to predict PV energy output based on historical meteorological data.

4. Results and Discussion

This section details the results achieved in this research. Python scripts used for the meteorological data pre-processing, model training, and evaluation. Where applicable, visual representations such as plots, charts, and tables of both pre-processed and predicted data are included to facilitate reference, illustration, and inference.

**4.1 Data Interpretation**

Fig. 2 to 5 show the plot of temperature, solar radiation, relative humidity, and wind speed recorded in the year 2022 for Uyo. From Fig. 2, it is observed that for year 2022, highest temperature of $30.2^{0}C$ was recorded in December and the lowest temperature $\left(26.7^{0}C\right)$ was observed in June while the yearly average was noted to be $28^{0}C$.

**Fig. 2. Average monthly temperature for year 2022**

Highest wind speed of $108.8 {m}/{s}$ was observed in the month of March while the lowest $64.3 {m}/{s}$ was recorded in October as illustrated in Fig. 3.

**Fig. 3. Average monthly wind speed for year 2022**

From Fig. 4, the lowest humidity $\left(70\%\right)$ was recorded in December while the highest $\left(92\%\right)$ was recorded in October.

**Fig. 4. Average monthly relative humidity for year 2022**

Solar radiation of $16.7{W}/{m^{2}}$ being the highest for the year 2022 is observed in month of January while the lowest $\left(11.2{W}/{m^{2}}\right)$ was recorded in the month of the August as illustrated in Fig. 5.

**Fig. 5. Average monthly solar irradiation for year 2022**

**4.2 Gaussian Distribution of Solar Irradiance**

Solar irradiance typically exhibits a near-Gaussian distribution pattern over the course of a day, peaking around noon when the sun is at its zenith. This characteristic behaviour is reflected in the PV output, which follows a similar trend under clear sky conditions. Fig. 6 illustrates a Gaussian curve fitted to the 24-hour measured PV energy output data recorded on February 8, 2022; it highlights the alignment between the theoretical distribution and observed performance.



**Fig. 6. Measure data vs Gaussian fit**

**4.3 Evaluation of Hybrid RL-LSTM and RL-A-LSTM Prediction Model**

Fig. 7 compared the predicted PV energy output generated by the hybrid RL-LSTM model against the actual recorded power output over 10 years (2012-2022). The predicted values (represented by the blue dots) follow a generally stable progression, reflecting the model's understanding of the dataset's patterns. However, the orange markers for the actual power output exhibit sudden changes, such as a sharp drop at the beginning and a rapid rise around the midpoint of the timesteps. These abrupt changes in actual power output indicate variations in the real-world data that may not have been fully captured or anticipated by the model.



**Fig. 7. RL-LSTM predicted versus actual (measured) PV power output**

This discrepancy between the predicted and actual values highlights potential gaps in the model's learning process or limitations in the input dataset. Though the model demonstrated a reasonable ability to maintain predictions within a stable range, the deviation during sharp transitions suggests that additional training data or improved feature engineering (such as incorporating more dynamic factors like weather anomalies or maintenance events) might enhance prediction accuracy. Fundamentally, the results demonstrated the hybrid model's capability to generalize trends in PV energy output but emphasize the need for further optimization to capture abrupt fluctuations and refine forecasting reliability.

The need for improved feature engineering as observed in Fig. 7 necessitated the inclusion of attention-enhanced mechanism to further boost the understanding of the training data by the RL-LSTM which resulted in the RL-A-LSTM model as illustrated in Fig. 8. The inference drawn from the output highlighted how the attention-enhanced RL-LSTM model significantly improved the predictive performance of the solar energy output system. The model's predictions closely matched the actual power output, indicating that the attention mechanism successfully highlighted critical parts of the input sequence, which the LSTM then used to make more accurate predictions. The reward structure and the RL agent's actions demonstrated effective learning, with the agent optimizing its policy to maximize the energy output over time.



**Fig. 8. RL-A-LSTM predicted versus actual (measured) PV power output**

Moreover, the stability of the deep deterministic policy gradient (DDPG) RL model during training, combined with meaningful predictions during the validation phase, revealed that the system was capable of adapting and generalizing to unseen data. Fig. 4.7 exhibited low error margins, indicating that the model was robust and could reliably forecast power output. This level of accuracy and adaptability offered substantial promise for real-world applications in energy management systems, enabling improved decision-making and efficient resource allocation.

**4.4 Gaussian Distribution of Solar Irradiance**

The performance comparison tabulated in Table 1 of various machine learning models, including SVM, MLR, ANN, DNN, GBT, hybrid CNN-LSTM as well as RL-LSTM and their attention-enhanced counterparts (CNN-A-LSTM, and RL-A-LSTM), provide important insights into their predictive capabilities for PV energy output.

**Table 1: Selected prediction models metric evaluation**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **MSE** | **SR** | **MAPE** | **MAE** | **RMSE** |
| SVM | 0.015 | 0.89 | 3.2 | 0.011 | 0.122 |
| MLR | 0.025 | 0.85 | 4.5 | 0.015 | 0.158 |
| CNN-LSTM | 0.008 | 0.93 | 2.8 | 0.007 | 0.089 |
| RL-LSTM | 0.006 | 0.95 | 2.5 | 0.006 | 0.077 |
| ANN | 0.012 | 0.91 | 3.0 | 0.010 | 0.109 |
| DNN | 0.010 | 0.92 | 2.9 | 0.009 | 0.100 |
| GBT | 0.007 | 0.94 | 2.6 | 0.007 | 0.084 |
| CNN-A-LSTM | 0.005 | 0.96 | 2.4 | 0.005 | 0.071 |
| RL-A-LSTM | 0.004 | 0.97 | 2.2 | 0.004 | 0.063 |

The performance metrics comparison presented in Fig. 9 provides a detailed view of how various machine learning models perform across key metrics, such as MSE, MAE, RMSE, MAPE, and SR. The CNN-LSTM and RL-LSTM models as illustrated in Fig. 9 show significantly better performance, with lower error metrics compared to conventional models like SVM and MLR. For instance, RL-A-LSTM achieved the lowest MSE (0.004), RMSE (0.063), MAE (0.004), and MAPE (2.2%), along with the highest Success Rate (SR) of 0.97. CNN-A-LSTM follows closely, with an MSE of 0.005, RMSE of 0.071, and an SR of 0.96. In contrast, Conventional models such as SVM and MLR performed poorly, with MSE values of 0.015 and 0.025, respectively, and RMSE values of 0.122 and 0.158. These results highlight the limitations of simpler models in capturing the complexities of solar energy data.



 **Fig. 9: Comparison of PV energy output prediction models**

Attention-enhanced models like CNN-A-LSTM and RL-A-LSTM outperformed standard deep learning models, demonstrating the added value of attention mechanisms in focusing on the most relevant features during prediction. For example, the CNN-LSTM model, while effective with an MSE of 0.008 and an SR of 0.93, is outperformed by CNN-A-LSTM, which reduced the MSE to 0.005 and increased the SR to 0.96. Similarly, RL-LSTM achieved an MSE of 0.006 and SR of 0.95, but RL-A-LSTM improved these metrics significantly. This comparison accentuates the importance of hybrid and advanced architectures in high-stakes applications like renewable energy forecasting, where precision and reliability are critical. The results demonstrated that attention mechanisms and reinforcement learning contribute significantly to improving prediction accuracy, making them indispensable for capturing temporal and nonlinear dependencies effectively.

**4.5 Computational Efficiency Discussion**

The computational efficiency of the evaluated models varied significantly. Traditional models (SVM, MLR) exhibited low training times (minutes) and minimal memory requirements due to their simple architectures, although at the cost of higher prediction errors (MSE: 0.015–0.025). Intermediate models like ANN and DNN required moderate resources (1–2 hours training on CPU), while GBT used ensemble techniques for faster convergence (45–60 minutes) but struggled with high-dimensional temporal data. Hybrid architectures, particularly CNN-LSTM and RL-LSTM, demanded substantial computational resources (4–6 hours on GPU) due to their dual CNN-RNN structures and reinforcement learning loops, with RL-A-LSTM further increasing overhead (6–8 hours) from attention mechanisms. Despite this, RL-A-LSTM achieved the highest accuracy (MSE: 0.004) by dynamically prioritizing critical features, extenuating its resource intensity for mission-critical applications. In contrast, CNN-A-LSTM balanced efficiency and performance (MSE: 0.005, 4.5 hours training), making it suitable for moderately resourced environments.

**4.6 Limitations and Future Research Directions**

This study’s limitations include its reliance on region-specific data (Uyo, Nigeria), which may limit generalizability to arid or temperate climates, and the computational cost of RL-A-LSTM, hindering real-time deployment on edge devices. Additionally, the dataset excluded extreme weather events (sandstorms, for instance), potentially underestimating model validity. Future work should optimize RL-A-LSTM through quantization or pruning for edge compatibility, integrate multi-regional datasets to enhance climatic adaptability, and explore federated learning for decentralized training. Extending the model to hybrid renewable systems (PV-wind) and incorporating satellite-derived irradiance data could further improve predictive granularity and grid integration feasibility.

**4.7 Discussion of Results**

The findings of the study demonstrated the superior performance of advanced machine learning models, particularly the RL-A-LSTM model, in accurately predicting PV energy output. The plotted outputs of the RL-LSTM and RL-A-LSTM model highlighted its ability to minimize prediction errors, closely aligning with actual energy values with remarkable precision. This supported the conclusion that the RL-A-LSTM model achieved the lowest MSE and RMSE, showcasing its capacity to handle complex temporal dependencies and nonlinear relationships inherent in PV system outputs. In contrast, CNN-LSTM and CNN-A-LSTM models also performed well but exhibited slightly higher prediction errors, which could be attributed to challenges in generalizing to unseen data. These results emphasized the validity of reinforcement learning when integrated with LSTM architectures, allowing for dynamic adaptation and effective handling of real-time data.

Conventional models like SVM and MLR, as revealed in the visual analysis and metric comparisons, struggled to capture the complex interactions between meteorological factors and PV output, resulting in significantly higher errors. The CNN-LSTM model demonstrated its strength in processing temporal and spatial dependencies, making it an effective alternative to reinforcement learning in scenarios with constrained computational resources. However, the RL-LSTM and RL-A-LSTM models’ closer alignment with actual PV outputs highlighted their potential for real-time PV system performance evaluation and optimization. These discoveries accentuated the efficacy of advanced hybrid models like RL-LSTM and RL-A-LSTM for renewable energy forecasting, paving the way for more reliable and efficient energy management strategies in the context of increasing global reliance on PV systems.

The results demonstrated the superior performance of advanced machine learning models, particularly the RL-A-LSTM model, in accurately predicting PV energy output. The plotted outputs of the RL-LSTM and RL-A-LSTM models highlighted their ability to minimize prediction errors, closely aligning with actual energy values with remarkable precision. This supported the conclusion that the RL-LSTM model achieved the lowest MSE and RMSE, showcasing its capacity to handle complex temporal dependencies and nonlinear relationships inherent in PV system outputs. In comparison, CNN-LSTM and CNN-A-LSTM models also performed well but exhibited slightly higher prediction errors, which could be attributed to challenges in generalizing to unseen data. These results emphasized the effectiveness of reinforcement learning when integrated with LSTM architectures, allowing for dynamic adaptation and effective handling of real-time data. Compared to recent works, like that of [23] and [24], which used hybrid CNN-LSTM and transformer-based architectures, the RL-LSTM model demonstrated a comparable or superior predictive accuracy, particularly in its ability to adapt dynamically using reinforcement learning.

As noted earlier, the CNN-LSTM model demonstrated its strength in processing temporal and spatial dependencies, aligning with findings from [25] and [26], which highlighted the efficacy of CNN-LSTM in handling time-series meteorological data. When compared to recent advancements, such as hybrid ensemble approaches (like Gradient Boosting Transformers) in renewable energy forecasting, the RL-LSTM model maintained its edge by integrating reinforcement learning, which allowed for dynamic decision-making and real-time adaptability. These findings accentuated the efficacy of advanced hybrid models like RL-LSTM for renewable energy forecasting, paving the way for more reliable and efficient energy management strategies in the context of increasing global reliance on PV systems.

4. Conclusion

The study demonstrated that advanced hybrid machine learning models, particularly the reinforcement learning-based LSTM (RL-LSTM), significantly outperformed conventional methods in PV energy forecasting. The RL-LSTM model achieved superior accuracy across key error metrics (MSE, RMSE, MAE, MAPE), closely mirroring actual generation data, due to its dual capability to dynamically adapt to meteorological fluctuations through reinforcement learning and capture long-term temporal dependencies through LSTM architecture. Although the CNN-LSTM hybrid model also performed well, the RL-LSTM model exhibited greater effectiveness, highlighting the value of integrating adaptive learning mechanisms. In contrast, conventional models like SVM and MLR struggled with nonlinear PV output relationships, revealing their limitations. These findings emphasize the critical role of advanced hybrid models in enhancing renewable energy forecasting, with RL-LSTM offering a viable pathway to optimize PV system efficiency, improve grid management, and inform future reinforcement learning applications in sustainable energy solutions.

Competing interests

The authors declare that no competing interest between the authors exist

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declares that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

References

1. Fernández-Guillamón, A., Gómez-Lázaro, E., Muljadi, E., & Molina-García, Á. (2019). Power systems with high renewable energy sources: A review of inertia and frequency control strategies over time. *Renewable and Sustainable Energy Reviews, 115*, 109369. <https://doi.org/10.1016/j.rser.2019.109369>
2. Khamisani, A. A. (2018). Design methodology of off-grid PV solar powered system (A case study of solar powered bus shelter). *Energy Reports, 4*, 1-12. <https://doi.org/10.1016/j.egyr.2018.08.001>
3. Patel, M. R., & Beik, O. (2021). *Wind and Solar Power Systems: Design, Analysis, and Operation*. CRC Press. <https://doi.org/10.1201/9781003042952>
4. Park, J., Kim, J., Lee, S., & Choi, J. K. (2022). Machine learning-based photovoltaic energy prediction scheme by augmentation of on-site IoT data. *Future Generation Computer Systems, 134*, 1–12. <https://doi.org/10.1016/j.future.2022.05.001>
5. Vidyanandan, K. V. (2017). An overview of factors affecting the performance of solar PV systems. *Energy Scan, 27*(28), 216.
6. Wang, K., Qi, X., & Liu, H. (2019). A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Applied Energy, 251*, 113315. <https://doi.org/10.1016/j.apenergy.2019.113315>
7. Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. (2019). Advances in deep learning for renewable energy systems. *IEEE Access, 7*, 137589-137607. <https://doi.org/10.1109/ACCESS.2019.2942415>
8. Yagli, G. M., Yang, D., & Srinivasan, D. (2019). Automatic hourly solar forecasting using machine learning models. *Renewable and Sustainable Energy Reviews, 105*, 487–498. <https://doi.org/10.1016/j.rser.2019.02.006>
9. Singh, G. (2013). Performance analysis of a grid connected solar photovoltaic plant. *Energy Conversion and Management, 78*, 505-512. <https://doi.org/10.1016/j.enconman.2013.11.017>
10. Skoplaki, E., & Palyvos, J. A. (2009). On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Solar Energy, 83*(5), 614-624. <https://doi.org/10.1016/j.solener.2008.10.008>
11. Mondol, J. D., Yohanis, Y. G., & Norton, B. (2005). Long-term performance analysis of a grid-connected photovoltaic system in Northern Ireland. *Energy Conversion and Management, 46*(6), 909-926. <https://doi.org/10.1016/j.enconman.2004.06.011>
12. Nguyen, N. Q., Bui, L. D., Van Doan, B., Sanseverino, E. R., Di Cara, D., & Nguyen, Q. D. (2021). A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks a case study in Vietnam. *Electric Power Systems Research, 199*, 107427. <https://doi.org/10.1016/j.epsr.2021.107427>
13. Ibrahim, H., Othman, M. Y., & Abdellatif, M. A. (2019). Impact of solar irradiance variation and seasons on the performance of grid-connected PV systems. *Renewable Energy, 143*, 635-644. <https://doi.org/10.1016/j.renene.2019.05.044>
14. Sani, M. B., & Sule, A. H. (2020). Effect of temperature on the performance of photovoltaic modules in a tropical climate. *Solar Energy, 207*, 1302-1311. <https://doi.org/10.1016/j.solener.2020.07.059>
15. Sarmah, R., Das, S., & Choudhury, S. (2023). A comprehensive analysis of meteorological parameters affecting solar photovoltaic performance: Correlation and predictive modeling. *Renewable and Sustainable Energy Reviews, 182*, 113367. <https://doi.org/10.1016/j.rser.2023.113367>
16. Ettah, E. B., Akpan, P. E., & Udo, S. O. (2015). Comparative study of the effect of relative humidity on solar electricity generation in Uyo and Port Harcourt, Nigeria. *Renewable Energy, 83*, 1234-1241. <https://doi.org/10.1016/j.renene.2015.05.049>
17. Abdelaziz, A., Kamel, R. M., & Jurado, F. (2021). Shading effects on photovoltaic modules: Simulation and experimental validation. *IEEE Journal of Photovoltaics, 11*(2), 521-529. <https://doi.org/10.1109/JPHOTOV.2020.3048533>
18. Kazem, H. A., & Chaichan, M. T. (2016). Experimental analysis of the effect of dust’s physical properties on photovoltaic modules in Northern Oman. *Solar Energy, 139*, 68-80. <https://doi.org/10.1016/j.solener.2016.09.019>
19. Chataut, R., & Akl, R. (2020). Massive MIMO systems for 5G and beyond networks—overview, recent trends, challenges, and future research direction. *Sensors, 20*(10), 2753. <https://doi.org/10.3390/s20102753>
20. Bamisile, O., Huang, Q., Hu, W., & Oluwasanmi, A. (2022). Comprehensive assessment, review, and comparison of AI models for solar irradiance prediction based on different time/estimation intervals. *Scientific Reports, 12*(1), 13652. <https://doi.org/10.1038/s41598-022-13652-w>
21. Al-Dahidi, S., Alrbai, M., Alahmer, H., Rinchi, B., & Alahmer, A. (2024). Enhancing solar photovoltaic energy production prediction using diverse machine learning models tuned with the chimp optimization algorithm. *Scientific Reports, 14*(1), 69544. <https://doi.org/10.1038/s41598-024-69544-8>
22. Ghazvinian, H., Mousavi, S. M., Karami, H., Farzin, S., & Ehteram, M. (2019). Integrated support vector regression and an improved particle swarm optimization-based model for solar radiation prediction. *PLOS ONE, 14*(5), e0217634. <https://doi.org/10.1371/journal.pone.0217634>
23. Zhang, Y., Li, X., & Wang, Z. (2023). A hybrid CNN-LSTM model for photovoltaic power forecasting with attention mechanism. *IEEE Transactions on Sustainable Energy, 14*(1), 432-442. <https://doi.org/10.1109/TSTE.2022.3214567>
24. Wang, L., Chen, J., & Zhang, H. (2022). Transformer-based deep learning models for solar irradiance prediction. *Applied Energy, 315*, 118952. <https://doi.org/10.1016/j.apenergy.2022.118952>
25. Li, T., Yang, J., & Liu, Y. (2022). Gradient boosting machine with feature engineering for short-term photovoltaic power forecasting. *IEEE Transactions on Power Systems, 37*(3), 2148-2159. <https://doi.org/10.1109/TPWRS.2021.3112345>
26. Hamad, S., Ghalib, M.A., Munshi, A., Majid A. & Mostafa, A. Evaluating machine learning models comprehensively for predicting maximum power from photovoltaic systems. *Sci Rep* 15, 10750 (2025). <https://doi.org/10.1038/s41598-025-91044-6>