Comparative Analysis of LSTM-Based PV Power Forecasting Models with Climate-Adaptive Feature Selection in Abuja, Nigeria

David Akpuluma[1,2], Wolf-Gerrit Früh[1], Neda Firoz[3], James Abam[4], Mohammed Umar Bello[1], Comfort Williams[1], Ambrose Onne Okpu[1]
[1]*Heriot–Watt University, Edinburgh, EH14 4AS, Scotland, United Kingdom*
[2]*Centre for Nuclear Energy Studies, University of Port Harcourt, East-West Road, PMB 5323, Port Harcourt, Nigeria*
[3]*Sam Higginbottom University of Agriculture Technology and Sciences, Allahabad, 211007 Uttar Pradesh, India*
[4]*Rivers State University, Nkpolu-Oroworukwo, PMB 5080, Port Harcourt, Nigeria*
{da2027, w.g.fruh}@hw.ac.uk, david.akpuluma@uniport.edu.ng, akpoebi@gmail.com

Keywords: PV Power Forecasting, LSTM Models, Climatic Feature Selection, Renewable Energy, Solar Energy Prediction, Abuja, Autocorrelation Analysis, Cross-Correlation Analysis, Sustainable Energy Solutions.

Abstract: In this research, we analyse how Long Short-Term Memory (LSTM) models can predict photovoltaic (PV) power output, in Abuja, Nigeria by selecting specific climate features and model configurations. The rising energy needs due to population growth and urbanisation emphasise the importance of sustainable energy sources. This study aims to improve the accuracy of PV power forecasts for integrating power into the current electrical grid and enhancing energy management strategies. By analysing data from the ERA5 dataset that includes various climatic features, we rigorously trained and assessed the LSTM models. Our results indicate that specific window sizes and combinations of features notably enhance forecasting accuracy with a window size of 6 and a mix of meteorological and solar radiation features showing the performance metrics (MAE, RMSE, R^2). The study also underscores the significance of autocorrelation and cross-correlation analyses in optimizing model setups. Our findings suggest that LSTM models can accurately predict PV power output offering insights for maximizing energy usage in urban areas with similar climates. This research contributes to efforts aimed at reducing reliance on fossil fuels and promoting sustainable energy solutions. Future endeavours will explore integrating real-time data and incorporating additional climatic features to further refine forecasting models.

1 INTRODUCTION

As population continues to grow, so does the energy demand, which has become a necessity to life in the 21st century. Like fuel to an automobile, so is energy to life, especially electrical power supply. Access to affordable, clean and reliable power supply is a sustainable development goal (SDG) of the United Nations [1]. Whilst other continents are making strides and seeking smoother transition, some West African countries seem to be lagging considerably with regards to electrical power infrastructure. Available infrastructure is outdated and requires an immediate and large-scale overhaul, which now is bereft of political will [2].

For several reasons generating power from fossil fuels is no longer sustainable, hence the need for renewable energy. In spite of the prospects of renewable energy, its reliability and integration to already existing infrastructure is dependent on a couple of factors, one of which is its availability owing to the fact that its generation is dependent on weather or meteorological factors [3, 4]. In tropical climates like Sub-Saharan Africa, renewable energy such as solar has a very high potential in addressing the electricity deficiency experienced in the region especially in the most populous African country, Nigeria. Off-grid solar homes systems (SHSs) and community microgrids seem to be the most effective and practical solution to accessing power supply for rural dwellers, with advantages of reduced cost, environmental sustainability and ease of deployment, especially when the centralized power infrastructure is not reliable and would be capital intensive for individuals without access to such funds [1].

The capital city of Nigeria, Abuja, is quite developed but not left out in electrical power infrastructure challenges the region faces. Unprecedented population growth and urbanization have resulted in an escalating need for dependable and cost-effective electricity. Nevertheless, the current power infrastructure in Abuja is overburdened and experiences frequent interruptions, impacting economic activities and quality of life. Moreover, the utilization of fossil fuels for electricity generation is not environmentally sustainable due to ecological degradation and rising fuel costs. Renewable energy, specifically solar power, offers a feasible resolution to these challenges. Abuja, positioned in a tropical climate with ample sunshine, boasts a significant potential for harnessing solar energy. However, the successful integration of solar power into the current grid and its dependability are contingent upon precise forecasting of photovoltaic (PV) power production. Accurate PV power prediction can enhance the efficacy of solar energy systems, improve grid reliability, and enable more effective energy management approaches [5, 6].

Predicting power generation with a high level of accuracy, would enable power users in these regions without reliable power supply; to embrace this option of solar energy in photovoltaics as a reliable source of power. It can improve the efficiency of smart community microgrids where peer to peer trading of energy is possible.

Photovoltaic power generation is dependent on certain weather conditions like solar radiation intensity, temperature, wind speed and direction, cloud cover, humidity etc [4, 7]. Several models have been used in forecasting the PV power generation of renewable energy systems, some of which were traditional techniques or even hybrid models demonstrating excellent accuracy depending on the peculiarity of the prevailing conditions [7].

Long Short-Term Memory (LSTM) has been used for different time scales in predicting PV power output for several plants using climatic data. This deep learning method is robust and flexible. In some cases, providing accuracy of over 18% better than other benchmarked methods [8]. The functionality of any smart grid is dependent on the efficiency of the energy management technique employed. Energy management strategies are based on timescales (hourly, daily, weekly, monthly or yearly) depending on the purpose of its design. An RNN generates its output predictions from both current input and past data or experience, and where the distance between the cells is significant and vanishing gradient may tend to lose some information, LSTM solves this challenge by adding three gates (input gate, forget gate and output gate) to the RNN cell. So, it captures nonlinear relationships improving accuracy of the model [9]. In this study a Recurrent Neural Network (LSTM - Long Short-Term Memory) is employed to predict the power output of a PV system from climatic conditions.

The objectives of this paper are to:

- Demonstrate the possible outcome of PV power forecasting using LSTM with different climatic features for an urban city.
- Determine the most effective feature selection that enhances the accuracy of the LSTM model in forecasting PV power output in Abuja.

The importance of this investigation lies in its emphasis on refining PV power forecasting through sophisticated machine learning methodologies. By formulating and assessing Long Short-Term Memory (LSTM) models customized to Abuja's climatic conditions, this study strives to offer more precise forecasts of solar power generation. Subsequently, this can accelerate the acceptance of solar energy solutions, diminish reliance on fossil fuels, and contribute to sustainable development objectives in the region.

2 RELATED WORKS

In recent years, significant advancements have been made in the field of PV power forecasting, particularly with LSTM models and other machine learning techniques. In this overview we briefly discuss a range of research projects showcasing progress made so far, while also looking into specific areas that need improvement.

Machine Learning Model Optimization: For PV power prediction it is essential to choose relevant input features. Research highlights the significance of including weather conditions and time details to improve forecast accuracy [10, 11]. LSTM models, specifically tailored for time series data analysis, have shown superior performance in predicting power generation often surpassing other neural networks like GRU and MLP [12]. Moreover, methods such as XGBoost have also been successful in forecasting PV power output outperforming statistical methods, like SARIMA [13].

LSTM-Based Models for PV Power Forecasting: LSTM models have become quite popular for their ability to effectively handle the intermittent nature of solar irradiation and incorporate it into power grids. Studies have shown that LSTM models outperform Artificial Neural Network (ANN) models, especially when it comes to short term predictions [14]. Single variable LSTM models, which rely on PV output data demonstrate good accuracy in predicting one step ahead while multivariable models that consider weather factors excel in forecasting multiple steps ahead [15]. Additionally, research suggests that

LSTM models outshine Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN) models when series data. According to a study by [16], employing two-week datasets produced the best outcomes. These methodologies indicate that LSTM based models are exceptionally proficient in providing dependable forecasts, for solar photovoltaic power across various time frames.

2.1 Gaps in Existing Research

Despite the progress made so far, some areas still need further investigation in the literature.

1) Selecting Features Adapted to Climate; While many studies stress the importance of choosing features, more research is required on selecting features that adapt to conditions especially tailored for regions like Abuja.
2) Ensuring Robustness Across Varied Climates; Most studies concentrate on weather patterns. Unique climate conditions in urban areas such as Abuja call for customized approaches to enhance forecasting accuracy.
3) Incorporating Local Data; Limited research exists on incorporating climate data with advanced machine learning algorithms for predicting PV power. This integration is vital for optimizing energy usage in areas with weather patterns.
4) Examining Window Sizes; Current studies often lack assessments of how different window sizes impact forecasting accuracy. This study aims to fill this void by examining the effects of window lengths on the precision of LSTM models in forecasting PV power.

By tackling these gaps this research strives to build models that utilize feature selection to climate and advanced machine learning methods customized for Abuja's specific climate conditions, in Nigeria.

3 METHODOLOGY

This section outlines the methodology employed in the study, to develop and evaluate an LSTM model for predicting surface solar radiation downwards (ssrd_5) using historical meteorological data from the ERA5 dataset, including data preprocessing, feature selection, and the development and evaluation of an LSTM model for time series forecasting.

3.1 Dataset Description

The data used in this study are sourced from the ERA5 reanalysis dataset, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) through the Copernicus Climate Change Service (C3S) [17]. The ERA5 dataset offers a comprehensive reanalysis of global climate data, providing hourly estimates of various atmospheric, oceanic, and land surface parameters. The specific dataset used for this study is available at: ERA5 Reanalysis Dataset. The dataset covers the period from January 1, 2020, to December 31, 2022, with hourly time stamps. The dataset for the years 2023 and 2024 was not available at the time of the analysis.

This dataset provides valuable insights into the climatic conditions of the Abuja region over the specified period. It can be used for various analyses including wind energy potential assessment, solar radiation analysis, temperature and humidity studies, and cloud cover observation, which are crucial for understanding local weather patterns and for planning renewable energy projects.

The dataset includes the following meteorological variables measured at two different heights (10 meters and 100 meters), solar radiation parameters, temperature, dew point, surface pressure, and cloud cover. These variables are recorded for nine different locations in the vicinity of Abuja, Nigeria. The dataset was in csv file format which contained the time series data for the specified period and locations, with each variable labelled accordingly with the location suffix (_i). Table 1 shows variables and description of the dataset.

The data were collected for nine specific locations, each identified by a unique suffix (i where i ranges from 1 to 9). These locations are situated around Abuja, Nigeria, with the central point being at Latitude 9.0° N and Longitude 7.5° E, as shown in Fig. 1.

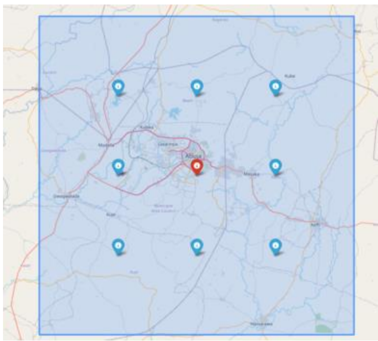

Figure 1: The geolocations of the study sites in Abuja, and a map of Nigeria.

Table 1: Data description.

Variable Type	Variable	Description
Meteorological Variables (Wind)	u10	10m_u_component_of_wind
	v10	10m_v_component_of_wind
	fg10	Wind gusts at 10m
	u100	100m_u_component_of_wind
	v100	100m_v_component_of_wind
Solar Radiation Variables	cdir	Clear sky direct solar radiation at surface
	fdir	Total sky direct solar radiation at surface
	ssrdc	Surface solar radiation downward clear sky
	ssrd	Surface solar radiation downwards
Temperature and Pressure Variables	t2m	2m temperature
	d2m	2m dew point temperature
	sp	Surface pressure
Cloud Cover Variables	lcc	Low cloud cover
	mcc	Medium cloud cover
	hcc	High cloud cover

3.2 Data Pre-Processing

Rows with missing values were dropped to ensure the dataset is complete and ready for analysis. The 'Timev' column, containing date and time information, was utilized to extract separate 'Date' and 'Time' columns. The 'Date' and 'Time' columns were converted to datetime formats to facilitate time series analysis.

1) Autocorrelation: Autocorrelation is a statistical measure used to analyze the degree of similarity between a given time series and a lagged version of itself over successive time intervals. It helps in understanding the repeating patterns, cycles, or trends within the data. For time series forecasting tasks, especially with LSTM (Long Short-Term Memory) models, autocorrelation analysis is crucial as it informs the selection of important features and appropriate model parameters. Before selecting features and building the model, it is essential to examine the autocorrelation of the target variable, which is, the surface solar radiation downwards (ssrd_5). By analyzing the autocorrelation plot, we can identify significant lags that exhibit strong correlations with the current time step. This information is valuable in determining the appropriate window size and the number of units for the LSTM model. The autocorrelation plot (shown in the Figure 2) indicates a strong correlation at a 24-hour repetition, suggesting a daily cycle in the data. The plot shows a clear 24-hour cycle, which means that the solar radiation values are highly correlated with the values from the previous day.
2) Feature Selection: Relevant features for the analysis were selected. These include meteorological parameters such as wind speed, temperature, solar radiation, and cloud cover at different time intervals for location 5 and location 8, since they had the highest correlation in our analysis.
3) Normalization: The numerical features were normalized to scale the data within a range of 0 to 1. This step was crucial for ensuring that the neural network training process converges more efficiently.

The dataset was further then transformed into sequences using a sliding window approach. The window size, which determines the number of time steps used to predict the next value, was experimented with different sizes ranging from 2 to 6. For each window size, the data was reshaped into overlapping sequences to create the input (X) and output (y) datasets.

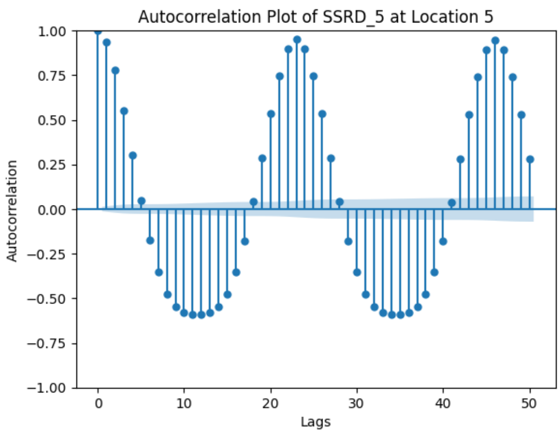

Figure 2: Autocorrelation of ssrd_5 at location 5.

4) Dataset Split: The dataset was divided into training, validation, and testing sets. The splitting approach used was chronological to preserve the temporal structure, ensuring the model is tested on unseen future data. This split helps in evaluating how well the model can predict future PV power output based on past data. Specifically, each sequence of weather parameters (X) was used to predict the subsequent solar radiation value (ssrd_5) (y). The sequences were split into training and testing sets using an 80-20 split ratio. The training set was further split into training and validation sets with an 80-20 split ratio for model evaluation.

3.3 Description/LSTM Model Architecture

The Long Short-Term Memory (LSTM) model is a type of recurrent neural network (RNN) designed to handle time series data and sequences. It addresses the vanishing gradient problem faced by traditional RNNs, enabling it to learn long-term dependencies.

Figure 3 represents the flow diagram of the LSTM approach proposed in this study.

3.3.1 Layers of the Proposed Model

1) Input Layer: The model takes sequences of shape (window_size, num_features), where window_size is the number of time steps in each input sequence, and num_features is the number of features in each time step.
2) LSTM Layer: The LSTM layer with 50 units processes the input sequences, capturing temporal dependencies:

$$\mathrm{LSTM}(x_t) \rightarrow (h_t). \tag{1}$$

3) Dense Layer: A Dense layer with a single unit produces the final output, predicting the target variable:

$$\mathrm{Dense}(h_t) \rightarrow (y_t). \tag{2}$$

4) Loss Function: The Mean Squared Error (MSE) is used as the loss function, measuring the average squared difference between the predicted values and the actual values:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{3}$$

3.3.2 Experimental Setup

The LSTM model was constructed using the Keras library, with an architecture comprising a single LSTM layer with 50 units and a ReLU activation function. The choice of 50 units was guided by the autocorrelation analysis. The number 50 was chosen to approximately cover two days (48 hours) of data, providing the model with sufficient information to capture the daily cycles effectively. While 50 is also a common default setting in many Keras examples, the decision to retain this number was supported by the autocorrelation findings from Figure 2. The strong 24-hour cycle observed in the data justifies using a window that spans multiple days.

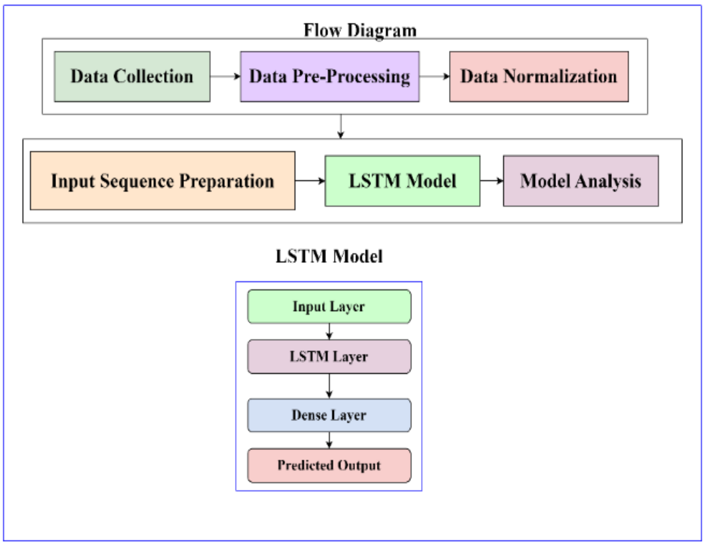

Figure 3: Flow diagram of the proposed model.

A Dense layer with a single neuron was added as the output layer to predict the solar radiation value. The model was compiled using the Adam optimizer and mean squared error (MSE) loss function. The model was trained on the training dataset with a batch size of 32 and a maximum of 100 epochs. Early stopping was employed with a patience of 5 epochs to prevent overfitting, ensuring the model retained the best weights based on validation loss. During cross-validation, different hyperparameters (e.g., number of LSTM units, batch size, learning rate) was tested to identify the optimal configuration that minimizes the validation loss. The model with the best performance across the cross-validation folds was selected as the final model. This model was then evaluated on the test set to confirm its performance. The dataset was divided into k folds (10). The model was trained and validated k times, each time using a different fold as the validation set and the remaining folds as the training set. For each fold, the same data preparation steps are followed: normalization, sequence preparation, and splitting into input (X) and target (y). The model is trained and evaluated on each fold, and performance metrics (e.g., MSE) are recorded. The performance metrics from all folds are averaged to provide a more accurate estimate of the model's performance. This average performance metric is considered the final evaluation of the model.

The trained model was evaluated on the test set to assess its performance using the mean squared error (MSE) metric. The training and validation loss over epochs were plotted to visualize the model's learning process and to identify any signs of overfitting. By following this methodology, the study aimed to develop a robust LSTM model for forecasting solar radiation using various meteorological features, contributing to more accurate weather predictions for the Abuja region. Table 2 presents the hyperparameters used in the proposed model.

Table 2: Hyperparameters description for the proposed model.

Hyperparameter	Window Size 2-6
Number of LSTM Units	50
Batch Size	32
Activation	ReLU
Epochs	100
Optimizer	Adam
Loss Function	Mean Squared Error (MSE)
Early Stopping Patience	5
Input Shape	(2, num_features)

3.4 Evaluation Metrics

1) Mean Absolute Error (MAE): It is the average of the absolute differences between the predicted and actual values. Mathematically, it can be represented as:

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

where:
y_i is the actual value,
$\hat{y}_i$ is the predicted value, and
n is the total number of data points.

2) Root Mean Squared Error (RMSE): It is the square root of the average of the squared differences between the predicted and actual values. Mathematically, it can be represented as:

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

3) R² (Coefficient of Determination): It is a statistical measure that represents the proportion of the variance for a dependent variable that is explained by an independent variable or variables in a regression model. Mathematically, it can be represented as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (6)$$

where:
SS_{res} is the sum of squares of residuals, i.e., $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$,
SS_{tot} is the total sum of squares, i.e., $\sum_{i=1}^{n}(y_i - \bar{y})^2$, and
$\bar{y}$ is the mean of the observed data.

4 RESULTS AND DISCUSSION

In this section we discuss the results of the experiments performed in this study.

4.1 Months Analysis

The prediction analysis focuses on several months throughout the dataset. The specific months highlighted include January 1 to February 29 and July 1 to August 31, for the years 2020 to 2022. These months were chosen to represent the dry and wet seasons characterised by the location. The varying weather conditions during these seasons were considered to test the robustness of the LSTM models.

4.2 Auto Correlation and Cross Correlation Analysis

Figure 2 shows the autocorrelation for location 5 with significant lags that influence the PV power output, determining the optimal number of input nodes for the LSTM model. This lag is crucial in setting up the sequence length for the LSTM.

The cross-correlation analysis in Figure 4 shows that locations 5 and 8 for SSRD variable are closely related. This relationship suggests that these locations experience similar climatic conditions, which can be leveraged in the model for improved accuracy by incorporating spatial dependencies.

For consistency and based on the cross-correlation findings, location 5 was used for both input and output. This consistency helps in reducing complexity and improving model stability, ensuring that the predictions are reliable and based on closely related data.

The histogram of the ratio of `SSRD_5` (Surface Solar Radiation Downwards at location 5) and `SSRDC_5` (Surface Solar Radiation Downwards Clear Sky at location 5) shown in Figure 5 provides insights into cloud cover variability. High variability in this ratio indicates significant cloud cover, which affects solar radiation and PV output. The histogram helps in understanding the distribution and frequency of different cloud cover conditions over the studied period.

The analysis was conducted using different window sizes (Window 2, 3, 4, 5, 6):

1) Window 2: Considered a very short-term prediction, capturing immediate past data.
2) Window 3, 4, 5, 6: Incrementally larger windows capturing longer past data sequences, helping in understanding how far back the model needs to look to make accurate predictions. Figure 6 visualises the analysis of the MSE vs Window lengths across the entire year, dry and wet seasons.
3) SSRD_5 and FDIR_5: Predictions for these features was performed. The results are shown in Figure 6.

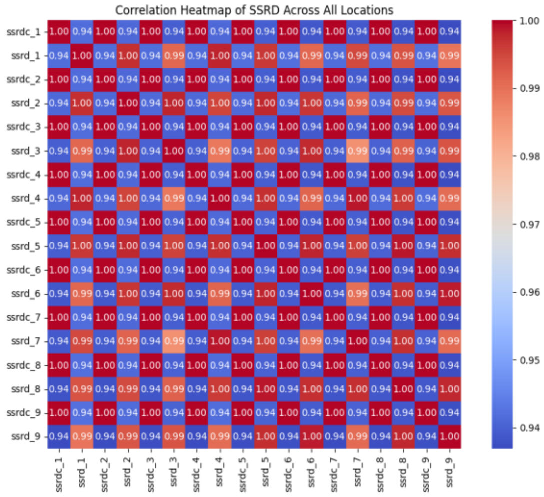

Figure 4: Correlation heatmap of SSRD across all locations.

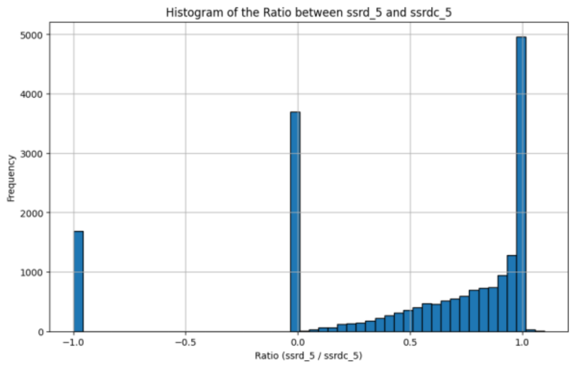

Figure 5: Ratio between ssrd_5 and ssrdc_5.

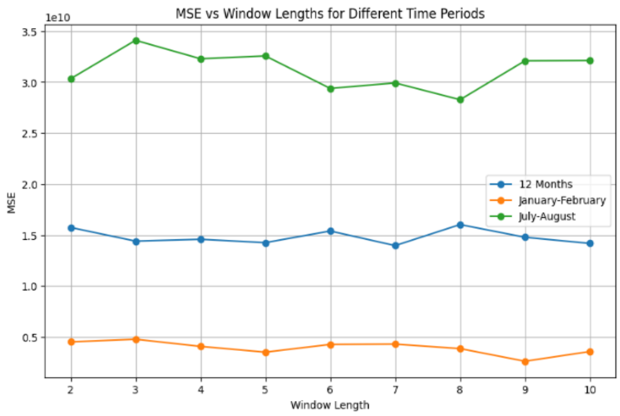

Figure 6: MSE vs window lengths for different time periods for year 2022.

4.3 Results Interpretation

The analysis demonstrates that LSTM models are highly effective in predicting PV power output in Abuja, leveraging weather data to account for both temporal and spatial factors. The model exhibits superior performance during stable weather seasons, such as the dry and rainy periods, which are characterized by consistent solar radiation patterns. This consistency allows the model to capture predictable dynamics, whereas transitional seasons pose greater challenges due to increased variability in weather conditions.

The 24-hour cycle identified through the autocorrelation study serves as a key insight for designing the model, providing a logical foundation for determining optimal window sizes. This alignment enables the model to effectively account for diurnal variations, improving accuracy and reliability. Moreover, the analysis highlights that incorporating data from multiple locations, especially those with strong spatial interdependencies like locations 5 and 8, enhances the model's ability to generalize across a wider range of conditions.

Figures 7, 8, and 9 provide evidence of how temporal and spatial configurations influence model performance. Window sizes tailored to the diurnal cycle result in reduced errors, as seen through lower MAE and RMSE metrics. Including features such as direct irradiance (FDIR_5) alongside solar radiation (SSRD_5) across all locations further boosts model accuracy, illustrating the value of combining complementary data sources.

The R^2 metric underscores the importance of spatial diversity in the input data, revealing that models trained with features from multiple locations achieve higher predictive accuracy and explanatory power. This reinforces the idea that broader spatial data not only improves generalization but also stabilizes performance across varying temporal scales.

While smaller window sizes fail to capture long-term dependencies and larger windows risk incorporating irrelevant information, an intermediate window size aligned with the 24-hour cycle strikes a balance between model complexity and accuracy. The non-linear relationship between window size and performance metrics, as observed in Figures 7 and 8, reflects this trade-off, emphasizing the need for careful tuning of input parameters.

Additionally, models leveraging multi-location data demonstrate greater resilience across changing window sizes, making them more robust for long-term forecasting. By incorporating direct measures of sunlight intensity, such as FDIR_5, the models further reduce errors and improve their explanatory power, highlighting the importance of precise feature selection [18].

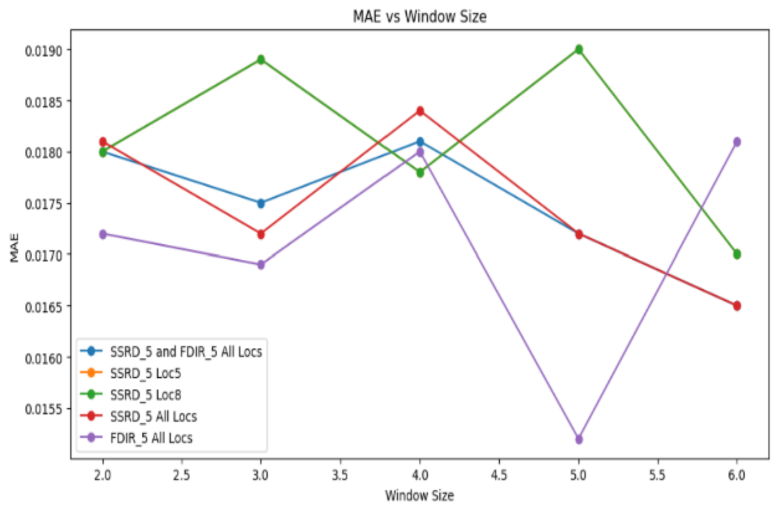

Figure 7: Performance metrics vs window sizes.

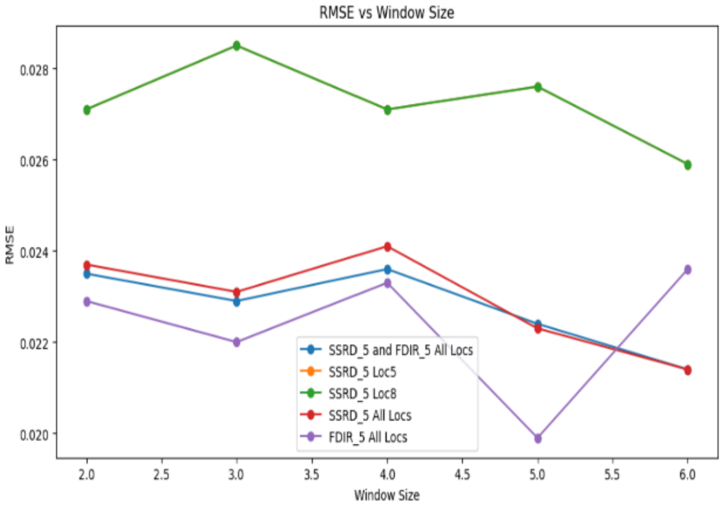

Figure 8: Performance metrics vs window sizes.

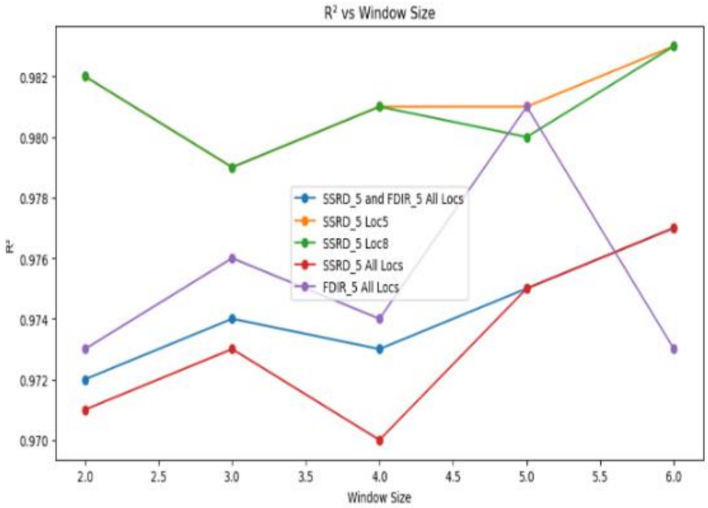

Figure 9: Performance metrics vs window sizes.

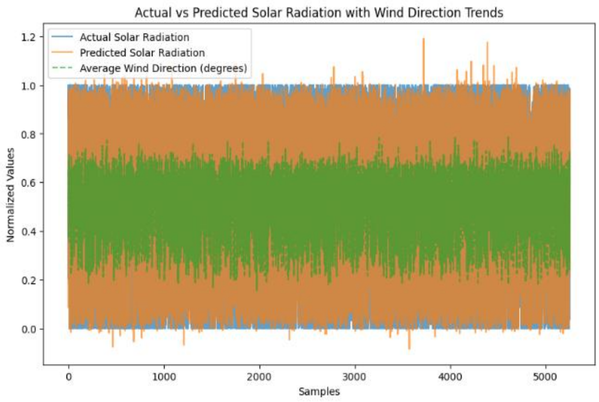

Figure 10: Actual vs Predicted Solar Radiation with Wind Direction Trends.

Figure 10 illustrates the relationship between actual solar radiation, predicted solar radiation, and average wind direction trends across the test dataset samples. The variations in solar radiation and wind direction trends reflect the dynamic behaviour of meteorological data. The green wind direction trendline indicates fluctuations in atmospheric conditions, which may have a subtle influence on solar radiation that the model captures. The predicted solar radiation values closely align with the actual values, showcasing the LSTM model's ability to effectively utilize temporal and meteorological features, including wind direction. Although wind direction does not display a direct cyclic pattern with solar radiation in this plot, its inclusion likely enhances the model's performance by accounting for atmospheric effects such as cloud movement or aerosol dispersion, which impact solar radiation.

Our findings reveal that a hybrid approach combining deep learning techniques with statistically correlated features outperforms a hybrid model that relies solely on machine-learning techniques without leveraging the statistical relationships between features, as demonstrated in the recent results reported by Alkandari and Ahmad [18].

Seasonal variability is another important factor, as spatially diverse configurations are better equipped to handle the uncertainties of transitional periods. These findings highlight that the integration of temporal cycles, spatial interdependencies, and diverse weather features is key to building a reliable and accurate LSTM model for solar energy forecasting in Abuja. Figures 7, 8, and 9 validate this approach, offering practical guidance for optimizing predictive performance through feature selection and model design.

4.3 Implications for PV Power Forecasting in Abuja and Similar Urban Settings

The use of LSTM models, for predicting PV power in Abuja and similar urban areas has shown promise for application in cities with comparable weather conditions. Precise forecasts of PV power generation can enhance the efficiency and reliability of energy systems leading to seamless energy management and grid stability. The findings from this research, which highlights the significance of choosing relevant features and optimal time windows can be utilized in other locations to accelerate the integration of renewable energy sources into current power systems. This in turn aligns with the overarching objectives of reducing dependence on fossil fuels and promoting sustainable development. Our study introduces an innovative method for real-time, low-cost solar energy forecasting on resource-constrained edge devices, focusing on the evaluation and optimization of deep learning models to improve energy management for both residential and industrial applications, aligning with findings reported in [19].

5 CONCLUSIONS

This study effectively showcased the use of LSTM models to predict PV power output by utilizing weather features specific to Abuja, Nigeria. Overall, the analyses revealed that LSTM models, with a window size of 6 and a combination of meteorological and solar radiation features resulted in the most accurate predictions. The analyses of autocorrelation and cross correlation played a role in identifying the optimal feature selection and window sizes thereby significantly enhancing the model's predictive accuracy. Through comparing setups, it became clear that certain combinations of features and window sizes produced better performance metrics (MAE, RMSE and R^2). These results emphasize the significance of choosing relevant features and appropriate window lengths to optimize the model configuration for predicting PV power in urban environments. The implications of this study go beyond Abuja as it provides insights for urban areas facing similar climate conditions. Accurate PV power prediction can lead to improved energy management, grid stability and greater adoption of energy sources aligning with sustainable development objectives.

Future research could explore real-time data feeds, larger datasets, and additional climatic features to further improve the model's accuracy and applicability in different regions.

ACKNOWLEDGEMENTS

This research paper would not have been possible without the support of various individuals and organizations. We gratefully acknowledge the Nigerian Petroleum Technology Development Fund (PTDF) for its immense support in bringing this work to fruition.

REFERENCES

[1] C. Etukudor, B. Couraud, V. Robu, W. Fruh, D. Flynn, and C. Okereke, “Automated negotiation for peer-to-peer electricity trading in local energy markets,” Energies, vol. 13, no. 4, 2020. [Online]. Available: https://doi.org/10.3390/en13040920.

[2] S. M. Kadiri, A. O. Bagre, M. B. Camara, B. Dakyo, and Y. Coulibaly, “Electrical power distribution status in west assessment and perspective overview,” in Proc. IEEE Int. Conf. Renewable Energy Res. Appl. (ICRERA), 2019. [Online]. Available: https://doi.org/10.1109/ICRERA47325.2019.8997112.

[3] W. Tsai, C. Tu, C. Hong, and W. Lin, “A review of state-of-the-art and short-term forecasting models for solar PV power generation,” Energies, vol. 16, no. 14, 2023. [Online]. Available: https://doi.org/10.3390/en16145436.

[4] K. Tu, H. Xiong, and F. Chen, “Comprehensive prediction of generation power in photovoltaic microgrid,” in Proc. IEEE Joint Int. Inf. Technol. Artif.

Intell. Conf., 2019. [Online]. Available: https://doi.org/10.1109/ITAIC.2019.8785746.

[5] O. T. Ibitoye, O. S. Agunbiade, T. W. Ilemobola, A. B. Oluwadare, P. C. Ofodu, K. O. Lawal, and J. O. Dada, “Nigeria electricity grid and the potentials of renewable energy integration: A concise review,” in Proc. IEEE 7th Int. Energy Conf. (ENERGYCON), 2022, pp. 1-4. [Online]. Available: https://doi.org/10.1109/ENERGYCON53164.2022.9830349.

[6] Y. N. Chanchangi, F. Adu, A. Ghosh, S. Sundaram, and T. K. Mallick, “Nigeria’s energy review: Focusing on solar energy potential and penetration,” Environ. Dev. Sustain., vol. 25, no. 7, pp. 5755-5796, 2023. [Online]. Available: https://doi.org/10.1007/s10668-022-02308-4.

[7] C. K. Rao, S. Sahoo, and F. F. Yanine, “Forecasting electric power generation in a photovoltaic power system for smart energy management,” in Proc. Int. Conf. Intell. Controller Comput. Smart Power (ICICCSP), 2022. [Online]. Available: https://doi.org/10.1109/ICICCSP53532.2022.9862396.

[8] M. N. Akhter et al., “An hour-ahead PV power forecasting method based on an RNN-LSTM model for three different PV plants,” Energies, vol. 15, p. 2243, 2022. [Online]. Available: https://doi.org/10.3390/EN15062243.

[9] Y. Wang, Y. Shen, S. Mao, X. Chen, and H. Zou, “LASSO and LSTM integrated temporal model for short-term solar intensity forecasting,” IEEE Internet Things J., vol. 6, pp. 2933-2944, 2019. [Online]. Available: https://doi.org/10.1109/JIOT.2018.2877510.

[10] F. Jin, “Photovoltaic power generation prediction based on k-means clustering analysis and GRO-CNN-LSTM attention,” Math. Model. Algorithm Appl., vol. 2, no. 1, pp. 61-68, 2024. [Online]. Available: https://doi.org/10.54097/qdcrpc46.

[11] R. Kabir, A. Elmouatamid, H. Elkhoukhi, and P. W. Pong, “Photovoltaic power forecasting using neural networks for short and medium-term dependencies,” in Proc. IEEE Texas Power Energy Conf. (TPEC), 2024, pp. 1-6. [Online]. Available: https://doi.org/10.1109/TPEC60005.2024.10472207.

[12] T. Magesh et al., “Prediction of solar power irradiance using machine learning and deep learning for smart grid integration,” in Proc. Int. Conf. Recent Adv. Electr. Electron. Ubiquitous Commun. Comput. Intell. (RAEEUCCI), 2024, pp. 1-6. [Online]. Available: https://doi.org/10.1109/RAEEUCCI61380.2024.10547750.

[13] C. K. Rodriguez-Leguizamon et al., “PV power generation forecasting based on XGBoost and LSTM models,” in Proc. IEEE Workshop Power Electron. Power Qual. Appl. (PEPQA), 2023, pp. 1-6. [Online]. Available: https://doi.org/10.1109/PEPQA59611.2023.10325757.

[14] H. He, R. Hu, Y. Zhang, Y. Zhang, and R. Jiao, “A power forecasting approach for PV plant based on irradiance index and LSTM,” in Proc. 37th Chin. Control Conf. (CCC), 2018, pp. 9404-9409. [Online]. Available: https://doi.org/10.23919/ChiCC.2018.8483960.

[15] T. Limouni et al., “Univariate and multivariate LSTM models for one step and multistep PV power forecasting,” Int. J. Renew. Energy Dev., vol. 11, no. 3, pp. 815-824, 2022. [Online]. Available: https://doi.org/10.14710/ijred.2022.43953.

[16] R. Ahmed et al., “Computationally expedient photovoltaic power forecasting: A LSTM ensemble method augmented with adaptive weighting and data segmentation technique,” Energy Convers. Manage., vol. 258, p. 115563, 2022. [Online]. Available: https://doi.org/10.1016/j.enconman.2022.115563.

[17] H. Hersbach et al., “ERA5 hourly data on single levels from 1940 to present,” Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2023. [Online]. Available: https://doi.org/10.24381/cds.adbb2d47.

[18] M. AlKandari and I. Ahmad, “Solar power generation forecasting using ensemble approach based on deep learning and statistical methods,” Appl. Comput. Informat., vol. 20, no. 3/4, pp. 231-250, 2024. [Online]. Available: https://doi.org/10.1016/j.aci.2019.11.002.

[19] A. M. Hayajneh, F. Alasali, A. Salama, and W. Holderbaum, “Intelligent solar forecasts: Modern machine learning models and TinyML role for improved solar energy yield predictions,” IEEE Access, vol. 12, pp. 10846-10864, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3354703.