FMDB Transactions on Sustainable Computer Letters

Advanced Time-Series Forecasting of Solar Energy Output Using Deep Learning for Improved Renewable Integration

S. Rubin Bose[1], J. Angelin Jeba[2,*], N. Christy Evangeline[3], R. Regin[4], J. Rahila[5], Madhan Raj Gopi Akila[6]

[1]Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[2]Department of Electronics and Communication Engineering, S.A. Engineering College, Chennai, Tamil Nadu, India.
[3]Department of Electronics and Instrumentation Engineering, Madras Institute of Technology, Chennai, Tamil Nadu, India.
[4]School of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[5]Department of Electrical and Electronics Engineering, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
[6]Department of Computer Science, University of Stuttgart, Stuttgart, Baden-Württemberg, Germany.
rubinbos@srmist.edu.in[1], angelinjeba@saec.ac.in[2], christy.evangeline@gmail.com[3], reginr@srmist.edu.in[4], rahila@dhaanishchennai.in[5], st199709@stud.uni-stuttgart.de[6]

Abstract: The increasing integration of photovoltaic (PV) systems into electrical grids presents significant challenges to grid stability due to the inherent volatility and intermittency of solar power generation. Accurate short-term power forecasting is therefore critical for effective grid management, optimal resource scheduling, and maintaining a reliable balance between energy supply and demand. This paper presents a deep learning methodology for forecasting plant-level solar power output at a 15-minute temporal resolution. A Long Short-Term Memory (LSTM) network, a specialised type of recurrent neural network, is developed to predict DC power output from a solar power plant. The model leverages a unified dataset comprising historical power generation data from 22 inverters and co-located meteorological sensor readings, including solar irradiation and ambient and module temperatures. A rigorous data pretreatment pipeline uses advanced time-series feature engineering techniques, such as lag features, rolling-window statistics, and sine-cosine transforms for cyclical temporal features. A recursive forecasting technique trains a stacked LSTM architecture to predict power output. The model's performance is assessed using regression measures such as MAE, RMSE, and R^2$ on an unseen test dataset. The model's high accuracy in capturing the complex, non-linear dynamics of solar power generation shows that LSTM-based approaches can improve the operational efficiency and reliability of modern power grids with high renewable energy penetration.

Keywords: Photovoltaic Power Forecasting; Long Short-Term Memory (LSTM); Time-Series Analysis (TSA); Renewable Energy Integration; Grid Stability; Machine Learning (ML).

Received on: 02/04/2025, **Revised on:** 21/06/2025, **Accepted on:** 28/07/2025, **Published on:** 03/01/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSCL

DOI: https://doi.org/10.69888/FTSCL.2026.000600

Cite as: S. R. Bose, J. A. Jeba, N. C. Evangeline, R. Regin, J. Rahila, and M. R. G. Akila, "Advanced Time-Series Forecasting of Solar Energy Output Using Deep Learning for Improved Renewable Integration," *FMDB Transactions on Sustainable Computer Letters*, vol. 4, no. 1, pp. 55–64, 2026.

*Corresponding author.

1. Introduction

In recent years, the global shift toward renewable energy has accelerated, driven by the need to meet rising energy demand and mitigate climate change. Among renewable sources, solar photovoltaic (PV) energy has emerged as a leading technology due to its scalability, decreasing costs, and environmental benefits. The global installed capacity of solar PV has grown rapidly, highlighting its critical role in the future energy mix. However, integrating solar PV into electrical grids presents challenges due to its intermittent nature. Power generation depends on weather conditions such as solar irradiance, cloud cover, and temperature, making it difficult to maintain a stable balance between supply and demand. Sudden fluctuations in solar output can affect grid stability and require grid operators to adopt flexible management strategies. Accurate short-term solar power forecasting is therefore essential. Forecasts enable efficient scheduling of backup power, optimal use of energy storage, and proactive management of supply fluctuations. This helps transform solar power from an unpredictable resource into a more manageable grid component. Traditional physical and statistical forecasting methods often struggle with the non-linear and dynamic patterns of solar generation. Machine learning and deep learning models, particularly Long Short-Term Memory (LSTM) networks, can capture temporal dependencies and improve prediction accuracy. This research aims to develop a deep learning framework using LSTM networks for short-term forecasting of solar power output. The proposed model integrates historical power generation and meteorological data to provide accurate predictions, supporting reliable and efficient operation of solar-integrated grids.

2. Literature Review

Belmahdi et al. [1] introduced a hybrid ARIMA–ANN method to forecast daily global solar radiation (DGSR) across three Moroccan cities. By combining statistical time-series techniques with neural networks, the hybrid model achieved superior accuracy compared to standalone ARIMA and ANN models, as measured by RMSE, MAPE, and R^2. The study demonstrated the effectiveness of hybridisation in handling complex meteorological patterns and its applicability for photovoltaic power forecasting. Elsaraiti and Merabet [2] proposed a deep learning framework for short-term solar power forecasting using Long Short-Term Memory (LSTM) networks. Performance evaluation against Multi-layer Perceptron (MLP) models, using MAE, RMSE, MAPE, and R^2, revealed that LSTM consistently outperformed MLP across different day categories. The approach highlighted the potential of deep learning to support efficient PV plant operation and advance sustainable energy management. Voyant et al. [3] conducted a comprehensive review of solar forecasting methods, comparing physical models, statistical approaches, and hybrid machine learning solutions. They highlighted that combining meteorological data with advanced time-series models yields more accurate short-term forecasts. Paletta and Lasenby [4] applied Convolutional Neural Networks (CNNs) to hemispherical sky images for short-term solar irradiance forecasting. Using 8 months of irradiance and image data from Palaiseau, France, at 2-minute resolution, the CNN models achieved a 10% improvement over smart-persistence baselines for 10-minute-ahead predictions.

Their findings emphasised the importance of same-day historical data and visual feature extraction in improving forecast skill. Ledmaoui et al. [5] conducted a comparative study of six machine learning algorithms—SVR, ANN, Decision Tree, Random Forest, GAM, and XGBoost—for solar energy production forecasting in Morocco. Evaluated with daily data from a solar plant in Benguerir, the ANN model achieved the lowest RMSE and MAE, outperforming other methods. The results demonstrated ANN's suitability for accurate solar forecasting and its potential for integration into predictive maintenance systems. Lauret et al. [6] applied probabilistic forecasting methods to solar power prediction, generating confidence intervals around predictions to support grid operation decisions. They demonstrated that probabilistic forecasts are more useful for risk-aware energy management than deterministic ones. Si et al. [7] proposed a novel photovoltaic power forecasting method using satellite images with cloud movement modelling and solar position adjustments. By developing nonlinear cloud-forecasting and sequential region-selection algorithms, they overcame limitations imposed by the low satellite update frequency. Coupled with XGBoost for power prediction, the method significantly improved accuracy compared to traditional benchmarks, validating the effectiveness of cloud-based feature integration. Marquez and Coimbra [8] developed a clear-sky index model to normalise irradiance data, enabling improved generalisation of forecasting models across different geographical locations.

This method was particularly effective in reducing site-specific biases. Baloch et al. [9] developed a solar energy forecasting framework using the Facebook Prophet model for Muscat, Oman. By incorporating features such as GHI, DNI, and GNI under both cloudy and clear-sky conditions, the model outperformed other machine learning approaches, including ARIMA, SVR, LSTM, and Random Forest. Prophet achieved the lowest MAE (9.876) and RMSE (18.762) with an R^2 of 0.991, demonstrating its robustness for both clean and cloudy weather scenarios. Eseye et al. [10] proposed a hybrid Wavelet–PSO–SVM model for short-term PV power forecasting. Using SCADA data from a microgrid PV system and meteorological inputs, wavelet transforms improved signal preprocessing, while PSO optimised the SVM parameters. Compared with seven alternative strategies, the hybrid model delivered superior forecasting accuracy, demonstrating the benefits of integrating signal decomposition, optimisation, and nonlinear regression for real-world PV systems. Sobri et al. [11] presented a comprehensive review of solar photovoltaic (PV) generation forecasting methods to address the challenges of uncertainty and power fluctuations in PV integration. The study analysed and compared three main categories of forecasting techniques—time-series statistical methods, physical models, and ensemble approaches. They also highlighted that ensemble techniques, which combine multiple models, further improve forecasting accuracy. The paper concluded that these advanced forecasting strategies are essential for efficient grid operation and optimal planning of solar photovoltaic systems.

Antonanzas et al. [12] conducted an extensive review of photovoltaic power forecasting techniques, addressing the growing challenges in grid management caused by solar resource variability. The study examined forecasting across multiple temporal horizons—from seconds to weeks ahead—and spatial scales, including both site-specific and regional forecasts. The authors analysed various deterministic and probabilistic forecasting approaches, emphasising the increasing relevance of probabilistic models for reducing prediction uncertainty and assessing their economic impact on grid operations. The paper also provided a detailed classification of forecasting studies by input sources and forecast horizons, along with a comprehensive summary of the performance metrics used across the literature. Overall, the review highlighted emerging trends and research directions to enhance forecast accuracy and grid reliability. Sammar et al. [13] presented a comprehensive review of AI-based solar irradiance prediction models to improve the reliability and efficiency of solar energy systems. The study discussed how artificial intelligence and machine learning techniques—such as regression models, neural networks, and ensemble methods—are increasingly utilised to capture complex patterns from historical weather data for accurate irradiance forecasting. The authors compared machine learning, numerical weather prediction, and hybrid approaches in terms of their accuracy, advantages, and limitations. The paper also emphasised the need for continued interdisciplinary collaboration and technological innovation to integrate solar energy into modern power systems. Overall, the review highlighted the potential of AI-driven models to enable more sustainable and resilient solar energy forecasting solutions.

Sahu [14] conducted a detailed study of global solar photovoltaic (PV) energy development and policy frameworks, focusing on the top 10 solar power–producing countries. The paper explored global trends in solar PV growth, per capita energy generation, and investments across major nations. It also analysed government initiatives such as supportive tariff rates, net metering, green certificates, and incentive policies that have driven large-scale adoption of solar energy. Additionally, the study discussed variations in solar cell efficiency achieved by different research laboratories worldwide. The findings revealed that these leading countries are successfully meeting their renewable energy projections through strategic policies and sustained technological advancements in solar PV systems. Diagne et al. [15] provided an in-depth review of solar irradiance forecasting methods with a specific focus on applications for small-scale insular grids. The study examined various forecasting techniques, including statistical models, cloud image–based methods, numerical weather prediction (NWP) models, and hybrid approaches. The authors emphasised the importance of accurate irradiance forecasting to improve grid reliability and optimise the integration of photovoltaic (PV) systems into power networks. Furthermore, the paper highlighted that hybrid and NWP-based methods have strong potential to address the variability of solar output. The review concluded with recommendations for developing advanced forecasting models tailored for efficient energy management in isolated or small-scale grid systems. The studies reviewed present a wide spectrum of approaches to solar power and irradiance forecasting.

Hybrid statistical–neural frameworks such as those by Belmahdi et al. [1] combine ARIMA and neural networks to capture both linear and nonlinear patterns, while Elsaraiti and Merabet [2] demonstrate the effectiveness of LSTM architectures for handling temporal dependencies in power data. Voyant et al. [3] highlight the role of ensemble methods in reducing prediction error by integrating multiple base learners. Paletta and Lasenby [4] show that convolutional neural networks applied to sky images can directly capture cloud dynamics for irradiance forecasting, complementing time-series approaches. Ledmaoui et al. [5] provide comparative insights into the performance of several machine learning techniques, emphasising trade-offs in accuracy and computational cost. Lauret et al. [6] focus on spatial solar irradiance forecasting and demonstrate the importance of accounting for spatial variability in distributed PV networks. Relying on satellite data, Si et al. [7] develop short-term irradiance predictions applicable globally. Marquez and Coimbra [8] contribute by developing probabilistic forecasting methods that explicitly quantify uncertainty in solar predictions. Baloch et al. [9] propose a location-specific forecasting framework for Pakistan that addresses challenges of data availability and climatic variation. Finally, Eseye et al. [10] employ hybrid optimisation techniques to enhance model performance, underlining the importance of parameter tuning and adaptive methods. Collectively, these studies highlight that while deep learning models such as LSTMs and CNNs, and ensemble techniques, significantly improve forecast accuracy, opportunities remain to integrate probabilistic forecasting, spatial methods, and ensemble learning into hybrid frameworks [3]. Future research should focus on real-time deployment, uncertainty quantification, and integration with smart grid systems to ensure reliability.

3. Methodology

The methodology adopted in this research centres on designing and implementing a deep learning–based forecasting system utilising a Long Short-Term Memory (LSTM) network. The primary objective is to accurately predict short-term solar power output from a photovoltaic (PV) plant by leveraging historical generation and meteorological data, including solar irradiance, temperature, humidity, and wind speed. The model development process involves several key stages, including data preprocessing, normalisation, and feature selection to ensure high-quality input for training. The LSTM network is specifically chosen for its capability to capture temporal dependencies and nonlinear relationships within time-series data, enabling it to model the dynamic fluctuations in solar energy output effectively. The proposed system is trained and validated on real-world datasets, and its performance is evaluated using statistical metrics such as MAE, RMSE, and R^2 to assess prediction accuracy. Overall, this methodology aims to enhance the reliability of solar energy forecasting and support optimal grid management and power scheduling in renewable energy systems. Figure 1 illustrates the complete workflow of the proposed solar power forecasting system. It begins with the input of raw solar data in CSV format, followed by data loading and preprocessing, where feature–target separation and an 80/20 train-test split are performed. The processed data is then used to train an LSTM network to directly forecast 96 timesteps. The trained model artifacts (saved as .keras and .gz files) are integrated with a Streamlit-based prediction application.

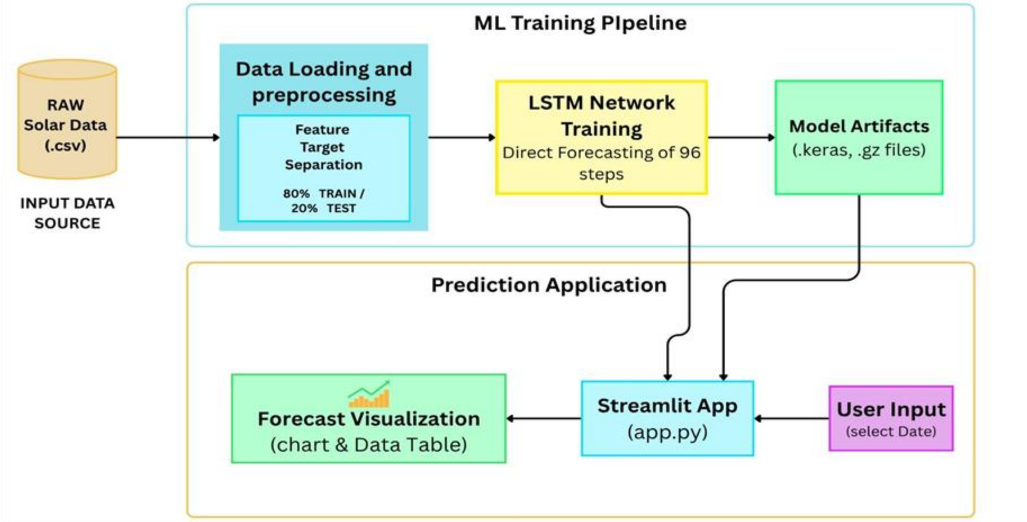

Figure 1: Block diagram

The app enables user interaction through date selection. It displays forecast outputs via visualisation charts and data tables, demonstrating the end-to-end integration of machine learning and user interface components for real-time solar forecasting. The core component of the proposed methodology is the Long Short-Term Memory (LSTM) network, an advanced variant of the Recurrent Neural Network (RNN) architecture specifically designed to learn and preserve long-term dependencies in sequential data. Unlike conventional RNNs, which often suffer from vanishing and exploding gradients, LSTMs incorporate a unique memory cell structure that effectively enables them to retain useful information over long time intervals. Each LSTM cell contains three key gates—forget, input, and output gates—that work together to regulate the flow of information. The forget gate decides which information from previous states to discard, the input gate determines which new information to store, and the output gate controls which information to pass on to the next time step. Due to these characteristics, LSTMs are particularly effective for time-series forecasting tasks such as solar power prediction, where capturing both short-term variations and long-term seasonal trends is essential for accurate and reliable performance.

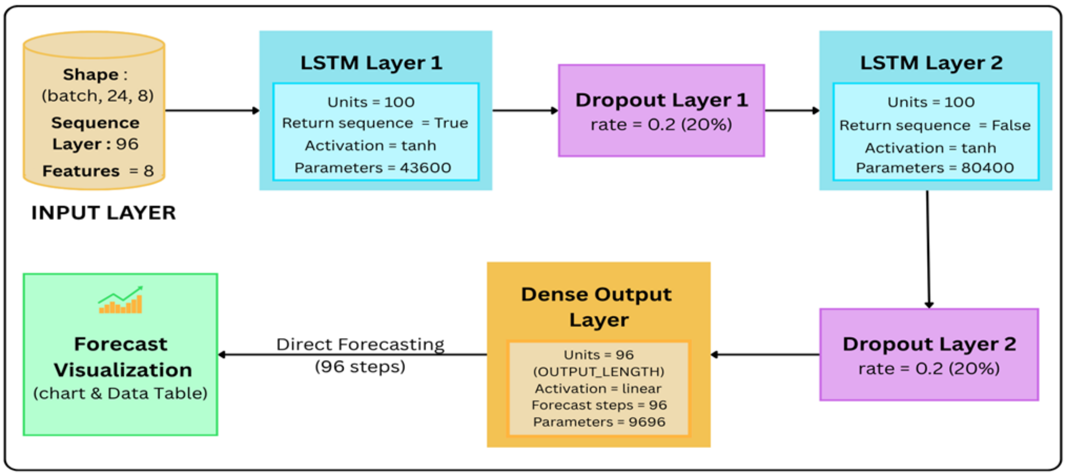

Figure 2: Architecture diagram

Figure 2 illustrates the architecture of the stacked LSTM model used for forecasting. The model processes input sequences with a shape of (batch, 24, 8), representing 8 features over 24 timesteps. It consists of two sequential LSTM layers, each with 100 units and tanh activation, separated by a 20% dropout layer for regularisation. The first LSTM layer returns the full sequence, while the second returns only the final output. After a second 20% dropout, the output is passed to a Dense layer with 96 units and linear activation, which directly forecasts the entire 96-step sequence. The proposed model employs a stacked Long Short-Term Memory (LSTM) architecture, which comprises multiple LSTM layers arranged sequentially to enhance the model's ability to learn complex temporal features from the input data. In this configuration, the output sequence of each LSTM layer serves as the input for the subsequent layer, allowing the network to capture both low-level and high-level temporal dependencies present in solar power time-series data. To improve model robustness and prevent overfitting, dropout layers are inserted between LSTM layers, thereby improving generalisation to unseen data. The forecasting process follows a recursive prediction strategy, wherein the model is initially trained to perform. One-step-ahead forecasting is then applied iteratively to generate multi-step-ahead predictions for future time intervals.

This method strikes an effective balance between computational efficiency and forecasting accuracy, making it particularly suitable for short-term solar power prediction, where rapid, reliable estimates are required for real-time energy management and grid stability. During the training phase, the preprocessed time-series data is transformed into a supervised learning format using a sliding window technique, enabling the model to learn temporal dependencies effectively. In this approach, each sliding window of consecutive historical observations is treated as an input sequence, while the subsequent time step represents the target output. This method allows the LSTM network to recognise sequential relationships between past and future values in the dataset. A sequence length of 96—corresponding to 24 hours of data recorded at 15-minute intervals—is used to predict the next 15-minute solar power output. This configuration ensures that the model captures both daily patterns and short-term fluctuations in power generation influenced by changing weather conditions. By framing the problem in this manner, the LSTM network can efficiently learn the temporal dynamics of solar power variation, leading to more accurate and consistent short-term forecasts. All input features are scaled using a MinMaxScaler to normalise them to the range [0, 1], ensuring stable and balanced model training. After prediction, the outputs are inverse-transformed to their original scale for performance evaluation. This structured methodology enables the LSTM model to effectively learn complex temporal relationships in the data, resulting in accurate and reliable solar power forecasts.

4. Experimental Setup

4.1. Training

The experimental design was structured to ensure a robust and reproducible evaluation of the LSTM model. The unified, feature-engineered dataset was split chronologically into three sets: a training set (70%), a validation set (15%), and a test set (15%). A chronological split is essential for time-series forecasting to prevent data leakage, ensuring that the model is trained only on past data and evaluated on future, unseen data, simulating a real-world deployment scenario. The model was compiled using the Adam optimiser, a widely adopted and effective algorithm for training deep learning models.13 The Mean Squared Error (MSE) was chosen as the loss function, which the optimiser seeks to minimise during training. MSE is a standard choice for regression problems as it strongly penalises larger prediction errors. The training process ran for a fixed number of epochs, each consisting of a full pass over the entire training dataset. The model's performance on the validation set was monitored at the end of each epoch. This validation performance is crucial for diagnosing issues such as overfitting and for selecting optimal hyperparameters.

4.2. Evaluation

To assess the model's forecasting accuracy on the unseen test set, a suite of standard regression metrics was used. Using multiple metrics provides a more nuanced understanding of the model's error characteristics. The following metrics were calculated.

4.2.1. Mean Absolute Error (MAE)

The metric in equation (1) measures the average absolute difference between the predicted values ($\hat{y}_i$) and the actual values (y_i). It is easily interpretable as the average magnitude of the forecast error in the original units of the data (kW). The equation (1) is defined as:

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

Where n represents the total number of observations, y_i is the actual value for the i^{th} observation, and $\hat{y}_i$ is the corresponding predicted value. The absolute difference $| y_i - \hat{y}_i |$ measures the magnitude of the prediction error for each observation, and the average of these differences gives an overall measure of prediction accuracy. A lower MAE value indicates better model performance, as it signifies that the predicted values are closer to the actual values on average.

4.2.2. Root Mean Squared Error (RMSE)

The metric in equation (2) is the square root of the average of the squared differences between predicted and actual values. By squaring the errors, RMSE gives a relatively high weight to large errors, making it particularly sensitive to outliers and significant forecast misses. The equation (2) is defined as:

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

Where n is the total number of observations, y_i represents the actual or observed value for the i^{th} data point, and $\hat{y}_i$ is the corresponding predicted value. The squared term $(y_i - \hat{y}_i)^2$ emphasises larger errors by giving them more weight, and taking the square root brings the result back to the same unit as the original data. A smaller RMSE value indicates higher prediction accuracy and better model performance.

4.2.3. Coefficient of Determination (R^2)

The metric in equation (3) represents the proportion of the variance in the dependent variable that is predictable from the independent variables. An R^2 value close to 1 indicates that the model explains a large portion of the variability in the target variable, signifying a good fit. The equation (3) is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

Where n is the total number of observations, y_i represents the actual or observed value for the i^{th} data point, $\hat{y}_i$ is the corresponding predicted value, and $\bar{y}$ is the mean of all actual values. The numerator $\sum(y_i - \hat{y}_i)^2$ represents the residual sum of squares (RSS), which measures the variance not explained by the model, while the denominator $\sum(y_i - \bar{y})^2$ represents the total sum of squares (TSS), which measures the total variance in the data. An R^2 A value closer to 1 indicates that the model explains most of the variability in the dependent variable, while a value near 0 indicates poor explanatory power.

4.2.4. Data Preprocessing

Min-Max Scaling. Before training, every feature in our dataset is normalised to a range using the Min-Max Scaling technique. The equation (4) is defined as:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4}$$

Where X represents the original value of a feature, X_{min} is the minimum value of that feature in the dataset, and X_{max} is the maximum value. The transformation subtracts the minimum value and divides by the range $(X_{max} - X_{min})$, ensuring that the smallest value becomes 0 and the largest becomes 1. This scaling method is useful for algorithms sensitive to data magnitudes, such as gradient descent-based models and distance-based algorithms, such as k-Nearest Neighbours (k-NN) and Support Vector Machines (SVM). Core Model: Long Short-Term Memory (LSTM).

4.3. Implementation

The entire paper was implemented in Python (version 3.8). The core deep learning model was built, trained, and evaluated using the TensorFlow framework (version 2.x) with its high-level Keras API. Data manipulation, preprocessing, and feature engineering were primarily handled using the Pandas and NumPy libraries. Data scaling was performed using the MinMaxScaler from the Scikit-learn library. All visualisations were generated using Matplotlib and Seaborn.

5. Results and Discussions

The performance of the trained LSTM model was rigorously evaluated on the chronologically held-out test set, which the model had not seen during training or validation. The evaluation metrics provide a quantitative measure of the model's predictive accuracy in a simulated real-world scenario. The results indicate high performance. The R^2 value of 0.98 is particularly noteworthy, indicating that the model explains 98% of the variance in total DC power output. This demonstrates an excellent fit to the data. The MAE of 265.31 kW indicates that, on average, the model's 15-minute ahead forecast deviates from the actual power output by approximately 265 kW. The RMSE of 488.15 kW, being larger than the MAE, reflects the model's occasional larger errors, which are penalised more heavily by the RMSE calculation. The 'Actual vs. Predicted' plot in Figure 3 for the LSTM model on a 24-hour training sample shows a strong correlation between the actual power (solid blue line) and the predicted power (dashed red line).

The model's prediction successfully captures the primary diurnal shape of solar generation, accurately identifying the zero-power night-time periods (timesteps 0-20 and 65-96) and the core ramp-up and ramp-down phases. The predicted line serves as a smoothed version of the actual data, indicating that the model has effectively learned the underlying solar curve. While it filters out some of the high-frequency volatility (the rapid "spikes" in the blue line), this strong fit confirms the LSTM network's capacity to learn the complex, non-linear patterns of power generation from the training dataset. The model's primary strength lies in its ability to learn and internalise the complex temporal dependencies of solar generation, from the fundamental 24-hour cycle encoded in cyclical features to the autoregressive nature captured by lag features.

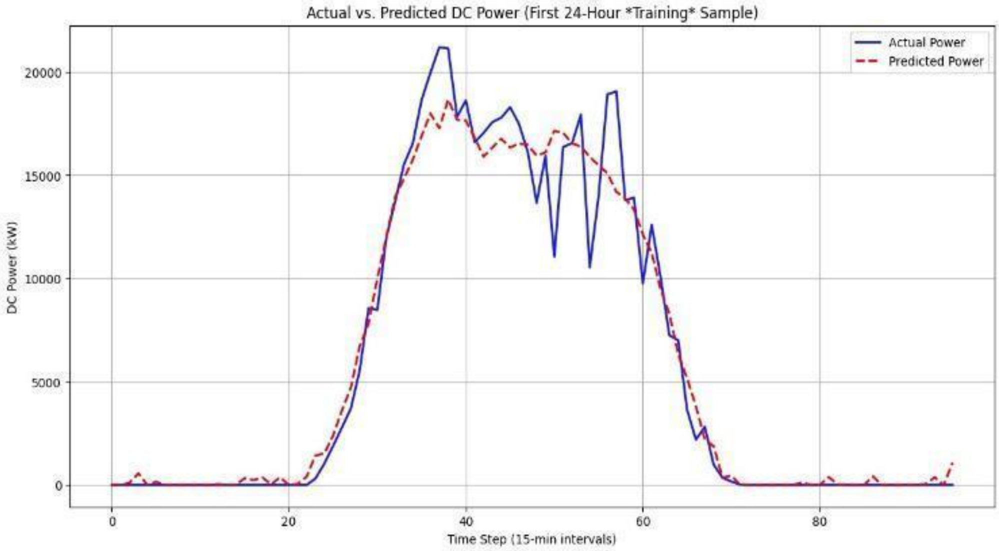

Figure 3: Actual vs. Predicted DC power

The model's inputs (past power and weather data) provide information about the immediate past. Still, they cannot perfectly predict a sudden, future change in cloud cover that has no precedent in the lookback window. This underscores the fundamental challenge in solar forecasting: distinguishing between predictable patterns and purely stochastic weather events.

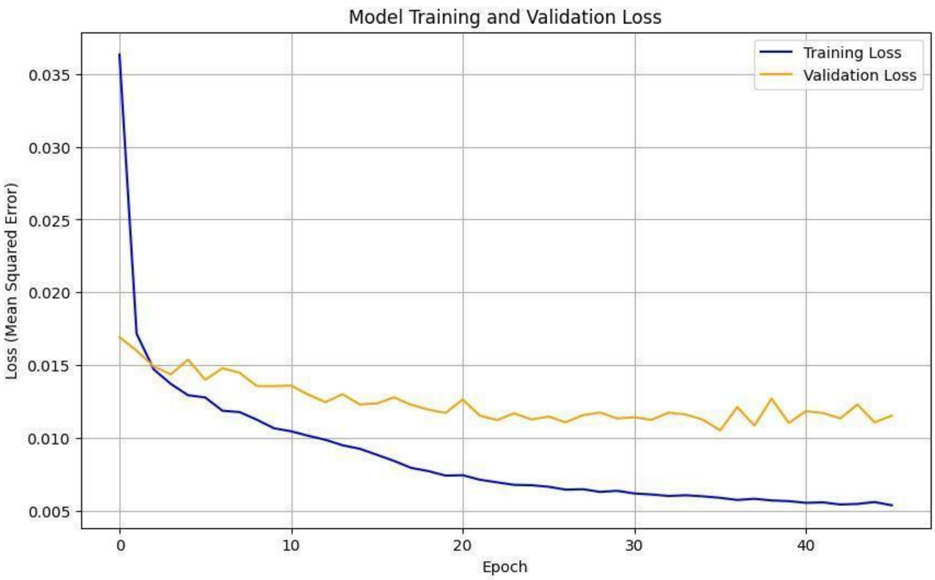

Figure 4: Model training and validation loss

Figure 4 illustrates the model's training and validation loss curves across 45 epochs. The training loss (dark blue line) shows a sharp, consistent decline, indicating the model's strong capacity to learn and fit the training data. The validation loss (orange line), however, follows a different trajectory: it decreases in parallel with the training loss for approximately the first 10-15 epochs, after which it stagnates and begins to fluctuate, failing to improve further. This pronounced divergence between the two curves is a clear sign of overfitting, where the model has begun to memorise noise and specific patterns in the training data rather than learning generalizable patterns. Furthermore, while efficient, a recursive forecasting strategy introduces the potential for error accumulation. While the 15-minute-ahead forecast is highly accurate, any small error in this prediction is fed back into the model as input for the next step. Over a longer forecast horizon (e.g., several hours), these small errors can compound, leading to a degradation in accuracy. Compared to the baseline models of Linear Regression, Random Forest, and Decision

Trees from the initial project, the LSTM model represents a significant methodological advancement. The LSTM's ability to maintain a cell state and use gates to regulate information flow is precisely what enables it to excel at this type of forecasting task, leading to more accurate and reliable predictions of the power generation curve.

5.1. Output

Visual analysis of forecasts generated by the Long Short-Term Memory (LSTM) model provides valuable qualitative insights into its predictive capabilities across varying environmental conditions. While quantitative metrics summarise overall performance, visual inspection offers a clearer understanding of how well the model captures real-world solar generation dynamics.

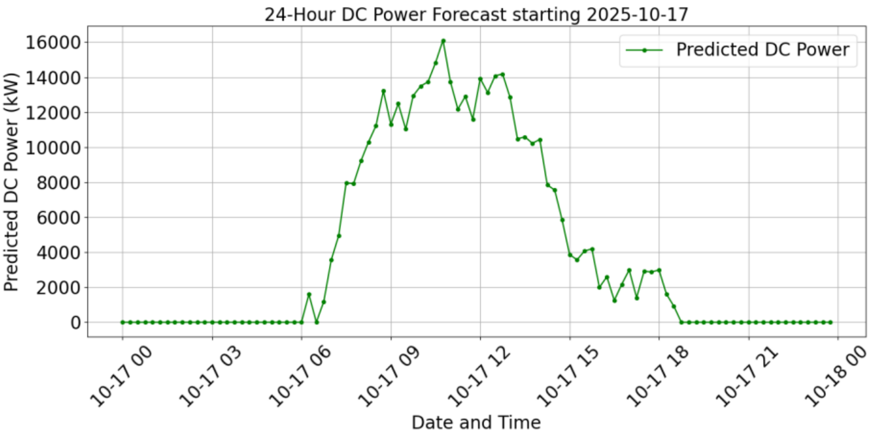

Figure 5: 24-Hour DC power forecast

Figure 5 shows the final 24-hour forecast for October 17, 2025, as generated by the predictive workflow. The model successfully applies learned temporal patterns, forecasting zero power at night and a characteristic solar curve with a peak output near 16,000 kW.

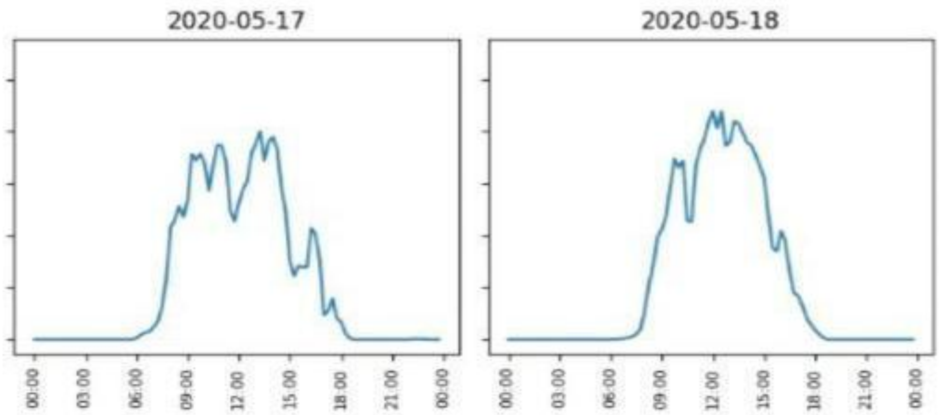

Figure 6: Daily irradiation patterns

A comparison plot of predicted and actual power values across the test period shows that the LSTM model effectively tracks the daily solar generation pattern. The predicted curve closely aligns with the actual power output, accurately capturing the diurnal rise and fall corresponding to the sun's movement. Figure 6 presents the daily irradiation patterns for two consecutive sample days, 2020-05-17 and 2020-05-18. While both plots show the expected diurnal cycle—rising after 06:00 and tapering to zero by 18:00—they also highlight significant high-frequency volatility. The pattern for 2020-05-17 is highly erratic, with numerous sharp peaks and troughs characteristic of intermittent cloud cover. In contrast, 2020-05-18 follows a smoother, more parabolic curve, despite a clear dip around 11:00. This inherent variability in the primary energy input feature (irradiation) demonstrates why simple statistical models are insufficient and justifies the use of an LSTM network capable of learning these complex, non-linear patterns. Further analysis of selected case studies highlights the model's adaptability. On clear, sunny days, the predictions exhibit remarkable accuracy, almost perfectly replicating the smooth parabolic power curve characteristic of stable solar conditions. This demonstrates the model's strong ability to learn the fundamental diurnal cycle when not influenced by weather irregularities. On cloudy days, where rapid fluctuations in power output occur, the model still performs well in identifying overall trends. It successfully predicts periods of reduced output and aligns with the timing of sharpest drops and recoveries. However, the model slightly smooths extreme short-term variations because it cannot precisely predict stochastic events, such as small, fast-moving clouds. Despite this, it effectively infers future volatility from recent data trends, demonstrating its robustness in dynamic weather conditions.

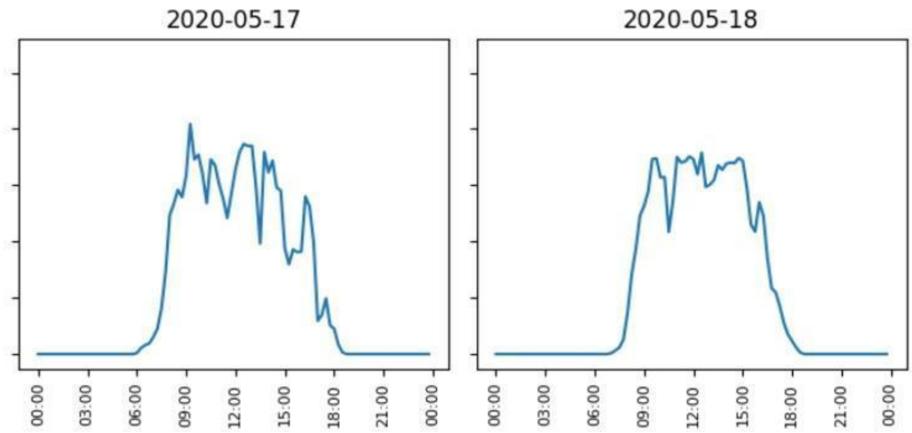

Figure 7: Daily DC power generation

Figure 7 displays the resulting DC power generation for the same two sample days, 2020-05-17 and 2020-05-18. The plots show that the power output (the target variable) is a direct, non-linear reflection of the irradiation input. The 2020-05-17 generation profile is highly erratic, characterised by constant, sharp fluctuations corresponding to the intermittent irradiation. In contrast, 2020-05-18 exhibits a smoother, more defined generation curve that also mirrors its irradiation pattern, including the noticeable dip around 11:00. This direct and complex correlation confirms that the LSTM's primary task is to accurately model the translation of the highly variable irradiation input into the final power output. Overall, the visual analysis confirms that the LSTM model not only delivers strong quantitative accuracy but also demonstrates high qualitative reliability in adapting to diverse solar and meteorological conditions.

6. Conclusion

The study on short-term solar power forecasting successfully addressed a major challenge in renewable energy prediction by developing a deep learning model based on a stacked Long Short-Term Memory (LSTM) architecture. Through a systematic workflow encompassing data cleaning, unification, and feature engineering, the research introduced lag and rolling window features to capture temporal dependencies and used sine-cosine encoding to represent the cyclical nature of time-based variables. The proposed model demonstrated remarkable predictive capability, achieving a Coefficient of Determination (R^2) of 0.98 on unseen test data. Furthermore, visual analyses confirmed the model's ability to accurately track daily generation curves across varying weather conditions, underscoring its potential to improve grid stability and operational efficiency amid solar intermittency. Future advancements in this field could focus on refining model architecture and enriching data sources. The integration of hybrid CNN-LSTM frameworks may enable automatic extraction of complex temporal-spatial patterns, while the inclusion of Attention mechanisms could enhance adaptability under highly variable conditions. A comparative analysis of recursive and direct forecasting strategies could offer deeper insights into the trade-offs between efficiency and robustness. Moreover, incorporating external datasets, such as Numerical Weather Prediction (NWP) outputs and satellite imagery, could significantly enhance predictive reliability. Finally, transitioning from point forecasts to probabilistic forecasting would enable uncertainty quantification, empowering grid operators with more informed, risk-aware decision-making for energy management and resource allocation.

Acknowledgment: The authors express their sincere gratitude to SRM Institute of Science and Technology at Ramapuram, S. A. Engineering College, Madras Institute of Technology, Dhaanish Ahmed College of Engineering, and University of Stuttgart for their invaluable support and academic guidance. Their contributions and resources have significantly facilitated the successful completion of this work.

Data Availability Statement: Data relevant to this study will be provided by the corresponding author upon reasonable request, where permissible.

Funding Statement: This study was conducted without any external financial support.

Conflicts of Interest Statement: The authors report no conflicts of interest, and all sources used in this original work have been duly acknowledged.

Ethics and Consent Statement: The study adhered to ethical guidelines, and informed consent was obtained from all participants.

References

1. B. Belmahdi, M. Louzazni, and A. E. Bouardi, "A hybrid ARIMA–ANN method to forecast daily global solar radiation in three different cities in Morocco," *European Physical Journal Plus*, vol. 135, no. 11, p. 925, 2020.
2. M. Elsaraiti and A. Merabet, "Solar power forecasting using deep learning techniques," *IEEE Access*, vol. 10, no. 3, pp. 31692–31698, 2022.
3. C. Voyant, G. Notton, S. Kalogirou, M. -L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, no. 5, pp. 569–582, 2017.
4. Q. Paletta and J. Lasenby, "Convolutional neural networks applied to sky images for short-term solar irradiance forecasting," *arXiv preprint*, 2020. [Accessed by 02/02/2025].
5. Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, and A. Chehri, "Forecasting solar energy production: A comparative study of machine learning algorithms," *Energy Reports*, vol. 10, no. 11, pp. 1004–1012, 2023.
6. P. Lauret, M. David, and H. T. C. Pedro, "Probabilistic solar forecasting using quantile regression," *Solar Energy*, vol. 10, no. 10, p. 1591, 2017.
7. Z. Si, M. Yang, Y. Yu, and T. Ding, "Photovoltaic power forecast based on satellite images considering effects of solar position," *Applied Energy*, vol. 302, no. 11, p. 117514, 2021.
8. R. Marquez and C. F. Coimbra, "Proposed metric for evaluation of solar forecasting models," *Journal of Solar Energy Engineering*, vol. 135, no. 1, p. 011016, 2013.
9. M. Baloch, M. S. Honnurvali, A. Kabbani, T. Ahmed, S. T. Chauhdary, and M. S. Saeed, "Solar Energy Forecasting Framework Using Prophet-Based Machine Learning Model: An Opportunity to Explore Solar Energy Potential in Muscat, Oman," *Energies*, vol. 18, no. 1, p. 205, 2025.
10. A. T. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and meteorological information," *Renewable Energy*, vol. 118, no. 4, pp. 357–367, 2018.
11. S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Conversion and Management*, vol. 156, no. 1, pp. 459-497,2018.
12. J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, no. 10, pp. 78-111, 2016.
13. M. J. Sammar, M. A. Saeed, S. M. Mohsin, S. M. A. Akber, R. Bukhsh, and M. Abazeed, "Illuminating the future: A comprehensive review of AI-based solar irradiance prediction models," *IEEE Access*, vol. 12, no. 5, pp. 114394 - 114415, 2023.
14. B. K. Sahu, "A study on global solar PV energy developments and policies with special focus on the top ten solar PV power producing countries," *Renewable and Sustainable Energy Reviews*, vol. 43, no. 3, pp. 621-634, 2015.
15. L. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, no. 11, pp. 65-76, 2013.