Scientific Research Publishing

Journal of Computer and Communications, 2026, 14(11), 54-71
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Performance Study of the RNN-LSTM-XGBoost Algorithm for Photovoltaic Power Forecasting

Zhengyu Yang

College of Energy and Power Engineering, Lanzhou University of Technology, Lanzhou, China
Email: 985821664@qq.com

How to cite this paper: Yang, Z.Y. (2026) Performance Study of the RNN-LSTM-XGBoost Algorithm for Photovoltaic Power Forecasting. *Journal of Computer and Communications*, **14**, 54-71.
https://doi.org/10.4236/jcc.2026.1411003

Received: October 9, 2026
Accepted: November 3, 2026
Published: November 6, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

Among the forces now driving the expansion of world electricity supply, solar photovoltaic (PV) generation occupies a leading position. Figures assembled by the UK energy think tank Ember for its Global Electricity Review 2026 indicate that worldwide solar output amounted to 2778 TWh during 2025, some 30% above the level of the preceding year and equivalent to about 8.8% of all electricity produced globally. Since PV output is closely tied to weather conditions and consequently varies a great deal, the continued growth of installed capacity makes the accuracy of power forecasting, and its adaptability to different forecast horizons, increasingly important. Against this background, the present study examines how the RNN-LSTM-XGBoost hybrid model behaves in practice at several forecasting horizons. The algorithmic principle of the model is discussed first, experiments on measured PV data follow, and the forecasts obtained are then set against the measured values at a number of time horizons, so that the forecasting capability of the model can be judged comprehensively and the manner in which its accuracy changes with the forecast horizon can be examined.

Keywords

Photovoltaic Power Generation, Power Forecasting, LSTM, RNN-LSTM, XGBoost, Performance Testing

1. Introduction

Profound changes are under way in the way energy is produced and consumed worldwide, and PV power generation, being one of the pillars of clean energy, has expanded its installed base continuously. In its Renewables 2024 report the International Energy Agency (IEA) anticipates that renewable sources will satisfy close to one half of world electricity demand by 2030 [1]; PV generation is expected to

be a leading player in that development [2]-[4]. The power delivered by a PV plant is nevertheless governed by a range of meteorological variables solar irradiance, air temperature, cloud cover and relative humidity among them and consequently shows marked intermittency, volatility and uncertainty. Difficulties of considerable magnitude therefore arise for the dispatching and operation of power systems [5] [6]. Forecasting PV power with high accuracy consequently represents an important prerequisite if grid dispatching is to become more intelligent and if power systems are to remain secure and stable in operation [7] [8].

Techniques founded on deep learning have been taken up widely for forecasting PV power in recent years, a development spurred by the fast progress of artificial intelligence. Of the models commonly adopted, the long short-term memory (LSTM) network holds a central place, for it succeeds in retaining dependencies that extend across long intervals of a time series [7]. The recurrent neural network (RNN), on the other hand, is a structure built for sequential inputs and is therefore well matched to data that possess temporal dependence. Drawing on both families, the combined RNN-LSTM model unites their respective strengths and in this way attains a higher level of temporal modelling capability [8]. Belonging to the field of ensemble learning [9], the extreme gradient boosting (XGBoost) algorithm combines high fitting precision with modest computational cost; it may therefore be used for PV power forecasting in its own right or else joined to the hybrid deep learning model, which yields the RNN-LSTM-XGBoost hybrid forecasting approach [5] [10].

The RNN-LSTM-XGBoost algorithm has gained acceptance as a mainstream forecasting option, yet comparatively few investigations have probed its behaviour systematically across a range of forecast horizons. Because the results scattered through the literature are fragmentary, they offer little on which a quantitative judgement of model performance might be founded. The analysis presented here therefore appraises the RNN-LSTM-XGBoost algorithm along three axes: the pipeline through which the data are processed, the architecture of the model itself, and the forecasting performance that results. The aim is to make clear how the algorithm behaves when put to practical use and to establish the working situations to which it is best adapted, thereby furnishing a point of reference of practical value to researchers and practitioners concerned with PV power forecasting [11] [12].

2. Data Analysis and Processing

2.1. Handling of Missing Values and Outliers

The records of past generation at a PV plant, taken together with the meteorological quantities measured at the same location, provide the empirical foundation on which PV power forecasting depends. External disturbances of many kinds affect data acquisition in practice, so that raw records commonly hold both gaps and readings that are anomalous. Were such records passed to a model untreated, both training and forecasting would suffer a serious loss of accuracy. Three stages

are therefore followed in the experiment reported below. Gaps are located and rebuilt through cubic spline interpolation; readings that are anomalous are afterwards detected and discarded according to the interquartile range (IQR) box-plot criterion; and only the records that survive this treatment are kept for the training of the model.

2.2. Feature Selection and Correlation Analysis

A PV plant is at one and the same time subject to several meteorological variables, and the magnitude of the response to each of them varies widely. Feature selection is consequently carried out so that only those input variables displaying a meaningful relation to PV output are retained, an operation that both protects the forecasting results from irrelevant disturbance and reduces the complexity of the model input. The Pearson correlation coefficient is in common use as an indicator of how strongly two variables are linearly associated, and it takes the form:

$$\rho_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

where: x_i, y_i denotes the i-th observation of the variable x, y; $\bar{x}, \bar{y}$ denotes the mean value of the variable x, y; n denotes the total number of samples; and $\rho_{x,y}$ denotes correlation coefficient of the variable x, y.

Values of the Pearson correlation coefficient lie between −1 and 1. A positive coefficient signifies that the two variables move in the same direction, whereas a negative coefficient signifies that they move in opposite directions. How closely the two variables are related is indicated by the magnitude of the coefficient: the nearer that magnitude approaches 1, the stronger the association.

Table 1 lists the coefficients obtained. Air temperature, diffuse horizontal irradiance (DHI), direct normal irradiance (DNI) and the wind speed measured at 10 m are all strongly and positively related to the actual power, while azimuth, dew point temperature, relative humidity and zenith angle display strong negative relations with it. The remaining candidates, cloud opacity, atmospheric precipitable water, snow depth and the wind direction at 10 m, yield coefficients that are too small to be of use and are accordingly treated as irrelevant and discarded. The input features adopted in this experiment are consequently air temperature, DHI, DNI, wind speed at 10 m, azimuth, relative humidity and zenith angle.

2.3. Dataset Partitioning

PV generation displays a marked diurnal pattern: no power whatsoever is produced during the night, when sunlight is absent. Records whose actual output is zero must accordingly be removed before training commences, so that observations of zero power drawn from the non-generating hours are prevented from distorting the accuracy of the model.

It is a sensible division of the dataset that permits the generalisation ability of a deep learning model to be assessed, and the same division also provides protection

against overfitting. So that both the operating performance and the forecasting accuracy of each model may be improved, the preprocessed records are apportioned among three parts in the ratio 8:1:1. The part reserved for training serves to learn the parameters of the model, the part reserved for validation is kept for the tuning of hyperparameters, and the part reserved for testing is withheld until the final assessment of forecasting performance is made.

Table 1. Pearson correlation coefficients between meteorological factors and actual power.

Meteorological variable	Correlation coefficient
Air temperature	0.69
Azimuth	−0.52
Cloud opacity	−0.21
DHI (diffuse horizontal irradiance)	0.72
DNI (direct normal irradiance)	0.97
Atmospheric precipitable water	−0.28
Relative humidity	−0.74
Snow depth	−0.10
Wind direction at 10 m	0.40
Wind speed at 10 m	0.83
Zenith angle	−0.79

3. Analysis of the Forecasting Model

3.1. Principle of the RNN-LSTM-XGBoost Forecasting Algorithm

The recurrent neural network (RNN) forms a family of networks built expressly for sequential data. Through its internal "memory", an RNN is able to carry information from earlier time steps forward so that it influences the output produced at the current step. The long short-term memory (LSTM) network represents a significant refinement of the RNN and was devised to overcome the vanishing and exploding gradient problems that affect conventional recurrent networks. XGBoost, for its part, is an efficient and flexible gradient boosting algorithm that ranks among the best-performing methods currently used in industrial practice [12].

By bringing the RNN and the LSTM together, the combined RNN-LSTM model exploits the strengths of both. The RNN is adept at modelling short-range temporal behaviour and reacts rapidly to local variations within a sequence, whereas the LSTM excels at capturing dependencies that extend over longer spans. The two are complementary, and their union enhances the capacity of the model to extract features at differing time scales from PV power time series.

XGBoost is an efficient and flexible gradient boosting algorithm. Derived from the gradient boosted decision tree (GBDT), it retains the additive, forward-stagewise boosting framework of GBDT while incorporating extensive engineering op-

timizations and algorithmic improvements, and it has been widely adopted in industrial applications, where it generally delivers strong performance.

Traditional GBDT relies solely on first-order gradient information. XGBoost, in contrast, applies a second-order Taylor expansion to the loss function and incorporates a regularization penalty into its objective, thereby limiting model complexity, improving generalization, and mitigating overfitting to the training data.

RNN-LSTM-XGBoost is a forecasting approach that merges deep learning with ensemble learning. Taking the RNN-LSTM deep learning model as its starting point, it incorporates the XGBoost ensemble learning model and combines the predictive strengths of the two through a combination strategy [13].

A weighted fusion scheme is applied to the outputs of the two models in order to arrive at the final forecast. To determine the optimal fusion ratio, a ratio of 0.4 RNN-LSTM to 0.6 XGBoost was adopted a priori as the starting point st first, reflecting the stronger standalone performance of the tree ensemble on this dataset and the role of the neural model as a feature generator [7] [12]. The ratio was then verified by a sensitivity analysis in which α, the weight of the neural forecast, was varied from 0 to 1 in steps of 0.05 and the validation period was scored with MAE, RMSE and R^2 for each value. The validation MAE is minimised at $\alpha = 0.25$ (minimum MAE value, 0.9350) and the validation RMSE at $\alpha = 0.40$ (minimum RMSE value, 2.1579), while R^2 remains constant at 0.9959 for α between 0.20 and 0.65. The pre-specified value $\alpha = 0.40$ therefore coincides with the minimum of the validation RMSE, so it was retained. And the specific test results for the prediction weight α as shown in Table 2:

Table 2. Results for the prediction weight α.

α	$1-\alpha$	MAE/MW	RMSE/MW	R^2
0.00	1.00	0.9609	2.2220	0.9957
0.05	0.95	0.9515	2.2076	0.9957
0.10	0.90	0.9443	2.1949	0.9958
0.15	0.85	0.9393	2.1841	0.9958
0.20	0.80	0.9363	2.1751	0.9959
0.25	0.75	0.9350	2.1680	0.9959
0.30	0.70	0.9355	2.1627	0.9959
0.35	0.65	0.9373	2.1593	0.9959
0.40	0.60	0.9403	2.1579	0.9959
0.45	0.55	0.9447	2.1583	0.9959
0.50	0.50	0.9503	2.1606	0.9959
0.55	0.45	0.9572	2.1648	0.9959
0.60	0.40	0.9652	2.1709	0.9959
0.65	0.35	0.9745	2.1789	0.9958
0.70	0.30	0.9850	2.1887	0.9958

Continued

0.75	0.25	0.9966	2.2003	0.9958
0.80	0.20	1.0092	2.2137	0.9957
0.85	0.15	1.0227	2.2289	0.9956
0.90	0.10	1.0373	2.2458	0.9956
0.95	0.05	1.0528	2.2644	0.9955
1.00	0.00	1.0692	2.2846	0.9954

In conclusion, the weighted fusion strategy in the present experiment is adopted as the actual forecasting output, and the weighting formula for the two models is:

$$\hat{y} = 0.4 \times \text{RNN_LSTM}_{\text{pred}} + 0.6 \times \text{XGB}_{\text{pred}} \quad (2)$$

where: $\hat{y}$ denotes the weighted forecasting result, $\text{RNN_LSTM}_{\text{pred}}$ denotes the forecasting result of the RNN-LSTM deep learning model, and XGB_{pred} denotes the forecasting result of the XGBoost ensemble learning model. The working principle of the forecasting algorithm is shown in Figure 1:

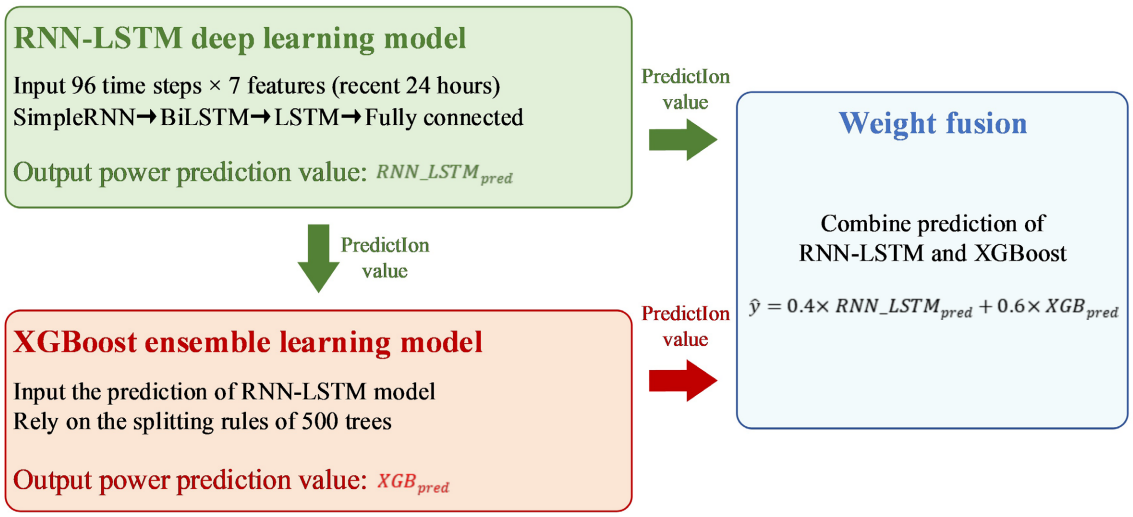


Figure 1. Working principle of the RNN-LSTM-XGBoost model.

3.2. Theoretical Analysis of the Forecasting Algorithm

A deep learning component (RNN-LSTM) and an ensemble learning component (XGBoost) are united within the RNN-LSTM-XGBoost combined forecasting model to give a single hybrid architecture. Considerable capacity for feature extraction and for non-linear fitting is possessed by the resulting structure, which makes it especially suitable for PV forecasting tasks marked by strong volatility and pronounced intermittency [3] [10] [11].

The RNN-LSTM-XGBoost combined model is endowed with a large parameter count and a correspondingly large capacity. Short-range fluctuations and long-range trends within a sequence can thus both be captured by it, and patterns of greater complexity can be learned. The danger of overfitting that accompanies a

single model is reduced to some degree when two models of different types are merged through a combination strategy, and that same strategy permits the outputs of the two models to be balanced against one another, so that overall robustness is improved [10] [14].

Taken as a whole, then, the RNN-LSTM-XGBoost combined model may be said to generalise well, and the results it yields in training are both accurate and credible.

4. Performance Evaluation of the Forecasting Model

4.1. Evaluation Metrics

Four metrics of a multi-dimensional character are adopted in this work so that the forecasting algorithm may be appraised from several directions at once, namely the mean absolute error (MAE), the root mean square error (RMSE), the mean absolute percentage error (MAPE) and the coefficient of determination (R^2) [15] [16]. Definitions of the four are given below:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{4}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{6}$$

where: y_i denotes the i-th true value, $\hat{y}_i$ denotes the i-th forecast value, $\bar{y}$ denotes the mean of the true values, and n denotes the number of data samples.

MAE conveys the average magnitude of the forecasting error without regard to sign, which renders it well suited to judging accuracy in overall terms; a smaller figure is to be preferred. By giving greater weight to large discrepancies, RMSE is able to describe how widely the forecasting errors are spread; here too a smaller figure is better. Stating the relative error as a percentage, MAPE allows quantities whose magnitudes differ greatly to be placed side by side, and a smaller value is again desirable. The coefficient of determination expresses the share of the variance in the true values that the forecasts can account for, so that a value approaching 1 signals a closer fit.

Two distinct sample populations are used in this study and are kept separate throughout. The estimation population is the set of records used to fit the models; in the preprocessing stage, records with zero output are removed from it, because they correspond to night-time intervals in which generation is structurally absent and their inclusion would only bias the fit towards the trivial solution. The evaluation population is the complete 15-minute grid of the forecast period, in which night-time zero-output intervals are retained. Retaining them is necessary because

the object of the assessment is a continuous power series rather than a daylight-only series, and because the error that matters operationally is defined at every instant, including intervals in which generation is zero. All metrics reported below are therefore computed on the evaluation population; the removal of zero-output records affects only model fitting and never the denominator of any reported metric.

Let the evaluation population contain n samples with measured values y_i and forecasts $\hat{y}_i$, and let $S_k = \{i \mid y_i > k \text{ MW}\}$, k donates threshold of MAPE. The four metrics are evaluated as Eq.(3) - (6): MAE, RMSE and R^2 are all computed on the full set of n samples. MAPE is the only one of the four whose value is governed by the denominator, so exact and near-zero values require an explicit convention. The convention adopted here is exclusion, not substitution, the adjustments to MAPE are as follows:

$$\text{MAPE} = \frac{100}{|S_k|} \sum_{i \in S_k} \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad S_k = \{i \mid y_i > k \text{ MW}\} \tag{7}$$

No artificial floor ε is added to the denominator. Substitution was rejected because the outcome then depends entirely on an arbitrary ε, the metric would quantify the choice of ε rather than the model. MAE, RMSE, R^2 are bounded at zero-output samples and are therefore computed over all n samples, including the zeros.

4.2. Design of the Comparative Analysis

PV power forecasting is often required over more than one time scale, so that comparisons can be drawn across scales between the short term (24 h), the medium term (72 h) and the long term (1 week). The statistical character of a power series alters appreciably from one scale to another, and a comparison of performance across forecasting scales accordingly supplies a basis on which the appropriate scale may be selected in practice.

PV forecasting is subject to many meteorological influences and therefore displays marked intermittency, volatility and uncertainty, from which it follows that the forecast curve and the actual power curve should be set against one another. Appreciating how closely the forecast curve tracks the measured one, and preparing plots that compare the two curves over various periods, makes the performance of the forecasting algorithm directly evident and furnishes a practical reference for engineering deployment.

4.3. Residual Analysis

By residual is meant the amount by which the actual observation differs from the value forecast or estimated by the model. Examining residuals in the context of PV power forecasting makes it possible to judge how well the model fits the data, a smaller residual indicating better performance. The residual is defined as:

$$e_i = y_i - \hat{y}_i \tag{8}$$

where y_i stands for the i-th true value, $\hat{y}_i$ stands for the i-th forecast value, and e_i stands for the i-th residual; $|e_i|$ the residual measures how far the forecast departs from the true value, e_i a positive value of it meaning under-prediction and a negative value meaning over-prediction.

The manner in which residuals are distributed throws light on both the systematic and the random portion of the forecasting error, and an analysis along these lines is consequently in wide use. Four aspects are usually taken into account: whether the mean equals zero, whether the variance remains constant, whether the distribution is normal, and whether the residuals are mutually independent [14] [17].

4.4. Framework of the Comparative Analysis

Drawing on the evaluation metrics and analytical dimensions set out above, the following framework for comparative analysis is established in this paper:

To begin with, the short, medium and long terms are taken as the time scales, and the overall forecasting performance achieved at each is set against the others on the basis of one common dataset and one common set of metrics; the assessment draws jointly on MAE, RMSE, MAPE and R^2.

Comparison plots of the forecast and actual power curves are then produced for each time scale, so that the groups can be distinguished directly in terms of how well they follow the trend of the power signal, how reliably they capture peaks and how promptly they react to abrupt changes.

Lastly, the residuals produced by the forecasting algorithm are examined with respect to their distribution, the mean, variance and shape of the residual distribution being considered, in order to judge how much of the forecasting deviation is systematic and how much is random.

5. Validation and Analysis of the Forecasting Model

5.1. Experimental Procedure

A fixed workflow is followed by the experiments reported here. It was adopted so that differences in working performance among the forecasting algorithms might be brought out while the influence of extraneous factors was excluded:

Zero values are first eliminated from the raw dataset and the features chosen; missing entries are afterwards compensated for and anomalous records discarded, which brings the preprocessing of the data to completion.

It should be noticed that during the preprocessing of dataset. To prevent information leakage during preprocessing, the data-handling procedure follows a strict split-then-fit order. The dataset is first partitioned by year into a training period (2015-2017), a validation period (2018) and a test period (2019). Every quantity that requires estimation—the interpolation function, feature correlations, outlier thresholds, the standardisation scalers and the power cap—is subsequently computed on the training period alone and applied to the validation and test periods as a frozen constant; no test-period observation enters any parameter estimate.

The interpolation step has three properties. First, it is executed within each split independently, so observations from one split are never used to fill another. Second, it is defined along the true timestamp axis rather than the record index, which prevents the temporal distortion that arises when sampling intervals are unequal. Third, it is forward-only and step-limited: only gaps of at most two consecutive missing steps are filled, and no backward filling is used anywhere in the pipeline.

The forecasts that emerge are then subjected to numerical evaluation, curve comparison and residual analysis at each of the three time scales (short, medium and long term), and from these a comprehensive assessment of the working performance of the model is finally assembled. This workflow is depicted in Figure 2.

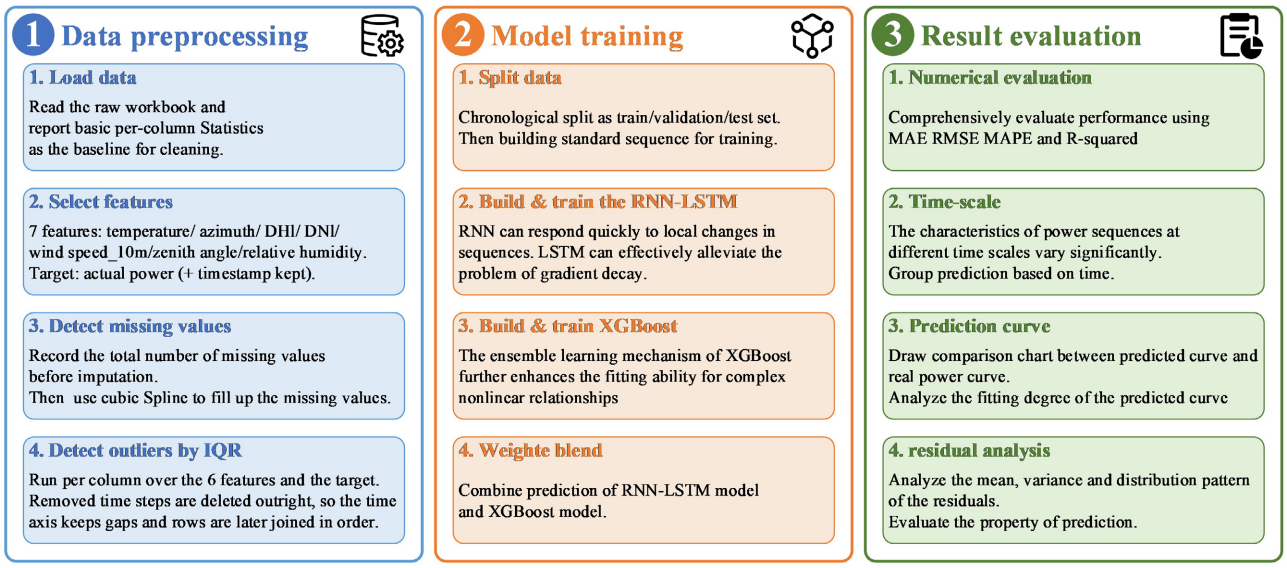


Figure 2. Flowchart of the experimental procedure.

5.2. Numerical Evaluation of Forecasting Performance

The measurements analysed in this work were supplied by a PV power station in Yumen City, Gansu Province, and they cover the years from 2015 to 2020. The dataset was obtained from publicly available online sources. It is sampled at 15-minute intervals, and the data of actual power is derived from irradiance. However, the installed capacity is not explicitly stated in the dateset.

Records of zero power recorded at night, together with observations that appeared anomalous, were discarded first. The surviving records were then apportioned among three subsets in the ratio 8:1:1: 17,152 records from 2015-2017 were used for training, 2144 records from 2018 for validation, and 2145 records from 2019 for testing. The entire 2020 dataset was set aside as an out-of-sample target that the model never sees during training. This arrangement was adopted so that the evaluation would reflect genuine generalisation ability.

Preprocessing follows a split-then-fit principle. The dataset is partitioned by

year into training (2015-2017), validation (2018) and test (2019) periods, and every subsequent statistic is estimated on the training period alone and then applied to the other splits as a frozen constant: interpolation is carried out within each split, along the true time axis, in a forward-only and step-limited manner (≤2 steps); feature correlations are computed solely on daytime training samples, with the final seven predictors fixed a priori; and the IQR thresholds are derived from the quantiles of daytime training samples, with no test-period data entering their estimation. The resulting outlier rates are 4.88%, 4.59% and 7.06% for the three splits, the difference reflecting inter-annual meteorological variability rather than adaptation of the thresholds to the test period.

For sequence construction, outliers are flagged as missing rather than deleted, so the full 15-minute time axis is preserved; gaps of at most two steps are filled by time-based interpolation, whereas longer gaps retain a missing mask and are not imputed. Sliding windows are built within each split only, and a window is retained only if all 96 steps are complete and its first-to-last time span equals exactly 95×15 minutes (23.75 h); windows containing gaps or temporal discontinuities are discarded.

For the RNN-LSTM branch, three recurrent stages are stacked in progressive order: SimpleRNN with 64 units, then a bidirectional LSTM with 128 units, and finally an LSTM with 64 units. Each stage is followed by batch normalisation and then by a Dropout layer whose rate lies between 0.2 and 0.3. The dense part of the network is shaped Dense(64)→Dense(32)→Dense(1). Optimisation uses Adam, with the learning rate set to 0.001, 64 samples in each batch, and at most 150 iterations. Two safeguards are added: early stopping with a patience of 20, and learning rate decay with a factor of 0.5 and a patience of 10.

For the XGBoost branch, 500 trees are grown to a greatest depth of 6, with a learning rate of 0.05 and sampling ratios of 0.8 for both rows and columns, while the L1 and L2 regularisation coefficients are set to 0.1 and 1.0. To combine the two branches, Equation (5) assigns a weight of 0.4 to the RNN-LSTM forecast and a weight of 0.6 to the XGBoost forecast.

Forecasting proceeds by rolling one-step-ahead recursion: with air temperature, DHI, DNI, wind speed at 10 m, azimuth and zenith angle as the meteorological input, the model outputs a single value every 15 minutes and then advances the input window by one step before predicting again.

Forecasting experiments were then run at three evaluation-window lengths: 24 hours, 72 hours and one week. Three periods were selected: 15 May 2020 for the short term, 22 - 24 February 2020 for the medium term, and 13 - 19 April 2020 for the long term. Table 3 collects the four metrics obtained at each scale.

Table 3 shows how the model performs at each scale. For the short term it achieves an MAE of 2.68 MW, an RMSE of 4.64 MW, a MAPE of 20.68% and an R^2 of 0.9899. Clearly the model captures the diurnal cycle of PV output closely.

Excluding exact zeros does not by itself stabilise MAPE, because the denomina-

tor stays small for samples immediately adjacent to sunrise and sunset. In the 72-h window, for instance, only 7 samples satisfy $0 < y_i < 1$ MW, yet they alone raise MAPE from 76.9% (over $y \geq 1$ MW) to 381.4% (over $y > 0$). A restricted variant is therefore reported alongside:

Table 3. Evaluation metrics of forecasting performance at different time scales.

Forecast horizon	MAE/MW	RMSE/MW	MAPE/%	MAPE (≥5 MW)/%	R^2
Short term (24 h)	2.68	4.64	20.68	14.30	0.9899
Medium term (72 h)	10.70	15.50	381.42	33.87	0.7652
Long term (1 week)	11.68	18.39	268.18	57.26	0.7744

$$\mathrm{MAPE}\left(\geq 5\ \mathrm{MW}\right) = \frac{100}{|S_5|}\sum_{i\in S_5}\left|\frac{y_i - \hat{y}_i}{y_i}\right|,\quad S_5 = \left\{i \mid y_i > 5\ \mathrm{MW}\right\} \tag{9}$$

This variant is a diagnostic of accuracy within the operating range; it is not presented as an improvement on MAPE and does not replace it. SO both values are reported for every window.

Extending the horizon changes the picture. The MAE rises to 10.70 MW at the medium term and 11.68 MW at the long term—roughly 4 times the short-term value. The RMSE grows by a factor of about 3.3, and R^2 falls from 0.9899 to 0.7652 and then 0.7744, a loss of about 0.22.

The value of MAPE depends heavily on how it is defined, as the three scales make clear. Samples taken at night, when the true power is zero, account for 41.2% - 53.6% of the data; because MAPE divides by the true value, its relative error explodes as that denominator approaches zero. Medium- and long-term MAPE is inflated for exactly this reason.

Filtering changes the numbers a great deal. Excluding samples below 1 MW cuts the medium-term MAPE to 76.89%. Keeping only samples of at least 5 MW cuts it further, to 33.87% at the medium term and 57.26% at the long term. MAPE (≥5 MW) is therefore reported when PV forecasting accuracy is assessed in this paper. Accuracy does not keep falling as the horizon grows. Medium- and long-term results are worse than the short-term result, yet both settle at a comparable level: an MAE of about 11 MW, an R^2 of about 0.77 and a MAPE (≥5 MW) of about 45.57%.

The 5 MW threshold was chosen on three grounds rather than by optimisation: it corresponds to 3.8% of peak output (131.17 MW) and 10.7% of mean daytime output (46.84 MW); it retains 85% - 91% of the non-zero samples across the three windows; and it caps the contribution of a 1 MW error at 20%, so the metric cannot be dominated by near-zero denominators. Table 4 shows that MAPE falls monotonically with the threshold (72-h window: 381.4% at $y > 0$ to 27.4% at ≥10 MW), confirming that the variation reflects the denominator rather than model behaviour and that no threshold is optimal.

Table 4. MAPE of each time scales under different thresholds.

Threshold	24-h MAPE/%	72-h MAPE/%	1-week MAPE/%
0	20.68%	381.42%	268.18%
0.5	20.68%	82.42%	153.99%
1	19.99%	76.89%	110.86%
2	18.03%	46.78%	88.12%
3	18.30%	45.07%	66.08%
5	14.30%	33.87%	57.26%
7	11.28%	30.35%	49.30%
10	11.28%	27.40%	37.55%

Once the forecast passes a single day, the model reaches a plateau. What limits its accuracy beyond that point is how faithfully the phase of the diurnal cycle can be reproduced, not the length of the forecasting step by itself.

5.3. Comparison of Forecasting Curves

The forecast and actual power curves are set against one another in Figures 3-5 for the short, medium and long terms respectively. Throughout the figures, the actual power appears as a black solid line and the model forecast as a red solid line.

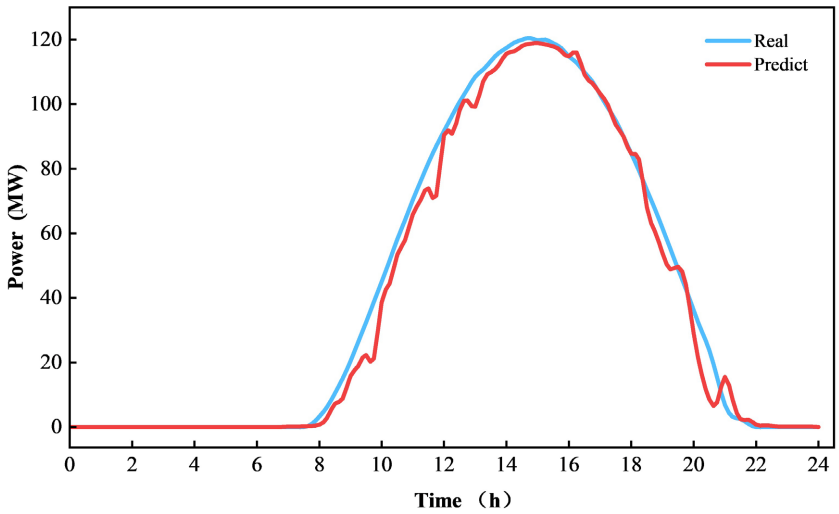


Figure 3. Comparison between forecast and actual power at the short-term (24 h) scale.

Consider first the short-term panel. Here the forecast tracks the measured curve most closely, and discrepancies of any consequence appear only in the morning and evening portions. The medium- and long-term panels still reproduce the general shape of the signal, but the forecast curve broadens the window between sunrise and sunset noticeably. Power is therefore still predicted during intervals in

which actual output has already fallen to zero, and a sustained positive deviation overnight is the result.

It is evident from the figures that the measured curve is followed most closely by the short-term forecast, discrepancies of any consequence being confined to the morning and evening portions. The general shape of the signal is still reproduced by the medium- and long-term forecasts, but the window between sunrise and sunset is distinctly broadened in the forecast curve, so that power continues to be predicted during intervals in which actual output has already fallen to zero; a sustained positive deviation overnight is the consequence.

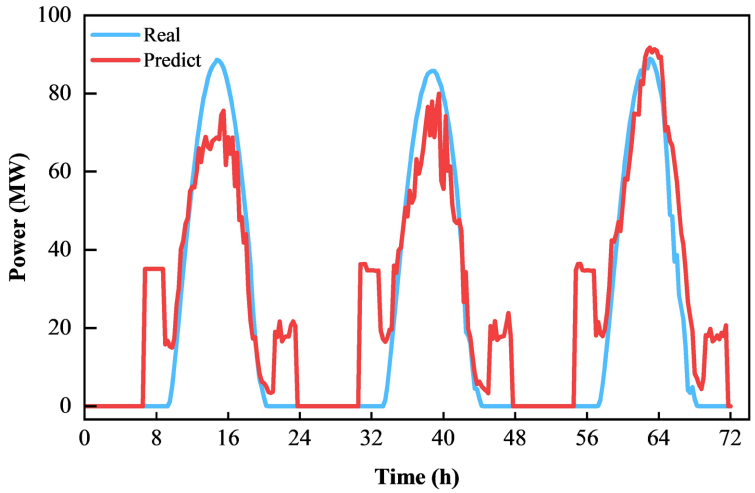


Figure 4. Comparison between forecast and actual power at the medium-term (72 h) scale.

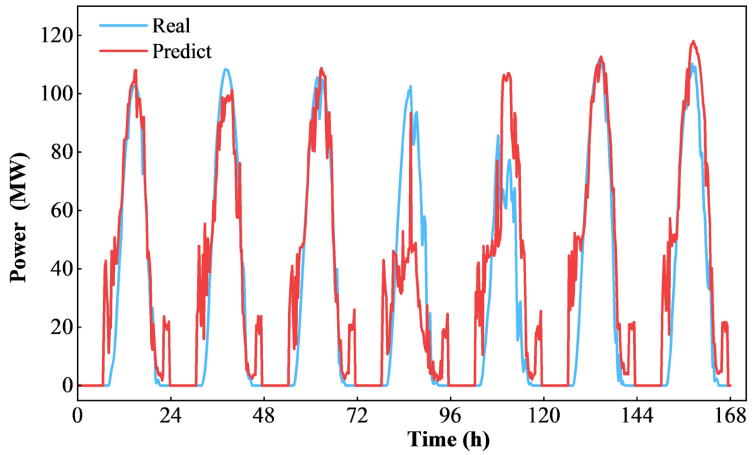


Figure 5. Comparison between forecast and actual power at the long-term (1 week) scale.

5.4. Analysis of Residual Distribution Characteristics

The forecasting error contains both a systematic and a random part. To separate them, the distribution of the residuals is examined below. Three aspects of that distribution receive attention: its centre, its spread, and its shape; its temporal dependence is considered afterwards. Table 5 gathers the resulting statistics.

Table 5. Statistics of residual distribution characteristics at different time scales.

Statistic	Short term (24 h)	Medium term (72 h)	Long term (1 week)
Residual mean ME/MW	+2.10	−6.88	−5.59
Residual standard deviation SD/MW	4.16	13.91	17.53
Residual range/MW	−8.66 - +17.52	−36.71 - +24.11	−66.81 - +56.30
Skewness	+1.3912	−0.5015	+0.3140
Excess kurtosis	+2.8728	−0.2681	+2.2359
Shapiro-Wilk test p-value	1.55×10^{-9}	2.11×10^{-12}	1.26×10^{-20}
Points of underestimation/overestimation	46/22	71/201	128/348
Net deviation as a share of total generation	+5.17%	−30.50%	−18.76%
Share of points with |residual| ≤ 5 MW	77.3%	43.8%	50.1%
Share of points with |residual| ≤ 10 MW	94.8%	58.3%	60.0%

Take the short-term residual first. Its mean is +2.10 MW, so the model tends to underestimate slightly, and that underestimation is concentrated in a small number of intervals rather than spread evenly. The errors are therefore strongly clustered.

The medium- and long-term residuals behave in the opposite way. Their means are −6.88 MW and −5.59 MW, and the corresponding net deviations amount to −30.50% and −18.76% of total generation. Across days, then, systematic overestimation dominates the error. Error dispersion widens sharply with the horizon. The residual standard deviation comes to 4.16 MW at the short term, 13.91 MW at the medium term and 17.53 MW at the long term.

The spread of absolute errors tells a similar story. For 94.8% of the short-term samples the absolute error stays within 10 MW, yet an absolute error within 10 MW holds for only 58.3% of the medium-term samples and 60.0% of the long-term samples. Large errors are accordingly no longer occasional at long horizons; they have become routine.

Shape and symmetry differ between the scales, as the skewness and kurtosis values reveal. The short-term residual has a skewness of +1.3912 and an excess kurtosis of +2.8728; the long-term residual takes +0.3140 and +2.2359. Both are right-skewed with heavy tails, which means that a few samples underestimate the power by a comparatively large margin while most samples err by very little.

The medium term behaves differently. Its skewness is −0.5015 and its excess kurtosis −0.2681, so the distribution leans to the left and sits comparatively flat. Moderate overestimation predominates and the errors are rather dispersed.

Normality cannot be assumed for any of the three scales; the p-value from the Shapiro-Wilk test falls far below 0.05 in every case. This has a practical consequence. Confidence intervals built on the normality assumption are unsuitable for describing the uncertainty of the forecasts, and empirical quantiles are used instead.

One feature of the dataset deserves emphasis. The "actual power" values used in these experiments correlate strongly and linearly with the diffuse horizontal irradiance (DHI) recorded at the same station.

The R^2 values of the order of 0.99 reported above therefore chiefly express how well the model reproduces that conversion relation, rather than the accuracy of an independent measurement. The conclusions drawn in this paper about performance decay across scales and about the structure of the residuals nevertheless hold. For engineering dispatching decisions, however, verification against independently measured power data would be advisable.

Two distinct sample populations are used in this study and are kept separate throughout. The estimation population is the set of records used to fit the models; in the preprocessing stage, records with zero output are removed from it, because they correspond to night-time intervals in which generation is structurally absent and their inclusion would only bias the fit towards the trivial solution. The evaluation population is the complete 15-minute grid of the forecast period, in which night-time zero-output intervals are retained. Retaining them is necessary because the object of the assessment is a continuous 24-hour, 72-hour and one-week power series rather than a daylight-only series, and because the error that matters operationally is defined at every instant, including intervals in which generation is zero. All metrics reported in Table 3 and Table 4 are therefore computed on the evaluation population; the removal of zero-output records affects only model fitting and never the denominator of any reported metric.

6. Conclusions

The performance of the RNN-LSTM-XGBoost PV power forecasting method has been investigated experimentally and systematically at three time scales—24 hours (short term), 72 hours (medium term) and 1 week (long term). The principal findings are set out below.

1) Forecasting accuracy declines sharply as the forecasting step grows longer, and a plateau effect is present. An MAE of 2.68 MW, an RMSE of 4.64 MW and an R^2 of 0.9899 are obtained for the short term, whereas the medium and long terms raise the MAE to 10.70 MW and 11.68 MW and lower R^2 to 0.7652 and 0.7744. The latter two figures are essentially equal, so the model may be said to reach a plateau in accuracy once the horizon passes one day, beyond which lengthening the step further brings no substantial deterioration.

2) MAPE is sensitive in PV power forecasting to the way it is defined. Night-time samples of zero power occupy a considerable share of the data, and MAPE is inflated as a result: the apparent medium-term value reaches 381.42%, falling to 33.87% once only samples whose true value is at least 5 MW are counted. MAPE ($\geq$5 MW) is accordingly recommended for evaluation.

3) The residual distribution is plainly non-normal and displays strong autocorrelation, and its systematic component grows with the forecasting step. The hypothesis of normality is rejected at all three scales. The mean residual moves from

+2.10 MW in the short term (net deviation +5.17%, underestimation predominating) to −6.88 MW in the medium term and −5.59 MW in the long term (net deviations of −30.50% and −18.76%, overestimation predominating).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] IEA (2024) Renewables 2024: Analysis and Forecast to 2030. International Energy Agency. https://www.iea.org/reports/renewables-2024

[2] Sadeghi, D., Golshanfard, A., Eslami, S., Rahbar, K. and Kari, R. (2023) Improving PV Power Plant Forecast Accuracy: A Hybrid Deep Learning Approach Compared across Short, Medium, and Long-Term Horizons. *Renewable Energy Focus*, **45**, 242-258. https://doi.org/10.1016/j.ref.2023.04.010

[3] Akhter, M.N., Mekhilef, S., Mokhlis, H., Almohaimeed, Z.M., Muhammad, M.A., Khairuddin, A.S.M., *et al.* (2022) An Hour-Ahead PV Power Forecasting Method Based on an RNN-LSTM Model for Three Different PV Plants. *Energies*, **15**, Article 2243. https://doi.org/10.3390/en15062243

[4] Armeanu, D.S., Joldes, C.C., Gherghina, S.C. and Andrei, J.V. (2021) Understanding the Multidimensional Linkages among Renewable Energy, Pollution, Economic Growth and Urbanization in Contemporary Economies: Quantitative Assessments across Different Income Countries' Groups. *Renewable and Sustainable Energy Reviews*, **142**, Article ID: 110818. https://doi.org/10.1016/j.rser.2021.110818

[5] Chahboun, S. and Maaroufi, M. (2021) Principal Component Analysis and Machine Learning Approaches for Photovoltaic Power Prediction: A Comparative Study. *Applied Sciences*, **11**, Article 7943. https://doi.org/10.3390/app11177943

[6] de Oliveira Santos, L., AlSkaif, T. and Carvalho, P.C.M. (2026) Hybrid Photovoltaic Power Forecasting by Combining Physical and Machine Learning Models. *Solar Energy*, **317**, Article ID: 114951. https://doi.org/10.1016/j.solener.2026.114951

[7] Ren, Y., Suganthan, P.N. and Srikanth, N. (2015) Ensemble Methods for Wind and Solar Power Forecasting—A State-of-the-Art Review. *Renewable and Sustainable Energy Reviews*, **50**, 82-91. https://doi.org/10.1016/j.rser.2015.04.081

[8] Zhao, T. and Huang, L. (2025). A Short-Term Photovoltaic Power Forecasting Method Based on Meteorological Data Using GRF-LSTM-XGBoost Model. 2025 7*th Asia Energy and Electrical Engineering Symposium* (*AEEES*), Chengdu, 28-31 March 2025, 1068-1073. https://doi.org/10.1109/aeees64634.2025.11019969

[9] Hossain, M.S. and Mahmood, H. (2020) Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast. *IEEE Access*, **8**, 172524-172533. https://doi.org/10.1109/access.2020.3024901

[10] Wang, K., Qi, X. and Liu, H. (2019) Photovoltaic Power Forecasting Based Lstm-Convolutional Network. *Energy*, **189**, Article ID: 116225. https://doi.org/10.1016/j.energy.2019.116225

[11] Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J. and Antonanzas-Torres, F. (2016) Review of Photovoltaic Power Forecasting. *Solar Energy*, **136**, 78-111. https://doi.org/10.1016/j.solener.2016.06.069

[12] Markovics, D. and Mayer, M.J. (2022) Comparison of Machine Learning Methods for Photovoltaic Power Forecasting Based on Numerical Weather Prediction. *Renewable*

and Sustainable Energy Reviews, **161**, Article ID: 112364.
https://doi.org/10.1016/j.rser.2022.112364

[13] Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B. and Talebiesfandarani, S. (2019) $PM_{2.5}$ Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere*, **10**, Article 373.
https://doi.org/10.3390/atmos10070373

[14] Wang, J., Zhang, Z., Xu, W., Li, Y. and Niu, G. (2025) Short-Term Photovoltaic Power Forecasting Using a BI-LSTM Neural Network Optimized by Hybrid Algorithms. *Sustainability*, **17**, Article 5277. https://doi.org/10.3390/su17125277

[15] Sur, E. (2026) A Hybrid 1D-CNN-LSTM-MHA Model for Short-Term Probabilistic Photovoltaic Power Forecasting. *Solar*, **6**, Article 30.
https://doi.org/10.3390/solar6030030

[16] Qiu, Y., Liu, Y. and Zheng, X. (2024) Research on Solar Photovoltaic Panel Power Generation Prediction and Comparative Analysis Based on Multiple Machine Learning Algorithms. 2024 3*rd International Conference on Energy, Power and Electrical Technology* (*ICEPET*), Chengdu, 17-19 May 2024, 83-86.
https://doi.org/10.1109/icepet61938.2024.10625860

[17] Bin Yousuf, K., Akter, A., Noor, H.A., Hoque, A. and Ahmed, A. (2026) A Hybrid XGBoost-LSTM Model with Physics-Informed Features and Uncertainty Quantification for Solar Power Forecasting. *Energy Reports*, **15**, Article ID: 109068.
https://doi.org/10.1016/j.egyr.2026.109068