

ATTENTION-ENHANCED LSTM-BASED SOLAR POWER MANAGEMENT FORECASTING: A COMPARATIVE ANALYSIS WITH RECURRENT NEURAL NETWORK ARCHITECTURES

Sasirekha P¹, Durgadevi Velusamy², Karthikeyan Ramasamy¹

¹Electrical and Electronics Engineering, M.Kumarasamy College of Engineering, Thalavapalayam, Karur, 639 113, Tamilnadu, India.

²*Information Technology, SSN School of Engineering, Shiv Nadar University, Chennai, Kalavakkam, Chennai, 603 110, Tamilnadu, India.

Corresponding Author: Sasirekha P
Sasirekhaapp@Gmail.Com

Abstract: The prediction of solar power production plays an important role in the proper functioning and efficient management of renewable energy systems. Nevertheless, the solar power generation varies continuously with changes in weather conditions, which make the process of forecasting difficult. The purpose of this paper is to evaluate the accuracy of forecasting with the application of Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (Bi-LSTM) models combined with the attention mechanism for solar power prediction. The efficiency of the models is measured by means of the coefficient of determination (R^2), Mean Squared Error (MSE), and Mean Absolute Error (MAE). The research results show that the best results are obtained with the Attention + LSTM model, with the highest mean values of R^2 of 0.9555, MSE of 0.4205, and MAE value is 0.4521. The mean R^2 for the Attention + Bi-LSTM model is 0.9468 and the conventional Bi-LSTM model shows only slightly better results in terms of mean R^2 value of 0.9473. Therefore, the implementation of attention does not necessarily improve the performance of LSTM-based architectures. The overall results of the comparative analysis reveal that the efficiency of attention is based on the particular recurrent architecture chosen. Of all the considered models, the one that delivers the best prediction results is the Attention + LSTM model, which is capable of identifying the temporal patterns for solar power generation.

Keywords: Bi-LSTM, LSTM, Soft Attention, Solar Energy, Power Generation, Forecasting, Time series.

1. INTRODUCTION

As the need for dependable, sustainable, and clean electricity grows, solar energy has emerged as a significant renewable energy source. Solar energy is among the most accessible renewable energy sources. Solar power is appropriate for both urban and rural places, including isolated and off-grid areas, because sunlight is naturally plentiful and can be used in a variety of settings [1]. Solar energy may generate electricity without directly releasing dangerous pollutants into the atmosphere, in contrast to traditional power generation techniques based on coal, gas, and other fossil fuels. Solar photovoltaic panels convert sunlight into electrical energy and can be installed on rooftops, open land, commercial buildings, educational institutions, and industrial facilities. This flexibility makes solar power a practical option for meeting the growing electricity demand while reducing dependence on conventional energy resources.

The increasing adoption of solar power has also been supported by technological developments and government initiatives. In India, the Ministry of New and Renewable Energy (MNRE) has been promoting the expansion of renewable energy as part of the country's long-term energy goals. According to the data presented in the source material, India is working towards achieving 500 GW of non-fossil fuel-based installed capacity by 2030. The Renewable Energy Statistical Report 2023–2024 reported that India's renewable energy capacity had reached 190.57 GW, representing nearly 43% of the country's total installed power generation capacity. Solar power has contributed significantly to this growth and has become one of the major components of India's renewable energy sector. Globally, solar power has also experienced substantial growth over the past decade, reflecting its increasing importance in the transition towards cleaner electricity generation. Therefore, solar energy can be considered a clean, scalable, and increasingly important source of electricity for India's future energy requirements.

Although solar energy offers several advantages, its power generation is not constant throughout the day. Changes in weather conditions such as temperature, solar irradiance, cloud cover, humidity, wind speed, and atmospheric pressure can cause considerable variations in the amount of electricity produced. These variations make solar power forecasting a challenging task. Accurate forecasting is important because it helps power system operators estimate future generation, balance electricity supply with demand, schedule energy storage systems, and manage power distribution more effectively. Reliable forecasts can also support the integration of larger amounts of solar power into the electrical grid while improving the overall economic operation of renewable energy systems [2].

Several machine learning techniques have been explored for solar power forecasting, including Artificial Neural Networks (ANN), Support Vector Machine Regression (SVMR), k-Nearest Neighbours (k-NN), Gaussian Regression (GR), k-means clustering, and Classification and Regression Trees (CART). These techniques have proven themselves able to recognize correlations between weather and solar power generation. However, solar power data often contains complex and nonlinear patterns that can be difficult to represent using conventional statistical approaches. For this reason, deep learning techniques have received increasing attention in recent years. Deep learning models are capable of learning complex relationships directly from large datasets and can automatically extract useful features from the available input variables. In this study, meteorological parameters such as humidity, temperature, tilt angle, wind speed, atmospheric pressure, cloud conditions, dew point, and solar irradiance are considered for predicting solar power generation [3–4].

Traditional linear regression methods may not be well suited to time-series solar power data because they can have difficulty representing nonlinear relationships, complex feature interactions, and the influence of abnormal observations. Machine learning methods such as decision trees, random forests, support vector machines, and gradient boosting can provide better representations of these relationships by learning patterns from the available data. Deep learning models can further improve this process because their multilayer structure allows them to learn different levels of representation from the input data. This ability is particularly useful when the dataset contains large amounts of time-dependent information and complex relationships between weather parameters and power generation.

Several deep learning architectures, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU), have been widely investigated for time-series forecasting because of their ability to learn patterns from sequential data [5]. In this work, LSTM, Bidirectional LSTM (Bi-LSTM), GRU, and Bidirectional GRU (Bi-GRU) models are considered for solar power prediction. An attention mechanism is incorporated into the recurrent models to allow the network to assign greater importance to the more relevant time steps in an input sequence. This helps the model focus on information that has a stronger influence on the predicted output and can be useful when dealing with long-term dependencies. Attention-based recurrent models are therefore particularly suitable for multivariate time-series forecasting, where several weather parameters may influence solar power generation at different time steps.

The dataset used in this study was collected from a 3 MW solar power plant located at Manjanaickenpatti in Karur. Weather information required for the forecasting process was obtained from the nearby Tiruchirapalli Airport. The collected data was initially pre-processed to improve its suitability for model development and to extract the required features. Min-max normalization was applied to bring the input variables into a common numerical range before they were provided to the prediction models. The performance of the developed models was then evaluated using different statistical measures, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), coefficient of determination (R^2), Adjusted R^2 , and Mean Absolute Percentage Error (MAPE). These measures provide different perspectives on the accuracy and reliability of the forecasting models.

In addition to individual deep learning models, this research considers ensemble-based approaches through stacking and multi-level stacking techniques for energy forecasting. The purpose of these approaches is to combine the strengths of different predictive models and obtain more reliable forecasting results. The hyperparameters of the selected models are adjusted using grid search and nature-inspired optimization techniques to identify suitable model configurations. Repeated K-fold cross-validation is also considered to improve the reliability of the evaluation and reduce the possibility of obtaining results that depend heavily on a particular division of the dataset.

A 10-fold cross-validation procedure is used to examine the generalization capability and robustness of the developed models. Unlike the conventional holdout approach, where the available data is divided into fixed training and testing sets, K-fold cross-validation uses multiple data partitions for training and validation. This provides a more consistent assessment of model performance across different subsets of the available data. Another important aspect of the proposed approach is the use of weighted averaging for combining the outputs of weak learners within the deep learning framework. The models are trained using an appropriately selected learning rate and the Adam optimizer to improve the learning process. Through these combined techniques, the study aims to develop and evaluate reliable models for predicting the energy output of a solar power plant and to identify the recurrent architecture that provides the most suitable forecasting performance.

The main contribution of this study is summarized as follows,

1. This paper proposes a comparative deep learning framework for solar power forecasting by combining attention mechanism with four recurrent neural network architectures including Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM). The proposed framework aims to study the capability of different recurrent architectures to learn complex temporal patterns in solar power generation.

2. The comparative analysis of LSTM-based models is performed through the use of traditional LSTM model, traditional Bi-LSTM model, Attention + LSTM model, and Attention + Bi-LSTM model to test the ability of attention models in solar power forecasting.
3. An attention mechanism is used to detect temporal information from the input sequence, thereby allowing the forecasting models to concentrate on relevant temporal relationships in the input sequence.
4. The developed forecasting models are evaluated using various performance indicators, such as the coefficient of determination (R^2), Mean Squared Error (MSE), and Mean Absolute Error (MAE), to provide a comprehensive assessment of prediction accuracy and model performance.

2. RELATED WORK:

The existing research on solar power forecasting shows a gradual shift from conventional statistical techniques towards machine-learning and deep-learning methods because photovoltaic power generation changes continuously with weather and environmental conditions. The study in [6] examined solar radiation forecasting using the ARIMA model and showed that historical solar-radiation data can be used to identify time-dependent patterns and estimate future radiation levels. Although statistical approaches are useful for analysing regular trends, their ability to represent complicated nonlinear variations in solar generation can be limited. In [7], a hybrid SARIMA-RVFL model supported by wavelet decomposition was developed for very short-term solar PV power forecasting. The wavelet technique was used to separate the solar-power signal into different components before forecasting, allowing the model to handle fluctuations more effectively.

The authors in [8] investigated different machine-learning techniques for predicting solar PV panel power under changing environmental conditions. Their work considered the influence of weather-related parameters on PV output and showed that machine-learning models can learn the relationship between environmental conditions and generated power without relying only on fixed mathematical relationships. In [4], a hybrid KNN-SVM model was proposed for solar power forecasting by combining two different learning approaches. The study showed that combining models can provide improved prediction capability because each algorithm can contribute different learning characteristics. The research presented in [10] concentrated on forecasting global solar irradiance and temperature for rooftop photovoltaic power generation using multi-step LSTM and GRU-based hybrid models. Since irradiance and temperature have a direct influence on PV output, accurately predicting these parameters can help improve the final power-generation forecast.

The use of LSTM and GRU also demonstrated the importance of recurrent neural networks in learning patterns from sequential weather data. In [11], machine-learning techniques were applied to rooftop photovoltaic power forecasting using historical PV and environmental information. The study emphasized that the accuracy of the prediction is influenced not only by the selected algorithm but also by the quality and relevance of the input variables. The authors in [12] further investigated solar PV power prediction under fluctuating environmental conditions using comparative machine-learning techniques. Their work demonstrated that solar power forecasting becomes more challenging when radiation, temperature and other environmental factors change rapidly, making data-driven methods useful for identifying hidden relationships in the available measurements.

The study in [13] investigated deep-learning approaches for short-term photovoltaic power forecasting and showed that recurrent architectures can learn the changing behaviour of solar power from previous observations. LSTM and GRU models are particularly useful in this application because their internal memory mechanisms allow information from previous time steps to influence future predictions. In [14], a CNN-LSTM hybrid model was developed for solar power forecasting. The CNN component was used to identify important patterns from the input data, while the LSTM component was responsible for learning the time-dependent behaviour of the extracted information.

This combination shows that using more than one type of neural-network architecture can be beneficial when the solar-power data contain both feature patterns and temporal dependencies. The authors in [15] proposed a combined LSTM and Gaussian Process Regression approach for short-term solar power forecasting. In this method, LSTM was used to learn the sequential behaviour of solar-power data, while Gaussian Process Regression provided an additional modelling capability. The study indicates that combining different forecasting techniques can provide a more flexible approach to dealing with the uncertainty and variability associated with solar generation. The research in [16] investigated both univariate and multivariate LSTM approaches for forecasting the output of a utility-scale photovoltaic system. The study considered historical PV power along with relevant meteorological information and examined different prediction intervals. The results indicated that the importance of individual input variables can change according to the forecasting period, which means that selecting suitable input data is an important part of developing an accurate forecasting model. In [17], deep-learning neural networks were investigated for short-term photovoltaic power prediction, with recurrent architectures being used to learn the temporal characteristics of solar generation.

The study demonstrated that deep-learning models are capable of handling nonlinear and rapidly changing PV-power patterns more effectively than many conventional approaches. The authors in [18] introduced an LSTM

model combined with an attention mechanism for short-term photovoltaic power forecasting. Instead of treating every part of the input sequence with equal importance, the attention mechanism allows the model to assign greater importance to information that is more useful for producing the prediction. This approach is particularly useful for solar forecasting because weather and power conditions at different previous time steps may not contribute equally to the future output.

In [19], an LSTM-TCN hybrid architecture was proposed for one-step and multi-step very-short-term PV power forecasting. The LSTM component was used to capture sequential dependencies, whereas the temporal convolution component helped extract patterns from the time-series data. The study demonstrated that combining different deep-learning structures can provide better representation of complex PV-power behaviour. The research in [20] also explored recurrent forecasting with attention-based learning, showing that attention can help a neural network identify the more significant information within a sequence. These developments indicate that solar power forecasting has progressed from simple statistical prediction towards more flexible machine-learning, recurrent and hybrid deep-learning models. Among the recurrent architectures, LSTM has the advantage of maintaining important information over longer sequences, while GRU provides a comparatively simpler structure with fewer gates and parameters. Bi-LSTM extends the conventional LSTM by processing the sequence in both forward and backward directions, allowing the model to obtain a broader representation of the available sequence. Similarly, Bi-GRU combines the simpler GRU structure with bidirectional processing and can therefore be investigated as an alternative recurrent architecture. Although many existing studies have examined individual LSTM, GRU, CNN-LSTM and attention-based models, a direct comparison of LSTM, Bi-LSTM, GRU and Bi-GRU under the same experimental conditions is useful for understanding which recurrent architecture is more suitable for solar power forecasting. Therefore, the present work focuses on implementing and comparing these four models using the same input data, preprocessing procedure and evaluation criteria. The comparison helps determine how the differences in network structure influence prediction accuracy and provides a clearer basis for selecting an appropriate deep-learning model for solar PV power forecasting

3. DATASET DESCRIPTION AND PREPROCESSING

3.1 SOLAR POWER DATASET

The experimental work was carried out using operational data obtained from a grid-connected photovoltaic (PV) power plant located at Manjanaickenpatti in Karur district, Tamil Nadu, India. The plant is located at approximately 11.1271° N latitude and 78.6569° E longitude and covers an area of nearly 50 acres. It is a ground-mounted solar PV installation with an installed capacity of 3 MW. The plant contains around 33,000 photovoltaic modules, with each module having a rated power of 305 W. The plant produces an average annual energy output of approximately 15,500 MWh. The dataset contains daily solar power generation data recorded between 1 April 2019 and 31 March 2023. Thus, the study covers almost four years of plant operation. The extended observation period is useful for capturing changes in solar generation across different seasons and years. It also provides the prediction models with a wide range of operating and environmental conditions for training and performance evaluation.

Solar irradiance measurements were obtained at the plant using two KIPP & ZONEN CMP11 thermopile pyranometers. The instruments were used to measure plane-of-array irradiance as well as global horizontal irradiance (GHI). These measurements provide important information about the amount of solar radiation available at the site and its relationship with the electrical power produced by the PV system. Along with solar irradiance and power generation, several weather-related parameters were considered to represent the environmental conditions affecting PV performance. Wind speed, module temperature, and ambient temperature were measured at the site using a Davis 6163 Wireless Vantage Pro2 Plus weather station. The station was equipped with a 24-hour fan-aspirated radiation shield and an AeroCone for weather monitoring. Additional meteorological information, including humidity, cloud cover, and other wind-related parameters, was obtained from the Weather Underground (Wunderground) platform using observations from the Tiruchirapalli International Airport weather station. The combination of these parameters provides a broader representation of the atmospheric conditions that can influence solar power generation.

The PV array was installed facing south and used a seasonally adjusted inclination angle to improve solar energy collection. The tilt angle was changed according to the seasonal variation in solar elevation. The panel inclination was maintained at 5° from April to September, increased to 10° during March, and set to 25° from October to February. The seasonal tilt settings used for the PV panels are presented in Table 1.

Table 1. Seasonal variation in PV panel inclination angle

Tilt Angle	Month
05	From April to September

10	March
25	From October to February

The measurements were automatically collected through the data acquisition system installed at the plant. Since the data was obtained from an operational solar power plant over a long period, some missing observations were present. These gaps were mainly associated with sensor degradation, equipment problems, physical damage, communication interruptions, and scheduled maintenance. Therefore, appropriate preprocessing was required before the dataset could be used for developing the forecasting models.

3.2 DATA IMPUTATION:

Data preprocessing is an important part of solar power forecasting because the quality of the input data directly influences the reliability of the prediction results. The collected dataset was first examined to identify missing values, incomplete observations, and other inconsistencies that could affect the subsequent modelling process. The missing observations were mainly related to sensor failures, aging measurement equipment, temporary communication problems, and maintenance activities. Rather than removing the records containing missing values, an imputation method was used so that as much of the original information as possible could be retained. This approach also helped maintain the continuity required for time-series forecasting.

A mean-based imputation method was applied separately to each feature containing missing observations. For every variable, the mean was calculated using only the available non-missing values. The calculated mean was then used to replace the missing entries corresponding to that particular feature. The same procedure was followed for all variables that contained incomplete observations. Mean imputation was selected because it provides a straightforward way of estimating missing continuous values while retaining the overall structure of the dataset. It also avoids unnecessary loss of records, which is particularly important when working with a long-term time-series dataset. The following method was used to determine each feature's median value expressed in equation 1,

$$X'_{imp} = \frac{1}{N} \sum_{j=1}^N x_j \quad (1)$$

Where, X'_{imp} is the imputed value in the input feature, N is the number of non-missing samples in the input feature and x_j is available values in the input feature. After completing the imputation process, the dataset contained continuous observations suitable for further analysis. The resulting data was then used as the input for training and evaluating the solar power forecasting models. This preprocessing step helped provide a consistent dataset while allowing the available operational information to be retained.

3.3 PREPROCESSING RATIONALE AND DATASET PREPARATION

The main purpose of preprocessing was to convert the raw measurements collected from the PV plant and weather data sources into a consistent format suitable for predictive modelling. Since the dataset combines solar power generation with several meteorological variables obtained from different measurement systems, variations in data availability and measurement conditions were expected. Handling the missing observations before model development was necessary to avoid interruptions in the time-series data and to ensure that the forecasting algorithms received complete input sequences. The mean-based imputation method allowed the missing values to be filled without removing the corresponding records from the dataset. This helped preserve the amount of usable data while maintaining a simple and computationally efficient preprocessing procedure. The prepared dataset provides the required foundation for training and evaluating the recurrent neural network models used for solar power forecasting. Proper preparation of the data is therefore an essential step in improving the stability of the models and obtaining dependable prediction results.

3.4 DATA NORMALIZATION

Before developing the models, the input features were normalized to reduce the effect of differences in their numerical ranges. This is important in solar power forecasting because the dataset includes variables such as solar irradiance, temperature, wind speed, humidity, and power generation, which can have considerably different scales. If these differences are not addressed, features with larger numerical values may have a greater influence on the model during training. In this study, Min - Max normalization was used to scale each input feature to a range between 0 and 1. This method maintains the relative pattern of the original data while bringing all features to a common scale. In this study, the min – max normalization techniques is used to scale the variable and the normalized data range indicated as $[0,1]$, using below mentioned Equation 2.

$$X'_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

Where, X'_i is the normalized data sample, X_i is the original value of the input feature variable, $X_{\min}$ and $X_{\max}$ are minimum and maximum value of minimum and maximum value of the input feature variable respectively in the dataset.

3.5 DATASET PARTITIONING AND CROSS-VALIDATION

To evaluate the predictive performance and generalization ability of the developed models, five-fold cross-validation was adopted. This approach allows the model to be tested on different portions of the dataset and reduces the dependence on a single train test split. The complete dataset was divided into five nearly equal parts, referred to as folds. In each iteration, four folds, representing approximately 80% of the data, were used for training, while the remaining fold, representing approximately 20%, was used for validation.

The process was done five times, and in each case, one fold acted as the validation set. The results obtained from all five iterations were then combined to obtain the overall model performance. Using five-fold cross-validation provides a more dependable estimate of model performance than relying on a single data split. It also reduces the possibility that the reported results are strongly influenced by a particular partition of the dataset.

3.6 HYPERPARAMETER OPTIMIZATION

The performance of a deep learning model depends significantly on the selection of suitable architectural and training parameters. Therefore, the main hyperparameters of the developed models were systematically varied and optimized during the model development stage. Both network-related and training-related parameters were considered. The hyperparameters investigated in this study included the number of neurons in the dense layer, optimizer, learning rate, dropout rate, batch size, and number of training epochs. These parameters were selected because they can directly affect the learning process, model complexity, convergence, and ability of the model to generalize to unseen data.

The combination of hyperparameters that produced the best validation performance was selected for further evaluation. The same evaluation procedure was followed for all the investigated architectures so that their performance could be compared under similar training conditions. This provided a fair basis for analyzing the effect of different recurrent architectures and attention mechanisms on solar power forecasting.

4. METHODOLOGY

The proposed methodology involves several stages, including data preprocessing, sequence formation, recurrent neural network development, integration of the attention mechanism, and performance evaluation. Different deep learning architectures were considered to study their ability to capture the temporal patterns present in solar power generation data. The investigated architectures include Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), Gated Recurrent Unit (GRU), and Bidirectional Gated Recurrent Unit (Bi-GRU), along with attention-based versions of the recurrent models.

The main objective is to compare these architectures and examine how effectively they can learn the temporal relationships in solar power data. The study also evaluates whether incorporating an attention mechanism can further improve the forecasting performance of the recurrent neural network models.

4.1 SEQUENCE FORMATION

Solar power generation is inherently a time-dependent process in which the current output is influenced by previous meteorological and generation conditions. Therefore, the preprocessed data were transformed into sequential input-output samples before being supplied to the recurrent neural networks.

A sliding-window approach was employed to construct the input sequences. For a selected time-step length of T , the input sequence can be represented in equation 3,

$$X_t = [x_{t-T+1}, x_{t-T+2}, \dots, x_t] \quad (3)$$

where X_t represents the input sequence, x represents the input feature vector at a particular time step, and T denotes the number of historical observations considered by the model. The corresponding target value is represented in equation 4,

$$y_{t+1} = f(X_t) \quad (4)$$

where y_{t+1} denotes the solar power generation to be predicted at the subsequent time step and $f(\cdot)$ represents the nonlinear mapping learned by the deep learning model. The sequential representation enables the recurrent networks to learn both short-term and long-term temporal relationships present in solar power generation data.

4.2 LONG SHORT-TERM MEMORY (LSTM)

The Long Short-Term Memory (LSTM) network is a specialized recurrent neural network designed to overcome the vanishing-gradient problem associated with conventional RNNs. LSTM uses a memory cell and three gates, namely the forget gate, input gate, and output gate, to regulate the flow of information through the network.

At time step t , the forget gate determines which information from the previous cell state should be retained and its explained in equation 5,

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (5)$$

where f_t represents the forget-gate activation, x_t is the current input, h_{t-1} is the previous hidden state, W_f and U_f are weight matrices, b_f is the bias vector, and σ represents the sigmoid activation function.

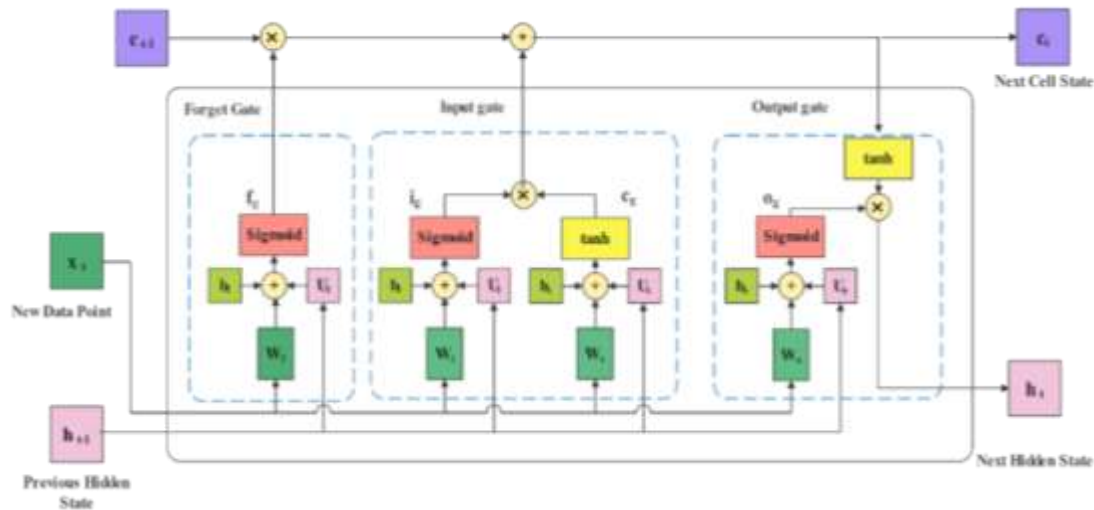


Fig 1. Basic Architecture of Long-Short term Memory

The input gate determines the new information that should be incorporated into the cell state expressed in equation 6,

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (6)$$

The candidate cell state is calculated using below equation 7,

$$C_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (7)$$

The cell state is then updated using below equation 8,

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t \quad (8)$$

where C_t represents the current cell state and $\odot$ denotes element-wise multiplication.

Finally, the output gate determines the information propagated to the next hidden state as in equation 9 and 10,

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (9)$$

$$h_t = o_t \odot \tanh(C_t) \quad (10)$$

Through this gated memory mechanism, the LSTM can retain relevant historical information while selectively discarding less important information, making it suitable for solar power time-series prediction.

4.3 BIDIRECTIONAL LSTM (BI-LSTM)

Although a conventional LSTM processes the sequence in a single temporal direction, a Bidirectional LSTM processes the input sequence in both forward and backward directions. This enables the network to obtain representations from both temporal directions.

The forward LSTM calculates the hidden state using below equation 11,

$$h_t^{\rightarrow} = LSTM(x_t, h_{t-1}^{\rightarrow}) \quad (11)$$

The backward LSTM processes the sequence in the reverse direction calculated using equation 12,

$$h_t^{\leftarrow} = LSTM(x_t, h_{t+1}^{\leftarrow}) \quad (12)$$

The forward and backward representations are then combined and expressed as in equation 13,

$$h_t = [h_t^{\rightarrow}; h_t^{\leftarrow}] \quad (13)$$

where $h_t^{\rightarrow}$ represents the forward hidden state, $h_t^{\leftarrow}$ represents the backward hidden state, and $[\cdot]$ denotes concatenation.

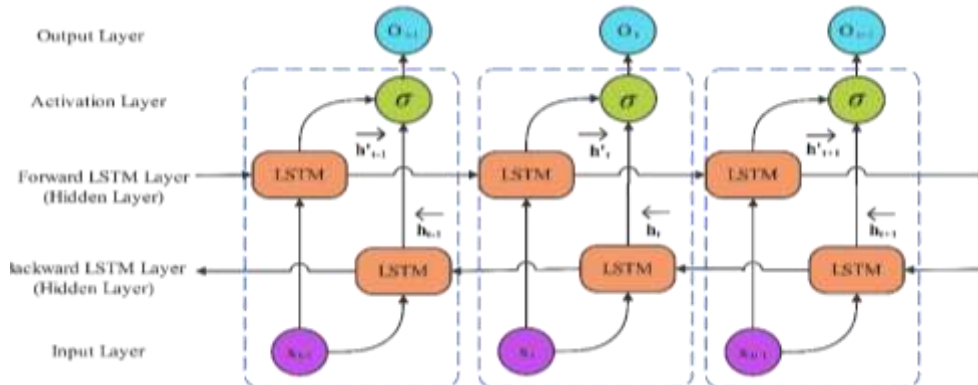


Fig 2. Bi-Directional Long-Short term Memory Architecture.

The combined representation provides the model with information from both directions of the input sequence. Consequently, Bi-LSTM can capture more comprehensive temporal patterns than a unidirectional LSTM, particularly when the relationships within the input sequence are complex and nonlinear. For solar power forecasting, this architecture enables the model to extract richer temporal representations from the available sequence and can therefore provide improved prediction capability under varying environmental conditions.

4.6 SCALED DOT – PRODUCT ATTENTION MECHANISM

Soft dot-product attention helps pick out the most relevant parts of the input sequence. The input features get converted into Query, Key, and Value representations. The attention score comes from multiplying the Query and Key vectors together using a dot product. These scores show how closely each input feature matches the current query. We use a scaling factor to keep numbers stable while training. The scaled scores go through a softmax function to produce smooth, normalized attention weights. The attention weights decide how much each Value vector contributes to the final output. The attention output is calculated by taking a weighted sum of the value vectors. This mechanism lets the model zero in on key time-based patterns while filtering out less important details. Soft dot-product attention helps the model better spot key relationships in sequences. The input vector gets split linearly into three parts: Query (Q_C), which tells you which features deserve attention. Key (K_C) captures how relevant each feature is for the input, while Values (V_C) assign weights to each input element or sequence to compute an average score used in the attention function as shown in Equation 14

$$\text{Attention sources (Q, K, V)} = \text{SoftMax} \left(\frac{Q_C K_C^T}{\sqrt{d}} \right) V_C \quad (14)$$

where, $\sqrt{d_K}$ represent the scaling factor (or) dimensionality of the query and key vector.

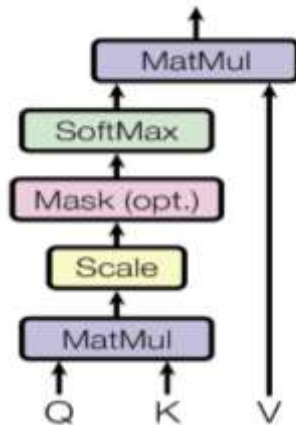


Fig 3. Dot-Product Attention.

5. MODEL PERFORMANCE EVALUATION METRICS:

To assess how effectively a deep learning model can learn from data and manage new, unseen inputs, you need to evaluate its performance. If not properly evaluated, a model may appear accurate but still fail to provide reliable solar power forecasts. To assess the performance of our deep learning model, we employed multiple evaluation metrics: MSE, MAE and R^2 .

The Mean Square Error (MSE) calculates the average of the squared differences between actual and predicted values; a lower MSE indicates better model performance. The formula is presented below equation 15,

$$MSE = \frac{1}{m} \sum_{i=1}^m (z_i - \hat{z}_i)^2 \quad (15)$$

Where, z_i is Actual values, $\hat{z}_i$ is predicted values and m is the number of samples in dataset.

Mean Absolute Error value (MAE) gives the absolute difference value of Actual and predicted value. which, measures the average magnitude of errors between predicted and actual values as represented as below equation 16,

$$MAE = \frac{1}{m} \sum_{i=1}^m |z_i - \hat{z}_i| \quad (16)$$

Where, z_i is Actual values, $\hat{z}_i$ is predicted values and m is the number of samples in dataset. The lower value of MAE indicates the high accurate mode.

R-squared (R^2) indicates how well the model's predictions align with the real data namely, the proportion of the target variable's variability that the model accounts for. An R^2 value closer to 1 indicates how closely the measured and predicted values align and this is shown in the equation 17,

$$R^2 = 1 - \frac{\sum_{i=1}^m (z_i - \hat{z}_i)^2}{\sum_{i=1}^m (z_i - \bar{z})^2} \quad (17)$$

Where, m is number of samples in dataset, z_i is the Actual value, $\hat{z}_i$ is predicted value and $\bar{z}$ is the mean of Actual value.

6.RESULTS AND DISCUSSION

This section presents the experimental setup and the results obtained for the prediction of the solar power generation of a plant. Moreover, the obtained results of the proposed model are compared with the state-of-the-art methodologies in the literature.

6.1 EXPERIMENTAL SETUP

The experiments are run on an 11th Gen Intel i3 dualcore processor with 8GB of main memory at 3.00 GHz speed on Windows 11 operating system. The algorithms used in deep learning are implemented in Python and simulated in Google Colab environment with T4 GPU [21]. The performance of the proposed algorithm is evaluated by 5 fold cross validation. The algorithm uses various Python libraries, including Numpy for mathematical and statistical calculations, pandas for data operations, Scikit-learn for different machine learning operations, preprocessing of data, and others [22]. The model performance evaluation analysed using multiple performance metrics discussed in Table 1

Table 1: Proposed model performance metrics for evaluation

Name of the Model	R-squared Value (kW)	MSE Value (kW)	MAE Value (kW)
Attention + LSTM	0.9555	0.4205	0.4521
Attention+Bi-LSTM	0.9468	0.4890	0.5102
Bi-LSTM	0.9473	0.5097	0.5186
LSTM	0.9223	0.7282	0.6749

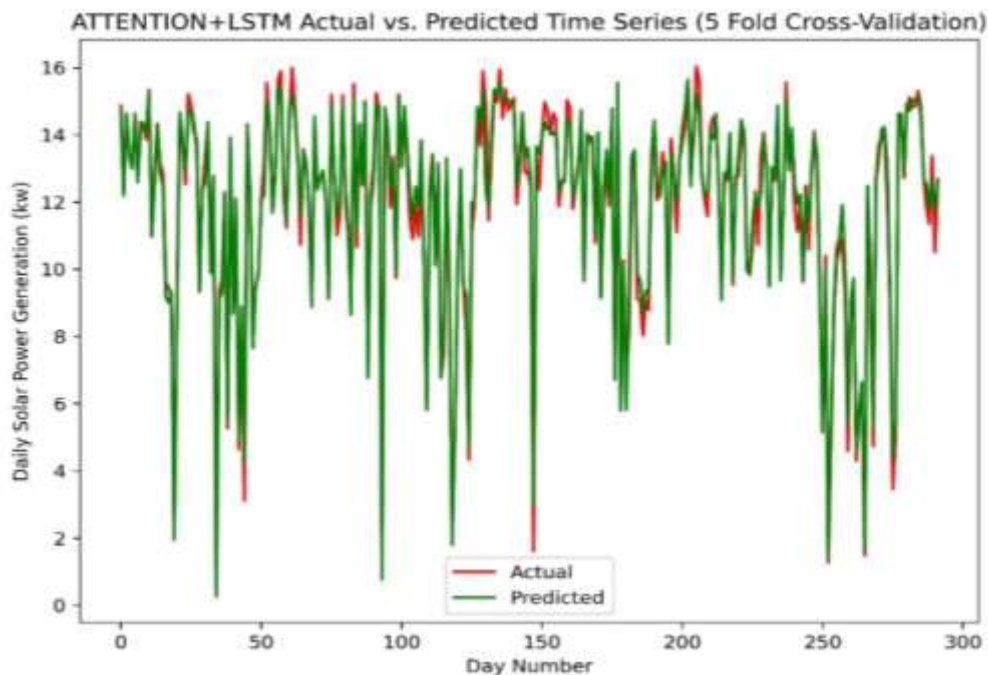


Fig 4. Predicted and Actual values of solar power generation using the proposed LSTM with Attention.

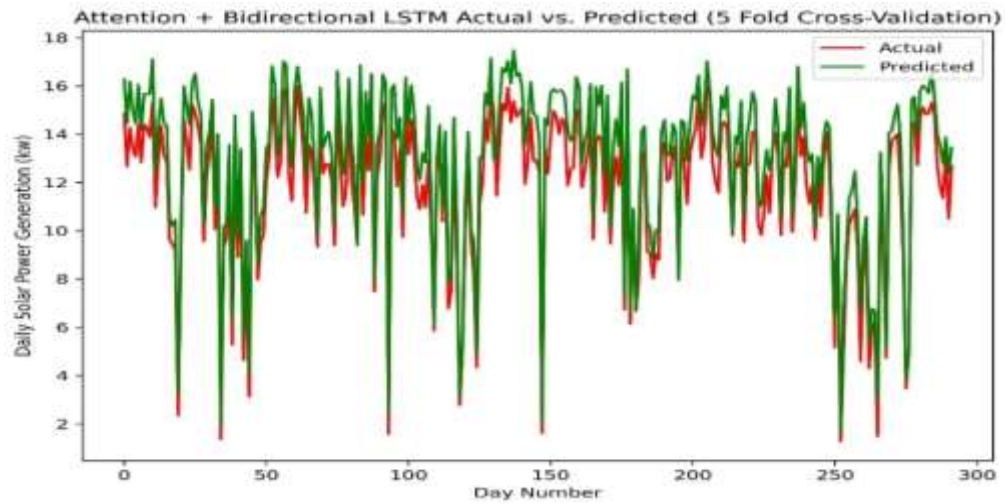


Fig 5. Predicted and Actual values of solar power generation using the proposed Bi-LSTM with Attention.

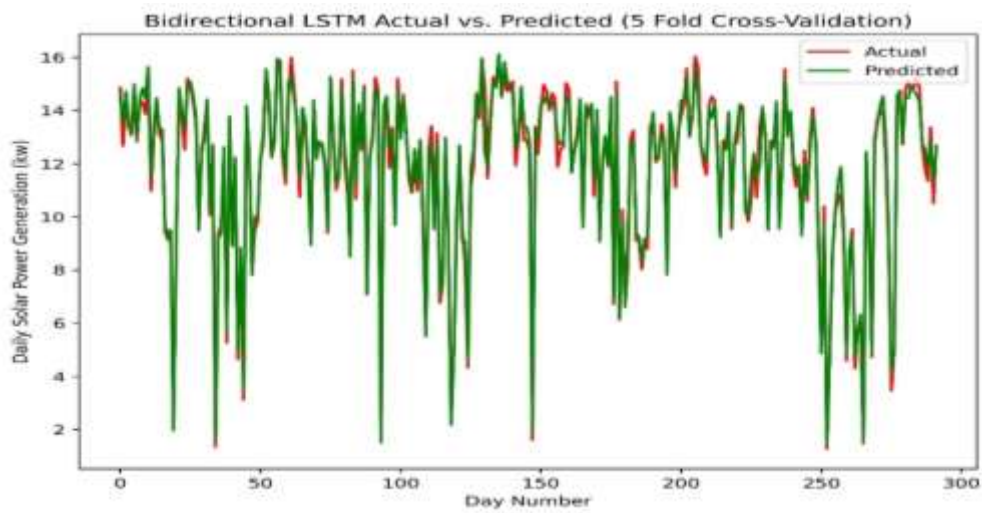


Fig 6. Predicted and Actual values of solar power generation Using Bi-LSTM.

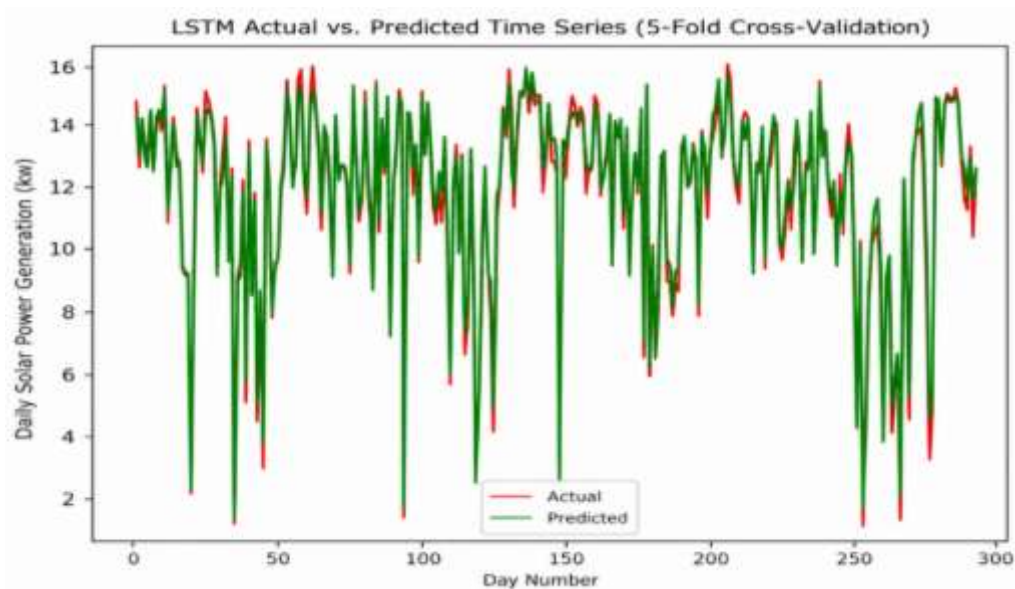


Fig 7. Predicted and Actual values of solar power generation Using LSTM.

6.2 DISCUSSION

The results obtained from the different models show that the attention mechanism does not have the same effect on every recurrent neural network architecture. Noticeable improvements were obtained when attention was combined with the LSTM and Bi-LSTM models, while the changes observed in the bidirectional models were comparatively smaller. This indicates that the usefulness of attention is influenced not only by its ability to focus on important time steps, but also by the way the underlying recurrent network learns and represents the temporal information in the solar power data.

Among all the models considered, the Attention + LSTM architecture produced the most consistent results across the evaluation measures. It achieved the highest R^2 value along with the lowest MSE and MAE, showing its ability to learn the changing patterns present in solar power generation. The combination of LSTM's capability to retain information over longer time periods and the attention mechanism's ability to give more importance to relevant portions of the input sequence appears to be well suited to this forecasting problem. The results from the Attention + Bi-LSTM model provide a different observation. Although this model achieved slightly lower MSE and MAE values, its R^2 value of 0.9468 was marginally lower than the 0.9473 obtained by the conventional Bi-LSTM model. This suggests that adding attention does not automatically result in better performance for every recurrent architecture. The finding may be related to the way bidirectional processing already represents information from both directions of the input sequence. As a result, the additional attention mechanism may provide only a limited advantage for this particular architecture and dataset.

The performance differences among the eight investigated architectures emphasize the need to compare different recurrent structures rather than assuming that a more complex model will always provide better forecasting results. Each architecture processes temporal information in a different way, and its performance can depend strongly on the characteristics of the dataset. In the present study, the combination of LSTM with attention provided a more effective representation of the solar power time series than the other investigated combinations.

Overall, the experimental results identify the Attention + LSTM model as the most suitable architecture among the models evaluated in this study. It achieved a mean R^2 of 0.9555, mean MSE of 0.4205, and mean MAE of 0.4521. These results demonstrate that incorporating attention with LSTM can improve the model's ability to represent the temporal variations associated with solar power generation. At the same time, the limited improvement obtained from Attention + Bi-LSTM shows that further work is required to better represent the highly variable and complex nature of solar power signals. These observations provide a useful starting point for developing more advanced forecasting approaches.

7. CONCLUSION AND FUTURE WORK

This study carried out a comparative evaluation of recurrent neural network models for solar power forecasting, with particular attention given to the integration of a soft attention mechanism with LSTM. LSTM and Bi-LSTM architectures were developed and evaluated both with and without attention using real-world solar power and meteorological data. Based on the experimental results, the Attention + LSTM model provided the best overall performance. The results indicate that combining the memory capability of LSTM with the ability of attention to assign greater importance to relevant information can be effective for learning the temporal patterns associated with solar power generation.

The comparison also revealed that the impact of attention depends on the recurrent architecture in which it is used. The addition of soft attention resulted in clear improvements for the LSTM model, whereas its effect on the Bi-LSTM model was comparatively small. Therefore, attention should not be considered as a technique that automatically improves every forecasting architecture. Its effectiveness is influenced by both the structure of the recurrent network and the characteristics of the time-series data being analysed. These findings can help in selecting suitable recurrent architectures for practical solar power forecasting applications.

Although the models developed in this study produced satisfactory prediction results, solar power generation remains highly variable. Short-term changes in solar irradiance, cloud movement, temperature, and other atmospheric conditions can cause rapid fluctuations in the generated power. Conventional recurrent architectures may not always capture all these variations effectively, particularly when patterns occur at different time scales. To address this limitation, future work will investigate the use of decomposition techniques as part of the forecasting framework. The original solar power signal can be separated into different temporal or frequency components, allowing each component to be modelled individually before the predicted components are combined to obtain the final forecast.

Future research will also examine the use of Multi-Head Attention (MHA) with recurrent neural network architectures. The soft attention mechanism used in the current study provides a single attention representation, whereas MHA can learn several attention representations at the same time. This may allow the model to identify different types of temporal relationships within the input sequence. As part of the proposed future work, models such as Decomposition–LSTM–Multi-Head Attention and related variants will be investigated and compared with the architectures developed in the present study.

The future models will be evaluated using the same dataset and performance measures to provide a consistent comparison with the current results. The combined use of signal decomposition and Multi-Head Attention is expected to provide a more detailed representation of both short-term fluctuations and longer-term trends in solar power generation. Therefore, the next stage of this research will focus on developing a decomposition-assisted and Multi-Head Attention-enhanced LSTM framework with the aim of improving forecasting accuracy, model robustness, and generalization under changing solar and weather conditions. This approach will build upon the findings of the present study and move towards a more reliable deep learning framework for solar power forecasting.

Reference

- [1]. K. Tripathi, M. Aruna, P. V. Elumalai, K. Karthik, S. A. Khan, M. Asif, K. S. Rao, Advancing solar pv panel power prediction: A comparative machine learning approach in fluctuating environmental conditions, *Case Studies in Thermal Engineering* 59 (2024) 104459. doi:10.1016/j.csite.2024.104459.
- [2]. Ministry of New and Renewable Energy (MNRE), Ministry of New and Renewable Energy, Government of India. Available online : https://mnre.gov.in/en/?utm_source=chatgpt.com
- [3]. M. Rupesh, J. Swathi Chandana, A. Aishwarya, C. Anusha and B. Meghana, Prediction of Solar Power Using Machine Learning Algorithm, *Intelligent Data Communication Technologies and Internet of Things*, Springer, 2022, pp. 529–539. doi:10.1007/978-981-16-7610-9_39.
- [4]. Ali Jassim Lari, Antonio P. Sanfilippo, Dunia Bachour and Daniel Perez-Astudillo, Using Machine Learning Algorithms to Forecast Solar Energy Power Output, *Electronics* 14 (2025) 866. doi:10.3390/electronics14050866.
- [5]. Seon Young Jang, Byung Tae Oh and Eunsung Oh, A Deep Learning-Based Solar Power Generation Forecasting Method Applicable to Multiple Sites, doi.org/10.3390/su16125240
- [6]. Ahzam Shadab, Shamshad Ahmad and Saif Said, Spatial forecasting of solar radiation using ARIMA model, *Remote Sensing Applications: Society and Environment* 20 (2020) 100427. doi:10.1016/j.rsase.2020.100427.
- [7]. Vishal Kushwaha and Naran M. Pindoriya, A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast, *Renewable Energy* 140 (2019) 124–139. doi:10.1016/j.renene.2019.03.020.
- [8]. Nishant Saxena, Rahul Kumar, Yarrapragada K. S. S. Rao, Dilbag Singh Mondloe, Nishikant Kishor Dhapekar, Abhishek Sharma and Anil Singh Yadav, Hybrid KNN-SVM machine learning approach for solar power forecasting, *Environmental Challenges* 14 (2024) 100838. doi:10.1016/j.envc.2024.100838.
- [9]. Mohsen Fathi, Sevda Allahyari, Abolfazl Ahmadi, Alireza Zahedi and Sasan Asiaei, Enhanced global solar irradiance and temperature forecasting for rooftop PV power generation prediction using multi-step LSTM and GRU based hybrid models, *Energy Conversion and Management: X* 28 (2025) 101351. doi:10.1016/j.ecmx.2025.101351.
- [10]. Kaleem Ullah, Waseem Akram, Aqib Hassan, Syed Ahmad Saleem Bokhari, Samain Abid, Hamza Yousaf and Adil Farooq, Hybrid CNN-Bi-GRU model with attention mechanism for enhanced short-term load forecasting, *Energy Reports* 14 (2025) 2352–4847. doi:10.1016/j.egyr.2025.09.035.
- [11]. Aml G. AbdElkader, Hanaa ZainElDin and Mahmoud M. Saafan, Optimizing wind power forecasting with RNN-LSTM models through grid search cross-validation, *Sustainable Computing: Informatics and Systems* 45 (2025) 2210–5379. doi:10.1016/j.suscom.2024.101054.
- A. K. Tripathi, M. Aruna, P. V. Elumalai, K. Karthik, S. A. Khan, M. Asif, K. S. Rao and Koppula Srinivas Rao, Advancing solar PV panel power prediction: A comparative machine learning approach in fluctuating environmental conditions, *Case Studies in Thermal Engineering* 59 (2024) 104459. doi:10.1016/j.csite.2024.104459.
- [12]. U. Singh, S. Singh, S. Gupta, M. A. Alotaibi and H. Malik, Forecasting rooftop photovoltaic solar power using machine learning techniques, *Energy Reports* (2023/2024). doi:10.1016/j.egyr.2023.09.018.
- [13]. Hangxia Zhou, Yujin Zhang, Lingfan Yang, Qian Liu, Ke Yan and Yang Du, Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism, *IEEE Access* 7 (2019) 78063–78074. doi:10.1109/ACCESS.2019.2923006.
- [14]. Adel Mellit, Alessandro Massi Pavan and Vanni Lughi, Deep learning neural networks for short-term photovoltaic power forecasting, *Renewable Energy* 172 (2021) 276–288. doi:10.1016/j.renene.2021.02.166.
- [15]. Su-Chang Lim, Jun-Ho Huh, Seok-Hoon Hong, Chul-Young Park and Jong-Chan Kim, Solar power forecasting using CNN-LSTM hybrid model, *Energies* 15 (2022) 8233. doi:10.3390/en15218233.
- [16]. Ying Wang, Bo Feng, Qing-Song Hua and Li Sun, Short-term solar power forecasting: A combined long short-term memory and Gaussian process regression method, *Sustainability* 13 (2021) 3665. doi:10.3390/su13073665.

- [17]. Noor Hasliza Abdul Rahman, Mohamad Zhafran Hussin, Shahril Irwan Sulaiman, Muhammad Asraf Hairuddin and Ezril Hisham Mat Saat, Univariate and multivariate short-term solar power forecasting of 25MWac Pasir Gudang utility-scale photovoltaic system using LSTM approach, *Energy Reports* 9 (2023). doi:10.1016/j.egy.2023.09.018.
- [18]. Reski Khelifi, Mawloud Guermoui, Abdelaziz Rabehi, Ayoub Taallah, Abdelhalim Zoukel, Sherif S. M. Ghoneim, Mohit Bajaj, Kareem M. AboRas and Ievgen Zaitsev, Short-term PV power forecasting using a hybrid TVF-EMD-ELM strategy, *International Transactions on Electrical Energy Systems* (2023) 6413716. doi:10.1155/2023/6413716.
- [19]. Tariq Limouni, Reda Yaagoubi, Khalid Bouziane, Khalid Guissi and El Houssain Baali, Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model, *Renewable Energy* 205 (2023) 1010–1024. doi:10.1016/j.renene.2023.01.118.
- [20]. Bisong, E Google Colaboratory. In: *Building machine learning and deep learning models on google cloud platform.*, Apress, Berkeley, CA. (2019). doi:10.1007/978-1-4842-4470-87.
- [21]. Pedregosa Scikit-learn: Machine learning in Python., *JMLR* 12 (2011). Available online: <http://scikit-learn.sourceforge.net/> Accessed 26 July 2024.