Forecasting Solar Power Generation Utilizing Machine Learning Models in Lubbock

Citation	Balal, A., Jafarabadi, Y.P., Demir, A., Igene, M., Giesselmann, M., & Bayne, S.. 2023. Forecasting Solar Power Generation Utilizing Machine Learning Models in Lubbock. *Emerging Science Journal, 7*(4). https://doi.org/10.28991/ESJ-2023-07-04-02
Citable Link	https://hdl.handle.net/2346/95988
Terms of Use	© 2023 by the authors. Licensee ESJ, Italy. cc-by

Forecasting Solar Power Generation Utilizing Machine Learning Models in Lubbock

Afshin Balal [1]*, Yaser Pakzad Jafarabadi [1], Ayda Demir [1], Morris Igene [2], Michael Giesselmann [1], Stephen Bayne [1]

[1] *Department of Electrical and Computer Engineering, Texas Tech Univeristy, Lubbock, Texas, United States.*

[2] *Department of Civil and Construction Engineering, Texas Tech Univeristy, Lubbock, Texas, United States.*

Abstract

Solar energy is a widely accessible, clean, and sustainable energy source. Solar power harvesting in order to generate electricity on smart grids is essential in light of the present global energy crisis. However, the highly variable nature of solar radiation poses unique challenges for accurately predicting solar photovoltaic (PV) power generation. Factors such as cloud cover, atmospheric conditions, and seasonal variations significantly impact the amount of solar energy available for conversion into electricity. Therefore, it is essential to precisely estimate the output of solar power in order to assess the potential of smart grids. This paper presents a study that utilizes various machine learning models to predict solar photovoltaic (PV) power generation in Lubbock, Texas. Mean Squared Error (MSE) and R^2 metrics are utilized to demonstrate the performance of each model. The results show that the Random Forest Regression (RFR) and Long Short-Term Memory (LSTM) models outperformed the other models, with a MSE of 2.06% and 2.23% and R^2 values of 0.977 and 0.975, respectively. In addition, RFR and LSTM demonstrate their capability to capture the intricate patterns and complex relationships inherent in solar power generation data. The developed machine learning models can aid solar PV investors in streamlining their processes and improving their planning for the production of solar energy.

1- Introduction

The utilization of solar power has become increasingly important in the fight against climate change due to its potential to significantly reduce greenhouse gas emissions. As a result, the solar power industry has grown rapidly to meet the increasing demand for renewable energy [1]. Accurate solar power forecasting is crucial for utility companies and power grid operators to effectively integrate solar power into the grid, plan maintenance and outage schedules, and reduce the need for expensive backup power [2, 3]. To achieve accurate solar power forecasting, machine learning (ML) models have shown great potential in predicting solar PV power generation, and many studies have utilized them for this purpose. These models can capture complex patterns in solar radiation and meteorological data that are difficult for humans to recognize [4–6]. However, the accuracy of the solar prediction depends on the quality and availability of the data. Thus, data preprocessing and feature selection are critical steps in developing ML models for solar forecasting [7]. This study aims to address the debate surrounding the most effective machine learning models for solar power forecasting. The debate revolves around determining which ML models provide the highest accuracy and reliability in predicting solar PV power generation. By evaluating and comparing the performance of different ML models, this

---

research contributes to the understanding of how these models can improve solar power forecasting, thereby benefiting solar energy operators and utility companies in making informed decisions about adopting ML-based solar power forecasting models [8]. Furthermore, the research seeks to enhance the theoretical discussion in the field of ML-based solar power forecasting by providing an in-depth analysis of the various ML algorithms used, the types of data sources employed, and the measures employed to assess their performance. By addressing this debate, the study aims to fill a knowledge gap and contribute to the ongoing advancements in accurate solar power forecasting techniques.

Recently, there have been numerous research studies in machine learning (ML)-based solar power forecasting to evaluate various ML algorithms. In this literature review, an overview of ML-based solar power forecasting is provided, including the different ML algorithms used, the types of data sources employed, and the measures employed to show how well these algorithms performed. In Kim et al. [9], the authors used various time-series techniques, such as deep learning and machine learning algorithms, to forecast the generation of PV power and promptly spot equipment and panel flaws. AI models were created using the South Korean data that was gathered between January 2017 and June 2021. Based on the findings, it was determined that the LSTM model had the greatest level of accuracy for predicting the hourly PV power generation in this 1.5 MW PV system. Elsaraiti & Merabet [10] suggested a deep learning-based approach for short-term PV power generation forecasting. Various performance criteria are used to analyze and contrast the Long Short-Term Memory (LSTM) algorithm and the Multi-Layer Perceptron (MLP) network. The outcomes demonstrate that the LSTM network works better than the MLP network and offers trustworthy data for the effective operation of PV power facilities. Deep learning and energy efficiency work well together to advance sustainability in the electrical industry.

In Li et al.'s [11] study, the author proposes a novel hybrid model for short-term PV energy prediction that combines WPD and LSTM networks. The model was validated using an actual solar system and outperformed other models in different seasons and meteorological conditions. The proposed approach may enhance the operating efficiency of regional power systems. Cabezón et al. [12] explore the implementation of ML techniques for short-term power forecasting in photovoltaic production plants. The study concludes that the tree-based XGB model provided the most accurate results in predicting the next hour's power demand. Bajpai & Duchon [13] present a hybrid approach using machine learning techniques for forecasting solar power generation by PV cells. According to the study, the model based on random forests outperformed other models, and it also suggested a unique method by merging several models according to certain weather circumstances. However, the hybrid approach requires a longer time to train the models. An RNN variant known as a gated recurrent unit (GRU) and LSTM have been employed by Sorkun et al. [14] to assess the effectiveness of the suggested model in terms of accurate solar irradiance prediction. Results show that the GRU and LSTM are more effective in forecasting time-series irradiance than a straightforward RNN.

Miraftabzadeh et al. [15] propose a framework that uses transfer learning to forecast the day ahead of energy production for recently constructed solar power stations. The framework involves training four predictive models based on different network architectures with a dataset, which are then transferred to the second phase to be retrained with the recently constructed PV dataset. The findings demonstrate that the transferred model outperforms other models and has the highest precision. Polasek & Čadík [16] propose a deep-learning (DL) model that predicts solar power plant production by combining elements of UNet with residual aggregation modules, using weather forecasts for training and augmentation. The dataset includes data for power stations throughout the year. By accurately assessing prediction uncertainty, the proposed architecture outperforms the base model by an extra 28.27%. Also, transfer learning is shown to be viable, enabling the provision of forecasts with just a few days ahead. In Mystakidis et al. [17], the author investigates the use of ML and DL algorithms for time-series data for energy generation forecasting. The study employs various metrics to evaluate the models' predicting capabilities and concludes that an ensemble method that integrates multiple ML and DL algorithms is the most accurate approach. The dynamic weighted average ensemble model outperformed all other standalone models, including Random Forest, LSTM, and XGBoost.

The main contribution of this paper is the successful utilization of machine learning techniques to accurately predict solar photovoltaic (PV) power generation in Lubbock, Texas. The study achieved a low Mean Squared Error (MSE) and high R^2 values, indicating the effectiveness of the RFR and LSTM models in capturing complex patterns and relationships in solar power generation data. These findings have practical implications for solar PV investors, providing valuable insights and aiding in the improvement of planning and decision-making processes for producing solar energy. In order to provide a clear and organized analysis, the remainder of this paper is structured into four main sections. Section 2 outlines the methodology and the processing of the data source utilized in this study, while Section 3 delves into a comprehensive analysis of the various machine learning models employed. Section 4 showcases the simulation and results generated from the analysis using Python. Finally, the paper concludes with a detailed discussion in Section 5, highlighting the final thoughts, implications, and future prospects of this study.

2- Research Methodology

The required processing steps, as shown in Figure 1, must be implemented to the dataset in order to estimate solar power with ML/DL models.
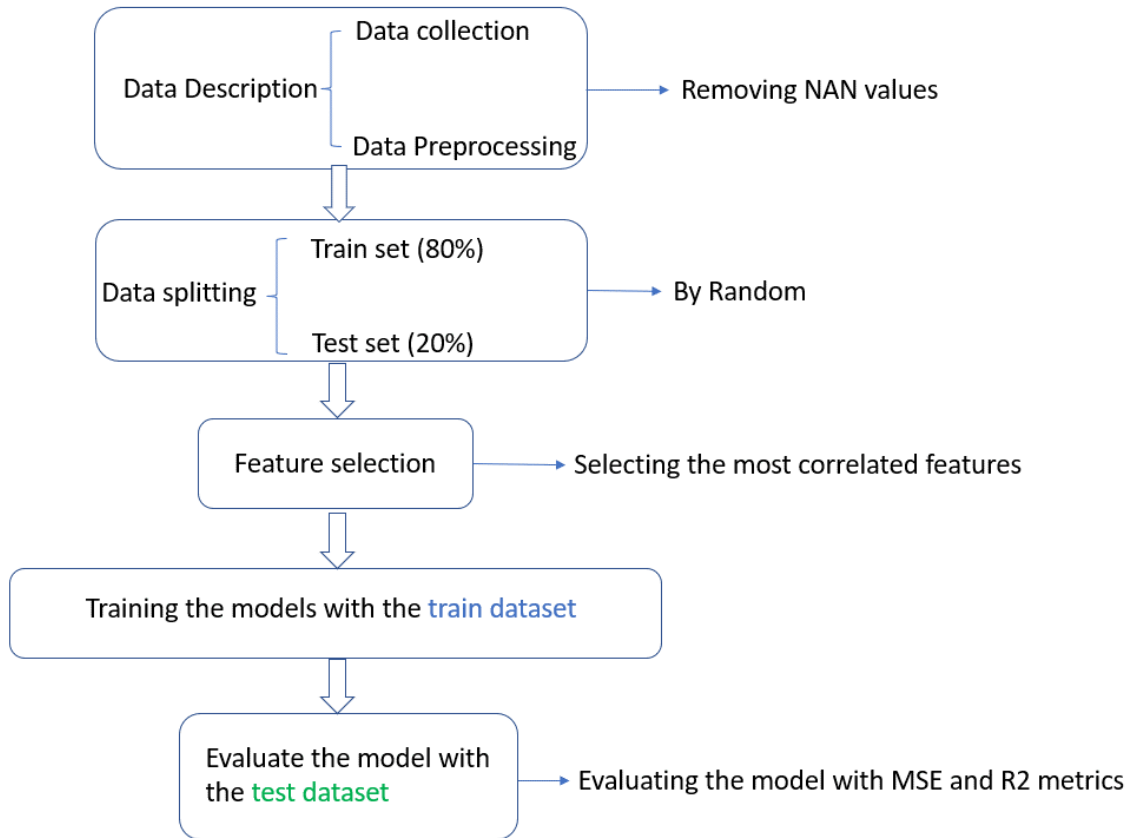
Figure 1. Required steps to predict solar power

Figure 1 summarizes the key procedures that must be followed and illustrates a methodical process that must be used to guarantee precise estimates of solar electricity. These essential actions are described as follows:

2-1- Data Description

This study used the West Texas Mesonet dataset for Lubbock, Texas. The West Texas Mesonet is a network of weather stations that provides high-quality meteorological data for various locations in West Texas. The dataset includes 5-minute measurements of humidity, temperature, wind direction, wind speed, and solar radiation for the years 2012–2022. The dataset was pre-processed and cleaned to remove any missing or erroneous data [18]. The solar radiation data were processed to calculate the solar irradiance on a horizontal surface, which was used as an input variable for the ML models. The West Texas Mesonet dataset is localized and provides more accurate and relevant meteorological data for Lubbock than other US-wide resources.

2-2- Data Preprocessing

Before training the ML models, the West Texas Mesonet dataset was pre-processed and cleaned to remove any missing or erroneous data. This step is crucial to ensuring the accuracy of the results and preventing errors in the models. The preprocessing steps included removing any NAN values, which are missing values in the dataset, and removing negative values for solar power, which is physically impossible [19]. After preprocessing, the dataset has been split into two sets: a training set and a test set. The training set was used to train the ML models, while the testing set was used to evaluate their performance. The training set contained a large portion of the dataset (80%), while the testing set contained a smaller portion (20%) to ensure that the models were not overfitting the data [20].

2-3- Feature Selection

Feature selection can play a crucial role in optimizing solar power generation. Feature selection is the process of determining the most important variables or features that contribute to a given outcome while ignoring irrelevant or redundant ones. Using feature selection techniques, researchers can identify the subset of variables that have the greatest impact on solar power output, thus reducing the dimensionality of the problem and improving the accuracy and efficiency of solar power generation models [21, 22]. This can also help investors reduce costs by focusing on the most influential factors rather than wasting resources on less significant ones. Therefore, feature selection is an important tool for optimizing solar power generation and achieving greater efficiency in renewable energy production [23]. In order to optimize solar power generation, it is important to understand the factors that affect solar power output. This can include

variables such as solar radiation, wind speed, temperature, humidity, and so on. Feature selection for extracting solar power can be done by examining the correlations between these variables and solar power output. Therefore, investors can gain insights into the relative importance of each factor and potentially improve the efficiency of solar power generation [24]. In this context, the correlation coefficients between solar power and the most related variables are shown in Table 1.

Table 1. List of correlation coefficients between solar power and related variables

Variables	Correlation coefficient
Solar power (kW)	1.000000
Solar Radiation (W/m^2)	0.989422
Wind Direction	0.604704
Temperature (C)	0.582534
Wind Speed (m/s)	0.570459
Humidity (%)	-0.517406

The above table shows that solar radiation has the highest positive correlation with solar output, while wind direction, temperature, and wind speed have moderate positive correlations. The humidity has a moderately negative correlation with solar power output. These findings suggest that solar radiation is the most important factor affecting solar power output. The information can be used to inform future research and practical applications in solar power generation [25]. The correlation matrix is depicted in Figure 2.
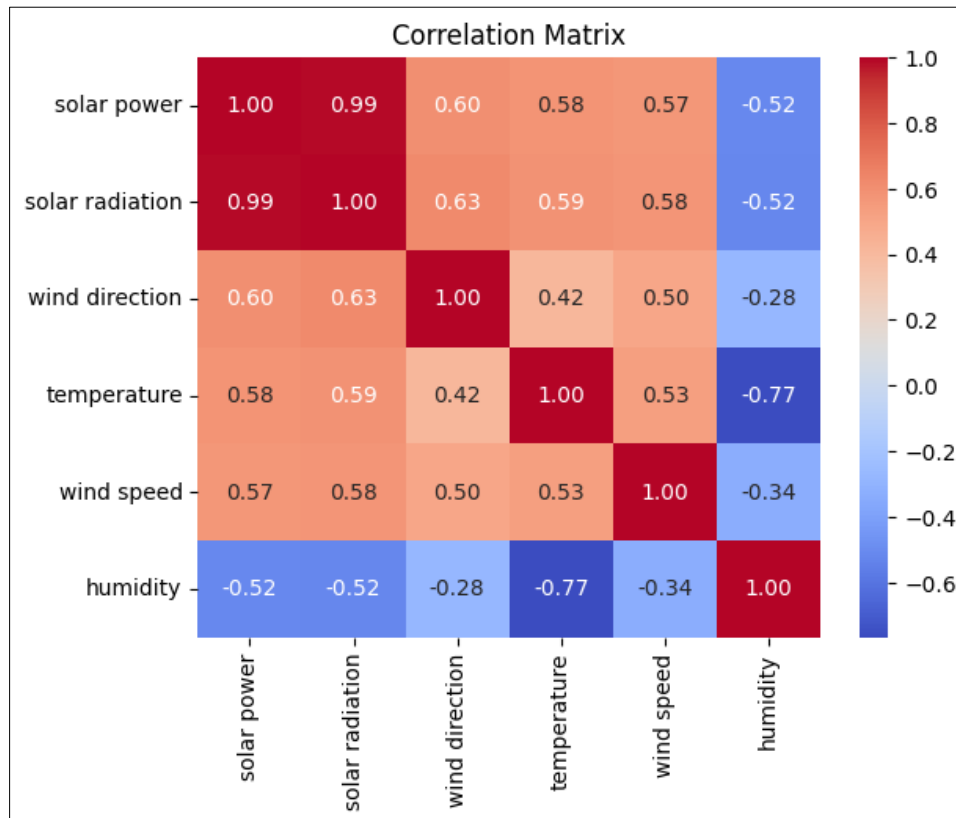

Figure 2. Correlation matrix between solar power and related variables

According to Figure 2, solar radiation has the most effect, and humidity has a low effect on solar power output. Furthermore, the scatter plot of the solar power regarding the correlated variables is depicted in Figure 3.

Based on Figure 3, the scatter plot matrix is utilized to visualize the pairwise relationship between the solar power output and the other weather-related variables like solar radiation, wind direction, temperature, wind speed, and humidity. The diagonal histograms show the distribution of each variable, while the off-diagonal scatter plots show the pairwise relationship between each pair of variables. For example, the scatter plot between solar power output and solar radiation shows a positive linear relationship, suggesting that solar radiation has the highest positive correlation with solar power output. The scatter plot between solar power and humidity shows a negative relationship, suggesting that humidity has a moderately negative correlation with solar power output [26].
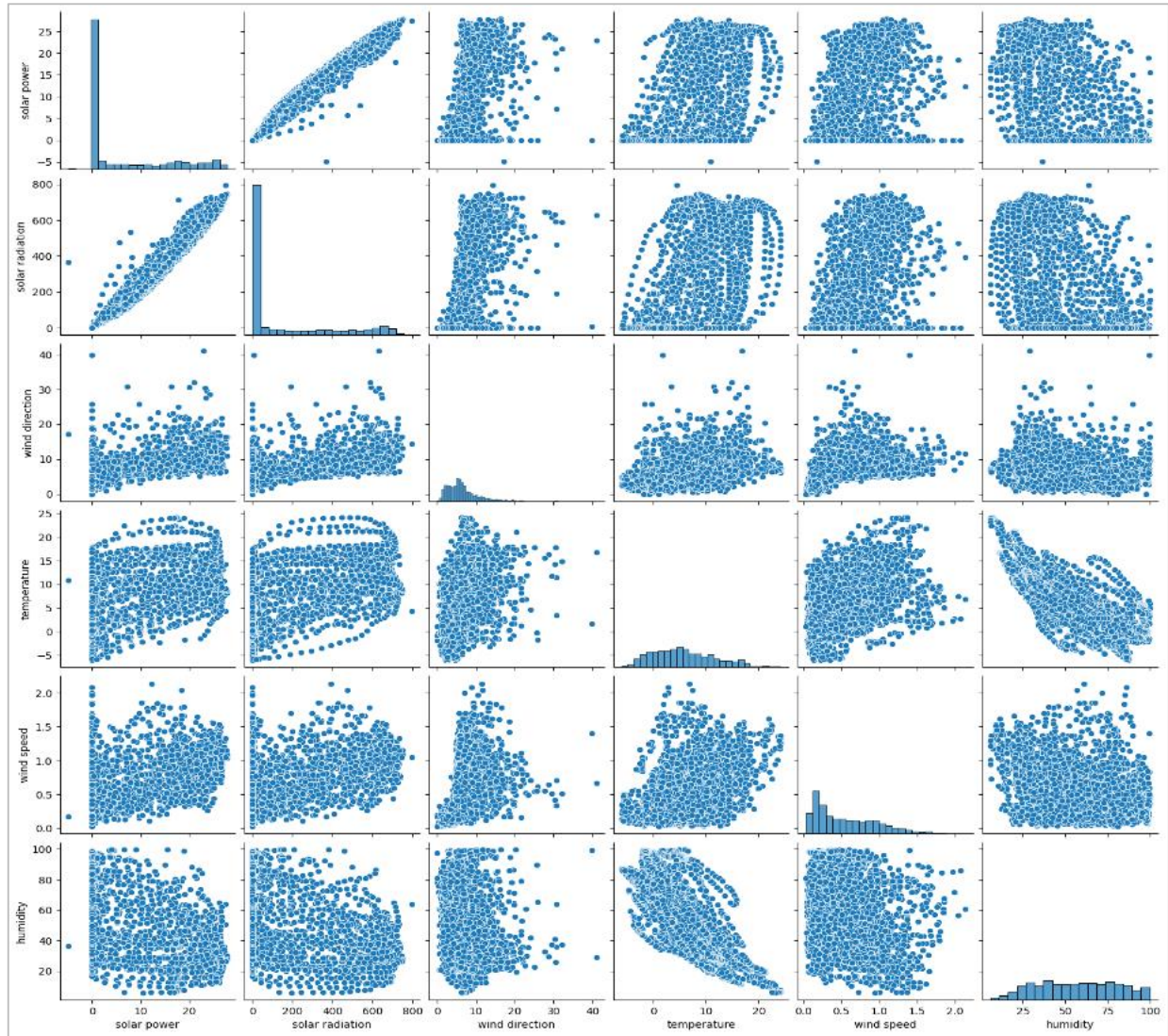
Figure 3. Scatter plot of the correlated weather variables

2-4-Evaluation of the Models

The following evaluation metrics were used to assess the performance of the hybrid RFR-LSTM approach and compare it with standalone RFR, LSTM, and other Machine Learning (ML) and Deep Learning (DL) models.

Mean Squared Error (MSE): MSE measures the average squared difference between the predicted and actual solar power values. It quantifies the overall accuracy of the predictions. The formula for MSE is:

$$\text{MSE} = \frac{1}{n} \times \sum(\text{predicted} - \text{actual})^2 \tag{1}$$

R-squared (R^2) Score: The R^2 score evaluates the proportion of variance in the target variable (solar power generation) that can be explained by the predictions. It indicates the goodness of fit of the model. The formula for the R^2 score is:

$$R^2 = 1 - \frac{\sum(\text{predicted} - \text{actual})^2}{\sum(\text{actual} - \text{mean})^2} \tag{2}$$

These evaluation metrics provide insights into the accuracy and performance of the models in predicting solar power generation. A lower MSE value indicates better accuracy, while a higher R^2 score indicates a better fit of the predictions to the actual data.

3- Machine Learning Models

In this study, eight ML models were used to forecast solar PV power generation: Linear Regression (LR), Polynomial Regression (PR), Decision Tree Regression (DTR), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Random Forest Regression (RFR), Gradient Boosting Regression (GBR), and Long Short-Term Memory (LSTM).

3-1- Linear Regression (LR)

Linear regression (LR) is a statistical approach for modeling the relationship between a target and features by fitting a linear equation to the observed data. In solar power forecasting, LR can be used to model the relationship of weather features and generated solar power. LR is a simple and interpretable model that can provide useful insights into the relationship among a target and some features [27].

3-2- Polynomial Regression (PR)

PR is a type of LR in which the relationship between the target and the features is modeled as an nth-degree function. Polynomial regression is a useful technique when the relationship between the target and the features is non-linear. In solar power forecasting, polynomial regression can be used to identify the nonlinear relationship between weather variables and solar power generation. Polynomial regression can be more flexible than linear regression, but it can also be prone to overfitting, particularly when the degree of the polynomial is high. Overfitting occurs when the model fits the training very closely, and as a result, it may not generalize well to new data. To avoid overfitting, it is important to use cross-validation and regularization techniques when fitting polynomial regression models [28].

3-3- Decision Tree Regression (DTR)

DTR is a non-parametric supervised learning technique implemented for either regression and classification applications. In the context of solar power forecasting, decision trees can be used to model the relationship of weather features and extracted solar power. The basic idea behind decision trees is to recursively partition the data into sub datasets regarding the input feature values. Each subset represents a node of the tree, and the tree is grown until a stopping point, like the lowest number of samples per loaf or the maximum node depth. Decision trees have several advantages, including their simplicity, interpretability, and ability to handle both continuous and categorical input variables. However, decision trees can be prone to overfitting, particularly when the tree is allowed to grow too deep. To avoid overfitting, it is important to use pruning techniques, such as minimum sample split or minimum impurity decrease [29].

3-4- Artificial Neural Network (ANN)

ANN is an ML model inspired by the function of the human brain. In the context of solar power forecasting, ANN can be used to model the relationship of weather features and generated solar output. ANN works by using a set of interconnected nodes (neurons) that are organized in layers. The output layer generates the output forecast after receiving the input data from the input layer. The hidden layers process the input data using a series of mathematical operations that transform the input data into a useful representation for the output layer. ANN can be trained using a variety of techniques, including supervised and unsupervised learning. ANN is capable of discovering patterns and relationships in input datasets autonomously. These networks possess several advantages, including their ability to handle non-linear relationships among input and target variables, as well as their flexibility in modeling complex data. However, training ANN requires careful hyperparameter tuning to achieve optimal performance. One potential challenge is overfitting, where the model fails to generalize well. To address this issue, techniques like regularization and early stopping are essential to prevent overfitting [30].

3-5- Convolutional Neural Network (CNN)

CNN is a DL model that performs well for image and time series data. In solar power forecasting, CNN can be used to identify the spatial and temporal dependencies in the input data, making it a powerful tool for solar power prediction. CNN works by applying convolutional filters to the input data, which extract features that are useful for the prediction task. The output of the convolutional layer is passed through a series of nonlinear activation functions and pooling layers, which help to lower the dimensionality and increase the computational efficiency of the model. The resulting features are then passed through fully connected layers, which produce the final prediction. One of the main advantages of CNN is its ability to automatically learn spatial and temporal patterns in the input data, which makes it well-suited for solar power forecasting. Additionally, CNN can handle large amounts of data and can generalize well to new data, making it a powerful tool for prediction tasks. However, CNN can be computationally expensive to train and requires large amounts of data to achieve good performance. It also requires careful tuning of hyperparameters to achieve optimal performance, and it can be sensitive to changes in the input data, such as changes in weather patterns or solar panel configurations [31].

3-6- Random Forest Regression (RFR)

RFR is an ensemble learning technique that combines multiple decision trees to create a strong learner. In the context of solar power forecasting, RFR can be used to model the relationship between weather features and the extracted solar power. RFR works by training multiple DR, with each tree using a random subset of the input features. The forecasts for the trees are then combined to make a final forecast. This approach can reduce the risk of overfitting and improve the generalization performance of the model. RFR has several advantages, including its ability to handle non-linear relationships among the target and output features, its robustness to outliers and noisy data, and its ability to capture complex interactions between variables. Additionally, RFR can handle missing data and categorical variables without requiring pre-processing. However, RFR can be computationally expensive and requires careful tuning of hyperparameters to achieve good performance. To overcome these challenges, it is important to use techniques such as regularization, early stopping, and cross-validation to prevent overfit [32].
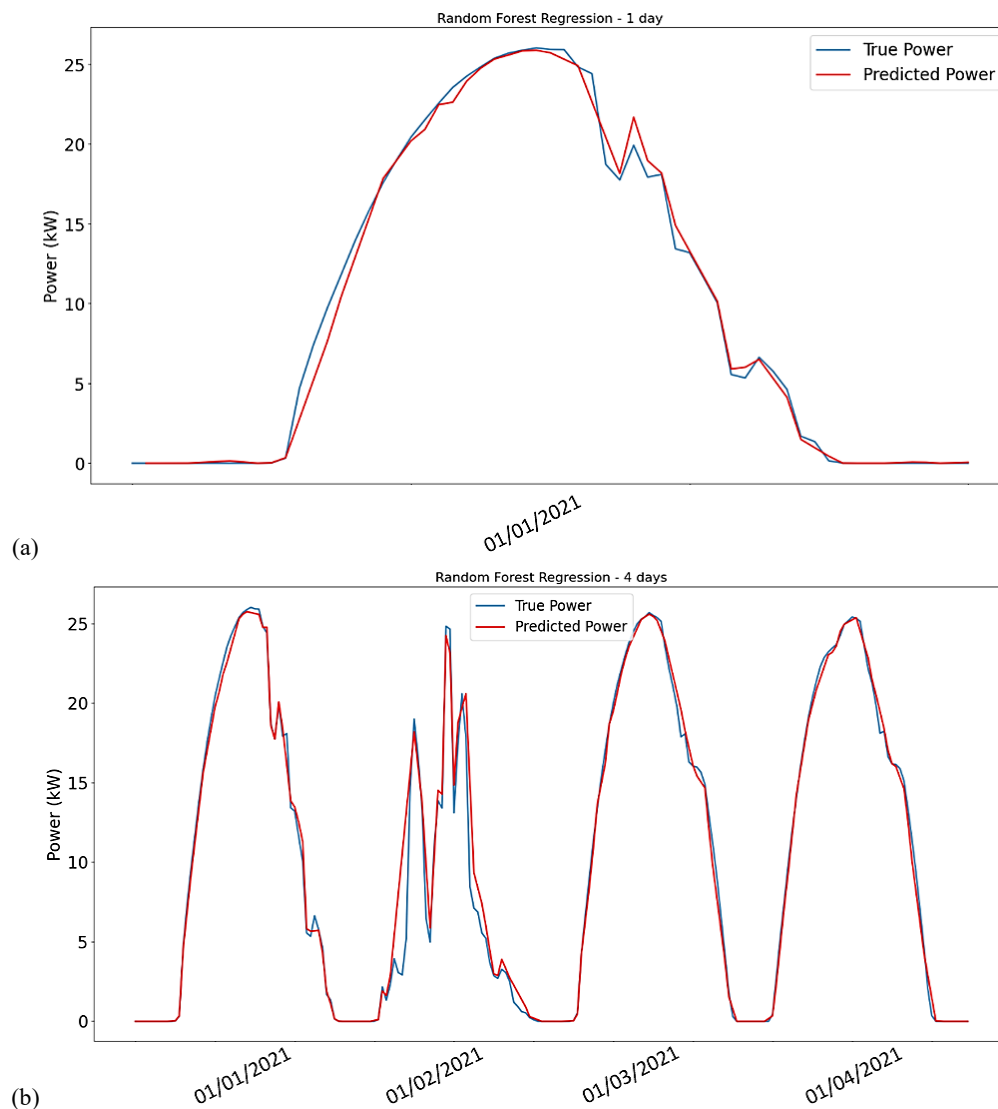
3-7-Gradient Boosting Regression (GBR)

GBR is another ensemble learning technique, and in the context of solar power forecasting, GBR can be used to model the relationship among weather features and generated solar power. GBR works by iteratively training decision trees to correct the errors of the previous trees. Each new tree is fitted to the residual errors of the previous trees, so that the overall model gradually improves in its ability to predict the target variable. GBR has several advantages, including its ability to handle non-linear relationships, its robustness to outliers and noisy data, and its ability to capture complex interactions between variables. However, like RFR, GBR can be computationally expensive and requires careful tuning of hyperparameters to achieve good performance [33].

3-8-Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network (RNN) that performs well for sequential data such as time series. LSTM networks are designed to overcome the vanishing gradient problem. LSTM networks use memory cells and gating mechanisms to selectively recall or forget details from earlier phases, allowing them to capture long-term dependencies more effectively. LSTM networks have several advantages, including their potential to handle input sequences of variable length and to learn complex relationships. In addition, it is important to use techniques such as dropout, early stopping, and regularization to prevent overfit problems [34].

4- Simulation Results

In this study, the performance of eight machine learning models for solar power forecasting was evaluated using two commonly used metrics for evaluating the performance of ML models: Mean Squared Error (MSE) and R-squared (R^2). MSE calculates the mean squared variation in the expected and true values. It is commonly used to evaluate the accuracy of regression models. A smaller MSE indicates that the model is better at predicting the outcome variable [35]. R^2 is a metric that shows how much of the variance in the target is accounted for by other variables. It gauges how well a regression model fits the data. R^2 has a 0 to 1 scale, with 1 denoting a perfect fit [36]. Figure 4 shows the performance of the Random Forest Model for one day, four days, ten days, and one month.
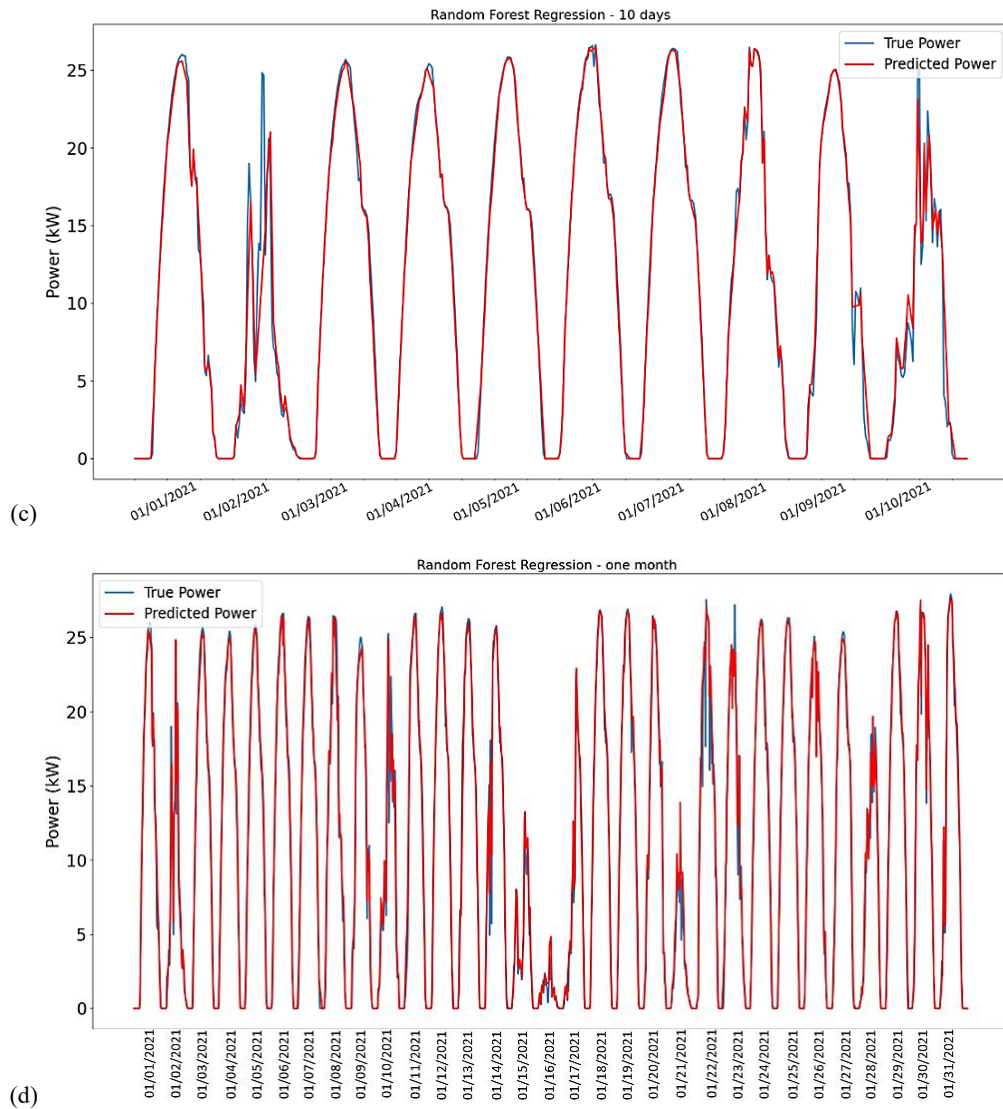

(a)

(b)

(c)

(d)

Figure 4. Random Forest Regression Model; (a) One day; (b) Four days; (c) Ten days; (d) One month

Figure 4 shows the performance of a Random Forest Regression Model used for forecasting solar power for different times, ranging from one day to one month. Time is shown on the x-axis, while actual and expected solar power output are shown on the y-axis. The blue line represents the true solar power output by the model, while the red line represents the predicted solar power output observed. The model's performance can be assessed by looking at the overall pattern and trend in the graph. If the red line consistently aligns closely with the blue line, it suggests the model performs well in accurately forecasting solar power output over different timeframes. On the other hand, significant disparities between the two lines indicate that the model's predictions may not be reliable enough. Figure 3 shows that the predicted solar power output (red line) generally follows the trend of the true solar power output (blue line). This indicates that the Random Forest Regression Model is able to capture the underlying patterns and variations in the solar power output to some extent. Furthermore, Table 2 shows the performance of each model.

Table 2. Performance of the ML models on the test dataset

ML Model	MSE	R^2
LR	3.25	0.9555
PR	2.41	0.9732
DTR	3.98	0.9569
ANN	2.35	0.9712
CNN	2.30	0.9741
RFR	2.06	0.9778
GBR	2.29	0.9756
LSTM	2.23	0.9760

According to Table 2, the ML/DL models chosen for the paper come from different categories, including traditional regression models, ensemble models, and neural network models. The results show that all eight ML models were able to predict solar PV power generation with good accuracy, as evidenced by the high R^2 values and low MSE values. The RFR and LSTM models performed well, achieving lower MSE and higher R^2 values than others. The MSE for RFR is 2.06, which is the lowest, and the R^2 value is 0.97%, which shows very high accuracy. Lower MSE indicates less deviation between predicted and actual values, and higher R^2 values indicate a better fit to the data compared to the other models. Overall, the results suggest that ML models, particularly ensemble methods such as RFR and GBR and neural network models such as LSTM, CNN, and ANN, are well-suited for forecasting solar PV power generation. Ensemble methods combine multiple ML models to improve predictive accuracy, while neural network models, including LSTM, CNN, and ANN, are designed to capture complex patterns and relationships in data.

5- Conclusion

Predicting solar energy generation is a challenging task due to the intermittent nature of weather conditions. Factors such as cloudy days, variations in wind, and the time of year can significantly impact the performance of solar panels and the amount of energy they generate. These dynamic elements introduce uncertainty and make accurate predictions more complex. Therefore, accurate ML/DL models are crucial for predicting solar power with precision. This paper has demonstrated the effectiveness of using various machine learning models to predict PV generation in Lubbock, Texas. The results indicate that the RFR and LSTM models performed better among other models, with the lowest MSE of 2.06% and 2.23%, respectively, and the highest R^2 values of 0.977 and 0.975, indicating their potential for helping solar energy operators optimize their operations and plan better for solar PV power generation. According to the research results, ensemble approaches like RFR and GBR and neural network models like LSTM, CNN, and ANN are optimal for predicting solar PV power generation. However, it is important to note that the dataset used in this study is limited to a specific location (Lubbock, Texas) and a particular period, and the models may perform differently in other locations or under different weather conditions. Future research endeavors could expand upon this study by investigating the performance of the models under varying weather conditions and in different geographical locations. This would provide valuable insights into the generalizability and robustness of the developed models. Additionally, exploring the impact of data pre-processing techniques on the models' performance would be a worthwhile avenue to pursue. By examining the effectiveness of different data preprocessing methods, researchers could potentially enhance the accuracy and reliability of PV generation predictions.

6- Declarations

6-1- Author Contributions

Conceptualization, A.B., Y.J., A.D., M.I., M.G., and S.B.; methodology, A.B. and A.D.; software, A.B., Y.J., and M.I.; validation, A.B., Y.J., and M.I.; formal analysis, A.B. and Y.J.; investigation, A.B., Y.J., and A.D.; resources, A.B., A.D., and M.I; data curation, A.B., Y.J., and A.D.; writing—original draft preparation, A.B.; writing—review and editing, A.B. and M.G.; visualization, A.B. and Y.J.; supervision, M.G. and S.B.; project administration, A.B. and M.G.; funding acquisition, M.G., and S.B. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6-3- Funding and Acknowledgements

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Not applicable.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Walmsley, M. R. W., Walmsley, T. G., & Atkins, M. J. (2015). Achieving 33% renewable electricity generation by 2020 in California. Energy, 92, 260–269. doi:10.1016/j.energy.2015.05.087.

[2] Ye, H., Yang, B., Han, Y., & Chen, N. (2022). State-Of-The-Art Solar Energy Forecasting Approaches: Critical Potentials and Challenges. Frontiers in Energy Research, 10, 268. doi:10.3389/fenrg.2022.875790.

[3] Balal, A., Abedi, M., & Shahabi, F. (2021). Optimized generated power of a solar PV system using an intelligent tracking technique. International Journal of Power Electronics and Drive Systems, 12(4), 2580. doi:10.11591/ijpeds.v12.i4.pp2580-2592.

[4] Harsito, C., Triyono, T., & Rovianto, E. (2022). Analysis of heat potential in solar panels for thermoelectric generators using ANSYS software. Civil Engineering Journal, 8(7), 1328-1338. doi:10.28991/CEJ-2022-08-07-02.

[5] Demir, A., Gutierrez, L. F., Namin, A. S., & Bayne, S. (2022). Solar Irradiance Prediction Using Transformer-based Machine Learning Models. 2022 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata55660.2022.10020615.

[6] Balal, A., & Giesselmann, M. (2021). Demand Side Management and Economic Analysis Using Battery Storage System (BSS) and Solar Energy. 2021 IEEE 4th International Conference on Power and Energy Applications (ICPEA), Busan, Korea, Republic of. doi:10.1109/icpea52760.2021.9639359.

[7] Solano, E. S., Dehghanian, P., & Affonso, C. M. (2022). Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection. Energies, 15(19), 7049. doi:10.3390/en15197049.

[8] Benti, N. E., Chaka, M. D., & Semie, A. G. (2023). Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects. Sustainability (Switzerland), 15(9), 7087. doi:10.3390/su15097087.

[9] Kim, E. G., Akhtar, M. S., & Yang, O. B. (2023). Designing solar power generation output forecasting methods using time series algorithms. Electric Power Systems Research, 216, 109073. doi:10.1016/j.epsr.2022.109073.

[10] Elsaraiti, M., & Merabet, A. (2022). Solar Power Forecasting Using Deep Learning Techniques. IEEE Access, 10, 31692–31698. doi:10.1109/ACCESS.2022.3160484.

[11] Li, P., Zhou, K., Lu, X., & Yang, S. (2020). A hybrid deep learning model for short-term PV power forecasting. Applied Energy, 259, 114216. doi:10.1016/j.apenergy.2019.114216.

[12] Cabezón, L., Ruiz, L. G. B., Criado-Ramón, D., J. Gago, E., & Pegalajar, M. C. (2022). Photovoltaic Energy Production Forecasting through Machine Learning Methods: A Scottish Solar Farm Case Study. Energies, 15(22), 8732. doi:10.3390/en15228732.

[13] Bajpai, A., & Duchon, M. (2019). A Hybrid Approach of Solar Power Forecasting Using Machine Learning. 2019 3rd International Conference on Smart Grid and Smart Cities (ICSGSC). doi:10.1109/icsgsc.2019.00-10.

[14] Sorkun, M. C., Paoli, C., & Incel, Ö. D. (2017). Time series forecasting on solar irradiation using deep learning. 2017 10th international conference on electrical and electronics engineering (ELECO), 30 November-2 December, 2017, Bursa, Turkey.

[15] Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foiadelli, F. (2023). A Day-Ahead Photovoltaic Power Prediction via Transfer Learning and Deep Neural Networks. Forecasting, 5(1), 213–228. doi:10.3390/forecast5010012.

[16] Polasek, T., & Čadík, M. (2023). Predicting photovoltaic power production using high-uncertainty weather forecasts. Applied Energy, 339, 120989. doi:10.1016/j.apenergy.2023.120989.

[17] Mystakidis, A., Ntozi, E., Afentoulis, K., Koukaras, P., Gkaidatzis, P., Ioannidis, D., Tjortjis, C., & Tzovaras, D. (2023). Energy generation forecasting: elevating performance with machine and deep learning. Computing, 1–23. doi:10.1007/s00607-023-01164-y.

[18] Zhang, X., Li, Y., Lu, S., Hamann, H. F., Hodge, B. M., & Lehman, B. (2019). A Solar Time Based Analog Ensemble Method for Regional Solar Power Forecasting. IEEE Transactions on Sustainable Energy, 10(1), 268–279. doi:10.1109/TSTE.2018.2832634.

[19] Mantri, V., Sharma, N., & Jayaraman, R. (2022). Solar Power Generation Prediction for Better Energy Efficiency using Machine Learning. 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India. doi:10.1109/icces54183.2022.9835749.

[20] Torres, J. F., Troncoso, A., Koprinska, I., Wang, Z., & Martínez-Álvarez, F. (2019). Deep Learning for Big Data Time Series Forecasting Applied to Solar Power. International Joint Conference SOCO'18-CISIS'18-ICEUTE'18. Advances in Intelligent Systems and Computing, 771, Springer, Cham, Switzerland. doi:10.1007/978-3-319-94120-2_12.

[21] Chaibi, M., Benghoulam, E. M., Tarik, L., Berrada, M., & El Hmaidi, A. (2022). Machine Learning Models Based on Random Forest Feature Selection and Bayesian Optimization for Predicting Daily Global Solar Radiation. International Journal of Renewable Energy Development, 11(1), 309–323. doi:10.14710/IJRED.2022.41451.

[22] Balal, A., & Giesselmann, M. (2022). PV to Vehicle, PV to Grid, Vehicle to Grid, and Grid to Vehicle Micro Grid System Using Level Three Charging Station. 2022 IEEE Green Technologies Conference (GreenTech), Houston, TX, USA. doi:10.1109/greentech52845.2022.9772041.

[23] Zhou, Y., Liu, Y., Wang, D., Liu, X., & Wang, Y. (2021). A review on global solar radiation prediction with machine learning models in a comprehensive perspective. Energy Conversion and Management, 235, 113960. doi:10.1016/j.enconman.2021.113960.

[24] Tang, N., Mao, S., Wang, Y., & Nelms, R. M. (2018). Solar Power Generation Forecasting With a LASSO-Based Approach. IEEE Internet of Things Journal, 5(2), 1090–1099. doi:10.1109/JIOT.2018.2812155.

[25] Qadir, Z., Khan, S. I., Khalaji, E., Munawar, H. S., Al-Turjman, F., Mahmud, M. A. P., Kouzani, A. Z., & Le, K. (2021). Predicting the energy output of hybrid PV–wind renewable energy system using feature selection technique for smart grids. Energy Reports, 7, 8465–8475. doi:10.1016/j.egyr.2021.01.018.

[26] Carrera, B., Sim, M. K., & Jung, J. Y. (2020). PVHybNet: A hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data. IET Renewable Power Generation, 14(12), 2192–2201. doi:10.1049/iet-rpg.2018.6174.

[27] Yagli, G. M., Yang, D., & Srinivasan, D. (2019). Automatic hourly solar forecasting using machine learning models. Renewable and Sustainable Energy Reviews, 105, 487–498. doi:10.1016/j.rser.2019.02.006.

[28] Abdul Karim, S. A., & Mahinder Singh, B. S. (2014). Global solar radiation modeling using polynomial fitting. Applied Mathematical Sciences, 8, 367–378. doi:10.12988/ams.2014.39516.

[29] Rahul, Gupta, A., Bansal, A., & Roy, K. (2021). Solar Energy Prediction using Decision Tree Regressor. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). doi:10.1109/iciccs51141.2021.9432322.

[30] Yadav, A. K., & Chandel, S. S. (2014). Solar radiation prediction using Artificial Neural Network techniques: A review. Renewable and Sustainable Energy Reviews, 33, 772–781. doi:10.1016/j.rser.2013.08.055.

[31] Jalali, S. M. J., Khodayar, M., Ahmadian, S., Shafie-khah, M., Khosravi, A., Islam, S. M. S., Nahavandi, S., & Catalao, J. P. S. (2021). A New Ensemble Reinforcement Learning Strategy for Solar Irradiance Forecasting using Deep Optimized Convolutional Neural Network Models. 2021 International Conference on Smart Energy Systems and Technologies (SEST). doi:10.1109/sest50973.2021.9543462.

[32] Khalyasmaa, A., Eroshenko, S. A., Chakravarthy, T. P., Gasi, V. G., Bollu, S. K. Y., Caire, R., Atluri, S. K. R., & Karrolla, S. (2019). Prediction of Solar Power Generation Based on Random Forest Regressor Model. 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). doi:10.1109/sibircon48586.2019.8958063.

[33] Tiwari, S., Sabzehgar, R., & Rasouli, M. (2018). Short Term Solar Irradiance Forecast Using Numerical Weather Prediction (NWP) with Gradient Boost Regression. 2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG). doi:10.1109/pedg.2018.8447751.

[34] Hong, Y. Y., Martinez, J. J. F., & Fajardo, A. C. (2020). Day-Ahead Solar Irradiation Forecasting Utilizing Gramian Angular Field and Convolutional Long Short-Term Memory. IEEE Access, 8, 18741–18753. doi:10.1109/ACCESS.2020.2967900.

[35] Elizabeth Michael, N., Mishra, M., Hasan, S., & Al-Durra, A. (2022). Short-Term Solar Power Predicting Model Based on Multi-Step CNN Stacked LSTM Technique. Energies, 15(6), 2150. doi:10.3390/en15062150.

[36] Jebli, I., Belouadha, F. Z., Kabbaj, M. I., & Tilioua, A. (2021). Prediction of solar energy guided by Pearson correlation using machine learning. Energy, 224, 120109. doi:10.1016/j.energy.2021.120109.