Quest Journals
Journal of Software Engineering and Simulation
Volume 10 ~ Issue 8 (2024) pp: 01-12
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

Forecasting Large Scale Solar Photovoltaic Power Generation Plant in Auchi Polytechnic Using Machine Learning Tools

Shaib Ismail Omade
Department of Statistics
School of Information and Communication Technology
Auchi Polytechnic, Auchi, Edo State, Nigeria
shaibomade@gmail.com

&

Suleiman Ibrahim Abubakar
Department of Agricultural Engineering
School of Engineering
Auchi Polytechnic, Auchi, Edo State, Nigeria

ABSTRACT
Solar energy is an abundant, clean, and renewable energy source, crucial for addressing the current global energy crisis. Efficiently harvesting solar power to generate electricity for smart grids is vital. However, the variability of solar radiation presents significant challenges in accurately forecasting solar photovoltaic (PV) power generation. Elements such as cloud cover, atmospheric conditions, and seasonal changes greatly influence the amount of solar energy available for electricity conversion. Accurate estimation of solar power output is therefore critical to evaluate the potential of smart grids. This study explores the use of various machine learning models to predict solar PV power generation in Lubbock, Texas. Performance is measured using Mean Squared Error (MSE) and R^2 metrics. The findings reveal that the Random Forest Regression (RFR) and Long Short-Term Memory (LSTM) models outperformed the others, achieving MSE values of 2.06% and 2.23%, and R^2 values of 0.977 and 0.975, respectively. These results indicate that RFR and LSTM are highly effective in capturing the complex patterns and relationships in solar power generation data. The developed machine learning models can assist solar PV investors in optimizing their processes and enhancing their planning for solar energy production.
Keywords: *Solar energy, Machine learning, Photovoltaic power generation, Smart grids, Forecasting accuracy*

***R**eceived 20 July, 2024; **R**evised 01 Aug., 2024; **A**ccepted 03 Aug., 2024 © The author(s) 2024.*
P**ublished with open access at **www.questjournas.org

I. INTRODUCTION

Due to its potential to drastically cut greenhouse gas emissions, the use of solar electricity has grown in importance in the fight against climate change. To fulfil the growing demand for renewable energy, the solar power industry has expanded quickly as a result (Walmsley et al., 2015). To successfully integrate solar power into the grid, manage maintenance and outages, and lessen the need for pricey backup power, utility companies and power grid operators depend on accurate solar power forecasts (A. T. Balal et al., 2021; Ye et al., 2022). Machine learning (ML) models have demonstrated significant promise in forecasting solar PV power generation, and numerous research have employed them to get precise solar power forecasting. These algorithms are able to identify intricate patterns in solar radiation and meteorological data that are not readily apparent to the human eye (Demir et al., 2022). However, the availability and quality of the data determine how accurate the solar projection is. Preprocessing data and choosing features are therefore essential phases in creating machine learning models for solar forecasting.

The purpose of this study is to investigate the controversy surrounding the best machine learning models for forecasting solar output. Determining which machine learning models offer the best accuracy and dependability for forecasting solar photovoltaic power generation is the main topic of discussion.

This research advances our understanding of how machine learning (ML) models can enhance solar power forecasting by assessing and contrasting their performances. This will help utility companies and solar energy operators make well-informed decisions regarding the adoption of ML-based models for solar power forecasting (Benti et al., 2023). The study also aims to improve the theoretical discourse in the area of machine learning (ML)-based solar power forecasting by offering a thorough examination of the different ML algorithms utilised, the kinds of data sources utilised, and the metrics applied to evaluate their efficacy. The study intends to close a knowledge gap and further the continuous development of precise solar power forecasting methods by tackling this debate.

In order to assess different machine learning (ML) methods, a number of research papers on ML-based solar power forecasting have been conducted recently. An overview of machine learning (ML)-based solar power forecasting is given in this literature review, together with information on the various ML algorithms that were used, the kinds of data sources that were used, and the metrics that were used to indicate how effectively these algorithms worked. In Kim et al. (2023), the authors forecasted the generation of PV power and quickly identified equipment and panel faults using a variety of time-series methodologies, including deep learning and machine learning algorithms. The South Korean data collected between January 2017 and June 2021 was used to construct AI models. Based on the results, it was concluded that the LSTM model predicted the hourly PV power generation in this 1.5 MW PV system with the highest degree of accuracy. A deep learning-based method was proposed by Elsaraiti & Merabet [10] for forecasting short-term PV power generation. The Multi-Layer Perceptron (MLP) network and the Long Short-Term Memory (LSTM) algorithm are compared and contrasted using a variety of performance metrics. The results show that the LSTM network performs better than the MLP network and provides reliable data for PV power facilities to operate efficiently. Energy efficiency and deep learning combine effectively to promote sustainability in the electrical sector.

The author of Li et al. (2020) paper suggests a novel hybrid model that blends LSTM and WPD networks for short-term PV energy prediction. The model performed better in all seasons and weather conditions and was verified using a real solar system. The suggested method might improve regional power systems' operational effectiveness. The use of ML approaches for short-term power forecasting in photovoltaic producing plants is investigated by Cabezón et al. (2022). According to the study's findings, the tree-based XGB model produced the best accurate predictions for the power demand for the upcoming hour. A hybrid method utilising machine learning techniques is presented by Bajpai & Duchon (2019) for forecasting solar power generation by photovoltaic cells. The study found that the random forest model performed better than other models and offered a novel approach by combining many models in accordance with specific meteorological conditions. But training the models takes longer using the mixed technique. Sorkun et al. (2017) implemented an LSTM and a gated recurrent unit (GRU) variant of RNN to evaluate the performance of the proposed model in terms of precise solar irradiance prediction. The results indicate that an RNN is not as effective as the GRU and LSTM in time-series irradiance forecasting.

A methodology that use transfer learning to predict the day-ahead energy production for freshly built solar power plants is put forth by Miraftabzadeh et al. (2023). Using a dataset, the framework trains four predictive models based on various network topologies. These models are then moved to the second phase and retrained using the newly created PV dataset. The results show that the transferred model has the highest precision and performs better than other models. Polasek & Čadík (2023) provide a deep-learning (DL) model that uses weather forecasts for training and augmentation. The model combines components of UNet with residual aggregation modules to predict the production of solar power plants. Data for power plants from all year round are included in the dataset. Through precise evaluation of prediction uncertainty, the suggested architecture surpasses the baseline model by an additional 28.27%. It is also demonstrated that transfer learning is feasible, allowing for the giving of forecasts with only a few days' notice. The application of ML and DL algorithms for time-series data for energy generation forecasting is examined by the author in Mystakidis et al. (2023). The study uses a number of criteria to assess the prediction skills of the models and comes to the conclusion that the most accurate strategy is an ensemble approach that combines numerous ML and DL algorithms. Including Random Forest, LSTM, and XGBoost, the dynamic weighted average ensemble model fared better than any other solo model.

This paper's primary contribution is the effective application of machine learning algorithms to precisely forecast Lubbock, Texas's solar photovoltaic (PV) power generation. The study's low Mean Squared Error (MSE) and high R2 values show how well the LSTM and RFR models captured intricate patterns and correlations in the data on solar power generation. These results have applications for solar photovoltaic investors, offering insightful information and supporting the enhancement of planning and decision-making procedures for solar energy generation. The rest of this work is divided into four main sections to give a clear and organised analysis. While Section 3 provides a thorough review of the several machine learning models used, Section 2 describes the technique and data source processing used in this study. The simulation and findings from the Python analysis are presented in Section 4. In Section 5, the report wraps up with a thorough discussion that highlights the conclusions, ramifications, and potential future directions of this research.

II. METHODOLOGY

To estimate solar power using ML/DL models, the dataset must be processed via the necessary processes, as indicated in Figure 1.

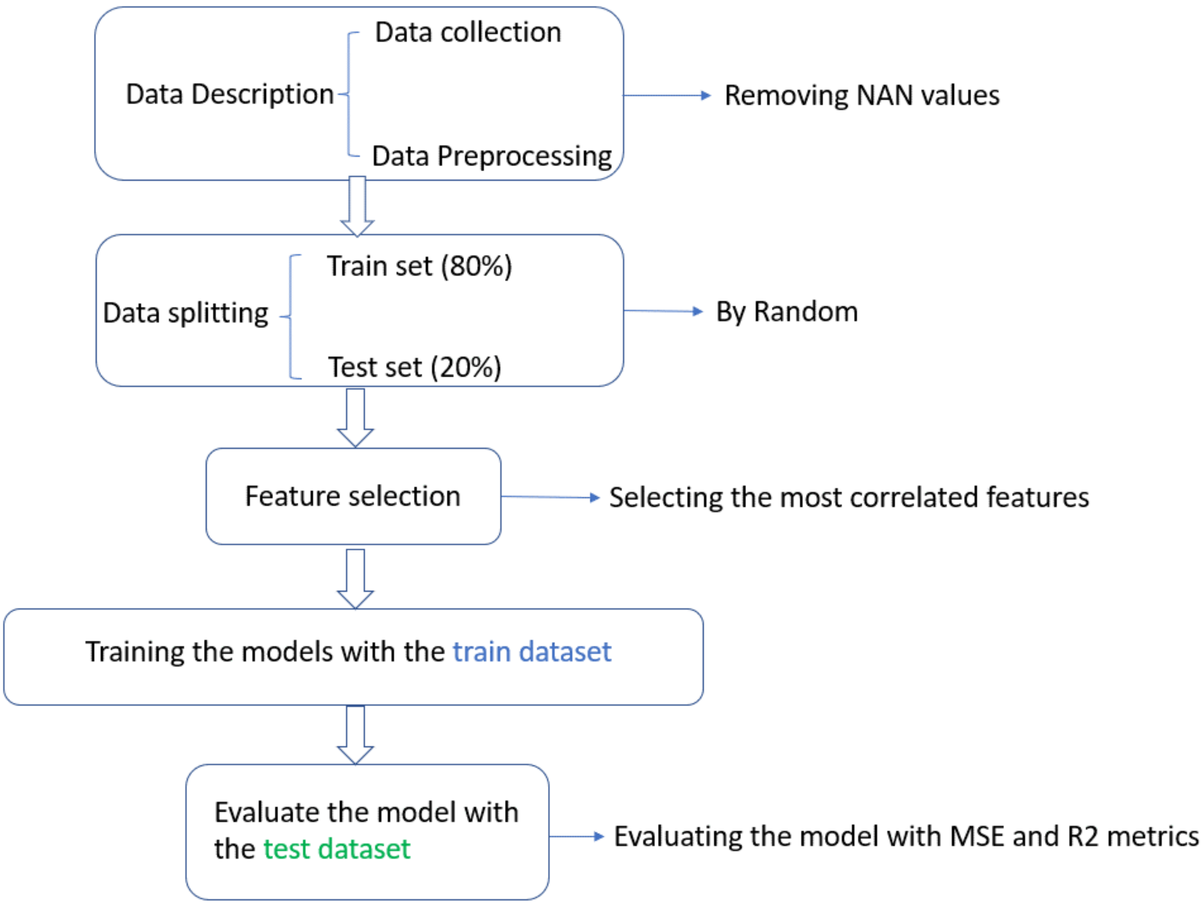


Figure 1: *The necessary actions to forecast solar energy*

In order to ensure accurate estimations of solar electricity, a methodical technique must be employed, as seen in Figure 1, which also summarises the essential procedures that must be followed. The following is an explanation of these crucial actions:

Data Description

The West Texas Mesonet dataset for Lubbock, Texas, was used in this investigation. A network of weather stations called the West Texas Mesonet offers top-notch meteorological data for a number of West Texas sites. For the years 2012–2022, the dataset contains 5-minute measures of temperature, humidity, wind direction, wind speed, and sun radiation. To get rid of any incorrect or missing data, the dataset was cleaned and pre-processed (Zhang et al., 2015). The solar irradiance on a horizontal surface was computed using the solar radiation data, and this value served as an input variable for the machine learning models. Compared to other US-wide resources, the West Texas Mesonet dataset is more relevant and accurate when it comes to meteorological data for Lubbock.

Data Preprocessing

The West Texas Mesonet dataset was cleaned and pre-processed to eliminate any missing or incorrect data before the ML models were trained. This stage is essential for guaranteeing the precision of the findings and avoiding mistakes in the models. As part of the preprocessing stages, any NAN values—missing values in the dataset—and negative values for solar power—which is physically impossible—were eliminated (Mantri et al., 2022). The dataset has been divided into a training set and a test set following preprocessing. The testing set was used to assess the ML models' performance after they had been trained on the training set. To make sure the models weren't overfitting the data, the testing set had 20% of the dataset, and the training set contained 80% of the dataset.

Feature Selection

Choosing the right features can be very important for maximising the production of solar energy. The practice of selecting the most crucial variables or features that influence a certain result while eliminating unnecessary or redundant ones is known as feature selection. Researchers can reduce the complexity of the issue and enhance the precision and effectiveness of solar power generation models by identifying the subset of factors that have the biggest influence on solar power output using feature selection approaches (A. Balal & Giesselmann, 2022; A. T. Balal et al., 2023). By concentrating on the most important variables rather than squandering money on the less important ones, this can also assist investors in cutting expenditures. As a result, feature selection is a crucial instrument for maximising solar power output and increasing the production of renewable energy with greater efficiency (Zhou & Esteban, 2018). It's critical to comprehend the variables influencing solar power output in order to maximise solar power generation. Variables like solar radiation, wind speed, temperature, humidity, and so on can be included in this. By looking at the relationships between these factors and solar power output, features can be chosen for solar power extraction. As a result, investors can learn more about the relative significance of each component and possibly increase the solar power generating efficiency (Tang et al., 2018). Table 1 displays the correlation coefficients in this context between solar power and the most relevant factors.

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

Variables	Correlation coefficient
Solar power (kW)	1.000000
Solar Radiation (W/m2)	0.989422
Wind Direction	0.604704
Temperature (C)	0.582534
Wind Speed (m/s)	0.570459
Humidity (%)	-0.517406

The aforementioned table demonstrates that although wind direction, temperature, and wind speed have moderately positive connections with solar output, solar radiation has the strongest positive link. There is a somewhat inverse relationship between humidity and solar power output. These results imply that the primary factor influencing solar power output is solar radiation. Future studies and useful applications in solar power generation can benefit from the information (Qadir et al., 2021).
In Figure 2, the correlation matrix is shown.

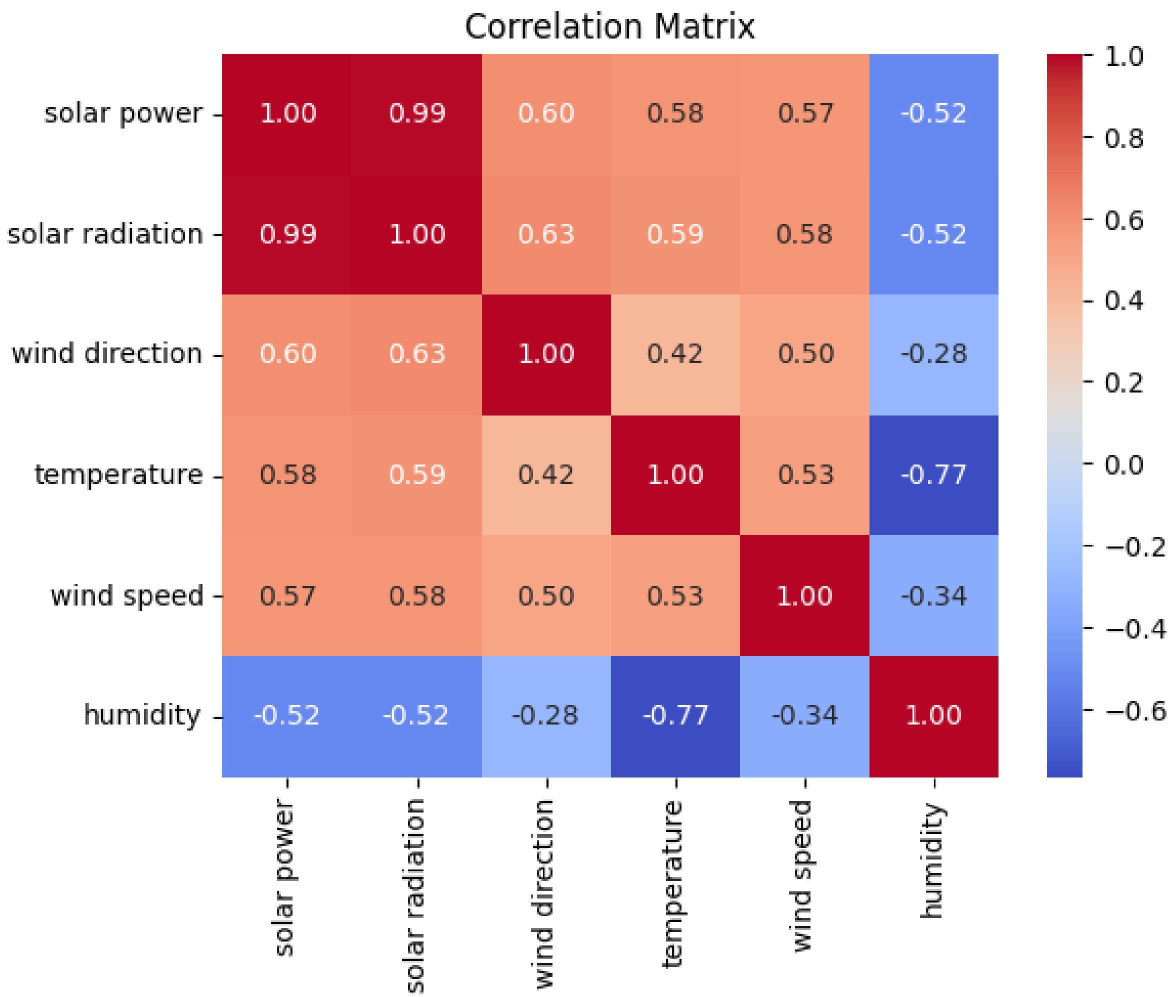


Figure 2: *Solar power and related variables' correlation matrix*

As seen in Figure 2, humidity has little impact on solar power generation, while solar radiation has the greatest effect. In addition, Figure 3 shows the solar power scatter plot with respect to the linked variables.

The pairwise link between the solar power output and the other weather-related variables, such as solar radiation, wind direction, temperature, wind speed, and humidity, is visualised using the scatter plot matrix, which is based on Figure 3. The distribution of each variable is displayed by the diagonal histograms, and the pairwise relationships between each pair of variables are displayed by the off-diagonal scatter plots. For instance, a positive linear relationship may be seen in the scatter plot comparing solar radiation and power output, indicating that solar radiation and solar power output are most positively correlated. According to the scatter plot between solar power and humidity, there is a moderately negative association between humidity and solar power output (Carrera et al., 2020).

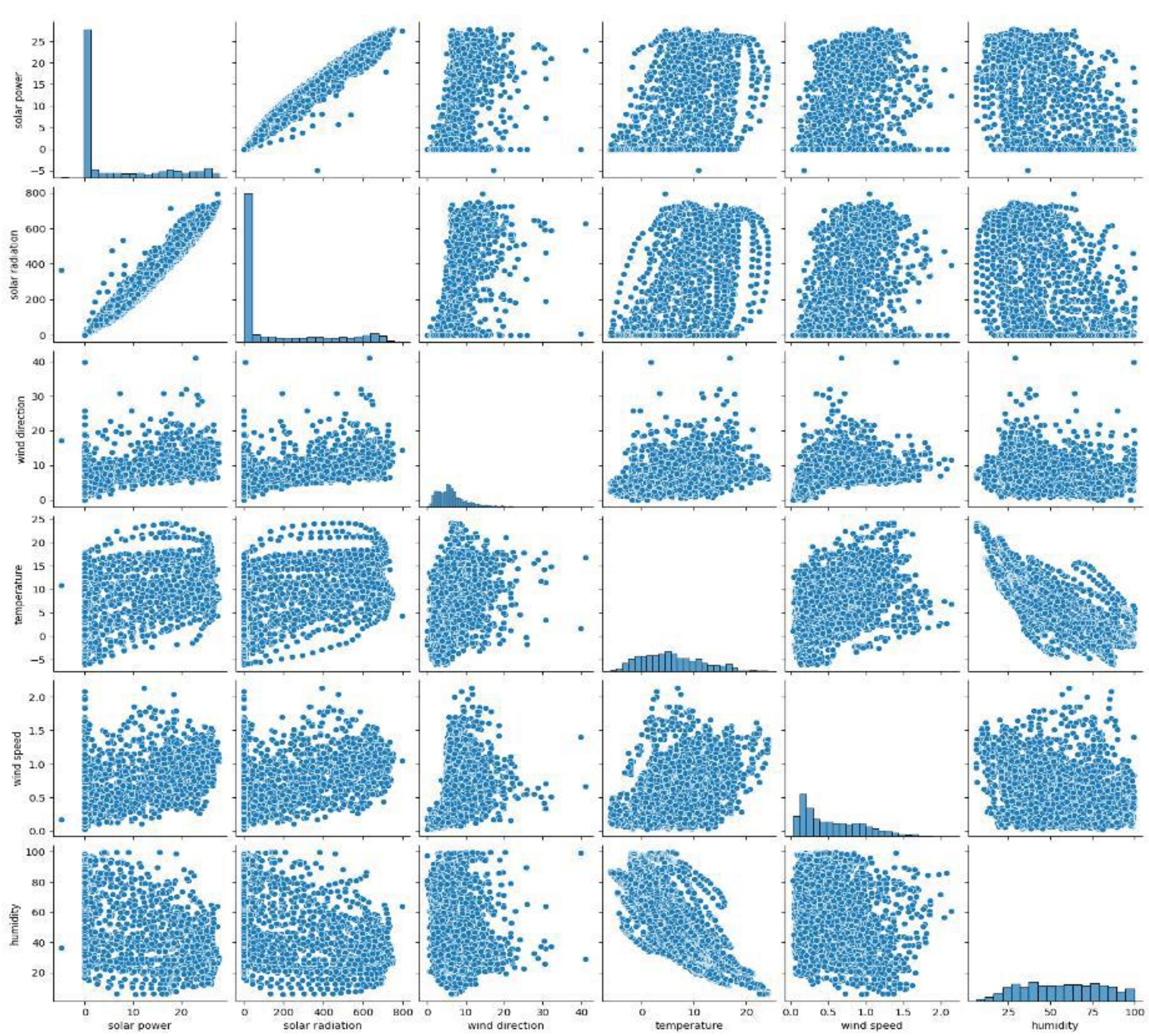


Figure 3. *Scatter plot illustrating the relationship between the weather variables*.

III. EVALUATION OF THE MODELS

The following evaluation metrics were utilized to gauge 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:

1. **Mean Squared Error (MSE):** This metric measures the average squared difference between predicted and actual solar power values, indicating the overall accuracy of the predictions.
2. **R-squared (R^2) Score:** This score evaluates the proportion of variance in the target variable (solar power generation) that is explained by the predictions, indicating the model's goodness of fit.

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

IV. MACHINE LEARNING MODELS

In this study, eight machine learning models were employed 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).

Linear Regression (LR)

A statistical method for simulating the relationship between a target and features is called linear regression (LR), which involves fitting a linear equation to the observed data. LR can be used to predict the link between meteorological characteristics and generated solar power in solar power forecasting. The link between a target and certain features can be better understood by using the straightforward and interpretable LR model (Sharkawy et al., 2023).

Polynomial Regression (PR)

Polynomial regression (PR) is a variant of linear regression (LR) where the relationship between the target variable and the features is represented as an nth-degree polynomial function. This method is beneficial when the relationship between the target and features is non-linear. In solar power forecasting, polynomial regression can help identify the non-linear connections between weather variables and solar power generation. While polynomial regression offers more flexibility compared to linear regression, it is susceptible to overfitting, especially with high-degree polynomials. Overfitting happens when the model closely matches the training data but fails to generalize to new data. To prevent overfitting, it is crucial to employ cross-validation and regularization techniques when fitting polynomial regression models (A. T. Balal et al., 2023).

Decision Tree Regression (DTR)

Decision Tree Regression (DTR) is a non-parametric supervised learning technique used for both regression and classification tasks. In solar power forecasting, decision trees can model the relationship between weather features and solar power output. The fundamental concept of decision trees involves recursively partitioning the data into sub-datasets based on input feature values. Each subset corresponds to a node in the tree, and the tree continues to grow until it reaches a stopping criterion, such as a minimum number of samples per leaf or a maximum node depth. Decision trees offer several benefits, including simplicity, interpretability, and the ability to handle both continuous and categorical input variables. However, they are susceptible to overfitting, especially when the tree grows too deep. To mitigate overfitting, pruning techniques like setting a minimum sample split or a minimum impurity decrease should be employed (Gupta et al., 2021).

Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are machine learning models inspired by the functioning of the human brain. In the context of solar power forecasting, ANNs can model the relationship between weather features and generated solar output. ANNs consist of interconnected nodes (neurons) organized in layers. The input layer receives the input data, which is then processed through hidden layers using a series of mathematical operations to transform the data into a useful representation for the output layer, where the forecast is generated. ANNs can be trained using various techniques, including supervised and unsupervised learning. They are capable of autonomously discovering patterns and relationships in input datasets. ANNs have several advantages, such as handling non-linear relationships between input and target variables and their flexibility in modeling complex data. However, training ANNs requires careful hyperparameter tuning for optimal performance. Overfitting, where the model fails to generalize well, is a potential challenge. To prevent overfitting, techniques like regularization and early stopping are essential (Asghar et al., 2024).

Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are deep learning models that excel with image and time series data. In solar power forecasting, CNNs can identify spatial and temporal dependencies in the input data, making them highly effective for predicting solar power output. CNNs operate by applying convolutional filters to the input data to extract features relevant to the prediction task. The output from the convolutional layers is processed through nonlinear activation functions and pooling layers, which reduce dimensionality and improve computational efficiency. The extracted features are then passed through fully connected layers to generate the final prediction. A key advantage of CNNs is their ability to automatically learn spatial and temporal patterns in the input data, making them well-suited for solar power forecasting. Additionally, CNNs can handle large datasets and generalize well to new data, enhancing their predictive power. However, training CNNs can be computationally intensive and requires substantial amounts of data for optimal performance. Hyperparameter tuning is crucial for achieving the best results, and CNNs can be sensitive to changes in input data, such as variations in weather patterns or solar panel configurations (Jalali et al., 2021).

Random Forest Regression (RFR)

Random Forest Regression (RFR) is an ensemble learning technique that combines multiple decision trees to create a strong predictive model. In solar power forecasting, RFR can model the relationship between weather features and solar power output. RFR operates by training multiple decision trees, each using a random subset of the input features. The predictions from these trees are then combined to produce a final forecast. This method helps to reduce the risk of overfitting and enhances the model's generalization performance. RFR offers several advantages, such as handling non-linear relationships between input and target features, robustness to outliers and noisy data, and the ability to capture complex interactions between variables. It can also manage missing data and categorical variables without extensive preprocessing. However, RFR can be computationally intensive and requires careful hyperparameter tuning for optimal performance. To address these challenges, techniques like regularization, early stopping, and cross-validation are essential to prevent overfitting (Khalyasmaa et al., 2019).

Gradient Boosting Regression (GBR)

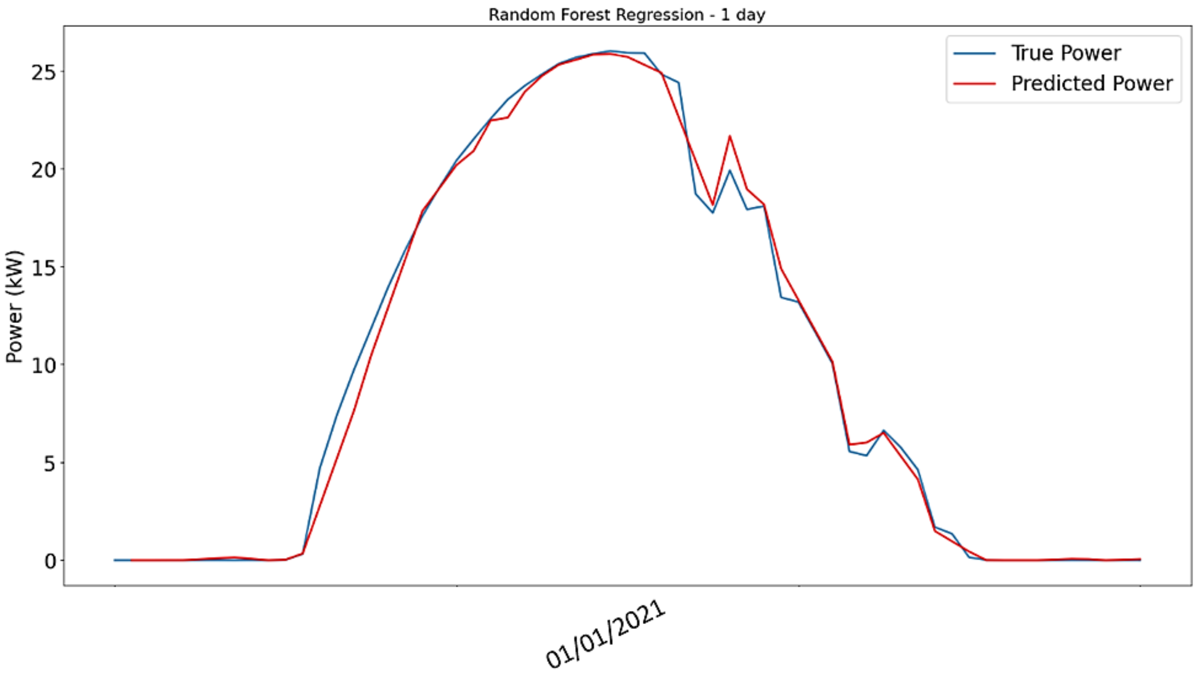
Another ensemble learning method is GBR, which can be used to predict the relationship between weather characteristics and generated solar power in the context of solar power forecasting. In order to fix the mistakes made by the prior trees, GBR trains decision trees iteratively. The residual errors of the previous trees are used to fit each subsequent tree, resulting in a model that progressively gets better at predicting the target variable. The ability of GBR to handle non-linear connections, its resilience to outliers and noisy data, and its capacity to capture intricate interactions between variables are just a few of its benefits. To attain good performance, GBR, like RFR, can be computationally expensive and requires careful hyperparameter adjustment (Tiwari et al., 2018).

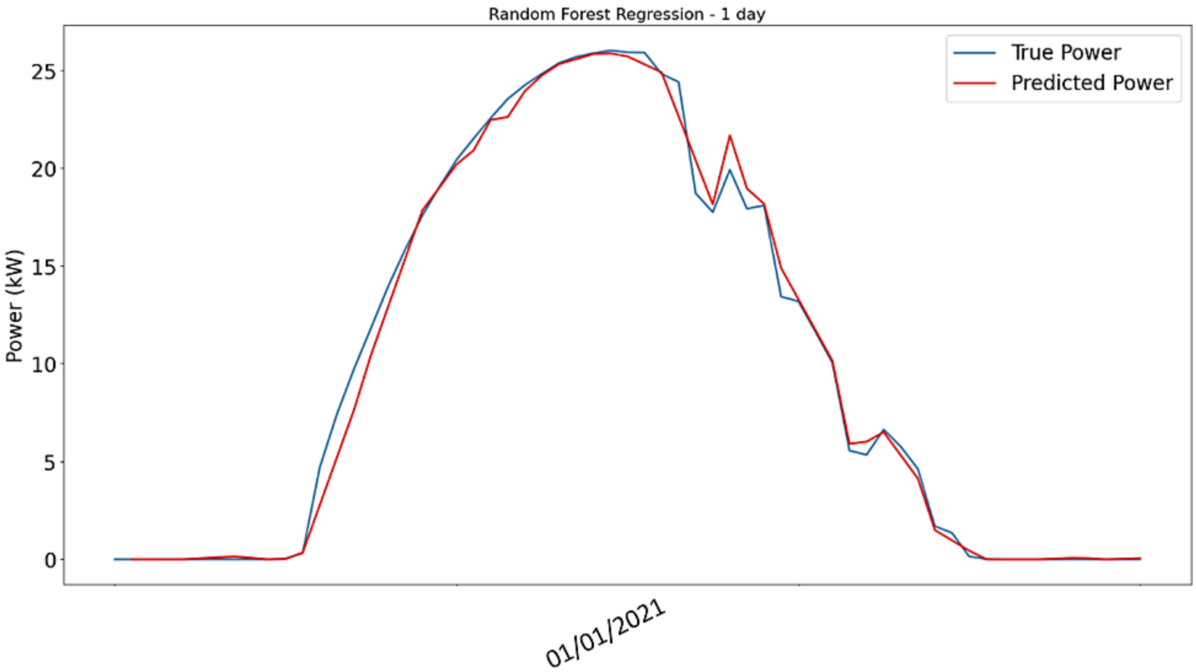
Long Short-Term Memory (LSTM)

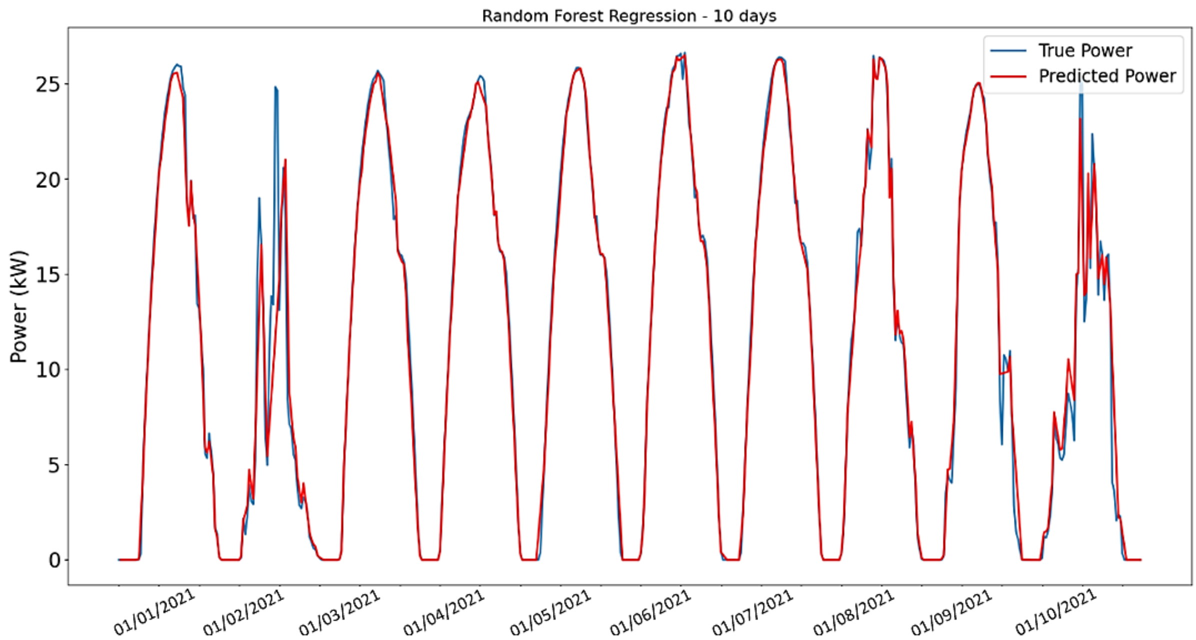
Recurrent neural networks (RNNs) like LSTM operate well with sequential data, such time series. The purpose of LSTM networks is to solve the vanishing gradient issue. Long-term dependencies can be captured more successfully by LSTM networks by using memory cells and gating mechanisms to selectively recall or forget details from earlier phases. LSTM networks offer a number of benefits, such as the ability to process input sequences of varying length and the capacity to understand intricate relationships. Furthermore, to avoid overfit issues, it's critical to employ strategies like regularisation, early halting, and dropout (A. T. Balal et al., 2023).

V. RESULTS

In this study, the performance of eight machine learning models for solar power forecasting was evaluated using two commonly used metrics: Mean Squared Error (MSE) and R-squared (R^2). MSE measures the average squared difference between the predicted and actual values, and it is commonly used to assess the accuracy of regression models. A smaller MSE indicates better predictive accuracy. R^2 measures the proportion of variance in the target variable explained by the model, indicating how well the regression model fits the data. It ranges from 0 to 1, with 1 indicating a perfect fit. Figure 4 illustrates the performance of the Random Forest Model over different time periods: one day, four days, ten days, and one month.




Random Forest Regression - 1 day
True Power
Predicted Power
Power (kW)
0
5
10
15
20
25
01/01/2021


Random Forest Regression - 10 days
True Power
Predicted Power
Power (kW)
0
5
10
15
20
25
01/01/2021
01/02/2021
01/03/2021
01/04/2021
01/05/2021
01/06/2021
01/07/2021
01/08/2021
01/09/2021
01/10/2021

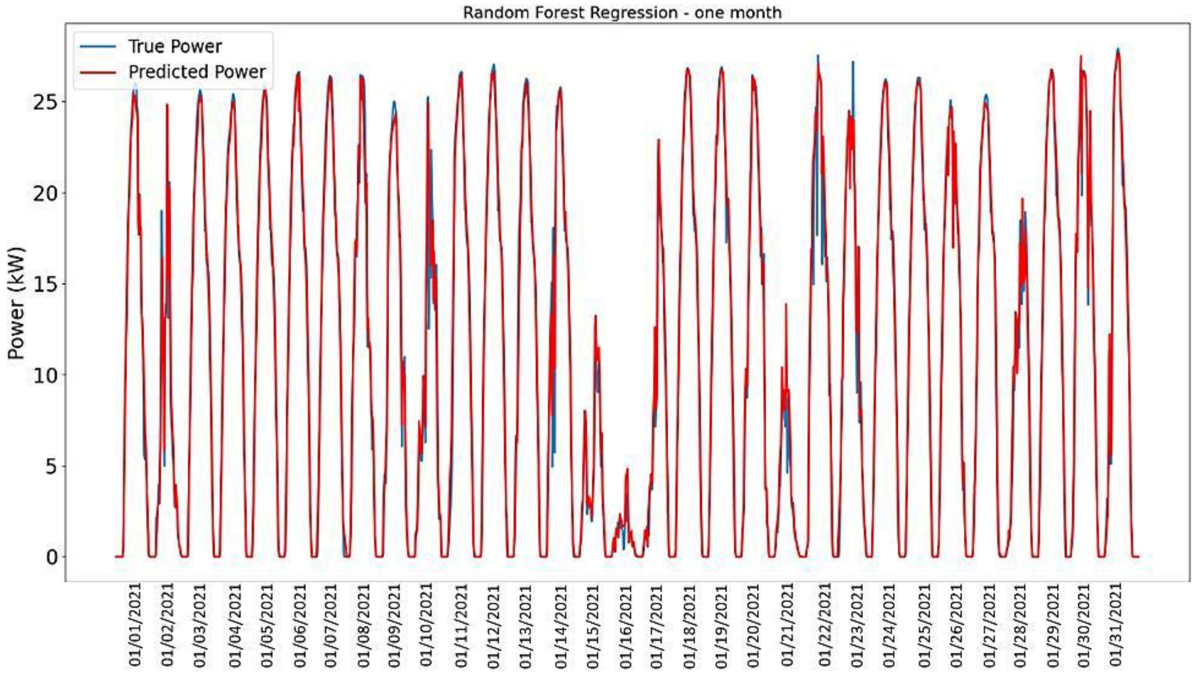


Figure 4. *Random Forest Regression Model; (a) One day; (b) Four days; (c) Ten days; (d) One month*

Figure 4 illustrates the performance of a Random Forest Regression Model for forecasting solar power over various time periods, ranging from one day to one month. The x-axis represents time, while the y-axis displays both actual and predicted solar power output. The blue line denotes the true solar power output, whereas the red line represents the model's predicted output. The model's effectiveness can be evaluated by observing the alignment between the red and blue lines. A close alignment suggests accurate forecasting, while significant discrepancies indicate less reliable predictions. Figure 3 shows that the predicted solar power output (red line) generally aligns with the true solar power output (blue line), suggesting that the Random Forest Regression Model captures the underlying patterns and variations in solar power output to some extent. Additionally, Table 2 presents the performance metrics of each model used in the study.

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 selected ML/DL models for the study include traditional regression models, ensemble models, and neural network models. The results indicate that all eight ML models achieved good accuracy in predicting solar PV power generation, as reflected by their high R^2 values and low MSE values. The Random Forest Regression (RFR) and Long Short-Term Memory (LSTM) models performed exceptionally well, with RFR achieving the lowest MSE of 2.06 and an R^2 value of 0.97, indicating very high accuracy. Lower MSE values suggest minimal deviation between predicted and actual values, while higher R^2 values signify a better fit to the data. Overall, the findings suggest that ML models, especially ensemble methods like RFR and Gradient Boosting Regression (GBR), as well as neural network models like LSTM, CNN, and ANN, are well-suited for forecasting solar PV power generation. Ensemble methods enhance predictive accuracy by combining multiple ML models, whereas neural network models are adept at capturing complex patterns and relationships in data.

VI. CONCLUSION

Predicting solar energy generation is challenging due to the intermittent nature of weather conditions. Factors like cloudy days, wind variations, and the time of year significantly affect the performance of solar panels and the amount of energy they generate. These dynamic elements introduce uncertainty, making accurate predictions more complex. Therefore, precise ML/DL models are essential for accurate solar power forecasting. This paper has demonstrated the effectiveness of various machine learning models in predicting PV generation in Lubbock, Texas.

The results show that the Random Forest Regression (RFR) and Long Short-Term Memory (LSTM) models outperformed others, achieving the lowest Mean Squared Error (MSE) values of 2.06% and 2.23%, respectively, and the highest R^2 values of 0.977 and 0.975. These findings suggest that these models have significant potential to help solar energy operators optimize operations and plan for solar PV power generation. According to the research, ensemble approaches like RFR and Gradient Boosting Regression (GBR), as well as neural network models like LSTM, Convolutional Neural Network (CNN), and Artificial Neural Network (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. The models might perform differently in other locations or under different weather conditions. Future research could expand upon this study by investigating the models' performance under varying weather conditions and in different geographical locations, providing valuable insights into their generalizability and robustness. Additionally, exploring the impact of data preprocessing techniques on the models' performance could enhance the accuracy and reliability of PV generation predictions, making it a worthwhile avenue for further research.

REFERENCES

[1]. Asghar, R., Fulginei, F. R., Quercio, M., & Mahrouch, A. (2024). Artificial Neural Networks for Photovoltaic Power Forecasting: A Review of Five Promising Models. IEEE Access. https://ieeexplore.ieee.org/abstract/document/10577259/

[2]. 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), 108–113. https://ieeexplore.ieee.org/abstract/document/8906549/

[3]. 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), 25–30. https://ieeexplore.ieee.org/abstract/document/9772041/

[4]. Balal, A. T., Abedi, M., & Shahabi, F. (2021). Optimized generated power of a solar PV system using an intelligent tracking technique. https://ttu-ir.tdl.org/items/45f3fdb5-94c1-4a04-b2e4-7965782d3f9a

[5]. Balal, A. T., Jafarabadi, Y. P. T., Demir, A. T., Igene, M. T., Giesselmann, M. T., & Bayne, S. T. (2023). Forecasting solar power generation utilizing machine learning models in Lubbock. https://ttu-ir.tdl.org/items/44138bc3-6a20-499c-b839-710f1b36293d

[6]. 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, 15(9), 7087.

[7]. 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.

[8]. Carrera, B., Sim, M., & Jung, J. (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. https://doi.org/10.1049/iet-rpg.2018.6174

[9]. Demir, A., Gutiérrez, 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), 2833–2840. https://ieeexplore.ieee.org/abstract/document/10020615/

[10]. 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), 489–495. https://ieeexplore.ieee.org/abstract/document/9432322/

[11]. Jalali, S. M. J., Khodayar, M., Ahmadian, S., Shafie-Khah, M., Khosravi, A., Islam, S. M. S., Nahavandi, S., & Catalão, J. P. (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), 1–6. https://ieeexplore.ieee.org/abstract/document/9543462/

[12]. 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), 0780–0785. https://ieeexplore.ieee.org/abstract/document/8958063/

[13]. Kim, E., Akhtar, M. S., & Yang, O.-B. (2023). Designing solar power generation output forecasting methods using time series algorithms. Electric Power Systems Research, 216, 109073.

[14]. Li, P., Zhou, K., Lu, X., & Yang, S. (2020). A hybrid deep learning model for short-term PV power forecasting. Applied Energy, 259, 114216.

[15]. 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), 1185–1191. https://ieeexplore.ieee.org/abstract/document/9835749/

[16]. 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.

[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, 105(8), 1623–1645. https://doi.org/10.1007/s00607-023-01164-y

[18]. Polasek, T., & Čadík, M. (2023). Predicting photovoltaic power production using high-uncertainty weather forecasts. Applied Energy, 339, 120989.

[19]. Qadir, Z., Khan, S. I., Khalaji, E., Munawar, H. S., Al-Turjman, F., Mahmud, M. 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.

[20]. Sharkawy, A.-N., Ali, M. M., Mousa, H. H., Ali, A. S., Abdel-Jaber, G. T., Hussein, H. S., Farrag, M., & Ismeil, M. A. (2023). Solar PV power estimation and upscaling forecast using different artificial neural networks types: Assessment, validation, and comparison. IEEE Access, 11, 19279–19300.

[21]. 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), 151–155. https://ieeexplore.ieee.org/abstract/document/8266215/

[22]. 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.

[23]. 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), 1–8. https://ieeexplore.ieee.org/abstract/document/8447751/

[24]. Walmsley, M. R., Walmsley, T. G., & Atkins, M. J. (2015). Achieving 33% renewable electricity generation by 2020 in California. Energy, 92, 260–269.

[25]. 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, 875790.

[26]. Zhang, D., Miao, X., Liu, L., Zhang, Y., & Liu, K. (2015). Research on development strategy for smart grid big data. Proceedings of the CSEE, 35(1), 2–12.

[27]. Zhou, W., & Esteban, M. (2018). Beyond balancing: China’s approach towards the Belt and Road Initiative. Journal of Contemporary China, 27(112), Article 112.