International Journal of Research and Innovation in Multidisciplinary (www.ijrim.com)
||Volume 3, Issue 8, August 2026 || Impact Factor: 5.2 |

An Efficient Machine Learning Technique for Solar Irradiance Forecasting

[1]Dharmjeet Kumar, [2]Mr. Sandeep Kumar Wasnik
[1]*M.Tech Scholar, School of Electrical and Electronics Engineering, SCOPE Global Skills University, Bhopal, India,*
[2]*Assistant Professor, School of Electrical and Electronics Engineering, SCOPE Global Skills University, Bhopal, India*

***Abstract*— The increasing use of solar photovoltaic (PV) systems in electrical power networks requires accurate solar irradiance forecasting for reliable and efficient power generation. This paper presents an efficient Random Forest-based Machine Learning technique for solar irradiance forecasting using historical irradiance and relevant meteorological parameters. The proposed model captures the nonlinear relationship between environmental conditions and solar irradiance and provides accurate prediction of available solar energy. The performance of the proposed technique is compared with a previous Deep Neural Network (DNN) method using accuracy and error rate as evaluation parameters. The Random Forest model achieves 99.67% accuracy with an error rate of 0.33%, while the previous DNN method achieves 94.70% accuracy with a 5.30% error rate. The results demonstrate that the proposed technique significantly reduces forecasting error and improves the reliability of solar irradiance prediction. Accurate forecasting can support PV power generation estimation, generation scheduling, power balancing, battery energy storage operation, voltage management, and renewable-energy integration. Overall, the proposed Random Forest technique provides an efficient and accurate solution for reducing uncertainty in solar PV generation and supporting reliable operation of modern electrical power systems.**

Keywords: Solar Irradiance, Machine Learning, Solar Forecasting, PV System, Renewable Energy.

I. INTRODUCTION

The rapid expansion of solar photovoltaic (PV) generation has made solar energy an important source of electrical power in modern power systems. However, PV generation is directly dependent on solar irradiance, which varies continuously with time and atmospheric conditions. Accurate solar irradiance forecasting is therefore essential for estimating future PV power output and maintaining reliable and economical operation of renewable-integrated electrical networks [1].

Solar irradiance directly determines the amount of electrical power that can be generated by a PV array. Variations in irradiance produce corresponding fluctuations in PV output, which can affect generation scheduling, power balancing, voltage profile, and network operating conditions. Accurate forecasting provides advance information about expected solar generation and helps system operators prepare suitable operational strategies [2].

The increasing penetration of distributed PV systems in distribution networks has further increased the importance of irradiance forecasting. Large variations in solar generation can create uncertainty in the generation-load balance and may require additional reserve power or energy-storage support. Reliable forecasting can reduce this uncertainty and improve coordination between solar generation and conventional generating resources [3].

Solar irradiance is influenced by several parameters, including solar position, temperature, humidity, cloud conditions, atmospheric characteristics, and seasonal variations. The nonlinear relationship between these parameters makes irradiance prediction a challenging electrical-energy forecasting problem. Conventional forecasting approaches may not always provide satisfactory results when irradiance changes rapidly due to changing atmospheric conditions [4].

Machine Learning (ML) techniques provide an alternative approach for developing accurate solar irradiance forecasting models. ML algorithms can learn relationships between historical irradiance measurements and relevant meteorological variables directly from available data. Techniques such as Artificial Neural Networks, Support Vector Regression, Decision Trees, Random Forest, and ensemble learning have been investigated for renewable-energy forecasting applications [5].

Among different ML approaches, Random Forest (RF) is particularly suitable for forecasting applications because it combines multiple decision trees to generate a reliable prediction. The ensemble structure can handle nonlinear relationships between input variables and solar irradiance and can provide stable prediction performance for varying operating conditions. Its capability to process multiple electrical and meteorological parameters makes it useful for solar-energy forecasting [6].

Accurate irradiance forecasting is also important for battery energy storage systems (BESS) connected with PV plants. Forecast information can be used to determine expected solar generation and plan battery charging and discharging operations. During periods of high expected solar generation, available storage capacity can be managed appropriately, while stored energy can support the electrical network during periods of low solar generation [7].

Solar irradiance forecasting also contributes to improved power quality and distribution-system operation. Rapid changes in PV generation can cause voltage variations and reverse power-flow conditions in distribution feeders with high PV penetration[8]. Forecasting future irradiance and corresponding PV output can assist in scheduling control actions and improving the coordination of PV inverters, storage systems, and other grid-supporting equipment [9].

Considering these requirements, this research proposes an efficient Random Forest-based technique for solar irradiance forecasting with a clear focus on solar PV and electrical power-system applications[10].

II. METHODOLOGY

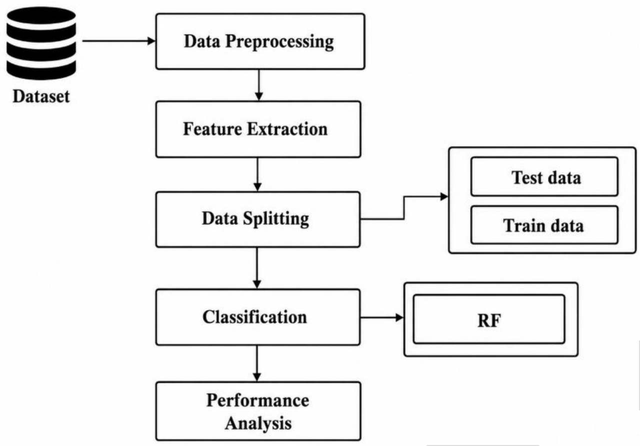


Figure 1: Flow chart

The proposed methodology follows a sequential process consisting of dataset collection, data preprocessing, feature extraction, data splitting, Random Forest-based forecasting, and performance analysis. The complete process is explained step-by-step below.

Step 1. Dataset Collection

The first step is to select a suitable solar irradiance dataset for developing the forecasting model. The dataset contains historical solar irradiance measurements along with relevant meteorological parameters such as temperature, humidity, wind speed, and other available environmental variables. Since solar irradiance directly affects PV power generation, historical irradiance values are considered the main input for forecasting future solar availability. The collected data provides the basis for learning the relationship between environmental conditions and solar irradiance.

Step 2. Data Preprocessing

The raw dataset is processed before applying the forecasting model. First, the data is checked for missing values, null values, duplicate records, abnormal observations, and measurement noise. Missing or invalid observations are handled appropriately to improve data reliability. The data may then be normalized to bring different input parameters into a comparable numerical range. Min-Max normalization can be expressed as:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where X represents the original value, while X_{min} and X_{max} represent the minimum and maximum values of the corresponding parameter. Proper preprocessing ensures that the forecasting model receives consistent and reliable input data.

Step 3. Feature Extraction

In this step, the important variables influencing solar irradiance are extracted from the preprocessed dataset. Important features may include previous solar irradiance, temperature, humidity, wind speed, time of day, month, solar position, and other available weather parameters. Historical irradiance values are particularly important because they provide information about the temporal behavior of solar radiation. Feature extraction reduces unnecessary information and allows the Random Forest model to focus on parameters that have a significant relationship with solar irradiance.

Step 4. Data Splitting

After feature extraction, the complete dataset is divided into training data and testing data. The training dataset is used to develop the Random Forest forecasting model, while the test dataset is used to evaluate its performance on unseen observations. For example, an 80:20 division can be used, where 80% of the observations are used for training and 20% for testing. For time-series solar data, chronological order should preferably be maintained so that future observations do not influence the training process.

Step 5. Random Forest-Based Solar Irradiance Forecasting

The main forecasting stage uses the Random Forest (RF) algorithm. Random Forest is an ensemble learning technique that combines multiple decision trees to generate a final prediction. Each decision tree learns a relationship between the selected input features and the target solar irradiance. Instead of depending on a single tree, multiple trees are generated using different samples and feature combinations.

For a regression-based solar irradiance forecasting problem, the final RF prediction can be represented as:

$$\hat{Y} = \frac{1}{N}\sum_{i=1}^{N} T_i(X)$$

where $\hat{Y}$ is the predicted solar irradiance, N is the number of decision trees, $T_i(X)$ is the prediction produced by the i^{th} tree, and X represents the input features.

The individual decision trees divide the input data according to feature conditions and learn the relationship between meteorological parameters and irradiance. The predictions from all trees are then averaged to obtain the final irradiance forecast. This ensemble process improves prediction stability and reduces the dependence on a single decision tree.

Step 6. Test Data-Based Prediction

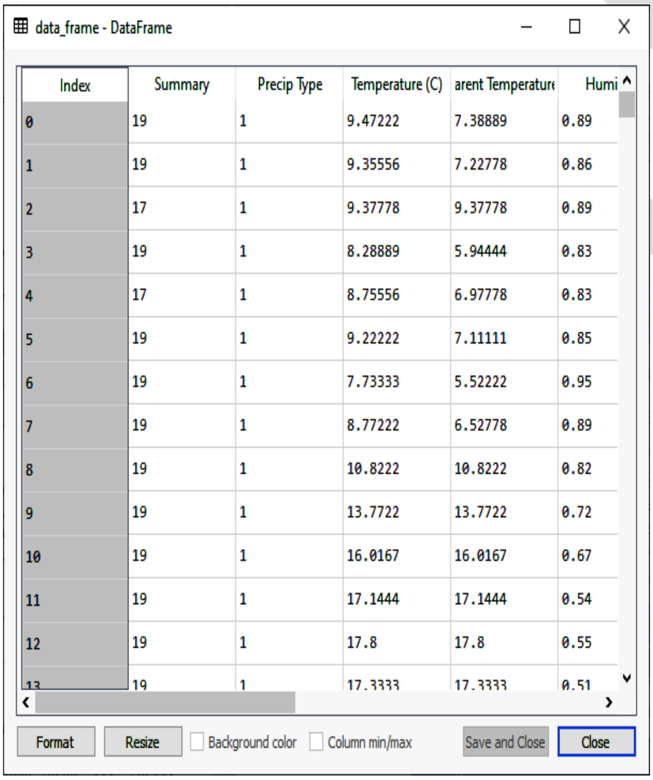
After the Random Forest model has been trained, the test dataset is supplied to the trained model. The model uses the selected meteorological and historical irradiance features to estimate the solar irradiance for the required future time interval. The predicted values are then compared with the actual measured irradiance values. This comparison determines how accurately the model can track changes in available solar radiation.

Step 7. Performance Analysis

In the final step, the performance of the proposed Random Forest-based solar irradiance forecasting model is evaluated using Accuracy and Error Rate. Accuracy represents the percentage of correctly predicted solar irradiance observations and indicates how closely the predicted values match the actual irradiance values.

III. SIMULATION RESULTS

The simulation results are performed using the python software.



Index	Summary	Precip Type	Temperature (C)	arent Temperature	Humi
0	19	1	9.47222	7.38889	0.89
1	19	1	9.35556	7.22778	0.86
2	17	1	9.37778	9.37778	0.89
3	19	1	8.28889	5.94444	0.83
4	17	1	8.75556	6.97778	0.83
5	19	1	9.22222	7.11111	0.85
6	19	1	7.73333	5.52222	0.95
7	19	1	8.77222	6.52778	0.89
8	19	1	10.8222	10.8222	0.82
9	19	1	13.7722	13.7722	0.72
10	19	1	16.0167	16.0167	0.67
11	19	1	17.1444	17.1444	0.54
12	19	1	17.8	17.8	0.55
13	19	1	17.3333	17.3333	0.51

Figure 2: Dataset

The figure 2 is showing the dataset, which is taken from the Kaggle machine learning website.

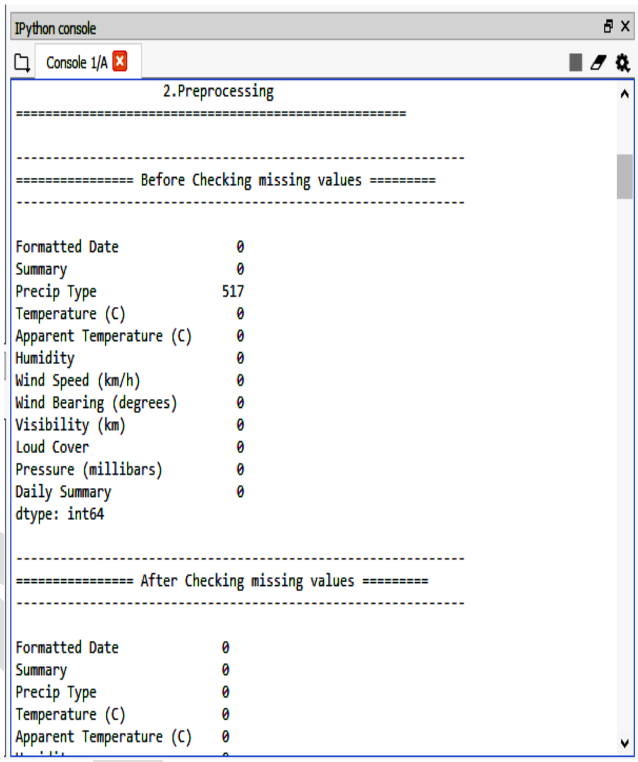


Figure 3: Process view in python console

Figure 3 is presenting the simulation process in the python console. First start with the dataset feature selection then preprocessing.

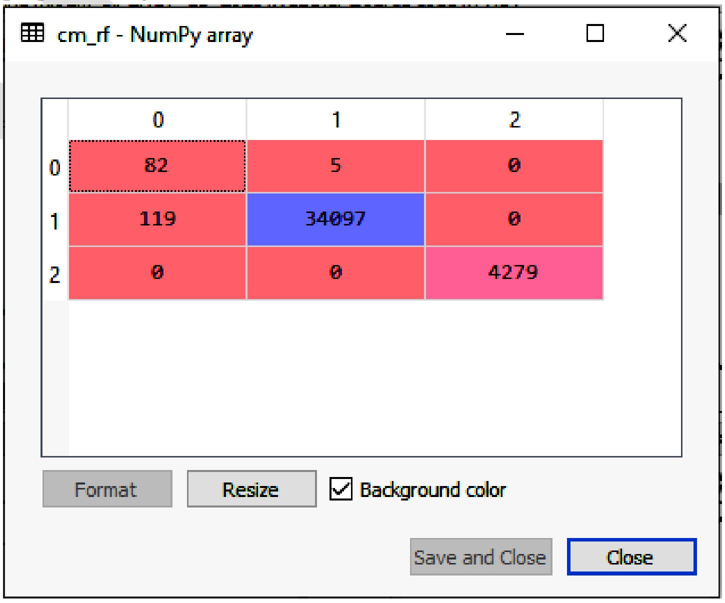


Figure 4: Confusion matrix

Figure 4 presents the confusion matrix for the proposed approach. A confusion matrix is a tabular representation used in machine learning classification to assess the performance of a classification algorithm by displaying the counts of true positives, false positives, true negatives, and false negatives.

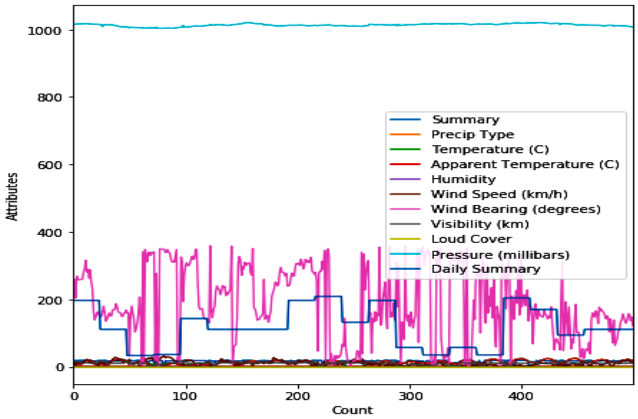


Figure 5: Performance attributes

Figure 5 is presenting performance attributes or metrics are quantitative measures used to assess the effectiveness and efficiency of a machine learning model. These metrics provide insights into how well the model is performing on a given task.

Table 1: Result Comparison

Sr. No.	Parameters	Previous Work [1]	Proposed Work
1	Method	DNN	Random Forest
2	Accuracy (%)	94.7	99.67
3	Error Rate (%)	5.3	0.33

IV. CONCLUSION

The increasing integration of solar PV generation into electrical power systems requires accurate solar irradiance forecasting for reliable power generation and grid operation. Accurate forecasting helps reduce uncertainty in PV output and supports generation scheduling, power balancing, and renewable-energy integration. This study proposed an efficient Random Forest-based Machine Learning technique for solar irradiance forecasting using relevant historical and meteorological data. The methodology included data preprocessing, feature extraction, training-testing data splitting, and Random Forest-based irradiance prediction. The model was evaluated using accuracy and error rate to determine its forecasting effectiveness. The proposed Random Forest technique achieved 99.67% accuracy, compared with 94.70% for the previous DNN method [1]. The error rate was reduced significantly from 5.30% to 0.33%, showing a clear improvement in forecasting performance. These results demonstrate the capability of Random Forest to provide reliable solar irradiance predictions for PV applications. The improved prediction can support PV power estimation, battery energy storage, generation scheduling, and power-system operation. Overall, the proposed method provides an effective approach for reducing solar-generation uncertainty and improving the reliable integration of solar PV into modern electrical power systems.

REFERENCES

[1] S. Roga, M. Lonkar, V. Mishra, P. Chakraborty, V. Dhanorkar and S. Bardhan, "Machine Learning Algorithms for Solar Irradiance Forecasting: A Custom Deep Neural Network Approach," 2025 IEEE 5th International Conference on Sustainable Energy and Future Electric Transportation (SEFET), Jaipur, India, 2025, pp. 1-6, doi: 10.1109/SEFET65155.2025.11255077.

[2] S. M. J. Jalali, S. Ahmadian, A. Kavousi-Fard, A. Khosravi and S. Nahavandi, "Automated Deep CNN-LSTM Architecture Design for Solar Irradiance Forecasting," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 1, pp. 54-65, Jan. 2022, doi: 10.1109/TSMC.2021.3093519.

[3] P. Singh, N. K. Singh and A. K. Singh, "Solar Photovoltaic Energy Forecasting Using Machine Learning and Deep Learning Technique," 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2022, pp. 1-6, doi: 10.1109/UPCON56432.2022.9986446.

[4] A. K. Parida, D. Kumar, R. R. Sahoo and B. K. Balabantaray, "Medium term Solar Power Prediction using stacked LSTM based deep learning Technique," 2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), 2022, pp. 19-26, doi: 10.1109/ICCCMLA56841.2022.9989067.

[5] U. T. Kartini, Hariyati, W. Aribowo and A. L. Wardani, "Development Hybrid Model Deep Learning Neural Network (DL-NN) For Probabilistic Forecasting Solar Irradiance on Solar Cells To Improve Economics Value Added," 2022 Fifth International Conference on Vocational Education and Electrical Engineering (ICVEE), 2022, pp. 151-156, doi: 10.1109/ICVEE57061.2022.9930352.

[6] C. N. Obiora, A. Ali and A. N. Hasan, "Finding the Optimum Horizon for Short-Term Solar Irradiance Forecasting Using Long Short-Term Memory (LSTM) Network," 2022 11th International Conference on

Power Science and Engineering (ICPSE), 2022, pp. 148-152, doi: 10.1109/ICPSE56329.2022.9935346.

[7] G. S. Thirunavukkarasu et al., "Very Short-Term Solar Irradiance Forecasting using Multilayered Long-Short Term Memory," 2022 7th International Conference on Smart and Sustainable Technologies (SpliTech), 2022, pp. 1-8, doi: 10.23919/SpliTech55088.2022.9854244.

[8] M. Pi, N. Jin, X. Ma and K. Yan, "Short-term Solar Irradiation Prediction Model Based on WCNN_ALSTM," 2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2021, pp. 405-412, doi: 10.1109/DASC-PICom-CBDCom-CyberSciTech52372.2021.00075.

[9] T. A. Fathima, V. Nedumpozhimana, J. Wu, Y. H. Lee and S. Dev, "Role of Temporal Information for Multi-step Ahead Forecasting of Solar Irradiance," 2021 Photonics & Electromagnetics Research Symposium (PIERS), 2021, pp. 2342-2346, doi: 10.1109/PIERS53385.2021.9695089.

[10] C. N. Obiora and A. Ali, "Effective Implementation of Convolutional Long Short-Term Memory (ConvLSTM) Network in Forecasting Solar Irradiance," IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, 2021, pp. 1-5, doi: 10.1109/IECON48115.2021.9589934.