International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

AI-Based Solar Power Forecasting Using Machine Learning Techniques

Satyawan Sundarrao Bahir[1], Prof. K. Chandra Obula Reddy[2]

1PG Scholar, Department of Electrical Engineering (Electrical Power System), MSS's College of Engineering and Technology, Jalna

2Assistant Professor, Department of Electrical Engineering, MSS's College of Engineering and Technology, Jalna

(Affiliated to Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra)

Abstract

Solar photovoltaic (PV) power generation is highly intermittent due to its dependence on weather-related parameters such as solar irradiation, ambient temperature, and module temperature, which creates significant challenges for grid stability and efficient energy management. Accurate solar power forecasting is therefore essential for improving renewable energy integration and supporting smart grid operations. This paper presents an AI-based solar power forecasting framework developed using the Kaggle Solar Power Generation Dataset. Four forecasting models — Linear Regression, Random Forest Regression, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) — were implemented and compared using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2 Score). Correlation analysis confirmed that irradiation (r = 0.95) is the most influential parameter affecting AC power output. Experimental results show that the LSTM model achieved the best forecasting accuracy (MAE = 1.74, RMSE = 2.35, R^2 = 0.97), outperforming the conventional machine learning models because of its ability to capture temporal dependencies in time-series data. The proposed framework offers an effective solution for intelligent solar energy prediction and smart renewable energy management.

Keywords: Solar PV Forecasting, Machine Learning, Deep Learning, LSTM, Renewable Energy Prediction.

I. INTRODUCTION

The growing demand for electricity and the environmental impact of fossil-fuel-based power generation have accelerated the global transition toward cleaner and more sustainable energy sources [1]. Among renewable technologies, solar photovoltaic (PV) systems have emerged as one of the most promising solutions because of their environmental friendliness, abundant availability, and continuous technological advancement [2]. Large-scale solar farms, rooftop installations, and distributed PV units are now widely integrated into electrical grids, reducing dependence on fossil fuels [3]. However, unlike conventional power plants, solar generation is highly intermittent and strongly dependent on environmental conditions; variations in irradiance, temperature, humidity, and cloud movement cause fluctuations in power output that affect grid stability and power quality [4].

Accurate forecasting of PV power generation helps grid operators and utilities balance electricity generation and demand, supports efficient scheduling, energy trading, reserve allocation, and demand-side management, and reduces operational costs and energy wastage [5], [6]. Traditional statistical forecasting methods, however, often fail to capture the complex, nonlinear relationships between weather parameters and solar output [7]. Recent advances in artificial intelligence (AI), machine learning (ML), and deep learning (DL) provide new opportunities for improving forecasting accuracy by learning hidden, nonlinear patterns from historical weather and generation data.

Traditional solar forecasting methods rely on statistical models or simplified assumptions. While easy to implement, such approaches are static and adapt poorly to changing environmental conditions, producing less accurate predictions in dynamic weather scenarios. Because solar generation is inherently variable, inaccurate forecasts create energy imbalances — either excess generation or unmet demand — that increase operational cost, reduce system efficiency, and can contribute to grid instability and voltage fluctuations. These limitations motivate the need for more flexible, data-driven forecasting techniques.

Artificial intelligence offers a promising solution because AI-based models can capture complex, nonlinear relationships between environmental variables and continuously improve as more data becomes available, unlike static conventional methods. This adaptability enables more reliable short-term predictions, supports better scheduling and decision-making, and enhances the integration of renewable sources into the grid.

This paper develops and compares four AI-based forecasting models — Linear Regression, Random Forest Regression, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) — for predicting solar PV output from irradiation and temperature data, and identifies the most accurate and reliable technique for smart renewable energy management applications such as energy scheduling, reserve allocation, and loss reduction.

II. LITERATURE REVIEW

Barhmi et al. [1] reviewed solar forecasting techniques and highlighted the growing role of artificial intelligence in improving prediction accuracy over classical statistical approaches. Markovics and Mayer [2] compared several machine learning methods for PV power forecasting based on numerical weather prediction and reported that ensemble and kernel-based methods outperform simple regression models. Alzain et al. [3] proposed a sustainable AI-based predictive model for solar power production, while Mehmood et al. [4] developed a hybrid AI framework combining machine learning and time-series analysis for accurate solar forecasting.

Ledmaoui et al. [5] presented a comparative study of machine learning algorithms for solar energy production forecasting, and Sammar et al. [6] reviewed AI-based solar irradiance prediction models, emphasizing the importance of feature selection and data quality. Meenal et al. [7] reviewed weather-forecasting techniques for renewable energy systems and noted that irradiance and temperature remain the most influential predictors of PV output.

Gaboitaolelwe et al. [8] reviewed machine learning approaches for short-term solar PV forecasting and compared model performance through simulation studies. Karthikeyan and Jagadeeshwaran [9] combined machine learning with ensemble techniques for PV power prediction and fault detection, achieving improved accuracy and processing speed. Rashid et al. [10] developed a hybrid CNN-BiLSTM model coupled with a genetic-algorithm-based dispatch strategy for PV–BESS coordination, reporting reduced losses and improved power quality. Pandu et al. [11] combined big-data analytics with an Elman neural network optimized by the Mayfly algorithm for solar irradiance prediction, while Bouquet et al. [12] applied an LSTM-based framework for solar-energy management and Smart Grid reliability, outperforming persistent forecasting baselines for horizons beyond two hours.

Said and Alanazi [13] proposed AI-based solar forecasting for smart-grid integration, showing that learned models adapt better to non-stationary weather patterns than fixed statistical baselines. Ozdemir et al. [14] applied explainable-AI techniques to interpret machine-learning-based solar power forecasts, improving the transparency of black-box model decisions for grid operators. Jain and Naik [15] compared AI models for solar power forecasting and confirmed that deep-learning architectures consistently outperform classical regression techniques across varying weather conditions.

Across these studies, irradiance and temperature consistently emerge as the dominant predictors of PV output, ensemble and kernel-based methods reliably outperform simple regression, and recurrent deep-learning models such as LSTM achieve the best accuracy on sequential generation data. However, most existing studies evaluate only one or two models on a given dataset rather than a systematic, multi-model comparison under identical preprocessing and evaluation conditions. This gap motivates the comparative evaluation of Linear Regression, Random Forest, SVM, and LSTM carried out in the present study using a common dataset and metric set.

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

III. METHODOLOGY

A. System Architecture

The proposed forecasting system follows a sequential workflow consisting of data collection, preprocessing, feature extraction, model training, forecast generation, and performance evaluation, as illustrated in Fig. 1. Solar generation and weather-sensor data are first collected from the Kaggle Solar Power Generation Dataset. Missing values, duplicate records, and noise are removed, timestamps are converted to a standard datetime format, and Min–Max normalization is applied. The dataset is then split into training and testing subsets (70:30) and supplied to four forecasting models, whose outputs are evaluated using MAE, RMSE, and R^2 Score.

This modular architecture separates data handling from model implementation, allowing the same preprocessed dataset and evaluation pipeline to be reused consistently across all four models. This design ensures a fair, like-for-like comparison of forecasting performance, since every model is trained and tested on identical inputs, splits, and metrics.

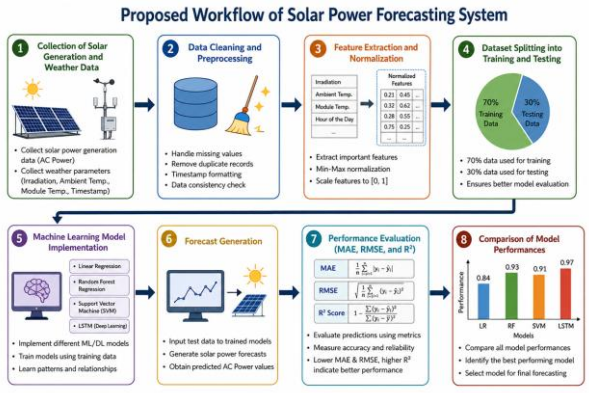


Fig. 1. Proposed workflow of the AI-based solar power forecasting system.

B. Dataset Description

The Kaggle Solar Power Generation Dataset provides real-time PV plant operational data and weather-sensor measurements at regular intervals, consisting of four files summarized in Table 1.

Dataset File	Description
Plant_1_Generation_Data.csv	Solar power generation data
Plant_1_Weather_Sensor_Data.csv	Weather sensor measurements
Plant_2_Generation_Data.csv	Additional generation data
Plant_2_Weather_Sensor_Data.csv	Additional weather data

TABLE 1. DATASET FILES DESCRIPTION

The forecasting model uses irradiation, ambient temperature, module temperature, and timestamp as input features and AC power as the output parameter, as summarized in Table 2. Irradiation is the most influential input because PV output directly depends on incident sunlight intensity.

Parameter	Description
Irradiation	Solar radiation intensity (input)
Ambient Temperature	Surrounding air temperature (input)
Module Temperature	Panel surface temperature (input)
AC Power	Generated solar power output (output)

TABLE 2. INPUT AND OUTPUT PARAMETERS

C. Data Preprocessing

Raw data containing missing, null, or duplicate records was removed to improve reliability, reduce prediction error, and eliminate unwanted noise. Timestamps were converted into standard datetime format, and additional temporal features (hour, day, month) were extracted, since solar generation varies systematically with time of day. Min–Max normalization was then applied to scale all inputs into the range [0, 1] so that no single feature dominates model training:

$$Xnorm = (X - Xmin) / (Xmax - Xmin) \quad ...(1)$$

Feature selection retained irradiation, ambient temperature, module temperature, and hour of day as the most relevant predictors of AC power, reducing computational complexity while preserving forecasting accuracy. The cleaned and normalized dataset was then

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

split into training and testing subsets, as summarized in Table 3.

Dataset Portion	Percentage
Training Data	70%
Testing Data	30%

TABLE 3. TRAINING AND TESTING DATASET DISTRIBUTION

The training subset allows each model to learn the relationship between irradiation, ambient temperature, module temperature, and AC power, while the held-out testing subset is used to evaluate generalization to unseen conditions and to prevent overfitting.

D. Forecasting Models

Linear Regression establishes a linear relationship between the input features and AC power:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn \quad \ldots(2)$$

where Y is the predicted solar power, Xn are the input features, βn the regression coefficients, and β0 the intercept. The model is simple, computationally efficient, and easy to interpret, making it a useful baseline, though it struggles with nonlinear environmental variations.

- Linear Regression: simple, fast, interpretable baseline.
- Random Forest: handles nonlinearity, robust to noise, reduces overfitting.
- SVM: strong generalization on multidimensional, nonlinear data.
- LSTM: captures long-term temporal dependencies in time-series data.

Random Forest Regression is an ensemble of decision trees whose averaged output improves accuracy and reduces overfitting relative to a single regressor. It handles nonlinear relationships efficiently, is robust to noise and outliers, and performs well on large datasets. Support Vector Regression (SVR) maps input features into a higher-dimensional space through kernel functions to fit a nonlinear hyperplane. It offers good generalization and handles multidimensional environmental data effectively, at the cost of somewhat higher computational complexity.

Long Short-Term Memory (LSTM) is a recurrent deep-learning architecture with memory cells that retain information across time steps, making it well suited to the sequential nature of solar generation data. Its hidden state at time t is computed as:

$$ht = f(Wh\ ht{-}1 + Wx\ xt + b) \quad \ldots(3)$$

where ht is the current hidden state, xt the current input, Wh and Wx the weight matrices, and b the bias term. Because it processes sequential inputs while retaining historical context, LSTM captures long-term temporal dependencies, handles nonlinear relationships, and is expected to outperform models that treat each observation independently.

E. Performance Evaluation Metrics

Model accuracy was evaluated using:

$$MAE = (1/n)\ \Sigma\ |yi - \hat{y}i| \quad \ldots(4)$$

$$RMSE = \sqrt{[(1/n)\ \Sigma\ (yi - \hat{y}i)^2]} \quad \ldots(5)$$

$$R^2 = 1 - [\Sigma(yi - \hat{y}i)^2 / \Sigma(yi - \dot{y})^2] \quad \ldots(6)$$

where yi is the actual value, ŷi the predicted value, and ẏ the mean of actual values. Lower MAE/RMSE and R^2 closer to 1 indicate better forecasting performance.

F. Software and Tools

Tool / Library	Purpose
Python	Programming
Pandas / NumPy	Data preprocessing
Scikit-learn	Machine learning models
TensorFlow / Keras	Deep learning (LSTM)
Matplotlib	Visualization

TABLE 4. SOFTWARE AND TOOLS USED

IV. RESULTS AND DISCUSSION

A. Correlation Analysis

Correlation analysis (Table 4) shows that irradiation has the strongest positive correlation with AC power (r = 0.95), confirming it as the dominant predictor, followed by module temperature (r = 0.81) and ambient temperature (r = 0.72).

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

Parameter	Correlation with AC Power
Irradiation	0.95
Module Temperature	0.81
Ambient Temperature	0.72

TABLE 5. CORRELATION MATRIX BETWEEN INPUT PARAMETERS AND AC POWER

A representative sample of the cleaned, time-stamped dataset used for model training and testing is shown in Table 6, illustrating how AC power tracks irradiation and temperature across a single day.

Timestamp	Irradiation	Amb. Temp (°C)	Mod. Temp (°C)	AC Power (kW)
06:00	0.12	24.5	26.3	15.2
09:00	0.48	29.1	34.7	82.5
12:00	0.91	35.4	46.2	210.8
15:00	0.63	33.8	41.5	154.7
18:00	0.08	27.2	29.6	10.4

TABLE 6. SAMPLE DATASET RECORDS

B. Data Visualization

Fig. 2 shows that solar irradiation rises after sunrise, peaks near midday, and falls toward evening. Fig. 3 shows that the AC power profile closely follows this irradiation trend, confirming the strong dependency between environmental conditions and PV output and validating the dataset's suitability for forecasting.

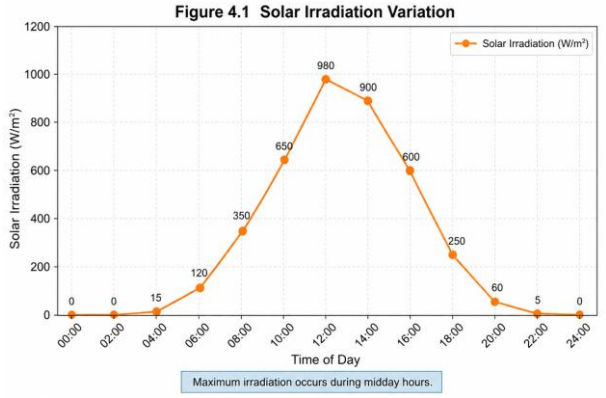


Fig. 2. Solar irradiation trend observed from the dataset.

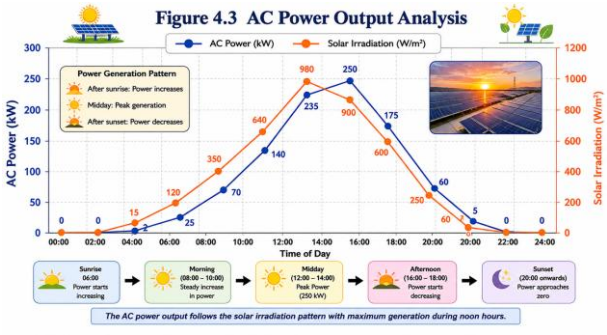


Fig. 3. AC power generation profile.

C. Individual Model Performance

Linear Regression, used as the baseline model, provided fast, simple, and easily interpretable predictions (MAE = 4.82, RMSE = 6.15, R^2 = 0.84) but showed larger deviations during sudden irradiation changes and peak generation hours because it assumes a purely linear input–output relationship. Random Forest Regression improved accuracy substantially (MAE = 2.91, RMSE = 3.84, R^2 = 0.93) through ensemble averaging across decision trees, which reduced overfitting and handled nonlinear irradiation–temperature interactions more effectively. Support Vector Regression achieved comparable gains (MAE = 3.26, RMSE = 4.28, R^2 = 0.91) by mapping features into a higher-dimensional kernel space, giving stable predictions across multidimensional weather data, though at higher computational cost than Random Forest.

The LSTM model achieved the best forecasting performance overall (MAE = 1.74, RMSE = 2.35, R^2 = 0.97). Its recurrent memory cells retained information from previous time steps, allowing it to track fluctuations in irradiation and temperature and to closely follow the actual solar generation profile with minimal deviation, including during rapid transitions around sunrise, midday peak, and sunset.

D. Comparative Analysis and Error Discussion

Table 7 summarizes the comparative performance of all four models, and Fig. 4 presents the corresponding MAE, RMSE, and R^2 comparison. Mean Absolute Error, which reflects average prediction error, was lowest for LSTM and highest for Linear Regression; Root Mean Square Error, which penalizes larger deviations more heavily, followed the same trend,

confirming that LSTM produced fewer large forecasting errors during rapid environmental changes. The Coefficient of Determination followed the inverse pattern, with LSTM explaining 97% of the variance in AC power output compared with 84% for Linear Regression.

Model	MAE	RMSE	R² Score
Linear Regression	4.82	6.15	0.84
Random Forest	2.91	3.84	0.93
SVM	3.26	4.28	0.91
LSTM	1.74	2.35	0.97

TABLE 7. COMPARATIVE ANALYSIS OF FORECASTING MODELS

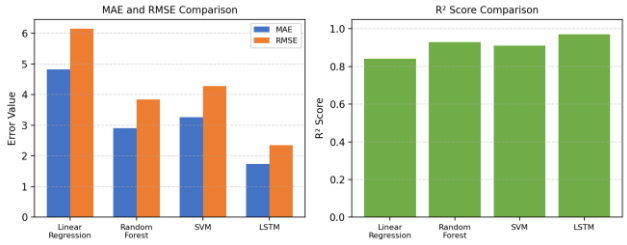


Fig. 4. Comparative MAE/RMSE and R² Score of the forecasting models.

E. Actual vs. Predicted Output

Graphical comparison of actual versus predicted AC power was performed for each model to assess how closely its forecasts followed the real generation profile. Linear Regression tracked the general rise-and-fall trend of daily generation but showed larger deviations near peak irradiation, since it cannot represent the nonlinear response of PV output to rapidly changing weather. Random Forest produced predictions that closely matched the actual curve, with smaller deviations during changing conditions, reflecting the benefit of averaging across multiple decision trees. SVM likewise gave stable, consistent predictions across the daily cycle, with only minor deviations during maximum-irradiation periods.

LSTM achieved the closest match to the actual generation profile of all four models, tracking sunrise, midday-peak, and sunset transitions with minimal deviation. This confirms that its memory-based architecture is particularly effective at learning the sequential dependencies characteristic of solar generation time series, and supports its selection as the preferred model for this forecasting task.

E. Discussion

The results confirm that model accuracy improves as architectural complexity increases from simple linear regression to ensemble and kernel-based methods, and is highest for the sequence-aware LSTM model. LSTM's memory cells allow it to retain information from prior time steps, enabling it to track fluctuations caused by rapidly changing irradiation and temperature more accurately than models that treat each observation independently. These findings agree with prior comparative studies [9], [10], [15], and confirm that deep-learning techniques are well suited to time-series solar forecasting, supporting smart-grid stability, efficient power scheduling, and reduced energy losses.

From a practical standpoint, the accuracy gains delivered by LSTM come with additional computational cost during training compared with Linear Regression or Random Forest, which should be weighed against forecasting-horizon requirements in real deployments. For applications requiring fast, interpretable estimates, Random Forest offers a strong accuracy-to-complexity trade-off, whereas LSTM is preferable where maximum accuracy is critical, such as short-term dispatch and reserve-scheduling decisions in smart-grid operations.

V. CONCLUSION

This paper presented an AI-based solar power forecasting framework built on the Kaggle Solar Power Generation Dataset. Four models — Linear Regression, Random Forest, SVM, and LSTM — were implemented and compared using MAE, RMSE, and R² Score. Correlation analysis confirmed irradiation as the most influential predictor of AC power (r = 0.95), followed by module temperature (r = 0.81) and ambient temperature (r = 0.72). Proper preprocessing, including missing-value removal, timestamp formatting, normalization, and feature selection, was essential to achieving reliable model performance.

Among the evaluated models, Linear Regression provided an acceptable but limited baseline, Random Forest and SVM offered clear accuracy improvements by capturing nonlinear relationships, and LSTM provided the best overall performance, with the lowest error (MAE = 1.74, RMSE = 2.35) and the highest R^2 Score (0.97), owing to its ability to capture temporal dependencies in sequential solar-generation data. The proposed framework offers an effective and reliable solution for intelligent PV power prediction and can support smart-grid stability, efficient power scheduling, and reduced energy losses in future renewable energy management systems.

Future Scope

Future work may integrate real-time weather data from IoT sensors and weather APIs to replace the historical-dataset training used here, and develop hybrid architectures such as CNN-LSTM, BiLSTM, or attention-based networks to improve feature extraction under rapidly changing environmental conditions. The framework can further be extended to multi-step, day-ahead, and week-ahead forecasting, seasonal solar-energy analysis, and large-scale solar-farm forecasting. Additional environmental parameters such as cloud cover, atmospheric pressure, wind direction, rainfall, and air-quality index could also be incorporated to further improve forecasting accuracy and support broader smart-grid deployment.

ACKNOWLEDGMENT

I am greatly obligated forever to my guide and HOD K .Chandra Obula Reddy and to all teaching and non-teaching staff who supported me directly and indirectly to complete my work. I am sincerely thankful to my principal Dr.S.K Biradar for their continued encouragement and active interest in my progress throughout the work I am grateful to be an M.Tech Electrical Power System student at Matsyodari Shikshan Sanstha ‘s College of Engineering and Technology, Jalna, Maharashtra

REFERENCES

[1] K. Barhmi, C. Heynen, S. Golroodbari, and W. van Sark, "A Review of Solar Forecasting Techniques and the Role of Artificial Intelligence," Solar, 2024, doi: 10.3390/solar4010005.

[2] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," Renew. Sustain. Energy Rev., vol. 161, p. 112364, 2022.

[3] E. Alzain, S. Al-Otaibi, T. H. H. Aldhyani, A. S. Alshebami, M. A. Almaiah, and M. E. Jadhav, "Revolutionizing Solar Power Production with Artificial Intelligence: A Sustainable Predictive Model," Sustainability, 2023, doi: 10.3390/su15107999.

[4] B. Mehmood et al., "Development of a Hybrid Artificial Intelligence Framework for Accurate Forecasting of Solar Power Generation Using Machine Learning Algorithms and Time-series Analysis," Spectr. Eng. Sci., vol. 3, no. 5, pp. 613–636, 2025.

[5] Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, and A. Chehri, "Forecasting solar energy production: A comparative study of machine learning algorithms," Energy Reports, vol. 10, pp. 1004–1012, 2023.

[6] M. J. Sammar et al., "Illuminating the Future: A Comprehensive Review of AI-Based Solar Irradiance Prediction Models," IEEE Access, vol. 12, pp. 114394–114415, 2024.

[7] R. Meenal et al., "Weather Forecasting for Renewable Energy System: A Review," Arch. Comput. Methods Eng., vol. 29, no. 5, pp. 2875–2891, 2022.

[8] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod, and A. O. Salau, "Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison," IEEE Access, vol. 11, pp. 40820–40845, 2023.

[9] G. Karthikeyan and A. Jagadeeshwaran, "Enhancing Solar Energy Generation: A Comprehensive Machine Learning-Based PV Prediction and Fault Analysis System for Real-Time Tracking and Forecasting," Electr. Power Compon. Syst., vol. 52, no. 9, pp. 1497–1512, 2024.

[10] T. H. M. S. Rashid et al., "Machine learning-driven PV forecasting and coordinated PV–BESS dispatch for optimizing distribution grid performance under high EV penetration," Energy Convers. Manag. X, vol. 29, p. 101418, 2026.

[11] S. B. Pandu et al., "Artificial intelligence based solar radiation predictive model using weather forecasts," Comput. Mater. Contin., vol. 71, no. 1, pp. 109–124, 2022.

[12] P. Bouquet, I. Jackson, M. Nick, and A. Kaboli, "AI-based forecasting for optimised solar energy management and smart grid efficiency," Int. J. Prod. Res., vol. 62, no. 13, pp. 4623–4644, 2024.

[13] Y. Said and A. Alanazi, "AI-based solar energy forecasting for smart grid integration," Neural Comput. Appl., vol. 35, no. 11, pp. 8625–8634, 2023.

[14] G. Ozdemir, M. Kuzlu, and F. O. Catak, "Machine learning insights into forecasting solar power generation with explainable AI," Electr. Eng., vol. 107, no. 6, pp. 7329–7350, 2025.

[15] N. Jain and D. Naik, "AI based Solar Power Forecasting," in Proc. 2024 15th Int. Conf. Computing Communication and Networking Technologies (ICCCNT), 2024, pp. 1–7.

Authors Biography

Author1



Mr. Satyawan Sundarrao Bahir is pursuing M.TECH degree in stream of Electrical Power System from Mss's college of Engineering and technology, Jalna, DR BATU University, Lonere. He was completed his B.E in the stream of Electrical Engineering from G.H.Raisoni institute of engineering and management jalgaon Affilated to Kaviyatri Bahinabai Choudhari North Maharashtra University Jalgaon
Email Id: satyawanbahir2018@gmail.com

Author2:



Prof Mr. K Chandra ObulaReddy receivedMtech in Power Electronics from the oxford college of engineering, Bangalore,VTU University,Belgaum. B.Tech in Electrical and Electronics Engineering from SVCET,chitoor,JNTU university Anantapur,A.P.He is currently working as HOD and Professor in Department of Electrical Engineering atMSS's College of Engineering and Technology,Jalna
Email Id: chandra6279@gmail.com