International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5347-5354
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0699

AI/ML-Based Short-Term Solar PV Power Forecasting: A Comparative Study of ANN, Random Forest, And SVR

Subash Ranjan Kabat[1], Priyadarshinee Das[2], Rashmita Lenka[3], Pragyanjit Jena[4]
[1]Professor, Department of Electrical Engineering, Radhakrishna Institute of Technology and Engineering, Odisha, India, 752057.
[2,3,4]Assistant Professor, Department of Electrical Engineering, Radhakrishna Institute of Technology and Engineering, Odisha, India, 752057.
Email Id: *srkabat.ee@riteindia.edu.in[1], pdas_ee@riteindia.edu.in[2], rlenka_ee@riteindia.edu.in[3], pjena_ee@riteindia.edu.in[4]*

Abstract
The increasing penetration of solar photovoltaic (PV) generation into modern power systems has created a growing need for accurate short-term PV power forecasting to support reliable grid operation, energy management, and renewable energy integration. This study presents a comparative analysis of three machine learning models, namely Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Regression (SVR), for short-term solar PV power forecasting. The forecasting framework utilizes solar irradiance, ambient temperature, module temperature, temporal variables, and historical PV power as input features. A chronological training-testing strategy is adopted to preserve the time-series characteristics of PV generation. The performance of the developed models is evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), normalized RMSE (nRMSE), and coefficient of determination (R^2). Initial results obtained from the experimental dataset demonstrate that Random Forest provides superior forecasting performance compared with ANN and SVR, achieving lower prediction errors and a higher coefficient of determination. The results indicate that ensemble-based machine learning techniques can effectively capture the nonlinear relationship between environmental conditions, historical generation, and PV power output. The proposed comparative framework provides a foundation for the development of advanced PV forecasting systems for grid-connected solar power plants. Future work will incorporate Long Short-Term Memory (LSTM) networks and validate the proposed models using real-world Indian PV plant data under different forecasting horizons and operating conditions.
Keywords: *Artificial Neural Network, Machine Learning, PV Power Forecasting, Random Forest, Renewable Energy, Short-Term Forecasting, Solar Photovoltaic, Support Vector Regression.*

1. Introduction
The increasing deployment of renewable energy technologies has significantly changed the structure and operation of modern electrical power systems. Among various renewable energy sources, solar photovoltaic (PV) generation has emerged as an important technology because of its scalability, declining cost, and availability of solar resources. However, unlike conventional power plants, PV generation is highly dependent on environmental conditions such as solar irradiance, ambient temperature, module temperature, cloud cover, and other atmospheric parameters. Consequently, the output of a PV plant can exhibit significant temporal variability and uncertainty [1]. The intermittent nature of solar generation presents challenges for power-system planning, scheduling, and real-time operation. Increasing penetration of variable renewable energy can result in rapid changes in net load and increased uncertainty for system operators. Accurate forecasting of PV power can therefore assist in unit commitment, economic dispatch, energy management, reserve allocation, voltage management, and reliable integration of renewable generation into the electrical grid [2]. PV power

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5347-5354
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0699

forecasting methods can generally be classified into physical, statistical, machine-learning, deep-learning, and hybrid approaches. Physical approaches generally use solar-resource and PV-system characteristics to estimate power output, whereas statistical and data-driven methods establish relationships between historical measurements and future generation. Direct PV power forecasting, in particular, has received considerable attention because the predicted power output can be directly utilized by grid operators and energy-management systems [3]. With the rapid development of artificial intelligence (AI), machine-learning techniques have become increasingly attractive for PV forecasting because they can model complex and nonlinear relationships between meteorological variables and PV power output without requiring an explicit physical model. Previous studies have extensively investigated techniques such as Artificial Neural Networks (ANNs), Support Vector Machines/Support Vector Regression (SVM/SVR), decision trees, ensemble learning, and other machine-learning approaches for solar forecasting. A systematic review reported that ANN and support-vector-based methods have played an important role in solar electricity forecasting, with meteorological parameters frequently used as model inputs [4]. Artificial Neural Networks are particularly suitable for PV forecasting because of their ability to approximate nonlinear input-output relationships. ANN-based approaches have consequently been widely investigated for solar-energy-system modeling and PV power prediction. However, ANN performance can depend strongly on network architecture, feature selection, parameter settings, and the characteristics of the training dataset [5]. Tree-based ensemble methods provide another promising approach. In particular, Random Forest (RF) can capture nonlinear interactions among multiple input variables while being relatively robust to variations in feature distributions. Previous large-scale comparisons of machine-learning methods for hourly solar forecasting have reported that tree-based methods can perform consistently well across different locations and climatic conditions. However, the optimal forecasting method cannot necessarily be identified independently of the dataset, location, forecasting horizon, and operating conditions, emphasizing the importance of comparative evaluation [6]. Similarly, Support Vector Regression (SVR) has been widely applied to nonlinear regression and solar-power forecasting problems. SVR can provide good generalization performance by mapping nonlinear relationships into a higher-dimensional feature space through kernel functions. Nevertheless, its performance depends on appropriate selection of parameters such as the penalty factor, kernel parameters, and insensitive loss function. Therefore, comparative assessment with other machine-learning models is necessary rather than assuming that a particular algorithm will always provide the best forecasting accuracy [7]. The selection of appropriate performance metrics is also important in PV forecasting research. MAE and RMSE provide information about absolute prediction errors, while MAPE provides a relative measure of forecasting accuracy. However, MAPE can become problematic when actual PV power approaches zero, particularly during nighttime or low-generation periods. Consequently, evaluation practices should consider the characteristics of the PV generation profile and should clearly describe the data preprocessing and validation methodology. Recent literature has emphasized the importance of consistent evaluation procedures and avoiding temporal data leakage when assessing PV forecasting models [8]. Although numerous studies have investigated individual forecasting algorithms, a systematic comparison of different machine-learning approaches under the same dataset, feature set, training-testing strategy, and evaluation metrics remains valuable. In particular, understanding the relative performance of ANN, RF, and SVR can help identify suitable models for short-term PV forecasting and provide a baseline for subsequent deep-learning approaches such as Long Short-Term Memory (LSTM) networks. Previous research has demonstrated the growing importance of deep-learning approaches for capturing temporal patterns in renewable-energy forecasting, while conventional machine-learning models can remain attractive because of their lower computational complexity and suitability for moderate-sized datasets [9]. Therefore, this study presents a comparative analysis of ANN,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5347-5354
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0699

Random Forest, and SVR models for short-term solar PV power forecasting. The models utilize relevant meteorological, temporal, and historical PV-generation variables to estimate PV power output. Their performance is evaluated using MAE, RMSE, MAPE, normalized RMSE (nRMSE), and R^2. A chronological training-testing strategy is adopted to preserve the temporal characteristics of the data. The objective is to identify the most effective machine-learning approach for the investigated PV forecasting problem and establish a reliable baseline for future integration of deep-learning models and PV forecasting applications in power-system operation.

2. Methodology

2.1.Overall Research Framework

The proposed methodology develops and evaluates machine-learning models for short-term solar photovoltaic (PV) power forecasting. The framework consists of five major stages: data acquisition, data preprocessing, feature engineering, machine-learning model development, and performance evaluation. The overall methodology is illustrated conceptually in Figure1.

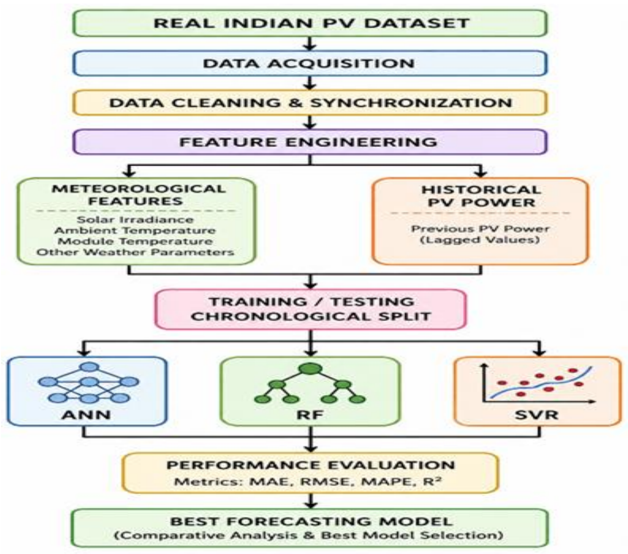

Figure 1 Proposed Methodology for Solar PV Power Forecasting Using Machine Learning Models

The objective is to determine which machine-learning model provides the most accurate short-term prediction of PV power under the investigated operating conditions.

2.2.PV Dataset

The study utilizes the publicly available Solar Power Generation Data obtained from an operational PV installation in India. The dataset contains generation measurements and weather-sensor measurements recorded at 15-minute intervals. The generation dataset contains parameters such as:

- DC power
- AC power
- Daily yield
- Total yield
- Date and time

The weather dataset contains:

- Irradiance
- Ambient temperature
- Module temperature
- Date and time

The two datasets are synchronized using their common timestamp and plant identifier. For the forecasting model, AC power is selected as the target variable, since it represents the electrical power delivered by the PV system after conversion through the inverter.

2.3.Data Preprocessing

Data preprocessing is performed to improve data quality and ensure that the machine-learning models receive consistent input information.

2.3.1. Date-Time Conversion

The date and time fields are converted into a standard date time format. The observations are then sorted chronologically.

2.3.2. Data Synchronization

The generation and weather datasets are merged using:

$$\boldsymbol{Date\_Time}$$

and the corresponding plant identifier. This produces a unified dataset containing both environmental conditions and PV generation.

2.3.3. Missing-Value Treatment

The dataset is examined for missing values. Records containing incomplete values in essential forecasting variables are removed or appropriately treated before model training.

2.3.4. Duplicate Removal

Duplicate observations are identified and removed to

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5347-5354
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0699

avoid giving excessive weight to repeated measurements.

2.3.5. Night-Time Data

PV power is approximately zero during nighttime. Since percentage-based forecasting metrics such as MAPE can become unstable when actual power approaches zero, daytime observations are considered separately when calculating daytime MAPE.

2.4.Feature Engineering

Feature engineering is performed to provide the models with both environmental and temporal information. The following variables are considered as input features Shown in Table 1:

Table 1 Features Used for Solar PV Power Forecasting

Category	Feature
Solar resource	Irradiance
Temperature	Ambient temperature
Temperature	Module temperature
Time	Hour
Time	Day
Time	Month
Historical generation	Previous 15-min PV power
Historical generation	Previous 30-min PV power
Historical generation	Previous 45-min PV power
Historical generation	Previous 1-hour PV power
Historical generation	Previous-day PV power

Historical PV power is incorporated through lagged variables. For example:

$$P_{t-1} = P(t - 15\text{ min})$$

$$P_{t-4} = P(t - 60\text{ min})$$

and an approximately one-day lag is represented by:

$$P_{t-96}$$

because the dataset has a 15-minute sampling interval. The inclusion of lagged variables allows the models to learn temporal relationships in PV generation.

2.5.Training and Testing Strategy

Because PV forecasting is a time-series problem, random shuffling of observations is avoided. The dataset is divided chronologically into:

- 80% training data
- 20% testing data

The training data are used to develop the models, while the final 20% are reserved for independent evaluation. This approach more closely represents the practical forecasting situation in which future PV power must be predicted using information available from the past.

2.6.Feature Scaling

Feature normalization is applied to models that are sensitive to the scale of input variables. Standardization is performed using:

$$X_{scaled} = \frac{X - \mu}{\sigma}$$

where:

- X= original feature value
- μ= mean of the training feature
- σ= standard deviation of the training feature.

The scaling parameters are obtained only from the training dataset and subsequently applied to the testing dataset to prevent information leakage. Random Forest does not require feature scaling, while ANN and SVR use standardized inputs.

2.7.Machine-Learning Models

Three machine-learning algorithms are investigated.

2.7.1. Artificial Neural Network

The ANN model is designed to capture nonlinear relationships between environmental variables, historical PV generation, and future PV power. The network consists of:

- Input layer
- Multiple hidden layers
- Output layer

A ReLU activation function is used in the hidden layers, while the output layer produces a continuous PV power prediction. The ANN learns the mapping:

IRJAEH
International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5347-5354
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0699

$$P_{PV} = f(G, T_a, T_m, H_t, P_{t-1}, P_{t-2}, \dots)$$

Where G represents irradiance, T_a represents ambient temperature, T_m represents module temperature, and Prepresents historical PV power.

2.7.2. Random Forest

Random Forest is an ensemble learning algorithm consisting of multiple decision trees. Each tree generates an individual prediction and the final regression output is obtained from the ensemble of trees:

$$\hat{P}_{RF} = \frac{1}{N}\sum_{i=1}^{N} \hat{P}_i$$

Where N is the number of decision trees. Random Forest is particularly suitable for this application because it can capture nonlinear relationships and interactions among irradiance, temperature, time variables, and historical PV generation.

2.7.3. Support Vector Regression

Support Vector Regression (SVR) is employed to model nonlinear relationships between input variables and PV power. The radial basis function (RBF) kernel is used:

$$K(x_i, x_j) = \exp(-\gamma \parallel x_i - x_j \parallel^2)$$

Where γ controls the influence of individual training samples. The main SVR parameters include:

- C: penalty parameter
- ϵ: insensitive loss parameter
- **γ:** kernel coefficient.

These parameters control the balance between model complexity and prediction accuracy.

2.7.4. Performance Evaluation

The forecasting models are evaluated using five performance metrics.

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

MAE represents the average magnitude of prediction error.

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

RMSE gives greater weight to larger prediction errors.

Mean Absolute Percentage Error

For daytime observations:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

Night-time observations are excluded from MAPE because actual PV output approaches zero.

Normalized RMSE

$$nRMSE = \frac{RMSE}{P_{rated}} \times 100$$

where P_{rated} represents the relevant PV system power rating.

Coefficient of Determination

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

An R^2 value closer to 1 indicates better agreement between predicted and actual PV power.

3. Results and Discussion

The performance of three machine-learning models, namely Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Regression (SVR), was evaluated for short-term solar photovoltaic (PV) power forecasting. The models were trained using solar irradiance, ambient temperature, module temperature, humidity, wind speed, time-related variables, and historical PV power information. The dataset was divided chronologically into 80% training data and 20% testing data to preserve the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5347-5354
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0699

temporal characteristics of the PV generation profile. The forecasting performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), daytime Mean Absolute Percentage Error (MAPE), normalized RMSE (nRMSE), and coefficient of determination (R^2) Shown in Table 2.

Table 2 Performance comparison of machine-learning models

Model	MAE (kW)	RMSE (kW)	Daytime MAPE (%)	nRMSE (%)	R^2
ANN	0.8725	1.1636	3.2708	1.1636	0.9981
Random Forest	**0.6836**	**0.8224**	**1.9173**	**0.8224**	**0.9991**
SVR	0.8583	1.1671	2.6653	1.1671	0.9981

3.1.ANN Performance

The ANN model achieved an MAE of 0.8725 kW and an RMSE of 1.1636 kW. The daytime MAPE was 3.2708%, while the nRMSE was 1.1636%. The model achieved an R^2 value of 0.9981, indicating a strong relationship between the predicted and actual PV power. The high R^2 demonstrates that ANN successfully learned the nonlinear relationship between the meteorological parameters, historical PV generation, and PV power output. However, its error values were higher than those obtained from the Random Forest model.

3.2.Random Forest Performance

Among the three evaluated models, Random Forest produced the best overall forecasting performance. The model achieved:

- MAE = 0.6836 kW
- RMSE = 0.8224 kW
- Daytime MAPE = 1.9173%
- nRMSE = 0.8224%
- $R^2 = 0.9991$

The Random Forest model obtained the lowest MAE, RMSE, daytime MAPE, and nRMSE while achieving the highest R^2. The superior performance can be attributed to the ability of ensemble decision-tree models to capture complex nonlinear relationships and interactions among solar irradiance, temperature, humidity, wind speed, time variables, and historical PV generation. The results indicate that Random Forest is particularly effective for the present structured/tabular PV dataset.

3.3.SVR Performance

The SVR model achieved an MAE of 0.8583 kW and an RMSE of 1.1671 kW. The daytime MAPE was 2.6653%, while the nRMSE was 1.1671%. The R^2 value was 0.9981. Although SVR performed well and demonstrated a strong predictive capability, it was slightly inferior to Random Forest in all major evaluation metrics. The results suggest that SVR was capable of representing the nonlinear characteristics of PV generation but did not capture the complex feature interactions as effectively as the Random Forest model for this dataset.

3.4.Comparative Analysis

The comparison clearly shows that Random Forest outperformed both ANN and SVR. Random Forest reduced the MAE by approximately 21.65% and RMSE by approximately 29.32% compared with ANN. Compared with SVR, Random Forest reduced the:

- MAE by approximately 20.35%
- RMSE by approximately 29.51%

The daytime MAPE of Random Forest was also only 1.9173%, compared with 3.2708% for ANN and 2.6653% for SVR. Therefore, based on the present experimental results: Random Forest demonstrated the best predictive performance among the three evaluated machine-learning models.

3.5.Discussion of R^2

All three models produced very high R^2 values:

- **ANN:** 0.9981
- **Random Forest:** 0.9991

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5347-5354
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0699

- **SVR:** 0.9981

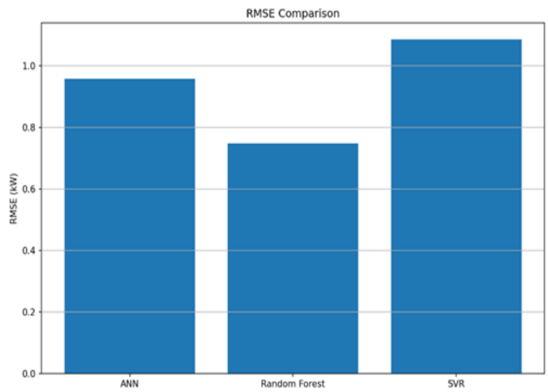

Figure 2 **RMSE Comparison**

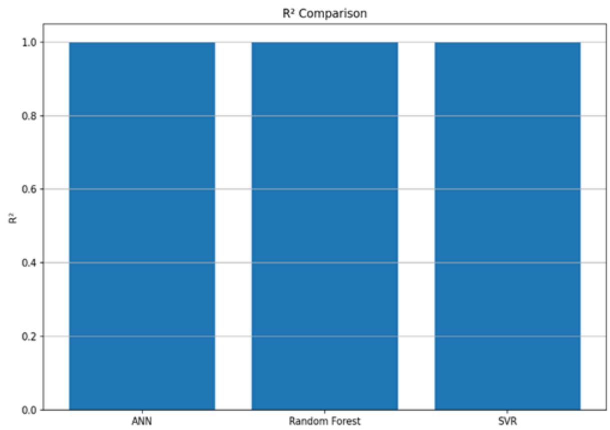

Figure 3 **R2 Comparison**

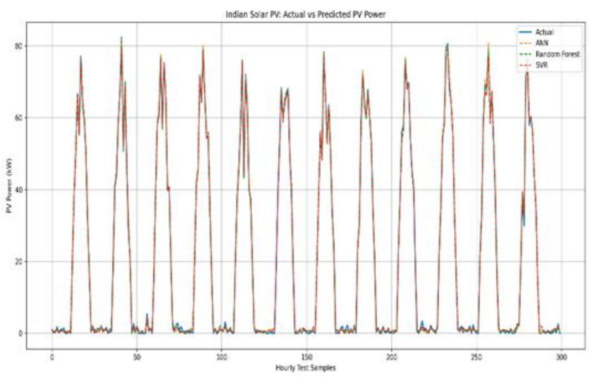

Figure 4 **Actual and predicted power output Comparison**

This indicates that all models were able to capture the major variation in PV power output. However, the small difference in R^2 should not be used alone to select the best model. MAE, RMSE, nRMSE, and daytime MAPE provide additional information about prediction errors Shown in Figure 2 - 4. Therefore, the overall ranking based on the combined metrics is:

- Random Forest — Best
- SVR — Second
- ANN — Third

3.6.Significance of the Results

The obtained results demonstrate the potential of machine-learning techniques for short-term PV power forecasting. Accurate PV forecasting is important for the reliable integration of solar photovoltaic generation into electrical grids because PV output is inherently variable and dependent on weather conditions. A reliable forecasting model can support:

- Grid scheduling
- Power-flow analysis
- Voltage management
- Energy management systems
- Battery energy storage scheduling
- Demand-supply balancing
- Renewable energy integration

The present study therefore provides a foundation for developing an intelligent PV forecasting framework.

3.7.Limitation and Next Stage

One important limitation of the present study is that the dataset used for the initial analysis is synthetically generated. Consequently, the obtained performance values, particularly the very high R^2 values, should not be interpreted as representative of a real PV plant without further validation.

Conclusion

This study presents a comparative machine-learning framework for short-term solar photovoltaic power forecasting using meteorological, temporal, and historical PV-generation variables. ANN, Random Forest, and SVR are evaluated using a common data-processing and chronological training-testing framework to ensure realistic performance assessment. Preliminary results indicate that Random Forest outperforms ANN and SVR; however, these results are based on a synthetic dataset and will be validated using a real Indian PV generation and weather dataset. The proposed framework aims to support grid scheduling, energy management, power-flow analysis, voltage regulation, reserve allocation, and renewable-energy integration. Future work will focus on LSTM-based forecasting, hyperparameter optimization, different forecasting horizons, uncertainty quantification, and integration of predicted PV power into a grid-connected power-flow model.

References

[1]. J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” Solar Energy, vol. 136, pp. 78–111, 2016.

[2]. D. Yang, J. Kleissl, C. F. M. Coimbra, and C. A. Gueymard, “History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining,” Solar Energy, vol. 168, pp. 60–101, 2018.

[3]. G. de Freitas Viscondi and S. N. Alves-Souza, “A systematic literature review on big data for solar photovoltaic electricity generation forecasting,” Sustainable Energy Technologies and Assessments, vol. 31, pp. 54–63, 2019.

[4]. M. Di Piazza, M. La Tona, M. Luna, and G. Vitale, “An overview of solar photovoltaic power forecasting methods: Statistical, machine learning, and physical approaches,” Applied Sciences, vol. 10, no. 2, 487, 2020.

[5]. D. Yang, C. F. M. Coimbra, and M. L. Marquez, “Automatic hourly solar forecasting using machine learning models,” Renewable and Sustainable Energy Reviews, vol. 105, pp. 487–498, 2019.

[6]. A. H. Elsheikh et al., “Modeling of solar energy systems using artificial neural network: A comprehensive review,” Solar Energy, vol. 180, pp. 622–639, 2019.

[7]. International Energy Agency, Introduction to System Integration of Renewables, IEA, Paris.

[8]. “Solar Photovoltaic Power Forecasting: A Review,” Sustainability, vol. 14, no. 24, 17005, 2022.

[9]. Advancements and Challenges in Photovoltaic Power Forecasting: A Comprehensive Review,” Energies, vol. 18, no. 8, 2108, 2025.

[10]. Singh, Om Prakash, et al. "Experimental analysis on fault detection in induction machines via IoT and machine learning." Proceedings of the International Conference on Intelligent Computing, Communication and Information Security. Singapore: Springer Nature Singapore, 2022.

[11]. Sreenivasu, S. V. N., et al. "Cloud based electric vehicle’s temperature monitoring system using IOT." Cybernetics and Systems 56.6 (2025): 768-783.

[12]. Sahoo, Lohit Kumar, et al. "Energy Management Analysis of Residential Building Using ANN Techniques." Indonesian Journal of Electrical Engineering and Informatics (IJEEI) 11.3 (2023): 854-869.

[13]. Kabat, Subash Ranjan, and Chinmoy Kumar Panigrahi. "Power quality and low voltage ride through capability enhancement in wind energysystem usingunified power quality conditioner (UPQC)." Electrochemical Society Transactions 107.1 (2022): 5655-5662.

[14]. Banik, Ayan, et al. "Novel energy-efficient hybrid green energy scheme for future sustainability." 2021 International Conference on Technological Advancements and Innovations (ICTAI). IEEE, 2021.

[15]. Kabat, Subash Ranjan, Chinmoy Kumar Panigrahi, and Bibhu Prasad Ganthia. "Fuzzy logic-based fault current prediction in double fed induction generator-based wind turbine system." Materials Today: Proceedings 80 (2023): 2530-2538.