AN INTEGRATED MACHINE LEARNING-BASED FRAMEWORK FOR WEATHER-DRIVEN SOLAR PHOTOVOLTAIC POWER FORECASTING AND INTELLIGENT ENERGY MANAGEMENT

Miss. SUBHAPRIYA PATTANAYAK
B. Tech Scholar
Department of Electrical and Electronics Engineering
Roland Institute of Technology, Berhampur

Mr. PINAKI BISOYI
Assistant Professor
Department of Electrical and Electronics Engineering
Roland Institute of Technology, Berhampur

Mr. SUBHAM MOHANTY
B. Tech Scholar
Department of Electrical and Electronics Engineering
Roland Institute of Technology, Berhampur

Mr. PRATIK RATH
B. Tech Scholar
Department of Electrical and Electronics Engineering
Roland Institute of Technology, Berhampur

Mr. SIBA SANKAR JENA
B. Tech Scholar
Department of Electrical and Electronics Engineering
Roland Institute of Technology, Berhampur

ABSTRACT - The increasing integration of renewable energy sources into modern power systems has created a need for accurate power forecasting and efficient energy management. This project presents a Solar Photovoltaic (PV) Power Prediction and Energy Management System that combines machine learning-based PV power forecasting with battery storage and grid energy management. A Random Forest Regression model is developed to predict PV power using important environmental and time-related parameters, including solar irradiance, temperature, humidity, wind speed, cloud cover, and temporal features. The proposed model is trained and evaluated using a 30-day hourly dataset containing 720 observations. The prediction model achieved an MAE of 0.0292 kW, RMSE of 0.0421 kW, and R2R^2 of 0.9976, demonstrating high prediction accuracy for the given dataset. Feature-importance analysis shows that solar irradiance is the dominant factor influencing PV power generation. The predicted PV power is subsequently supplied to an Energy Management System that coordinates energy flow among the PV system, electrical load, battery, and utility grid. Excess solar energy is used for battery charging and grid export, while battery discharge and grid import are utilized when PV generation is insufficient to satisfy the load. During the analyzed period, the system achieved 91.96 kWh of total PV generation, 81.17 kWh of load energy, 37.49 kWh of grid export, and 27.59 kWh of grid import, with a calculated energy cost of ₹108.23. The proposed framework demonstrates the potential of combining machine learning and energy management techniques to improve solar-energy utilization, reduce grid dependency, and support reliable and economical operation of PV-battery systems.

KEYWORDS - Solar Photovoltaic Power Forecasting, Intelligent Energy Management System, Random Forest Regression, Machine Learning, Meteorological Data Analysis, Renewable Energy Integration, Battery Energy Storage System, Solar Irradiance Prediction, Grid-Connected Photovoltaic System, Energy Optimization, Smart Grid, Renewable Energy Utilization.

I. INTRODUCTION

The increasing demand for electrical energy, rising environmental concerns, and the depletion of conventional energy resources have accelerated the adoption of renewable energy sources, particularly solar photovoltaic (PV) systems. Solar energy is clean, abundant, and widely available, making solar PV generation an important component of modern sustainable power systems. However, the output power of a solar PV system is highly dependent on environmental conditions such as solar irradiance, temperature, humidity, cloud cover, wind speed, and precipitation. These weather-dependent variations make solar power generation intermittent and difficult to predict accurately, which can create challenges in maintaining power system reliability, balancing supply and demand, and efficiently utilizing available energy.

Solar PV power prediction plays a significant role in addressing these challenges by estimating the amount of electricity that a PV system as shown in Fig. 1 is expected to generate over a particular period. Accurate prediction enables better scheduling of energy resources, improves grid stability, reduces dependence on conventional backup generation, and supports efficient operation of energy storage systems. Weather data provides valuable information for predicting PV output because parameters such as irradiance and temperature directly influence the electrical characteristics and power-producing capability of photovoltaic modules. By analyzing historical weather conditions along with corresponding PV generation data, predictive models can identify relationships between environmental variables and solar power output and subsequently estimate future generation.



Figure 1: A SOLAR PV SYSTEM

Energy management is another essential aspect of effective solar PV utilization. Since solar generation varies throughout the day and may not always coincide with the electrical load demand, an intelligent energy management system is required to coordinate PV generation, energy storage, grid power, and load demand. The predicted PV power can be used to determine appropriate energy management decisions, such as when to supply power directly to the load, when to charge or discharge a battery, and when to import or export electricity from the grid. This predictive approach can improve energy utilization, reduce unnecessary grid dependency, minimize energy losses, and enhance the overall efficiency and reliability of the PV-based power system.

The proposed system, therefore, focuses on the integration of solar PV power prediction with energy management using weather data. Weather parameters are collected and processed to develop a prediction model capable of forecasting PV power generation. The predicted output is then utilized by an energy management strategy to optimally coordinate the available energy sources and storage system according to generation and demand conditions. Such an integrated approach can contribute to more reliable and efficient utilization of solar energy and provide a foundation for intelligent renewable-energy systems, smart grids, and sustainable power management. By combining data-driven solar forecasting with effective energy management, the system aims to improve the utilization of renewable energy while maintaining a stable and economical supply of electricity.

II. LITERATURE REVIEW

Solar photovoltaic (PV) power prediction has become an important research area due to the intermittent nature of solar energy and the increasing penetration of PV systems into modern electrical networks. The power generated by a PV system depends strongly on environmental conditions, particularly solar irradiance and module temperature. A simplified relationship between PV output power and irradiance can be expressed as

$$P_{PV} = \eta_{PV} A G \quad (1)$$

where P_{PV} is the generated PV power, η_{PV} is the PV conversion efficiency, A is the effective area of the PV array, and G is the incident solar irradiance. However, PV efficiency is not constant and varies with cell temperature and operating conditions. Therefore, researchers have increasingly incorporated weather parameters such as irradiance, ambient temperature, humidity, wind speed, cloud cover, and rainfall into prediction models to improve forecasting accuracy.

Early studies on PV power prediction primarily employed statistical and mathematical techniques. Methods such as linear regression, multiple linear regression, autoregressive models, and autoregressive integrated moving average (ARIMA) models were used to establish relationships between historical weather conditions and PV output. Multiple linear regression, for example, can be represented as

$$P_{PV} = \beta_0 + \beta_1 G + \beta_2 T + \beta_3 H + \beta_4 W + \epsilon \quad (2)$$

where G represents solar irradiance, T is temperature, H is humidity, W is wind speed, β_i are regression coefficients, and ϵ represents prediction error. These approaches are relatively simple and computationally efficient, but their performance may decrease when the relationship between weather variables and PV generation becomes highly nonlinear.

With the development of machine learning, several researchers have investigated artificial neural networks (ANN), support vector regression (SVR), random forests, decision trees, and other data-driven techniques for PV power forecasting. ANN-based models are particularly useful because they can approximate complex nonlinear relationships between weather variables and PV generation. A simplified neural-network representation is

$$\hat{P}_{PV} = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \quad (3)$$

where x_i represents an input weather parameter, w_i is its corresponding weight, b is the bias, and $f(\cdot)$ is the activation function. Deep learning techniques, including recurrent neural networks (RNN), long short-term memory (LSTM), and convolutional neural networks (CNN), have further improved forecasting capabilities by learning temporal and spatial patterns from large historical datasets. LSTM models are particularly suitable for solar forecasting because PV generation at a given time is related to weather and generation conditions at previous time steps.

The accuracy of a PV prediction model is generally evaluated using statistical error measures. Commonly used performance indicators include mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). These can be expressed as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - \hat{P}_i|$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - \hat{P}_i)^2} \quad (4)$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{P_i - \hat{P}_i}{P_i}\right|$$

where P_i is the actual PV power, $\hat{P}_i$ is the predicted PV power, and N is the total number of observations. RMSE gives greater importance to larger prediction errors, while MAE provides a direct measure of the average absolute deviation between predicted and actual generation.

In addition to conventional machine learning and deep learning methods, recent research has explored hybrid and optimization-based forecasting approaches. Hybrid models combine two or more techniques to exploit their individual advantages. For example, weather-data preprocessing can be combined with ANN, LSTM, fuzzy logic, or optimization algorithms to improve prediction accuracy. Feature selection and data preprocessing are particularly important because weather datasets may contain missing values, measurement noise, redundant variables, and rapidly changing conditions. Normalization of input variables is commonly performed using

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (5)$$

which scales the input data into a suitable range for machine-learning algorithms.

Alongside PV power prediction, energy management has received considerable attention because accurate forecasting alone does not guarantee efficient utilization of generated solar energy. In a PV-based energy management system, the predicted PV generation is used to determine the operating states of the battery, local loads, and utility grid. The fundamental energy balance can be represented as

$$P_{PV} + P_{grid} + P_{dis} = P_{load} + P_{ch} \quad (6)$$

where P_{PV} is PV power, P_{grid} is grid power, P_{dis} is battery discharge power, P_{load} is load demand, and P_{ch} is battery charging power. The energy management controller attempts to maximize the utilization of renewable energy while maintaining battery constraints and satisfying the load demand.

Battery state of charge (SOC) is one of the most important variables in energy management. A simplified discrete-time battery model can be expressed as

$$SOC(k+1) = SOC(k) + \frac{\eta_{ch}P_{ch}(k)\Delta t}{E_{bat}} - \frac{P_{dis}(k)\Delta t}{\eta_{dis}E_{bat}} \quad (7)$$

where SOC(k) is the battery state of charge at time k, E_{bat} is the rated battery capacity, η_{ch} and η_{dis} are charging and discharging efficiencies, and Δt is the time interval. The battery must operate within predefined limits:

$$SOC_{min} \le SOC(k) \le SOC_{max} \quad (8)$$

Similarly, charging and discharging powers are restricted by

$$0 \le P_{ch} \le P_{ch,max}$$
$$0 \le P_{dis} \le P_{dis,max} \quad (9)$$

Research in predictive energy management has demonstrated that incorporating forecasted PV power can enable the controller to make decisions before changes in solar generation occur. For example, if the prediction model indicates high PV generation during the following period, the energy management system can prioritize local loads and battery charging. Conversely, when low solar generation is predicted, the system can preserve battery energy or schedule grid usage more effectively. This predictive approach is generally more effective than purely reactive energy management because it considers future generation conditions rather than responding only to instantaneous PV output.

Optimization techniques have also been widely considered for energy management. The objective may involve minimizing electricity cost, reducing grid dependency, maximizing renewable-energy utilization, and minimizing battery degradation. A general optimization objective can be formulated as

$$J = \sum_{k=1}^{N}[C_{grid}(k)P_{grid}(k) + C_{bat}(k)P_{bat}(k) + C_{curt}(k)P_{curt}(k)] \quad (10)$$

where C_{grid} represents the cost of grid electricity, C_{bat} represents a battery-related cost, C_{curt} represents the cost associated with PV curtailment, and P_{curt} is curtailed PV power. The optimization is subject to power-balance, battery SOC, charging/discharging, and operational constraints.

More recently, artificial intelligence and reinforcement learning (RL) have emerged as promising approaches for energy management because they can learn control policies from interactions with the energy system. In an RL-based system, the controller observes the current state, selects an action, and receives a reward according to the quality of that action. A possible state vector is

$$s_t = [P_{PV,t}, P_{load,t}, SOC_t, G_t, T_t, H_t] \quad (11)$$

where the state includes PV power, load demand, battery SOC, irradiance, temperature, and humidity. The action may be defined as

$$a_t = [P_{ch,t}, P_{dis,t}, P_{grid,t}] \quad (12)$$

and the objective is to maximize the cumulative discounted reward

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \quad (13)$$

where r_t is the instantaneous reward and γ is the discount factor. A suitable reward function can penalize grid electricity consumption, unmet load demand, excessive battery operation, and PV curtailment while rewarding renewable-energy utilization.

Overall, the literature indicates a clear progression from conventional statistical forecasting methods toward machine learning, deep learning, hybrid prediction, optimization, and intelligent energy-management techniques. Weather data plays a central role in improving PV forecasting because variations in irradiance, temperature, cloud conditions, humidity, and other environmental parameters directly influence PV generation. However, many existing approaches focus either on PV power prediction or on energy management independently. An integrated framework that uses weather-based PV forecasting as an input to an intelligent energy management system can provide better coordination between PV generation, battery storage, load demand, and grid power. Therefore, the proposed work on Solar PV Power Prediction and Energy Management Using Weather Data aims to combine accurate weather-based PV prediction with an effective energy management strategy to improve renewable-energy utilization, reduce grid dependency, and enhance the overall efficiency and reliability of the PV-based energy system.

III. PROBLEM STATEMENT

The increasing integration of solar photovoltaic (PV) systems into electrical power networks has created significant opportunities for clean and sustainable energy generation. However, solar PV power generation is inherently intermittent because it is strongly influenced by changing weather conditions such as solar irradiance, temperature, cloud cover, humidity, wind speed, and rainfall. Sudden variations in these parameters can cause considerable fluctuations in PV output, making it difficult to accurately predict the available solar power and effectively match generation with load demand. Inaccurate PV power prediction may result in inefficient utilization of renewable energy, unnecessary dependence on grid electricity, improper battery charging and discharging, and increased energy losses. Therefore, there is a need for a reliable prediction system that can utilize historical and real-time weather data to estimate future PV generation accurately. Furthermore, the predicted PV power should be integrated with an intelligent energy management strategy capable of coordinating PV generation, battery storage, electrical loads, and grid power. The major problem addressed in this work is therefore the development of an integrated solar PV power prediction and energy management system using weather data that can accurately forecast PV generation and use the prediction to make efficient energy-management decisions. The system aims to maximize solar energy utilization, minimize unnecessary grid dependence, maintain battery operation within safe limits, and ensure reliable supply of the connected load.

IV. PROPOSED METHODOLOGY

The proposed methodology presents an integrated Python-based framework for Solar PV Power Prediction and Intelligent Energy Management Using Weather Data. The system combines meteorological data analysis, machine-learning-based photovoltaic power forecasting, battery energy storage management, and optimized power distribution to improve renewable-energy utilization and reduce grid dependency. The methodology involves data acquisition, preprocessing, feature engineering, prediction model development, performance evaluation, and intelligent energy scheduling. Python libraries such as Pandas, NumPy, Scikit-learn, Matplotlib, and suitable optimization tools are employed to implement and evaluate the proposed framework.

4.1. Data Collection and Preprocessing

The initial stage involves collecting historical meteorological data and corresponding solar PV power generation measurements. The primary input parameters include solar irradiance, ambient temperature, relative humidity, wind speed, cloud cover, rainfall, and temporal information. The corresponding PV power output is considered the target variable for developing the prediction model. Data may be acquired from weather monitoring stations, photovoltaic installations, publicly available databases, or validated simulation environments.

The collected dataset undergoes preprocessing to improve its quality and consistency. Missing values, duplicate records, abnormal measurements, and inconsistent timestamps are identified and appropriately handled. Numerical features are normalized or standardized when required by the selected learning algorithm. The processed dataset is subsequently arranged in chronological order to preserve the temporal characteristics of photovoltaic generation.

4.2. Exploratory Data Analysis and Feature Engineering

Exploratory data analysis is performed to investigate the relationships between meteorological parameters and photovoltaic power generation. Statistical analysis and graphical visualization techniques are employed to examine variations in solar irradiance, temperature, humidity, and other environmental conditions. Correlation analysis helps identify influential parameters, while graphical

representations provide insights into daily and seasonal PV generation patterns.

Feature engineering is performed to extract meaningful information from the available measurements. Additional features, including hour of the day, month, seasonal indicators, historical PV generation, and lagged weather parameters, may be incorporated to improve prediction accuracy. Feature selection techniques are subsequently applied to retain relevant predictors and reduce unnecessary computational complexity.

4.3. Development of the Solar PV Prediction Model

A supervised machine-learning model is developed to predict photovoltaic power generation using meteorological and temporal input features. Suitable algorithms, including Random Forest Regression, Support Vector Regression, Gradient Boosting, and Artificial Neural Networks, are considered based on their ability to capture nonlinear relationships between weather conditions and solar power output.

The general prediction model is represented as

$$\hat{P}_{PV,t} = f_\theta(X_t) \tag{14}$$

The selected model is developed using Python-based machine-learning libraries, and its parameters are optimized to achieve reliable forecasting performance under varying meteorological conditions.

4.4. Model Training and Performance Evaluation

The processed dataset is chronologically divided into training, validation, and testing subsets. The training dataset is used to establish relationships between weather conditions and photovoltaic generation, while the validation dataset supports model tuning. The testing dataset evaluates prediction performance using previously unseen observations.

The forecasting accuracy is assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination R^2. Lower prediction errors and higher R^2 values indicate improved model performance. MAPE is evaluated only for suitable nonzero generation values to avoid numerical instability.

The predicted PV generation profiles are compared with actual measurements to examine forecasting accuracy, temporal variations, and model reliability. The best-performing model is subsequently integrated with the proposed energy management system.

4.5. Weather-Based Solar PV Power Forecasting

Following model training and validation, the developed prediction system estimates photovoltaic power generation using available weather forecasts and relevant historical observations. The forecasting module generates expected PV power profiles for the selected prediction horizon.

These forecasts provide advance information regarding solar energy availability under different environmental conditions. The predicted generation is subsequently supplied to the energy management module to support battery charging, discharging, and grid power scheduling. This enables the proposed system to anticipate variations in renewable-energy availability rather than relying exclusively on instantaneous power measurements.

4.6. Intelligent Energy Management System

The proposed energy management system coordinates photovoltaic generation, electrical load demand, battery storage, and utility grid interaction. The primary objective is to maximize solar energy utilization while maintaining reliable power supply and minimizing unnecessary grid consumption.

The power balance of the integrated system is expressed as

$$P_{PV} + P_{grid} + P_{dis} = P_{load} + P_{ch} + P_{curt} \tag{15}$$

During periods of excess solar generation, the available power is utilized to satisfy load demand and charge the battery. When solar generation becomes insufficient, the battery supplies the deficit within its permissible operating limits. If the available battery energy is inadequate, the remaining demand is supplied by the utility grid. Surplus PV energy may be exported where permitted or curtailed when storage capacity is unavailable.

4.7. Battery Energy Storage Management

Battery energy storage is incorporated to compensate for fluctuations in photovoltaic generation and improve the continuity of electrical supply. The battery management module monitors the state of charge (SOC) and regulates charging and discharging operations according to predicted solar generation and load requirements.

The battery SOC is estimated using

$$SOC_{t+1} = SOC_t + \frac{\eta_{ch} P_{ch,t} \Delta t}{E_B} - \frac{P_{dis,t} \Delta t}{\eta_{dis} E_B} \tag{16}$$

The battery controller maintains SOC within predefined operating limits and prevents simultaneous charging and discharging. By considering predicted PV generation, the proposed system can schedule battery operation more effectively and maintain stored energy for anticipated periods of reduced solar availability.

4.8. Optimization of Energy Distribution

An optimization-based scheduling mechanism is employed to determine efficient energy allocation among photovoltaic generation, battery storage, electrical loads, and the utility grid. The optimization objective is to minimize electricity procurement costs and unnecessary battery usage while promoting renewable-energy utilization.

The generalized objective function is expressed as

$$\min J = \sum_{t=1}^{T} [C_{grid,t} P_{imp,t} + C_{deg}(P_{ch,t} + P_{dis,t})] \, \Delta t \qquad (17)$$

The optimization is performed subject to power-balance requirements, battery SOC limits, charging and discharging constraints, and available PV generation. The resulting schedule determines suitable battery and grid operating conditions for the predicted energy availability.

4.9. Python-Based Simulation and Performance Analysis

The complete methodology is implemented in Python by integrating meteorological data processing, machine-learning prediction, battery modelling, and energy management optimization. Pandas and NumPy are utilized for data handling, Scikit-learn supports prediction model development, and Matplotlib and Seaborn are employed for performance visualization.

The simulation generates comparative plots of actual and predicted photovoltaic power, battery SOC, load demand, grid power exchange, and energy distribution. The overall system performance is evaluated using forecasting accuracy, grid energy consumption, solar self-consumption, battery utilization, and operating cost.

The proposed approach is compared with conventional forecasting and rule-based energy management strategies to assess the effectiveness of prediction-assisted energy scheduling under varying environmental and load conditions.

4.10. Overall Proposed Framework

The proposed framework establishes a coordinated relationship between weather-dependent photovoltaic forecasting and intelligent energy management. Meteorological observations are initially processed to develop a reliable prediction model, after which the estimated solar generation is utilized for battery scheduling and optimized power distribution.

The overall methodological sequence is summarized as follows:

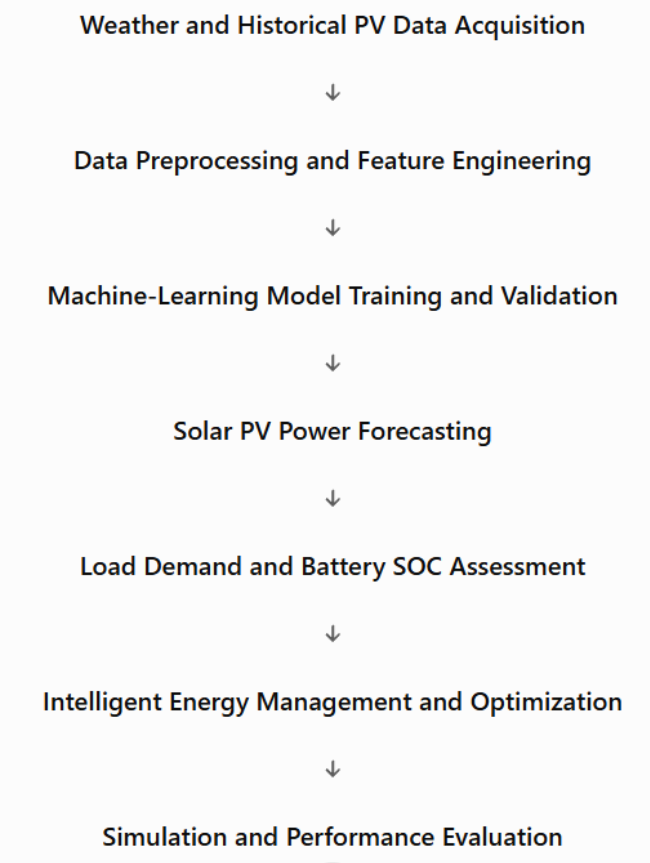


Figure 2: Overall methodological framework for weather-driven solar PV power forecasting and intelligent energy management.

The principal contribution of the proposed methodology is the integration of machine-learning-based photovoltaic power prediction with battery-aware energy scheduling and constrained power optimization. By utilizing anticipated solar generation to guide energy management decisions, the framework aims to improve renewable-energy utilization, reduce dependence on utility-grid electricity, and enhance the operational efficiency of photovoltaic energy systems. The proposed Python-based implementation provides a flexible and reproducible platform for evaluating the effectiveness of intelligent forecasting and energy management under variable meteorological conditions.

V. MODEL IMPLEMENTATION

The uploaded dataset, solar_pv_weather_dataset_30days (2).csv, as shown in Fig. 3 is a time-series dataset developed for analyzing solar photovoltaic (PV) power generation, weather conditions, and electrical load demand. It contains 720 observations and 8 variables. Since the dataset covers 30 days with measurements recorded at every hour, there are 24 observations per day, giving 30×24=72030\times24=720 records. The dataset therefore provides a continuous hourly representation of the solar-energy system over the 30-day period.

DateTime	Irradiance	Temperatu	Humidity_	WindSpee	CloudCove	PV_Power	Load_kW
01-01-2025 00:00	0	10.75	78.61	3.12	2.13	0	0.478
01-01-2025 01:00	0	9.16	80.99	2.73	5.67	0	0.532
01-01-2025 02:00	0	8.96	76.26	2.2	16.41	0.067	0.427
01-01-2025 03:00	0	8.31	75.57	2.22	10.37	0.014	0.38
01-01-2025 04:00	0	11.36	79.47	1.36	12.26	0	0.268
01-01-2025 05:00	0	12.24	80.6	3.72	8.69	0	0.391
01-01-2025 06:00	0	13.47	74.1	2.51	6.7	0	0.643
01-01-2025 07:00	51.88	12.38	69.47	4.06	69.56	0.215	0.493
01-01-2025 08:00	286.26	16.5	65.49	1.84	11.73	1.431	0.671
01-01-2025 09:00	330.35	21.42	59.54	3.04	33.13	1.541	0.409
01-01-2025 10:00	358.92	21.2	59.12	3.94	54.62	1.748	0.447
01-01-2025 11:00	666.39	24.09	54.63	2.7	14.66	3.012	0.435
01-01-2025 12:00	731.99	23.26	61.66	3.3	9.63	3.273	0.463
01-01-2025 13:00	678.41	27.45	62.26	3.74	17.65	3.018	0.481
01-01-2025 14:00	275.3	25.83	69.1	4.73	60.54	1.189	0.528
01-01-2025 15:00	457.15	26.08	71.58	1.93	11.81	2.101	0.391
01-01-2025 16:00	263.93	27.72	82.37	3.45	30.82	1.226	0.431
01-01-2025 17:00	115.78	25.48	73.81	2.02	13.4	0.587	0.492
01-01-2025 18:00	0	21.63	77.07	3.31	4.19	0	0.889
01-01-2025 19:00	0	19.36	76.95	2.86	11.63	0.046	0.923
01-01-2025 20:00	0	17.61	76.25	3.14	10.19	0.033	0.916
01-01-2025 21:00	0	16.54	75.57	3.06	10.22	0.05	0.954

Figure 3: SCREENSHOT OF THE UPLOADED DATASET FOR SOLAR PV POWER PREDICTION

The dataset utilized in the proposed study comprises comprehensive meteorological observations, photovoltaic power generation measurements, and electrical load demand information collected over a continuous 30-day period, extending from January 1, 2025, to January 30, 2025. The observations are recorded at hourly intervals, resulting in 720 time-indexed records. The dataset incorporates eight essential attributes representing temporal characteristics, solar irradiance, ambient temperature, relative humidity, wind speed, cloud cover, photovoltaic power generation, and electrical energy demand. These parameters collectively provide the foundation for developing a weather-driven solar power forecasting model and an intelligent energy management framework.

Temporal information plays a significant role in characterizing the periodic behavior of photovoltaic generation. The hourly observations enable the identification of daily solar generation patterns, variations in environmental conditions, and changes in electricity demand. The chronological arrangement of the dataset also facilitates the extraction of relevant temporal characteristics, allowing the prediction model to recognize variations in solar energy availability throughout the day. Furthermore, maintaining the chronological sequence supports reliable model training and evaluation without introducing information from future observations.

Solar irradiance represents one of the most influential meteorological parameters affecting photovoltaic power generation. The recorded irradiance varies between 0 and 808.96 watts per square metre, with an average value of approximately 148.79 watts per square metre. The presence of zero irradiance values primarily corresponds to nighttime periods when solar energy is unavailable. Variations in irradiance directly influence the electrical energy generated by photovoltaic modules, making this parameter particularly important for forecasting solar power under changing atmospheric conditions.

Ambient temperature is another significant environmental factor influencing photovoltaic performance. The recorded temperature varies from 4.17℃ to 31.72℃, with an average of approximately 17.87℃. Although solar irradiance primarily determines the availability of incident solar energy, temperature influences photovoltaic conversion efficiency through its effect on module operating conditions. Elevated module temperatures can reduce electrical conversion efficiency, whereas favorable thermal conditions may support improved performance. Incorporating ambient temperature therefore enables the forecasting model to account for temperature-related variations in photovoltaic generation.

Relative humidity is included to represent atmospheric moisture conditions that may influence solar radiation availability. The recorded humidity ranges from 53.40% to 91.25%, with an average value of approximately 73.08%. Variations in atmospheric humidity can be associated with changes in cloud formation, atmospheric absorption, and prevailing weather conditions. Although humidity does not directly determine photovoltaic output, its inclusion provides additional environmental information that may improve the model's ability to recognize complex meteorological relationships.

Wind speed is considered because of its potential influence on photovoltaic module cooling and thermal performance. The recorded wind speed ranges from 1.33 to 4.88 metres per second, with an average of approximately 2.99 metres per second. Air movement can facilitate convective heat transfer from photovoltaic module surfaces, thereby influencing operating temperature and electrical conversion efficiency. The inclusion of wind speed allows the prediction framework to consider the combined effects of environmental temperature and atmospheric movement on photovoltaic performance.

Cloud cover is incorporated as an important indicator of atmospheric conditions influencing the availability of solar radiation. The recorded cloud cover varies between 0% and 71%, with an average value of approximately 23.87%. Increased cloud coverage generally reduces the direct solar radiation reaching photovoltaic surfaces, although the resulting effect depends on cloud characteristics and prevailing atmospheric conditions. Since variations in cloud cover can produce significant fluctuations in photovoltaic generation, this parameter contributes valuable information for developing a weather-responsive forecasting system.

Photovoltaic power generation constitutes the principal prediction target of the proposed machine-learning framework. The recorded electrical power varies from 0 to

3.598 kilowatts, with an average generation of approximately 0.704 kilowatts. These variations reflect changes in solar energy availability and environmental operating conditions throughout the observation period. The forecasting model utilizes meteorological and temporal information to identify relationships between atmospheric conditions and photovoltaic output, enabling the estimation of solar power generation for previously unseen observations.

Electrical load demand represents an essential component of the proposed energy management framework. The recorded demand varies between 0.250 and 1.052 kilowatts, with an average of approximately 0.559 kilowatts. Unlike photovoltaic generation, electrical demand is not considered the primary prediction target; instead, it serves as an operational input for determining energy distribution requirements. By comparing anticipated photovoltaic generation with electrical demand, the energy management system can identify periods of surplus and deficient solar power and coordinate energy supply through direct photovoltaic utilization, battery charging or discharging, and grid interaction.

Table 1. Descriptive statistical characteristics of the meteorological, photovoltaic generation, and electrical load dataset.

Parameter	Minimum	Maximum	Mean
Solar irradiance (W/m²)	0	808.96	148.79
Ambient temperature (°C)	4.17	31.72	17.87
Relative humidity (%)	53.40	91.25	73.08
Wind speed (m/s)	1.33	4.88	2.99
Cloud cover (%)	0	71	23.87
PV power generation (kW)	0	3.598	0.704
Electrical load demand (kW)	0.250	1.052	0.559

Overall, the dataset provides a structured representation of the interaction between meteorological conditions, photovoltaic electricity generation, and electrical load requirements. The combination of environmental measurements and operational energy information supports the development of an integrated forecasting and energy management framework. By incorporating these parameters into the proposed machine-learning approach, the study aims to improve photovoltaic power estimation and enable more informed energy distribution decisions. The dataset also establishes a consistent basis for evaluating forecasting accuracy, renewable-energy utilization, battery storage operation, and grid dependency under varying environmental and demand conditions.

The proposed system establishes a comprehensive computational framework that integrates weather-driven photovoltaic power forecasting with intelligent battery storage management and grid-connected energy distribution. The implementation employs Random Forest Regression to estimate photovoltaic generation from meteorological and temporal information, followed by a rule-based energy management strategy that coordinates solar generation, electrical demand, battery storage, and utility-grid interaction.

The operational framework begins with dataset acquisition and systematic preprocessing, followed by exploratory analysis, temporal feature extraction, and chronological partitioning of observations. A Random Forest model comprising 300 decision trees is trained using 80% of the available observations, while the remaining 20% is reserved for independent performance assessment. The forecasting accuracy is evaluated through established statistical indicators, and the resulting predictions are subsequently incorporated into the energy management module.

The energy management strategy prioritizes direct photovoltaic utilization, followed by battery charging during surplus generation and battery discharging during energy deficits. Grid import is employed when available renewable generation and permissible battery discharge cannot satisfy electrical demand, whereas surplus electricity is exported after battery charging requirements have been fulfilled. Battery operation is regulated within prescribed state-of-charge and power limits to maintain operational feasibility.

The complete implementation concludes with quantitative evaluation of forecasting performance, renewable-energy utilization, grid dependency, battery operation, and electricity expenditure. Graphical visualization and automated export of simulation results provide a reproducible basis for assessing the effectiveness of the integrated photovoltaic forecasting and energy management framework.

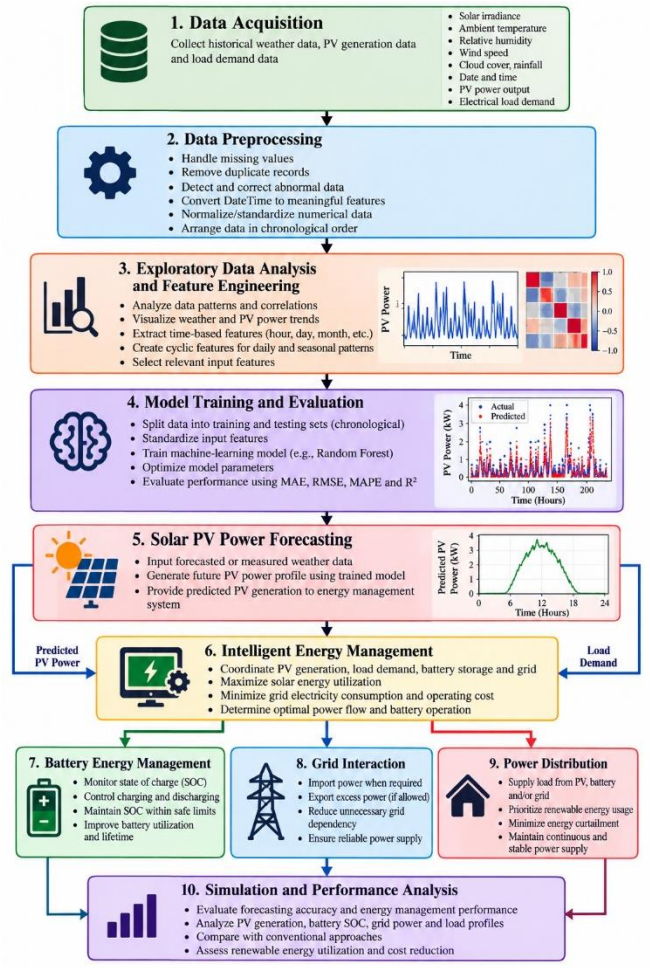
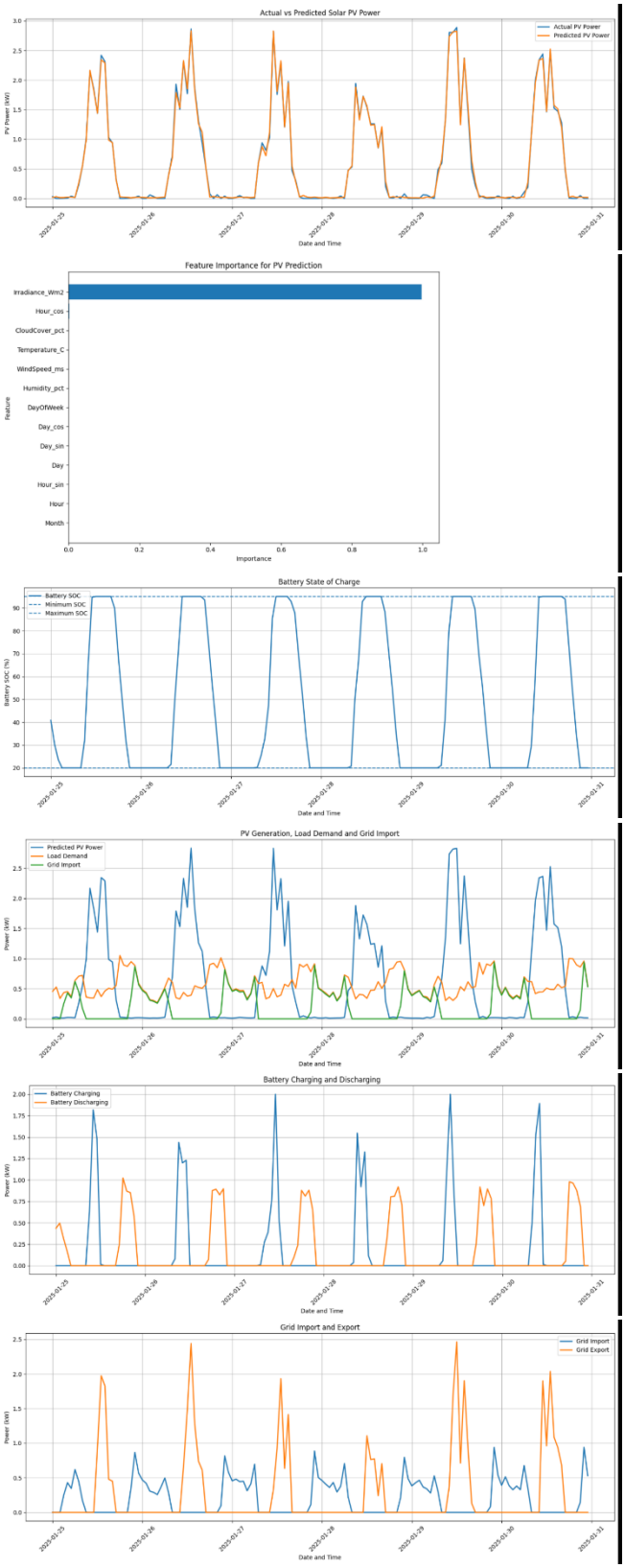


Figure 4: Comprehensive operational flowchart of the proposed weather-driven solar photovoltaic power forecasting and intelligent energy management system incorporating Random Forest Regression, battery storage coordination, electrical load management, and utility-grid interaction.

The flowchart illustrates the sequential integration of meteorological data acquisition, preprocessing, feature engineering, machine-learning-based forecasting, and intelligent energy distribution. The framework employs a Random Forest Regression model to estimate photovoltaic generation, which subsequently guides battery charging and discharging, grid electricity import and export, and load supply decisions. The final simulation stage evaluates forecasting accuracy, battery operating characteristics, renewable-energy utilization, grid dependency, and electricity expenditure, providing a comprehensive assessment of the proposed system's technical and operational performance.

Journal accuracy note: The supplied implementation uses a rule-based energy management strategy rather than a formal mathematical optimization solver. Accordingly, the framework should be described as prediction-assisted intelligent energy management unless a separate optimization algorithm is implemented.

VI. RESULTS AND DISCUSSION



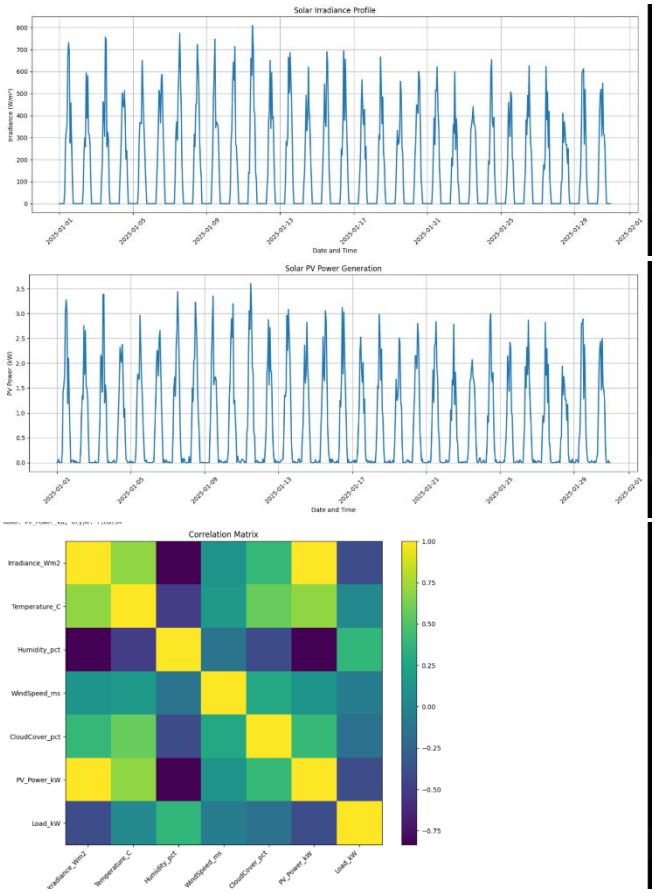


Figure 5: SCREENSHOT OF THE OUTPUT OF THE PYTHON CODE IMPLEMENTING SOLAR PV POWER PREDICTION AND ENERGY MANAGEMENT USING WEATHER DATA

The output of the proposed Solar PV Power Prediction and Energy Management System indicates that the complete program has executed successfully, starting from dataset uploading and preprocessing to PV power prediction, model evaluation, energy management, correlation analysis, visualization, and result export. The uploaded dataset contains 720 observations and 8 variables, namely DateTime, solar irradiance, temperature, humidity, wind speed, cloud cover, PV power, and electrical load. The program confirms that all the required columns are available and that preprocessing was completed successfully without reducing the dataset size.

The statistical analysis of the dataset shows that the average solar irradiance is 148.79 W/m², with a maximum value of 808.96 W/m². The average temperature is approximately 17.87°C, while the average humidity is 73.08% and the average wind speed is 2.99 m/s. The average cloud cover is 23.87%. The actual PV power has an average value of 0.704 kW, with a maximum generation of 3.598 kW, whereas the electrical load has an average value of 0.559 kW and a maximum of 1.052 kW. These values indicate that the PV generation varies significantly with solar availability, while the electrical load remains present even during periods of low or zero solar generation.

For machine-learning development, the dataset was divided into 576 training samples and 144 testing samples, corresponding to an 80:20 chronological train-test split. The first 80% of the data was used to train the Random Forest Regression model, while the remaining 20% was used to evaluate its prediction performance. Feature engineering was also completed before model training, allowing the DateTime information to be transformed into useful temporal characteristics.

The Random Forest model produced very good prediction results. The obtained Mean Absolute Error (MAE) is 0.0292 kW, which means that the average absolute difference between the actual and predicted PV power is approximately 0.0292 kW. The Root Mean Squared Error (RMSE) is 0.0421 kW, indicating that the model has relatively small prediction errors, including when larger errors are given greater importance. The R^2 value is 0.9976, which indicates that the model explains approximately 99.76% of the variation in PV power within the testing data. These results demonstrate that the Random Forest model provides a very strong prediction of PV power for this dataset.

The MAPE value is 43.19%, which is considerably higher than the other error indicators. This value should be interpreted carefully because the dataset contains many observations where PV generation is zero or very small, particularly during nighttime and periods of low solar irradiance. Percentage-based errors can become very large when the actual value is close to zero. Therefore, the high MAPE does not necessarily contradict the very low MAE and RMSE or the high R^2 value. For this dataset, the MAE, RMSE, and R^2 provide a more representative indication of the overall prediction performance, while MAPE should be considered with caution.

The feature-importance results show that solar irradiance is by far the most influential variable for predicting PV power. Its importance value is 0.996737, while the next most important feature, Hour_cos, has an importance of only 0.002142. Cloud cover, temperature, wind speed, humidity, day of the week, and other temporal variables have very small importance values in comparison. This result indicates that the Random Forest model relies primarily on irradiance to determine PV generation. The result is also consistent with the physical behavior of photovoltaic systems because the amount of solar radiation reaching the PV modules has a direct and dominant influence on electricity generation.

The Energy Management System results show that the total PV generation during the analyzed period was 91.96 kWh, while the total electrical load energy was 81.17 kWh. The system directly utilized 30.78 kWh of PV energy to supply

the load. In addition, 23.68 kWh was used for battery charging and 22.80 kWh was obtained through battery discharging. The system imported 27.59 kWh from the electrical grid when the available PV and battery energy were insufficient to meet the load, while 37.49 kWh of surplus energy was exported to the grid.

The PV utilization is reported as 33.48%, which indicates that approximately one-third of the total PV generation was directly utilized according to the definition used in the program. The remaining PV generation was associated with battery charging and grid export. Therefore, the lower direct PV utilization does not necessarily indicate that the remaining solar energy was wasted; rather, a considerable portion was stored or exported. This demonstrates the importance of battery storage and grid interaction in managing excess solar generation.

The grid dependency is 33.99%, meaning that approximately one-third of the total load energy was supplied through grid imports according to the program's calculation. The remaining energy demand was supported by the PV system and battery. This indicates that the combination of PV generation and battery storage reduced the dependence on the utility grid, although the system still required grid energy during periods when PV generation and available battery energy were insufficient.

The calculated total energy cost is ₹108.23. This value is determined by considering the cost of electricity imported from the grid and the financial benefit obtained from exporting surplus PV energy to the grid. The program therefore considers both grid purchases and grid-export revenue when determining the net energy cost. This provides an indication of the economic performance of the proposed PV-battery-grid system under the assumed electricity import and export prices.

The final battery State of Charge (SOC) is 20%, which corresponds exactly to the minimum SOC limit specified in the Energy Management System. This indicates that the battery was significantly utilized during the simulation and eventually reached its minimum permitted operating level. Once the battery reaches this limit, the Energy Management System prevents further discharge so that the battery does not operate below the defined SOC constraint.

The correlation analysis further supports the importance of solar irradiance in PV power prediction. The correlation coefficient between irradiance and PV power is 0.999085, indicating an extremely strong positive relationship. Temperature has a correlation of 0.697387, showing a moderately strong positive relationship with PV power, while humidity has a correlation of −0.834895, indicating a strong negative relationship in this dataset. Cloud cover has a correlation of 0.393587, wind speed has a relatively weak positive correlation of 0.115310, and load has a negative correlation of −0.399472 with PV power. These are statistical relationships within this dataset and should not be interpreted as proof of direct causation.

Overall, the output demonstrates that the proposed system successfully integrates machine-learning-based PV power prediction with battery and grid energy management. The Random Forest model achieves highly accurate prediction performance, particularly demonstrated by the low MAE and RMSE and the very high R^2 value. The feature-importance and correlation analyses identify solar irradiance as the dominant factor influencing PV generation. The Energy Management System then uses the predicted PV power to supply the load, charge and discharge the battery, import energy from the grid when required, and export excess solar energy when available. The final results, therefore, demonstrate the capability of the proposed system to perform both accurate solar PV forecasting and effective energy-flow management.

VII. CONCLUSION

In conclusion, the proposed Solar PV Power Prediction and Energy Management System successfully integrates weather-based solar power prediction with intelligent energy management involving the PV system, battery storage, electrical load, and utility grid. The developed system uses a Random Forest Regression model to predict PV power from solar irradiance, temperature, humidity, wind speed, cloud cover, and time-related features. The results demonstrate excellent prediction performance, achieving an MAE of 0.0292 kW, RMSE of 0.0421 kW, and an R^2 value of 0.9976, indicating that the model can accurately reproduce the PV generation pattern for the given dataset. The analysis also identifies solar irradiance as the dominant factor influencing PV generation, with a feature importance of 0.996737 and a correlation of 0.999085 with PV power.

The developed Energy Management System effectively manages the predicted solar energy according to the instantaneous load and battery conditions. During periods of excess PV generation, solar energy is first utilized to satisfy the load, followed by battery charging and grid export. During periods of insufficient PV generation, the battery is discharged to support the load, and the remaining demand is supplied by the grid. For the analyzed period, the system produced 91.96 kWh of PV energy against a total load demand of 81.17 kWh, while 23.68 kWh was used for battery charging, 22.80 kWh was supplied through battery discharge, 27.59 kWh was imported from the grid, and 37.49 kWh was exported to the grid. The system achieved a 33.99% grid dependency and a calculated net energy cost of

₹108.23, demonstrating the economic and operational benefits of coordinated PV, battery, and grid operation.

Overall, the results confirm that the proposed approach provides a reliable framework for solar PV forecasting and energy management. Accurate PV prediction enables better planning of energy flows, while battery storage helps shift excess solar energy to periods of lower solar availability and reduces dependence on the grid. Although the reported MAPE of 43.19% is relatively high because percentage errors become sensitive when actual PV generation is zero or very small, the low MAE and RMSE and exceptionally high R^2 demonstrate strong overall predictive performance. Therefore, the proposed system can serve as a useful foundation for developing intelligent, cost-effective, and more self-sufficient solar PV energy systems, with potential for further improvement through larger datasets, real-time measurements, advanced optimization or reinforcement-learning-based energy management, and more detailed battery and PV system modeling.

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] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," Renewable Energy, vol. 105, pp. 569–582, 2017.

[3] M. Q. Raza, M. Nadarajah, and C. Ekanayake, "On recent advances in PV output power forecast," Solar Energy, vol. 136, pp. 125–144, 2016.

[4] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[5] T. G. Dietterich, "Ensemble methods in machine learning," in Multiple Classifier Systems. Berlin, Germany: Springer, 2000, pp. 1–15.

[6] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[7] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, pp. 273–297, 1995.

[8] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," Progress in Energy and Combustion Science, vol. 34, no. 5, pp. 574–632, 2008.

[9] A. Mellit, A. M. Pavan, and M. Benghanem, "Least squares support vector machine for short-term prediction of meteorological time series," Theoretical and Applied Climatology, vol. 111, pp. 297–307, 2013.

[10] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 2, no. 1, pp. 2–10, 2009.

[11] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," Progress in Energy and Combustion Science, vol. 39, no. 6, pp. 535–576, 2013.

[12] H. T. C. Pedro and C. F. M. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," Solar Energy, vol. 86, no. 7, pp. 2017–2028, 2012.

[13] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," IEEE Transactions on Industry Applications, vol. 48, no. 3, pp. 1064–1069, 2012.

[14] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," Solar Energy, vol. 83, no. 10, pp. 1772–1783, 2009.

[15] C. Monteiro, T. Santos, L. A. Fernandez-Jimenez, I. J. Ramirez-Rosado, and M. S. Terreros-Olarte, "Short-term power forecasting model for photovoltaic plants based on historical similarity," Energies, vol. 6, no. 5, pp. 2624–2643, 2013.

[16] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," Renewable and Sustainable Energy Reviews, vol. 27, pp. 65–76, 2013.

[17] J. A. Ruiz-Arias, H. Alsamamra, J. Tovar-Pescador, and D. Pozo-Vazquez, "Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions," Energy Conversion and Management, vol. 51, no. 5, pp. 881–893, 2010.

[18] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, "Photovoltaic and solar forecasting: State of the art," IEA PVPS Task 14, Rep. IEA-PVPS T14-01:2013, 2013.

[19] D. Yang, J. Kleissl, C. A. Gueymard, H. T. C. Pedro, and C. F. M. Coimbra, "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.

[20] D. Yang et al., "A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality," Renewable and Sustainable Energy Reviews, vol. 161, Art. no. 112348, 2022.

[21] G. M. Masters, Renewable and Efficient Electric Power Systems, 2nd ed. Hoboken, NJ, USA: Wiley, 2013.

[22] R. Messenger and J. Ventre, Photovoltaic Systems Engineering, 3rd ed. Boca Raton, FL, USA: CRC Press, 2010.

[23] S. A. Kalogirou, Solar Energy Engineering: Processes and Systems, 2nd ed. Oxford, U.K.: Academic Press, 2014.

[24] M. A. Green, Solar Cells: Operating Principles, Technology, and System Applications. Englewood Cliffs, NJ, USA: Prentice-Hall, 1982.

[25] A. Luque and S. Hegedus, Eds., Handbook of Photovoltaic Science and Engineering, 2nd ed. Chichester, U.K.: Wiley, 2011.

[26] T. Esram and P. L. Chapman, "Comparison of photovoltaic array maximum power point tracking techniques," IEEE Transactions on Energy Conversion, vol. 22, no. 2, pp. 439–449, 2007.

[27] M. A. Eltawil and Z. Zhao, "Grid-connected photovoltaic power systems: Technical and potential problems—A review," Renewable and Sustainable Energy Reviews, vol. 14, no. 1, pp. 112–129, 2010.

[28] J. M. Carrasco et al., "Power-electronic systems for the grid integration of renewable energy sources: A survey," IEEE Transactions on Industrial Electronics, vol. 53, no. 4, pp. 1002–1016, 2006.
[29] J. P. Barton and D. G. Infield, "Energy storage and its use with intermittent renewable energy," IEEE Transactions on Energy Conversion, vol. 19, no. 2, pp. 441–448, 2004.
[30] H. Ibrahim, A. Ilinca, and J. Perron, "Energy storage systems—Characteristics and comparisons," Renewable and Sustainable Energy Reviews, vol. 12, no. 5, pp. 1221–1250, 2008.
[31] B. Dunn, H. Kamath, and J.-M. Tarascon, "Electrical energy storage for the grid: A battery of choices," Science, vol. 334, no. 6058, pp. 928–935, 2011.
[32] A. Nottrott, J. Kleissl, and B. Washom, "Energy dispatch schedule optimization and cost benefit analysis for grid-connected, photovoltaic-battery storage systems," Renewable Energy, vol. 55, pp. 230–240, 2013.
[33] M. R. Alam, M. St-Hilaire, and T. Kunz, "Computational methods for residential energy cost optimization in smart grids: A survey," ACM Computing Surveys, vol. 49, no. 1, Art. no. 2, 2016.
[34] A. Parisio, E. Rikos, and L. Glielmo, "A model predictive control approach to microgrid operation optimization," IEEE Transactions on Control Systems Technology, vol. 22, no. 5, pp. 1813–1827, 2014.
[35] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," IEEE Transactions on Industrial Informatics, vol. 7, no. 3, pp. 381–388, 2011.
[36] A. S. O. Ogunjuyigbe, T. R. Ayodele, and O. A. Akinola, "Optimal allocation and sizing of PV/wind/split-diesel/battery hybrid energy system for minimizing life cycle cost, carbon emission and dump energy of remote residential building," Applied Energy, vol. 171, pp. 153–171, 2016.
[37] F. Marra, G. Yang, C. Træholt, J. Østergaard, and E. Larsen, "A decentralized storage strategy for residential feeders with photovoltaics," IEEE Transactions on Smart Grid, vol. 5, no. 2, pp. 974–981, 2014.
[38] M. L. Tuballa and M. L. Abundo, "A review of the development of Smart Grid technologies," Renewable and Sustainable Energy Reviews, vol. 59, pp. 710–725, 2016.
[39] H. Lund, A. N. Andersen, P. A. Østergaard, B. V. Mathiesen, and D. Connolly, "From electricity smart grids to smart energy systems—A market operation based approach and understanding," Energy, vol. 42, no. 1, pp. 96–102, 2012.
[40] J. Weniger, T. Tjaden, and V. Quaschning, "Sizing of residential PV battery systems," Energy Procedia, vol. 46, pp. 78–87, 2014.