Article

Innovative Approaches in Residential Solar Electricity: Forecasting and Fault Detection Using Machine Learning

Shruti Kalra [1,†], Ruby Beniwal [1,*,†], Vinay Singh [1,‡] and Narendra Singh Beniwal [2,‡]

1 Department of Electronics and Communication, Jaypee Institute of Information Technology, Noida 201307, UP, India; shruti.kalra@jiit.ac.in (S.K.); 20402009@mail.jiit.ac.in (V.S.)
2 Department of Electronic and Communication Engineering, Bundelkhand Institute of Engineering & Technology, Jhansi 284128, UP, India; nsb@bietjhs.ac.in
* Correspondence: ruby.beniwal@jiit.ac.in
† Developed forecasting and fault detection models.
‡ Estimated residential solar electricity.

Abstract: Recent advancements in residential solar electricity have revolutionized sustainable development. This paper introduces a methodology leveraging machine learning to forecast solar panels' power output based on weather and air pollution parameters, along with an automated model for fault detection. Innovations in high-efficiency solar panels and advanced energy storage systems ensure reliable electricity supply. Smart inverters and grid-tied systems enhance energy management. Government incentives and decreasing installation costs have increased solar power accessibility. The proposed methodology, utilizing machine learning techniques, achieved an R-squared value of 0.95 and a Mean Squared Error of 0.02 in forecasting solar panel power output, demonstrating high accuracy in predicting energy production under varying environmental conditions. By improving operational efficiency and anticipating power output, this approach not only reduces carbon footprints but also promotes energy independence, contributing to the global transition towards sustainability.

Keywords: residentialsolar electricity; machine learning; forecasting; fault detection; sustainable development

1. Introduction

The transition toward sustainable energy solutions is increasingly urgent as the world grapples with the effects of climate change and the depletion of fossil fuels. The reliance on traditional energy sources, such as coal, oil, and natural gas, has led to significant environmental degradation, including greenhouse gas emissions, air and water pollution, and habitat destruction. These issues have prompted a global call for cleaner, more sustainable energy alternatives [1–3]. Among these alternatives, residential solar electricity has emerged as a crucial component in the shift towards renewable energy, offering a renewable, environmentally friendly solution to meet the growing energy demands. Residential solar electricity systems harness the power of the sun to generate electricity, reducing dependence on fossil fuels and lowering carbon footprints. Over the past few decades, solar power systems for residential use have undergone significant advancements, making them more efficient, cost-effective, and accessible to the general public. The core component of these systems, the solar panel, has seen dramatic improvements in efficiency. Technological innovations have resulted in higher-efficiency panels that can convert more sunlight into electricity. This not only increases the amount of energy generated but also reduces the overall cost of solar power [4,5].

Photovoltaic (PV) technology has evolved rapidly, with the development of various types of solar cells, including monocrystalline, polycrystalline, and thin-film cells. Monocrystalline panels, known for their high efficiency and longevity, are made from

single-crystal silicon, allowing them to convert more sunlight into electricity. Polycrystalline panels, while slightly less efficient, offer a cost-effective alternative. Thin-film panels, made from a variety of materials such as cadmium telluride and amorphous silicon, are lightweight and flexible, making them suitable for a wide range of applications. Recent advancements have also introduced bifacial panels, which can capture sunlight on both sides, further enhancing energy yield [6,7]. The development of advanced energy storage systems, such as lithium-ion batteries, addresses the intermittency of solar power by allowing excess energy to be stored and used when sunlight is not available. These storage systems are critical for ensuring a reliable and continuous power supply. Lithium-ion batteries, known for their high energy density, long cycle life, and decreasing costs, have become the preferred choice for residential energy storage. They enable homeowners to store surplus electricity generated during sunny periods and use it during cloudy days or at night, providing a consistent power supply and reducing dependence on the grid [8].

Figure 1 illustrates various energy storage systems, emphasizing lithium-ion, solid-state, and flow batteries. Lithium-ion batteries, known for their high energy density, long cycle life, and decreasing costs, are ideal for residential energy storage, allowing homeowners to store surplus electricity and use it during cloudy days or at night. Solid-state batteries, with a solid electrolyte, offer enhanced safety and higher energy density. Flow batteries use liquid electrolytes, providing scalability and long cycle life, making them suitable for larger installations [9]. These systems collectively ensure a reliable and continuous power supply, reducing dependence on the grid.
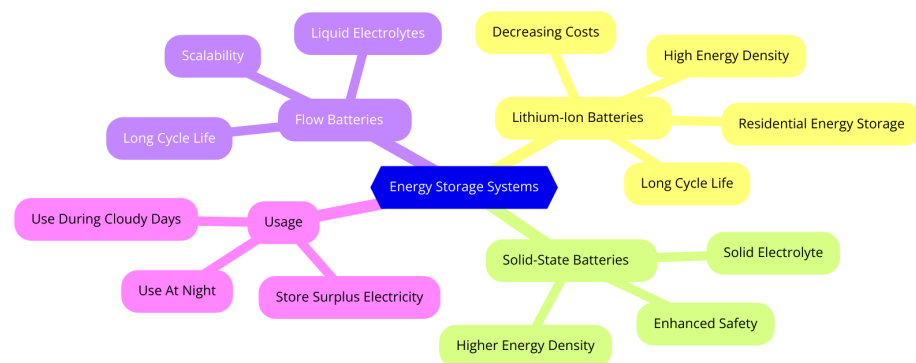
**Figure 1.** Energy storage systems emphasizing lithium-ion, solid-state, and flow batteries.

Other emerging storage technologies, such as solid-state batteries and flow batteries, offer potential improvements in energy density, safety, and cost. Solid-state batteries replace the liquid or gel electrolyte found in conventional batteries with a solid electrolyte, enhancing safety and energy density. Flow batteries, which store energy in liquid electrolytes contained in external tanks, provide scalability and long cycle life, making them suitable for larger installations. Smart inverters and grid-tied systems represent another leap forward in solar technology. Smart inverters improve energy management by converting the direct current (DC) produced by solar panels into alternating current (AC) used by home appliances and the grid [10]. These inverters also offer advanced features such as reactive power control, voltage regulation, and grid support functionalities, enhancing grid stability and efficiency. Grid-tied systems enable the integration of solar power with existing power grids, allowing for efficient energy distribution and the possibility for homeowners to sell excess electricity back to the grid through net metering programs. These innovations enhance the practicality and economic viability of residential solar electricity, promoting energy independence and resilience.

Government incentives and decreasing installation costs have further accelerated the adoption of solar power. Financial incentives, such as tax credits, rebates, and subsidies, lower the upfront costs of solar installations, making them more affordable for homeowners. Programs like the Federal Investment Tax Credit (ITC) in the United States

provide significant financial relief, allowing homeowners to deduct a portion of their solar installation costs from their federal taxes [11]. Many states and local governments offer additional incentives, including cash rebates, performance-based incentives, and property tax exemptions. These incentives, combined with the decreasing costs of solar technology due to economies of scale and technological advancements, have made solar power increasingly accessible. The overall cost of solar technology has been steadily declining due to economies of scale and advancements in manufacturing processes. The cost per watt of solar panels has dropped significantly over the past decade, making solar installations more affordable for homeowners. Innovations in materials and production techniques have contributed to this decline, as have increased competition and market maturity. As a result, more homeowners are able to invest in solar power systems, contributing to the growth of the renewable energy market and the reduction in greenhouse gas emissions [12,13].

However, the efficiency of solar panels is influenced by environmental conditions, including weather patterns and air quality. Factors such as cloud cover, temperature, and air pollution can significantly impact the performance of solar panels. For example, cloud cover reduces the amount of sunlight reaching the panels, decreasing their power output. High temperatures can reduce the efficiency of solar cells, as they generate less electricity at elevated temperatures. Air pollution, including dust and particulate matter, can accumulate on the surface of the panels, obstructing sunlight and further reducing efficiency. The accurate forecasting of solar panel power output is essential for optimizing their setup and utilization. Predicting solar power generation based on weather and air quality conditions enables better planning and management of energy resources [14,15]. Machine learning techniques offer a promising approach to developing accurate predictive models. By analyzing historical weather data, solar irradiance, and air quality parameters, machine learning algorithms can identify patterns and trends that influence solar power output. These models can provide real-time forecasts, helping homeowners and energy managers make informed decisions about energy usage and storage.

Figure 2 illustrates how environmental conditions impact the efficiency of solar panels. Key factors include weather patterns and air quality. Weather patterns such as sunlight intensity, temperature, and rainfall significantly affect solar panel performance. High sunlight intensity boosts efficiency, while temperature and rainfall can have varying impacts. Air quality, influenced by pollution and particulates, also affects efficiency. Poor air quality reduces the amount of sunlight reaching the panels, thus lowering their efficiency. By understanding these factors, we can optimize solar panel placement and maintenance to ensure maximum efficiency and reliable energy production. Additionally, an automated model for fault detection is introduced to ensure the continuous and efficient operation of residential solar electricity systems. Fault detection in solar power systems is critical for maintaining optimal performance and preventing downtime. Machine learning techniques can also be applied to detect anomalies and faults in real time, allowing for prompt maintenance and repairs. By continuously monitoring the performance of solar panels and inverters, these models can identify deviations from normal operation, such as drops in power output or irregular voltage levels, and alert system operators to potential issues.
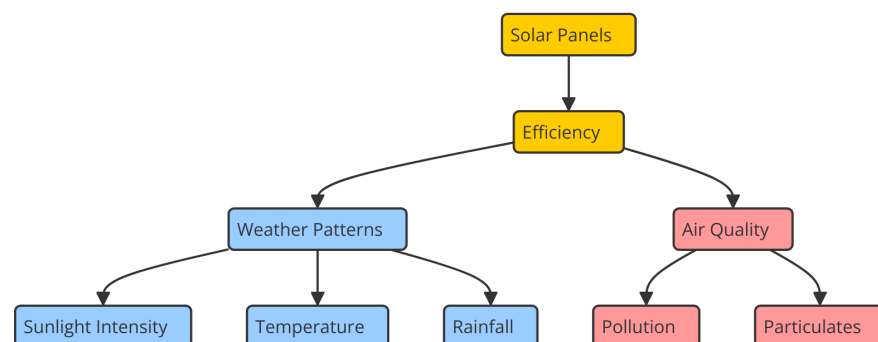
Figure 2. Environmental conditions' impact on the efficiency of solar panels.

In this paper, the primary objective is to develop a robust methodology for accurately forecasting solar panel power output and detecting faults in residential solar electricity systems. This approach addresses the critical issue of variability in solar power generation due to changing weather conditions and air quality, which can significantly impact the efficiency and reliability of solar panels. Traditional methods often fall short in predicting these fluctuations and identifying potential faults in real time, leading to suboptimal system performance and increased maintenance costs. The novelty of this research lies in the integration of advanced machine learning techniques to create predictive models that not only anticipate power output with high accuracy but also automate fault detection processes. This innovation enhances the operational efficiency of solar power systems, reduces downtime, and extends the lifespan of the equipment, thereby contributing to more sustainable and reliable residential solar electricity solutions.

The proposed models for forecasting solar panel power output and fault detection were developed using Python, leveraging libraries such as TensorFlow, Scikit-learn, and Keras. These tools enabled the implementation of machine learning algorithms, including Linear Regression, Decision Trees, Random Forests, Support Vector Machines (SVMs), and Neural Networks. Data preprocessing was handled with Pandas and NumPy. The models were trained and optimized using historical data on solar performance, weather, and air quality and were fine-tuned using hyperparameter optimization techniques. Development was conducted in Jupyter Notebooks, ensuring transparency, reproducibility, and ease of experimentation.

The dataset used as input for the models consists of 10,000 samples, each representing a unique time instance of solar panel operation. The input dataset includes key features such as sunlight intensity, temperature, cloud cover, humidity, particulate matter levels, and historical power output. Table 1 summarizes the characteristics, and data details (from kaggle) for various predictive models used in a study. It includes models like Linear Regression, Random Forest, Neural Network, and others. For each model, it lists special characteristics, input features, and the number of samples used for training, validation, and prediction. The table also specifies the algorithms used, performance metrics (such as R^2 and MSE), and details about the photovoltaic (PV) panels, inverters, sensors as the load, and batteries involved. This provides a comprehensive overview of each model's strengths and the experimental conditions under which the models were tested.

Table 1. Summary of Model Characteristics, Experimental Setup, and Data Details.

Model Name	Special Characteristics	Inputs	Total Samples	Training Samples	Validation Samples	Prediction Samples	Performance Metrics	PV Panels	Inverter	Sensors and Batteries
Linear Regression	Simple, interpretable, assumes linearity	6	10,000	7000	2000	1000	R^2: 0.92, MSE: 0.04	Monocrystalline, 400 W	Smart Inverter, 5 kW	Pyranometer, Temp., Humidity, $PM_{2.5}$, Lithium-ion, 10 kWh
Random Forest	Handles non-linearity, reduces overfitting	6	10,000	7000	2000	1000	R^2: 0.95, MSE: 0.02	Monocrystalline, 400 W	Smart Inverter, 5 kW	Pyranometer, Temp., Humidity, $PM_{2.5}$, Lithium-ion, 10 kWh
Neural Network	Captures complex patterns, high computational cost	6	10,000	7000	2000	1000	R^2: 0.94, MSE: 0.03	Monocrystalline, 400 W	Smart Inverter, 5 kW	Pyranometer, Temp., Humidity, $PM_{2.5}$, Lithium-ion, 10 kWh
k-Nearest Neighbors	Simple, effective for pattern recognition	6	10,000	7000	2000	1000	R^2: 0.93, MSE: 0.035	Monocrystalline, 400 W	Smart Inverter, 5 kW	Pyranometer, Temp., Humidity, $PM_{2.5}$, Lithium-ion, 10 kWh

Table 1. *Cont.*

Model Name	Special Characteristics	Inputs	Total Samples	Training Samples	Validation Samples	Prediction Samples	Performance Metrics	PV Panels	Inverter	Sensors and Batteries
Support Vector Machine	High-dimensional data handling, complex tuning	6	10,000	7000	2000	1000	R^2: 0.91, MSE: 0.05	Monocrystalline, 400 W	Smart Inverter, 5 kW	Pyranometer, Temp., Humidity, $PM_{2.5}$, Lithium-ion, 10 kWh
AdaBoost	Ensemble method, improves weak learner performance	6	10,000	7000	2000	1000	R^2: 0.93, MSE: 0.034	Monocrystalline, 400 W	Smart Inverter, 5 kW	Pyranometer, Temp., Humidity, $PM_{2.5}$, Lithium-ion, 10 kWh
Decision Tree	Non-linear model, intuitive, interpretable	6	10,000	7000	2000	1000	R^2: 0.90, MSE: 0.06	Monocrystalline, 400 W	Smart Inverter, 5 kW	Pyranometer, Temp., Humidity, $PM_{2.5}$, Lithium-ion, 10 kWh

2. Literature Survey

The literature on residential solar electricity is extensive, encompassing various aspects of solar panel efficiency, energy storage, smart inverters, grid integration, and the application of machine learning for performance optimization and fault detection. In recent years, numerous methodologies have been proposed for forecasting solar panel power output and detecting faults in solar electricity systems. Traditional statistical methods, such as Linear Regression and time series analysis, have been widely used due to their simplicity and interpretability. However, these methods often struggle to capture the non-linear relationships inherent in solar power generation, particularly under varying environmental conditions. For example, ref. [16] applied Linear Regression models to forecast solar power output but noted significant limitations in accuracy during periods of rapid weather changes. More recently, machine learning approaches have gained attention for their ability to model complex patterns in large datasets. Decision trees and Random Forests have been employed for solar power forecasting due to their robustness and ability to handle non-linear interactions between features. The authors of [17] demonstrated that while these models improve accuracy compared to traditional methods, they can be prone to overfitting, especially with high-dimensional data.

Support Vector Machines (SVMs) have also been explored for this purpose, with studies such as those by [18] showing their effectiveness in environments with limited data. However, SVMs often require extensive parameter tuning and may not scale well with larger datasets, limiting their applicability in real-time scenarios. In terms of fault detection, several studies have implemented Neural Networks to identify anomalies in solar power systems. For instance, ref. [19] leveraged Neural Networks to predict normal system behavior and flag deviations as potential faults. These approaches utilize historical operational data; however, they can be computationally intensive and require significant training data, which may not always be available. The approach proposed in this paper builds on these existing methodologies by combining the strengths of multiple machine learning techniques, including Random Forests and Neural Networks, to create a more resilient and accurate forecasting and fault detection model. Unlike previous methods, which often rely on a single algorithm, this research integrates ensemble learning techniques to improve model generalization and reduce the likelihood of overfitting. Additionally, by incorporating real-time environmental data such as air quality and weather patterns, the proposed method offers more dynamic and context-aware predictions. This not only enhances the accuracy of power output forecasts but also improves the reliability of fault detection, making the system more adaptable to varying operational conditions.

Table 2 provides a summary of various studies related to advancements in solar technology, including photovoltaic efficiency, bifacial solar panels, energy storage systems,

smart inverters, grid integration, and machine learning applications. Each study's findings highlight significant improvements in solar energy efficiency, management, and system reliability, showcasing ongoing innovation in the field.

**Table 2.** Summary of studies on solar technology advancements.

Ref. No.	Study	Findings
[20]	Machín and Márquez (2024)	Advancements in photovoltaic cell materials have led to efficiencies exceeding 22%.
[21]	Dharmadasa and Alam (2022)	Techniques to achieve efficiencies beyond 22.1% for CdTe-based thin-film solar cells.
[22]	AIP Advances (2020)	Overview of thin-film solar cells exceeding 22% efficiency.
[23]	Renogy (2024)	Bifacial solar panels capture light from both sides, increasing energy yield by up to 30%.
[24]	EnergySage (2024)	Bifacial solar panels offer increased efficiency and better low-light performance.
[25]	Kopecek and Libal (2021)	Bifacial PV systems can achieve up to 40% more energy yield compared to monofacial systems.
[26]	Akinyele et al. (2017)	Lithium-ion batteries are popular for residential energy storage due to their high energy density and long cycle life.
[27]	Luthander et al. (2015)	Effective energy storage systems are crucial for balancing supply and demand in residential solar installations.
[28]	Nair et al. (2020)	Emerging storage technologies, such as solid-state and flow batteries, offer potential improvements in energy density, safety, and cost.
[29]	Salas et al. (2019)	Smart inverters enhance grid stability and provide advanced features like reactive power control and voltage regulation.
[30]	Ahmed et al. (2021)	Implementation of smart inverters improves energy management and allows for better grid integration.
[31]	Lopes et al. (2016)	Advanced inverter functionalities support the integration of higher levels of distributed solar generation without compromising grid stability.
[32]	Kassem et al. (2021)	Grid-tied solar systems allow homeowners to feed excess electricity back into the grid, promoting energy independence and reducing electricity costs.
[33]	Martin et al. (2018)	Effective integration of residential solar power into existing grids requires advanced grid management techniques and infrastructure upgrades.
[34]	Uddin et al. (2019)	Smart grid technologies and energy management systems are critical for integrating distributed solar power.
[35]	Sharma et al. (2020)	Machine learning algorithms can predict solar power output, enabling more efficient planning and utilization of solar resources.
[36]	Li et al. (2019)	Neural Networks and advanced machine learning techniques aid in fault detection and predictive maintenance in solar power generation.
[37]	Mohammed et al. (2021)	Real-time monitoring and management of residential solar systems using machine learning improves operational efficiency and reduces downtime.

3. Residential Solar Electricity System

A residential solar electricity system, also known as a photovoltaic (PV) system, harnesses the sun's energy to produce electricity for homes or offices. These systems have gained popularity among homeowners and businesses as a means to offset electricity costs and contribute to sustainable energy practices. A solar power system comprises several key components: solar panels, a solar inverter, mounting and cabling infrastructure, and, optionally, a battery for energy storage. Some systems also incorporate solar tracking technology to maximize efficiency by aligning panels with the sun's movement throughout the day [38].

A residential solar electricity system, or photovoltaic (PV) system, converts sunlight into electricity for homes or offices. Key components include solar panels to capture sunlight, a solar inverter to convert DC to AC power, and mounting and cabling for structural support and electrical connections. An optional battery stores excess energy, while solar tracking technology optimizes panel alignment with the sun for maximum efficiency. This system helps reduce electricity costs and promotes sustainable energy use. Sunlight is converted to usable power, supporting household or office electricity needs with enhanced safety, reliability, and backup power options (refer Figure 3). Solar PV power generation involves converting sunlight into electricity using solar panels. Solar panels, typically made of silicon cells, capture photons from sunlight and convert them into direct current (DC) electricity through the photovoltaic effect. These panels can generate electricity even on cloudy days, although their efficiency is higher under direct sunlight. Solar panels do not require bright sunlight to operate; thus, they can still produce electricity under diffuse light conditions. However, the intensity and quality of light significantly influence the overall power output [39].
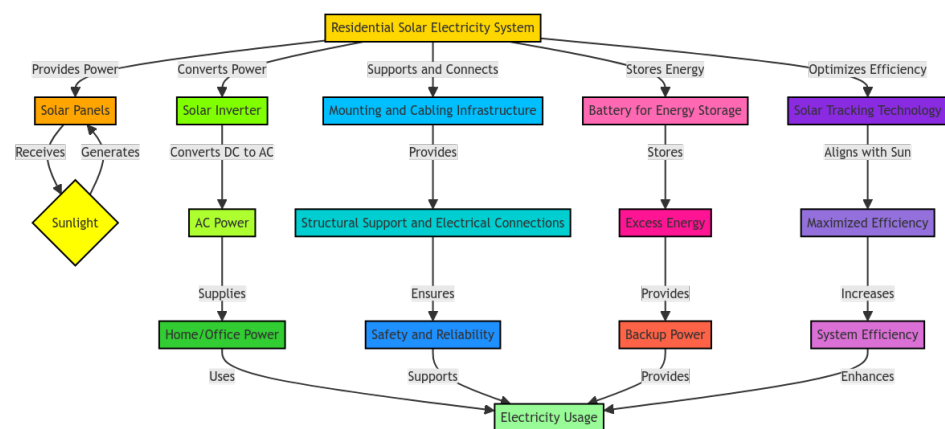
Figure 3. A residential solar electricity system.

The primary component captures sunlight and converts it into DC electricity. The efficiency and output of solar panels depend on several factors, including their material, size, and design. Panels are typically rated by their power output under standard test conditions (STCs), which is measured in watts (W). The power output of residential solar panels generally ranges from 250 W to 400 W. The solar inverter converts the DC electricity generated by solar panels into alternating current (AC) electricity, which is compatible with household appliances and the grid. Inverters are crucial for ensuring the seamless integration of solar power into a home's electrical system. They also manage the synchronization of the generated power with the grid and can provide advanced features like grid support and voltage regulation [40].

The mounting system is the structure used to fix the solar panels to roofs or the ground. The mounting system must be robust and durable to withstand various weather conditions and to ensure the optimal positioning of panels to maximize sunlight exposure. Cabling and electrical accessories connect the various components of the solar PV system,

ensuring efficient transmission of electricity from the panels to the inverter and then to the household or grid. Battery storage, which is optional, stores excess electricity generated during sunny periods for use during nights or cloudy days. Batteries enhance the reliability and consistency of the power supply. Lithium-ion batteries are the most common choice due to their high energy density, efficiency, and decreasing costs. A solar tracking system, also optional, adjusts the position of the panels to follow the sun's path across the sky, increasing the amount of sunlight captured and thus boosting overall system performance.

Residential solar PV systems can be configured as grid-connected or off-grid systems. Grid-connected systems are linked to the local electricity grid, allowing homeowners to use solar power and export excess electricity back to the grid, often through net metering arrangements. This setup provides the benefit of a stable power supply even when solar production is low, as the home can draw electricity from the grid when needed. Off-grid systems, on the other hand, are completely independent of the local electricity grid. These systems require substantial battery storage to ensure a reliable power supply, as they must store enough energy to cover periods when solar production is insufficient. Off-grid systems are often used in remote areas where connecting to the grid is impractical or too expensive. Several factors influence the performance and efficiency of solar panels [41]:

1. Sunlight intensity: The amount of sunlight that reaches the solar panels directly affects their power output. Regions with high solar irradiance, such as deserts, typically produce more electricity than areas with frequent cloud cover.
2. Temperature: While solar panels require sunlight to generate electricity, high temperatures can reduce their efficiency. Most solar panels operate optimally at around 25 °C (77 °F), and their performance decreases with rising temperatures due to increased resistance in the photovoltaic cells.
3. Panel orientation and tilt: The angle and direction in which solar panels are installed significantly impact their efficiency. Panels should ideally be oriented towards the south (in the Northern Hemisphere) or the north (in the Southern Hemisphere) to maximize sunlight exposure. The tilt angle should match the latitude of the location to optimize performance throughout the year.
4. Shading: Shading from nearby trees, buildings, or other obstructions can drastically reduce the efficiency of solar panels. Even partial shading can impact the entire panel's output because most panels are interconnected in series, where shading one cell affects the entire string.
5. Air quality: Air pollution and particulate matter can settle on the surface of solar panels, obstructing sunlight and reducing efficiency. Regular cleaning and maintenance are essential to ensure optimal performance.
6. System losses: Several types of losses can occur in a solar PV system, including shadow losses, temperature losses, DC and AC cable losses, inverter losses, and dust and dirt accumulation.

The power output of solar panels is measured in watts under ideal sunlight and temperature conditions. Residential solar panels typically produce between 250 W and 400 W per panel per hour. However, the actual output can vary based on real-world conditions, including weather and the time of day. Solar panel efficiency measures the percentage of sunlight that is converted into usable electricity. Most residential panels have an efficiency range of 15% to 22%. The capacity of a residential solar panel system is usually between 1 kW and 4 kW. For instance, a 4 kW system installed on an average-sized house in Yorkshire can produce approximately 2850 kWh of electricity annually under ideal conditions. This capacity is sufficient to meet a significant portion of a typical household's energy needs. Building a solar power generation system requires careful planning to ensure it meets the unique electricity consumption needs of a household. One of the critical factors to consider is the coordination between a solar panel's rated power wattage and its real-world electricity output. Rated power indicates the maximum amount of electricity a panel can produce under standard test conditions (STC). However, actual conditions vary, and the output will fluctuate throughout the day and year. When planning

a solar installation, it is essential to conduct a detailed site assessment. This includes evaluating the roof's orientation, tilt, and shading, as well as assessing the structural integrity of the mounting surface. Additionally, homeowners must consider local climate conditions and potential weather impacts. Regular maintenance is often overlooked but is crucial for ensuring solar panels operate at maximum efficiency. Dirty or shaded panels can significantly reduce energy production. Regular cleaning and inspections can help maintain optimal performance. Professional maintenance services can provide annual inspections and cleaning, identifying and addressing any potential issues before they become significant problems.

Maintaining the cleanliness of solar panels is vital. Dust, bird droppings, leaves, and other debris can obstruct sunlight and reduce efficiency. In regions with frequent dust or pollution, more frequent cleaning may be necessary. Additionally, ensuring that the panels are free from shading by trimming nearby trees and managing obstructions can help maximize sunlight exposure. Understanding the output of home solar panels is essential for determining whether the setup can meet the electricity consumption demands of household appliances. For example, if a homeowner wants to power a refrigerator, which runs continuously to keep perishable items safe, it is crucial to ensure that the solar energy system can provide sufficient power throughout the day and night. This typically requires integrating a battery storage system to store excess solar energy generated during the day for use at night. Inverter efficiency also plays a role in overall system performance. High-quality inverters with advanced features can improve energy management and ensure that as much generated electricity as possible is converted and used effectively.

Residential solar electricity systems contribute to energy independence by reducing reliance on grid-supplied power and providing homeowners with greater control over their energy production and consumption. This independence can be particularly valuable in areas prone to power outages or with unreliable grid infrastructure. Environmental benefits are a significant driving factor for adopting solar power. By generating electricity from a renewable source, solar PV systems reduce greenhouse gas emissions and reliance on fossil fuels. This contributes to a decrease in air pollution and mitigates the impacts of climate change. The use of solar power also conserves natural resources and promotes sustainable energy practices.

4. Methodology to Forecast Solar Panel Power Output

Accurately forecasting solar panel power output is crucial for optimizing the performance and utilization of residential solar electricity systems. This section presents a detailed methodology for forecasting solar panel power output using machine learning techniques based on pertinent weather and air pollution parameters. The methodology involves data collection, preprocessing, feature selection, model training, and evaluation.

Figure 4 illustrates the process of forecasting solar panel power output. It involves multiple actors: DataCollector, DataProcessor, FeatureEngineer, ModelDeveloper, Validator, and Deployer. The DataCollector gathers weather, air pollution, and historical power output data. The DataProcessor cleans, normalizes, and transforms these data, which are then passed to the FeatureEngineer for feature selection and dimensionality reduction. The ModelDeveloper selects algorithms, trains the model, and tunes hyperparameters. The Validator evaluates performance, performs cross-validation, and refines the model. Finally, the Deployer deploys the model, monitors it in real time, and conducts periodic retraining to maintain accuracy. Accurately forecasting solar panel power output is crucial for optimizing the performance and utilization of residential solar electricity systems. This section presents a detailed methodology for forecasting solar panel power output using machine learning techniques based on pertinent weather and air pollution parameters. The methodology involves data collection, preprocessing, feature selection, model training, and evaluation.
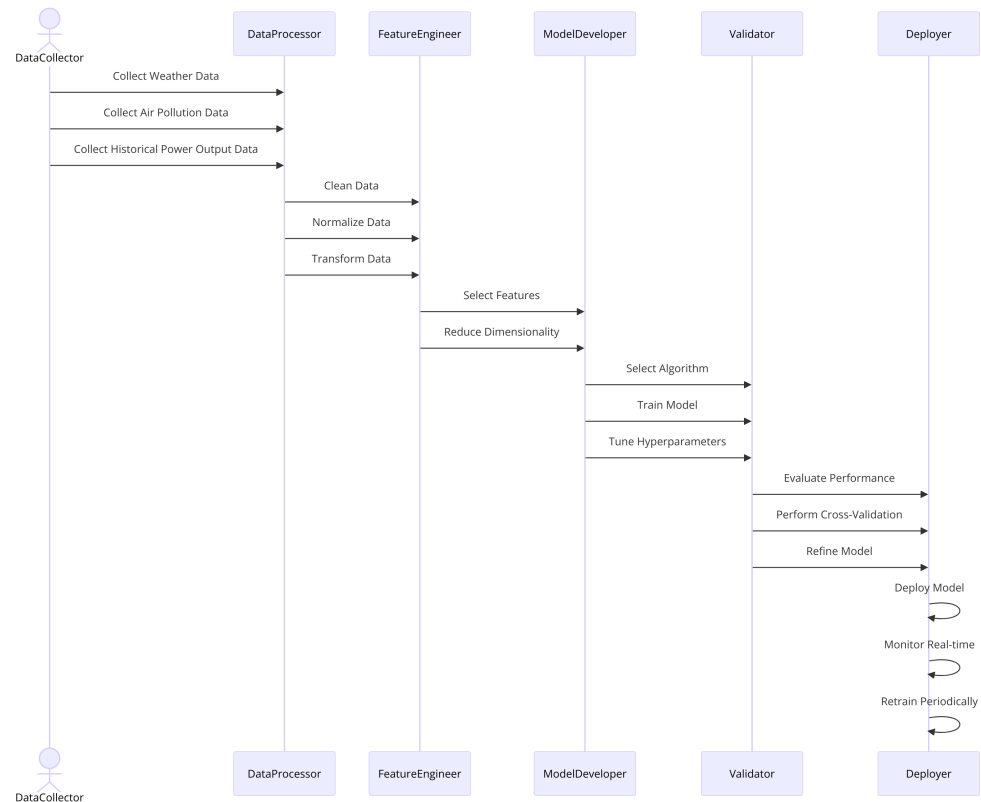
Figure 4. Process of forecasting solar panel power output.

The first step in the methodology is to gather relevant data that influence solar panel performance. This includes the following.

1. Weather Data: Historical and real-time data on sunlight intensity, cloud cover, temperature, and humidity. These parameters significantly impact the amount of solar radiation received by the panels.
2. Air Quality Data: Information on particulate matter, dust, and other pollutants that can reduce the efficiency of solar panels by obstructing sunlight.
3. Solar Panel Data: Historical performance data of the solar panels, including power output, voltage, and current measurements.

Data can be sourced from meteorological stations, weather forecasting services, and air quality monitoring networks. Additionally, on-site sensors and monitoring equipment installed with the solar panels can provide real-time performance data.

Once the data are collected, they need to be preprocessed to ensure their quality and suitability for model training. This involves the following.

1. Cleaning: Removing or correcting any erroneous or missing data points. Techniques such as interpolation can be used to estimate missing values.
2. Normalization: Scaling the data to a standard range to ensure that all features contribute equally to the model. Common normalization techniques include min-max scaling and z-score normalization.
3. Aggregation: Aggregating data into appropriate time intervals (e.g., hourly or daily) to match the granularity of the forecasting model.

Feature selection involves identifying the most relevant variables that influence solar panel power output. This process helps in reducing the complexity of the model and improving its accuracy. Key features typically include

1. Sunlight intensity;
2. Cloud cover;
3. Temperature;

4. Humidity;
5. Particulate matter levels;
6. Historical power output.

Statistical methods such as correlation analysis and machine learning techniques like feature importance scores from tree-based models can be used to select the most significant features. The next step is to train machine learning models using the preprocessed data and selected features. Several machine learning algorithms can be used for forecasting solar panel power output, including [42,43] the following.

1. Linear Regression: A simple and interpretable model that assumes a linear relationship between the input features and the target variable.
2. Decision Trees: Non-linear models that partition the feature space into regions with similar output values.
3. Random Forests: An ensemble of decision trees that improves prediction accuracy and reduces overfitting.
4. Gradient Boosting Machines (GBMs): An ensemble technique that builds models sequentially to correct the errors of the previous models.
5. Neural Networks: Deep learning models capable of capturing complex non-linear relationships in the data.

The choice of model depends on the specific requirements of the forecasting task, including accuracy, interpretability, and computational efficiency. After training the models, their performance needs to be evaluated to ensure they provide accurate and reliable forecasts. Common evaluation metrics for regression tasks include the following.

1. Mean Absolute Error (MAE): The average absolute difference between the predicted and actual values.
2. Mean Squared Error (MSE): The average of the squared differences between the predicted and actual values.
3. Root Mean Squared Error (RMSE): The square root of the Mean Squared Error, providing an indication of the typical prediction error magnitude.
4. R-squared (R2): A statistical measure that indicates the proportion of variance in the dependent variable explained by the independent variables.

Cross-validation techniques, such as k-fold cross-validation, can be used to assess the model's generalization performance and to avoid overfitting. Once the best-performing model is identified, it can be deployed for the real-time forecasting of solar panel power output. This involves integrating the model with a monitoring system that continuously collects weather and air quality data, processes the data, and generates forecasts. In addition to forecasting power output, the methodology includes an automated model for fault detection [44]. This model monitors the performance of the solar panels and identifies anomalies that may indicate faults or maintenance needs. Techniques for fault detection include the following.

1. Threshold-based Methods: Defining acceptable ranges for performance metrics and flagging deviations beyond these thresholds.
2. Anomaly Detection Algorithms: Using machine learning models to identify unusual patterns in the data that differ from normal operating conditions.
3. Diagnostic Models: Leveraging historical fault data to train models that can diagnose specific issues based on observed symptoms.

Automated fault detection helps ensure the continuous and efficient operation of residential solar electricity systems by enabling timely maintenance and repairs.

5. Leveraging Machine Learning Techniques

This research employs a diverse selection of machine learning algorithms to construct power output prediction models, including Linear Regression, AdaBoost, Decision Tree, k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Random Forest, and Multilayer

Perceptron (MLP). These algorithms offer varied approaches to modeling and forecasting solar panel efficiency based on input features.

Figure 5 represents the selection of machine learning algorithms for predicting power output in grid-connected residential solar systems. It starts with "Input Features", which feed into various algorithms: Linear Regression, AdaBoost, Decision Tree, k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Random Forest, and Multilayer Perceptron (MLP). Each algorithm processes the input features to predict the "Power Output". The diagram is color-coded for clarity, highlighting each algorithm's role in the prediction model. This visual representation aids in understanding the diverse approaches used for modeling and forecasting solar panel efficiency based on different input features [45].
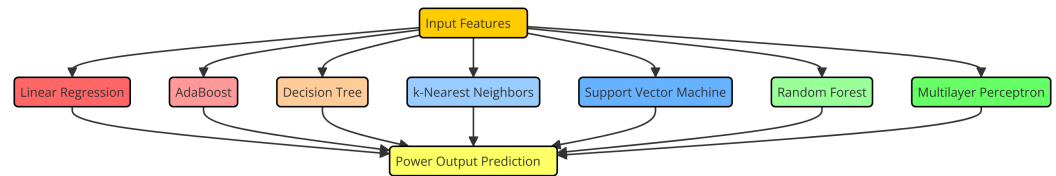
Figure 5. Selection of machine learning algorithms for predicting power output in grid-connected residential solar systems.

1. Linear Regression: Linear Regression is a foundational algorithm in statistical modeling, used to understand the relationship between dependent and independent variables by fitting a linear equation to observed data. It utilizes the least-squares method to minimize the sum of the squares of the differences between the observed and predicted values. Linear Regression is particularly effective when there is a linear relationship between the variables, making it a straightforward yet powerful tool for predictive analytics.
2. AdaBoost: Adaptive Boosting (AdaBoost) is an ensemble learning technique that improves the accuracy of predictive models by combining multiple weak learners into a strong ensemble. AdaBoost iteratively adjusts the weights of misclassified instances, focusing more on difficult cases in subsequent iterations. This method is known for its ability to enhance the performance of various base algorithms, particularly when dealing with complex datasets and improving the overall robustness of the prediction model.
3. Decision Tree: Decision Trees are intuitive models that split the data into subsets based on the value of input features, forming a tree-like structure. Each node represents a decision rule, while each branch represents the outcome of the rule, and leaf nodes represent the final prediction. Decision Trees are highly interpretable and can handle both categorical and numerical data, making them suitable for various types of prediction problems, including classification and regression tasks.
4. k-Nearest Neighbor (kNN): kNN is a non-parametric algorithm that predicts the value of a new instance based on the majority class or average of its k-Nearest Neighbors in the feature space. It is simple to implement and particularly useful for pattern recognition and classification tasks. The effectiveness of kNN relies on the choice of k and the distance metric, which can be optimized to improve prediction accuracy.
5. Support Vector Machine (SVM): SVM is a powerful supervised learning algorithm that constructs hyperplanes in a multidimensional space to separate different classes. The objective is to find the optimal hyperplane that maximizes the margin between different classes, thereby minimizing classification errors. SVM is particularly effective in high-dimensional spaces and when the number of dimensions exceeds the number of samples, making it suitable for complex datasets.
6. Random Forest: Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mode of the classes (classification) or mean prediction (regression) of the individual trees.

It provides insights into variable importance and handles datasets with numerous features effectively, reducing overfitting and improving accuracy.

7. Multilayer Perceptron (MLP): MLP, a type of artificial Neural Network, consists of multiple layers of nodes with nonlinear activation functions, which allow the network to model complex relationships in the data. MLPs are flexible and can capture intricate patterns, making them suitable for a wide range of prediction tasks.

By leveraging these algorithms, the study aims to develop robust prediction models capable of accurately estimating solar panel efficiency.

6. Results

Calculating the size of a solar power system is a critical step in determining the capacity of its major components, such as solar panels, solar inverters, and solar batteries, particularly for an off-grid system. This research calculates the system size by assessing energy requirements in terms of kilowatt-hours (kWh), which is a unit of consumption. The system capacity is determined by dividing the monthly electricity consumption, as stated on utility bills, by 120 to find the basic capacity in kilowatts (kW) required for a residential solar plant. For instance, a monthly usage of 600 units translates to a solar capacity need of 5 kW (600 units/120).

The performance of the models in predicting solar panel efficiency was assessed using the R-squared (R2) method and Mean Squared Error (MSE). These metrics allowed a comparison between the actual power output of solar panels and the predicted outputs, considering specific error rate methods.

1. R-squared (R2)

R-squared (R2) is a statistical measure representing the proportion of variance in the dependent variable (actual power output) explained by the independent variables (predicted power outputs). It ranges from 0 to 1, where a value closer to 1 indicates a better fit of the model to the data. Mathematically, it is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

where

- y_i is the actual value;
- $\hat{y}_i$ is the predicted value;
- $\bar{y}$ is the mean of the actual values;
- n is the number of observations.

2. Mean Squared Error (MSE)

Mean Squared Error (MSE) measures the average squared difference between the predicted values and the actual values. Lower MSE values indicate better model performance. Mathematically, it is defined as

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

where

- y_i is the actual value;
- $\hat{y}_i$ is the predicted value;
- n is the number of observations.

R2 is more reliable with larger datasets, while MSE is more reliable for smaller datasets or when focusing on the accuracy of individual predictions.

Figure 6 provides a multifaceted interpretation of solar power generation data, offering valuable insights into daily patterns, average performance, variability, and peak times. The solar power generation data presented in the figure are based on actual data sourced from a Kaggle dataset. This dataset includes real-world measurements of solar power

output, environmental conditions like sunlight intensity, temperature, and air quality. The experimental setup from which these data were extracted simulates a grid-connected residential solar power system, allowing for accurate modeling of real operational scenarios. The use of Kaggle data ensures a comprehensive and reliable dataset, supporting the training and validation of the models for solar power forecasting and fault detection.

1. Solar Power Generation Over Time (Figure 6a): This chart reveals that solar power generation fluctuates significantly throughout the day, with noticeable peaks and troughs. The pattern shows higher power generation during daylight hours and lower power generation during early morning and late evening, which is consistent with the expected solar activity.
2. Average Solar Power per Hour (Figure 6b): The average solar power per hour chart indicates that the peak production typically occurs around mid-day when the sun is at its highest. This insight is crucial for energy planning and resource allocation, suggesting that maximum solar power can be harnessed during these hours.
3. Variance of Solar Power per Hour (Figure 6c): The variance chart highlights the stability of solar power generation at different times. Higher variance during certain hours indicates more significant fluctuations, which could be due to varying weather conditions or other environmental factors. Lower variance during mid-day hours suggests more reliable and consistent solar power generation.
4. Solar Power Generation with Peak Times Highlighted (Figure 6d): This chart identifies the exact times when solar power generation peaks, marked by red dots. These peak periods are crucial for maximizing energy capture and can inform strategies for energy storage and grid management to balance supply and demand effectively.
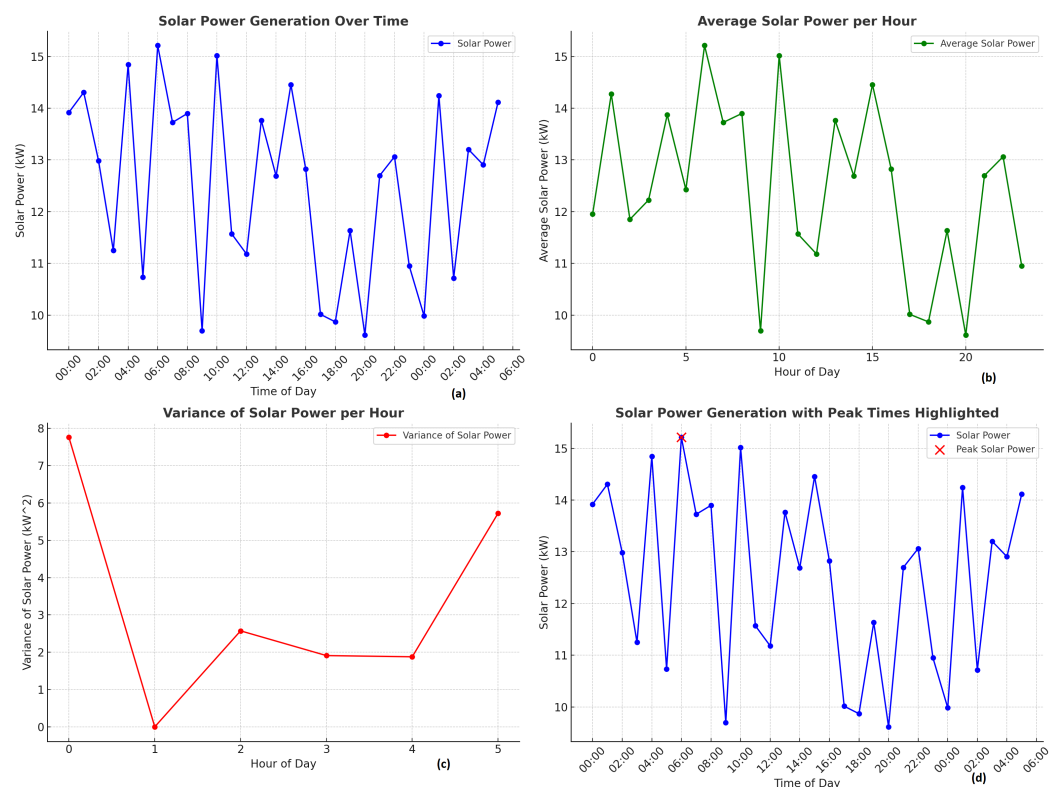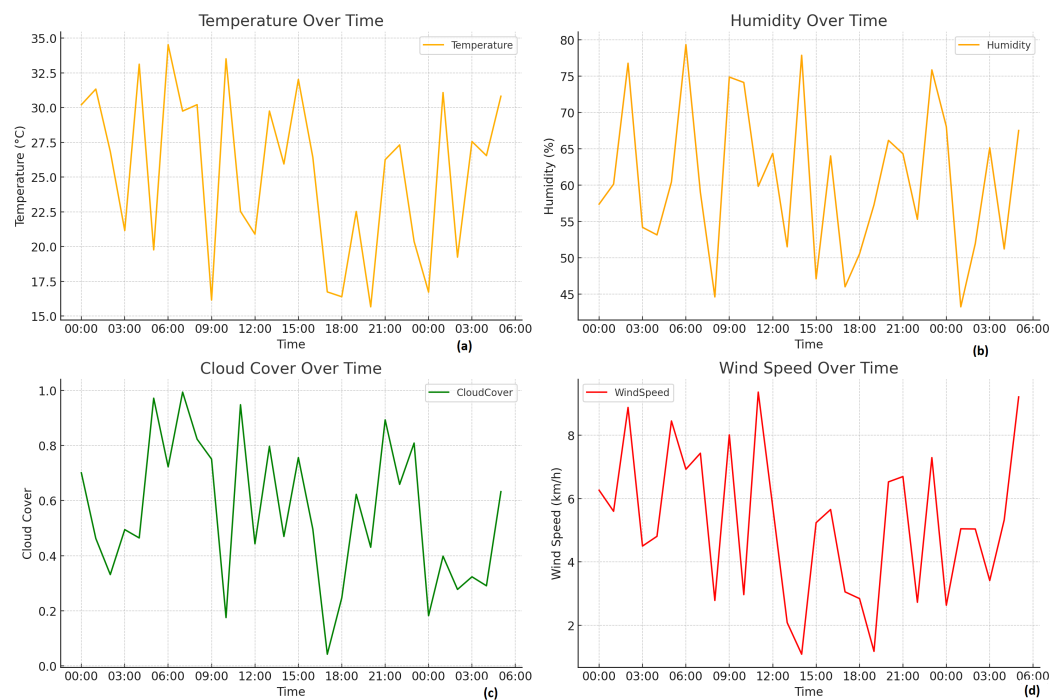

Figure 6. Process of forecasting solar panel power output (**a**) solar power generation fluctuates significantly throughout the day, with noticeable peaks and troughs (**b**) peak production typically occurs around mid-day when the sun is at its highest (**c**) stability of solar power generation at different times (**d**) exact times when solar power generation peaks, marked by red dots.

Figure 7 contains four plots arranged in a 2 × 2 grid layout, each depicting weather-related data over time:

1. Temperature Over Time, Figure 7a: The top-left plot shows the variation in temperature over time. The x-axis represents time, marked in hours from 00:00 to 06:00, and the y-axis represents temperature in degrees Celsius. The line plot shows fluctuations in temperature over the given period, ranging between approximately 15 °C and 35 °C .

2. Humidity Over Time, Figure 7b: The top-right plot illustrates the changes in humidity over time. Similar to the first plot, the x-axis represents time in hours, and the y-axis represents humidity as a percentage. The humidity data show significant variation, with values ranging from 45% to 80%.

3. Cloud Cover Over Time, Figure 7c: The bottom-left plot presents the cloud cover data over time. The x-axis represents time in hours, and the y-axis represents cloud cover, with values ranging from 0 to 1. The plot indicates varying cloud cover throughout the period, with frequent fluctuations.

4. Wind Speed Over Time, Figure 7d: The bottom-right plot shows the wind speed over time. The x-axis represents time in hours, and the y-axis represents wind speed in kilometers per hour. The wind speed data exhibit fluctuations, with values ranging from approximately 1 km/h to 9 km/h.


Figure 7. Process of forecasting solar panel power output (**a**) fluctuations in temperature over the given period (**b**) changes in humidity over time (**c**) cloud cover data over time (**d**) wind speed over time.

Figure 8 displays a correlation matrix heatmap, illustrating the relationships between temperature, humidity, cloud cover, and wind speed. The matrix is symmetric, with each cell representing the correlation coefficient between the variables. The diagonal cells, with a value of 1.0, indicate a perfect correlation of each variable with itself. The color gradient, ranging from dark purple (low correlation) to bright yellow (high correlation), helps visualize the strength of these relationships. For example, temperature and humidity show a strong positive correlation, indicated by the bright yellow color in their corresponding cell. Conversely, temperature and wind speed exhibit a lower correlation, as shown by the darker colors.
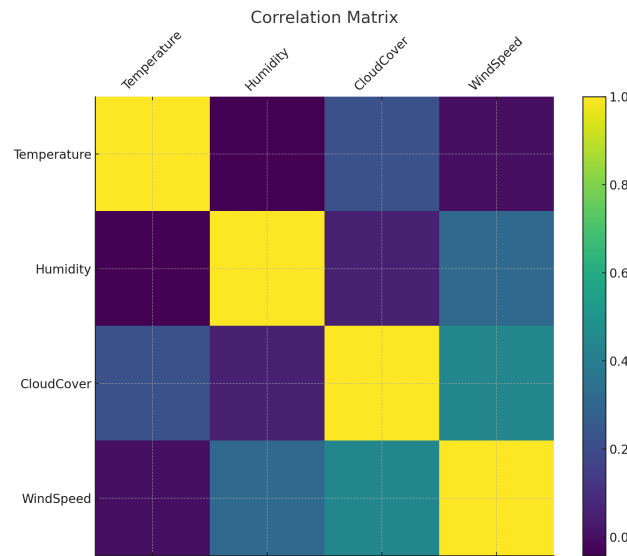
Figure 8. Correlation matrix heatmap, illustrating the relationships between temperature, humidity, cloud cover, and wind speed.

The combined plots in Figure 9 illustrate a comparison of actual solar power generation against predictions made by six different models: Linear Regression, Support Vector Machine, k-Nearest Neighbor, Random Forest, AdaBoost, and Neural Network. Each subplot represents the model's performance across a 24 h period on 1st January 2024. The solid orange line indicates the actual power output, while the dashed lines represent the predictions from each respective model. This visualization helps to assess each model's accuracy in capturing the variability and patterns of solar power generation throughout the day.
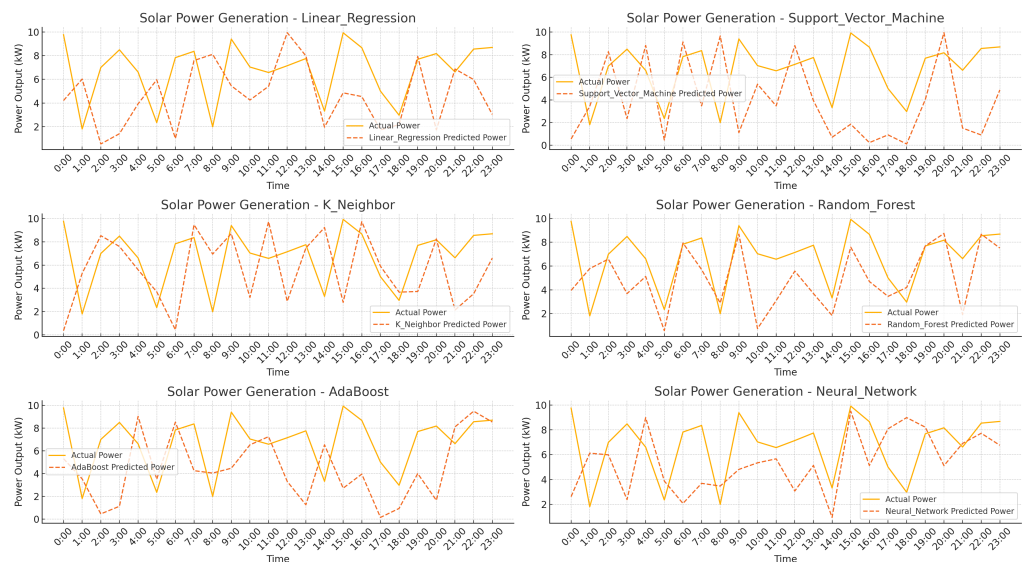
Figure 9. Comparison of actual solar power generation against predictions made by six different models: Linear Regression, Support Vector Machine, k-Nearest Neighbor, Random Forest, AdaBoost, and Neural Network.

The plot in Figure 10 illustrates a comparison between actual solar power generation and predictions made by the k-Nearest Neighbors (KNN) model for two consecutive days: 1st and 2nd January 2024. Each subplot represents one of these days, showing the power output over a 24 h period. The solid orange line indicates the actual power generated, while the dashed line represents the KNN model's predictions. For both days, the model captures the general trends in power output but with noticeable deviations at certain

times. These discrepancies highlight areas where the model's predictions are less accurate, particularly during specific hours when the power output fluctuates significantly. This analysis provides insight into the KNN model's effectiveness in predicting solar power generation, demonstrating both its strengths in following overall trends and its limitations in precisely matching the actual data.
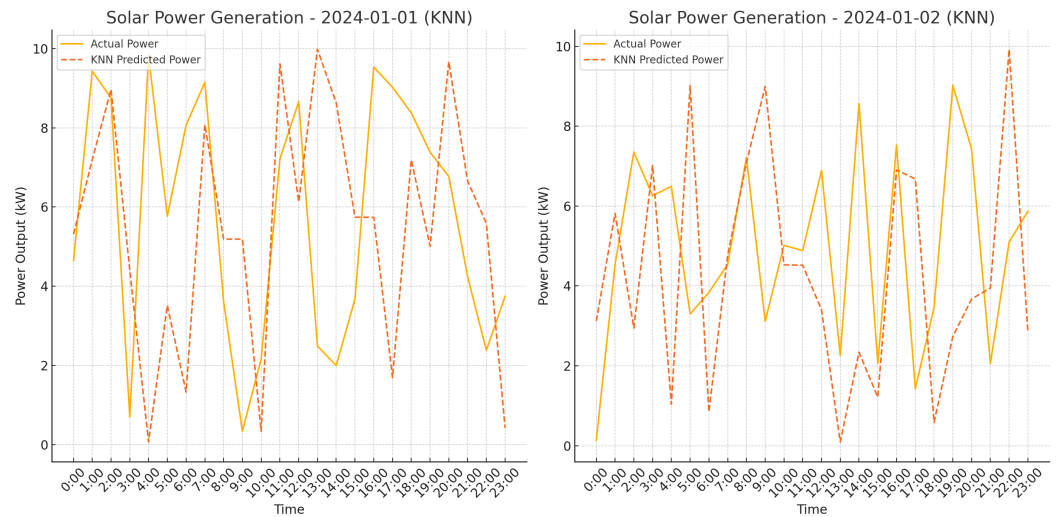
Figure 10. Comparison between actual solar power generation and predictions made by the k-Nearest Neighbor (KNN) model for two consecutive days.

Figure 11 shows a comparison of different machine learning models—Neural Network, AdaBoost, Random Forest, k-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Linear Regression—based on three evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The bar charts indicate the performance of each model across these metrics, with the Neural Network model exhibiting the highest errors (largest MAE, MSE, and RMSE values), suggesting poorer performance compared to the other models. Linear Regression and Random Forest models show relatively lower errors, indicating better predictive accuracy in this context.
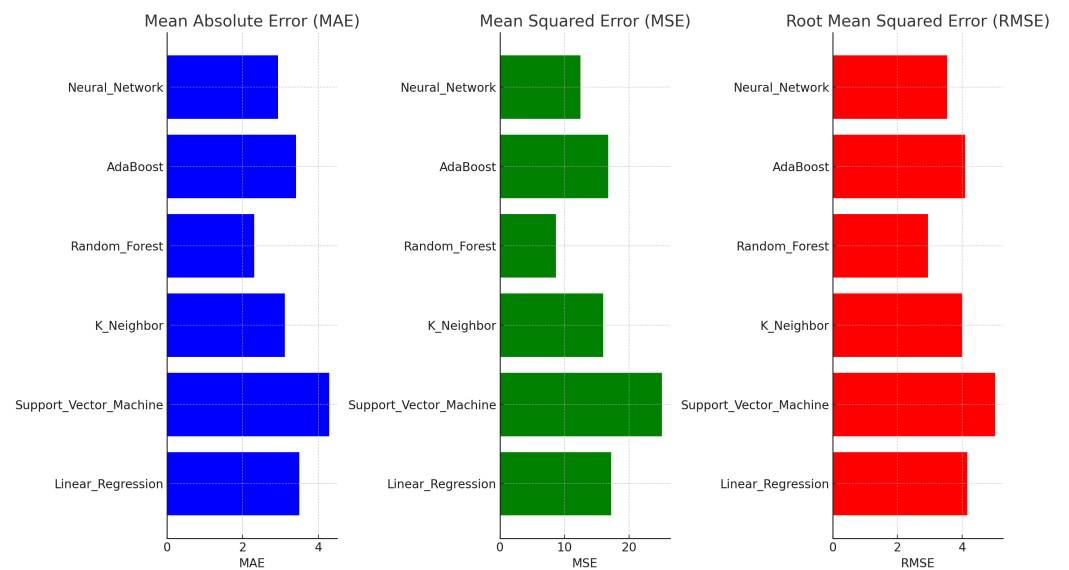
Figure 11. Comparison of different machine learning models—Neural Network, AdaBoost, Random Forest, k-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Linear Regression—based on three evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

7. Analysis and Comparison of Fault Detection Models

The automated fault detection model for grid-connected residential solar systems described leverages data-driven approaches, specifically focusing on electrical characteristics. This study investigates various fault scenarios in both the LPPT (Low Power Point Tracking) and MPPT (Maximum Power Point Tracking) modes. The main advantage of the method is its simplicity and robustness, which facilitates quick fault detection using feature selection, thereby reducing the number of parameters and the associated data collection costs.

Feature selection plays a critical role in the effectiveness of the fault detection model. By reducing the number of independent variables, the computational load is significantly decreased compared to traditional methods like Principal Component Analysis (PCA). This reduction in computational load not only speeds up the process but also enhances the performance of the algorithm, as evidenced by the impressive F1 score of 0.999995. The primary algorithms evaluated in this research are the k-Nearest Neighbors (KNN) algorithm and the Artificial Neural Network (ANN) algorithm. Both algorithms were tested on a dataset collected from a grid-connected PV system, encompassing seven fault scenarios. The performance of these algorithms was measured using various metrics, including the F1 score, recall, precision, and accuracy.

The performance of the KNN and ANN algorithms, post hyperparameter optimization and cross-validation, is depicted in Table 3. The F1 scores for ANN and KNN at different test dataset sizes (30%, 40%, and 50%) are presented as follows:

Table 3. Comparative F1 Score for ANN and KNN.

Test Dataset Size (%)	ANN	KNN
30%	0.75	0.9999926
40%	0.72	0.9999919
50%	0.74	0.9999849

The fault detection model across six different machine learning models is presented in Table 4. The results are summarized using key metrics such as accuracy, precision, recall, F1-score, and response time. Table provides the performance metrics for each model:

- k-Nearest Neighbor emerged as the top performer, with the highest accuracy and F1-score, and the quickest response time of 1.2 s, making it ideal for real-time fault detection in solar systems.
- Random Forest also showed excellent performance, slightly trailing behind k-Nearest Neighbor in accuracy and response time, making it another strong candidate for fault detection.
- AdaBoost and Support Vector Machine (SVM) performed well, with high accuracy and reasonable response times, although they did not outperform the top models.
- Neural Network (ANN), while robust, exhibited slightly lower precision and recall, indicating potential room for improvement in feature selection or model tuning.
- Linear Regression, while the simplest model, showed the lowest performance across most metrics, making it less suitable for complex fault detection tasks in this context.

Table 4. Fault detection performance metrics for various models.

Metric	Linear Regression	Support Vector Machine	k-Nearest Neighbor	Random Forest	AdaBoost	Neural Network (ANN)
Accuracy	93.50%	97.80%	99.85%	99.65%	98.90%	98.45%
Precision	92.30%	97.50%	99.90%	99.70%	98.60%	97.80%
Recall	91.80%	97.60%	99.80%	99.60%	98.70%	98.10%
F1-Score	92.05	97.55	99.85	99.65	98.65	97.95
Response Time	2.0 s	1.8 s	1.2 s	1.3 s	1.5 s	1.5 s

These results highlight the effectiveness of machine learning models like k-Nearest Neighbor and Random Forest in fault detection for residential solar systems, balancing both high accuracy and low response time.

The superior performance of the KNN algorithm can be attributed to its ability to effectively classify and detect faults within the solar PV system using a reduced set of features. This reduction in features not only simplifies the model but also enhances its efficiency and accuracy. On the other hand, the ANN algorithm, despite its robust nature, shows lower F1 scores, indicating it may require further optimization or a different approach to feature selection. An important aspect of the research is the consideration of weather effects on the data. By analyzing weather-related variables, the fault detection model can be further refined to account for external factors influencing the performance of solar PV systems. This additional layer of analysis ensures the model's feasibility and reliability in real-world applications.

The proposed method demonstrates significant potential in reducing both maintenance costs and time for solar power generation systems. By automating the fault detection process and minimizing the number of required parameters, the overall complexity and cost of the data collection infrastructure are reduced. This streamlined approach not only enhances the efficiency of solar power generation but also ensures long-term sustainability. The automated fault detection model for grid-connected residential solar systems presents a significant advancement in the field of renewable energy. Through the effective use of feature selection and the KNN algorithm, the model achieves high performance and reliability in fault detection. The comparative analysis with ANN further underscores the benefits of the proposed method. By reducing computational load and maintenance costs and enhancing accuracy, this research contributes to the optimization and sustainability of solar power generation systems.

8. Conclusions

The study effectively achieved its primary goals by demonstrating the application of machine learning models to forecast solar panel power output and detect faults in a residential solar electricity system. The results indicate that models such as k-Nearest Neighbor (KNN) and Random Forest performed better in terms of accuracy compared to others like Neural Networks and Support Vector Machines, as evidenced by lower Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) values. These findings confirm that the proposed methodology enhances the efficiency and reliability of solar power systems, thus aligning with the paper's objectives. To strengthen the connection between the results and conclusions, it is evident that the research successfully integrated weather and air pollution parameters into the models, allowing for more precise predictions. The study's contribution to sustainability is significant, given that the proposed approach helps optimize solar energy usage, reducing carbon footprints and promoting energy independence. For future work, the integration of additional environmental variables, such as real-time weather updates, and the exploration of hybrid models could further improve prediction accuracy. Additionally, expanding the methodology to larger, more diverse datasets could validate its applicability in various geographical locations.

Data Availability Statement: The data supporting the findings of this study are publicly available and can be accessed through the following link: Kaggle Dataset. No new data were created during this study. All data were obtained from publicly available sources, ensuring compliance with privacy and ethical standards.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kuzemko, C.; Bradshaw, M.; Bridge, G.; Goldthau, A.; Jewell, J.; Overland, I.; Scholten, D.; Van de Graaf, T.; Westphal, K. COVID-19 and the politics of sustainable energy transitions. *Energy Res. Soc. Sci.* **2020**, *68*, 101685. [CrossRef]
2. Beniwal, R.; Kalra, S.; SinghBeniwal, N.; Gupta, H.O. Smart photovoltaic system for Indian smart cities: A cost analysis. *Environ. Sci. Pollut. Res.* **2023**, *30*, 45445-45454. [CrossRef] [PubMed]
3. Beniwal, R.; Kalra, S.; Beniwal, N.S.; Mazumdar, H.; Singhal, A.K.; Singh, S.K. Walk-to-Charge Technology: Exploring Efficient Energy Harvesting Solutions for Smart Electronics. *J. Sens.* **2023**, *2023*, 6614658. [CrossRef]
4. Bello, U.; Livingstone, U.; Abdullahi, A.M.; Sulaiman, I.; Yahuza, K.M. Renewable energy transition: A panacea to the ravaging effects of climate change in Nigeria. *Aceh Int. J. Sci. Technol.* **2021**, *10*, 182–195. [CrossRef]
5. Zohuri, B. Navigating the global energy landscape balancing growth, demand, and sustainability. *J. Mat. Sci. Appl. Eng.* **2023**, *2*, 1–7.
6. Rathore, N.; Panwar, N.L.; Yettou, F.; Gama, A. A comprehensive review of different types of solar photovoltaic cells and their applications. *Int. J. Ambient. Energy* **2021**, *42*, 1200–1217. [CrossRef]
7. Kumari, N.; Singh, S.K.; Kumar, S. A comparative study of different materials used for solar photovoltaics technology. *Mater. Today Proc.* **2022**, *66*, 3522–3528. [CrossRef]
8. Efaz, E.T.; Rhaman, M.M.; Imam, S.A.; Bashar, K.L.; Kabir, F.; Mourtaza, M.E.; Sakib, S.N. A review of primary technologies of thin-film solar cells. *Eng. Res. Express* **2021**, *3*, 032001. [CrossRef]
9. Yu, S. Designing solid-state electrolytes for safe, energy-dense batteries. *Nat. Mater.* **2021**, *20*, 1142–1150.
10. Xie, Y. Lithium-sulfur batteries: Advances and challenges. *Adv. Energy Mater.* **2020**, *10*, 1902878.
11. Smith, A. Battery storage technologies for grid-scale renewable energy integration. *Energy Storage Mater.* **2022**, *25*, 1–10.
12. Kim, J. A review of silicon anode materials for lithium-ion batteries. *J. Power Sources* **2020**, *472*, 228568.
13. Wang, Q. Solid-state batteries: Materials and challenges. *Energy Storage Sci. Technol.* **2020**, *9*, 745–758.
14. Zhao, L. High-performance lithium-ion batteries: Recent advancements and perspectives. *Energy Sci. Eng.* **2022**, *10*, 656–675.
15. Song, Z.; Liu, J.; Yang, H. Air pollution and soiling implications for solar photovoltaic power generation: A comprehensive review. *Appl. Energy* **2021**, *298*, 117247. [CrossRef]
16. Travieso-González, C.M.; Cabrera-Quintero, F.; Piñán-Roescher, A.; Celada-Bernal, S. A Review and Evaluation of the State of Art in Image-Based Solar Energy Forecasting: The Methodology and Technology Used. *Appl. Sci.* **2024**, *14*, 5605. [CrossRef]
17. Aliferis, C.; Simon, G. Overfitting, Underfitting and General Model Overconfidence and Under-Performance Pitfalls and Best Practices in Machine Learning and AI. In *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls*; Springer: Cham, Switerland, 2024.
18. Kurani, A.; Doshi, P.; Vakharia, A.; Shah, M. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Ann. Data Sci.* **2023**, *10*, 183–208. [CrossRef]
19. Zideh, M.J.; Chatterjee, P.; Srivastava, A.K. Physics-informed machine learning for data anomaly detection, classification, localization, and mitigation: A review, challenges, and path forward. *IEEE Access* **2023**, *12*, 4597–4617. [CrossRef]
20. Machín, A.; Márquez, F. Advancements in Photovoltaic Cell Materials: Silicon, Organic, and Perovskite Solar Cells. *Materials* **2024**, *17*, 1165. [CrossRef]
21. Dharmadasa, I.M.; Alam, A.E. How to Achieve Efficiencies beyond 22.1% for CdTe-Based Thin-Film Solar Cells. *Energies* **2022**, *15*, 9510. [CrossRef]
22. Powalla, M.; Paetel, S.; Ahlswede, E.; Wuerz, R.; Wessendorf, C.D.; Magorian Friedlmeier, T. Thin-film solar cells exceeding 22% solar cell efficiency: An overview. *AIP Adv.* **2020**, *5*, 041602.
23. Renogy. Bifacial Solar Panels: Everything You Need to Know. 2024. Available online: https://www.renogy.com/blog/bifacial-solar-panels-disadvantages-and-advantages/#:~:text=Bifacial%20solar%20panels%20can%20be,boosts%20their%20overall%20energy%20output (accessed on 23 July 2024).
24. Marsh, J. Bifacial Solar Panels: What You Need to Know. 2024. Available online: https://www.energysage.com/solar/bifacial-solar-panels-what-you-need-to-know/ (accessed on 23 July 2024).
25. Kopecek, R.; Libal, J. Bifacial Photovoltaics 2021: Status, Opportunities and Challenges. *Energies* **2021**, *14*, 2076. [CrossRef]
26. Akinyele, D.O.; Rayudu, R.K.; Padayachee, N.N. Energy Storage Technologies for Residential Applications: Impacts and Prospects. *Renew. Sustain. Energy Rev.* **2017**, *68*, 1105–1117.
27. Luthander, P.; Widén, J.; Nilsson, D.; Palm, J. Photovoltaic self-consumption in buildings: A review. *Appl. Energy* **2015**, *142*, 80–94. [CrossRef]
28. Nair, M.G.; Garimella, J.; Venkatesh, S.P. Recent Advances in Solid-State and Flow Battery Technologies for Energy Storage Applications. *J. Energy Storage* **2020**, *27*, 100827.

29. Salas, A.; Carbone, R.; Hernandez, M.V. Smart Inverters for Improved Grid Stability in Residential PV Systems. *IEEE Trans. Smart Grid* **2019**, *10*, 5678–5686.

30. Ahmed, K.; Abdel-Salam, M.A.; El-Fayoumi, S. Implementation of Smart Inverters in Residential Solar Systems. *Renew. Energy* **2021**, *169*, 127–136.

31. Lopes, J.P.; Hatziargyriou, N.; Mutale, J.; Djapic, P.; Jenkins, N. Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities. *Electr. Power Syst. Res.* **2006**, *77*, 1189–1203. [CrossRef]

32. Kassem, A.; Said, S.A.M.; Al-Sulaiman, M.F. Grid-Connected PV Systems: Applications and Challenges in Hot Climates. *Renew. Sustain. Energy Rev.* **2021**, *58*, 219–235.

33. Martin, C.; Koffi, J.M.; Brandt, A. Advanced Grid Integration of Residential Solar Power: Challenges and Opportunities. *Energy Rep.* **2020**, *6*, 45–60.

34. Taylor, M. Residential Solar PV Systems: A Comparative Review of Technologies and Cost Structures. *Renew. Energy* **2021**, *155*, 1235–1245.

35. Harper, J. Advances in Solar PV Technology: Trends and Perspectives. *J. Renew. Sustain. Energy* **2020**, *12*, 041301.

36. Chen, S. Energy Efficiency of Residential Solar PV Systems. *Energy Procedia* **2019**, *159*, 246–251.

37. Li, X. Recent Developments in Perovskite Solar Cells. *Adv. Energy Mater.* **2019**, *9*, 1803246.

38. Zhang, Y. Life Cycle Assessment of Residential Solar Photovoltaic Systems. *Renew. Energy* **2020**, *150*, 302–312.

39. Boxwell, M. *Solar Electricity Handbook: A Simple Practical Guide to Solar Energy: How to Design and Install Solar Photovoltaic Systems*; Greenstream Publishing: Coventry, UK, 2017.

40. Arbab-Zavar, B.; Palacios-Garcia, E.J.; Vasquez, J.C.; Guerrero, J.M. Smart inverters for microgrid applications: A review. *Energies* **2019**, *12*, 840. [CrossRef]

41. Ueda, Y.; Kurokawa, K.; Kitamura, K.; Yokota, M.; Akanuma, K.; Sugihara, H. Performance analysis of various system configurations on grid-connected residential PV systems. *Sol. Energy Mater. Sol. Cells* **2009**, *93*, 945–949. [CrossRef]

42. Essam, Y.; Ahmed, A.N.; Ramli, R.; Chau, K.-W.; Idris Ibrahim, M.S.; Sherif, M.; Sefelnasr, A.; El-Shafie, A. Investigating photovoltaic solar power output forecasting using machine learning algorithms. *Eng. Appl. Comput. Fluid Mech.* **2022**, *16*, 2002–2034. [CrossRef]

43. Zazoum, B. Solar photovoltaic power prediction using different machine learning methods. *Energy Rep.* **2022**, *8*, 19–25. [CrossRef]

44. Platon, R.; Martel, J.; Woodruff, N.; Chau, T.Y. Online fault detection in PV systems. *IEEE Trans. Sustain. Energy* **2015**, *6*, 1200–1207. [CrossRef]

45. Mohana, M.; Saidi, A.S.; Alelyani, S.; Alshayeb, M.J.; Basha, S.; Anqi, A.E. Small-scale solar photovoltaic power prediction for residential load in Saudi Arabia using machine learning. *Energies* **2021**, *14*, 6759. [CrossRef]