Data-driven Artificial Intelligence Techniques for Enhancing Solar Photovoltaic System Performance: A Systematic Review

Leonardo Alves Messias[1], José Luis Domingos[1], Thiago Augusto Mendes[1], Saymon Fonseca Santos Mendes[2,3] and Raphael de Aquino Gomes[1,*]

[1]*Instituto Federal de Goiás, Goiânia, Brazil*
[2]*Secretaria-Geral de Governo de Goiás, Goiânia, Brazil*
[3]*Universidade Federal de Goiás, Goiânia, Brazil*

Abstract

Solar energy is increasingly recognized as a clean and renewable power source. In this context, artificial intelligence (AI) techniques have emerged as effective tools for enhancing the efficiency, reliability, and management of photovoltaic (PV) systems. This article presents a systematic review of recent literature on AI applications in solar PV, focusing on data-driven methods for energy production forecasting, fault detection, and system monitoring and control. The reviewed studies span multiple countries and employ diverse datasets, prediction horizons, and parameters to train AI models. By summarizing the state-of-the-art approaches, this review highlights how AI can optimize performance, improve reliability, and support the scalability of solar energy solutions, ultimately contributing to more sustainable and cost-effective energy systems.

Keywords

Energy Efficiency, Photovoltaic Energy Generation, Artificial intelligence, Solar Energy Forecasting

1. Introduction

The growing concern for environmental preservation has intensified the search for renewable energy sources, among which photovoltaic (PV) energy stands out as a notable option. Solar energy is recognized as one of the most promising clean and renewable alternatives, with significant potential to reduce greenhouse gas emissions [1].

According to [2], Brazil benefits from a privileged geographic location that provides exceptional potential for solar power generation. The lowest solar radiation index recorded in the country is approximately 1,500 kWh/m² per year, on the northern coast of Santa Catarina. At the same time, the highest reaches 2,350 kWh/m² in the northern region of Bahia. These values are notably higher than the annual averages in countries with consolidated solar energy markets, such as Germany (900–1,250 kWh/m²) and Spain (1,200–1,950 kWh/m²), highlighting Brazil's strong comparative advantage in this sector.

Public policies have played a crucial role in promoting the expansion of solar energy in Brazil. Tax incentives and subsidy programs have been key drivers of the sector's growth; however, further improvements are still needed to enhance economic feasibility and competitiveness.

From January to August 2023, Brazil's installed electricity capacity increased by 7.0 Gigawatts (GW), of which 6.2 GW came from solar and wind sources. This period marked the most significant recorded growth in solar generation, with renewable sources now accounting for 83.79% of the national energy

ICAIW 2025: Workshops at the 8th International Conference on Applied Informatics 2025, October 8–11, 2025, Ben Guerir, Morocco
*Corresponding author.
Leonardoalves.eng@hotmail.com (L. A. Messias); jose.domingos@ifg.edu.br (J. L. Domingos); thiago.mendes@ifg.edu.br (T. A. Mendes); saymon.mendes@goias.gov.br (S. F. S. Mendes); raphael.gomes@ifg.edu.br (R. d. A. Gomes)
https://teclimpa.ifg.edu.br/ (J. L. Domingos); https://teclimpa.ifg.edu.br/ (T. A. Mendes); https://goias.gov.br/governo/ (S. F. S. Mendes); https://teclimpa.ifg.edu.br/ (R. d. A. Gomes)
0009-0007-7263-2495 (L. A. Messias); 0000-0003-3634-3462 (J. L. Domingos); 0000-0001-6910-5722 (T. A. Mendes); 0009-0004-7699-2242 (S. F. S. Mendes); 0000-0002-6036-7678 (R. d. A. Gomes)

matrix, consolidating Brazil as a global leader in clean energy production. Solar power alone contributed 3.0 GW to the National Interconnected System (SIN) [3].

In this context, Artificial Intelligence (AI) techniques have been increasingly applied to improve both the efficiency and reliability of PV systems. As research advances, significant progress has been made in methodologies and approaches, resulting in notable improvements in forecasting, control, and fault diagnosis. Andrade et al. (2021) presented two Machine Learning (ML) tools capable of achieving up to 90% accuracy in PV generation forecasting, although at a high computational cost. Such techniques have broad applicability, including power output prediction, solar irradiance estimation, water heating, efficient PV system sizing, Maximum Power Point Tracking (MPPT) optimization, and fault detection [4].

The literature review identifies two primary approaches for analyzing PV systems. The first involves visual methods, which require specific equipment, such as thermal imaging cameras, and specialized labor for techniques like electroluminescence and UV fluorescence. These are used to detect issues such as dirt accumulation, hotspots, degradation, and cell breakage [5]. The second approach involves non-visual electrical methods, which are based on measurements of electrical parameters or, in some cases, meteorological data. This approach is particularly advantageous as it enables automation and integration into fault detection and diagnosis processes [6], while leveraging sensors already installed in PV facilities, such as current and voltage sensors, and solar irradiation meters.

This study aims to conduct a systematic literature review on the application of AI methods to PV systems using PV modules as sensors, leveraging solar radiation data and meteorological information from stations near existing PV plants. The research is justified by the need to highlight emerging technologies in the field, present different AI applications in PV systems, and assess which methods are most efficient and accurate for solar energy generation forecasting, considering data availability and quality. The findings will allow the identification of research gaps and key literature within the study scope, thereby contributing to the development of a more sustainable and environmentally responsible energy system.

2. Methodology

contributions to

The study was conducted in three distinct phases. The first focused on a bibliographic survey of fault detection in PV Plants, emphasizing methods that employ AI techniques for data analysis. The second phase focused on forecasting solar power generation. Finally, the third phase concentrated on identifying systems that monitor and control PV plants through AI-based techniques.

Table 1 outlines the methodological steps adopted for selecting bibliographic sources in this review, ensuring both the quality and relevance of the studies analyzed and following the guidelines provided by [7, 8]. The choice of databases — SCOPUS, SciELO Citation Index, and Web of Science — reflects their reputation for indexing peer-reviewed, high-impact publications across diverse fields, including renewable energy and artificial intelligence. The search string was carefully designed to capture a comprehensive set of studies at the intersection of photovoltaic systems and AI, incorporating multiple synonyms and related terms to minimize the risk of omitting relevant research. The filters applied, covering the period from 2017 to 2024 and focusing on titles, abstracts, and keywords, ensured that only recent and directly relevant contributions were included.

The inclusion criteria prioritize works that integrate AI techniques into photovoltaic systems, present quantitative performance metrics, and address at least one of the three focal areas of this review: generation forecasting, fault detection, or monitoring and control. Notably, only studies with full-text access that lacked a direct connection were excluded, guaranteeing that the analysis could be conducted with complete methodological and results information.

The exclusion criteria were established to maintain thematic and methodological coherence. Studies relying solely on thermal or photographic image analysis without AI integration were excluded, as were works focused on other renewable energy sources that lacked a direct connection to solar energy.

Table 1
Methodological steps for selecting bibliographic sources

Item	Details
Data collection	SCOPUS (Elsevier), SciELO Citation Index (Web of Science), Web of Science - Main collection (Clarivate Analytics)
Search string	("PHOTOVOLTAIC" OR "DISTRIBUTED GENERATION" OR "SOLAR POWER GENERATION") AND ("ARTIFICIAL INTELLIGENCE" OR "MACHINE LEARNING" OR "DEEP LEARNING" OR "DATA ANALYSIS") AND ("FORECASTING" OR "PREDICTION") AND ("MAINTENANCE" OR "FAULT DETECTION")
Filters used	Period: 2017 to 2024, Fields: title, abstract, keywords
Inclusion Criteria	• Papers that apply AI techniques to photovoltaic systems. • Studies that present quantitative performance data (e.g., RMSE, MAE, AUC, accuracy). • Papers that address generation forecasting, fault detection, or monitoring and control. • Full text available.
Exclusion Criteria	• Papers that exclusively address thermal or photographic image analysis without AI integration. • Studies applied to other renewable sources not directly related to solar. • Duplicate publications in databases. • Non-peer-reviewed documents (e.g., technical reports, white papers).

Duplicate publications were removed to avoid redundancy, and non-peer-reviewed documents were excluded to ensure scientific rigor. These choices collectively aim to build a robust, reliable, and focused dataset, enhancing the validity of the conclusions drawn from the literature review.

The selection of research databases is essential to ensure the quality and comprehensiveness of the results. The journals and indexing services used in this study were selected for their recognition as reputable references in the research field. Scopus, Web of Science, and SciELO Citation Index were selected due to their extensive coverage of peer-reviewed journals, books, and conference proceedings, as well as their relevance in providing high-impact publications.

Following the selection steps described earlier, a total of 202 articles with potential for inclusion were initially identified. However, several studies did not align with the specific objective of this article — for instance, research focused on thermographic and image analysis, electric vehicles, and smart cities. To refine the selection, an initial screening of titles and abstracts was conducted, resulting in the inclusion of 15 articles. Figure 2 presents the flowchart of the article selection process, summarizing the systematic filtering steps that led to the final set of studies included in this review. The identification phase began with 202 records retrieved from three major databases — SCOPUS (108), Web of Science Main Collection (85), and SciELO Citation Index (9). The first exclusion stage removed 87 papers due to duplication, being review articles, or lacking full-text availability, leaving 52 eligible studies.

In the eligibility phase, an initial screening based on titles and abstracts resulted in 31 papers; however, 18 of these were excluded because they focused exclusively on thermal or photographic image analysis without AI integration, which did not align with the study's scope. The remaining 15 articles were then assessed against the thematic criteria of the review, ensuring they addressed at least one of the three focal areas — solar generation prediction, fault detection in photovoltaic plants, or solar plant monitoring and control. No further exclusions were needed at this stage, and all 15 papers were included in the final review.

This process demonstrates a rigorous application of inclusion and exclusion criteria, ensuring that the final dataset is both methodologically sound and thematically aligned with the research objectives.

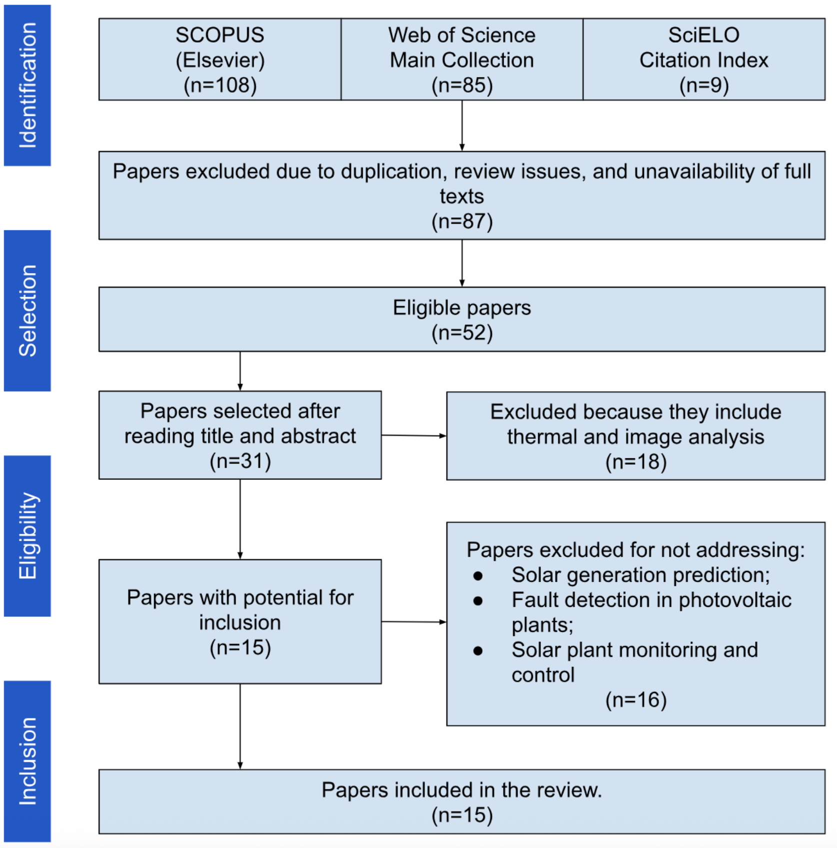

Figure 1: Flowchart of the paper selection process.

It also highlights that a substantial number of initially retrieved papers did not directly fit the scope, underscoring the importance of a structured selection approach in systematic literature reviews.

3. Discussion

In the context of technological advancements aimed at optimizing solar energy, the techniques described above play a crucial role in enhancing the efficiency of this renewable energy source. The application of these methodologies — in fault detection for PV plants, solar generation forecasting, and the monitoring and control of solar facilities — creates numerous opportunities for the solar energy market.

The significance of these techniques lies not only in optimizing technical aspects but also in enabling the economic and environmental feasibility of solar power, thereby consolidating it as a cornerstone for a sustainable future.

Figure 2 illustrates the distribution of the selected articles between 2019 and 2023, showing a peak in publications in 2020, followed by a moderate decline in subsequent years. The highest number of articles was recorded in 2020 (five publications), which may reflect a period of intensified research activity in AI applications for photovoltaic systems. While there is a recovery in 2022, with four publications, 2023 shows a slight decrease. It is noteworthy that no works are included for 2024, which is likely due to the exclusion criterion adopted in this review — only studies with full-text availability were considered. This criterion may have limited the inclusion of the most recent research, as many publications from 2024 might not yet have been fully accessible at the time of data collection.

The collaboration network shown in Figure 3 provides a visual representation of the research relationships among the authors whose studies were analyzed in this review. Each node represents an individual researcher, and the edges between nodes indicate co-authorship ties, with the numerical labels on the edges representing the number of joint publications within the corpus under review.

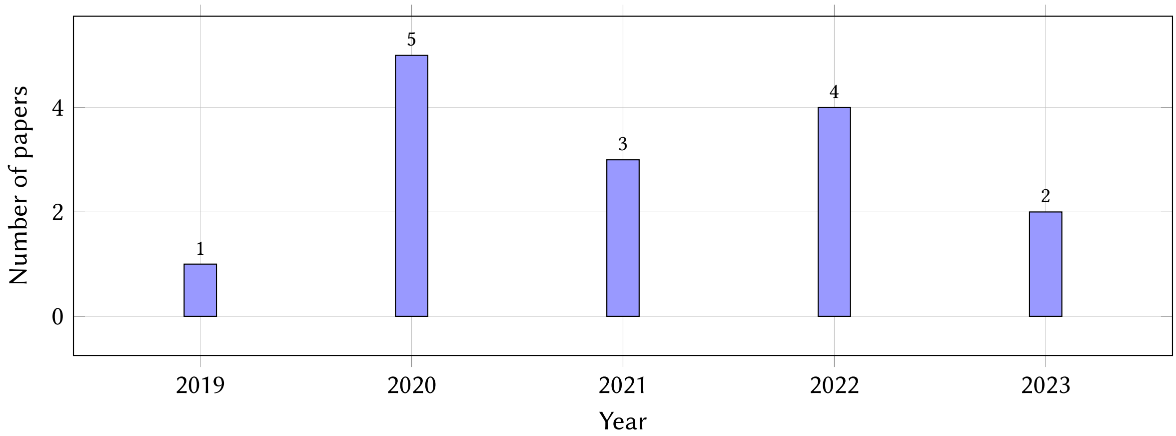

Figure 2: Distribution of reviewed papers by year of publication

This structure reveals the collaborative dynamics in the field of Artificial Intelligence applications for photovoltaic (PV) systems, highlighting both concentrated partnerships and isolated contributions.

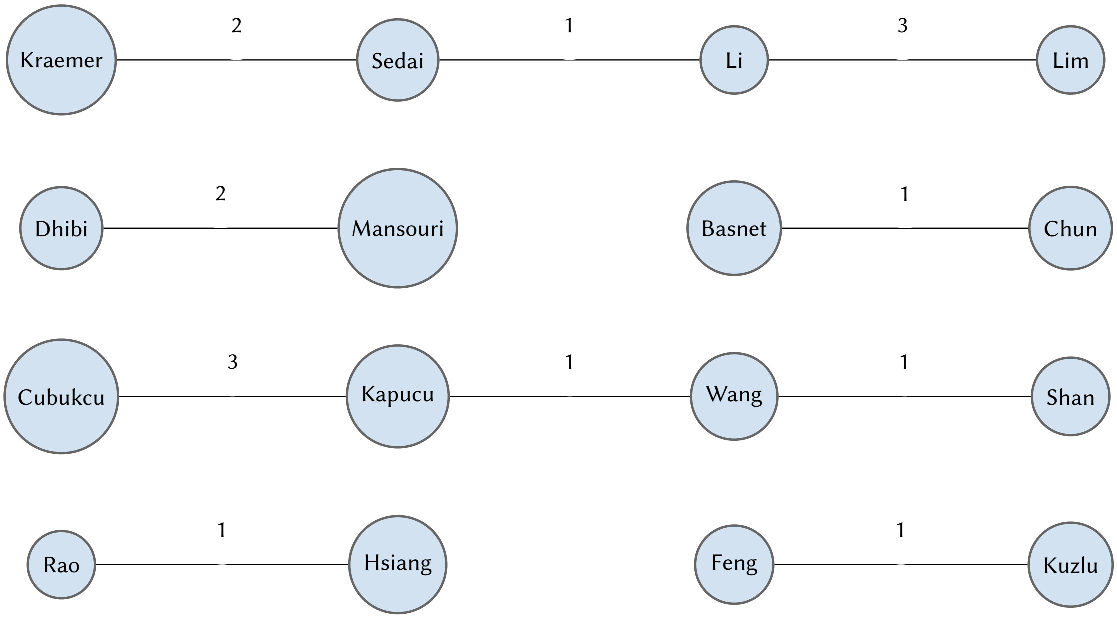

Figure 3: Co-authorship network between authors of reviewed papers, with number of interactions indicated on the edges

Several subgroups emerge, often reflecting thematic or methodological specializations. For example, Kapucu and Cubukcu form a tightly connected pair, indicating recurring joint work, most likely focused on supervised ensemble learning approaches for PV fault diagnosis. Similarly, Dhibi and Mansouri are linked through strong collaboration ties, suggesting a shared research trajectory, potentially centered on enhanced ensemble learning methods for grid-connected systems. Other pairs, such as Wang with colleagues in the KNN-based fault warning domain, and Basnet with Chun in the probabilistic neural network (PNN) approach, show targeted but less extensive collaborations.

The network also includes a set of authors — Kraemer, Sedai, Shan, Lim, Li, Feng, Kuzlu, Rao, and Hsiang — whose links reflect fewer but still relevant co-authorships. These individuals are often associated with distinct methodological contributions, such as hybrid CNN-LSTM forecasting models, feature engineering for stacked machine learning models, or explainable AI (XAI) approaches. Their lower number of shared works suggests they may be working within smaller, specialized research groups or on isolated projects that nonetheless contribute significantly to the knowledge base.

Notably, the network structure reveals moderate fragmentation, with limited interconnections between clusters. This feature indicates that while strong bonds exist within specific teams, cross-pollination of ideas between groups is less frequent. Such a pattern may result from differences in research focus (e.g., fault detection vs. forecasting vs. monitoring), geographical separation of institutions, or the use of different datasets and experimental conditions.

From a research development perspective, this fragmentation highlights opportunities for fostering broader interdisciplinary collaboration. Integrating the expertise of tightly knit groups could lead to the creation of more comprehensive AI frameworks capable of addressing multiple aspects of PV system optimization simultaneously — from early fault detection to predictive maintenance, performance forecasting, and integrated control systems. Enhanced cooperation across these clusters could also accelerate standardization of datasets, promote reproducibility, and encourage the development of hybrid methodologies that combine the strengths of different AI techniques. In the context of global challenges in renewable energy integration, such expanded collaboration has the potential to significantly advance the field and contribute to more efficient, reliable, and sustainable solar energy solutions.

3.1. Fault Detection in PV Plants

This research highlights the application of Artificial Intelligence (AI) as a promising tool for diagnosing a wide range of faults in PV systems. AI's ability to learn and process large volumes of data has proven to be particularly effective. The studies we've analyzed consistently report high accuracy rates in fault detection, instilling confidence in the effectiveness of these advanced methods to enhance the reliability and efficiency of PV systems.

Table 2 presents studies that explore AI techniques for fault diagnosis in PV systems. We will provide detailed information about each study, including the databases used, types of input data, historical time series length, study locations, data formats, and the results obtained through AI performance evaluation metrics. This will ensure that you, as our audience, are fully informed and knowledgeable about the research.

Table 2
Selected papers on Fault Detection in PV Plants

Paper title	IA technique	Authors	Journal
A Supervised Ensemble Learning Method for Fault Diagnosis in Photovoltaic Strings [9]	EL	Kapucu, K.; Cubukcu, M.	Energy
An Enhanced Ensemble Learning-Based Fault Detection and Diagnosis for Grid-Connected PV Systems [10]	EL	Dhibi, K. et al.	IEEE Access
Photovoltaic Power System Fault Warning Based on State Assessment [11]	KNN	Wang, D. et al.	IOP Publishing
An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems [12]	PNN	Basnet, B.; Chun, H.; Bang, J.	Sensors

Kapucu and Cubukcu [9] presented a system installed in a residential building in Muğla, Turkey, that collects electrical data on voltage, current, power, and irradiance from PV modules, in addition to climatic data. The setup employed Arduino™ Mega R3 and ESP8266 microcontrollers to transmit data to the cloud. The measured variables included global and tilted global radiation (in W/m^2), temperature, and humidity, all of which were stored at 30-second intervals. The specific data collection period was not reported. The approach applied 10-fold cross-validation, achieving an accuracy of 0.9767. The authors concluded that the technique could be used to identify defective PV systems, with its cloud-based implementation enabling real-time diagnostics and email notifications, thereby expanding its practical applicability.

In this context, Dhibi et al. [10] also applied Machine Learning (ML) methods for fault diagnosis in grid-connected (On-Grid) PV systems. The data originated from PV system simulators, with sampling intervals ranging from 5 to 15 seconds, although the total data collection period was not specified. The approach analyzed system behavior under adverse conditions. Simulated faults included open-circuit defects in one of the inverter switches, sensor-level issues indicating wiring or reading failures, grid disconnection faults at the common coupling point, permanent partial shading between 10% and 20%, and open-circuits or short-circuits in PV cell connections. Using a 10-fold cross-validation method, the approach achieved 100% accuracy and precision. For comparison, Recurrent Neural Network (RNN) methods yielded accuracies and precisions of 68.24% and 68.26%, respectively, while Probabilistic Neural Network (PNN) methods recorded accuracies and precisions of 62.18% and 62.21%, respectively.

Wang et al. [11] introduced an approach for early fault warning using the K-Nearest Neighbors (KNN) technique. In a study conducted on the operation of a PV plant in China, the analyzed parameters included direct current (DC) input flow, total DC input power, three-phase grid currents and voltages, frequency, total energy generation, and CO_2 emissions. Measurement intervals ranged from 5 to 30 minutes. The study also examined the impact of weather conditions on energy production, categorizing the data into sunny, cloudy, and rainy days. The technique demonstrated promising capability in classifying fault severity levels as usual, mild, and severe. The Receiver Operating Characteristic (ROC) curve yielded an Area Under the Curve (AUC) value of 0.81, a metric summarizing model effectiveness on a scale from 0 to 1, where 1 represents a perfect model.

Basnet, Chun, and Bang [12] developed a PNN-based model for analyzing data collected from a 1.8 kWp On-Grid PV plant located at Jeonbuk National University in Jeonju, South Korea. The dataset included current, voltage, irradiation levels, and temperature from sensors integrated into the PV array. Trained on data gathered under various environmental conditions, the model achieved 100% fault prediction accuracy, even in challenging scenarios, and demonstrated superior predictive performance compared to other Deep Learning (DL) classification methods.

3.2. Solar Generation Prediction

Forecasting in PV systems is essential to improve operational efficiency and maximize return on investment, addressing aspects ranging from efficient management to the mitigation of energy losses. The incorporation of AI into PV plants stands out for enhancing the prediction of maintenance costs and enabling a predictive maintenance approach. Through data analysis, it is possible to conduct an in-depth assessment of the technical and economic feasibility of PV systems, taking into account variations in energy generation capacity according to regional specificities.

Table 3 presents studies that discuss prediction systems with various purposes, indicating the datasets used, prediction periods, AI techniques applied, validation metrics, and stated objectives.

Kraemer et al. [13] proposed a study investigating short-term PV energy generation forecasting with a horizon of up to 24 hours. The dataset was obtained from Sigrist et al., comprising energy records from six distinct locations over 961 days, spanning from 2017 to 2020, which included both power and meteorological variables. Emphasis was placed on the use of Random Forest Regression (RFR), resulting in a Symmetric Mean Absolute Percentage Error (SMAPE) of 22.13% with a maximum of 30 days of training data, indicating a relatively low forecasting error. However, when replacing observed irradiance data with available irradiance forecasts, the prediction error increased to 37.79%, suggesting that most of the forecasting error could be attributed to inaccuracies in weather forecasts—specifically, discrepancies between predicted and observed irradiance.

In a complementary study, Sedai et al. [14] performed a performance analysis of multiple statistical models for long-term solar energy forecasting. The methods utilized a set of input variables, including the expected PV system power output, Direct Normal Irradiance (DNI), Global Horizontal Irradiance (GHI), Diffuse Horizontal Irradiance (DHI), and meteorological data such as temperature and wind speed. The Random Forest (RF) model outperformed others, achieving a 50% improvement in accuracy compared to univariate statistical models and a 10% improvement compared to multivariate models. Furthermore, adding five additional input variables improved the accuracy of both ML and DL models

Table 3
Selected papers on Solar Generation Prediction

Paper title	IA technique	Authors	Journal
Online Machine Learning for 1-Day-Ahead Prediction of Indoor Photovoltaic Energy [13]	RF	Kraemer, F. A. et al.	IEEE Access
Performance Analysis of Statistical, Machine Learning, and Deep Learning Models in Long-Term Forecasting of Solar Power Production [14]	RF	Sedai, A. et al.	Forecasting
Assessing the Modelling Approach and Datasets Required for Fault Detection in Photovoltaic Systems [15]	RF	Bird, M. et al.	IEEE IAS
A Deep-Learning Based Solar Irradiance Forecast Using Missing Data [16]	RNN	Shan, S. et al.	IET RPG
A Photovoltaic Power Prediction Approach Enhanced by Feature Engineering and Stacked Machine Learning Model [17]	SM	Abdelmoula, I. A. et al.	Energy Rep.
Solar Power Forecasting Using CNN-LSTM Hybrid Model [18]	CNN and LSTM	Lim, S. et al.	Energies
Photovoltaic Power Forecasting with a Hybrid Deep Learning Approach [19]	CNN and LSTM	Li, G. et al.	IEEE Access
Machine Learning Models to Quantify and Map Daily Global Solar Radiation and Photovoltaic Power [20]	PSO-ELM	Feng, Y. et al.	RSER
Gaining Insight Into Solar Photovoltaic Power Generation Forecasting Utilizing Explainable Artificial Intelligence Tools [21]	XAI	Kuzlu, M. et al.	IEEE Access

by 36%.

Bird et al. [15] used PV generation data from an extensive database covering over 100 commercial rooftop PV installations. Collected data included irradiance, solar position, wind direction and speed, ambient temperature, and relative humidity from four geographically distinct measurement sites. PV power generation, measured in kWh, was recorded hourly from 2013 to 2018. The study compared the performance of several algorithms, including Support Vector Machines (SVM), which achieved a Relative Mean Error (RME) ranging from 3.27% to 4.81%. The RF algorithm stood out, with an RME ranging from 2.54% to 4.88%, depending on the proximity of the weather stations. The study suggests a cost-effective approach to PV asset management, particularly when combined with nearby meteorological stations.

Shan et al. [16] developed a multiple-input single-output model based on Recurrent Neural Networks (RNN) to forecast DNI. The dataset, from the National Renewable Energy Laboratory, included DNI, wind speed, temperature, cloud cover, and humidity. These parameters are critical for PV plant analysis, as they allow verification of module performance against datasheet specifications. The approach was notable for its application at 4-hour intervals, even when 40% of data was missing. The authors emphasized that missing data posed minimal issues due to the larger training sample size, offering a solution for solar energy forecasting in scenarios with limited historical data availability.

In Ait Abdelmoula et al. [17], the potential of AI models for power generation forecasting was explored through a comparative evaluation of different techniques. Conducted in Benguerir, Morocco, the study analyzed data including DC power, module temperature, global horizontal irradiance, global tilted irradiance, ambient temperature, relative humidity, and wind speed, collected from October 2021 to March 2022. This period covered autumn and winter seasons, with a minimum daily threshold of 210 records, corresponding to an average of 7 generation hours per day and 30 records per hour. Among the tested models—Support Machine (SM), RF, and XGBoost—the SM achieved the highest accuracy (99%), effectively synthesizing the strengths of all base models.

Another relevant approach involved the use of Long Short-Term Memory (LSTM) networks. Lim et al. [18] proposed a hybrid model combining Convolutional Neural Networks (CNN) and LSTM for PV generation forecasting. Data from a PV plant in Busan, South Korea, were collected between September 10, 2019, and July 22, 2021, including the average annual PV generation, wind speeds, and temperatures. Weather conditions were classified into two categories: sunny and cloudy. Results showed a Root Mean Square Error (RMSE) of 4.58 on sunny days and 7.06 on overcast days, indicating better performance under sunny conditions. Similarly, Li et al. [19] also employed CNN-LSTM integration, utilizing output power data from July 1 to 10. The resulting RMSE ranged from 3.206 to 33.405 in spring, 1.664 to 25.166 in summer, 1.448 to 19.806 in autumn, and 2.062 to 21.223 in winter—demonstrating greater effectiveness than other benchmark methods.

Feng et al. [20] employed a model integrating Particle Swarm Optimization (PSO) and Extreme Learning Machine (ELM) to forecast daily solar radiation. Data were collected from seven stations on the Loess Plateau, China, including photosynthetically active photon flux density, maximum and minimum temperatures, atmospheric pressure, global solar radiation, and sunshine duration. Results showed that PSO-ELM outperformed five other ML models, especially when complete, on-site meteorological data were used as inputs, achieving the best Mean Absolute Error (MAE) of 1.649 $\mathrm{MJm^{-2}day^{-1}}$.

Finally, Kuzlu et al. [21] explored the use of Explainable AI (XAI) tools for PV energy generation forecasting. Using data from the Global Energy Forecasting Competition (GEFCOM), the study analyzed hourly PV generation and numerical weather prediction data from April 1, 2012, to July 1, 2014. The input variables included atmospheric pressure, relative humidity, cloud cover, horizontal and vertical wind speeds, air temperature, surface solar radiation, surface terrestrial radiation, total solar radiation received, total precipitation, and time of day. The goal was to interpret the internal mechanisms of an AI-based forecasting model. Results indicated an RMSE of 7.23%, demonstrating the potential of XAI for applications in smart grids.

3.3. Monitoring and Control of Solar Power Plants

Monitoring solar power plants plays a central role in the efficient management of photovoltaic (PV) systems, enabling the development of control panels and advanced management platforms that streamline and enhance the supervision of these facilities. The implementation of such systems extends beyond technical knowledge of the energy sector, requiring multidisciplinary expertise in information technology, including programming, database management, and front-end and back-end development.

As shown in Table 4, the analyzed systems are designed to manage multiple PV installations by integrating and centralizing data into a single platform. This approach offers a comprehensive view of operational performance, supports informed strategic decision-making, and enhances both preventive and predictive maintenance practices.

Table 4
Selected papers on Monitoring and Control of Solar Power Plants

Paper's title	IA technique	Authors	Journal
Machine Learning for Solar Array Monitoring Optimization and Control [22]	ML	Rao, S. et al.	Springer Nature
A Case Study on Experiment Site Selection for PV Energy Generation Forecast [23]	DL	Yu, H. et al.	ICS

Rao et al. [22] presented a monitoring and control system for PV solar power plants. The system integrates machine learning algorithms and signal processing to optimize the performance of solar panels located at the Arizona State University Research Park in the United States. It was developed using a ZigBee microcontroller connected to a wireless network, devices, and a storage server. Through sensors, the 104 solar panels of an 18 kW solar array testbed provide real-time information on MPPT voltage, MPPT current, temperature, irradiance, fill factor, temperature coefficient, maximum PV module power, open-circuit voltage, and short-circuit current, enabling fault detection and allowing changes in the connection topology. Additionally, the system employs shading predictions based on historical data and kernel regression techniques. The results showed an accuracy of 91.62% in training and 89.34% in testing—significant indicators of the method's efficiency, which benefits from in-situ data collection.

The study by Hsiang et al. [23] described the development of a web platform designed to serve as a real-time meteorological information management system and database storage solution. Data are automatically collected from sources such as OPENDATA from the Central Weather Bureau and PV generation data from Yeou-Ghi Industry Corp. in the Dou Liu district. The system processes variables such as solar radiation, wind speed and direction, temperature, and humidity, which are updated every 5 to 10 minutes. It identifies the most accurate dataset for a specific location and applies the most suitable weather forecast to the PV plant located in the Zhongzheng district, Taipei. Deep learning methods are applied, taking into account the consistency between meteorological data and energy generation, as well as the proximity of the measuring stations.

3.4. Comparison

Table 5 presents a comparative summary of the main works reviewed. A clear trend emerges in the predominance of supervised learning algorithms, with Random Forest (RF) models, various neural network architectures (CNN, LSTM, PNN), and hybrid approaches that combine classical Machine Learning techniques with Deep Learning standing out. This methodological diversity reflects an active search for solutions that can optimize both predictive accuracy and computational efficiency.

In terms of performance evaluation, accuracy remains the most widely reported metric for classification tasks. At the same time, RMSE and MAE are frequently used in forecasting studies, while AUC is employed in classification scenarios where the trade-off between sensitivity and specificity is relevant. These metrics vary according to the scope of the study. For instance, [9] and [10] reported perfect accuracy (100%) under simulated or controlled environmental conditions, while [11] achieved a relatively low SMAPE value (22.13%) for short-term generation forecasting. On the other hand, approaches such as XAI ([21]) and PSO-ELM ([20]) demonstrated competitive error rates while contributing additional interpretability or optimization features.

Despite the high accuracy levels reported, the limitations identified suggest that many of these methods face challenges in scalability and robustness when applied to real-world PV plants. For example, some works ([9, 10]) rely heavily on simulated or idealized datasets, which may not capture the full variability of field conditions. Others, such as [13] and [18], demonstrate dependence on high-quality, complete meteorological datasets, which can be costly or difficult to obtain in certain regions. These constraints underscore the importance of developing models that can maintain acceptable performance under conditions of incomplete, noisy, or regionally limited data.

Another noteworthy observation is the vast diversity of application contexts — from small-scale residential arrays ([9, 10]) to large commercial installations ([13]) and integrated monitoring platforms ([20, 21]). This diversity underscores the broad applicability of AI in the PV sector. However, it also reveals a lack of standardized datasets and evaluation protocols that could facilitate the comparison of methods across different contexts. Moreover, the coexistence of fault detection, forecasting, and monitoring/control studies suggests an opportunity for integrated frameworks that combine these functionalities in a single, unified AI system, potentially enhancing operational efficiency and reducing maintenance costs.

Table 5
Comparison of the main reviewed works on the application of AI in photovoltaic systems

Ref.	Study Objective	Dataset	AI Technique	Metric	Main Results	Limitations
[9]	Fault diagnosis in residential PV modules	Electrical data (V, I, P, irradiance) + weather, Muğla (Turkey)	Supervised Ensemble Learning	Accuracy (97.67%)	Real-time cloud-based diagnosis with automatic notifications	Data collection period not reported
[10]	Fault diagnosis in on-grid systems	Simulated PV system data	Ensemble Learning	Accuracy (100%), Precision (100%)	Outperformed RNN and PNN in all scenarios	Use of simulated data limits direct application
[11]	Early fault warning with severity classification	PV plant in China, electrical and meteorological variables	KNN	AUC (0.81)	Good classification of fault severity	Low robustness under extreme weather conditions
[12]	Fault prediction under various environmental conditions	1.8 kWp plant, South Korea	PNN	Accuracy (100%)	Better performance than other DL methods	Scalability not evaluated
[13]	Indoor PV generation forecasting (24h)	961 days of data from 6 locations	Random Forest Regression	SMAPE (22.13%)	Low error with observed irradiance	Dependence on accurate meteorological data
[14]	Long-term solar generation forecasting	Meteorological variables and irradiance	RF, SVM, DL	Percentage Error Variation	RF outperformed statistical models by up to 50%	Requires multiple input variables
[15]	Fault detection with different datasets	100 commercial facilities, 5 years of data	RF, SVM	MRE (2.54–4.88%)	RF showed best cost-effectiveness	Dependence on nearby weather stations
[16]	DNI forecasting with missing data	NREL, climate variables	RNN	RMSE (–)	Robust even with 40% missing data	Short forecasting periods (4h)
[17]	PV forecasting with feature engineering	Plant in Morocco	Stacked ML (SM, RF, XGBoost)	Accuracy (99.00%)	SM combined best performance of models	Data collected in only 2 seasons
[18]	Forecasting using CNN-LSTM	Plant in Busan, South Korea	CNN + LSTM	RMSE (4.58–7.06)	Best performance on sunny days	Significant drop on cloudy days
[19]	Seasonal hybrid CNN-LSTM forecasting	10 days per season	CNN + LSTM	RMSE (1.448–33.405)	Outperformed benchmarks	Limited temporal sample
[20]	Daily solar radiation forecasting	7 stations in China	PSO-ELM	MAE (1.649 MJ/m²d)	Best performance among 5 ML models	Requires complete meteorological data
[21]	XAI for PV generation forecasting	GEFCOM 2012–2014	XAI	RMSE (7.23%)	High model interpretability	High computational complexity
[22]	Monitoring and control of experimental PV array	104 panels, Arizona (USA)	ML + Kernel Regression	Accuracy (89.34%)	Connection optimization and shading detection	Scope limited to experimental environment
[23]	Web platform for PV generation forecasting	CWB data and plant in Taipei	DL	–	Integration of climate and energy data	Dependence on external data sources

4. Final Remarks

When applying AI tools for the management and sizing of solar energy solutions, several important conclusions emerge. First, AI can substantially enhance the performance of solar systems. This improvement is achieved through the optimization of solar panel positioning, accurate prediction of energy demand, and facilitation of preventive maintenance. These factors contribute to a more efficient use of solar energy, leading to increased energy production and lower operational costs. Such improvements not only enhance the economic value of solar power but also underscore AI as an indispensable ally in addressing the technical and operational challenges faced by the renewable energy sector.

Another significant benefit is enhanced sustainability. The application of AI in solar energy management is a crucial step toward transitioning to renewable energy sources and reducing dependence on fossil fuels. By maximizing system efficiency, AI plays a vital role in mitigating climate change and promoting more sustainable energy practices. This advancement, therefore, not only meets contemporary energy needs but also safeguards natural resources for future generations, underscoring AI's essential role in shaping a greener and more sustainable energy future.

However, challenges must be addressed when implementing AI tools for solar energy management. The complexity of implementation is a significant factor, as integrating AI systems often requires substantial technical expertise and resources. Additionally, the initial cost of implementation can be high and prohibitive for some organizations, particularly those with limited resources.

Data security is also a critical concern. Protecting the data used by AI tools—including information on solar system performance and usage patterns—is essential to ensure confidentiality and integrity. Nonetheless, these challenges are not insurmountable. With the continuous advancement of technology and the development of more accessible and secure management solutions, it is now possible to maximize AI's potential in transforming the solar energy sector.

In summary, the implementation of AI tools in solar energy management marks a significant milestone in the pursuit of more efficient, sustainable, and reliable energy solutions. Despite initial challenges, the potential for performance improvement, sustainability promotion, and resource optimization makes AI an indispensable transformative force in the renewable energy sector.

Acknowledgments

Fundação Sitawi supported the work of Leonardo Alves Messias through a permanence scholarship (Commitment Term n. CRF_ALPB_2023_108), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo a Pesquisa do Estado de Goiás (FAPEG) supported the work of Raphael de Aquino Gomes.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] O. B. Ogundipe, A. C. Okwandu, S. A. Abdulwaheed, Recent advances in solar photovoltaic technologies: Efficiency, materials, and applications, GSC Advanced Research and Reviews 20 (2024) 159–175.

[2] A. R. O. da Rosa, F. P. Gasparin, Panorama da energia solar fotovoltaica no brasil, Revista brasileira de energia solar 7 (2016) 140–147.

[3] Ministério de Minas e Energia, Planos e Informações Energéticas, Available in https://www.gov.br/mme/pt-br/planos-e-informacoes-energeticas, 2024. Access in: 4 aug 2025.

[4] S. Kalogirou, A. Sencan, Artificial intelligence techniques in solar energy applications, Solar collectors and panels, theory and applications 15 (2010) 315–340.

[5] G. M. Tina, F. Cosentino, C. Ventura, Monitoring and diagnostics of photovoltaic power plants, in: Renewable Energy in the Service of Mankind Vol II: Selected Topics from the World Renewable Energy Congress WREC 2014, Springer, 2015, pp. 505–516.

[6] S. Ahmed, A. Ali, J. A. Ansari, S. A. Qadir, L. Kumar, A comprehensive review of solar photovoltaic systems: Scope, technologies, applications, progress, challenges and recommendations, IEEE Access (2025).

[7] S. Keele, et al., Guidelines for performing systematic literature reviews in software engineering, Technical Report, Technical report, ver. 2.3 ebse technical report. ebse, 2007.

[8] B. Kitchenham, S. Charters, et al., Guidelines for performing systematic literature reviews in software engineering (2007).

[9] C. Kapucu, M. Cubukcu, A supervised ensemble learning method for fault diagnosis in photovoltaic strings, Energy 227 (2021) 120463.

[10] K. Dhibi, M. Mansouri, K. Bouzrara, H. Nounou, M. Nounou, An enhanced ensemble learning-based fault detection and diagnosis for grid-connected pv systems, IEEE Access 9 (2021) 155622–155633.

[11] D. Wang, X. Xie, Z. Na, W. Shen, X. Fan, Photovoltaic power system fault warning based on state assessment, in: IOP Conference Series: Earth and Environmental Science, volume 632, IOP Publishing, 2021, p. 032063.

[12] B. Basnet, H. Chun, J. Bang, An intelligent fault detection model for fault detection in photovoltaic systems, Journal of Sensors 2020 (2020) 6960328.

[13] F. A. Kraemer, H. A. Asad, K. Bach, B.-C. Renner, Online machine learning for 1-day-ahead prediction of indoor photovoltaic energy, IEEE access 11 (2023) 38417–38425.

[14] A. Sedai, R. Dhakal, S. Gautam, A. Dhamala, A. Bilbao, Q. Wang, A. Wigington, S. Pol, Performance analysis of statistical, machine learning and deep learning models in long-term forecasting of solar power production, forecasting 5 (2023) 256–284.

[15] M. Bird, S. Acha, N. Le Brun, N. Shah, Assessing the modelling approach and datasets required for fault detection in photovoltaic systems, in: 2019 IEEE Industry Applications Society Annual Meeting, IEEE, 2019, pp. 1–6.

[16] S. Shan, X. Xie, T. Fan, Y. Xiao, Z. Ding, K. Zhang, H. Wei, A deep-learning based solar irradiance forecast using missing data, IET Renewable Power Generation 16 (2022) 1462–1473.

[17] I. Ait Abdelmoula, S. Elhamaoui, O. Elalani, A. Ghennioui, M. El Aroussi, A photovoltaic power prediction approach enhanced by feature engineering and stacked machine learning model, Energy Reports 8 (2022) 1288–1300.

[18] S.-C. Lim, J.-H. Huh, S.-H. Hong, C.-Y. Park, J.-C. Kim, Solar power forecasting using cnn-lstm hybrid model, Energies 15 (2022) 8233.

[19] G. Li, S. Xie, B. Wang, J. Xin, Y. Li, S. Du, Photovoltaic power forecasting with a hybrid deep learning approach, IEEE access 8 (2020) 175871–175880.

[20] Y. Feng, W. Hao, H. Li, N. Cui, D. Gong, L. Gao, Machine learning models to quantify and map daily global solar radiation and photovoltaic power, Renewable and Sustainable Energy Reviews 118 (2020) 109393.

[21] M. Kuzlu, U. Cali, V. Sharma, Ö. Güler, Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools, Ieee Access 8 (2020) 187814–187823.

[22] S. Rao, S. Katoch, V. Narayanaswamy, G. Muniraju, C. Tepedelenlioglu, A. Spanias, Machine learning for solar array monitoring, optimization, and control, Springer Nature, 2022.

[23] H. Y. Hsiang, T. S. Yuan, P. Wang, L. W. Hui, L. H. Chung, A case study on experiment site selection for pv energy generation forecast, in: 2020 International Computer Symposium (ICS), IEEE, 2020, pp. 468–472.