Big-Data-Driven AI in Predictive Maintenance for Photovoltaic Plants

Qing Ma[1,*], Jingran He[2], Yuan Jing[1], Xiangxue Lv[3], Jia Liao[3], Haimin Li[3,4]

[1] Gas Transmission Management Department of Petro China Southwest Oil & Gasfield Company, PetroChina Southwest Oil & Gas Field Company, Chengdu 621000, China

[2] Gas Transmission Management Department of Petro China Southwest Oil & Gasfield Company, PetroChina Southwest Oil & Gas Field Company, Nanchong 637000, China

[3] School of New Energy and Materials & Modern Industry College of Photovoltaic New Energy, Southwest Petroleum University, Chengdu 610500, China

[4] College of Materials and Energy, Guang'an Institute of Technology, Guang'an 638550, China

Corresponding Author: Qing Ma (maqiang02@petrochina.com.cn)

Abstract: As photovoltaic (PV) systems rapidly scale and decentralize, conventional manual inspection is proving inadequate in coverage, timeliness, and cost-efficiency. A paradigm shift in operation and maintenance (O&M) is urgently needed—one that deeply integrates big data predictive analytics with artificial intelligence (AI) to enable proactive, autonomous, and scalable management. This review systematically consolidates recent advances and proposes a comprehensive data driven intelligent inspection framework that unifies multi source sensing, predictive fault diagnosis, spatiotemporal risk mapping, and closed loop task orchestration. Beyond synthesizing the state of the art, we critically examine persistent challenges including data drift, class imbalance, and cross domain generalization, and survey emerging strategies such as online incremental learning and transfer learning. To accelerate future research and deployment, we advocate for the establishment of standardized benchmark datasets spanning diverse climatic and topological conditions, along with a unified fault labeling system and physics aware evaluation metrics. This review aims to provide both a technical roadmap and a critical foundation for transitioning PV O&M toward intelligent, predictive, and sustainable practices.

Keywords: Photovoltaic power plants; AI-driven; Big data; Intelligent inspection

1. Introduction

1.1 Research Background

With the global energy transition accelerating, PV plants—now the backbone of renewable generation—are expanding at an exponential rate. Yet the same scale that underpins their economic appeal simultaneously magnifies the long-standing limitations of manual inspection: human patrols struggle to cover array fields that frequently exceed several thousand hectares and are often suspended during sandstorms or blizzards(Abubakar et al. 2021a); moreover, inspection outcomes remain highly idiosyncratic, resulting in recurrent misdetections and omissions that fall short of the precision demanded by modern operation and maintenance (O&M)(Abubakar et al. 2021b). To overcome these bottlenecks, unmanned-aerial-vehicle (UAV) and robotic solutions have been

successively deployed. UAVs leverage elevated vantage points and rapid mobility to acquire high-resolution visible and thermal imagery, thereby extending the effective inspection radius; however, both flight-path design and defect interpretation still hinge on human input, preventing closed-loop intelligent decision-making. Conversely, ground robots can operate 24 h within narrow intra-station corridors, yet their limited reach precludes full-site coverage. Consequently, neither technology alone reconciles the dual objectives of "wide-area coverage" and "fine-grained perception."

Embedding artificial-intelligence (AI) techniques has emerged as the pivotal step toward bridging this gap. Deep-learning-based computer-vision models now enable automatic parsing of the heterogeneous, multi-modal data collected by UAVs and robots, yielding substantial gains in the detection accuracy of hotspot, crack, and soiling defects. Moving beyond static recognition, large-scale temporal analytics and graph-neural-network models forecast defect propagation trajectories, shifting the maintenance paradigm from reactive repair to proactive warning and markedly reducing downtime losses and O&M expenditures (Aghaei 2022). Reinforcement-learning-driven path-planning algorithms further refine system-level energy efficiency by dynamically optimizing UAV routes and robot dispatch strategies in real time, integrating meteorological forecasts, equipment health indicators, and task priorities. Figure 1 shows the schematic diagram of a photovoltaic system with AI-based MPPT control.

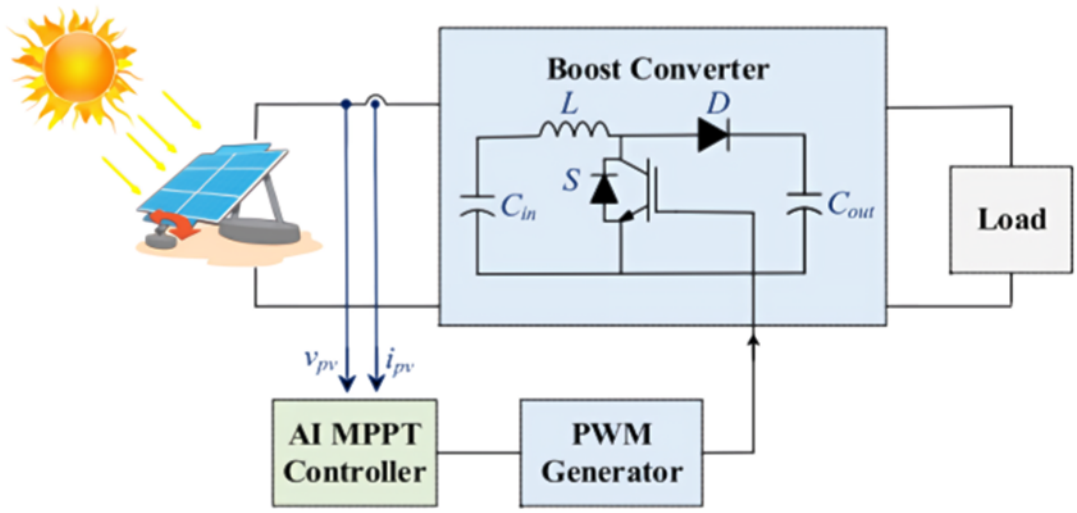


Figure 1: A photovoltaic system with an AI-based MPPT controller: PWM-Pulse width modulation; MPPT-Maximum power point tracking.

Nevertheless, current AI-enabled PV inspection systems confront three intertwined challenges. First, task allocation and spatiotemporal coordination between UAVs and robots remain inefficient, depressing overall resource utilization (Ahan et al. 2021). Second, edge-device computational constraints coexist with cloud-side latency, frustrating the simultaneous demands of real-time analytics and continual model updating for massive inspection streams. Third, algorithmic robustness and interpretability under harsh scenarios—sandstorms, snow cover, and other environmental extremes—still lack systematic validation. Therefore, a research agenda centered on the synergistic coupling of "big-data prediction" and "AI orchestration" is both theoretically valuable and urgently

needed. Priority should be given to multi-agent collaborative frameworks, edge-cloud elastic computing architectures, and adaptive model mechanisms under extreme conditions to elevate PV plant O&M efficiency and safeguard the reliability of the broader energy system.

1.2 Research Objectives

Drawing on multi-source heterogeneous data and state-of-the-art algorithms, this review proposes a big-data-driven AI inspection framework for PV plants. The framework enables real-time monitoring of module-level operating conditions, early-stage fault warning and closed-loop optimization of O&M decisions (Yao et al. 2021). Specifically, three objectives are pursued. (1) To overcome the limitations of conventional manual inspection—namely its dependence on experiential judgement, delayed response and incomplete coverage—predictive models are constructed that can learn the temporal evolution of array-level power generation, ambient parameters and device-ageing patterns, thereby identifying latent failure risks before they propagate (Alam et al. 2023). (2) To address the poor generalizability and high false-alarm rates of existing AI algorithms in complex field scenarios, a deep-learning-based adaptive inspection architecture is developed that synergistically fuses thermal infrared, visible-light and UAV remote-sensing data to extract fault-related features and perform dynamic classification with high fidelity (Ali and Dasgupta 2023). (3) To break the information silos created by single-source data, a PV-specific digital-twin system is built that deeply integrates real-time sensor streams with historical O&M logs, meteorological forecasts and grid-dispatch information (Ananthasaravanan and Megala 2026). Ultimately, the review provides a theoretical basis and technical roadmap for data-driven, predictive maintenance of PV plants, facilitating the transition of renewable-energy O&M toward intelligent, prognostic paradigms (Ansari et al. 2025).

1.3 Research Significance

Against the backdrop of China's "dual-carbon" strategy, cumulative installed PV capacity has surpassed 500 GW. Yet conventional manual inspections remain inefficient: a 100 MW plant typically requires 15–20 days for full coverage, suffers a 15% miss rate, and consequently incurs annual yield losses of 3–8%—equivalent to hundreds of millions of renminbi in foregone revenue (Ayturan et al. 2025). Hence, an AI-driven inspection paradigm that couples big-data forecasting with edge computing is urgently needed.

By integrating heterogeneous data streams—meteorological, electrical, and high-resolution imagery—and leveraging deep-learning models optimized for edge deployment, this framework achieves predictive fault detection and centimeter-level localization (Bello et al. 2024). The proposed approach has been shown to reduce O&M costs while enhancing generation efficiency (Bouchetob and Nadji 2025, Cardinale-Villalobos et al. 2023, Carletti et al. 2020). Beyond offering a replicable smart-O&M template for large-scale PV plants, these findings provide pivotal theoretical and technical support for the digital transformation of the energy sector (Chen 2024).

2. Data Of PV Power Stations

2.1 Sensor

Within the research framework of intelligent monitoring for PV systems, the sensing layer is regarded as the primary conduit for acquiring raw operational data; the heterogeneity of its configuration directly governs the fidelity of all subsequent analytics and fault-diagnosis routines.

S. Chorozoglou et al , focusing on string-level electrical characteristics, emphasize that a PV cell establishes a highly non-linear and time-varying relationship between its environmental conditions and its power output (Chorozoglou et al. 2023) Along the current-voltage (I-V) curve of a solar cell (Figure 2(a)), the point at which power is maximized-commonly referred to as the maximum power point (MPP)—is intrinsically dynamic. Irradiance, cell temperature, and the cell's own health continuously reshape the curve, rendering the MPP a non-stationary target. To capture this variability, synchronized I-V curve tracers are deployed. Operating at millisecond refresh rates, these devices record high resolution voltage–current trajectories that serve as electrical fingerprints for both maximum power point tracking (MPPT) and component level health assessment. (De Riso et al. 2024) (Figure 2(b) for an overview of imaging techniques used in PV module inspection).

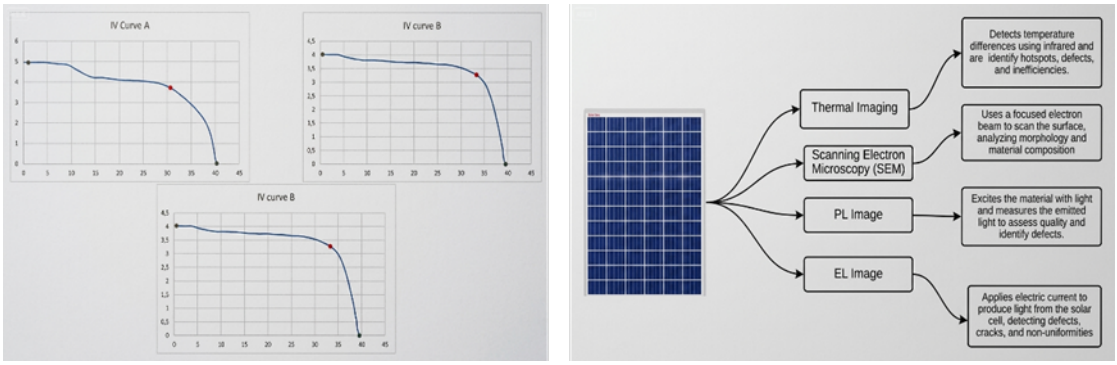


Figure 2: (a) Different IV curves (Chorozoglou et al. 2023); (b) Example of various imaging techniques to identify defects in PV module.

To detect both surface and subsurface defects in PV modules, three complementary optical modalities were integrated: infrared (IR) thermography, visible-light imaging, and electroluminescence (EL). IR thermography identifies latent soldering defects or localized shading by mapping thermal anomalies; a high-resolution RGB camera localizes surface features such as cracks, foreign objects, and dust accumulation; and EL imaging, performed under dark conditions with reverse-injected current, provides high-contrast visualization of micro-cracks and cell fractures. The synergistic use of these techniques significantly enhances defect detection accuracy(Devanathan et al. 2025).

Accurate environmental data are crucial for fault attribution in PV systems. To this end, each research site was equipped with a micro-automatic weather station to synchronously record key parameters, including total irradiance on both horizontal and inclined planes, ambient temperature, relative humidity, wind speed and wind direction, thereby providing dynamic boundary conditions for the power prediction model and the analysis of hot spot formation mechanisms(Djabali et al. 2024).

Furthermore, to overcome the limited spatial coverage of ground-based sensors, the latest research has introduced UAV platform with Light Detection and Ranging (LiDAR) and multispectral sensors. This approach enables the reconstruction of centimeter-level 3D point clouds and the measurement of spectral reflectance. The LiDAR data precisely estimates module tilt angles, heights and occlusion geometry, while the multispectral information supports the quantitative assessment of large-scale dust distribution and vegetation overgrowth. Collectively, UAV-based sensing compensates for the blind spots inherent in ground-based point measurements at the system level (El

Karch et al. 2024).

2.2 PV O&M for Economic Synergy

The research demonstrates the integration of technology and operational management through a clear logical framework. Regarding data acquisition and monitoring systems, N. El Yanboiy et al. noted that the photovoltaic system utilizes a Meteocontrol monitoring system, with data collected by a "blue 'Log XM-200' data logger" at configurable intervals (e.g., 5 minutes, hourly, daily), with further technical details provided in Table 1 (El Yanboiy et al. 2025). Complementing this, T. A. Felipe, et al. described a system where sensors continuously monitor component voltages and upload data in real-time to a Raspberry Pi controller. Together, these studies outline a comprehensive three-tier monitoring architecture—encompassing on-site sensors, edge controllers, and centralized data logging—which establishes a robust data foundation for subsequent analysis and decision-making (Felipe et al. 2021).

Table 1: Technical information of the PV system at Universidad de los Andes.

Technical Spec	System A	System B
AC Power (kW)	51.24	28.82
Tilt (°)	10	10
Azimuthal (°)	180	180
PV Panels per String	16	18
Strings per Inverter	8	4
Inverter	ABB TRIO 50.0TL OUTD US	ABB TRIO 27.6TL OUTD US
PV Panel	LG Electronics LG400N2W-A5	LG Electronics LG400N2W-A5

The defect handling process also yields significant benefits for operational management, particularly in spare parts management and economic dispatch. In the domain of spare parts, Wang, Y et al. emphasized that early defect detection during manufacturing allows for the replacement or repair of damaged cells, preventing substantial losses in plant efficiency and performance (Wang and Xing 2026). Thus, proactive component management is essential. Beyond physical maintenance, effective defect diagnosis and power forecasting are directly tied to economic optimization. (Wasser et al. 2025) As F. Gao et al. point out, accurate prediction of PV power generation enables the formulation of advance dispatch and operational strategies, which is crucial for optimizing the supply-demand balance between photovoltaic power and the conventional grid and, ultimately, for enhancing economic returns (Gao et al. 2024).

Looking ahead, A. Ghahremani et al. state in their abstract that "the continuous advancement of artificial intelligence technology and hardware will promote the rapid global adoption of solar energy and other renewables, with the potential to gradually displace traditional energy sources"(Ghahremani et al. 2025). This vision of widespread renewable integration underscores the critical importance of high-precision forecasting. Such forecasting capabilities serve a dual purpose: it provides essential inputs for economic dispatch on the grid side, while enabling individual photovoltaic plants to maximize profitability amid electricity price fluctuations. Ultimately, it translates technical research advances into a tangible competitive advantage in market-driven operations.

2.3 Multifaceted Data Quality Challenges

As PV systems evolve toward ultra-large-scale and distributed architectures, the volume of operation data has rapidly reached the Petabyte (PB) level and exhibits significant heterogeneity-encompassing cross-domain signals such as electrical, infrared, meteorological, and image data, which poses substantial challenges to unified modeling (Ghenai 2024). Concurrently, data loss issues are growing more severe. These include both "physical loss" due to sensor failure and "effective loss" caused by a sudden drop in the signal to noise ratio under cloudy, rainy, or snowy conditions; both types can undermine the robustness of supervised learning models. Furthermore, data drift—jointly induced by equipment aging and environmental fluctuations—causes the training set distribution to gradually deviate from that of online data, leading to performance degradation in earlier models. To address these challenges, H. Haddad et al. implemented a fault detection algorithm; its information flow steps (numbered 1–6) are illustrated in Figure 3. A common research direction for mitigating the data drift problem is the adoption of online incremental learning or transfer learning strategies. However, these approaches typically come at the cost of increased demand for computing resources and annotated data (Haddad et al. 2024).

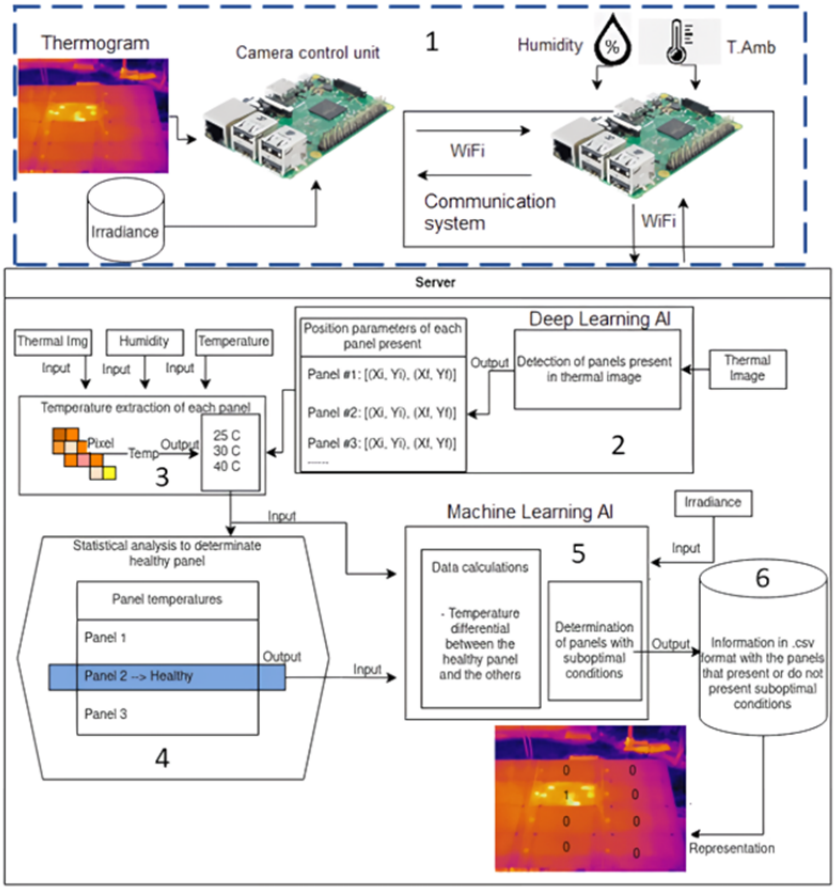


Figure 1: Implemented fault detection algorithm. The flow of information is listed from 1 to 6. (Haddad et al. 2024)

It is worth noting that the spatio-temporal synchronization problem further exacerbates the inconsistency of data quality. On the one hand, the sampling clock differences among different sub-sites can reach the millisecond level, and direct superposition will generate false transients. On

the other hand, there are differences in spatial resolution and viewing Angle among multi-platform observations such as satellites, ground, and unmanned aerial vehicles, which leads to misrepresentation of the same fault in thermal images and I-V curves.

3. Big Data Prediction Technology Stack

3.1 Irradiance-centric PV forecasting with hybrid AI models

Recent evidence indicates that artificial intelligence continues to play a central role in photovoltaic power systems (Zhang et al. 2025a). Across prediction models, solar irradiance is consistently regarded as the primary meteorological factor determining PV output power. Hamz, A et al. pointed out that irradiance has a non-linear positive correlation with component power: when irradiance increases, the component current significantly increases while the voltage only slightly increases, thereby enhancing the overall power (Hamz et al. 2025). Zhang ZX et al. further explained through the I-V/ power-voltage (P-V) characteristic curve that when the irradiance increased from 400 $W \cdot m^{-2}$ to 1000 $W \cdot m^{-2}$ (Figure 4(a) and 4(b)), the current increase at the MPP could reach 2.5 times, while the voltage change was less than 5%. This physical mechanism provides a theoretical basis for subsequent prediction models to take irradiance as the primary input feature (Zhang et al. 2022).

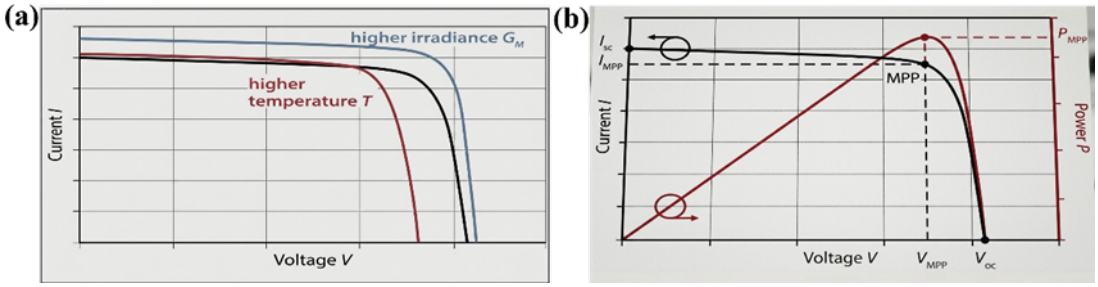


Figure 4: (a)Effect of irradiance (GM) or temperature (T) on PV Module I-V curves. (Durán et al. 2012) (b) Standard I-V/ P-V curve with the maximum power point. (Durán et al. 2012)

At intra-hour and intra-day scales, the literature generally adopts high-resolution irradiance sampling to capture rapid fluctuations. For instance, J. Hu et al. performed power prediction (with forecast horizons from 5 minutes to 1 hour) using a Long Short-Term Memory (LSTM) network. Their model utilized 1-minute irradiance data, combined with real-time temperature, humidity, and historical power data, to construct a multi-dimensional input tensor, which was then normalized via the Min-Max method to address feature scale differences(Hu et al. 2024). The experimental results show that after introducing the high-frequency component of irradiance, the Root Mean Square Error (RMSE) of the LSTM-100 epoch model on the test set drops to 1.06%, and the R^2 (R: Pearson correlation coefficient) increases to 0.968, which is significantly better than the control model without incorporating the dynamic characteristics of irradiance. As shown in Table 2. (Test Mse: Test Mean Squared Error, Train Mae: Training Mean Absolute Error, Test Mae: Test Mean Absolute Error, Train R^2: Training Coefficient of Determination, Test R^2: Test Coefficient of Determination). The RMSE is de fined as:

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

Table 2: Evaluation Metrics for LSTM Models with Epochs (20,50,100)

Test Mse	1.2878	1.1574	1.1212
Train Mae	0.5834	0.4936	0.5054
Test Mae	0.5833	0.4964	0.5136
Train R^2	0.9651	0.9693	0.9704
Test R^2	0.9634	0.9671	0.9681

Furthermore, in their review, H. K. Huneria et al. note that the PSO ANN hybrid algorithm can control power prediction error within ±3 %. This is achieved by online updates of the irradiance weight coefficient, enabling the algorithm to track irradiance fluctuations caused by cloud shading at a 10 minute timescale (Huneria et al. 2021).

The processed data was fed into a coupled LSTM-CNN (Xue et al. 2024) (CNN: Convolutional Neural Network) model, which was trained for 20, 50, and 100 epochs (Figures 5(a),5(b), and, 5(c)). Among all configurations, the model trained for 100 epochs achieved the best performance, reducing the Mean Absolute Percentage Error (MAPE) to 4.9% for the 24-hour-ahead prediction task. This result validates the significant contribution of spatial irradiance downscaling to improving mid-term forecast accuracy (Ibrahim et al. 2026).

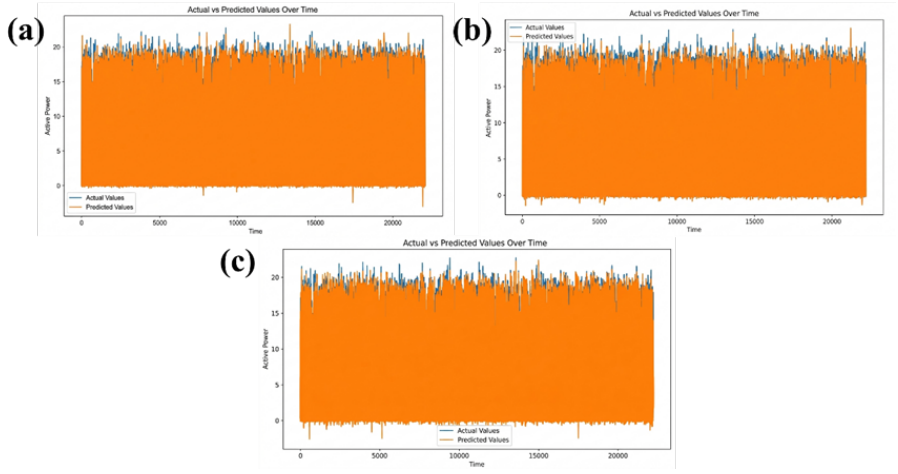


Figure 5: (a) Predicted PV Output Using LSTM Model (20 Epochs) (Ibrahim et al. 2026); (b) Predicted PV Output Using LSTM Model (50 Epochs); (c) Predicted PV Output Using LSTM Model (100 Epochs).

S. M. Jaafar et al. further demonstrated that integrating satellite derived regional irradiance fields can increase the medium term prediction model's R^2 from 0.81 to 0.93 in cloudy climates (Jaafar et al. 2025).

To enhance prediction robustness across different time scales, a common strategy in the literature is to adopt multivariate modeling that jointly considers irradiance, temperature, wind speed, and humidity. For example, Tang, W et al. (Tang et al. 2023) implemented an Internet of Things (IoT)-cloud architecture where irradiance and temperature sensors are deployed on each PV module. Data is uploaded to a cloud-based Artificial Neural Network (ANN) every 30 seconds to perform rolling forecasts with a 15-minute prediction horizon. Their results show that with the introduction of

real-time irradiance correction, the Relative Root Mean Square Error (RRMSE) was reduced from 5.68% to 4.03%.

D. F. Josè et al. compared four maximum power point tracking (MPPT) strategies: Perturb and Observe (P&O), Incremental Conductance (INC), Fuzzy Logic Controller (FLC), and Artificial Neural Network (ANN). They found that within an irradiance range of 500–1000 $W \cdot m^{-2}$, the ANN model—by learning the irradiance power mapping online—reduced tracking error to approximately 2 %, compared to 5–7 % for traditional methods. This improvement indirectly enhances the input data accuracy for subsequent power prediction models (Josè et al. 2025). Among these, the P&O algorithm is one of the most used MPPT methods. The INC MPPT rules are briefly overviewed in Figure 6. The flowcharts for fuzzy logic control and ANN based MPPT are presented in Figures7(a) and 7(b), respectively.

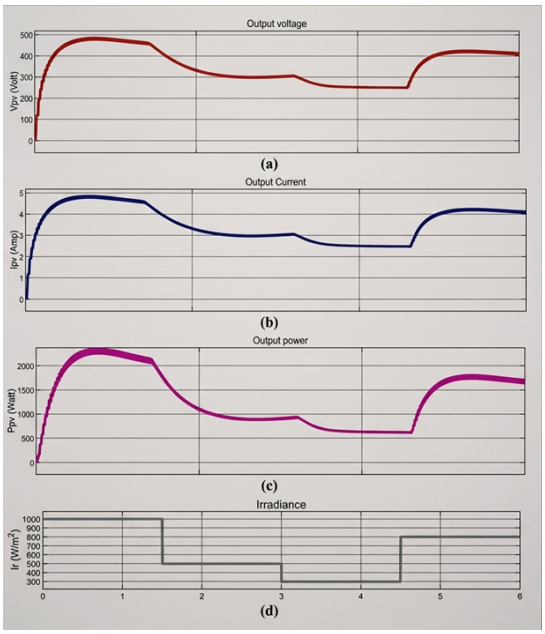


Figure 6: Simulation result for PV module under different shading/irradiance condition by using Perturb and Observe algorithm: (a) Variation of output voltage with irradiance; (b) Variation of output current with irradiance; (c)Variation of output power with irradiance; (d) Variation of irradiance with time. (Sahoo et al. 2020)

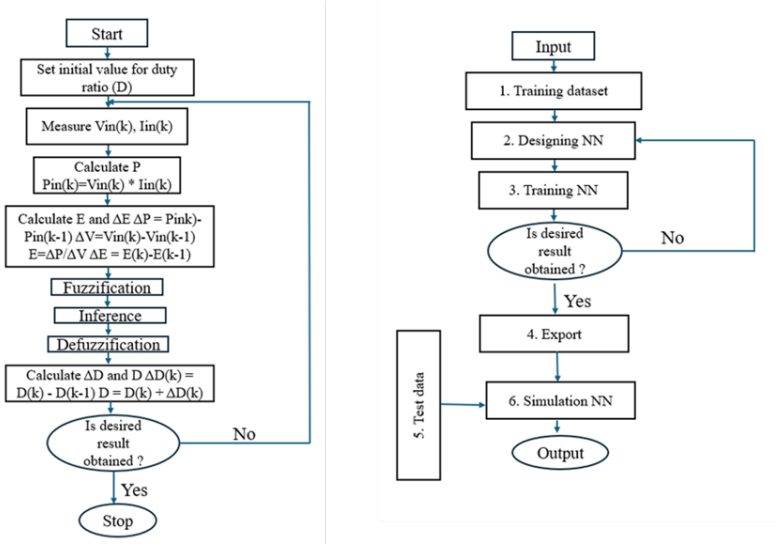


Figure 7: (a) Flowchart of the fuzzy logic control.(Josè et al. 2025); (b) Flowchart of an artificial neural network. (Josè et al. 2025)

3.2 Quantification of Uncertainty

The uncertainty of PV systems originates from the intermittency of solar irradiation and dynamic weather changes, which makes traditional point predictions inadequate for risk assessment. To handle these challenges, Zhang, L et al propose an incremental probabilistic prediction method based on a Bayesian stochastic configuration network (BSCN) and personalized federated learning (PFL)(Zhang et al. 2025b). The results of this AI based digital twin (DT) framework are presented in Figure 8(a).

During evaluation, prediction interval calibration and sharpness were assessed using PICP, MPIW, and CRPS, with the reliability of the 95% confidence interval validated via Monte Carlo simulation. At the application level, the predicted distribution's upper and lower quantiles are fed in real time into a robust MPPT controller and grid dispatch system (Yap et al. 2020; Ibrahim et al. 2025), enabling dynamic adjustment of the maximum power point search space and providing risk-aware constraints for electricity market bidding. In related irradiation studies, Jandaghian et al. investigated photovoltaic cells irradiated by a proton beam in a plasma focus device (Jandaghian et al. 2023). Future work should investigate non-parametric probabilistic prediction based on deep generative models to capture more complex, multimodal uncertainty structures.

Figure 8(b) compares the accuracy of various uncertainty quantification models. Related work also treats uncertainty explicitly. For instance, Kumar A et al. systematically reviewed artificial intelligence techniques for photovoltaic systems, including evaluation and benchmarking analyses of the principal model families (Kumar et al. 2024a). In a related maintenance context, Kumar RK et al. developed an AI-enabled robotic maintenance system for solar panel health monitoring (Kumar et al. 2024b). Similarly, V. S. B. Kurukuru et al. introduced uniform noise into fault detection data to generate multiple operational scenarios, using confusion matrices to delineate the uncertainty boundaries of fault classification (Kurukuru et al. 2021).

4. Big Data×AI

Firstly, in the “prediction” stage, research focuses on leveraging imaging data for early anomaly detection and power forecasting.

For power prediction based on AI recognition, Ge et al established hybrid eXtreme Gradient Boosting (XGBoost) and Bidirectional LSTM (Bi LSTM) architecture performs rolling 5 minutes ahead power curve predictions, which serve as input for subsequent risk classification (Ge et al. 2025). In parallel, hybrid deep-learning models such as LSTM-XGBoost have been widely applied to short-term photovoltaic power prediction (Zhu et al. 2025), while popular vision algorithms like YOLO, whose basic architecture is shown in Figure 9, often support such recognition tasks.

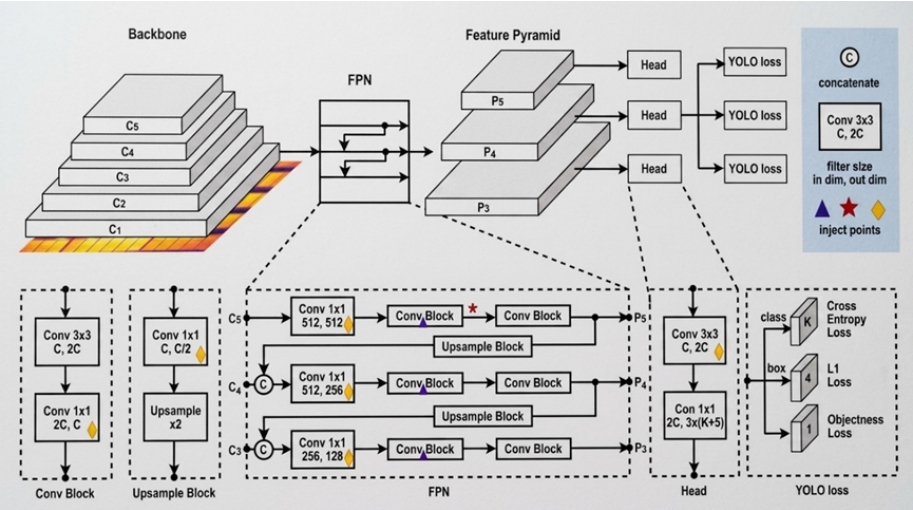


Figure 9: Basic YOLO Architecture

Subsequently, at the "task assignment" level, the UAV AI (Wang et al. 2025) construction management system translates the risk map directly into route planning and maintenance work orders. Specifically, the system automatically adjusts UAV flight altitude, overlap rate, and shooting frequency based on the risk level and generates a task list—including geographic coordinates, defect types, and priorities—on the backend BIM platform (Quater et al. 2014).

Finally, in the "defect label update" stage, Carletti et al. defined a public dataset containing 14,215 component images, 3,654 local hotspots, and 76 global hotspots. They consolidated detection results into new defect labels by applying the PASCAL VOC benchmark criterion, using a bounding box intersection over union (IoU) ratio of $\geqslant$0.4 (Carletti et al. 2020; Pandey and Dhavale 2024). This process enables closed loop label updating.

5. Conclusion

The integration of big data predictive analytics with AI driven inspection is reshaping the operation and maintenance (O&M) paradigm for utility scale photovoltaic plants. This review has consolidated recent advances and proposed a unified framework that couples multi source sensing, predictive fault diagnosis, spatiotemporal risk mapping, and closed loop task orchestration.

Three core benefits emerge from this transformation: (i) timeliness—online I V curve monitoring and real time anomaly detection enable early warnings without interrupting generation; (ii) precision—deep learning models (e.g., CNN LSTM, autoencoder LSTM) achieve fault localization accuracy above 97% and F1 scores exceeding 0.92 for gradual degradation, substantially outperforming conventional heuristics; and (iii) sustainability—AI driven dynamic optimization reduces unnecessary manual inspections by up to 38% and boosts energy yield under partial shading

by 8–12%, while extending equipment lifetime through condition based strategies.

Nevertheless, several critical barriers remain unresolved. Dataset diversity is insufficient; most benchmarks are dominated by normal operating conditions, with scarce samples for rare fault types, undermining model generalization. Data quality and annotation inconsistency—arising from varying acquisition conditions, sensor noise, and environmental interference—further degrade classification reliability. Most importantly, the lack of standardized benchmarks and a unified fault labeling system prevents fair cross study comparison and hampers reproducible research.

To accelerate practical deployment, we advocate for (a) constructing open access benchmark datasets that span multiple climate zones, topologies, and fault categories; (b) establishing a consistent fault taxonomy and annotation protocol; and (c) developing physics aware evaluation metrics that reflect both predictive accuracy and operational impact. Addressing these open issues will be pivotal for transitioning PV O&M from reactive repair to truly intelligent, predictive, and sustainable stewardship.

Acknowledgments

The authors gratefully acknowledge the Technology Project for Production Application of Gas Transmission Management Department of PetroChina Southwest Oil & Gas Field Company for their technical support.

Funding

We would like to express our sincere appreciation for the generous financial support provided by Technology Project for Production Application of Gas Transmission Management Department of Petro China Southwest Oil & Gas field Company (No. 2025D1-105-01).

Authors' Contributions

Qing Ma, Jingran He, and Yuan Jing contributed to data collection and industrial expertise. Xiangxue Lv performed the data analysis and participated in manuscript writing. Jia Liao performed the data analysis and drafted the manuscript. Haimin Li conceived and supervised the study, and revised the manuscript. All authors read and approved the final manuscript.

References

[1] Abubakar A, Almeida CFM, Gemignani M (2021a) A review of solar photovoltaic system maintenance strategies. In: 2021 14th IEEE International Conference on Industry Applications (INDUSCON). IEEE, pp 1400–1407. https://doi.org/10.1109/INDUSCON51756.2021.9529669

[2] Abubakar A, Almeida CFM, Gemignani M (2021b) Review of artificial intelligence-based failure detection and diagnosis methods for solar photovoltaic systems. Machines 9(12):328. https://doi.org/10.3390/machines9120328

[3] Aghaei M (2022) Autonomous monitoring and analysis of photovoltaic systems. Energies 15(14):5011. https://doi.org/10.3390/en15145011

[4] Ahan MR, Nambi A, Ganu T, Nahata D, Kalyanaraman S (2021) AI-assisted cell-level fault detection and localization in solar PV electroluminescence images. In: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems (SenSys '21). ACM, pp 485–491. https://doi.org/10.1145/3485730.3493455

[5] Alam MM, Alshahrani T, Khan F, Hakami J, Shinde SM, Azim R (2023) AI-based efficiency analysis technique

for photovoltaic renewable energy system. Phys Scr 98(12):126006. https://doi.org/10.1088/1402-4896/ad0bb4

[6] Ali MH, Dasgupta D (2023) AI-based approach for behind-the-meter solar photovoltaic power forecasting. In: 2023 IEEE Conference on Artificial Intelligence (CAI). IEEE, pp 15–18. https://doi.org/10.1109/CAI54212.2023.00015

[7] Ananthasaravanan N, Megala V (2026) IoT-based hybrid power system with an AI-driven MPPT controller for Z-source boost converters. Iran J Sci Technol Trans Electr Eng 50(1):551–571. https://doi.org/10.1007/s40998-025-00856-9

[8] Ansari SA, Naick BK, Vishwakarma SK (2025) A comparison of AI and conventional MPPT technique for solar photovoltaic system. In: 2025 IEEE 1st International Conference on Smart and Sustainable Developments in Electrical Engineering (SSDEE). IEEE, pp 1–6. https://doi.org/10.1109/SSDEE64538.2025.10967896

[9] Ayturan K, Sarikamis B, Aksahin MF, Kutbay U (2025) SPHERE: benchmarking YOLO vs. CNN on a novel dataset for high-accuracy solar panel defect detection in renewable energy systems. Appl Sci 15(9):4880. https://doi.org/10.3390/app15094880

[10] Bello R-W, Owolawi PA, van Wyk EA, Du C (2024) Photovoltaic module dataset for automated fault detection and analysis in large photovoltaic systems using photovoltaic module fault detection. Data Brief 57:111184. https://doi.org/10.1016/j.dib.2024.111184

[11] Bouchetob E, Nadji B (2025) Deep learning and WBG devices combining to improve PV system efficiency: ANFIS-based MPPT controller. Rev Roum Sci Techn – Électrotechn Énerg 70(4):453–458. https://doi.org/10.59277/RRST-EE.2025.70.4.4

[12] Cardinale-Villalobos L, Jimenez-Delgado E, Garcia-Ramirez Y, Araya-Solano L, Solis-Garcia LA, Mendez-Porras A, Alfaro-Velasco J (2023) IoT system based on artificial intelligence for hot spot detection in photovoltaic modules for a wide range of irradiances. Sensors 23(15):6749. https://doi.org/10.3390/s23156749

[13] Carletti V, Greco A, Saggese A, Vento M (2020) An intelligent flying system for automatic detection of faults in photovoltaic plants. J Ambient Intell Humaniz Comput 11(5):2027–2040. https://doi.org/10.1007/s12652-019-01212-6

[14] Chen Y (2024) Application of big data analysis technology in industrial intelligent inspection system. In: Fourth International Conference on Testing Technology and Automation Engineering (TTAE 2024). Proc. SPIE 13439, art. 134392E. https://doi.org/10.1117/12.3055503

[15] Chorozoglou S, Sarmas E, Marinakis V (2023) An integrated ML-Ops framework for automating AI-based photovoltaic forecasting. In: 2023 IEEE International Conference on Big Data (BigData). IEEE, pp 3921–3928. https://doi.org/10.1109/BigData59044.2023.10386197

[16] De Riso M, Dhimish M, Guerriero P, Daliento S (2024) Design of a portable low-cost I-V curve tracer for on-line and in situ inspection of PV modules. Micromachines 15(7):896. https://doi.org/10.3390/mi15070896

[17] Devanathan B, Suyampulingam A, Selvaraj P, Karuppasamy I, Ilamparithi T (2025) A cloud-integrated virtual framework for LSTM-driven solar forecasting and residential energy management. IEEE Access 13:151055–151071. https://doi.org/10.1109/ACCESS.2025.3601722

[18] Djabali S, Sangroy R, Taheri S, Cretu A-M (2024) An optimized shade-tolerant MPPT for photovoltaic systems using AI-classification. In: 2024 IEEE 12th International Conference on Smart Energy Grid Engineering (SEGE). IEEE, pp 206–210. https://doi.org/10.1109/SEGE62220.2024.10739530

[19] Durán E, Andújar JM, Enrique JM, Pérez-Oria JM (2012) Determination of PV generator I-V/P-V characteristic curves using a DC-DC converter controlled by a virtual instrument. Int J Photoenergy 2012:843185. https://doi.org/10.1155/2012/843185

[20] El Karch H, El Gouri R, Natij Y, Benally M, Mezouari A (2024) AI-based smart real-time PV panels soiling

recognizing system using deep neural network framework on NVIDIA Jetson Nano embedded GPU. Ing Syst Inf 29(5):1687–1699. https://doi.org/10.18280/isi.290503

[21] El Yanboiy N, Khala M, Elabbassi I, Elhajrat N, Eloutassi O, Hassouani YE, Messaoudi C (2025) Artificial intelligence approaches for smart anomaly detection in solar power systems. In: Hagras H, Bennani Y, Nemiche M (eds) Intelligent Systems and Advanced Computing Sciences. Communications in Computer and Information Science, vol 2255. Springer, Cham, pp 205–215. https://doi.org/10.1007/978-3-031-93448-3_17

[22] Felipe TA, Melo FC, Freitas LCG (2021) Design and development of an online smart monitoring and diagnosis system for photovoltaic distributed generation. Energies 14(24):8552. https://doi.org/10.3390/en14248552

[23] Gao F, Wu Z, Han Z, Wang Q, He X, Sun Y, Zhao D, Zhao W, Sun M (2024) Research on distribution inspection planning strategy based on artificial intelligence image recognition. In: Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition. ACM, pp 275–280. https://doi.org/10.1145/3700906.3700951

[24] Ge WQ, Wang XT, Sun YB (2025) Study on photovoltaic power prediction with TSO-LSTM-XGBoost coupled model accounting for weather factors. Int J Green Energy 22(9):1725–1739. https://doi.org/10.1080/15435075.2024.2390159

[25] Ghahremani A, Adams SD, Norton M, Khoo SY, Kouzani AZ (2025) Advancements in AI-driven detection and localisation of solar panel defects. Adv Eng Inform 64:103104. https://doi.org/10.1016/j.aei.2024.103104

[26] Ghenai C (2024) Artificial intelligence based forecasting models for solar PV/hydrogen fuel cell power system: role of short-term AI-based forecasting for the energy transition. In: 2024 International Conference on Sustainable Energy: Energy Transition and Net-Zero Climate Future (ICUE). IEEE, pp 1–7. https://doi.org/10.1109/ICUE63019.2024.10795602

[27] Haddad H, Jerbi F, Smaali I (2024) Performance evaluation of AI-driven photovoltaic output forecasting. In: 2024 IEEE PES/IAS PowerAfrica. IEEE, Johannesburg, South Africa, pp 1–5. https://doi.org/10.1109/PowerAfrica61624.2024.10759493

[28] Hamz A, Ali Z, Dudley S, Saleem K, Uneeb M, Christofides N (2025) A multi-stage review framework for AI-driven predictive maintenance and fault diagnosis in photovoltaic systems. Appl Energy 393:126108. https://doi.org/10.1016/j.apenergy.2025.126108

[29] Hu J, Lim B-H, Tian X, Wang K, Xu D, Zhang F, Zhang Y (2024) A comprehensive review of artificial intelligence applications in the photovoltaic systems. CAAI Artif Intell Res 3:9150031. https://doi.org/10.26599/air.2024.9150031

[30] Huneria HK, Yadav P, Shaw RN, Saravanan D, Ghosh A (2021) AI and IoT-based model for photovoltaic power generation. In: Mekhilef S, Favorskaya M, Pandey RK, Shaw RN (eds) Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2021. Lecture Notes in Electrical Engineering, vol 756. Springer, Singapore, pp 697–706. https://doi.org/10.1007/978-981-16-0749-3_55

[31] Ibrahim A, Farh HMH, Al-Shamma'a AA (2025) A comprehensive review of MPPT strategies for hybrid PV-TEG systems: advances, challenges, and future directions. Mathematics 13(17):2900. https://doi.org/10.3390/math13172900

[32] Ibrahim MS, Gharghory SM, Kamal HA (2026) An advanced transformer model-based LSTM-CNN for power consumption and PV power generation forecasting in modern microgrid. Iran J Sci Technol Trans Electr Eng (online first). https://doi.org/10.1007/s40998-026-01047-w

[33] Jaafar SM, Al Musalhi MS, Al Jazeeri HN, Al Dhabari AA, Rajakannu A (2025) A review on artificial intelligence-based visualization and cleaning of PV panels using uncrewed air vehicle for desert plants in Oman. Preprints.org, preprint. https://doi.org/10.20944/preprints202504.1784.v1

[34] Jandaghian B, Bidabadi BS, Gourabi H (2023) Studies on irradiated photovoltaic cells by proton beam in a plasma focus device. Appl Radiat Isot 193:110663. https://doi.org/10.1016/j.apradiso.2023.110663

[35] Josè DF, Janeiro FM, Pires VF, Pires AJ, Martins JF (2025) Artificial intelligence for fault detection in photovoltaic panels. In: 2025 IEEE 19th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG). IEEE, pp 1–6. https://doi.org/10.1109/CPE-POWERENG63314.2025.11027226

[36] Kumar A, Dubey AK, Segovia Ramírez I, Muñoz del Río A, García Márquez FP (2024a) Artificial intelligence techniques for the photovoltaic system: a systematic review and analysis for evaluation and benchmarking. Arch Comput Methods Eng 31(8):4429–4453. https://doi.org/10.1007/s11831-024-10125-3

[37] Kumar RK, Prabhakaran S, Moorthy SK, Shreeram S, Veeramani R (2024b) AI-enabled robotic maintenance system for solar panel health monitoring. In: 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, pp 2185–2189. https://doi.org/10.1109/ICACCS60874.2024.10716831

[38] Kurukuru VSB, Haque A, Khan MA, Sahoo S, Malik A, Blaabjerg F (2021) A review on artificial intelligence applications for grid-connected solar photovoltaic systems. Energies 14(15):4690. https://doi.org/10.3390/en14154690

[39] Ma W, Wang W, Wu XZ, Hu RN, Tang F, Zhang WG (2019) Control strategy of a hybrid energy storage system to smooth photovoltaic power fluctuations considering photovoltaic output power curtailment. Sustainability 11(5):1324. https://doi.org/10.3390/su11051324

[40] Pandey H, Dhavale S (2024) A survey on artificial intelligence based photovoltaic cell defect detection techniques. In: 2024 IEEE Pune Section International Conference (PuneCon). IEEE, pp 1–6. https://doi.org/10.1109/PuneCon63413.2024.10895155

[41] Quater PB, Grimaccia F, Leva S, Mussetta M, Aghaei M (2014) Light unmanned aerial vehicles (UAVs) for cooperative inspection of PV plants. IEEE J Photovolt 4(4):1107–1113. https://doi.org/10.1109/JPHOTOV.2014.2323714

[42] Sahoo J, Samanta S, Bhattacharyya S (2020) Adaptive PID controller with P&O MPPT algorithm for photovoltaic system. IETE J Res 66(4):442–453. https://doi.org/10.1080/03772063.2018.1497552

[43] Tang W, Yang Q, Hu X, Yan W (2023) Edge intelligence for smart EL images defects detection of PV plants in the IoT-based inspection system. IEEE Internet Things J 10(4):3047–3056. https://doi.org/10.1109/JIOT.2022.3150298

[44] Wang F, Chen SS, Sun WQ, Zhang RY, Zhang LH, Lv SC, Ren GC (2025) Research on photovoltaic construction management based on UAV AI recognition technology. In: Proceedings of the 6th International Conference on Advances in Civil and Ecological Engineering Research. Lecture Notes in Civil Engineering, vol 575. Springer, Singapore, pp 91–103. https://doi.org/10.1007/978-981-96-1627-5_8

[45] Wasser HP, Wenge C, Bagheri F, Suciu G, Beceanu C (2025) Review of AI methods for fault diagnosis and predictive maintenance in solar panels. In: 2025 IEEE 19th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG). IEEE, Antalya, Türkiye, pp 1–7. https://doi.org/10.1109/CPE-POWERENG63314.2025.11027253

[46] Xue HL, Ma JW, Zhang JL, Jin PH, Wu J, Du F (2024) Power forecasting for photovoltaic microgrid based on multiscale CNN-LSTM network models. Energies 17(16):3877. https://doi.org/10.3390/en17163877

[47] Yao SY, Kang Q, Zhou MC, Abusorrah A, Al-Turki Y (2021) Intelligent and data-driven fault detection of photovoltaic plants. Processes 9(10):1711. https://doi.org/10.3390/pr9101711

[48] Yap KY, Sarimuthu CR, Lim JMY (2020) Artificial intelligence based MPPT techniques for solar power system: a review. J Mod Power Syst Clean Energy 8(6):1043–1059. https://doi.org/10.35833/MPCE.2020.000159

[49] Zhang HR, Yu XH, Gao ZX (2025a) Is the end of AI in photovoltaic power? Evidence from China. Energy Econ 145:108423. https://doi.org/10.1016/j.eneco.2025.108423

[50] Zhang L, Chen ZY, Zhu JZ, Lin KX, Huang LY (2025b) Online incremental probability power prediction for distributed PVs in heterogeneous and dynamic data environments. Appl Energy 394:126110. https://doi.org/10.1016/j.apenergy.2025.126110

[51] Zhang ZX, Ma MY, Ma WT, Zhang R, Wang J (2022) A data-driven photovoltaic string current mismatch fault diagnosis method based on I-V curve. Microelectron Reliab 138:114705. https://doi.org/10.1016/j.microrel.2022.114705

[52] Zhu CY, Tu YB, Wei QY, Zang Y, Zhou P, Yang LX, Wang J, Yu YJ, Lv RR, Du JX, Yan WS (2025) Short-term power prediction of photovoltaic power station based on LSTM-XGBoost model. Sol Energy 300:113819. https://doi.org/10.1016/j.solener.2025.113819