

A Review of Machine Learning Models and Evaluation Techniques for Single String Anomaly Detection in Malaysian Large-Scale Photovoltaic Systems

Muhammad Ashraf Bin Ghazali¹, Mohd Hezri Fazalul Rahiman¹, Zuhaila Mat Yasin^{2*}

¹Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia

²Solar Research Institute (SRI), Universiti Teknologi MARA, 40450 Shah Alam, Malaysia

Corresponding author's email: zuhai730@uitm.edu.my

Article info:

Received: 30 December 2025

Revised: 27 March 2026

Accepted: 5 May 2026

DOI:

[10.69650/rast.2026.265480](https://doi.org/10.69650/rast.2026.265480)

Keywords:

Anomaly
Fault Detection
Machine Learning
Photovoltaic
Predictive Maintenance

ABSTRACT

The rapid expansion of solar photovoltaic (PV) systems has increased the need for reliable and intelligent fault detection techniques to maintain system performance and operational efficiency. This paper presents a comprehensive review of machine learning (ML) models and evaluation techniques for single-string anomaly detection in large-scale PV systems, with particular emphasis on applications in Malaysia's tropical environment. The review examines the transition from conventional monitoring approaches, such as I–V curve tracing and threshold-based diagnostics, to advanced data-driven methods. Various ML and deep learning models, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and hybrid frameworks such as CNN–SVM and AE–GRU, are analysed in terms of their anomaly detection capabilities and performance. Common evaluation metrics, including RMSE, MAE, MAPE, R^2 , and F1-score, are also reviewed to assess model effectiveness and reliability. In addition, recent Malaysian studies employing K-Means–LSTM clustering, Random Forest-based fault diagnosis, Vision Transformers (ViT) and Vertices Principal Component Analysis (VPCA) are discussed, highlighting their potential for PV anomaly detection under tropical operating conditions. The review identifies key challenges, particularly the dependence on labelled datasets, which limits scalability and early fault detection. Emerging opportunities in unsupervised and semi-supervised learning approaches, including autoencoders, Isolation Forests, and clustering-based reconstruction methods, are also explored as promising solutions for real-time and adaptive anomaly detection. The findings provide insights into current research trends, existing limitations, and future directions for enhancing the reliability and maintenance of large-scale PV systems in Malaysia.

1. Introduction

To ensure the accuracy of a solar energy system, thorough analysis must be conducted for each individual panel, encompassing voltage, current, temperature, and irradiance, to optimize its efficiency. However, the challenge with PV systems lies in the cost-effectiveness of monitoring every single solar panel connected in the system, given the large number of interconnected panels in an array. Thus, a new method of analysis is required to monitor the entire PV System. Fault in PV system arrays often cause maintenance costs to rise and efficiency to drop. Because solar panels are usually set up in remote areas, fixing these issues can be expensive and time-consuming. Researchers have also discovered that implementing appropriate fault detection solutions can significantly enhance the reliability of PV systems and substantially decrease operational and maintenance expenses [1].

Renewable energy adoption has accelerated globally in the past decade, with solar PV emerging as the dominant technology. In 2022 alone, solar PV generation grew by 270 TWh, representing a 26% annual increase and bringing total generation to approximately 1,300 TWh. This expansion enabled solar to surpass wind power as the largest

renewable electricity capacity worldwide, largely driven by rapid deployment in China, the European Union, and the United States. The growth of solar reflects declining module costs, rising efficiency, and strong policy incentives, making it the most scalable renewable source for meeting future energy demand. In Southeast Asia, countries such as Vietnam and Thailand have recorded significant solar adoption, while the Philippines and Indonesia are beginning to scale up utility-scale installations [2].

Malaysia's Renewable Energy Roadmap, under the National Energy Transition Plan, sets ambitious targets of achieving 31% renewable energy in the generation mix by 2025 and 40% by 2035. Solar PV is recognised as the central pillar of this transition, supported by favourable solar radiation levels of 4,400–5,000 Wh/m² per day. Long-term projections estimate solar capacity could reach 83 GW by 2050 in the Transformed Energy Scenario and up to 153 GW under the 1.5°C pathway. However, the variability and intermittency of solar output present major challenges for reliable grid integration. To meet these goals, complementary technologies such as Battery

Energy Storage Systems (BESS) and Demand Response (DR) are increasingly emphasised. At the same time, advanced fault detection and monitoring systems are critical to ensure that large-scale PV installations operate efficiently and deliver the performance required to achieve Malaysia's renewable energy roadmap [3].

To date, numerous studies have been conducted on fault detection using artificial intelligence (AI) methods to enhance system reliability and reducing the maintenance costs. Among these methods, ANN emerges as one of the most extensively employed techniques in fault detection and classification processes [4]. However, a significant portion of the research has primarily focused on classifying types of faults in solar systems, rather than prioritizing fault detection itself. Early studies, such as [5] in 2015, explored fault classification in solar PV systems using semi-supervised learning. While these approaches demonstrated potential, more recent works [6] highlight the limitations of classification models in achieving fast and scalable fault detection, especially under real-time constraints.

Despite substantial progress in AI-based solar PV monitoring, most research has concentrated on fault classification using supervised learning, which requires large volumes of labelled data that are often unavailable in real-world utility-scale plants. Furthermore, many studies validate their models using simulated or small-scale datasets, limiting their applicability under diverse operational conditions. In Malaysia, the challenge is compounded by the tropical climate, characterised by high humidity, haze, and intermittent cloud cover, which increases variability and complicates model generalisation. These gaps underscore the urgent need for robust, unsupervised, and real-time fault detection approaches that can operate effectively without reliance on labelled data, while being scalable to large PV systems. Addressing this research gap is critical to ensuring the reliability of Malaysia's rapidly expanding solar infrastructure and supporting the national target of achieving 31% renewable energy in the generation mix by 2025.

In particular, this study focuses on single-string anomaly detection, which provides a more cost-effective and scalable alternative compared to panel-level monitoring while offering higher resolution fault identification than system-level approaches.

1.1 Literature Review Methodology

This review was conducted through a structured literature search to identify relevant studies on machine learning-based PV fault detection. Major academic databases including IEEE Xplore, and Scopus were used to collect publications related to PV system monitoring and anomaly detection.

The search was performed using keywords such as "photovoltaic fault detection", "solar PV anomaly detection", "machine learning in photovoltaic systems", and "solar PV fault diagnosis". Only studies related to photovoltaic systems employing machine learning techniques for fault detection were considered.

The screening process involved reviewing titles, abstracts, and full texts to ensure relevance to the scope of this review. Studies focusing on non-photovoltaic systems, unrelated monitoring techniques, or non-machine learning approaches were excluded.

The selected publications were further analysed based on their fault detection methodology, dataset characteristics, validation techniques, and reported performance metrics. This approach ensures that the review provides a comprehensive overview of current machine learning techniques applied to photovoltaic fault detection.

2. Types of electrical faults

PV systems are prone to several types of electrical faults that can significantly affect performance, reliability, and safety. Among the most common are open-circuit, short-circuit, ground, and arc faults, each of which produces distinct signatures on the current–voltage (I-V) and power–voltage (P-V) characteristics. Understanding these differences is crucial for accurate fault diagnosis and the development of effective detection methods.

2.1 Open Circuit (OC) fault

An open-circuit (OC) fault in a PV system happens when the electrical continuity of a module or string is broken, preventing normal current flow. Unlike short-circuit faults, which immediately trigger protection, OC faults are more subtle and may allow the system to continue operating while silently degrading performance. In terms of the current–voltage (I-V) characteristic, the most noticeable effect is a significant reduction in current output. The short-circuit current (I_{sc}) drops sharply because the path for current is interrupted, while the open-circuit voltage (V_{oc}) remains almost unchanged since it is determined mainly by the semiconductor properties of the PV cells. This creates a distortion in the I-V curve: instead of the smooth curve expected under healthy operation, the curve flattens and bends downward, especially near the maximum power point region, showing a clear loss in usable current. Consequently, the fill factor (FF) decreases, and the maximum power point (P_{mpp}) shifts, leading to reduced energy harvest from the PV system. If multiple strings or modules are affected, the overall array curve may exhibit abrupt deviations or distortions, further highlighting the imbalance. These distortions in the I-V curve make OC faults detectable, but only if proper monitoring techniques are applied, since the V_{oc} alone does not provide a clear indication of the problem [7].

2.2 Short Circuit (SC) fault

A short-circuit (SC) fault in a PV system occurs when one or more cells are bypassed or directly connected, causing their voltage contribution to drop to nearly zero. This fault reduces the effective voltage of the string or array, leading to significant mismatch losses between the actual output and the theoretical healthy output. On the I-V curve, a short-circuit fault is characterized by a sharp decrease in open-circuit voltage (V_{oc}), while the short-circuit current (I_{sc}) remains almost unchanged because it is mainly governed by irradiation and temperature. This results in the curve shifting horizontally towards the origin, reflecting voltage collapse but stable current at short-circuit. In contrast, an open-circuit (OC) fault shows the opposite pattern: the current (I_{sc}) drops sharply due to the interruption of the current path, while V_{oc} stays nearly the same, producing a flattened I-V curve with suppressed current flow. Both faults lower the maximum power point (P_{mpp}), but SC faults reduce power through loss of voltage, whereas OC faults reduce power through loss of current. On the P-V curve, this distinction is clear: SC faults cause the power peak to shift leftward due to voltage reduction, while OC faults cause the peak to drop downward due to current reduction [8].

2.3 Ground fault

A ground fault in a PV system occurs when one of the conductors in a string unintentionally comes into contact with the ground, creating a leakage path for current. This condition not only reduces system efficiency but also introduces serious safety risks such as electric shock and fire. On the I-V curve, a ground fault resembles a short-circuit fault: the open-circuit voltage (V_{oc}) drops significantly because part of the array is bypassed to ground, while the short-

circuit current (I_{sc}) remains nearly unchanged since it is still dictated by irradiation. As a result, the curve shifts leftward, and the maximum power point (P_{mpp}) decreases due to reduced operating voltage. In comparison, an open-circuit fault primarily causes a sharp drop in I_{sc} while V_{oc} remains almost the same, flattening the I-V curve along the current axis, whereas a short-circuit fault causes V_{oc} to collapse with I_{sc} nearly constant. Thus, while both SC and ground faults show voltage reduction and OC faults show current reduction, ground faults are more critical because they not only degrade performance but also allow continuous leakage to earth, posing higher reliability and safety concerns [9].

2.4 Arc fault

An arc fault in a PV system occurs when electrical contacts are loose, corroded, or damaged, creating a high-resistance gap where current continues to flow through ionized air. This produces unstable voltage and current fluctuations, intense heat, and electromagnetic emissions, which can ignite fires if undetected. In the I-V characteristics, arc faults are different from conventional open-circuit (OC), short-circuit (SC), or ground faults. Instead of a stable shift in V_{oc} or I_{sc} , arc faults produce chaotic, fluctuating voltage behaviour with irregular drops and recoveries, while the current shows transient spikes before stabilizing. By contrast, an OC fault causes a sharp reduction in I_{sc} with V_{oc} nearly unchanged, flattening the curve toward the current axis. An SC fault causes V_{oc} to collapse with I_{sc} nearly constant, shifting the curve leftward. A ground fault resembles SC behaviour on the I-V curve (V_{oc} drop, I_{sc} stable) but adds continuous leakage to earth, raising safety risks. Arc faults are therefore the most dangerous among these, since they not only reduce power output but also release significant thermal energy that can melt connectors and cause fires, making them a critical focus for PV safety and requiring advanced detection methods [10].

3. Types of environmental faults

Apart from electrical faults, PV systems are also vulnerable to environmental faults that arise from external factors such as shading, dust accumulation, temperature stress, or high system voltage interacting with ambient conditions. These faults are equally critical as they directly affect module performance, reduce overall energy yield, and in some cases pose safety risks. The most common environmental faults include partial shading, soiling, hotspot formation, and potential induced degradation (PID), each with distinct mechanisms and impacts on PV reliability and efficiency.

3.1 Partial shading fault

Partial shading in PV systems occurs when parts of the array receive nonuniform irradiance due to factors such as passing clouds, nearby structures, accumulated dust, dirt, or bird droppings. This condition reduces the current output of the shaded modules compared to unshaded ones, which in turn causes power mismatch losses across the array. The main effects include a significant reduction in output power, the formation of hotspots on shaded cells or modules, and accelerated degradation of PV panels when shading persists. Partial shading can be classified as static, where the shadow remains fixed on the PV surface for a long period (e.g., dirt or debris), or dynamic, where the shadow changes over time (e.g., clouds or smoke). Although often temporary, partial shading is considered an intermittent fault, and it can lead to an estimated 10–20% annual reduction in power generation if not addressed. Moreover, it complicates maximum power point tracking (MPPT) by introducing multiple peaks in the power–voltage curve, making it harder for the system to operate at the true maximum power point. Overall, partial shading

lowers system efficiency, reduces reliability, and increases the risk of long-term damage to PV modules [11].

3.2 Soiling fault

Soiling in PV systems occurs when dust, dirt, or bird droppings accumulate on the panel surface, gradually blocking sunlight and reducing current generation even though external irradiance readings may appear unchanged. This condition develops slowly over time, leading to uniform power loss across the array, lower efficiency, and potential long-term degradation if left unaddressed. In contrast, partial shading is caused by external obstructions such as clouds, trees, or nearby structures, creating nonuniform irradiance across modules that results in hotspots, mismatch losses, and multiple peaks in the P–V curve. While shading often occurs dynamically and produces abrupt fluctuations in system output, soiling tends to be static and progressive, making it harder to detect but equally detrimental in reducing system performance and reliability [12].

3.3 Hotspot fault

A hotspot fault in a PV system happens when a solar cell or part of a module becomes significantly hotter than its surrounding areas due to uneven operating conditions. This usually occurs when certain cells are partially shaded, contaminated with dirt, cracked, or have internal defects that restrict current flow. Under these conditions, the affected cell is forced to operate in reverse bias, dissipating power as heat instead of generating electricity. The excessive localized heating reduces the overall power output of the module, accelerates material degradation, and can create hotspots that permanently damage cells or, in severe cases, pose a fire hazard. Over time, hotspot faults decrease system efficiency, shorten module lifespan, and compromise both the reliability and safety of PV installations [13].

3.4 Potential Induced Degradation fault

Potential Induced Degradation (PID) is a progressive fault in PV systems that arises when a high voltage difference between solar cells and the grounded frame causes leakage currents through the encapsulant, leading to electrochemical stress that gradually reduces performance. This degradation is accelerated by factors such as high system voltage, humidity, and elevated operating temperatures, and it manifests as a uniform reduction in open-circuit voltage (V_{oc}), fill factor (FF), and maximum power point (P_{mpp}). Because PID develops silently across multiple cells, strings, or even the entire array, it is harder to detect without continuous monitoring, yet it can cause significant long-term energy yield losses and shorten module lifespan. In contrast, hotspot faults are localized issues that occur when cells are partially shaded, contaminated, cracked, or defective, forcing them into reverse bias and causing concentrated heating on specific regions of the module. The affected cells dissipate power as heat instead of generating electricity, resulting in localized high temperatures that can be detected with infrared thermography. While PID uniformly reduces electrical output through widespread electrochemical degradation, hotspots lead to immediate localized overheating, which not only lowers efficiency but can also accelerate material damage and, in severe cases, present safety risks such as fire hazards [14].

Table 1 summarizes the major electrical and environmental faults that commonly occur in PV systems and their corresponding impacts on system performance, reliability, and safety. Electrical faults such as open-circuit (OC), short-circuit (SC), ground, and arc faults directly affect current and voltage behavior, leading to noticeable distortions in the I–V and P–V characteristics. OC faults primarily cause

a sharp reduction in short-circuit current (I_{sc}) while leaving the open-circuit voltage (V_{oc}) nearly constant, whereas SC and ground faults result in a significant drop in V_{oc} with minimal change in I_{sc} . Arc faults, on the other hand, generate unstable voltage and current fluctuations

accompanied by transient spikes, posing serious fire risks. In contrast, environmental faults such as partial shading, soiling, hotspot formation, and potential induced degradation (PID) arise from external and operational factors that gradually degrade performance.

Table 1 Types of Faults in PV systems.

Fault Type	Description / Cause	Effect on I-V / P-V Characteristics	Impact on Performance / Safety
Open-Circuit (OC) Fault	Occurs when electrical continuity in a module or string is broken, interrupting current flow.	Sharp drop in short-circuit current (I_{sc}); V_{oc} remains nearly unchanged; I-V curve flattens and bends downward; reduced fill factor (FF) and shifted P_{mpp} .	Reduces current and power output; may go undetected without proper monitoring; leads to hidden performance degradation.
Short-Circuit (SC) Fault	Caused by bypassed or directly connected cells resulting in near-zero voltage contribution.	Significant decrease in V_{oc} while I_{sc} remains stable; I-V curve shifts leftward; P-V curve peak shifts left due to voltage loss.	Causes voltage collapse, mismatch losses, and power reduction; can trigger protection devices.
Ground Fault	Occurs when a live conductor unintentionally contacts ground, forming a leakage path.	Similar to SC fault with V_{oc} drop and stable I_{sc} ; leftward shift in I-V curve; decreased P_{mpp} .	Leads to energy loss, electric shock risk, and potential fire hazard due to continuous current leakage.
Arc Fault	Caused by loose, corroded, or damaged connections forming high-resistance arcs through ionized air.	Produces unstable and fluctuating voltage and current with transient spikes; irregular I-V curve behavior.	Generates heat and electromagnetic emissions; high fire risk; requires advanced, high-speed detection.
Partial Shading Fault	Caused by uneven irradiance due to clouds, buildings, dirt, or bird droppings on module surfaces.	Multiple peaks appear in the P-V curve; reduced current from shaded cells; mismatch between strings.	Reduces output power (10–20% annually); increases risk of hotspots and long-term degradation.
Soiling Fault	Gradual dust or dirt accumulation blocking sunlight, reducing irradiance on panel surface.	Uniform current reduction across all modules; smooth but lower I-V curve; decreased P_{mpp} .	Causes continuous power loss and efficiency drop; harder to detect since irradiance readings remain constant.
Hotspot Fault	Local overheating of shaded or defective cells operating in reverse bias.	Localized reverse current flow; unaffected cells maintain output but affected cells dissipate heat.	Leads to permanent cell damage, accelerated degradation, and fire hazards.
Potential Induced Degradation (PID)	Caused by high voltage difference between PV cells and grounded frame, leading to leakage currents.	Uniform reduction in V_{oc} , FF, and P_{mpp} across modules; consistent I-V curve drop.	Long-term power degradation across multiple modules; difficult to detect without continuous monitoring.

Partial shading introduces multiple peaks in the P-V curve and mismatch losses, while soiling causes uniform power reduction across modules. Hotspots lead to localized heating and potential cell damage, whereas PID causes widespread degradation and long-term energy loss. Overall, these faults alter the electrical signatures of PV modules, reducing output efficiency and accelerating system degradation, underscoring the need for continuous monitoring and intelligent fault detection mechanisms.

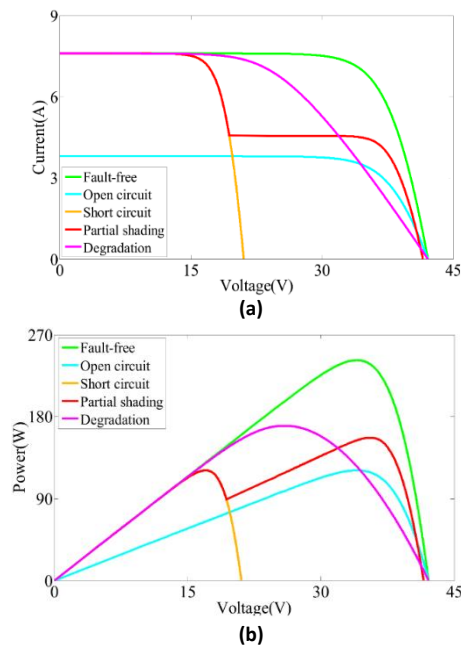


Fig. 1 (a) I-V Curve (b) P-V Curve [15].

The I-V characteristics under various operating and fault conditions, as illustrated in (a), show clear deviations from fault-free behaviour, where electrical and environmental faults lead to current reduction, knee shifting, and boundary effects such as voltage or current collapse compared to normal operation.

Correspondingly, the P-V characteristics shown in (b) highlight the impact of these faults on power generation, where faults result in reduced maximum power output, shifted peak power points, and distorted power curves relative to the fault-free condition.

4. Fault Detection Method

In this section, the reviewed fault detection methods are analysed based on their applicability to different monitoring levels, including system-level, array-level, and single-string level detection, with particular emphasis on methods suitable for single-string anomaly detection.

4.1 Supervised Learning–Based Fault Detection Methods

Rao et al. [16] present a detailed exploration of fault detection in PV arrays using customized neural network algorithms. This study emphasizes the necessity of accurate fault detection to maintain the efficiency and reliability of PV systems, which are susceptible to faults such as ground faults, arc faults, open circuits, and partial shading. The authors propose a feedforward neural network model designed to detect and classify eight different types of faults by analyzing critical parameters like voltage, current, and temperature. The neural network model was trained and validated using extensive data from a simulated PV array, demonstrating high accuracy and efficiency in fault detection and classification. The research highlights the effectiveness of neural networks in real-time

fault management, significantly reducing the mean time to repair (MTTR) and enhancing the operational stability of utility-scale PV systems. This approach underscores the potential of advanced machine learning techniques in optimizing the performance and reliability of solar energy installations.

Rao *et al.* [17] present a sophisticated approach to fault detection and classification in PV arrays using advanced neural network techniques, specifically focusing on dropout and pruned neural networks. The study highlights the critical need for accurate fault detection to maintain the efficiency and reliability of PV systems, which can suffer from various faults such as shading, soiling, degraded modules, and short circuits. The authors propose using dropout neural networks to prevent overfitting and improve generalization, alongside pruning strategies to optimize model complexity without significantly compromising accuracy. By training the neural networks on a comprehensive dataset with custom input features, the method demonstrates high accuracy in classifying different fault types. The results show that concrete dropout neural networks outperformed traditional machine learning classifiers and achieved superior fault classification accuracy. This research underscores the potential of employing neural network techniques for real-time, precise fault detection and classification in PV arrays, enhancing operational efficiency and reducing maintenance costs.

Gupta Bakshi *et al.* [18] demonstrated the application of ANN in diagnosing faults in medium-scale PV systems, particularly in identifying string, string-to-ground, and string-to-string faults. In their study, an ANN model was implemented to diagnose faults in a 250 kW solar PV array, achieving 100% accuracy for both training and validation, and 98% for testing. This approach, using electrical parameters optimized by random search, showed superior performance, highlighting the potential for smart fault-detection in real-world PV plants.

Devi *et al.* [19] present an in-depth analysis of fault detection and classification in islanded microgrids using LSTM networks. The study focuses on a microgrid comprising solar PV, fuel cells, and battery storage systems. By leveraging voltage and current data, the LSTM model was trained to identify various fault types, demonstrating high accuracy and robustness. The model's efficacy was validated using real-time simulations on OPAL-RT platforms, ensuring practical applicability. Compared to traditional methods, the LSTM approach offers superior adaptability to dynamic system conditions, highlighting its potential for enhancing the reliability and safety of decentralized energy systems. This research underscores the importance of advanced machine learning techniques in managing and protecting modern power systems.

Dhoundiyal *et al.* [20] study presents a neural network-based approach for fault detection and classification in large-scale solar photovoltaic arrays. The research emphasizes the critical need for efficient fault management in PV systems due to the increasing scale of installations and the associated risks of faults, which can lead to significant performance degradation, accelerated system depreciation, and fire hazards. The proposed method leverages ANN to diagnose and classify faults such as open circuit, short circuit, and partial shading. Using MATLAB Simulink for simulation, the study demonstrates the ANN model's high accuracy and low computation time in identifying fault types and their locations. This approach, which includes extensive data training and validation, highlights the potential for integrating ANN-based systems to enhance the reliability, safety, and efficiency of large-scale PV arrays.

Pa *et al.* [21] present a CNN-based fault detection scheme for PV solar panels, highlighting its high accuracy in identifying faults. The study addresses the challenges of traditional fault detection methods and emphasizes the importance of improving the efficiency and reliability of PV systems. The proposed CNN model classifies PV module images into faulty and normal categories, achieving a success rate of 91.1% for binary classification and 88.6% for multi-classification (shadowy, cracked, or dusty cells). This model outperforms previous CNN models by a significant margin. The CNN-based approach leverages image data to enhance fault detection accuracy, offering a straightforward and effective solution for real-time fault management in large-scale PV systems. The study underscores the potential of CNNs in advancing fault detection methodologies, thereby contributing to the operational stability and maintenance efficiency of solar energy systems.

Jayasinghe *et al.* [22] study presents an advanced fault detection, classification, and localization system for microgrids using ANN combined with discrete wavelet transform (DWT) and wavelet energy entropy. The research addresses the limitations of conventional fault detection methods, particularly in microgrids with diverse operational scenarios and lower fault currents from inverter-interfaced distributed generators (IIDGs). The proposed system uses DWT to extract key features from current signals, which are then processed by ANNs for precise fault classification and localization. MATLAB/Simulink simulations demonstrate the high accuracy of the model in identifying various fault types and their locations. This approach enhances the reliability and efficiency of microgrids by ensuring rapid fault isolation and power restoration, offering significant improvements over traditional protection schemes. The study underscores the potential of integrating ANN-based fault management systems in large-scale PV arrays and microgrids to ensure stable and efficient operations.

Dhimish [23] presented an advanced ANN model designed for detecting bypass diode faults in PV panels. This research emphasizes the importance of accurate fault detection to enhance the reliability and efficiency of PV systems. The proposed ANN model utilizes only three inputs/ output power, short-circuit current, and open-circuit voltage to detect both short and open bypass diode faults with high precision. Experimental validation demonstrated that the model achieves an average detection accuracy of 93.4%, with precision and sensitivity rates of 96.4% and 92.6%, respectively. This ANN-based approach not only identifies the presence of faults but also classifies the fault types, significantly improving fault management in PV systems. The study underscores the potential of integrating ANN models for real-time fault detection, thereby enhancing the operational stability and safety.

Bendale *et al.* [24] present a deep learning-based approach for fault detection and maintenance in solar photovoltaic PV systems using CNN. The study addresses the challenge of identifying faults, which are crucial for maintaining the efficiency and reliability of PV systems. The proposed method analyzes electrical parameters such as voltage, current, and power output to accurately detect and classify faults. The CNN model is trained on historical data and demonstrates high accuracy in identifying various fault types, achieving a detection accuracy of 97.5%. This approach improves fault detection enhancing real-time monitoring and maintenance. The study highlights the potential of integrating advanced neural network techniques to improve the operational efficiency and reliability of solar PV systems, thereby contributing significantly to the advancement of smart renewable energy technologies.

Shubham Kumar et al. [25] presents a fault detection and location estimation method for photovoltaic PV arrays using an ensemble learning model. The study addresses the limitations of conventional protection schemes such as overcurrent protection devices and ground fault interrupters, which often fail to detect low-mismatch and high-impedance faults. The proposed ensemble learning technique combines multiple decision tree classifiers to improve accuracy and reliability in identifying faults. Using string current and array voltage data simulated through MATLAB/Simulink for a 3 MW PV system, the model was trained and tested across various fault types including line-to-ground (L-G), line-to-line (L-L), and open-circuit faults, as well as partial shading conditions. The results demonstrated that the model successfully detected both minor and severe faults while accurately distinguishing healthy strings, achieving up to 100% classification accuracy in most cases. Additionally, the method effectively located the faulted string without misclassifying healthy ones. This research highlights the potential of ensemble learning approaches in overcoming the blind spots of traditional protection systems, offering a fast, accurate, and data-driven solution for real-time PV fault detection and localization, thereby enhancing system safety, reliability, and operational performance.

Eya Gassab et al. [26] presents a comprehensive study on fault detection and classification in PV systems using a multi-class Support Vector Machine (SVM) algorithm with a Radial Basis Function (RBF) kernel. The research focuses on improving early and accurate fault diagnosis to enhance PV system reliability and performance. A detailed model of a grid-connected PV system with a storage element was developed in MATLAB/Simulink to simulate both healthy and faulty conditions, including partial shading, open-circuit, short-circuit, and degradation faults. The proposed SVM model, trained on features extracted from I-V and P-V characteristics, effectively identifies and classifies these faults with high precision. Experimental results show that the SVM achieved up to 99.85% accuracy in detecting partial shading faults and maintained an overall precision exceeding 90% across all fault types. The study also compared the One-Versus-One (OVO) and One-Versus-All (OVA) classification schemes, demonstrating consistent accuracy in both. This work highlights the effectiveness of SVM-based approaches in handling non-linear data and complex fault patterns, confirming its strong capability for reliable and efficient fault detection and classification in PV systems.

S. Ragul et al. [27] proposes a fault detection and classification method for PV systems using a Logistic Regression (LR) classifier. The study addresses the increasing demand for reliable fault diagnosis in PV systems that are prone to open-circuit and short-circuit faults due to environmental and operational stresses. A pre-tuned dataset generated through MATLAB simulations under various fault conditions was used to train and test the model, eliminating the need for post-simulation hyperparameter tuning. The LR classifier was trained on 80% of the dataset and validated on the remaining 20%, achieving a testing accuracy of 88.7%. The model efficiently distinguishes between healthy and faulty states, offering a simple yet effective approach for real-time PV fault detection. The paper also highlights the comparative advantages of logistic regression, including low model complexity, reduced risk of overfitting, and high interpretability compared to neural networks. This research demonstrates that logistic regression can serve as a lightweight yet reliable solution for identifying PV faults, minimizing system downtime, and supporting predictive maintenance strategies. Overall, the proposed approach enhances PV system resilience by ensuring early detection and improving the long-term stability and efficiency of solar power generation.

J. Jayapriya and R. Kalaivani [28] presents a hybrid deep learning model combining CNN with Support Vector Machines (SVM) for accurate detection of cracks and surface faults in PV solar panels. The proposed model integrates the VGG16 CNN architecture for feature extraction with the SVM classifier for fault categorization, leveraging the strengths of both models to improve detection accuracy. The system processes solar panel images and classifies them into normal and defective categories, followed by sub-classification into partial shading, cracks, dusty surfaces, and bird droppings. Experimental results demonstrated an overall classification accuracy of 88.74%, outperforming standalone CNN approaches. For the cracked versus partial shading condition, the model achieved 91.3% accuracy, while classification between dusty and bird-dropping conditions attained 86.4% accuracy. The results confirm that integrating SVM enhances the CNN's classification capabilities by improving generalization and reducing misclassification errors. This hybrid CNN-SVM framework provides a reliable and efficient solution for real-time solar panel monitoring, enabling early fault detection and minimizing maintenance costs, thereby extending the operational lifespan and efficiency of photovoltaic systems.

Suganya Govindarajan et al. [29] introduces an advanced fault detection and classification system for solar PV arrays using a Recurrent Neural Network (RNN) integrated with Internet of Things (IoT) technology. The study aims to address the limitations of traditional fault detection techniques that rely on multiple sensors and high computational costs by proposing a cost-effective and sensorless approach. The system employs an IoT-based web platform for real-time monitoring of individual PV panels, coupled with an RNN transformer model for fault classification. The RNN model efficiently captures temporal dependencies in voltage and current signals, enabling accurate detection of open-circuit, short-circuit, and line-to-ground faults. Experimental validation using an ESP32 microcontroller and prototype PV setup achieved a fault classification accuracy of 99.67%, outperforming conventional ANN and Feedforward Neural Network (FFNN) models, which recorded accuracies of 94.3% and 93.2%, respectively. The integration of RNN with IoT significantly enhances the precision, speed, and reliability of PV fault diagnosis while minimizing hardware requirements. This research highlights the potential of intelligent IoT-driven deep learning models to improve predictive maintenance, ensure grid stability, and optimize the operational efficiency of modern solar PV systems.

Rong-Jong Wai and Xuerong Cai [30] presents a DC arc-fault detection strategy for PV systems that integrates Empirical Mode Decomposition (EMD) with a Gated Recurrent Unit Neural Network (GRU-NN). The proposed hybrid approach addresses the limitations of conventional time- and frequency-domain methods that rely on fixed thresholds and exhibit poor adaptability across different PV stations. In this method, EMD is employed to decompose the PV current signal into multiple intrinsic mode functions, from which six statistical features—Shannon entropy, energy, variance, crest factor, kurtosis, and number of extreme points—are extracted to characterize arc signatures. These features are then arranged sequentially and input into a GRU-NN, which captures temporal dependencies and effectively mitigates modal confusion inherent in EMD-based methods. Furthermore, an online updating mechanism is introduced to continuously fine-tune the model, maintaining accuracy under varying conditions such as system aging and environmental changes. Experimental validation conducted in two PV stations in Taiwan and China achieved an average detection accuracy exceeding 98.7%, with a fast execution time of 8 milliseconds and strong resistance to interference from inverter startup, shadow occlusion, and airflow

disturbances. This study demonstrates the potential of combining signal decomposition with recurrent learning for high-speed, adaptive, and reliable DC arc-fault detection in modern PV systems.

These supervised approaches demonstrate high fault classification accuracy but rely heavily on labelled datasets and predefined fault categories. However, most supervised approaches are applied at system-level or array-level monitoring, as they rely on labelled datasets that are typically collected at aggregated levels rather than individual string resolution.

4.2 Semi-Supervised Fault Detection Methods

The paper by Tsai *et al.* [31] discusses an anomaly detection mechanism for solar generation using a semi-supervised learning model. The study employs clustering algorithms to filter normal events and an autoencoder neural network for classification. The proposed model was trained and validated using data from a 500kW solar power plant. The results showed varying performance across different clusters, with recall, precision, and F1 scores ranging from 0.5 to 0.8, 0.67 to 1.0, and 0.57 to 0.8, respectively. This semi-supervised approach efficiently identifies abnormal events, reducing the need for extensive labeled data and enhancing the reliability of solar panel maintenance and management.

Semi-supervised fault detection methods aim to reduce the dependency on extensive labelled fault data by learning normal operating behaviour prior to anomaly identification. These approaches are particularly useful when only limited fault labels are available, although their effectiveness depends on the quality of clustering and assumptions regarding normal system behaviour. Semi-supervised approaches offer moderate suitability for single-string detection, although their effectiveness depends on the availability of partially labelled or clustered data.

4.3 Unsupervised and Anomaly Detection-Based Method

The paper by Seyed Mahdi Miraftebadeh *et al.* [32] proposes an anomaly detection method for PV systems using a deep learning autoencoder. The model uses only measured PV power production data, eliminating the need for additional equipment data collection. The autoencoder demonstrated a high F1-score accuracy of 82%, significantly outperforming other machine learning models such as random forest, K-nearest neighbors, and support vector classification. The model effectively identifies anomalies by reconstructing input data and comparing it against a defined error threshold. This approach proves to be reliable and efficient for fault detection in PV systems, showcasing its potential for improving the reliability and performance of solar power generation.

Yadav and Tummuru [33] propose a signature-based short-circuit fault detection and isolation scheme for low-voltage DC microgrids (LV-DCMG) using filter capacitor current signatures. Unlike conventional differential or resistance-based protection schemes, this approach relies solely on monitoring the transient and steady-state dynamics of the filter capacitor current to identify low- and high-impedance short-circuit faults. A fault index (δi_C) is computed from the difference between current samples and compared with a defined threshold to determine fault severity. The proposed scheme successfully detects both line-to-line and line-to-ground faults with a rapid isolation time of approximately 1.5 milliseconds and demonstrates high sensitivity and low false detection rates under various loading and converter operating conditions. Simulation and experimental validation confirm the model's robustness and reliability while significantly reducing sensor and communication

requirements. Overall, this current-signature-based method provides a fast, cost-effective, and accurate alternative for PV-integrated DC microgrids, enhancing fault resilience and operational safety without relying on complex machine learning architectures.

Shaharier Kabir *et al.* [34] presents an unsupervised anomaly detection and fault localization method for PV systems using the Isolation Forest (iForest) algorithm combined with a rule-based fault detection approach. The study aims to enhance reliability and operational efficiency in large-scale PV power plants by identifying both internal and external anomalies without requiring labeled training data. The iForest model isolates outliers by constructing random decision trees from PV sensor data—including AC/DC power, irradiance, and inverter temperature—collected over 34 days with 15-minute intervals. Anomalous data points indicate potential faults, which are then localized through rule-based logic that compares irradiance–power and DC–AC conversion relationships to pinpoint defective panels or inverter faults. The proposed method detected 453 anomalies out of 45,740 observations, identifying six faulty panels with an overall accuracy score of 98.86%, outperforming previous models that achieved only 89.63% accuracy. This hybrid approach provides a fast, data-driven, and sensor-efficient solution for PV fault diagnosis, capable of detecting and localizing both inverter and panel-level anomalies. The research demonstrates that the iForest algorithm is a robust and effective tool for unsupervised anomaly detection and fault localization in large-scale PV systems, achieving high detection accuracy without requiring labeled training data.

Unsupervised anomaly detection methods are well suited for single-string monitoring, where labelled fault data are unavailable, and faults may emerge gradually rather than abruptly. Unsupervised anomaly detection methods are particularly suitable for single-string monitoring, as they can detect subtle deviations without requiring labelled data and can operate effectively at fine-grained string-level measurements.

Table 2 summarizes the recent developments in fault detection models for PV systems, demonstrating a clear evolution from traditional threshold-based diagnostic methods to advanced machine learning and deep learning frameworks. Conventional approaches such as I–V curve tracing and thermal monitoring are often limited in identifying low-mismatch or complex multi-string faults. In contrast, AI-based methods—including ANN, CNN, and Recurrent Neural Networks (LSTM, GRU)—have achieved remarkable performance in detecting and classifying diverse faults such as open circuit, short circuit, shading, soiling, ground, and arc faults. Hybrid and semi-supervised models, including CNN–SVM, EMD–GRU, and IoT-integrated architectures, have further enhanced diagnostic speed and adaptability while minimizing reliance on labeled datasets. These models reported accuracies ranging between 88.7% and 100%, with notable improvements in precision, recall, and reduced Mean Time to Repair (MTTR). Although challenges remain—such as overfitting, dependency on data quality, and limited interpretability—AI-driven models have demonstrated superior scalability, robustness, and real-time applicability. Validation through MATLAB/Simulink simulations, OPAL-RT platforms, and IoT-based hardware testing confirms their practical relevance in large-scale PV arrays, highlighting their potential to significantly enhance operational efficiency, predictive maintenance, and long-term reliability in modern solar energy systems. Among these approaches, unsupervised and lightweight models are more aligned with single-string level monitoring due to

their low dependency on labelled data and ability to operate with limited sensing infrastructure. Table 3 compares commonly used machine learning techniques for photovoltaic fault detection based on input data requirements, computational complexity, and reported performance. Supervised learning models such as ANN, CNN, LSTM, and ensemble learning generally achieve very high classification accuracy when labelled datasets are available. However, obtaining labelled PV fault data in real operating conditions can be challenging.

Unsupervised approaches such as autoencoders and isolation forest provide alternative solutions by detecting anomalies without requiring labelled datasets. Furthermore, lightweight models such as logistic regression, support vector machines, and isolation forest offer lower computational complexity and are more suitable for deployment in embedded monitoring systems or edge-based PV diagnostics.

Table 2 Summary Fault Detection Model.

Category	Details
Types of PV Faults	Open circuit, short circuit, ground faults, arc faults, bypass diode faults, shading, soiling, degraded modules, string-to-string, and line-to-ground faults.
Impact on PV Performance	Alters I–V and P–V characteristics, reducing voltage, current, and power output. Causes mismatch losses, accelerated degradation, reduced efficiency, and potential fire or electric hazards.
Traditional Methods	Basic electrical monitoring, I–V curve tracing, and threshold analysis; limited in detecting low-mismatch or complex multi-string faults.
AI Techniques Used	• Artificial Neural Networks (ANN) Feedforward, Dropout, Pruned, Random Search, Wavelet–ANN. • Convolutional Neural Networks (CNN) Image-based classification (VGG16, CNN–SVM hybrid). • Recurrent Models LSTM, GRU, RNN with IoT integration. • Autoencoders Deep, Semi-supervised, LSTM–DAE. • Traditional ML SVM (RBF kernel), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Ensemble Learning. • Hybrid/Advanced CNN–SVM, EMD–GRU, LSTM–SVM–LR, IoT-integrated frameworks. • Unsupervised/Signature-based Isolation Forest, Filter Capacitor Current Signatures.
Input Parameters	Voltage, current, temperature, irradiance, power output, frequency-domain and time-series data, image data (EL/Thermal), and statistical features (entropy, energy, kurtosis, crest factor).
Model Applications	Fault detection, classification, localization, anomaly detection, degradation diagnosis, and predictive maintenance for large-scale PV arrays and microgrids.
Advantages of AI Models	High detection accuracy, adaptability to varying fault conditions, low computational cost, real-time capability, reduced dependency on labeled data, and strong generalization under dynamic environments.
Limitations & Challenges	Overfitting risk in deep models, dependency on data quality, limited interpretability, computational load, and reduced transferability across different PV systems.
Real-World Validation	MATLAB/Simulink and OPAL-RT simulations, IoT hardware testing (ESP32), EL image datasets (PVEL-AD), and PV arrays from 250 kW to 3 MW under real irradiance and temperature variations.

Table 3 Comparison of Machine Learning Models for PV Fault Detection.

Model / Technique	Input Data Type	Label Requirement	Computational Complexity	Typical Accuracy	Deployment Suitability	Reference
ANN with Dropout / Pruning	Electrical parameters	Labelled	Medium	High classification accuracy	Suitable for fault classification	[17]
Artificial Neural Network (ANN)	Voltage, current, temperature	Labelled	Medium	100% (training), 98% testing	Suitable for real-time PV monitoring	[18]
LSTM-based Fault Detection	Voltage and current time-series	Labelled sequential data	Medium–High	High classification accuracy	Suitable for dynamic microgrid monitoring	[19]
CNN-based Image Classification	Electroluminescence / thermal / panel images	Large labelled dataset	High	91.1% (binary), 88.6% (multi-class)	Mainly cloud or GPU systems	[21]
ANN with Wavelet Transform	Current signal features	Labelled	Medium	High fault detection accuracy	Suitable for microgrid protection systems	[22]
CNN-based Electrical Parameter Detection	Voltage, current, power output	Labelled	High	97.5%	Suitable for centralized monitoring	[24]
Ensemble Learning (Decision Trees)	String current and array voltage	Labelled	Medium	Up to 100%	Suitable for large PV plants	[25]
Support Vector Machine (SVM-RBF)	I–V and P–V characteristics	Labelled	Low–Medium	99.85%	Suitable for embedded PV diagnostics	[26]
Logistic Regression	Electrical parameters	Labelled	Low	88.7%	Lightweight deployment	[27]
RNN with IoT Monitoring	Voltage and current signals	Labelled	Medium	99.67% ³	Suitable for IoT-based PV monitoring	[29]
Autoencoder-based Anomaly Detection	PV power output data	Unlabelled	Low–Medium	82% (F1-score)	Suitable for anomaly detection	[32]
Isolation Forest	Sensor data (power, irradiance, temperature)	Unlabelled	Low	98.86%	Highly suitable for large-scale monitoring	[34]

5. Model validation or assessment

Model validation is a critical process in the development and deployment of machine learning models, particularly in the context of PV fault detection and power forecasting. It ensures that a model not only fits the training data well but also generalizes effectively to unseen real-world scenarios. Without rigorous validation, models

are prone to overfitting, resulting in inaccurate predictions and unreliable system behavior in operational environments.

Common validation techniques—such as cross-validation, confusion matrices, and statistical performance metrics like Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2)—serve to objectively assess a model's predictive

performance and robustness. In PV applications, where operational reliability and precision are paramount, proper validation is essential to guarantee the practical utility and safety of deployed models. The model validation in the paper by Rahman *et al.* [35] effectively demonstrates the superiority of SolarNet over traditional forecasting methods. Using real-world datasets, the authors evaluated the model performance through key metrics such as R-squared, RMSE, MAE, and MAPE. The results show that SolarNet, with its ensemble of a 1D hybrid CNN, attention-based Bidirectional Long Short-Term Memory (BiLSTM), and linear regression, achieved an R-squared value of 0.924 and an RMSE of 0.072 in hour-based predictions, indicating high accuracy. This thorough validation confirms SolarNet's robustness and reliability, highlighting its potential as a leading tool for solar power forecasting.

The paper by Ponraj *et al.* [36] primarily focuses on the design of machine learning models for PV power prediction, exploring the effectiveness of Decision Tree (DT) and Random Forest (RF) algorithms. The study outlines how model structure, algorithm selection, and feature inputs play a vital role in achieving accurate forecasts. To evaluate the reliability of these designs, the authors applied model validation techniques using performance metrics such as MAE and R-squared. The Random Forest model demonstrated strong validation results with an MAE of 4.41 and an R-squared of 0.91, whereas the Decision Tree model yielded an MAE of 55.96 and a negative R-squared value of -7.21. This contrast reinforces the importance of combining thoughtful model design with rigorous validation to ensure dependable PV power forecasting outcomes.

The paper by Yoga Saputra *et al.* [37] investigates feature selection for electricity power forecasting in solar power plants using deep learning models. The study employs LSTM and BiLSTM models. Model validation is carried out using RMSE and R^2 as the primary metrics. The LSTM model, with optimized hyperparameters, achieved an RMSE of 34.122 and an R^2 of 0.842, indicating high prediction accuracy. Hyperparameter tuning was performed using GridSearch with Cross Validation (CV), which significantly improved the model's performance.

The paper by Frederick Sidney Correa *et al.* [38] focuses on the design and comparative analysis of deep learning architectures for solar power forecasting, particularly exploring combinations such as Autoencoder (AE)-GRU and CNN-LSTM. The study emphasizes how architectural choices—such as feature extraction through autoencoders and temporal learning through recurrent networks—impact model effectiveness.

To assess the performance of these models, validation techniques using MAE and Mean Squared Error (MSE) were employed. Among the models tested, the AE-GRU architecture achieved the best results, with an MAE of 0.0277 and an MSE of 0.0341, demonstrating its superior accuracy in solar power prediction.

The paper by Vira Shendryk *et al.* [39] presents short-term solar power generation forecasting methods. The study utilizes RMSE, MAE, and Mean Absolute Percentage Error (MAPE) as the primary validation metrics to evaluate the performance of different forecasting models. Specifically, the models employed include ANN and LSTM networks. The ANN model achieved an RMSE of 4.1% and an MAPE of 3.7%, demonstrating its accuracy in short-term predictions. The LSTM model also showed robust performance, validating the effectiveness of these metrics in assessing the accuracy and reliability of forecasting models.

The paper by Khushbu *et al.* [40] explores the design and comparative performance of various machine learning models for fault detection and classification in PV array systems. The study examines classifiers such as Decision Tree, XGBoost, Random Forest, and Neural Networks, emphasizing how algorithm selection and configuration affect fault diagnosis accuracy.

Azimi *et al.* [41] utilized a fine-tuned VGG-16 CNN model for fault detection in photovoltaic systems, achieving an accuracy of 98.76%. The validation techniques included generating two-dimensional scalograms using continuous wavelet transform (CWT) from the photovoltaic data, which were then used as input for the CNN. The pre-trained VGG-16 model was validated through a process involving the freezing of initial convolutional layers while training additional classification layers. This fine-tuning method ensured robust fault detection performance. Accuracy was assessed using metrics such as confusion matrices and average accuracy calculations, which confirmed the model's effectiveness in detecting faults across various scenarios.

The study by Deekshita *et al.* [42] centers on the design of a CNN model for fault detection in photovoltaic systems, utilizing wavelet transforms to convert current signals into scalograms for image-based analysis. This innovative preprocessing step enables the CNN to effectively learn and classify fault patterns from visual signal representations. To evaluate the model's performance, the authors employed a confusion matrix along with key metrics such as precision, recall, and F1-score. These validation results confirmed the model's high accuracy and robustness across different fault scenarios.

In the study by Meghana Ramesh *et al.* [43], a CNN was utilized to classify fault zones in distribution feeders with photovoltaics. The model's validation involved using confusion matrices and performance metrics such as precision, recall, and F1-score, achieving an accuracy of 99.93%. These methods ensure rigorous assessment of the model's ability to accurately classify faults from different protection zones, confirming its effectiveness in practical scenarios.

Across the reviewed studies, a range of model validation techniques have been employed to assess the performance and reliability of machine learning and deep learning models in solar power forecasting and PV fault detection as shown in Table 4. These validation methods collectively enhance the robustness, transparency, and reliability of the models, allowing researchers and practitioners to make informed decisions about model deployment in real-world PV monitoring and forecasting applications.

To validate the effectiveness of these models, cross-validation was employed as the primary technique, ensuring consistent and unbiased model assessment. Through this process, the Neural Network classifier emerged as the most accurate, achieving the lowest mean squared error (MSE) among all models evaluated.

In PV anomaly detection, the selection of evaluation metrics should consider the presence of class imbalance and the need for early fault detection. Since fault events are typically rare, relying solely on accuracy can be misleading, as models may achieve high accuracy by predominantly predicting normal conditions. Therefore, metrics such as precision, recall, and F1-score are more appropriate. Recall is particularly important for early anomaly detection, as it ensures that actual fault conditions are correctly identified and not missed. Meanwhile, precision helps reduce false alarms, which is critical to avoid unnecessary maintenance actions. The F1-score provides a balanced evaluation between fault detection capability and false alarm reduction. For regression-based approaches such as

reconstruction error methods, metrics like MAE, RMSE, and MSE remain useful but should be complemented with appropriate thresholding strategies to effectively distinguish between normal and anomalous behaviour. Table 4 presents commonly used validation metrics include R^2 , MAE, RMSE, MAPE, and Mean Squared Error (MSE) providing insight into different aspects of prediction accuracy. Additionally, confusion matrices, precision, recall, and F1-score are

frequently applied in classification tasks to evaluate model sensitivity and specificity. Several studies also adopt cross-validation (e.g., k-fold or GridSearch CV) to ensure generalizability and prevent overfitting, especially when tuning hyperparameters.

Table 4 Summary Model Validation Method.

Validation Method / Metric	Purpose	Advantages
R-squared (R^2)	Measures proportion of variance explained by the model	Indicates how well the model fits the observed data
MAE (Mean Absolute Error)	Measures average magnitude of errors	Easy to interpret and less sensitive to outliers
RMSE (Root Mean Squared Error)	Penalizes larger errors more than MAE	Highlights large prediction errors effectively
MAPE (Mean Absolute Percentage Error)	Measures percentage error relative to actual values	Useful for comparing prediction accuracy across datasets
MSE (Mean Squared Error)	Measures average squared error	Strongly penalizes large prediction deviations
Confusion Matrix	Provides overview of classification outcomes (TP, FP, FN, TN)	Offers detailed insight into classification performance
Precision	Measures correct positive predictions	Important when false positives must be minimized
Recall (Sensitivity)	Measures how many actual positives were predicted correctly	Important for detecting rare or critical faults
F1-score	Harmonic mean of precision and recall	Balances precision and recall in classification tasks
Cross-Validation (e.g., k-fold)	Splits data into multiple training/testing sets	Improves model generalization and reduces overfitting

These validation methods collectively enhance the robustness, transparency, and reliability of the models, allowing researchers and practitioners to make informed decisions about model deployment in real-world PV monitoring and forecasting applications. To validate the effectiveness of these models, cross-validation was employed as the primary technique, ensuring consistent and unbiased model assessment. Through this process, the Neural Network classifier emerged as the most accurate, achieving the lowest mean squared error (MSE) among all models evaluated.

In PV anomaly detection, the selection of evaluation metrics should consider the presence of class imbalance and the need for early fault detection.

Since fault events are typically rare, relying solely on accuracy can be misleading, as models may achieve high accuracy by predominantly predicting normal conditions. Therefore, metrics such as precision, recall, and F1-score are more appropriate. Recall is particularly important for early anomaly detection, as it ensures that actual fault conditions are correctly identified and not missed. Meanwhile, precision helps reduce false alarms, which is critical to avoid unnecessary maintenance actions. The F1-score provides a balanced evaluation between fault detection capability and false alarm reduction. For regression-based approaches such as reconstruction error methods, metrics like MAE, RMSE, and MSE remain useful but should be complemented with appropriate thresholding strategies to effectively distinguish between normal and anomalous behaviour.

6. Malaysian approach to fault detection model

In recent years, Malaysia has seen a surge in research focusing on the application of machine learning techniques for fault detection in PV systems, driven by the growing adoption of large-scale solar (LSS) projects and the need for efficient, predictive maintenance solutions. Researchers are leveraging advanced algorithms to enhance fault detection accuracy, reduce operational costs, and improve the overall reliability of PV systems operating under Malaysia's unique environmental conditions. Below are some of the research has been done in Malaysia.

In a study conducted in Malaysia, researchers introduced a predictive maintenance strategy for large-scale solar (LSS) PV plants by combining K-Means clustering with LSTM neural networks [44]. K-Means was employed to group PV string current data according to environmental variables such as solar irradiance and module temperature. Within each cluster, the LSTM model predicted future current values and identified anomalies. When compared to an ANN, the LSTM achieved a lower average relative error (4.316% vs. 4.363%) along with superior performance in terms of MAE, MSE, and RMSE. These results indicate that this hybrid approach provides a more accurate and cost-efficient method for early fault detection and predictive maintenance in LSS PV systems.

In another study, a Smart Fault Diagnosis and Classification (SFDC) model was developed using the Random Forest algorithm enhanced with 10-fold cross-validation and a carefully selected set of feature extraction techniques [45]. The model was tested using real-world irradiance data from the KMSB Solar PV plant in Kelantan, Malaysia, across small (10 kW), medium (250 kW), and large (2 MW) system scales. Features such as maximum power, mean, standard deviation, RMS, skewness, waveform length, and autoregressive coefficients contributed significantly to classification accuracy. The model achieved 100% accuracy during training and over 98% during testing for detecting faults including line-to-line faults, open circuits, ground faults, and partial shading. These outcomes demonstrate the model's effectiveness and reliability in real PV operating conditions.

A separate investigation utilized a Decision Tree (DT) algorithm, implemented via MATLAB's Classification Learner, to diagnose PV system faults [46]. The model achieved an accuracy of 85.40%, with recall, precision, and F1 scores of 94.32%, 87.37%, and 90.71% respectively. The study concluded that the DT method is effective in classifying faults such as false alarms (FA), faulty modules (FM), and partial shading (PS) in PV systems, showing promising potential for broader applications and future exploration with other classification algorithms for enhanced performance.

Further research applied a LSTM Recurrent Neural Network (RNN) to detect high impedance faults (HIF) in solar PV-integrated

power systems [47]. The model achieved an overall classification accuracy of 91.21% and a success rate of 92.42% in detecting high impedance faults (HIF) in a solar PV integrated power system. The study concludes that the proposed LSTM model significantly outperforms other classifiers such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), J48 Decision Tree (JDT), and Naïve Bayes (NB) based on various performance indices like Kappa statistics, precision, recall, and F-measure, making it an effective and robust solution for accurate HIF detection.

Another recent study employed a Vision Transformer (ViT) deep learning architecture for the detection of overlapping and multi-label faults in PV modules using electroluminescence (EL) imaging [48]. The model leveraged transformer-based self-attention mechanisms to capture long-range spatial relationships across PV cells, enabling precise identification of concurrent defects such as cracks, dislocations, and black cores that are often missed by conventional CNNs. Trained on the PVEL-AD dataset containing twelve defect categories, the ViT achieved a classification accuracy of 93.5%, outperforming benchmark CNN architectures including MobileNet (90.4%), EfficientNet (90.3%), and ResNet (89.6%). The model also demonstrated superior generalization, stable convergence, and robustness against overlapping faults, showing its capability to handle complex real-world PV defect scenarios effectively. These findings emphasize the growing potential of Vision Transformers as a powerful and scalable approach for automated PV fault detection, offering improved accuracy, faster convergence, and enhanced adaptability for predictive maintenance in large-scale solar power installations.

A recent study proposed a Vertices Principal Component Analysis (VPCA) approach for detecting partial shading faults in grid-connected PV systems [49]. The VPCA method extends the conventional Principal Component Analysis (PCA) by incorporating interval-valued data to account for measurement uncertainties from sensors such as irradiance, temperature, and voltage. Using data collected from a 2.025 kW monocrystalline PV array installed at the Power Electronics and Renewable Energy Research Laboratory (PEARL), University of Malaya, the model successfully identified partial shading faults under four distinct shading scenarios. Compared to standard PCA, the VPCA achieved a 17% improvement in detection performance, significantly reducing missed detections while maintaining comparable false alarm rates. The method demonstrated the ability to capture system uncertainties, improving reliability in detecting dynamic and intermittent shading conditions. However, the study noted that the VPCA may produce slightly higher false alarms due to its transformation of interval-valued data into single-valued form. Despite this limitation, the proposed VPCA approach provides a robust, data-driven framework for partial shading fault detection in PV systems, offering better adaptability and accuracy for monitoring real-world solar installations under Malaysia's tropical environmental conditions.

7. Conclusion

The literature shows a clear transition in PV fault research from conventional threshold-based diagnostics and electrical testing toward intelligent, data-driven methods. Supervised machine learning and deep learning models—particularly ANN, CNN, and recurrent architectures such as LSTM and GRU—now dominate, often augmented by hybrid frameworks like CNN-SVM, AE-GRU, and IoT-RNN. These approaches consistently demonstrate high accuracy in detecting and classifying open-circuit, short-circuit, ground, arc,

shading, soiling, and related faults across simulated and real operating conditions.

Validation practices are robust and multi-faceted. Regression tasks commonly report RMSE, MAE, MAPE, MSE, and R^2 to quantify predictive fidelity, while classification studies rely on confusion matrices, precision, recall, F1-score, and receiver-operating characteristics. Cross-validation (e.g., k-fold) and hyperparameter search (e.g., GridSearch CV) are frequently employed to mitigate overfitting and ensure generalization—an essential requirement for deployment in heterogeneous PV fields.

Within Malaysia's rapidly expanding solar ecosystem, empirical studies increasingly apply these models to large-scale and panel-level data under tropical conditions. Work spanning K-Means-LSTM predictive maintenance, Random-Forest-based fault diagnosis, Decision Trees, LSTM-RNN detection of high-impedance faults, Vision Transformers for EL imaging, and statistical methods like VPCA underscores strong performance and practical promise for grid resilience and cost-effective operations.

Despite significant progress in AI-based PV fault detection, a critical gap remains in the development of methods specifically tailored for single-string anomaly detection, where labelled data are scarce and early-stage faults are often subtle and difficult to identify. Most existing approaches are designed for system-level or array-level monitoring, limiting their effectiveness in detecting fine-grained anomalies at the string level. Methods such as autoencoders, Isolation Forests, clustering (e.g., K-Means), and reconstruction-based anomaly detection remain underutilized in this context. Advancing these approaches—potentially through hybrid frameworks incorporating lightweight pre-filters and online adaptation—offers a scalable pathway toward real-time, data-efficient fault detection. This is particularly important for improving predictive maintenance, reducing operational expenditure, and enhancing the long-term reliability of Malaysia's large-scale PV infrastructure.

In addition to model performance, practical deployment considerations play an important role in determining the suitability of fault detection approaches for real-world PV applications. Factors such as the availability of labelled data, the level of measurement granularity (e.g., string-level versus array-level monitoring), resilience to tropical environmental variability, and computational constraints must be taken into account. While supervised models often achieve high accuracy, their dependence on labelled datasets may limit their applicability in large-scale PV systems. In contrast, unsupervised approaches such as autoencoders and Isolation Forests offer greater flexibility in data-scarce environments and are more suitable for single-string level monitoring. Furthermore, lightweight models are preferable for edge-based implementations due to lower computational and latency requirements, whereas more complex architectures can be utilized in centralized cloud-based systems. Considering these aspects enables more practical, scalable, and cost-efficient deployment of fault detection solutions in modern PV plants.

Acknowledgements

The authors acknowledge Universiti Teknologi MARA (UiTM) for providing financial support for this research work.

References

- [1] Zhao, Y., Li, D., Lu, T., Lv, Q., Gu, N. and Shang, L., Collaborative Fault Detection for Large-Scale Photovoltaic Systems. *IEEE Transactions on Sustainable Energy*. 11 (2020) 2745-2754, doi: <https://doi.org/10.1109/TSTE.2020.2974404>.

- [2] Ahmed, S., Ali, A., Ansari, J. A., Qadir, S. A. and Kumar, L., A Comprehensive Review of Solar Photovoltaic Systems: Scope, Technologies, Applications, Progress, Challenges, and Recommendations. *IEEE Access*. 13 (2025) 69723-69750, doi: <https://doi.org/10.1109/ACCESS.2025.3558539>.
- [3] Zahari, N. E. M., Mokhlis, H., Mubarak, H., Mansor, N. N., Sulaima, M. F., Ramasamy, A. K., Zulkapli, M. F., Ja'apar, M. A. B., Jaafar, M. and Marsadek, M. B., Integrating Solar PV, Battery Storage, and Demand Response for Industrial Peak Shaving: A Systematic Review on Strategy, Challenges and Case Study in Malaysian Food Manufacturing. *IEEE Access*. 12 (2024) 106832-106856, doi: <https://doi.org/10.1109/ACCESS.2024.3420941>.
- [4] Gao, W. PV Array Fault Detection Based on Deep Neural Network. in *2021 IEEE Green Technologies Conference (GreenTech)*. (2021), 42-47, doi: <https://doi.org/10.1109/GreenTech48523.2021.00018>.
- [5] Baruah, D., Subbiah, S. and Chouhan, S. Exploiting Context for Fault Detection in Solar Photovoltaic System using Context-Sensitive Recurrent Neural Network. in *2023 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy (PESGRE)*. (2023), 1-6, doi: <https://doi.org/10.1109/PESGRE58662.2023.10404256>.
- [6] Aryan, B. K., Sobhana, O., Prabhakar, G. C. and Reddy, N. A. Fault Detection and Classification in Micro Grid Using AI Technique. in *2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)*. (2022), 1-6, doi: <https://doi.org/10.1109/ICMACC54824.2022.10093359>.
- [7] M. R, M. K. P., Ahmad, M. W. and Gorla, N. B. Y. Sliding Mode Observer Based Open-Circuit Fault Detection in the HERIC PV Inverter. in *2022 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*. (2022), 1-5, doi: <https://doi.org/10.1109/PEDES56012.2022.10080686>.
- [8] Zhang, J., Li, Y., Li, B., Chen, Q., Zhang, J. and Jiao, T. Theoretical Mismatch Degree Calculation of PV Array by Traversal Cell-Based Short-Circuit Characteristic Fault Topologies. in *2023 5th International Conference on Power and Energy Technology (ICPET)*. (2023), 422-428, doi: <https://doi.org/10.1109/ICPET59380.2023.10367601>.
- [9] Pillai, D. S. and Rajasekar, N., An MPPT-Based Sensorless Line-Line and Line-Ground Fault Detection Technique for PV Systems. *IEEE Transactions on Power Electronics*. 34 (2019) 8646-8659, doi: <https://doi.org/10.1109/TPEL.2018.2884292>.
- [10] Li, W., Li, T. and Su, J. DC Series Arc Fault Detection Method for Photovoltaic Array Based on Random Forest. in *2024 IEEE 2nd International Conference on Control, Electronics and Computer Technology (ICCECT)*. (2024), 138-143, doi: <https://doi.org/10.1109/ICCECT60629.2024.10545829>.
- [11] Abdulmawjood, K. and Morsi, W. G. Comparative Analysis of Machine Learning Algorithms for Identifying Partial Shading Conditions on PV Array. in *IECON 2024 - 50th Annual Conference of the IEEE Industrial Electronics Society*. (2024), 1-6, doi: <https://doi.org/10.1109/IECON55916.2024.10906002>.
- [12] Pujara, D., Ramirez, D., Tepedelenlioglu, C., Srinivasan, D. and Spanias, A. Real-time PV Fault Detection using Embedded Machine Learning. in *2024 IEEE 7th International Conference on Industrial Cyber-Physical Systems (ICPS)*. (2024), 1-5, doi: <https://doi.org/10.1109/ICPS59941.2024.10640018>.
- [13] Ishak, N. H. B., Isa, I. S. B., Osman, M. K. B., Daud, K. and Jadin, M. S. B. Hotspot Detection of Solar Photovoltaic System: A Perspective from Image Processing. in *2023 IEEE 3rd International Conference in Power Engineering Applications (ICPEA)*. (2023), 263-267, doi: <https://doi.org/10.1109/ICPEA56918.2023.10093148>.
- [14] Martinez-Loran, E., Gastrow, G. v., Clenney, J., Meier, R., Bandaru, P., Bertoni, M. I. and Fenning, D. P., Finite Element Simulation of Potential-Induced Degradation Kinetics in p-Type Silicon Solar Modules. *IEEE Journal of Photovoltaics*. 12 (2022) 45-52, doi: <https://doi.org/10.1109/JPHOTOV.2021.3123870>.
- [15] Pei, T. and Hao, X., A Fault Detection Method for Photovoltaic Systems Based on Voltage and Current Observation and Evaluation. *Energies*. 12 (2019) 1712, doi: <https://doi.org/10.3390/en12091712>.
- [16] Rao, S., Spanias, A. and Tepedelenlioglu, C. Solar Array Fault Detection using Neural Networks. in *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*. (2019), 196-200, doi: <https://doi.org/10.1109/ICPHYS.2019.8780208>.
- [17] Rao, S., Muniraju, G., Tepedelenlioglu, C., Srinivasan, D., Tamizhmani, G. and Spanias, A., Dropout and Pruned Neural Networks for Fault Classification in Photovoltaic Arrays. *IEEE Access*. 9 (2021) 120034-120042, doi: <https://doi.org/10.1109/ACCESS.2021.3108684>.
- [18] Bakshi, B. G., Wasif, S., Sarkar, S. and Wasim, S. A study on fault diagnosis in solar photovoltaic array using artificial neural network. in *2024 IEEE 3rd International Conference on Control, Instrumentation, Energy & Communication (CIEC)*. (2024), 124-128, doi: <https://doi.org/10.1109/CIEC59440.2024.10468482>.
- [19] Devi, A. D., Sinha, N., Saikia, L. C. and Singh, S. R. Detection and Classification of Faults in An Islanded Microgrid Using LSTM Model and its Real Time Validation. in *2023 IEEE Silchar Subsection Conference (SILCON)*. (2023), 1-6, doi: <https://doi.org/10.1109/SILCON59133.2023.10404990>.
- [20] Dhoundiyal, P., Kumar, Y., Negi, S. and Barthwal, A. Fault Detection and Classification in Solar Photovoltaic Array. in *2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT)*. (2022), 1-6, doi: <https://doi.org/10.1109/GlobConPT57482.2022.9938265>.
- [21] Pa, M., Uddin, M. N. and Kazemi, A. A Fault Detection Scheme Utilizing Convolutional Neural Network for PV Solar Panels with High Accuracy. in *2022 IEEE 1st Industrial Electronics Society Annual On-Line Conference (ONCON)*. (2022), 1-5, doi: <https://doi.org/10.1109/ONCON56984.2022.10126746>.
- [22] Jayasinghe, J. A. R. R., Malindi, J. H. E., Rajapaksha, R. M. A. M., Logeeshan, V. and Wanigasekara, C. Enhanced Fault Classification and Localization in Microgrids Using Machine Learning. in *2023 IEEE World AI IoT Congress (AllIoT)*. (2023), 0542-0547, doi: <https://doi.org/10.1109/AllIoT58121.2023.10174440>.
- [23] Dhimish, M. and Tyrrell, A. M., Photovoltaic Bypass Diode Fault Detection Using Artificial Neural Networks. *IEEE Transactions on Instrumentation and Measurement*. 72 (2023) 1-10, doi: <https://doi.org/10.1109/TIM.2023.3244230>.
- [24] Bendale, H., Aswar, H., Bamb, H., Desai, P. and Aher, C. N. Deep Learning for Solar Panel Maintenance: Detecting Faults and Improving Performance. in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. (2023), 1-5, doi: <https://doi.org/10.1109/ICCCNT56998.2023.10307465>.
- [25] Kumar, S., Mohanty, S. K. and Nayak, P. K. PV Array Fault Detection and Location Estimation Using Ensemble Learning Model. in *2025 IEEE 1st International Conference on Smart*

- and Sustainable Developments in Electrical Engineering (SSDEE). (2025), 1-6, doi: <https://doi.org/10.1109/SSDEE64538.2025.10968531>.
- [26] Gassab, E., Zaidi, N. and Khedher, A. PV System Faults Detection and Classification Using Multi-class Support Vector Machine Algorithm. in *2024 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*. (2024), 1-6, doi: https://doi.org/10.1109/IC_ASET61847.2024.10596185.
- [27] Ragul, S., Tamilselvi, S., Elamurugan, P. and Bharathidasan, S. Enhancing Photovoltaic System Resilience: A Logistic Regression Approach to Fault Diagnosis. in *2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*. (2024), 1-5, doi: <https://doi.org/10.1109/RAEEUCCI61380.2024.10547898>.
- [28] Jayapriya, J. and Kalaivani, R. Performance Evaluation of Hybrid CNN-SVM model for Efficient Detection of Cracks and Faults in Solar Panels. in *2025 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)*. (2025), 1-5, doi: <https://doi.org/10.1109/AMATHE65477.2025.11081220>.
- [29] Govindarajan, S., Subramanian, V., K, R., A, J., E, G. and P, P. Enhanced Fault Detection and Classification in Solar PV Systems Using RNN and IoT Integration. in *2024 10th International Conference on Electrical Energy Systems (ICEES)*. (2024), 1-4, doi: <https://doi.org/10.1109/ICEES61253.2024.10776922>.
- [30] Wai, R. J. and Cai, X. Newly-Designed DC-Arc Fault Detection Strategy Via EMD Signal Analysis and GRU Neural Network for Solar PV Power Generation System. in *2023 5th International Conference on Electrical, Control and Instrumentation Engineering (ICECIE)*. (2023), 1-8, doi: <https://doi.org/10.1109/ICECIE58751.2024.10457366>.
- [31] Tsai, C. W., Yang, C. W., Hsu, F. L., Tang, H. M., Fan, N. C. and Lin, C. Y. Anomaly Detection Mechanism for Solar Generation using Semi-supervision Learning Model. in *2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)*. (2020), 9-13, doi: <https://doi.org/10.1109/Indo-TaiwanICAN48429.2020.9181310>.
- [32] Miraftabzadeh, S. M., Longo, M., Foadelli, F. and Bracco, S. Anomaly Detection in Photovoltaic Systems via Deep Learning Autoencoder. in *2023 International Conference on Smart Energy Systems and Technologies (SEST)*. (2023), 1-6, doi: <https://doi.org/10.1109/SEST57387.2023.10257517>.
- [33] Yadav, N. and Tummuru, N. R., Short-Circuit Fault Detection and Isolation Using Filter Capacitor Current Signature in Low-Voltage DC Microgrid Applications. *IEEE Transactions on Industrial Electronics*. 69 (2022) 8491-8500, doi: <https://doi.org/10.1109/TIE.2021.3109523>.
- [34] Kabir, S., Shufian, A. and Zishan, M. S. R. Isolation Forest Based Anomaly Detection and Fault Localization for Solar PV System. in *2023 3rd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. (2023), 341-345, doi: <https://doi.org/10.1109/ICREST57604.2023.10070033>.
- [35] Rahman, M. T., Pall, S. C., Rashid, R. B., Hossen, M. S., Das, S. K. and Cheragee, S. H. SolarNet- An efficient multi-model ensemble approach for Solar Power Forecasting based on Artificial Intelligence. in *2023 6th International Conference on Electrical Information and Communication Technology (EICT)*. (2023), 1-6, doi: <https://doi.org/10.1109/EICT61409.2023.10427929>.
- [36] Ponraj, R. P., Ravindran, V., Chittibabu, K., A, T., C, A. and Louis, E. Machine Learning based Solar PV Power Prediction. in *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*. (2023), 1646-1652, doi: <https://doi.org/10.1109/ICAISS58487.2023.10250567>.
- [37] Saputra, I. N. Y., Adytia, D. and Aditya, I. A. Feature Selection for Electricity Power Forecasting of Solar Power Plants. in *2023 3rd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*. (2023), 225-230, doi: <https://doi.org/10.1109/ICICyTA60173.2023.10428929>.
- [38] Correa, F. S. Deep Learning Based Solar Power Forecasting with Improved Representation. in *2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC)*. (2023), 57-61, doi: <https://doi.org/10.1109/PEEIC59336.2023.10452052>.
- [39] Shendryk, V., Parfenenko, Y., Kholiavka, Y., Pavlenko, P., Shendryk, O. and Bratushka, L. Short-term Solar Power Generation Forecasting for Microgrid. in *2022 IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC)*. (2022), 1-5, doi: <https://doi.org/10.1109/SAIC57818.2022.9922982>.
- [40] Pahwa, K., Sharma, M., Saggu, M. S. and Mandpura, A. K. Performance Evaluation of Machine Learning Techniques for Fault Detection and Classification in PV Array Systems. in *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*. (2020), 791-796, doi: <https://doi.org/10.1109/SPIN48934.2020.9071223>.
- [41] Azimi, S. and Manthouri, M. Fault Detection of Photovoltaic Systems using Pre-Train CNN-VGG16. in *2023 9th International Conference on Control, Instrumentation and Automation (ICCIA)*. (2023), 1-7, doi: <https://doi.org/10.1109/ICCIA61416.2023.10506372>.
- [42] Sriraman, D. and Ramaprabha, R. Application of Machine Learning and Convolutional Neural Networks for the Fault Detection and Classification Monitoring System in PV Plants. in *2023 9th International Conference on Electrical Energy Systems (ICEES)*. (2023), 694-699, doi: <https://doi.org/10.1109/ICEES57979.2023.10110284>.
- [43] Ramesh, M., Chatterjee, K., Glover, D., Follum, J., McDermott, T. and Reiman, A. Convolutional Neural Network-Based Protection-Zone Classification of Faults in Distribution Feeders with Photovoltaics. in *2024 IEEE Green Technologies Conference (GreenTech)*. (2024), 173-177, doi: <https://doi.org/10.1109/GreenTech58819.2024.10520556>.
- [44] Zulfauzi, I. A., Dahlan, N. Y., Sintuya, H. and Setthapun, W., Anomaly detection using K-Means and long-short term memory for predictive maintenance of large-scale solar (LSS) photovoltaic plant. *Energy Reports*. 9 (2023) 154-158, doi: <https://doi.org/10.1016/j.egy.2023.09.159>.
- [45] Sujod, M. Z., Mohd Ghazali, S. N. A., Abdul Kadir, M. F. and Al-Shetwi, A. Q., PV Fault Classification: Impact on Accuracy Performance Using Feature Extraction in Random-Forest Cross Validation Algorithm. *Journal of Advanced Research Design*. 123 (2024) 66-78, doi: <https://doi.org/10.37934/ard.123.1.6678>.
- [46] Ridzuan, N. M., Shariff, K. K. M. and Muhammad, N. Fault Detection and Diagnosis in Photovoltaic Systems Using Decision Tree Method. in *2024 IEEE Symposium on Wireless*

- Technology & Applications (ISWTA)*. (2024), 221-224, doi: <https://doi.org/10.1109/ISWTA62130.2024.10651909>.
- [47] Veerasamy, V., Wahab, N. I. A., Othman, M. L., Padmanaban, S., Sekar, K., Ramachandran, R., Hizam, H., Vinayagam, A. and Islam, M. Z., LSTM Recurrent Neural Network Classifier for High Impedance Fault Detection in Solar PV Integrated Power System. *IEEE Access*. 9 (2021) 32672-32687, doi: <https://doi.org/10.1109/ACCESS.2021.3060800>.
- [48] Khan, U., Khan, S. F., Khalil, I. U. and Ahsan, M., Deep learning based vision transformer approach for detecting overlapping PV faults using multi labeling. *Measurement*. 256 (2025) 118537, doi: <https://doi.org/10.1016/j.measurement.2025.118537>.
- [49] Rouani, L., Harkat, M. F., Kouadri, A. and Mekhilef, S., Shading fault detection in a grid-connected PV system using vertices principal component analysis. *Renewable Energy*. 164 (2021) 1527-1539, doi: <https://doi.org/10.1016/j.renene.2020.10.059>.