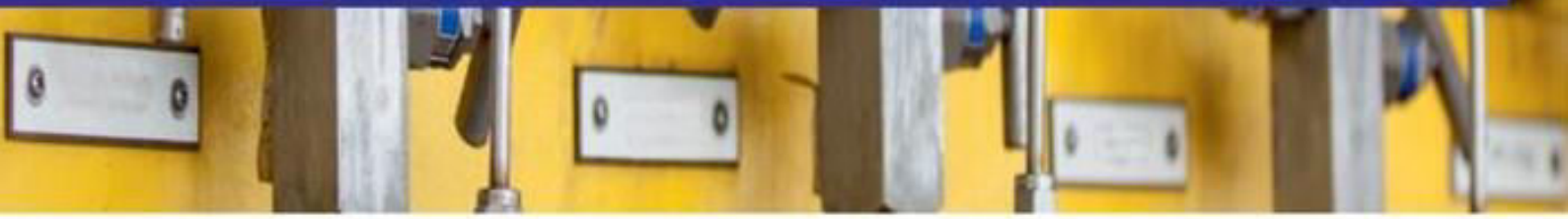


IJAREEIE

e-ISSN: 2278-8875
p-ISSN: 2320-3765

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

Volume 15, Issue 11, November 2026



INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 9.867

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

The Review on AI and Machine Learning-Based Protection Techniques for Solar Photovoltaic Power Plants

K.Poojitha, J.Yogitha, V.Premkumar, Ch.Charan, M.Rambabu

B. Tech Scholar, Department of EEE, GMRIT-Deemed to be University, Vizianagaram, Andhra Pradesh, India

B. Tech Scholar, Department of EEE, GMRIT Deemed to be University, Vizianagaram, Andhra Pradesh, India

B. Tech Scholar, Department of EEE, GMRIT Deemed to be University, Vizianagaram, Andhra Pradesh, India

B. Tech Scholar, Department of EEE, GMRIT Deemed to be University, Vizianagaram, Andhra Pradesh, India

Associate Professor, Department of EEE, GMRIT- Deemed to be University, Vizianagaram, Andhra Pradesh, India

ABSTRACT: The increasing use of solar photovoltaic (PV) plants makes dependable fault detection and protection important for safe operation, energy production and maintenance planning. PV output changes naturally with irradiance, temperature, shading and loading, so a simple fixed threshold may confuse normal variation with an actual fault. This review explains how artificial intelligence (AI), machine learning (ML) and deep learning can support conventional PV protection by analysing electrical measurements, I–V and P–V characteristics, time-series signals and inspection images. It discusses common PV faults, measurement methods, data preparation, classical ML algorithms, neural networks, image-based diagnosis, hybrid approaches and practical deployment. The reviewed literature shows that model suitability depends on the fault, data type, operating environment and available computing resources; no single algorithm is best for every application. Important barriers include limited labelled field data, class imbalance, differences between plants, sensor errors, false alarms and model ageing. A layered protection arrangement is therefore recommended: established electrical devices remain responsible for safety-critical fault clearing, while AI provides additional detection, classification, localization and maintenance information.

KEYWORDS: Solar photovoltaic systems; artificial intelligence; machine learning; fault detection; fault diagnosis; deep learning; PV protection; predictive maintenance.

I. INTRODUCTION

Solar photovoltaic generation is now used in homes, commercial buildings and large power stations. A PV plant is made up of many connected components: cells form modules, modules are connected into strings, strings form arrays, and the generated DC power is processed by inverters before being supplied to an AC load or the utility grid. This modular structure makes PV plants easy to expand. At the same time, it creates many possible points where an electrical, mechanical or environmental problem can affect performance. A loose connector, damaged cell, failed bypass diode or inverter abnormality may first appear as a small change in current, voltage, temperature or power.

Protection in a PV plant has two related purposes. The first is to prevent dangerous conditions from damaging equipment or creating a risk to people. The second is to identify developing problems early enough that maintenance can be planned before a major loss of generation occurs. Conventional devices such as fuses, circuit breakers, disconnectors, insulation monitoring systems and inverter protection functions are essential for the first purpose. However, these devices generally respond to defined electrical limits; they do not always explain the underlying cause of an abnormal operating pattern. Reviews by Hong and Pula, Gaviria et al. and Islam et al. describe the growing use of data-driven methods to add diagnostic capability to PV monitoring.

The main difficulty is that the electrical output of a healthy PV array is not constant. Irradiance changes with cloud cover, cell temperature changes throughout the day, and partial shading may affect only a small part of an array. These normal changes can resemble faults when only total power is observed. On the other hand, some faults develop slowly and may not cross a conventional protection threshold immediately. AI-based diagnosis can help by learning patterns from multiple measurements and comparing the present behaviour with examples of healthy and faulty operation.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

This paper takes an engineering-student perspective. It explains the physical meaning of the signals, describes how different AI methods can be selected, and discusses the practical steps needed before a model is trusted in an operating plant. It does not claim that AI can replace electrical protection. Instead, it treats AI as an additional decision-support layer whose output must be checked against engineering limits and plant safety requirements.

II. REVIEW APPROACH AND SCOPE

This is a narrative engineering review based on the technical material and bibliography in the supplied PV protection term-paper source, with additional cross-checking of selected review and research publications available from journal and publisher pages. The sources cover conventional PV fault detection, machine learning for PV applications, visual and thermal inspection, multimodal diagnosis, predictive maintenance and dataset limitations. Particular attention is given to papers by Hong and Pula, Gaviria et al., Islam et al., Thakfan and Bin Salamah, Balachandran et al., Afrasiabi et al., Chen et al., and Meena and Gawre.

The review compares methods by the type of input data they use, the fault or condition they are intended to identify, their engineering strengths and their limitations. It is not a PRISMA systematic review and does not claim to have independently screened every paper in the wider research field. Where the available source material describes a method only at review level, the discussion is kept at that level rather than inventing experimental results or performance values.

III. PV SYSTEM OPERATION AND COMMON FAULTS

3.1 Basic PV operation

A PV cell converts sunlight into electrical energy through the photovoltaic effect. Photons absorbed by the semiconductor create charge carriers, and the internal electric field of the cell helps separate them. The practical electrical behaviour is often represented by a current source, a diode and resistive elements. The photocurrent mainly depends on irradiance, while cell temperature has a strong influence on voltage. Series resistance represents losses in contacts and conductors; shunt resistance represents leakage paths within the cell or module.

The current–voltage (I–V) curve shows how current changes as terminal voltage is varied. The power–voltage (P–V) curve is obtained from the product of voltage and current and indicates the maximum-power operating point. A healthy module has a characteristic curve under a given irradiance and temperature. When a fault or defect occurs, the curve may shift, become distorted or develop features such as multiple local power peaks. These changes can be used as diagnostic information, provided the environmental conditions are considered.

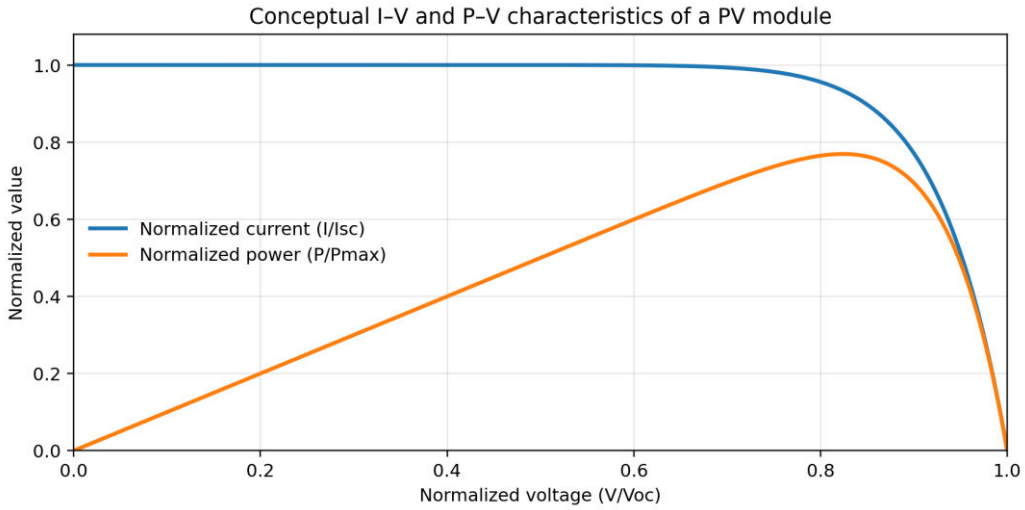


Figure 1. Conceptual normalized I–V and P–V characteristics. The curves are illustrative and are not measured experimental results.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

3.2 PV plant structure and protection zones

At plant level, modules are connected in series to obtain the required DC voltage and several strings may be connected in parallel. Combiner boxes collect string outputs and may include fuses, surge protection and monitoring. The inverter converts DC power to AC power and also provides functions related to grid synchronization, voltage and frequency limits, and internal equipment protection. Transformers and switchgear connect the plant to the distribution or transmission network.

This structure creates different protection zones. A fault inside a module may be visible through cell temperature or a change in one string's I–V curve. A fault in a DC cable or combiner box may affect several modules connected to the same circuit. An inverter fault may appear in DC measurements, AC measurements, internal temperature or event logs. For this reason, measurements should be associated with their physical location. A diagnostic system that only reports 'PV plant fault' is less useful than one that can identify the affected inverter, string or module group.

3.3 Main electrical faults and defects

Ground faults occur when a conductor unintentionally connects to earth or to a grounded structure. Depending on the system arrangement and fault path, they can create leakage current, insulation problems and safety hazards. Line-to-line faults occur when two conductors at different potentials become connected through an unintended low-resistance path. These faults can produce abnormal current and require prompt action by correctly coordinated protection devices.

Open-circuit faults may result from broken conductors, disconnected connectors, damaged solder joints or failed module interconnections. They can reduce the current in a string or disconnect part of the array. Short-circuit and high-resistance connection faults have different signatures: a low-resistance fault can create a high current path, while a resistive connection can cause local heating and power loss. Series arc faults are particularly challenging because the current may not always rise enough to operate ordinary overcurrent protection. Their detection can require analysis of high-frequency signal components and noise patterns.

Partial shading and soiling are not always equipment faults, but they can cause mismatch and reduce energy production. Bypass diodes are used to limit the effect of reverse bias in shaded cell groups. If a bypass diode fails open or short, the module's electrical behaviour can change and local heating may occur. Other physical defects include cracks, delamination, corrosion, discolouration, moisture ingress and hot spots. Some defects are visible in photographs, while others are easier to identify using infrared or electroluminescence images.

3.4 Why fault detection is difficult

A reduction in output power does not automatically mean that a fault has occurred. The same reduction may be caused by cloud cover, higher module temperature, temporary shading, dust accumulation or a change in load. In addition, faults may occur together. For example, a shaded module may also have a damaged bypass diode, making its electrical signature different from that of simple shading. A useful diagnostic method therefore needs more than one signal and should consider the operating condition at the time of measurement.

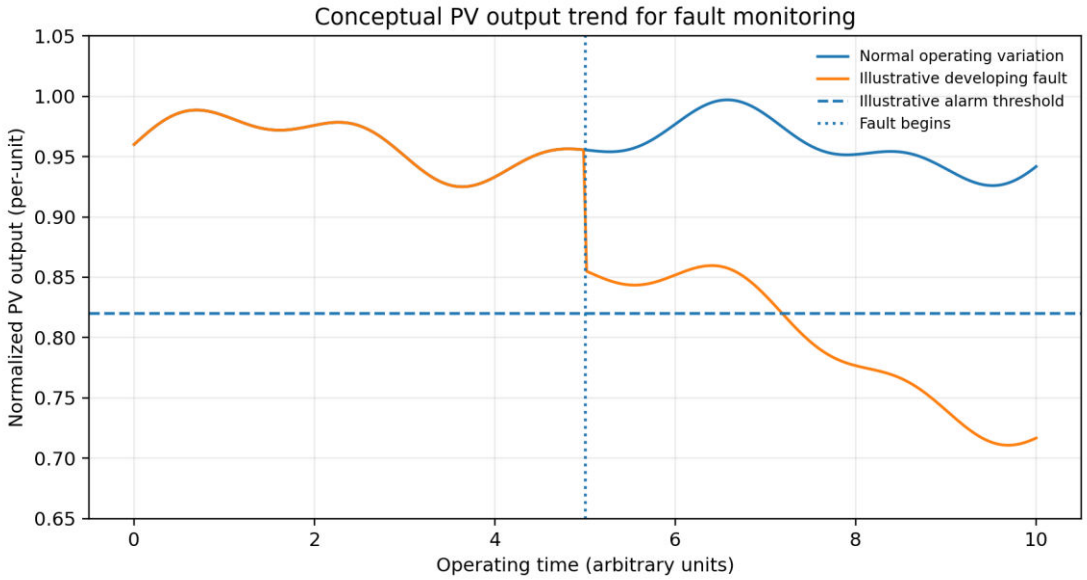
Another difficulty is the difference between detection, classification and localization. Detection asks whether the system is behaving abnormally. Classification estimates the type of fault. Localization identifies where the fault is likely to be. Severity assessment estimates how serious the condition is, while prognosis attempts to estimate how the condition may develop. These are separate tasks and may require different input data and algorithms. A model that detects an anomaly accurately may still be unable to identify its exact cause.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

Figure 4. Conceptual PV output trend illustrating normal variation and a developing fault.



Justification: The graph is included to explain why a single fixed threshold can be unreliable in a PV plant. The normal trace contains small operating fluctuations, while the illustrative fault trace begins to decline after the fault-onset marker. A monitoring method should distinguish a sustained abnormal trend from short-term changes caused by irradiance or temperature. The plotted values are normalized and deliberately illustrative; they are not experimental measurements and should not be interpreted as performance data from any cited study. This distinction is consistent with review literature describing the difficulty of separating operating-condition changes from PV faults.

IV. DATA SOURCES AND MEASUREMENT METHODS

4.1 Electrical measurements

Voltage, current and power are the most common signals available from PV monitoring systems. They can be collected at inverter, combiner or string level depending on the installed instrumentation. A sudden change in current may indicate an open circuit, mismatch or abnormal operating condition. Voltage deviation may indicate a change in string configuration or inverter operating point. Power trends are useful for long-term performance monitoring, but power alone may not identify the fault because it combines the effects of both voltage and current.

Electrical measurements are suitable for continuous monitoring because they can often be collected automatically. Their main limitation is that plant-level measurements may hide a small local defect. If the current and voltage are measured only at the inverter, a single weak module may have a small influence on the total signal. More detailed measurements at string or module level improve localization but increase sensor cost, wiring and maintenance requirements.

4.2 I–V and P–V curve analysis

I–V curve tracing provides a more detailed view of PV module or string behaviour than a single operating-point measurement. Features such as short-circuit current, open-circuit voltage, maximum-power current, maximum-power voltage, curve slope and fill factor can be extracted. A change in series resistance may affect the curve near the high-voltage region, while shunt-related problems can influence the low-voltage region. Partial shading can create steps in the I–V curve and multiple peaks in the P–V curve.

Machine learning can use either the complete curve or a set of features calculated from it. Using the complete curve may preserve more information, but it can require a larger input representation and careful normalization. Feature-based approaches are often easier to interpret and may need less computation. A practical limitation is that continuous I–V tracing across every module in a large plant may require additional hardware or temporary operating changes. Therefore, the method should be selected according to the plant size and inspection objective.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

4.3 Environmental and operating data

Irradiance and temperature are important supporting measurements. Irradiance helps explain the expected photocurrent, while module temperature helps explain voltage changes. Ambient temperature, wind speed, humidity, soiling level and tracker position may also be useful, depending on the plant. Inverter status, alarm logs, DC-link variables and AC-side measurements can help distinguish a PV array problem from a converter or grid-related event.

The value of environmental data is that they provide context. For example, a current reduction during a rapid drop in irradiance may be normal. The same reduction under stable irradiance may deserve investigation. A diagnostic model trained without environmental information may learn a misleading relationship between power and fault labels.

4.4 Visible, infrared and electroluminescence images

Visible-light images are useful for identifying surface contamination, broken glass, discoloration, vegetation shading and some forms of physical damage. Infrared thermography measures surface temperature patterns and can reveal hot spots, abnormal heating at connectors, bypass-diode problems and cell mismatch. Electroluminescence imaging is generally used under controlled conditions to reveal cracks and electrically inactive regions that may not be visible in ordinary photographs.

Image-based inspection is especially useful in large PV plants because a camera can inspect many modules without measuring each one electrically. Oliveira et al. demonstrated the use of aerial infrared thermography and orthomosaicking for utility-scale PV inspection. However, the quality of thermal images depends on irradiance, wind, camera calibration, viewing angle, emissivity and the time of inspection. A hot region is evidence that needs interpretation; it is not automatically proof of one specific fault.

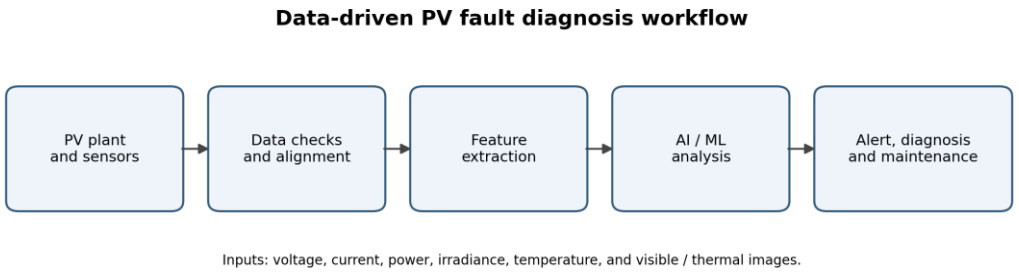


Figure 2. AI-based PV fault diagnosis workflow, from sensing and data preparation to diagnosis and maintenance action.

4.5 Combining multiple data sources

Electrical and image data provide different kinds of evidence. Electrical measurements show how the system performs as a circuit, while images show where a physical or thermal abnormality is located. A multimodal system can use electrical data to identify an affected string and an image model to inspect the modules in that string. The outputs can also be cross-checked: an electrical anomaly without a visible defect may require further electrical testing, while a hot spot with normal string-level measurements may indicate a localized issue that is too small to affect the total output.

Afrasiabi et al. describe a multimodal deep-learning approach for PV fault diagnosis. The wider lesson is that combining data is useful only when the measurements are synchronized, correctly labelled and relevant to the same equipment. If an image is taken at a different time from the electrical event, the model may connect unrelated observations and produce an incorrect diagnosis.

V. AI AND MACHINE LEARNING METHODS

5.1 Classical machine learning

Classical ML methods are often suitable when the input is a table of measured values or engineered features. These methods can be easier to train and deploy than large deep-learning models, especially when the available dataset is not very large.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

5.1.1 Support vector machine

A support vector machine (SVM) separates classes by finding a decision boundary with a wide margin between them. Kernel functions allow it to represent nonlinear boundaries. In PV diagnosis, an SVM may use features such as voltage deviation, current residual, power loss, I–V curve slope or wavelet coefficients. It can work well with a moderate number of carefully selected features. Its limitations include sensitivity to kernel and parameter selection and increased training cost for very large datasets.

5.1.2 Decision tree and random forest

A decision tree classifies a measurement by following a sequence of simple rules. For example, it may first check whether current deviation exceeds a limit and then consider temperature or voltage behaviour. This makes the result relatively easy for an engineer to follow. However, a single tree can change considerably when the training data change and may overfit if it grows too deep.

A random forest combines many decision trees trained on different samples and feature subsets. The final decision is based on the votes of the trees. This approach can capture nonlinear relationships in sensor data and is often a useful baseline for PV fault classification. Feature-importance scores can indicate which inputs influenced the model, but they should not be treated as proof of physical cause.

5.1.3 k-nearest neighbours and logistic regression

The k-nearest neighbours (KNN) method compares a new measurement with similar examples in the training dataset. It is straightforward, but the result depends on the distance measure, feature scaling and number of neighbours. Prediction can become slow when the dataset is very large. Logistic regression is a simpler probabilistic classifier and can serve as a baseline. If a complex model does not improve meaningfully over this baseline, its extra complexity may not be justified.

5.2 Artificial neural networks

An artificial neural network (ANN) contains connected processing units called neurons. Each neuron combines its inputs using weights and a bias, then applies an activation function. During training, the weights are adjusted to reduce the difference between predicted and correct outputs. ANNs can model nonlinear relationships among irradiance, temperature, voltage, current and power. They are useful when the relationship is too complicated for a small set of fixed rules, but they need representative training examples and can be difficult to interpret.

5.3 Convolutional neural networks

A convolutional neural network (CNN) is designed to learn patterns in images. Its filters identify local features such as edges, shapes, textures and temperature patterns. In PV inspection, CNNs can classify images as healthy or faulty, identify the type of visible defect, or locate a defect within an image. They can be applied to visible, infrared and electroluminescence images, but the training set must represent the cameras, environmental conditions and module types expected in practice. If a model is trained on images from one camera and tested on a different camera, its performance may decrease.

5.4 Recurrent neural networks and LSTM

PV measurements are time-dependent. A fault may develop gradually, appear intermittently or produce a pattern that is only clear when several minutes or hours of data are considered. Recurrent neural networks (RNNs) process a sequence while retaining information from earlier time steps. Long short-term memory (LSTM) networks use gates to control what information is retained and can learn longer-term dependencies. Appiah et al. applied LSTM-based automatic feature extraction to PV array fault diagnosis. Such models are relevant when the timing and evolution of measurements matter, although they require careful sequence construction and sufficient training data.

5.5 Ensemble and hybrid approaches

Ensemble learning combines predictions from several models. Bagging trains models in parallel on different samples, while boosting trains a sequence of models that gives additional attention to previous errors. Ensembles can improve stability, but they may increase memory and computation requirements.

Hybrid approaches combine methods with complementary strengths. A wavelet transform may first extract frequency-related features from a signal, after which an SVM classifies the condition. A CNN may analyse an image while an LSTM examines the associated time series. A physical PV model may estimate expected output, and an ML model may classify

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

the residual between expected and measured behaviour. These combinations are attractive because physical knowledge can guide the model while learning methods handle patterns that are difficult to describe with fixed equations.

5.6 Choosing an appropriate model

The model should be selected according to the engineering problem, not simply because a particular algorithm is popular. For structured sensor measurements, an SVM or random forest may be sufficient. For image inspection, a CNN is a natural option. For evolving time-series behaviour, an LSTM or temporal CNN may be suitable. For a task involving both electrical signals and images, a multimodal or hybrid model may be considered. The final choice should also account for training-data size, inference time, available hardware, explainability and the consequence of a wrong prediction.

Data / task	Possible methods	Main advantage	Main caution
Voltage, current, power and weather	SVM, random forest, ANN	Works with routine monitoring data	Needs environmental context and good labels
I–V / P–V curve features	SVM, RF, ANN	Uses physically meaningful curve changes	Curve acquisition may need extra instrumentation
Time-series measurements	RNN, LSTM, temporal CNN	Can learn evolving or intermittent patterns	Sequence quality and data volume matter
Visible / infrared images	CNN and vision models	Can detect and localize visual or thermal defects	Sensitive to camera and inspection conditions
Electrical plus image data	Hybrid / multimodal models	Combines circuit and spatial evidence	Synchronization and fusion design are difficult

Table 1. General relationship between PV data types and suitable AI model families. The table is a qualitative guide, not a ranking of measured accuracy.

VI. DATA PREPARATION, TRAINING AND VALIDATION

6.1 Data cleaning and synchronization

An AI model learns from the data supplied to it. If the measurements contain errors, the model may learn sensor problems instead of PV behaviour. Before training, the data should be checked for missing values, impossible readings, duplicated records, outliers and calibration problems. Measurements from different devices should use consistent timestamps and units. Electrical signals, environmental measurements and images must be aligned to the same operating period if they are to be combined.

6.2 Feature extraction and normalization

Feature extraction converts raw measurements into information that is useful for diagnosis. Examples include mean current, voltage variation, rate of change, signal standard deviation, power residual, I–V curve slope, fill factor and temperature difference between neighbouring modules. Frequency-domain or time–frequency features can be useful when arc faults or switching-related abnormalities are being investigated. Features should have a physical or diagnostic reason for inclusion. Normalization is also important because current, voltage, temperature and power have different numerical scales.

6.3 Class imbalance and rare faults

In real plants, healthy operating data are collected continuously, while confirmed fault events may be rare. As a result, a dataset may contain thousands of healthy examples but only a small number of examples for a serious fault. A classifier can appear accurate by predicting the majority class most of the time, even though it fails to recognize the rare fault. Possible responses include collecting more fault examples, using class-weighted learning, carefully applying resampling methods and reporting class-wise performance. Artificially generated or simulated examples can help, but they should not be treated as a complete substitute for real field observations.

6.4 Training, testing and validation

Training data are used to fit the model, validation data help select model settings, and test data provide a final estimate of performance. The split should avoid placing nearly identical measurements from the same operating event in both training and test sets. A stronger test is to train on one period or plant and evaluate on a different period or plant. This better represents the situation in which a model is installed at a site it has not seen before. Chen et al. highlight the fragmentation and inconsistency of PV reliability datasets, which makes fair comparison between methods difficult.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

6.5 Performance measures

Accuracy is the fraction of all predictions that are correct. Precision indicates how many predicted fault cases are actually faults, while recall indicates how many real faults are detected. The F1-score combines precision and recall. In protection-related work, false-alarm rate and missed-fault rate should also be reported because the consequences are different: too many false alarms can lead to unnecessary inspections and loss of operator confidence, while a missed serious fault can threaten equipment or safety. A confusion matrix helps show which fault classes are being confused.

Reported accuracy values from different papers should not be compared directly unless the datasets, fault classes, operating conditions and test procedures are similar. A model evaluated on a small laboratory dataset may report excellent results but still perform poorly on a different plant. This is one reason why field validation and transparent reporting are essential.

VII. MATHEMATICAL BASIS FOR PV DIAGNOSIS

Only a few basic relationships are needed to understand the main measurements used in AI-assisted PV protection. The electrical output power is:

$$P = V \times I$$

The single-diode model is commonly used to describe the current–voltage behaviour of a practical PV cell or module:

$$I = I_{ph} - I_0[\exp((V + I \cdot R_s)/(n \cdot V_t)) - 1] - (V + I \cdot R_s)/R_{sh}$$

Here, I_{ph} is photocurrent, I_0 is diode saturation current, R_s is series resistance, R_{sh} is shunt resistance, n is the diode ideality factor and V_t is thermal voltage. The equation is nonlinear because current appears on both sides. In practical diagnosis, the model can estimate the expected electrical response for measured irradiance and temperature.

A simple residual can be calculated by comparing the measured output with the model prediction:

$$r(t) = y_{measured}(t) - y_{predicted}(t)$$

A residual close to zero indicates that the measured value is near the expected value. A persistent residual may indicate a fault, degradation or an error in the model or sensor. Therefore, a residual should be interpreted together with environmental data and other measurements. It should not be used as an automatic trip command without a validated protection design.

For a classification model, the predicted class is obtained from the input feature vector. During training, model parameters are adjusted to reduce a loss function. For a multi-class problem, cross-entropy loss is commonly used. The important engineering point is that mathematical optimization only reduces error on the training objective; it does not guarantee safe operation in every real-world condition.

VIII. PROPOSED LAYERED AI-ASSISTED PROTECTION ARCHITECTURE

A practical PV protection system should separate safety-critical protection from data-driven diagnosis. The primary protection layer contains established devices and inverter functions that operate according to applicable electrical standards and the plant's protection coordination. This layer must remain functional even if the AI computer, communication network or sensor system fails.

Above the primary layer, a measurement layer collects voltage, current, power, irradiance, temperature, inverter status and, where available, image data. A data-quality block checks the readings for missing samples, implausible values, sensor failure and timing mismatch. An anomaly-detection block then identifies behaviour that differs from the expected operating range. If an anomaly is detected, a classification model estimates the likely fault type. A localization stage uses the measurement location, string arrangement or image coordinates to identify the affected area.

The operator interface should not display only a fault label. It should also provide the measurements that triggered the warning, the time of occurrence, the confidence or uncertainty of the model, the likely affected equipment and a recommended inspection step. When confidence is low or measurements disagree, the system should request further inspection rather than presenting a guess as a confirmed fault. Maintenance records can later be used to check whether the diagnosis was correct and to improve the model.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

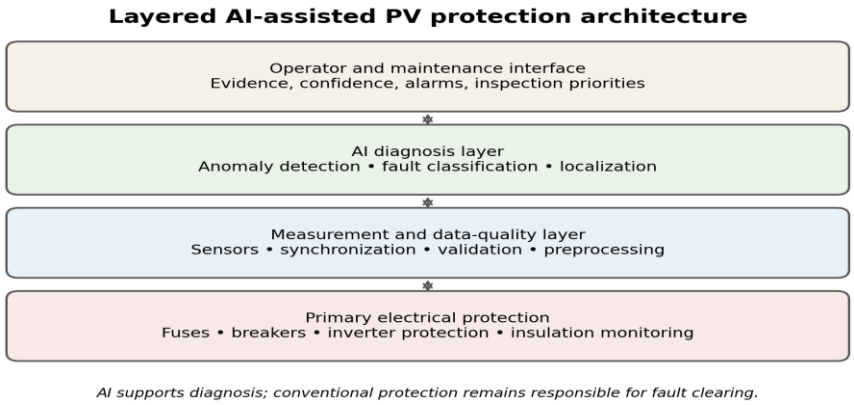


Figure 3. Layered AI-assisted protection architecture. Conventional protection remains responsible for safety-critical fault clearing; AI supports monitoring and diagnosis.

8.1 Edge and cloud implementation

Edge computing means processing data close to the equipment, for example in an inverter controller, combiner monitoring unit or local industrial computer. It can reduce communication delay and allow basic diagnosis to continue when internet connectivity is poor. Cloud or central-server processing can be used for long-term trend analysis, model training and comparison across multiple plants. A combined edge–cloud design is often practical: time-sensitive checks are performed locally, while larger datasets and images are analysed centrally.

8.2 Cybersecurity and system reliability

A connected PV plant may exchange data with remote monitoring platforms, maintenance systems and control networks. Communication links and devices therefore need suitable access control, secure updates, logging and network separation. Sensor failure must also be considered. A model may produce a confident but incorrect result if a sensor becomes stuck or drifts slowly. Monitoring the health of sensors and the AI model is consequently part of the protection design, not an optional extra.

IX. DISCUSSION OF THE REVIEWED LITERATURE

The reviewed publications show a progression from electrical thresholds and physical-model residuals toward machine-learning classifiers and deep-learning methods. Hong and Pula provide a broad classification of PV fault-detection approaches, while Gaviria et al. place PV diagnosis within the wider use of machine learning for monitoring and operation. Islam et al. organize AI-based fault identification research across classical ML and deep-learning approaches. These reviews support the view that the field is not based on one single diagnostic technique; instead, researchers choose methods according to the available data and the fault under study.

A second trend is the increased use of images. Thakfan and Bin Salamah discuss thermal imaging and I–V curve analysis as important routes for PV fault diagnosis. Balachandran et al. examine imaging methods and the preprocessing needed before visual diagnosis. Oliveira et al. show how aerial infrared thermography and orthomosaicking can be used to inspect utility-scale PV generators. These methods are valuable because they add spatial information, but their results depend on inspection conditions and image quality.

A third trend is the combination of data types and methods. Afrasiabi et al. investigate multimodal deep learning for PV fault diagnosis, illustrating the potential of combining complementary evidence. Hybrid approaches can also join signal processing with machine learning, or use physical-model residuals as inputs to a classifier. Such designs can be more informative than relying on a single measurement, although they introduce additional requirements for synchronization, calibration and data management.

Recent work places more emphasis on deployment and the quality of the evidence used to train models. Chen et al. examine open datasets for PV reliability and identify the importance of dataset availability and consistency. Meena and Gawre review a large body of PV fault-analysis studies and highlight standardization, benchmarking and field validation



| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

as continuing needs. These findings are important because a high score on one dataset does not establish that a model will remain reliable at another plant, in another climate or with a different module technology.

Overall, the literature suggests that model selection should follow the physical problem. Electrical measurements are appropriate for continuous monitoring; I–V curves provide detailed electrical signatures; thermal and visible images help locate physical defects; and time-series models can identify developing behaviour. Combining these sources may improve diagnosis, but only if the data are properly aligned and the system is tested under realistic conditions.

X. PRACTICAL CHALLENGES AND FUTURE DIRECTIONS

10.1 Limited field data and domain shift

Many models are developed using simulated or laboratory data because faults are easier to reproduce under controlled conditions. Actual plants experience changing weather, dust, ageing, different installation practices and equipment from different manufacturers. These differences create domain shift: the data seen during deployment do not fully match the training data. Models should therefore be tested on independent plants and across seasons. Periodic checks are needed to identify when performance begins to deteriorate.

10.2 Explainability and uncertainty

Engineers need to understand why a system raised an alarm. Feature importance, rule-based explanations, saliency maps and example-based explanations can help show which measurements or image regions influenced a prediction. However, an explanation is not proof that the diagnosis is correct. The system should also report uncertainty and provide a way to request human inspection when the evidence is weak.

10.3 Real-time processing and low-cost deployment

Large deep-learning models may require more memory and processing power than a small PV monitoring device can provide. A practical solution may use a smaller model, feature selection, quantization or edge hardware designed for industrial use. The processing delay must be matched to the task: fast electrical protection cannot wait for a cloud-based image model, whereas maintenance planning can use slower analysis.

10.4 Research priorities

Future work should prioritize shared datasets that include fault labels, operating conditions, sensor details and documented test procedures. Researchers should report more than overall accuracy and should test models on data from different plants or time periods. Further opportunities include physics-informed learning, uncertainty-aware diagnosis, multimodal fusion, digital-twin support, federated learning and automated model-health monitoring. These methods should be evaluated for engineering usefulness, not only algorithmic novelty.

XI. CONCLUSION

AI and machine learning can improve the monitoring and diagnosis of solar PV plants by identifying relationships among electrical measurements, environmental conditions, time-series behaviour and inspection images. Classical ML methods remain useful for structured sensor data, CNNs support image-based inspection, recurrent and LSTM models can analyse temporal patterns, and hybrid systems can combine several sources of evidence. The reviewed studies also show that no model is universally best; the choice depends on the fault, data availability, operating environment and computing resources.

The main challenge is not simply to build a model with high accuracy. The model must work with realistic field data, handle rare faults, avoid excessive false alarms, provide useful information to engineers and remain dependable as the plant ages. For this reason, conventional electrical protection should remain the final safety barrier. AI should be used as an additional layer for early warning, classification, localization and maintenance support. With better datasets, transparent validation and careful integration, AI can make PV protection more informative without weakening established electrical safety practices.

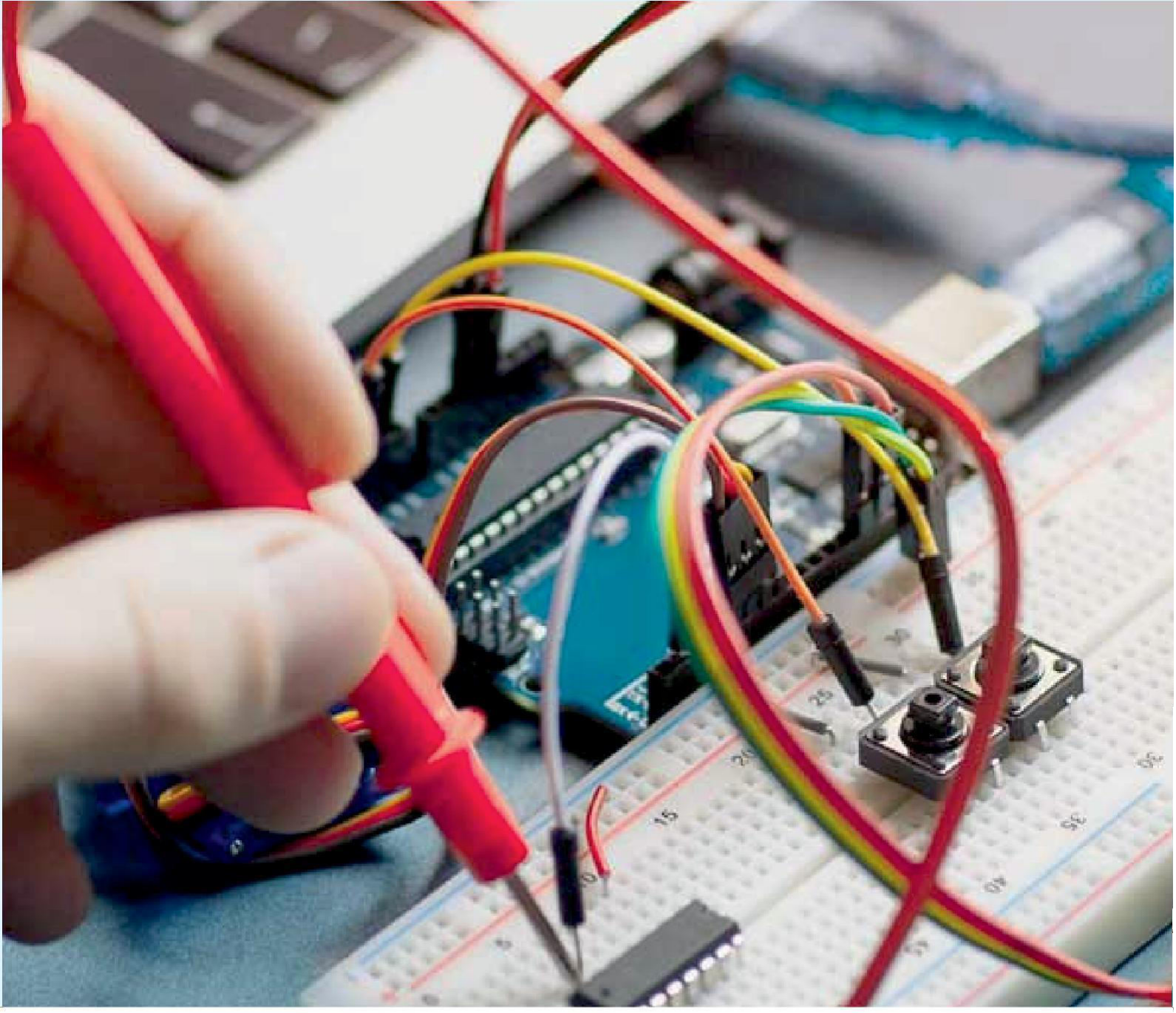
| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 9.867 | A Monthly Peer Reviewed & Refereed Journal |

|| Volume 15, Issue 11, November 2026 ||

| DOI:10.15662/IJAREEIE.2026.1511016 |

REFERENCES

1. J. F. Gaviria, G. Narváez, C. Guillen, L. F. Giraldo, and M. Bressan, “Machine learning in photovoltaic systems: A review,” Renewable Energy, vol. 196, pp. 298–318, 2022, doi: 10.1016/j.renene.2022.06.105.
2. M. Islam et al., “Artificial Intelligence in Photovoltaic Fault Identification and Diagnosis: A Systematic Review,” Energies, vol. 16, no. 21, Art. no. 7417, 2023, doi: 10.3390/en16217417.
3. Y.-Y. Hong and R. A. Pula, “Methods of photovoltaic fault detection and classification: A review,” Energy Reports, vol. 8, pp. 5898–5929, 2022, doi: 10.1016/j.egyr.2022.04.043.
4. W. Tang, Q. Yang, Z. Dai, and W. Yan, “Module defect detection and diagnosis for intelligent maintenance of solar photovoltaic plants: Techniques, systems and perspectives,” Energy, vol. 297, Art. no. 131222, 2024, doi: 10.1016/j.energy.2024.131222.
5. G. B. Balachandran, M. Devisridhivyadharshini, M. E. Ramachandran, and R. Santhiya, “Comparative investigation of imaging techniques, pre-processing and visual fault diagnosis using artificial intelligence models for solar photovoltaic system – A comprehensive review,” Measurement, vol. 232, Art. no. 114683, 2024, doi: 10.1016/j.measurement.2024.114683.
6. X. Chen et al., “Open data sets for assessing photovoltaic system reliability,” Applied Energy, vol. 395, Art. no. 126132, 2025, doi: 10.1016/j.apenergy.2025.126132.
7. A. Appiah, X. Zhang, B. B. K. Ayawli, and F. Kyeremeh, “Long Short-Term Memory Networks Based Automatic Feature Extraction for Photovoltaic Array Fault Diagnosis,” IEEE Access, vol. 7, pp. 30089–30101, 2019, doi: 10.1109/ACCESS.2019.2902949.
8. IEC, IEC 62109 series, Safety of Power Converters for Use in Photovoltaic Power Systems, International Electrotechnical Commission.
9. IEC, IEC 62446-1, Photovoltaic Systems—Requirements for Testing, Documentation and Maintenance, International Electrotechnical Commission.
10. Y. Liu and Y. Wu, “Fault diagnosis of photovoltaic modules: A review,” Solar Energy, vol. 293, Art. no. 113489, 2025, doi: 10.1016/j.solener.2025.113489.
11. IEC, IEC 61724-1, Photovoltaic System Performance—Part 1: Monitoring, International Electrotechnical Commission.
12. B. Afrasiabi, S. Allahmoradi, M. Afrasiabi, X. Liang, C. Y. Chung, and J. Aghaei, “A Robust Multi-Modal Deep Learning-Based Fault Diagnosis Method for PV Systems,” IEEE Power and Energy Technology Systems Journal, vol. 11, pp. 583–594, 2024, doi: 10.1109/OAJPE.2024.3497880.
13. M. K. Alam, F. Khan, J. Johnson, and J. Flicker, “A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques,” IEEE Journal of Photovoltaics, vol. 5, no. 3, pp. 982–997, 2015, doi: 10.1109/JPHOTOV.2015.2397599.
14. A. Thakfan and Y. Bin Salamah, “Artificial-Intelligence-Based Detection of Defects and Faults in Photovoltaic Systems: A Survey,” Energies, vol. 17, no. 19, Art. no. 4807, 2024, doi: 10.3390/en17194807.



INNO SPACE
SJIF Scientific Journal Impact Factor

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

9940 572 462 6381 907 438 ijareeie@gmail.com

Scan to save the contact details