Machine Learning for PV System Operational Fault Analysis: Literature Review*

Tarikua Mekashaw Zenebe[1,2], Ole-Morten Midtgård[1,3], Steve Völler[1,4], and Ümit Cali[1,5]

[1] Norwegian University of Science and Technology, Trondheim, Norway
[2] tarikua.zenebe@ntnu.no
[3] ole-morten.midtgard@ntnu.no
[4] steve.voller@ntnu.no
[5] umit.cali@ntnu.no

Abstract. This review paper aims to discover the research gap and assess the feasibility of a holistic approach for photovoltaic (PV) system operational fault analysis using machine learning (ML) methods. The analysis includes the detection and diagnosis of operational faults in a PV system. Even if standard protective devices are installed in PV systems, they fail to clear various faults because of low current during low mismatch levels, high impedance fault, low irradiance, etc. This failure will increase the energy loss and endanger the PV system's reliability, stability, and security. As a result of the ML method's ability to handle a non-linear relationship, distinguishing features with similar signatures, and their online application, they are getting attractive in recent years for fault detection and diagnosis (FDD) in PV systems. In this paper, a review of literature on ML-based PV system FDD methods is provided. It is found that considering their simplicity and performance accuracy, Artificial Neural networks such as Multi-layer Perceptron are the most promising approach in finding a central PV system FDD. Besides, the review paper has identified main implementation challenges and provides recommendations for future work.

Keywords: Ensemble learning · Fault detection and diagnosis · Machine learning · PV system fault · Transfer learning.

1 Introduction

Owning to the various advantages PV system can provide, the global market for PV has been increasing sharply. According to [1], the cumulative globally installed capacity in 2019 increased to about 627 GW. Assuming a medium scenario where cases like COVID-19 pandemic considered, [14] estimated the total global installed PV generation capacity to exceed 1.2 TW by 2022. In addition, the price of electricity from a PV system is constantly decreasing [20]. This all shows a promise for further increase in the PV market in the coming years.

* Supported by NTNU

Nomenclature

AC	Alternative Current
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DA	Discriminant Analysis
DC	Direct Current
DL	Deep Learning
DT	Decision Tree
DWT	Discrete Wavelet Transform
EL	Ensemble Learning
FDD	Fault Detection and Diagnosis
G	Irradiance at Array
$GCPVS$	Grid Connected PV System
$GFDI$	Ground Fault Detection and Interruption
I	Current
I_{MPP}	Current at Maximum Power Point
$IGBT$	Insulated Gate Bipolar Transistor
$KELM$	Kernel Based Extreme Learning Machine
$LSTM$	Long Short Term Memory
$MIMO$	Multiple Input Multiple Output
ML	Machine Learning
MLP	Multi-layer Perceptron
$MPPT$	Maximum Power Point Tracker
$OCPD$	Over Current Protection Devices
RBF	Radial Basis Function
RF	Random Forest
$SAPVS$	Stand Alone PV System
$SCADA$	Supervisory Control and Data Acquisition
SOC	State of Charge
STC	Standard Test Condition
SVM	Support Vector Machine
T	Module Temperature
TL	Transfer Learning
V	Voltage
V_{MPP}	Voltage at Maximum Power Point

With the increase in PV system deployment for electricity production, ensuring the system's reliability, stability, and safety is crucial. However, despite the advancement in technology and implementation of standards such as the National Electric Code (NEC) article 690 [33], still, faults are problems for the efficient and effective operation of a PV system. Because there are various conditions where the protective devices fail to clear fault on time. For example, according to NEC, the fuse rating should be greater than 2.1 times the short-circuit current at standard test condition (STC) in PV system [31]. However, if line to line (LL) fault occurred at low mismatch and high impedance level, the fuse will not be able to clear the fault as the current will not be enough to blow the fuse [33]. In addition, due to the blocking diodes, which prevent the string from back-feed current, the protective devices may fail to interrupt the fault current even under STC. Moreover, these diodes may fail and lead to serious damage [12].

As mentioned in [27], the annual energy loss due to various faults might go up to 18.9%. This reported power loss is very significant as the efficiency of a typical PV cell range between 15-21%. Unless the faults cleared on time, they might also cause additional damages to other property in case of fire. Therefore, detecting and clearing the fault on time is an indispensable solution to mitigate these losses while ensuring the reliability and security of a PV system.

Up to now, several techniques have been created for FDD in a PV system. However, the demand for techniques which is simple and cheap, can handle non-linear nature of the PV modules, can be remotely applied and can differentiate features with similar signature are the primary motivation to move to data-driven methods like that of machine learning (ML) [27] for many researchers in recent years.

If ML is used to analyze the fault in a PV system, as much as possible, there should be a holistic method that is used to detect and diagnosis at least all the most frequent and dangerous faults. Nevertheless, most of the literature focused on PV array fault. Besides, only one paper tried to implement the ML method in a programmable logic device based on the author's knowledge. Thus, this paper aim at answering the reason behind. To the best of the author's knowledge, this review paper is the first to review literature, keeping in mind the feasibility of a holistic fault analysis approach for a PV system specifically for standalone PV system (SAPVS) as well as categorizing and analyzing FDD into methods based on ML and deep learning, ensemble learning, and transfer learning.

The paper is organized as follows: After providing a summary of review papers, the first part of section 2 gives detailed information about various faults commonly occurring in PV system components. Then, the second part of section 2 provides a comprehensive literature review on PV FDD. Whereas, section 3 presents and discuss all the findings. Finally, the paper concludes by summarizing the main findings and providing some recommendations.

2 Literature Review

Pillai et al. [33] provided a comprehensive literature review on PV faults and advanced detection techniques. The paper tried to review literature, including all PV faults. However, most of the discussion focused on PV array faults. Mellit et al. [31] presented very detailed information about PV faults, including FDD methods. But similar to [33], the main focus of the paper was on PV array faults. [40] reviewed papers on the role of artificial intelligence on modeling, sizing, control, fault diagnosis, and output estimation of PV systems. Whereas, Li et al. [24] reviewed recent work specifically applying Artificial Neural Network (ANN) and hybrid ANN for FDD based on the fault they analyzed, the type and amount of data they used, their model's configuration, and its performance. Besides, they highlighted the major challenges and prospects of the methods. [16] is among the papers dedicated to explaining the PV system faults in a wider spectrum. Fault detection methods on grid-connected PV system (GCPVS) were studied comprehensively in [26]. Contrary to most of the review papers, the current paper focuses only on the advanced data-driven approach that of ML.

This section will discuss different PV system faults classified based on components and the various ML methods used in PV system FDD.

2.1 PV System Fault

To design an efficient and effective fault detection and diagnosis method, it is necessary to know about the character of each fault, including their protection challenges [33].

SAPVS comprises of PV array, inverter, battery, charge controller, MPPT, connection wires, and other additional protection and safety devices.

PV Array Fault Some of the PV array faults are discussed below.

Open Circuit Fault (OC) OC fault is an intentional disconnection of a closed-loop that results in interruption of current flow due to breakage of the cable that connects two strings, any object falling on panels, loose connection between two points, or an accidental disconnection at a current-carrying conductor [4]. In addition, broken cells, physical breakdown of cable joints, loose connections, and aged power cables near terminal may lead to OC fault [27]. Due to the presence of a bypass diode, current flow will be kept even if an OC fault occurs. In addition, it results in a substantial power loss due to the reduction in voltage in a string [27].

Line to Line Fault (LL) LL fault is an unintentional connection between lines with different potential difference [4, 31, 41] due to cable insulation failures, mechanical damage, water ingress, D-junction box corrosion, and hot spots caused by the back-sheet failures [31]. LL fault could lead to serious problems like fire hazards in addition to degrading PV arrays lifetime. LL fault is very hard to identify by the conventional protection devices such as Over Current Protection Devices (OCPD) that is mainly: 1) as a result of the decrease in current in cases of LL fault during high impedance and low mismatch level [31, 33], 2) due to the presence of a blocking diode as it blocks back-fed current [12, 33], 3) as the presence of MPPT decreases the current to optimize the power output and difficult to distinguish it from normal cases [12, 33], 4) its similarity with ground fault [33], 5) as a result of low current at low irradiance values [33].

Ground Fault (GF) GF occurs when a current-carrying wire/cell/module connected with a ground accidentally. It can be detected by Ground Fault Detection and Interruption (GFDI) and Ground Fault Protection Devices in a normal scenario. However, during high impedance cases, detection is challenging as the current will be low. In addition, there are scenarios where it looks like SC fault [33]. Thus, this fault also needs an efficient method to detect and distinguish it from other faults.

Arc Fault (AF) AF is a fault where current flows in the air or dielectric outside the conductor due to loose connection. It could be a series arc in case of a connection between modules or a parallel arc in case of a closely placed conductor at different potential differences [4, 33, 31]. On the contrary to other faults, arc fault has little effect on the I-V or I-P characteristic of PV arrays. Nevertheless, it leads to a severe distortion in the output current and voltage waveform [33]. Arc Fault Circuit Interrupters and Arc Fault Detectors are recommended for clearing this fault. However, multiple of them have to be installed to clear the fault correctly. Moreover, when they are installed at the inverter side, they fail to protect the fault as attenuated arc signals reach them. Beside detecting arc fault, identifying which arc fault is occurring is important as the measure taken for one will increase the impact of the other [33].

Partial Shading (PS) In addition to decreasing and resulting in continuous fluctuating PV output power [4, 6], PS facilitates the degradation of PV arrays [33].

Even it can lead to destruction due to fire hazards as a result of cell/module temperature increase due to the dissipated energy [27]. Furthermore, as it results in multiple peaks in I-V characteristic curve, identifying the maximum power point by MPPT will be challenging [33]. Besides, unless a time factor is used, it is hard to differentiate it from OC fault as their effect on power output characteristic has similarity [27]. Furthermore, to mitigate the problem, bypass diodes are installed at each module, but this will increase the installation cost [33].

Others In addition to the above main PV array faults others may include degradation faults [18, 34], hot spot fault [39], fault in bypass diode which could be OC or SC fault [33], and blocking diodes faults [30].

Solar Battery Fault The battery takes around 43% of the life cycle cost of SAPVS [37]. As a result, it shall get attention, and a good working condition shall be provided. The main faults that could happen in this PV component includes external short-circuit fault [36], degradation fault [35], internal fault which could be GF and SC fault [32], overcharging (over-voltage), undercharging (under-voltage) and open circuit (total voltage to zero) [39]. The impact of those faults in a battery may range from decreasing its performance, shorten its lifetime, and increased maintenance cost to fire hazard explosion [32]. The lack of guidelines on how to select fuse and circuit breaker is mentioned in [32] as one of the main challenges in detecting internal faults. Moreover, the gradual change of current and voltage of a battery makes detecting faults on time extremely difficult.

Inverter Fault Inverter faults may include the OC of switches, SC of switches, filter failure, and gating failure [31]. For instance, the gate failure could be an incipient fault of the Insulated Gate Bipolar Transistor (IGBT). IGBT is the most critical component in an inverter. It is also one of the main reasons for the failure of inverters. So if the incipient faults of the IGBT can be identified, the reliability of the PV system can be enhanced. Nevertheless, a procedure is needed to generate this fault to train and validate ML algorithms. Thus, Ismail et al. [21] provided the way to generate this fault.

MPPT Fault An MPPT control system comprises various sensors to get irradiance, temperature, current, and voltage measurements and an optimization algorithm to search the maximum power point to operate the PV array and boost the PV system yield. Thus, any error in any part of the MPPT will lead to a wrong operating power point, which significantly decreases the PV system's output power. Sensor failure and lack of an efficient and effective MPPT algorithm are the most common fault in MPPT [27, 29].

2.2 PV System Fault Detection and Diagnosis Methods (FDD)

In this paper, fault detection indicates the process of identifying a fault occurrence, while fault diagnosis comprises the process of finding the type of fault and localizing the occurrence. This section closely look at the literature on ML application for FDD in PV systems by classifying them into methods based on 1) machine learning and deep learning, 2) ensemble learning, and 3) transfer learning.

Methods based on Machine Learning and Deep Learning Similar findings and a comprehensive explanation about ML and deep learning with respective of PV application can be found in the book chapter in [30]. PV system fault has been detected and diagnosed using supervised machine learning such as Support Vector Machine (SVM), Naive Bayes (NB), k-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), Discriminant Analysis (DA), Radial Basis Function (RBF), and deep learning like Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN).

Among other works [38] used SVM with a higher classification accuracy using a climate corrected performance ratio. However, the paper fails to mention which fault has been analyzed specifically besides it is a fault or normal operation. Similarly Dong et al. [11] proposed an FDD method based on SVM while using available SCADA (Supervisory Control and Data Acquisition) data. In addition, as an input they used an index called anomaly detection index. However, this paper focuses only on PV string faults.

In order to analyze multiple faults Hajji et al. [17] tried to include a feature extraction and selection stage using principal component analysis. They have tested various classifiers like KNN, RF, DA, NB, DT, and SVM to classify fault in GCPVS. To evaluate the performance of the classifier, they have used additional metrics. They all have achieved an accuracy greater than 96%. In addition, the execution time of each classifier was evaluated. Relative to other papers, they have included inverters, MPPT, and DC-DC converter in addition to PV array. Nevertheless, battery fault is not analyzed.

In [23], the authors used a comparison between model and real system output to identify fault occurrence for GCPVS. First, they have tested various linear and nonlinear models of PV system. Then, they used ML techniques like KNN, DT, SVM, and MLP to identify faults such as SC, OC, degradation, and shadowing. MLP is found to be a suitable and more accurate ML algorithm. Basnet et al.[5] has used MLP, to detect and classify LL and GF in GCPVS as well. They could achieve 100% training accuracy. As input parameters, voltage, current, irradiance, average temperature of each module, and weather conditions were utilized. Despite getting a good accuracy, both papers [23, 5] application's is limited to PV array fault.

[10] is among the few papers which evaluate the ML algorithm's performance based on both accuracy and execution time. Faults like module SC, MPPT fault, OC, PS, and degradation have been detected and classified using five ML techniques: kNN, DT, SVM, and ANN. ANN resulted in higher accuracy (99.65%) even though it took longer computational time. Nevertheless, for generalization, ANN should be tested by incorporating other faults.

A cascaded Probabilistic neural network (PNN), due to its robustness to noise, has been used in [15] to detect and classify a different number of module SC and string OC faults. In addition, the result compared with a feed-forward back-propagation ANN with both noisy and noiseless data. As input features, they have utilized temperature, tilted irradiance, current (I_{MPP}), and voltage (V_{MPP}) at MPP. The training data set is generated from a validated one diode PV system model. Despite the effort made to bring a robust method, the literature focused on faults on a PV array and DC side of the GCPVS. With a similar focus on GCPVS, DT was used in [25] for detecting PS, inverter, and bypass diode failure. The authors has also used other methods for detecting and classifying these faults. They achieved an average classification accuracy of 98.7%.

A kernel-based extreme learning machine (KELM) was used in [7] to detect and classify degradation fault, OC, SC, and PS. Features that enable the identification of faults are extracted after examining their impact on I-V characteristics. Besides, the PV model was validated before using it to simulate the fault to generate a training and test data set. In addition to the simulation data, a real laboratory PV array data set has been used. In general, even if a very efficient and accurate method is devised here, the determination of I-V characteristics of the PV array in an online scenario might be problematic.

[35] is one of the few papers which has focused specifically on faults in SAPVS. They proposed a fault diagnosis method using MLP feed-forward neural network to detect and classify faults. The faults include SC of PV module, OC of PV module, and external SC of a battery where the fuse fails to clear them in low irradiance condition. Though most papers entirely focus on PV array fault, this paper included the battery and load fault. Only electrical measurements like current and voltage are used for validation using experimental data from an existing PV system in Algeria. They have achieved 96% test and 97.8% training accuracy. Nevertheless, to consider as a valid method for complete SAPVS fault analysis, it shall be verified including the missing other faults like inverters fault, MPPT fault, and others.

Chine et al. [8] used a combination of threshold method and ML to detect and classify eight faults, including SC, bypass diode fault, OC, connection fault, shadow effect and etc. From ML, MLP and RBF have been compared. Extracted attributes like the current, voltage, and peaks from I-V characteristics were used as input. In addition to showing the method's feasibility, this is the only paper encountered that shows a prototype by implementing the ML in a Field Programmable Gate Array (FPGA). However, they used simulation data for training and testing the models. The other drawback of this paper is that they also applied threshold method which can be very much dependent on system parameters and the accuracy of the threshold limits.

The authors in [12] focused on the detection of LL fault in PV system under high impedance fault and low mismatch condition, which is one of the cases where protection devices fail to clear a fault. The authors used an SVM classifier that is resistant to model error and computational efficiency, based on the features extracted by analyzing I-V characteristics of a PV array. As the methods were validated for LL fault, further analysis is needed to apply it for multiple input and output (MIMO) cases in SAPVS. Moreover, validation is needed with real PV system data.

Instead of threshold method, Ahmad et al. [2] has used a combination of transformation for feature extraction and ML algorithm for detecting and classifying PS condition as a modular fault, DC-DC converter switch SC, inverter switch OC, inverter switch SC with LCL filter failure and gating circuit failure. They used discrete wavelet transform (DWT) due to its less computational time and complexity, as well as it enables us to work both in the time and frequency domain. Whereas the ML algorithm is MLPNN. The data set was obtained from a simulated a PV system. They can achieve an accuracy greater than 99%. However, the battery and MPPT fault is missing. Furthermore, other faults in PV array such as LL has to be checked.

[3] used a hybrid features-based support vector machine (SVM) model in order to detect and classify hot spot fault in PV array using infrared thermography as an input image for the model. The model could detect and classify with 96.8% and 92% training and testing accuracy, respectively. This paper is dedicated to one fault only. For small-scale SAPVS using this individual method will not be cost-effective.

Improper operation is One of the reasons for the short lifetime of solar batteries. In addition to the available energy from solar or demanded load, to decide whether the battery has to be charged or discharged, knowing the battery capacity accurately is a determining factor. There are various statistical estimation techniques, but recently SOC (state of charge) estimation using ML is getting attractive as it exhibits non-linear input-output characteristics. [9] presented an ML-based SOC estimation method for the most common solar battery, which is a lead-acid battery. The proposed methods are based on a feed-forward neural network, a recurrent neural network, and an adaptive neuro-fuzzy inference system. As an input feature for the model, voltage and current data were used. The findings of this paper could be used for further studying FDD methods in PV batteries. However, the paper did not mention how the training SOC data is obtained.

In [39] internal resistance effect and overcharging problem in lead-acid battery in a PV system was detected using solar radiation data estimated from satellite image analysis. The paper showed the impact of overcharging and internal resistance fault on the battery voltage and SOC. Even if the paper used the ML for estimating the solar irradiation, from the finding, there is an indication for using battery voltage to detect and classify faults in a battery. In another study in [19], the author has used a long short-term memory (LSTM) recurrent neural network for state prediction and fault prognosis for battery in an electric vehicle. This approach could also be used for solar batteries from the knowledge domain, and its finding is significantly important though its feasibility has to be checked in a solar battery.

One of the challenges in using ML, especially in analyzing inverter fault in a PV system, is the lack of methods that guide us in generating the faults in simulation as the faults do not occur frequently, but they are responsible for the majority of inverter failures. Thus, Ismail et al. [21] used a feed-forward back propagation neural network to detect SC incipient faults by first modeling a way to generate this fault for GCPVS. For using it in SAPVS, others fault has to be incorporated, and the method has to be verified for its performance for other PV system components fault.

In addition to supervised learning, unsupervised ML methods has been used in PV system FDD. In [37] an internal fault detection for solar battery using unsupervised ML algorithm based on anomaly detection method has been proposed. The intuition for using unsupervised learning is whenever it is difficult to obtain a labeled data set, which is the case in solar battery fault analysis for using ML. The internal faults investigated are SC and GF. The data set was generated using simulation of SAPVS using irradiance and temperature data from Algeria. They have used readily available current and voltage data set. As the work is only for internal fault, it is important to incorporate it/hybridize it with other methods to identify other faults in SAPVS.

Methods based on ensemble learning (EL) [13], similarly to [12] the model is based on I-V characteristics and focus on LL fault at different mismatch and impedance level. However, here they used probabilistic ensemble learning model comprising of SVM, NB, and KNN. For decision, the average of all the results of the algorithm was used. They could achieve an average of 99% and 99.5% for detecting and classifying LL fault. Moreover, they have evaluated the model with simulation and experimental data set. In [22] EL method with DT, RF, DA, etc., was used to detect PS and SC fault, but the focus is still PV array. They have used electrical parameters as input features.

Methods based on Transfer Learning (TL) In order to detect and classify PV system faults, in [4], the concept of transfer learning has been employed by using a pre-trained AlexNetCNN for feature extraction and classification to minimize the impact of low data set in the model performance. They also proposed a deep 2-D CNN to extract 2-D scalograms generated from a PV system. The authors analyzed faults like PS, LL, OC, high-impedance series /arc fault, and faults in PS with the presence of MPPT. A detecting accuracy of 73.53% and a classification accuracy of 70.45% were achieved. They have noticed the decreasing of performance as the number of class increase. They have used deep learning in-depth and also made a comparison with classical ML models. They handled MIMO data. Though it needs to be verified, the methods they followed seem promising for SAPVS fault analysis. Even though it is only for inverter's fault in GCPVS, in [28], TL was also used to detect faults like SC, OC using ResNet with an accuracy greater than 97%.

3 Result and Discussion

As we can see from Table 1 the majority of the papers, greater than 80%, has analyzed the fault in PV array. However, faults in an inverter and a battery have also been investigated. Relatively, faults in GCPVS have got special attention than SAPVS. SC, OC, and PS in PV array are the most investigated type of fault using ML methods though faults like GF, AF, and LL, are the most severe.

Among other ML methods, SVM and MLP, in general, have been used extensively to detect and classify faults in a PV system. For evaluating the models, accuracy and confusion metrics are the most employed performance indices. However, some have utilized their own metrics and execution time. Due to ML's random nature, it is very important to report performance after conducting a reasonable number of model execution though it takes time.

Looking at the data source where the experimental PV system was installed, Algeria took the lead, China and Korea take second place. It is also very important to analyze the performance of the ML methods under different climatic and geographical conditions before utilizing them. This is because the challenge for the PV array and the battery is different depending on the geographical location. For instance, while snow is a big problem in the polar region, dust, soiling, and higher operating temperature are huge problems in the equatorial region.

Most of the papers depend on the input features which has been generated from a simulated PV system. Whereas only a few have included experimental data. This is because of the difficultly of setting up a PV system only for collecting data. Furthermore, when an available PV system exists as the environmental condition can not be controlled, it is tedious and time-consuming to generate a data set that will enable the model to acquire a generalization capacity. Irradiance, temperature, and major points from I-V characteristics are the most utilized input features in case of a fault in a PV array. In comparison, current and voltage data are used in case of a fault in a battery, inverter, MPPT, and others. Electrical and meteorological data are mostly used in ML, whereas image data are the most common input features for deep learning algorithms such as CNN. However, recently as 1-D can be transformed to 2-D data, electrical and meteorological data are also employed for deep learning algorithms in general.

Faults like arc fault that does not reflect its effect on I-V characteristics of PV arrays, a method that includes the analysis of signal waveform (some kind of transformation, for example, wavelet) which could show signal distortion effect, might be an appropriate method to capture most of the faults in a PV system. Moreover, in most papers, prepossessing of data like normalization has resulted in better accuracy. Nevertheless, whenever this is not possible deep learning models are efficient due to their capacity in extracting features automatically.

Even if major progress has been seen in the research area in using the ML method for FDD in a PV system, only one paper has implemented the ML method in prototype based on the literature review. Furthermore, so far, this method is not commercialized. Thus, the authors have identified the following main challenges.

Table 1: Summary of reviewed literature on PV system FDD using ML methods

Reference	Year	Methods		Data Set			Fault identified		Comment	
		Type	Algorithm	Input Data	Simulated	Real system	Components	Faults	Accuracy	Type of PV system
[28]	2021	TL	ResNet-50 CNN	I, V	✓	×	Inverter	Norma, PS and SC	>97%	GCPVS
[22]	2021	EL	DT, RF, DA ...	String I, V, P, T, G	×	✓	PV array	Norma, OC and SC	>97%	GCPVS
[18]	2021	ML	SVR, GPR	T, G, P_{peak}	✓	✓	PV array	OC, SC, LL, PS, Degradation	-	GCPVS
[38]	2020	ML	SVM	-	-	-	-	Normal and Faulty	-	-
[17]	2020	ML	KNN, RF, DA, NB, DT, SVM	-	-	-	PV array	OC, PS, Sensor	>96%	GCPVS
							Inverter	IGBT, Grid connection		
[23]	2020	ML	KNN, DT, SVM, ANN(MLP)	-	✓	✓	PV array	OC, PS, SC, Degradation	-	GCPVS
[5]	2020	ML	KNN, DT, SVM, ANN(MLP)	-	✓	✓	PV array	OC, PS, SC, Degradation	-	GCPVS
[12]	2020	ML	SVM, GA	From I-V char	✓	×	PV array	LL	97%	-
[3]	2020	ML	SVM	Image (Infrared thermography)	×	✓	PV array	Hotspot	>92%	SAPVS
[37]	2020	ML	Anomaly detection	-	-	-	Battery	Internal SC, GF	-	SAPVS
[13]	2020	EL	SVM, NB, KNN	I-V char	✓	✓	PV array	LL	>99 %	SAPVS
[4]	2020	TL	AlexNetCNN	T,G, I-V char,	✓	✓	PV array	LL, OC, AF, PS with MPPT	>70.45 %	SAPVS
				Boost converter I_{max}, V_{max}, P_{max}						
		ML	SVM, RF, LSTM, Bi-LSTM							
[10]	2019	ML	KNN, SVM, ANN	-	-	-	PV array	Module SC, OC, PS, Degradation	99.65%(ANN)	-
							MPPT	MPPT fault		
[35]	2018	ML	MLP	I, V	✓	✓	PV array	SC, OC	>96 %	SAPVS
							Battery	External SC		
[2]	2018	ML	MLP	DWT voltage data	✓	×	PV array	PS	>99.1%	-
							DC-DC converter	Switch SC		
							Inverter	Switch SC, OC with LCL filter failure and Gating circuit failure		
[27]	2018	ML	KNN	T, G, V_mpp, I_mpp, and P_mpp	✓	✓	PV array	OC, LL, PS	98.7%	GCPVS
[11]	2017	ML	SVM	SCADA data	×	✓	PV array	PS, Hotspot	-	-
[21]	2017	ML	ANN	-	✓	×	Inverter	SC incipient	-	GCPVS
[25]	2017	ML	DT	-	-	-	PV array	PS,By pass diode failure	>95.3 %	GCPVS
							Inverter	failure		
[15]	2017	ML	PNN, ANN	T, G, I_{mppt} , V_{mppt}	✓	✓	PV array	Module SC, String OC	-	GCPVS
[7]	2017	ML	KELM	I-V char	✓	✓	PV array	SC, OC, PS, Degradation	-	-
[8]	2016	ML	MLP, RBF	I, V, peaks from I-V char	✓	×	PV array	SC,bypass diode, OC, connection, PS	-	-
[39]	2014	ML	-	-	-	-	Battery	Internal resistance fault, overcharging	-	-

Note: - : Not given, ✓: Is used, ×: Not used G: Irradiance, T: Temperature and for other abbreviations please refer to the document

- Training, validation, and test data set that fit at least major fault in a PV system, PV type and size are very rare to find.
- Even if most researchers have developed their own data set, most of them are simulation data. Besides, in DL-based methods, gathering the image data using a camera and drone is very expensive.
- Many measuring devices and sensors are needed due to the absence of a proper method for effective input feature selection.
- There is a lack of knowledge on how to generate rare but severe faults.
- Selection of model configurations is done with try and error.
- The model devised so far does not have the modularity and generalization capacity; as a result, ML model selection varies depending on fault type, the size, and type of input data.
- Studies that guide integrating the methods with the existing protective devices are not developed very well. Moreover, all the paper does not go in-depth on how to clear the faults. Once the fault is classified, a method and strategy are needed to coordinate it with protective devices for clearing the fault automatically and/or convey the message to the operators for solutions.
- The model's accuracy is variable as it depends on the data size, data quality, and the number of input and output features.
- For comparing ML methods based on accuracy, cost, execution time, memory usage, there are no standards or common testing platforms.

4 Conclusion and Recommendation

We found that SVM and MLP are the most utilized ML methods in recent literature. In addition, only a few literature used ensemble and transfer learning. As input features, electrical, meteorological, and image data have been used. Furthermore, the majority of ML techniques have resulted in an accuracy of greater than 90%. Besides, PV array faults such as SC, OC, and PS are the most investigated faults in a PV system. Challenges related to a data set, model configuration selection, and integration of the ML method with the existing PV system are identified. For SAPVS, it can be concluded that there is a lack of a holistic approach for critical faults in its components. Therefore, extensive research is required to see the implementation of those methods, and less investigated algorithms have to be studied. Moreover, for efficient and effective research, sharing of training, validation, and testing data set shall be encouraged.

Acknowledgements

We would like to thank our Electrical Power Engineering department at NTNU for funding this work as part of a PhD. project.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Snapshot of Global PV Markets - 2020 p. 20
2. Ahmad, S., et al.: Fault Classification for Single Phase Photovoltaic Systems using Machine Learning Techniques. In: 2018 8th IEEE India International Conference on Power Electronics (IICPE). pp. 1–6. https://doi.org/10.1109/IICPE.2018.8709463
3. Ali, M.U., et al.: A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography **208**, 643–651. https://doi.org/10.1016/j.solener.2020.08.027
4. Aziz, F., et al.: A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays **8**, 41889–41904. https://doi.org/10.1109/ACCESS.2020.2977116
5. Basnet, B., et al.: An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems **2020**, e6960328. https://doi.org/10.1155/2020/6960328
6. Bognár, A., et al.: An unsupervised method for identifying local PV shading based on AC power and regional irradiance data **174**, 1068–1077. https://doi.org/10.1016/j.solener.2018.10.007
7. Chen, Z., et al.: Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and IV characteristics. Applied Energy **204**, 912–931 (2017)
8. Chine, W., et al.: A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. Renewable Energy **90**, 501–512 (2016)
9. Cho, T.H., et al.: Comparison of Intelligent Methods of SOC Estimation for Battery of Photovoltaic System **9**(9), 8
10. Da Costa, C.H., et al.: A Comparison of Machine Learning-Based Methods for Fault Classification in Photovoltaic Systems. In: 2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America). pp. 1–6. https://doi.org/10.1109/ISGT-LA.2019.8895279
11. Dong, A., et al.: Fault Diagnosis and Classification in Photovoltaic Systems Using SCADA Data. In: 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC). pp. 117–122. https://doi.org/10.1109/SDPC.2017.31
12. Eskandari, A., et al.: Autonomous Monitoring of Line-to-Line Faults in Photovoltaic Systems by Feature Selection and Parameter Optimization of Support Vector Machine Using Genetic Algorithms **10**(16), 5527. https://doi.org/10.3390/app10165527
13. Eskandari, A., et al.: Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on I-V characteristics **211**, 354–365. https://doi.org/10.1016/j.solener.2020.09.071
14. Europe, S.P.: Global market outlook for solar power/2020–2024. Solar Power Europe: Brussels, Belgium (2020)
15. Garoudja, E., et al.: An enhanced machine learning based approach for failures detection and diagnosis of PV systems. Energy conversion and management **151**, 496–513 (2017)
16. Ghaffarzadeh, N., Azadian, A.: A comprehensive review and performance evaluation in solar (PV) systems fault classification and fault detection techniques
17. Hajji, M., et al.: Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems . https://doi.org/10.1016/j.ejcon.2020.03.004
18. Harrou, F., et al.: Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes **11**(3), 806–818. https://doi.org/10.1109/JPHOTOV.2021.3057169
19. Hong, J., et al.: Fault prognosis of battery system based on accurate voltage abnormity prognosis using long short-term memory neural networks **251**, 113381. https://doi.org/10.1016/j.apenergy.2019.113381
20. IRENA: Renewable power generation costs in 2018. International Renewable Energy Agency, Abu Dhabi (2019)
21. Ismail, N., et al.: Short-Circuit Incipient Faults Detection from Single Phase PWM Inverter using Artificial Neural Network **10**, 10

22. Kapucu, C., et al.: A supervised ensemble learning method for fault diagnosis in photovoltaic strings **227**, 120463. https://doi.org/10.1016/j.energy.2021.120463
23. Lazzaretti, A.E., et al.: A Monitoring System for Online Fault Detection and Classification in Photovoltaic Plants **20**(17), 4688. https://doi.org/10.3390/s20174688
24. Li, B., et al.: Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review **138**, 110512. https://doi.org/10.1016/j.rser.2020.110512
25. Livera, A., et al.: Advanced Failure Detection Algorithms and Performance Decision Classification for Grid-Connected PV Systems. https://doi.org/10.4229/EUPVSEC20172017-6BV.2.13
26. Madeti, S.R., Singh, S.N.: A comprehensive study on different types of faults and detection techniques for solar photovoltaic system **158**, 161–185. https://doi.org/10.1016/j.solener.2017.08.069
27. Madeti, S.R., Singh, S.N.: Modeling of PV system based on experimental data for fault detection using kNN method **173**, 139–151. https://doi.org/10.1016/j.solener.2018.07.038
28. Malik, A., et al.: Transfer Learning-Based Novel Fault Classification Technique for Grid-Connected PV Inverter. In: Mekhilef, S., et al. (eds.) Innovations in Electrical and Electronic Engineering. pp. 217–224. Lecture Notes in Electrical Engineering, Springer. https://doi.org/10.1007/978-981-16-0749-3-16
29. Mano, M., et al.: Classification and detection of faults in grid connected photovoltaic system. Int. J. Sci. Eng. Res **7**(4), 149–154 (2016)
30. Mellit, A.: Recent Applications of Artificial Intelligence in Fault Diagnosis of Photovoltaic Systems. In: Mellit, A., Benghanem, M. (eds.) A Practical Guide for Advanced Methods in Solar Photovoltaic Systems, pp. 257–271. Advanced Structured Materials, Springer International Publishing. https://doi.org/10.1007/978-3-030-43473-1-13
31. Mellit, A., et al.: Fault detection and diagnosis methods for photovoltaic systems: A review. Renewable and Sustainable Energy Reviews **91**, 1–17 (2018)
32. Nailen, R.L.: Battery protection-where do we stand? IEEE transactions on industry applications **27**(4), 658–667 (1991)
33. Pillai, D.S., Rajasekar, N.: A comprehensive review on protection challenges and fault diagnosis in PV systems. Renewable and Sustainable Energy Reviews **91**, 18–40 (2018)
34. Sabbaghpur Arani, M., Hejazi, M.A.: The Comprehensive Study of Electrical Faults in PV Arrays **2016**, 1–10. https://doi.org/10.1155/2016/8712960, https://www.hindawi.com/journals/jece/2016/8712960/
35. Sabri, N., et al.: Intelligent fault supervisory system applied on stand-alone photovoltaic system. In: 2018 International Conference on Applied Smart Systems (ICASS). pp. 1–5. https://doi.org/10.1109/ICASS.2018.8651950
36. Sabri, N., et al.: Faults diagnosis in stand-alone photovoltaic system using artificial neural network. In: 2018 6th International Conference on Control Engineering & Information Technology (CEIT). pp. 1–6. IEEE (2018)
37. Sabri, N., et al.: Battery internal fault monitoring based on anomaly detection algorithm. Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems p. 187 (2020)
38. Shin, J.H., Kim, J.O.: On-Line Diagnosis and Fault State Classification Method of Photovoltaic Plant **13**(17), 4584. https://doi.org/10.3390/en13174584
39. Tadj, M., et al.: An innovative method based on satellite image analysis to check fault in a PV system lead–acid battery **47**, 236–247. https://doi.org/10.1016/j.simpat.2014.06.010
40. Youssef, A., et al.: The role of artificial intelligence in photo-voltaic systems design and control: A review **78**, 72–79. https://doi.org/10.1016/j.rser.2017.04.046
41. Zhao, Y., et al.: Line-line fault analysis and protection challenges in solar photovoltaic arrays. IEEE transactions on Industrial Electronics **60**(9), 3784–3795 (2012)