**4th International Conference on Intelligent Technologies and Applications INTAP 2021 manuscript No.**
(will be inserted by the editor)

# Machine Learning for PV System Operational Fault Analysis

## Literature Review

**Tarikua Mekashaw Zenebe · Ole-Morten Midtgård · Steve Völler · Ümit Cali**

Received: date / Accepted: date

**Abstract** The aim of this review paper is to find out the research gap and assesses the feasibility of a holistic approach for photovoltaic (PV) system operational fault analysis using machine learning (ML) methods. The analysis includes detection and diagnosis of operational faults in PV system. Even if standard protective devices are installed in PV systems, they fail to clear various faults as a result of low current during low mismatch level, high impedance fault, low irradiance and etc. This will not only increase the energy loss but also endanger the reliability, stability and security of the PV system. Furthermore, in extreme case if a fire hazards occurred an additional asset will be destroyed. In search of other techniques, fault detection and diagnosis (FDD) in PV system using ML methods is getting attraction in recent years. This is due to their ability to handle non-linear relationship, distinguish features with similar signature and their online application. In this paper a review of literature on ML based PV system FDD methods are 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. Beside, the aforementioned finding, the review paper has identified main implementation challenges and also provide some recommendation for future work.

Tarikua Mekashaw Zenebe
Department of Electric Power Engineering, Norwegian University of Science and Technology, Trondheim, Norway
Tel.: +4792972271
E-mail: tarikua.zenebe@ntnu.no

Ole-Morten Midtgård
Department of Electric Power Engineering, Norwegian University of Science and Technology, Trondheim, Norway
E-mail: ole-morten.midtgard@ntnu.no

**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 627GW. However, due to the COVID-19 pandemic, it shows decline in 2020. Assuming a medium scenario, [2] estimated the total global installed PV generation capacity to exceed 1.2TW by 2022. In addition, the prices of electricity from PV system is decreasing constantly and becoming competent to the conventional electricity sources [3]. This all show a promise for further increase in PV market in the coming years.

With the increase in the deployment of PV for electricity production, ensuring the reliability, stability and safety of the system is very crucial. However, in spite of the advancement in technology, still fault are a headache for efficient and effective operation of PV system. Different standards mandate the installation of protective device in PV system. For instance, the National Electric Code (NEC) article 690 demanded the installation of protective devices in order to safeguard the DC side of a PV system from over-current faults, ground faults and arcing faults [4]. Nevertheless, there are various condition where the devices fail to clear fault on time. For example according to NEC, the fuse rating should be greater than 2.1 $I_{SC}$ (short-circuit current at standard test condition (STC)) in PV system [5]. However, if line to line (LL) fault occurred at low mismatch and high impedance level, the fuse will not able to clear the fault as the current will not be enough to blow the fuse [4]. In addition, due to the presence of 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 a serious damage [6]. The nonlinear characteristics of PV arrays, high impedance, low mismatch level, low-irradiance, maximum power point tracker (MPPT), faults impedance, degradation and presence of blocking diodes are some of the factors that prevent protection devices to trip under certain conditions as mentioned in [7].

The above aforementioned faults are on the PV array. If we look other components of PV system there are also cases where detection of fault is difficult, for instance finding internal faults in battery [8] and incipient faults in inverters [9]. Thus, the failure of the protective device to clear fault lead to unreliable and unsecured PV system. In addition to this, in [10], it is mentioned that the annual power loss due to various fault might goes up to 18.9%. Taking in to account the efficiency of a typical PV cell which is in between 15-21%, this reported power loss is very significant. Unless faults can be detected and cleared on time, they have the potential to affect the efficiency, energy yield, security and reliability of the PV system [5]. Furthermore, they might cause additional damages to other property for example house in case of fire at rooftop PV system. Therefore, detecting and clearing the fault on time is an indispensable solution in order to mitigate this huge losses while ensuring reliability and security of the PV system.

The distinguishing feature among fault detection and diagnosis (FDD) methods in PV systems are the time consumed to detect the malfunctioning, the input data required and the capability to differentiate between faults [11]. In [11], the FDD techniques are categorized in to image processing and electrical methods.

We can say that, the demand for techniques which is simple and cheep, can handle non-linear nature of the PV modules, can be remotely applied and is able to differentiate features with similar signature are the main motivation to move to data-driven methods like that of machine learning (ML) [10] for many researchers in recent years.

If ML is used to analyze the fault in the PV system, as much as possible there should be a holistic method that can be used to detect and diagnosis at least all the most frequent and dangerous faults. However, most of the literature focused on PV array fault. In addition to that most of the papers used simulation data to train and validate their ML model. Further more, based on the authors knowledge only one paper tried to implement the ML method in a programmable logic devices. Thus, this paper aimed at answering the reason behind all this and see the feasibility of a central FDD method through literature review.

To best of the authors knowledge this review paper is the first in reviewing literature keeping in mind the feasibility of a holistic fault analysis approach for PV system specifically for SAPVS. It is also the first review paper that has categorized and analyzed FDD in to methods based on ML and deep learning, ensemble learning and transfer learning.

The paper is organized as follow: After providing a summary of review papers, a detailed information about various faults commonly occurring in the components of PV system is dealt in the first part of section 2. Then, in the second part, a comprehensive literature review on PV faults detection and diagnosis will be provided. Whereas, in section 3, all the finding with their discussion including the research gap will be presented. Finally, the paper will conclude by giving a summary of the main finding and highlighting what has to be done in order to see the realization of ML based PV system fault analysis in general.

## 2 Literature Review

[4] provided a comprehensive literature review on possible PV faults and on advanced detection techniques. The paper tried to review literature including all PV faults. However, most of the discussion focused on PV array faults. With a similar approach to [4], Mellit et al. [5] presented a very detailed information about the fault type, cause and effect in PV system including FDD methods. Likewise, the main focus of the paper was on PV array faults. [12] reviewed papers on the role of artificial intelligence on modeling, sizing, control, fault diagnosis and output estimation of PV systems. Whereas, Li et al. [13] reviewed recent work specifically applying Artificial Neural Network (ANN) and hybrid ANN for FDD based on the fault they analysed, the type and amount of data they used, their models configuration and its performance. Beside, they highlighted the major challenges and prospects on the methods. [14] is among the papers that dedicated in explaining the faults in PV system in wider spectrum. A fault detection methods on grid connected PV system (GCPVS) were studied comprehensively in [15]. Fault was classified based on DC and AC side instead of component wise. In contrary to most of the review papers, the current paper focus only on advanced data driven approach that of ML.

In this section we will see different PV system faults classified based on components and then the varies ML methods which has been used in PV system FDD.

## 2.1 PV System Fault

In order to design an efficient and effective fault detection and diagnosis method, it is necessary to know about the characteristic of each faults including their protection challenges [4].

SAPVS comprise of PV array, inverter, storage device (like battery), charge controller, MPPT, connection wires and other additional protection and safety devices. A fault that could happen in any of the above listed component will put the reliability and safety of the PV system in danger. The main PV system fault are summarized in Figure 1.

### *2.1.1 PV array fault*

Fault in PV array can be classified in a various way. It could be permanent or temporary, electrical or environmental [11]. Electrical faults consist of open circuit(OC) module or string fault, short circuit(SC) fault which could be line to line(LL) or line to ground [10]. It also comprises arc fault (AF), bypass and blocking diodes faults [11].

*Line to Line fault (LL):* LL fault is an unintentional connection between lines with different potential difference [5, 7]. This may occur in PV array when two nodes unintentionally connected [16] for instance due to cable insulation failures, mechanical damage, water ingress, D-junction box corrosion, and hot spots caused by the back-sheet failures [5]. This is one of the PV fault that might lead to serious problem like fire hazard in addition to degrading PV arrays life time. LL fault is very hard to identify by the conventional protection devices such as Over Current Protection Devices (OCPD) like fuse and CB that is mainly: 1) as a result of the decrease in current in cases of LL fault during high impedance and low mismatch level [4, 5], (the mismatch level indicates the number of module under LL fault [7]), 2) due to the presence of a blocking diode as it block back-fed current [4, 6], 3) as the presence of MPPT decreases the current to optimize the power output and difficult to distinguish it from normal cases [4, 6], 4) its similarity with ground fault [4], 5) as a result of low current at low irradiance values, "any line-line fault occurring during the night time may remain hidden in a PV system forever "[4].

*Open Circuit fault (OC):* OC fault is an intentional disconnection of a closed loop which result in interruption of current flow. This might occur 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 [7]. In addition, broken cells, physical breakdown of cable joints, loose connections and aged power cables near terminal may lead to OC fault [10]. Due to the presence of bypass diode current flow will be kept even if OC fault occur. However, it results in a substantial power loss due to the reduction in voltage in a string [10].

*Partial Shading (PS):* PS does not only decreases (due to static shading like dust, soil, bird dropping) and results in the continuous fluctuation of the PV output power (due to dynamic shading which change as a result of sun position includes shading as result of building trees etc.) [7], but it also facilitate the degradation of PV arrays [4]. PS even can lead to complete destruction due to fire hazard as a result of cell/module temperature increase due to the dissipated energy [10]. Furthermore, as it results multiple peaks in I-V characteristic curve, it makes the job of MPPT in identifying the maximum power point extremely challenging [4]. Beside, unless time factor is used, its is hard to differentiate with OC fault as their effect on power output characteristic has similarity [10]. Furthermore, to mitigate the problem bypass diode are installed at each module but this will increase the installation cost [4]. Thus, identifying the existence of PS on time is very essential.

*Ground fault (GF):* GF occurs when a current carrying wire/cell/module connected with a ground accidentally. In a normal scenario this can be detected by Ground Fault Detection and Interruption (GFDI) and Ground Fault Protection Devices (GFPDs). However, during high impedance case, detection is challenging as the current will be low. In addition, there are scenarios where it look like SC fault [4]. Thus, this fault also need an efficient method to detect and distinguish it from other faults.

*Arc fault (AF):* AF is a fault where current flow in the air or dielectric outside the conductor. This might be due to loose connection as a result of poor solder joints or corrosion of conductors or degraded cable insulation due to mechanical damage. It could be series arc in case of connection between modules or parallel arc in case of closely placed conductor at different potential difference [4, 5, 7]. On contrary to others fault, arc fault have little effect on I-V or I-P characteristic of PV arrays. But it leads to a serious distortion in the output current and voltage wave form [4]. Unless it is cleared on time, it might lead to serious damage even cause fire. For clearing this fault, Arc Fault Circuit Interrupters (AFCIs) and Arc Fault Detectors (AFDs) are recom-


Fig. 1: Summary of the major type of faults in PV system especially in SAPVS

mended. However, multiple of them has to be installed in order to clear the fault properly. Beside, when they are installed at inverter side, they fail to protect the fault as attenuated arc signal reach to them. Moreover, those device are not capable of clearing parallel fault even if it is sever and hard to interrupt. GFDIs and OCPDs are expected to be efficient enough to address majority of parallel arc faults to the ground . But this redundancy in protection devices will increase the cost. Not only detecting arc fault but also detecting which arc fault is occurring is very important as the measure taken for one will increase the impact of the other. For example disconnecting the terminal of the inverter can clear series arc fault but on the other hand it increases circulating current if in case a parallel arc fault occurs since the voltage is increasing to open circuit voltage in the array [4].

*Others:* In addition to the above main PV array faults others may include degradation faults [17], hot spot fault [18], fault in bypass diode which could be OC or SC fault [4].

As we will see in the next section, the most investigated faults from PV system components using ML methods is that of PV array [11]. The reason could be: 1) it is the most affected PV system component by fault, 2) it is the place where identification of fault is difficult as it is highly dependent on the drastically varying weather condition. Furthermore, it is one of the component that has to be place outdoor in harsh environmental condition. Nevertheless, it is worth to mention the importance of including all component faults in PV system as the failure of one of the components will lead to total power loss and affect the reliability, as well as the security of the PV system.

#### *2.1.2 Solar battery fault*

The battery take around 43% of the life cycle cost of SAPVS [19]. As a result, it shall get an attention and a good working condition shall be prepared. One solution is identifying fault early as possible and schedule proactive maintenance. The main faults that could happen in this PV component includes external short-circuit fault [20], degradation fault [21], internal fault which could be GF and SC fault [8], overcharging (over voltage), undercharging (under voltage ) and open circuit (total voltage to zero) [18]. The impact those faults in a battery may range from decreasing its performance, shorten its lifetime and increased maintenance cost to fire hazard explosion [8]. The lack of guideline how to select fuse and circuit breaker is mentioned in [8] as one of the main challenges in detecting internal faults. Moreover, the gradual change of current and voltage of a battery make detection of fault on time extremely difficult.

#### *2.1.3 Inverter fault*

Inverter fault in PV system is classified as AC stage fault. It may include the OC of switches, SC of switches, filter failure and gating failure [5]. For instance the gate failure could be an incipient faults of the Insulated Gate Bipolar Transistor (IGBT). IGBT is the most critical component in an inverter. It is also one of the main reason 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. [9] provided the way to generate this fault.

#### *2.1.4 MPPT fault*

MPPT control system comprised of various sensor to get irradiance, temperature, current and voltage measurements as well as an optimization algorithm that can search the maximum power point in order to operate the PV array at this point and boost the PV system yield. Thus, any error in any part of the MPPT will lead to a wrong operating power point which in turn significantly decrease the output power of the PV system. Sensor failure and lack of an efficient and effective MPPT algorithm are the most common fault in MPPT [22, 23].

### 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 comprise of the process of finding the type of fault and localizing the occurrence.

From the previous section's discussion, we can conclude that there is a need for other methods in order to assist protective devices whenever they fail to clear faults. With similar reason, a number of research has been conducted on PV system FDD in recent years. These researchers have devised methods that can fall in to the following categories: signal processing, performance comparison and machine learning according to [7].

In this paper in order to show the other possible direction of ML application for better performance, the reviewed literature are classified in to methods based on 1) machine learning and deep learning, 2) ensemble learning and 3) transfer learning.

#### *2.2.1 Methods based on machine learning and deep learning*

In this section papers which used machine learning and deep learning algorithms are presented by classifying them into supervised and unsupervised learning. From supervised machine learning Support Vector Machine (SVM), Naive Bayes (NB), k-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), Discriminant Analysis (DA) and Radial Basis Function (RBF) are found to be used in PV system FDD. Whereas, from unsupervised machine learning anomaly detection algorithm is employed. Finally, from deep learning, methods based on Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN) are the most developed. Similar finding and a comprehensive explanation about ML and deep learning with respective of PV application is given in the book chapter in [11].

*Methods based on supervised machine learning*

Among other works in [24], it used one-SVM to detect anomaly and fault in PV system. Higher accuracy while classifying faulty and abnormal operation was achieved while using climate corrected performance ratio instead of only using performance ratio. However, the paper fail to mention which fault has been analyzed specifically beside saying it is fault or normal operation. Identifying which fault happened/is happening is very important to diagnose fault for fast recovery thus to enhance the resilience of the PV system.

The major limiting factor in using supervised ML are the need for labeled data set and the cost related to measuring device for collecting historic data. For solving this, Dong et al. [25] proposed a FDC method based on SVM while using available scada data. In addition, as an input for training the SVM model, they have used index called anomaly detection index. The index is derived based on weather corrected performance ratio. They were able to distinguish between recoverable fault like shading and unrecoverable fault like hot-spot. However, this paper focuses only on PV string fault.

If we would like to consider multiple faults as an output, it is difficult to find input features that can lead to a separable output. Hajji et al. [26] has tried to tackle this problem by including a feature extraction and selection stage using principal component analysis before providing the input to ML algorithm. They have tested various classifiers like KNN, RF, DA, NB, DT and SVM to classify fault in GCPVS. In order to evaluate the performance of the classifier, they have used metrics like false alarm rate, missed detection rate and good isolation rate. They all have achieved an accuracy greater than 96%. In addition, the execution time of each classifier were evaluated. We can clearly see that relative to other papers they have included inverters, MPPT and DC-DC converter in addition to PV array. Nevertheless, battery fault is not analysed.

In [27] a monitoring, fault detection and classification system for GCPVS is proposed. The authors used comparison between model and real system output to identify fault occurrence. First, they have tested varies linear and nonlinear model of PV to suggest the suitable one for fault detection. 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 suitable and more accurate ML algorithm for fault classification. The paper is limited in analyzing PV array fault.

Basnet et al.[28] has used one of the most commonly applied machine learning model, MLP or a probability neural network as it is refereed in the paper, to detect and classify LL and GF in GCPVS. They could achieve 100% training accuracy. As an input parameters, voltage, current, G, average T of each module and weather condition were utilized. In spite of getting a good accuracy, its application is limited as its main focus were PV array. Therefore, further analysis is needed by including other PV system faults.

[29] is among the few papers which evaluate the ML algorithm's performance based on both accuracy and execution time. In the mentioned paper, faults like module SC, MPPT fault, OC, PS, and degradation has been detected and classified using five different ML techniques such as 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 in cooperating other faults.

A cascaded Probabilistic neural network (PNN) due to its robustness to noise has been used in [30] to detect and classify 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 an input features, they have utilized temperature, tilted irradiance, current ($I_{MPP}$) and voltage ($V_{MPP}$) at MPP. The training data set are generated from validated PSIM™/Matlab™ Cosimulation one diode PV system model taking an existing 9.54kWp PV system parameter located in Algeria. In spite of the effort made to brought a robust method, here also the literature focused on faults on PV array and DC side of the GCPVS. However, one peculiar aspect of this paper is that they have tested the proposed method in the aforementioned real PV system.

For the first time (as the authors claim) kernel based extreme learning machine (KELM) was used in [31] to detect and classify degradation fault, OC, SC and PS. In this paper features that enable identification of the fault are extracted after examining their impact on I-V characteristics. Beside, the simulink PV model was validated before using it in simulating the fault in order to generate training and test data set. In addition to the simulation data, a real laboratory PV array data set has been used. The challenge of using real PV system is that we do not have control over the irradiance and temperature data and it is hard to train the model and able the model to generalize. Likewise to other papers, however it is not verified that this method could also be used in order to identify faults considering other components of PV system. In general even if a very efficient and accurate method is devised here, the determination of I-V characteristics of the PV array in online scenario might be problematic.

[21] is one of the few papers which has focused specifically on faults in SAPVS. They proposed a fault diagnosis method based on a supervised learning using MLP feed forward neural network to detect and classify faults such as SC of PV module, OC of PV module and external SC of battery where the fuse fail 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 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. [32] used a combination of threshold method and ML to detect and classify eight faults including SC, inversed and shunted bypass diode fault, OC, connection fault, shadow effect with normal operation, with bypass fault and connection fault. SC, inversed and shunted bypass diode fault, and connection fault are founded to have the same combination of attributes. From ML, MLP and RBF has been compared. For threshold method, measured and simulated PV array output power were compared. For ML extracted attribute like current, voltage and peaks from I-V characteristics were used as input. In addition to showing the feasibility of the method, this is the only paper encountered that show 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 focus of this paper is also on fault on PV array The other drawback of this paper is that the applied threshold method is very much dependent on system parameter and the accuracy of the threshold limits.

The authors in [6] focused on detection of LL fault in PV system under high impedance fault and low mismatch condition which is one of the case where protection devices fail to clear fault. The author used SVM classifier which is resistance to model error and computational efficiency, based on the features extracted by analysing I-V characteristics of PV array. The PV array was modeled in Simulink.

They were able to detect and classify LL fault with greater than 99.3% training and 89.16% validation accuracy with various kernel function for SVM and using less number of data set. They have selected Gaussian kernels with detection and classification accuracy of 97% approximately.

Even if the methods in the above two papers were validated for LL fault, still further analysis is needed in order to apply it for MIMO case in SAPVS. Moreover, it has to be validated with real PV system data in addition to the simulation.

In stead of threshold method, Ahmad et al. [33] has used a combination of transformation for feature extraction and ML algorithm for detecting and classifying PS condition as modular fault, DC-DC converter switch SC, inverter switch OC, inverter switch SC with LCL filter failure and gating circuit failure. The transformation used is discrete wavelet transform (DWT) due to its less computational time and complexity as well as it enable us to work both in time and frequency domain. Whereas, the ML algorithm is MLPNN. The data set was obtained from simulation of a PV system in MATLAB/Simulink environment. They could achieve 99.3 %, 99.1 % and 100 % training, validation and testing accuracy respectively. However for SAPVS, the battery and MPPT fault is missing. Further more, the most difficult scenario of LL fault which is high impedance and low mismatch has to be checked. Beside, the geographical location irradiance and temperature data has taken is not specified.

[34] used hybrid features based support vector machine (SVM) model in order to detect and classify hot spot fault in PV panels using infrared thermography as input image for the model. The model could detect and classify between healthy, non-faulty hot spot and faulty hot spot faults 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 effect. Nevertheless, for large scale PV system, the output of the result will have a significant role.

In [35], DT was used for detecting PS, inverter and bypass diode failure in GCPVS. In addition to other methods, the author has used decision tree for detecting and classifying these faults. They achieved an average classification accuracy of 98.7%.

One of the reason for short life time of solar battery is it's improper operation. In addition to the available energy from solar or demanded load, in order to decide whether the battery has to be charged or discharged, knowing accurately how much is in the battery is a determining factor. There are various statistical estimation techniques but recently SOC estimation using ML is getting attraction as that of PV panel as it exhibit a non-linear input out put characteristics. [36] presented ML based SOC estimation method for the most common solar battery which is lead-acid battery. They proposed methods based on a feed-forward neural network (FFNN), a recurrent neural network (RNN) and an adaptive neuro-fuzzy inference system (ANFIS). As an input feature for the model, voltage and current data was used. However, the paper did not mentioned how the training SOC data is obtained, expect it is obtained from an experimental setup. The out put of this paper could be used for further studying FDD methods in PV battery.

Internal resistance effect and overcharging problem in lead-acid battery in PV system, was detected using solar radiation data estimated from satellite image analysis in [18]. The satellite data is taken from three location, two from Algeria and one in Spain. In the paper we can clearly see that 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, eventually from the finding it is a good indication on the use of this battery voltage for detecting and classifying those faults in battery.

In other study in [37] the author has used long short-term memory (LSTM) recurrent neural network for state prediction and fault prognosis for battery in electric vehicle. From the knowledge domain, this approach could also be used for solar battery and its finding is significantly important though its feasibility has to be checked in solar battery.

One of the challenge in using ML specially in analyzing inverter fault in PV system is the lack of methods that guide us how to generate the faults in simulation as the faults does not occur frequently but they are responsible for the majority of inverter failure. Thus, Ismail et al. [9] used feed-forward back propagation neural network to detect SC incipient faults by first modeling a way to generate this fault for GCPVS. This paper is among the few literature that concentrate on PV inverter faults. For using it in SAPVS, other fault has to be incorporated and the method has to be verified for it performance for other PV system components fault.

*Methods based on unsupervised machine learning (UML)*

In [19] 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 labelled 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 SAPSS in Matlab/Simulink using irradiance and temperature data from from the Centre de Development des Energies renewable (CDER), Algeria. They have used ready available current and voltage data set. As the work is only for internal fault, it is important to incorporate it/hybridize it with other method for the whole identification of faults that could happen in SAPVS.

### *2.2.2 Methods based on ensemble learning (EL)*

[38], similarly to [6] 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 result of the algorithm was used. They could achieve an average of 99% and 99.5% for detecting and classifying LL fault. And they have evaluated the model with simulation and experimental data set.

### *2.2.3 Methods based on transfer learning (TL)*

In order to detect and classify PV system faults, [7] proposed a deep 2-D CNN to extract 2-D scalograms generated from PV system. The authors analysed 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 notice the decreasing of performance as the number of class increase. In this paper the concept of transfer learning has been employed by using a pre-trained AlexNetCNN for feature extraction and classification in order to minimize the impact of low data set in the model performance. In addition, AlexNetCNN has been selected based on its less execution time than GoogleLeNt and ResNet. As the CNN model require 2-D input data, 1-D data has been transformed to 2-D scalograms using continuous wavelet transform. In addition to CNN, ML algorithms like SVM, RF and from DL algorithms like LSTM, Bi-LSTM have been used to compare the model performance. The input features are irradiance, temperature, main points from I-V characteristics of the PV array and maximum current, voltage and power from boost converter. Based on the author survey, this is the only paper that have used deep learning in this depth and made comparison with classical ML models. In this paper instead of using separate model for detecting and classifying fault, it uses only one method considering normal case as one of the output class. This approach seems promising in handling MIMO data, however it still needs to be verified for SAPVS fault analysis. Beside the simulation data set, the method has to be verified using real data. It is worthy to mention that this paper has clearly indicate which software(version), platform and computer specification had been used.

## 3 Result and Discussion

In this section, the authors will try to present and discuss the main findings from the performed literature review. Besides, later in this section, the research gap which has to get a greater attention in the future researches are highlighted.

As we can see from Fig 2 the number of publication on ML based FDD in PV system has increased tremendously in 2020. This shows the special attention that has been given for ML based FDD methods in PV system in recent years.

As we can see from Table 1 the majority of the papers, greater than 80%, has analysed mainly the fault in PV array. Even if they are very less, faults in inverter

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 |
| [24] | 2020 | ML | SVM | - | - | - | - | Normal and Faulty | - | - |
| [25] | 2017 | ML | SVM | Scada data | ✗ | ✓ | PV array | PS, Hotspot | - | - |
| [26] | 2020 | ML | KNN, RF, DA, NB, DT, SVM | - | - | - | PV array | OC, PS, Sensor | >96% | GCPVS |
| | | | | | | | Inverter | IGBT, Grid connection | | |
| [27] | 2020 | ML | KNN, DT, SVM, ANN(MLP) | - | ✓ | ✓ | PV array | OC, PS, SC, Degradation | - | GCPVS |
| [28] | 2020 | ML | KNN, DT, SVM, ANN(MLP) | - | ✓ | ✓ | PV array | OC, PS, SC, Degradation | - | GCPVS |
| [29] | 2019 | ML | KNN, SVM, ANN | - | - | - | PV array | Module SC, OC, PS, Degradation | 99.65%(ANN) | - |
| | | | | | | | MPPT | MPPT fault | | |
| [30] | 2017 | ML | PNN, ANN | T, G, $I_{mppt}$ , $V_{mppt}$ | ✓ | ✓ | PV array | Module SC, String OC | - | GCPVS |
| [31] | 2017 | ML | KELM | I-V char | ✓ | ✓ | PV array | SC, OC, PS, Degradation | - | - |
| [21] | 2018 | ML | MLP | I, V | ✓ | ✓ | PV array | SC, OC | >96 % | SAPVS |
| | | | | | | | Battery | External SC | | |
| [32] | 2016 | ML | MLP, RBF | I, V, peaks from I-V char | ✓ | ✗ | PV array | SC,bypass diode, OC, connection, PS | - | - |
| [6] | 2020 | ML | SVM, GA | From I-V char | ✓ | ✗ | PV array | LL | 97% | - |
| [33] | 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 | | |
| [17] | 2021 | ML | SVR, GPR | T, G, $P_{peak}$ | ✓ | ✓ | PV array | OC, SC, LL, PS, Degradation | - | GCPVS |
| [10] | 2018 | ML | KNN | T, G, $V_mpp$, $I_mpp$, and $P_mpp$ | ✓ | ✓ | PV array | OC, LL, PS | 98.7% | GCPVS |
| [9] | 2017 | ML | ANN | - | ✓ | ✗ | Inverter | SC incipient | - | GCPVS |
| [35] | 2017 | ML | DT | - | - | - | PV array | PS,By pass diode failure | >95.3 % | GCPVS |
| | | | | | | | Inverter | failure | | |
| [18] | 2014 | ML | - | - | - | - | Battery | Internal resistance fault, overcharging | - | - |
| [34] | 2020 | ML | SVM | Image (Infrared thermography) | ✗ | ✓ | PV array | Hotspot | >92% | SAPVS |
| [19] | 2020 | UML | Anomaly detection | - | - | - | Battery | Internal SC, GF | - | SAPVS |
| [38] | 2020 | EL | SVM, NB, KNN | I-V char | ✓ | ✓ | PV array | LL | >99 % | SAPVS |
| [7] | 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 | | | | | | | |

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

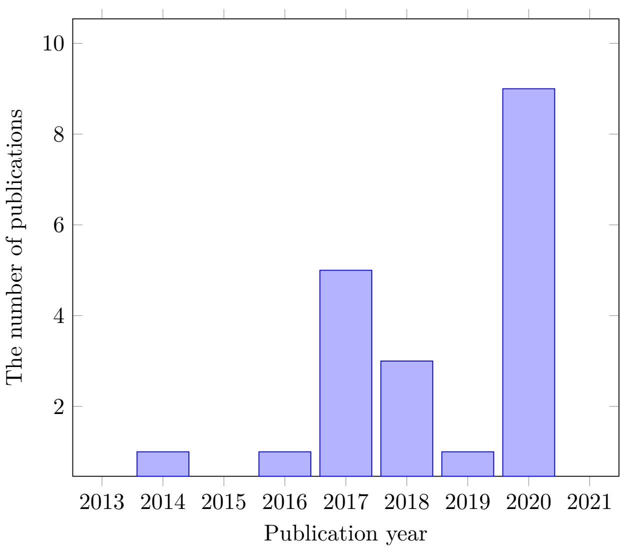


Fig. 2: The number of publication on ML based FDD in PV system versus publication year

and battery has been also analysed. It is worthy to mention here that non of the papers reviewed has studied all components at once. Relatively, faults in GCPVS has got a special attention than SAPVS. SC, OC and PS are the most investigated type of fault using ML methods. However, double ground fault, arc fault, LL fault are the most sever one. On the other hand those faults are less frequent. Thus, methods that protect the PV system during those conditions shall be devised.

Among other ML methods, SVM and MLP in general has been used extensively to detect and classify faults in PV system. Selection of hyper parameters for the algorithm like batch size, epoches, momentum are mostly set by try and error. For evaluating the models, accuracy and confusion metrics are the most employed performance indices. However, some has devised their own metrics and also used execution time. It is very important to mention here that due to ML random nature, it is good 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 installed, Algeria took the lead, China and Korea take the second places. It is also very important to analyze the performance of the ML methods under different climatic and geographical condition in order to device a general method that fit all if possible. The reason behind is that depending on the geographical location the challenge for the PV panel and the battery is different. For instance while snow is a big problem in the polar region, dust, soiling as well as higher operating temperature are the huge problems in equatorial region.

Most of the papers depend on the input features which has been generated from simulated PV system. Whereas, only few have included experimental data. This is because of the difficultly in setting up a PV system only for collecting data. Furthermore, when available PV system exist 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. G, T and major points from I-V characteristics are the most utilized input features in case of fault in PV array. While current and voltage data are used in case of fault in 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 algorithm 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.

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

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

- Training, validation and test data set that fit at least major fault in PV system, PV type and size are very rare to find.
- Even if most researchers has developed their own data set, the majority of them are simulation data. Beside, in case of deep learning based methods, gathering the image data using camera and drone is very expensive.
- Many measuring devices and sensors are needed due to the absence of proper method for effective input feature selection.
- There is lack of knowledge on how to generate the rare but sever faults.
- Selection of model configurations are 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 how to integrate 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 is needed to coordinate it with protective devices for clearing the fault automatically and/or convey the message to the operators for solutions.
- The accuracy of the model is variable as it depends on the data size, data quality and number of input and output feature.
- For comparing ML methods based on accuracy, cost, execution time, memory usage, there is no standards or common testing platforms.

## 4 Conclusion and Recommendation

We have seen that SVM and MLP are the most utilized ML methods. In addition electrical and meteorological data have been used as an input features. Furthermore, the majority of ML techniques has resulted in an accuracy of greater than 90%. Not only this, we have also found that SC, OC and PS are the most investigated faults in PV system.

Challenges related to data set, model configuration selection, and integration of ML method with the existing PV system are identified.

For SAPVS specifically, it can be concluded that there is a lack of a holistic approach for critical faults that can happen in each component of SAPVS.

In general, as it is mentioned in [4], a reliable fault detection techniques/algorithm must be able to detect multiple faults without interfering with power production, capable of distinguish and localize faults, economical and flexible to be integrated with existing PV systems easily, simple in structure, modular and able to generalize irrespective of the type and size of PV systems.

Therefore, in order to see the real implementation of those methods an extensive research is required. In addition, algorithm which has not be investigated so far such as RNN has to be studied in case if they resulted a better performance. Beside, 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, [Online]. Available: https://iea-pvps.org/snapshot-reports/snapshot-2020/ (visited on Apr. 26, 2021) (cit. on p. 1).

[2] S. P. Europe, "Global market outlook for solar power/2020–2024," *Solar Power Europe: Brussels, Belgium*, 2020. [Online]. Available: https://www.solarpowereurope.org/global-market-outlook-2020-2024/ (cit. on p. 1).

[3] IRENA, "Renewable power generation costs in 2018," *International Renewable Energy Agency, Abu Dhabi*, 2019. [Online]. Available: https://www.irena.org/publications/2019/May/Renewable-power-generation-costs-in-2018/ (cit. on p. 1).

[4] D. S. Pillai and N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in pv systems," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 18–40, 2018 (cit. on pp. 1–4, 10).

[5] A. Mellit, G. M. Tina, and S. A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 1–17, 2018 (cit. on pp. 1–3, 5).

[6] A. Eskandari, J. Milimonfared, M. Aghaei, and A. H. M. E. Reinders, "Autonomous Monitoring of Line-to-Line Faults in Photovoltaic Systems by Feature Selection and Parameter Optimization of Support Vector Machine Using Genetic Algorithms," *Applied Sciences*, vol. 10, p. 5527, Jan. 2020 (cit. on pp. 2, 3, 7–9).

[7] F. Aziz, A. Ul Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays," *IEEE Access*, vol. 8, pp. 41 889–41 904, 2020 (cit. on pp. 2, 3, 5, 8, 9).

[8] R. L. Nailen, "Battery protection-where do we stand?" *IEEE transactions on industry applications*, vol. 27, no. 4, pp. 658–667, 1991 (cit. on pp. 2, 5).

[9] N. Ismail, F. H. Nordin, and Z. A. M. Sharrif, "Short-Circuit Incipient Faults Detection from Single Phase PWM Inverter using Artificial Neural Network," *Indian Journal of Science and Technology*, vol. 10, p. 10, 2017 (cit. on pp. 2, 5, 8, 9).

[10] S. R. Madeti and S. N. Singh, "Modeling of PV system based on experimental data for fault detection using kNN method," *Solar Energy*, vol. 173, pp. 139–151, Oct. 1, 2018 (cit. on pp. 2, 3, 9).

[11] A. Mellit, "Recent Applications of Artificial Intelligence in Fault Diagnosis of Photovoltaic Systems," in *A Practical Guide for Advanced Methods in Solar Photovoltaic Systems*, A. Mellit and M. Benghanem, Eds., Cham: Springer International Publishing, 2020, pp. 257–271 (cit. on pp. 2–5).

[12] A. Youssef, M. El-Telbany, and A. Zekry, "The role of artificial intelligence in photo-voltaic systems design and control: A review," *Renewable and Sustainable Energy Reviews*, vol. 78, pp. 72–79, Oct. 1, 2017 (cit. on p. 2).

[13] B. Li, C. Delpha, D. Diallo, and A. Migan-Dubois, "Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review," *Renewable and Sustainable Energy Reviews*, vol. 138, p. 110 512, Mar. 1, 2021 (cit. on p. 2).

[14] N. Ghaffarzadeh and A. Azadian, "A comprehensive review and performance evaluation in solar (PV) systems fault classification and fault detection techniques," 2018 (cit. on p. 2).

[15] S. R. Madeti and S. N. Singh, "A comprehensive study on different types of faults and detection techniques for solar photovoltaic system," *Solar Energy*, vol. 158, pp. 161–185, Dec. 1, 2017 (cit. on p. 2).

[16] Y. Zhao, J.-F. De Palma, J. Mosesian, R. Lyons, and B. Lehman, "Line–line fault analysis and protection challenges in solar photovoltaic arrays," *IEEE transactions on Industrial Electronics*, vol. 60, no. 9, pp. 3784–3795, 2012 (cit. on p. 3).

[17] F. Harrou, A. Saidi, Y. Sun, and S. Khadraoui, "Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes," *IEEE Journal of Photovoltaics*, vol. 11, pp. 806–818, May 2021 (cit. on pp. 4, 9).

[18] M. Tadj, K. Benmouiza, and A. Cheknane, "An innovative method based on satellite image analysis to check fault in a PV system lead–acid battery," *Simulation Modelling Practice and Theory*, vol. 47, pp. 236–247, Sep. 1, 2014 (cit. on pp. 4, 5, 7, 9).

[19] N. Sabri, A. Tlemçani, and A. Chouder, "Battery internal fault monitoring based on anomaly detection algorithm," *Advanced Statistical Modeling, Forecasting, and Fault Detection in Renew-*

*able Energy Systems*, p. 187, 2020 (cit. on pp. 5, 8, 9).

[20] N. Sabri, A. Tlemçani, *et al.*, "Faults diagnosis in stand-alone photovoltaic system using artificial neural network," in *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*, IEEE, 2018, pp. 1–6 (cit. on p. 5).

[21] N. Sabri, A. Tlemçani, and A. Chouder, "Intelligent fault supervisory system applied on stand-alone photovoltaic system," in *2018 International Conference on Applied Smart Systems (ICASS)*, Nov. 2018, pp. 1–5 (cit. on pp. 5, 6, 9).

[22] S. R. Madeti and S. Singh, "Modeling of pv system based on experimental data for fault detection using knn method," *Solar Energy*, vol. 173, pp. 139–151, 2018 (cit. on p. 5).

[23] M. Mano, R. Pau, R. Mahalakshmi, M. Karuppasamypandiyan, A. Bhuvanesh, and R. Ganesh, "Classification and detection of faults in grid connected photovoltaic system," *Int. J. Sci. Eng. Res*, vol. 7, no. 4, pp. 149–154, 2016 (cit. on p. 5).

[24] J.-H. Shin and J.-O. Kim, "On-Line Diagnosis and Fault State Classification Method of Photovoltaic Plant," *Energies*, vol. 13, p. 4584, Jan. 2020 (cit. on pp. 5, 9).

[25] A. Dong, Y. Zhao, X. Liu, L. Shang, Q. Liu, and D. Kang, "Fault Diagnosis and Classification in Photovoltaic Systems Using SCADA Data," in *2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, Aug. 2017, pp. 117–122 (cit. on pp. 6, 9).

[26] M. Hajji, M.-F. Harkat, A. Kouadri, K. Abodayeh, M. Mansouri, H. Nounou, and M. Nounou, "Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems," *European Journal of Control*, Apr. 9, 2020 (cit. on pp. 6, 9).

[27] A. E. Lazzaretti, C. H. da Costa, M. P. Rodrigues, G. D. Yamada, G. Lexinoski, G. L. Moritz, E. Oroski, R. E. de Goes, R. R. Linhares, P. C. Stadzisz, J. S. Omori, and R. B. dos Santos, "A Monitoring System for Online Fault Detection and Classification in Photovoltaic Plants," *Sensors*, vol. 20, p. 4688, Jan. 2020 (cit. on pp. 6, 9).

[28] B. Basnet, H. Chun, and J. Bang, "An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems," *Journal of Sensors*, vol. 2020, e6960328, Jun. 9, 2020 (cit. on pp. 6, 9).

[29] C. H. da Costa, G. L. Moritz, A. E. Lazzaretti, B. M. Mulinari, H. C. Ancelmo, M. P. Rodrigues, E. Oroski, and R. E. de Goes, "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)*, Sep. 2019, pp. 1–6 (cit. on pp. 6, 9).

[30] E. Garoudja, A. Chouder, K. Kara, and S. Silvestre, "An enhanced machine learning based approach for failures detection and diagnosis of pv systems," *Energy conversion and management*, vol. 151, pp. 496–513, 2017 (cit. on pp. 6, 9).

[31] Z. Chen, L. Wu, S. Cheng, P. Lin, Y. Wu, and W. Lin, "Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and iv characteristics," *Applied Energy*, vol. 204, pp. 912–931, 2017 (cit. on pp. 6, 9).

[32] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. M. Pavan, "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks," *Renewable Energy*, vol. 90, pp. 501–512, 2016 (cit. on pp. 7, 9).

[33] S. Ahmad, N. Hasan, V. S. Bharath Kurukuru, M. Ali Khan, and A. Haque, "Fault Classification for Single Phase Photovoltaic Systems using Machine Learning Techniques," in *2018 8th IEEE India International Conference on Power Electronics (IICPE)*, Dec. 2018, pp. 1–6 (cit. on pp. 7, 9).

[34] M. U. Ali, H. F. Khan, M. Masud, K. D. Kallu, and A. Zafar, "A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography," *Solar Energy*, vol. 208, pp. 643–651, Sep. 15, 2020 (cit. on pp. 7, 9).

[35] A. Livera, A. Phinikarides, G. Makrides, J. Sutterlueti, and G. Georghiou, *Advanced Failure Detection Algorithms and Performance Decision Classification for Grid-Connected PV Systems*. Sep. 1, 2017 (cit. on pp. 7, 9).

[36] T.-H. Cho, H.-R. Hwang, J.-H. Lee, and I.-S. Lee, "Comparison of Intelligent Methods of SOC Estimation for Battery of Photovoltaic System," *International Journal of Advanced Computer Science and Applications*, vol. 9, p. 8, 2018 (cit. on p. 7).

[37] J. Hong, Z. Wang, and Y. Yao, "Fault prognosis of battery system based on accurate voltage abnormity prognosis using long short-term memory neural networks," *Applied Energy*, vol. 251, p. 113 381, Oct. 1, 2019 (cit. on p. 8).

[38] A. Eskandari, J. Milimonfared, and M. Aghaei, "Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on I-V characteristics," *Solar Energy*, vol. 211, pp. 354–365, Nov. 15, 2020 (cit. on pp. 8, 9).