Artificial neural network based photovoltaic module diagnosis by current-voltage curve classification

Marica Laurino, Michel Piliougine and Giovanni Spagnuolo

Artificial neural network based photovoltaic module diagnosis by current-voltage curve classification

Marica Laurino, Michel Piliougine*, and Giovanni Spagnuolo

Università degli Studi di Salerno. Via Giovanni Paolo II 132. 84084 Fisciano (SA), Italy

ABSTRACT

In this paper a model-based procedure for fault detection and diagnosis of photovoltaic modules is presented. A four-layered feedforward artificial neural network learns the correlation between the features of the current vs. voltage curve and the environmental variables, which are the irradiance and the temperature. This correlation describes the behavior of the module in normal conditions. Moreover, the effect of anomalous variation of some parameters is learnt and correlated to the shape of the same curve, thus associated to a specific failure mechanism. The neural network is trained by using synthetic curves simulated with the single diode model and some well assessed and validated translation formulae. The approach is also validated by means of an experimental set of measurements.

Keywords:
Artificial neural network, Current-voltage curve, Fault classification, Fault detection, Multilayer perceptron, Single diode model.

1. Introduction

Photovoltaic modules are one of the most promising renewable energy sources and this is confirmed by the exponential growth of PV plants installed all over the world in the last decades. Indeed, if in 2011 the total installed capacity was only 72 GW, by the end of 2020 grew up to almost 708 GW (IRENA, 2021). Despite the Covid-19 pandemic, it is expected that the global electricity from PV sources in 2021 will rise by 18%, reaching the amount of 1 000 TWh (IEA, 2021). In the longer-term operation, the modules are subject to ageing and to various faults since they are exposed to bad weather, to corrosion and to ultraviolet radiation. These faults may lead to abnormal operation and cause safety problems and fire hazards (Sabbaghpur Arani and Hejazi, 2016), as well as big power losses. To fulfil the need for higher performance, efficiency, reliability, power availability and safe operation of PV systems, techniques for monitoring, fault detection and diagnostic purposes are today more and more proposed in the literature (Chine, et al., 2016) to detect and rapidly clear off any occurring degradation and fault. These methods are generally classified as visual, imaging solutions or electrical, depending on the fact that they are based on an operator inspection, on an automatic detection through algorithms analyzing images in the visible or infrared spectra, and on the processing of electrical measurements performed at the module terminals. In this latter group, which is of interest for this paper, many techniques are based on ANNs[1].

ANNs gained a huge success in multi-class classification problems, as well as for pattern recognition and regression tasks. For this reason, many researchers use ANNs as a FDD[2] method to efficiently diagnose and classify the degradation level and different type of faults occurring in PV systems. Above all the different topologies, MLP[3] and RBF[4] networks gained high popularity (Driss, et al., 2017; Sug, 2009)

In most of the cases, fault identification and localization are performed by feeding the classifier with the electrical parameters defining the SDM[5]. Thus, starting from a current vs. voltage (*I-V*) curve, a parameter identification procedure is applied by using a well-established and reliable algorithm. Consequently, the FDD action is performed in the parameters space, not in the *I-V* plane. These approaches also require the measurement of the actual irradiance and the operating temperature of the module. For example, in (Zhu, et al., 2018) and (Garoudja, et al., 2017), some attributes of the *I-V* characteristics of the PV strings are calculated with a simulation model and provided as input to train a probabilistic neural network, together with the irradiance G [Wm^{-2}] and the cell temperature T_c [K]. This leads to the identification of possible faulty operating conditions since different faults show different effects on the *I-V* curve attributes. The same approach is followed in (Aziz, et al., 2020) and (Basnet, et al., 2020), where a convolutional neural network and a multilayer perceptron are respectively used. Unfortunately, the reliability of the approaches working in the parameters space depends on the performance of the parameter identification algorithm. It is well known that computing such parameters is difficult and it might be demanding from a computational point of view.

[1] ANN: Artificial Neural Networks
[2] FDD: Fault Detection and Diagnosis
[3] MLP: Multilayer Perceptron
[4] RBF: Radial Basis Function
[5] SDM: Single Diode Model

* Corresponding author. Tel: +39 371 4201 756.
E-mail addresses: m.laurino7@studenti.unisa.it (M. Laurino); mpiliougineroca@unisa.it (M. Piliougine*); spagnuolo@unisa.it (G. Spagnuolo).

In this paper, an ANN-based method operating only in the *I-V* plane is presented. It takes as input the duly sampled *I-V* curve of the PV module and assigns it to a class corresponding to the normal operation or to a specific fault of a given extent. The training is performed by providing to the ANN a number of synthetic curves generated through a well assessed model. The degraded conditions are simulated by infringing the natural correlation between the *I-V* curve features and the exogenous variables G and T_c to some extent, by varying the model parameters set that, according to the literature, is affected differently by each degradation mechanism. The idea underlying the proposed approach is that, regardless of G and T_c values, which are not required for each *I-V* curve, the ANN learns the "natural" shape of a healthy PV module *I-V* curve. Moreover, it is able to properly classify those curves having a shape indicating the abnormal variation of some parameters due to the module degradation. Sixteen operating status are considered, characterized by different environmental conditions in terms of irradiance G and cell temperature T_c. Concerning the network architecture, the ANN is a four-layered feedforward neural network, and it is trained with the Scaled Conjugate Gradient backpropagation algorithm.

The major advantage of this approach is that there is no need for a reliable algorithm to extract the SDM parameters or to know in advance the exact value of the irradiance and cell temperature. Indeed, in this case the SDM is directly used to generate synthetic samples, reproducing the shape of *I-V* curves in healthy and faulty operating conditions, which are later fed into the network. Thus, starting from the analysis of the shape of the *I-V* curve, the network is able to correctly distinguish the cases where the distortion of the curve is due to the natural variation of environmental factors from those where it is due to an occurring fault.

The paper is organized as follows: Section 2 describes the model used to generate the synthetic data to train the ANN. Section 3 gives a summary of the literature results concerning the correlation between different type of degradations inspected in this paper and the model parameters. Section 4 is dedicated to the description of the ANN-based algorithm and in Section 5 the training results are presented and discussed. It also shows the classification results obtained by using an experimental set of curves.

2. The Single Diode Model (SDM)

The SDM is a well assessed model used to generate PV modules *I-V* curves in different environmental conditions (Sera, et al., 2007; Stein, et al., 2010). In this paper, the fast explicit formulation based on Lambert *W*-function, denoted by $W(\theta_I)$, is considered (Petrone, et al., 2017). The PV module current is expressed as an explicit function of voltage, as it is shown in Eq. (1):

$$I = \frac{R_{sh}\cdot\left(I_{ph}+I_s\right)-V}{R_s+R_{sh}} - \frac{\eta V_{th}}{R_s}\cdot W\left(\theta_I\right) \tag{1}$$

where

$$\theta_I = \frac{\left(R_s \,||\, R_{sh}\right) I_s \exp\left(\dfrac{R_{sh}\,R_s\left(I_{ph}+I_s\right)+R_{sh}\,V}{\eta V_{th}\left(R_s+R_{sh}\right)}\right)}{\eta V_{th}} \tag{2}$$

and

$$R_s \,||\, R_{sh} = \frac{R_s\cdot R_{sh}}{R_s+R_{sh}} \tag{3}$$

In these equations, I_{ph} [A] is the photo-generated current and I_s [A] denotes the saturation current. R_s [Ω] is the module series resistance, R_{sh} [Ω] is the module shunt resistance and η [-] is the diode ideality factor. Finally, V_{th} [V] is the thermal voltage of the PV device, defined as

$$V_{th} = N_s \frac{\mathrm{k_b}\,T_c}{\mathrm{q}} \tag{4}$$

where q = 1.602e−19 [C] is the electric charge and k_b = 1.3806503e−23 [JK^{-1}] is the Boltzmann constant, and N_s is the number of cells in series.

The five parameters {I_{ph}, I_s, R_s, R_{sh}, η} appearing in Eq. (1) have a natural dependency on G and T_c that can be expressed through a set of translating equations with respect to a reference operating condition. The ones listed in Table 1 are widely validated in the recent literature.

Table 1
Translating equations for SDM parameters

Formula	Eq.
$I_{ph}(G,T_c) = \frac{G}{G_{stc}}\left[I_{ph,ref} + \mu_{I_{sc}}(T_c - T_{ref})\right]$ (Anani and Ibrahim, 2020; Petrone, et al., 2017)	(5)
$I_s(T_c) = C_0 T_c^{\left(\frac{3}{\eta}\right)} \exp\left(-E_g(T_c) \Big/ \frac{\eta k_b T_c}{q}\right)$ (Piliougine, et al., 2020; Wilcox, et al., 2011)	(6)
$E_g(T_c) = E_{g,stc}\left(1 - \rho\cdot(T_c - T_{stc})\right)$ (De Soto, et al., 2006; Petrone, et al., 2017)	(7)
$R_s(T_c) = R_{s,ref} + \kappa \cdot (T_c - T_{ref})$ (Piliougine, et al., 2021)	(8)
$R_{sh}(G) = R_{sh,base} + (R_{sh,0} - R_{sh,base}) \cdot \exp\left(-R_{sh,exp}\frac{G}{G_{ref}}\right)$ (Sauer, et al., 2014)	(9)

Assuming that μ_{Isc} [AK^{-1}] is the temperature coefficient of the short-circuit current, the photo-generated current I_{ph} is a function of the irradiance G and cell temperature T_c, as it can be seen in Eq. (5). The saturation current I_s defined in Eq. (6) varies with respect to T_c and to the bandgap energy E_g; C_0 is an experimental fitting parameter. The energy bandgap E_g [eV] in Eq. (7) depends on the temperature and cell technology, *as* well as ρ [K^{-1}]. Moreover, the series resistance R_s is assumed to depend on cell temperature as expressed in Eq. (8), where κ [ΩK^{-1}] is a constant whose value is identified with a fitting procedure applied on experimental data. On the other side, the shunt resistance R_{sh} is calculated using Eq. (9), where:

$$R_{sh,base} = \frac{R_{sh,ref} - R_{sh,0}\exp(-R_{sh,exp})}{1 - \exp(-R_{sh,exp})} \tag{10}$$

and $R_{sh,0}$ and $R_{sh,exp}$ can be determined from experimental data. The assessment of the cell temperature T_c requires accurate measuring sensors, since it depends on many factors (cell technology, type of installation of the PV module, wind speed of the site, etc.). The empirical relationship proposed in (Kratochvil, et al., 2004) is used in this paper to calculate the cell temperature T_c by means of the backside temperature of the PV module, T_b:

$$T_c = T_b + \frac{G}{G_{stc}}\cdot\Delta T \tag{11}$$

where G_{stc} = 1 000 Wm^{-2} is the standard irradiance and ΔT [K] is the difference between internal and external temperature at the standard irradiance. The value of ΔT depends on the cell technology, encapsulating materials, and type of installation. It can be computed by means of a fitting procedure on experimental measurements of the module backside and cell temperatures, from both external and internal sensors (last one in direct contact to the cell).

3. Overview of PV system faults

Faults in PV modules can be categorized as: physical (damages to the glass, frame or bypass diodes, cracks in PV cells, degradation, etc.); environmental (temporary shading, permanent shading, hot-spots); electrical (short-circuit, open-circuit, line-line faults either in PV modules, arrays, or in the whole PV system). The types of faults that are addressed in the present work, are discussed in the following.

3.1. Physical Faults

3.1.1. Degradation of PV module. Degradation faults are due to abnormal ageing occurring in PV panels and they can give rise to pronounced increase in series resistance R_s, decrease in fill factor (FF) as well as increased heating in the module and localized hot-spot (Ahmad, et al., 2019). In (Spataru, et al., 2016) it is shown that the cell degradation occurs due to different reasons like poor soldering, mechanical stress, thermal cycling, or moisture ingress (that can cause corrosion of the module connectors); some of them manifested as a growth in the PV module series resistance and power loss. Another source of degradation of a PV module is the deterioration of the anti-reflective coating with time and this is highlighted by the change in its color from white to yellow and sometimes then to brown. This process leads to the permanent faults of Yellowing and Browning. All the mentioned issues may occur also in an early phase of the PV panel lifespan due to humidity and high temperature (Spataru, et al., 2016) and may leave visual impact on the surface. Thus, they can be easily detected through visual inspection.

3.2. Electrical Faults

3.2.1. Short-circuit and Open-circuit. Short-circuit and open- circuit faults are mainly related to the bypass diodes. If a PV module is equipped with three bypass diodes, the underperformance is 1/3 of the maximum power and of open-circuit voltage in case of a short-circuited bypass diode, while an open-circuited bypass diode does not have any significant effect of the operation of the PV panel if it is not shaded (Ahmad, et al., 2019). This fault can also be related to PV modules, when disconnection problems appear in one PV string or more. Most of the disconnection problems are due to ageing after a long-term operation and poor soldering in string interconnections (Spataru, et al., 2016). The open-circuit fault leads to a decline of the current flowing in that string and a loss of power, depending on the number of disconnected strings. The output voltage remains stable assuming that all the solar panels have the same technical specifications and are under the same level of irradiance and temperature (Vallejo, et al., 2019). On the other side, short-circuit fault significantly reduce the voltage and the maximum power delivered to the load, while the current is not affected.

3.2.2. Potential Induced Degradation. Potential induced degradation is related to high voltage gradients between PV cells and the module frame. Due to the very short distance between solar cells and the frame, and the possible presence of impurities in the encapsulant material, the formation of leaking currents may occur. These currents can flow from the PV cells to the frame (positively biased PV cell) or vice versa (negatively biased PV cell), causing a strong reduction in shunt resistance, as well as lower V_{oc} and FF (Ahmad, et al., 2019). Moreover, in (Spataru, et al., 2016) is observed that an important I_{mpp} current loss and FF degradation is commonly associated with PID, while the current generation (I_{sc}) is not significantly affected, even for relatively large extents of degradation. However, this can change R_{sh} is drastically reduced, which can be the case in highly PID-sensitive modules.

4. Proposed fault detection technique

In the following, a detailed explanation of the proposed FDD technique based on a four-layered feedforward ANN for PV fault classification is provided, as well as a more specific description of the PV device under study.

4.1. PV module under study

Several *Isofotón* I-53 PV modules were installed on the terrace of the Laboratory of Photovoltaic Systems of the University of Málaga (Spain). The modules include 36 pseudo square single-crystalline silicon (sc-Si) cells connected in series, with two bypass diodes on an overlapped configuration (Ziar, et al., 2014). The main specifications of *Isofotón* I-53 provided by the manufacturer under STC conditions (characterized by a cell temperature equal to 25 °C and an irradiance of 1 000 Wm^{-2}) are listed in Table 2.

One of these modules was available for the experimental measurements. However, it is not in its infant phase, because it has been operating since 1996 and experienced ageing phenomena. The current electrical parameters of this specific PV panel are obtained by applying a parameter identification procedure at a reference condition characterized by G_{ref} = 807 Wm^{-2} and T_{ref} = 315.95 K. The values of G_{ref} and T_{ref} are selected so that the reference condition is close to NOCT conditions (an irradiance G = 800 Wm^{-2} and an ambient temperature T_{amb} = 20 °C). The values of the identified parameters are reported in Table 3.

For the PV panel under study, the parameters μ_{Isc}, C_0, κ, $R_{sh,0}$, and $R_{sh,exp}$ are estimated by fitting the experimental data. More precisely, μ_{Isc} = 0.00092 AK^{-1}, C_0 = 38.53, κ = 0.0043 ΩK^{-1}, $R_{sh,0}$ = 172.76 Ω, and $R_{sh,exp}$ = 1.385. In addition, ρ and $E_{g,stc}$ depend on the cell technology. For sc-Si technology, ρ = 0.0002677 K^{-1} and $E_{g,stc}$ = 1.121 eV, being T_{stc} = 298.15 K (25 °C). Finally, a value ΔT = 5.6 K was estimated previously for a similar PV module (same manufacturer and technology) installed also on an open-rack structure.

Table 2
Main characteristics of module Isofotón I-53

Technology	monocrystalline Si
Length	1304 mm
Width	340 mm
Thickness	39.5 mm
Weight	5.5 kg
N_s (number of cells, in series)	36
I_{sc} (short-circuit current)	3.27 A
V_{oc} (open-circuit voltage)	21.6 V
P_{mpp} (maximum power)	53 W
I_{mpp} (current at max. power)	3.05 A
V_{mpp} (voltage at max. power)	17.4 V
NOCT	47 °C

Table 3
Values of SDM parameters at reference condition

$I_{ph,ref}$ [A]	$I_{s,ref}$ [A]	$R_{s,ref}$ [Ω]	$R_{sh,ref}$ [Ω]	η_{ref} [-]
2.166	2.2896e−7	0.348	109.7	1.23

Table 4
Selected operating conditions of the PV module

Class symbol	Class description and affected parameters	Variation range
F1	Healthy state: Natural distortion of the *I-V* curves due to daily variation of environmental factors G and T_c	
F2	Abnormal increase of series resistance (R_s, I_{mpp}, V_{mpp})	$R_s \rightarrow$ +[20%, 200%]
F3		$R_s \rightarrow$ +[200%, 350%]
F4		$R_s \rightarrow$ +[350%, 500%]
F5	Abnormal decrease of shunt resistance (R_{sh}, I_{mpp}, V_{mpp})	$R_{sh} \rightarrow$ +[30%, 50%]
F6		$R_{sh} \rightarrow$ +[50%, 70%]
From F7 to F12	Combination of F2, F3, F4, F5, F6	
F13, F14	Short-circuit in 1 or 2 bypass diodes (V_{oc}, V_{mpp})	
F15, F16	Open-circuit of 1 or 2 sub-modules (I_{sc} , I_{mpp})	

4.2. Operating conditions

Sixteen operating conditions are selected for the classification problem at hand, one corresponding to a healthy state of the PV modules, the others corresponding to different faulty states, as listed in Table 4. Class F1 denotes the healthy state. Classes F2, F3 and F4 are related to an abnormal increase of the series resistance R_s. More in detail, each of these three classes corresponds to a different range of variation of R_s with respect to its original value. The same procedure is used for classes F5 and F6, denoting an abnormal reduction of the shunt resistance R_{sh}. Finally, classes from F7 to F12 correspond to combinations of the previous faults. Indeed, the ANN model should be able to correctly classify not only the type of fault and its entity, but also combinations of faults, i.e., when faults affecting series and shunt resistances occur simultaneously.

A synthetic dataset of *I-V* curves corresponding to each class is generated using the equations from Eq. (1) to Eq. (11). Fig. 1 shows some examples of the generated synthetic *I-V* curves. The simulation of these curves is achieved by intentionally acting on the SDM parameters affected by each type of failure mechanism. Moreover, different levels of irradiance and cell temperature are considered, to reproduce the electrical characteristics in different environmental conditions. Specifically, the backside temperature T_b of the module and the irradiance G range between 30 and 45 °C, and between 300 and 900 Wm^{-2} respectively. The T_c range corresponding to the assigned T_b one is computed through Eq. (11).

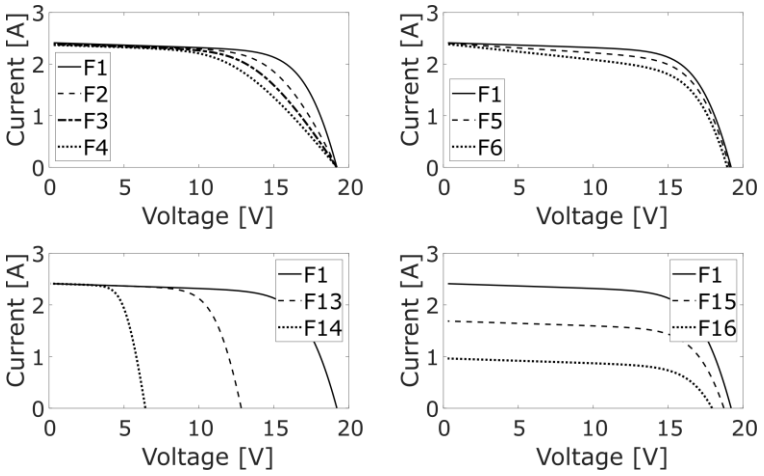

Fig. 1. Effect of the different faults on the *I-V* curves: (top-left) healthy state vs abnormal variation of R_s, (top-right) healthy state vs abnormal variation of R_{sh}, (bottom-left) healthy state vs short-circuit fault, (bottom-right) healthy state vs open-circuit fault.

4.3. Sampling method

Each curve is sampled in a discrete number of points. The choice of the sampling method is of primary importance since the inputs of the network are the sampled *I-V* curves. Thus, the more uniformly distributed are the sampling points, the better the curve shape reproduction will be, and a more accurate classification is expected from the ANN.

As a first attempt, the simulated *I-V* curves are sampled at a fixed voltage step ΔV. As expected, the part of the *I-V* curve that approaches the open circuit voltage is more coarsely sampled, because of its high slope at high voltage values (see Fig. 3a). The same problem, even more evident, would arise if a constant current step is used for the *I-V* curve sampling. In order to have a better sampling of the synthetic curves, a different approach is adopted, and it can be summarized in three main steps.

1st step: normalization. For each *I-V* curve generated by Eq. (1), the value of I_{sc} is calculated by setting $V = 0$ V and V_{oc} is computed by means of Eq. (12):

$$V_{oc} = \eta V_{th} T_c \Big/ \log\left(1 + \frac{I_{ph}}{I_s}\right) \tag{12}$$

Then, the parameters defining Eq. (1) are normalized as reported in Eq. (13), where the subscript *n* denotes the normalized parameters. This guarantees that the values of current *I* and voltage *V* fall within the range [0,1]:

$$R_{s,n} = R_s \frac{I_{sc}}{V_{oc}} \qquad R_{sh,n} = R_{sh} \frac{I_{sc}}{V_{oc}} \qquad I_{ph,n} = \frac{I_{ph}}{I_{sc}} \tag{13}$$

$$I_{s,n} = \frac{I_s}{I_{sc}} \qquad V_{th,n} = N_s \frac{k_b T_{stc}}{q V_{oc}}$$

2nd step: sampling. The normalized *I-V* curves are then sampled at their intersection with the straight lines defined by Eq. (14):

$$I = tan(\alpha) \cdot V \tag{14}$$

where α varies in the range [0.1°, 89.9°] with a step $\Delta\alpha = 0.1°$. In Fig. 2 a normalized *I-V* curve and some of the generated straight lines are depicted.

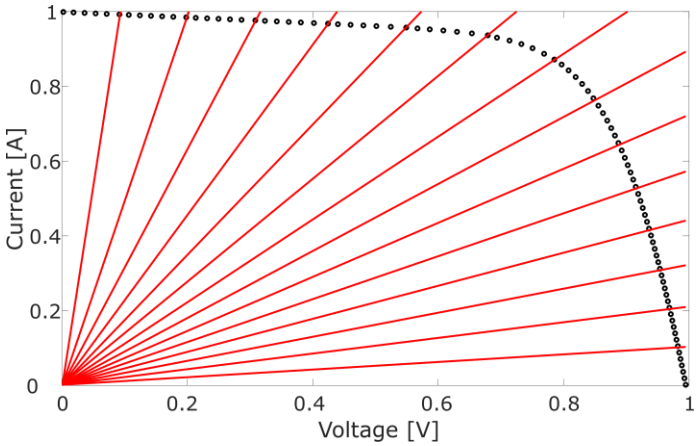

Fig. 2. Sampling of the normalized *I-V* curve.

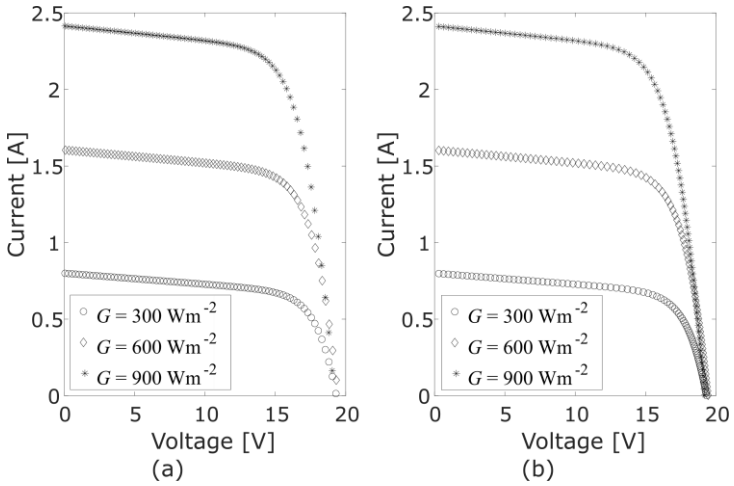

Fig. 3. Comparison between the distribution of sampling points obtained by applying the sampling method with constant ΔV (a) or using the proposed approach (b).

3rd step: de-normalization. Finally, the discretized *I-V* curves are de-normalized by scaling up the generated samples by I_{sc} and V_{oc} . Fig. 3b shows the final results of this procedure. It can be noticed that it ensures a uniform distribution of sampling points, thus a better description of the synthetic *I-V* curves.

4.4. ANN tuning and training

Once the synthetic *I-V* curves related to the selected operating conditions are collected, they are fed into the network for training. The main parameters defining the final neural network model are listed in Table 5. Furthermore, a sensitivity analysis is performed to choose the network size (i.e., number of hidden neurons) and the size of the training set.

Table 5
ANN parameters

Parameter	Value
Training algorithm	Scaled Conjugate Gradient
Activation functions	Hyperbolic Tangent (Hidden Layers) SoftMax (Output Layer)
Number of hidden layers	2
Max number epochs	1200
Learning Rate	0.4
Loss function	Cross-entropy

4.4.1. Hidden Neurons. The objective is to tune the size of the ANN by setting the number of neurons in each hidden layer, depending on the complexity of the classification task and on the amount of training data (Dębska and Guzowska-Świder, 2011). For the input and output layers the number of neurons is automatically determined by the dimension of the input matrix and the number of output classes. Specifically, the input layer is made up of 180 neurons, since this is the number of coordinates of sampling points for each input curve (90 voltage coordinates, 90 current coordinates); the output layer is characterized by 16 neurons, as the number of classes. Concerning the number of hidden nodes, a sensitivity analysis is conducted. The initial adopted configuration consists of 50 neurons in each hidden layer. Then, the network is expanded by adding more nodes, in order to analyze the effect of this hyperparameter on the classifier accuracy. Several configurations are examined, but for the sake of brevity, in the next section only the results related to two particular cases are reported. They are characterized by 50 and 100 neurons for each hidden layer respectively and they will be referred with the notation [180,50,50,16] and [180,100,100,16].

4.4.2. Dataset Size. The effect of the training set size on the classifier accuracy is also investigated. As reported in (Sug, 2009), in general the accuracy of the classifier increases with the size of the training set and the performance follows a logarithmic trend. Thus, a second sensitivity analysis is performed, comparing the accuracy of the algorithm when trained with 300, 400 and 500 curves for each class. During the learning phase, the 70% of synthetic *I-V* curves for each class is randomly selected for training, 15% for validation and the remaining for testing. This splitting is very important as it allows to examine the network prediction performance on new instances and avoid overfitting. To summarize, six configurations are considered by varying the number of neurons in both hidden layers (50/100) and the size of the training set (300/400/500 per class).

5. Training results

Table 6 shows the training results in terms of average training accuracy, average training time and final value of the loss function for the six configurations. The training procedure run on a Dell Inspiron 13 5000 with a microprocessor Intel(R) Core (TM) i7-8550U, 256 GB of storage, and 16 GB of principal memory.

As expected, the training accuracy reaches an all-time high in correspondence of a training set with 500 samples per class and hidden layers with 100 neurons each. In this case the training accuracy is equal to 98.55%. This value drops to 98.10% and 97.92% if 400 and 300 curves per class are considered, respectively. A similar trend but with slightly lower accuracy values is observed when 50 neurons per hidden layer are used. These results confirm what is shown in (Sug, 2009). It could seem surprising that the higher is the number of hidden neurons, the shorter is the training time. However, it is well known that with a higher number of neurons, a lower number of epochs are required to reach the convergence.

Table 6
Main performance parameters of the classifier obtained from the sensitivity analysis

Dataset size	Network architecture	Avg. training accuracy	Avg. training time[min]	Avg. final value of the cross-entropy
500×16	[180 50 50 16]	98.25%	2.40	0.008881
	[180 100 100 16]	98.55%	2.22	0.007844
400×16	[180 50 50 16]	98.02%	1.40	0.009972
	[180 100 100 16]	98.10%	1.13	0.009003
300×16	[180 50 50 16]	97.85%	0.94	0.010520
	[180 100 100 16]	97.92%	0.83	0.009387

From the results, it is clear that the best performing configuration is the one characterized by two hidden layers with 100 units each and trained with 500 training samples per class. Indeed, it achieves the highest classification accuracy, and it has an acceptable training time around 2 min. It must be pointed out that the ANN training phase is assumed to be done off-line, while the classification of a new experimental *I-V* curve can be done on-line since it requires a low computational effort. The algorithm achieves convergency at the 690th epoch with a cross-entropy of 0.0075196. In this configuration, the true positive rate for each class are reported in the bar chart in Fig. 4 . The ANN achieves an effective classification accuracy for all the selected faulty operating conditions. Indeed, the minimum accuracy is 92.4% and it is obtained for class F8, while the maximum value is 100% for classes F1, F2, F4, F5, F13, F14, F15, F16. The confusion matrix related to the training phase is depicted in Fig. 5.

Two additional quality indicators used for multi-class classification problems are computed in Eq. (15) and Eq. (16), i.e., Macro Average Precision and Recall, to assess the quality of the classification. They are the arithmetic means of precision and recall evaluated for single classes (Grandini, et al., 2020):

$$MacroAveragePrecision = \frac{\sum_{k=1}^{16} Precision_k}{16} = 98.52\% \tag{15}$$

$$MacroAverageRecall = \frac{\sum_{k=1}^{16} Recall_k}{16} = 98.52\% \tag{16}$$

Thus, the computation of macro F1-score is:

$$Macro\ F1 - Score = 2 \cdot \left(\frac{MacroAveragePrecision \cdot MacroAverageRecall}{MacroAveragePrecision + MacroAverageRecall} \right) = 98.52\% \tag{17}$$

5.1. Experimental set

The ANN trained on synthetic data is also tested by using experimental *I-V* curves acquired from a real operating PV field at the University of Málaga, Spain. The experimental curves fall into categories F1, F2, F3 or F4, characterized by an increase of the module series resistance. These faulty conditions were reproduced using four additional resistors with a known value connected in series with the PV module. Unfortunately, the emulation of the faults corresponding to the other classes is not possible since no experimental *I-V* curves for those classes are currently available. In Fig. 6, the four groups of experimental curves and the corresponding values of the additional series resistances are presented.

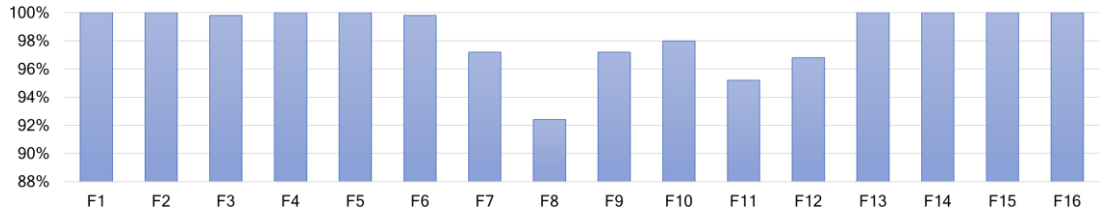

Fig. 4. Comparison of the per-class true positive rates.

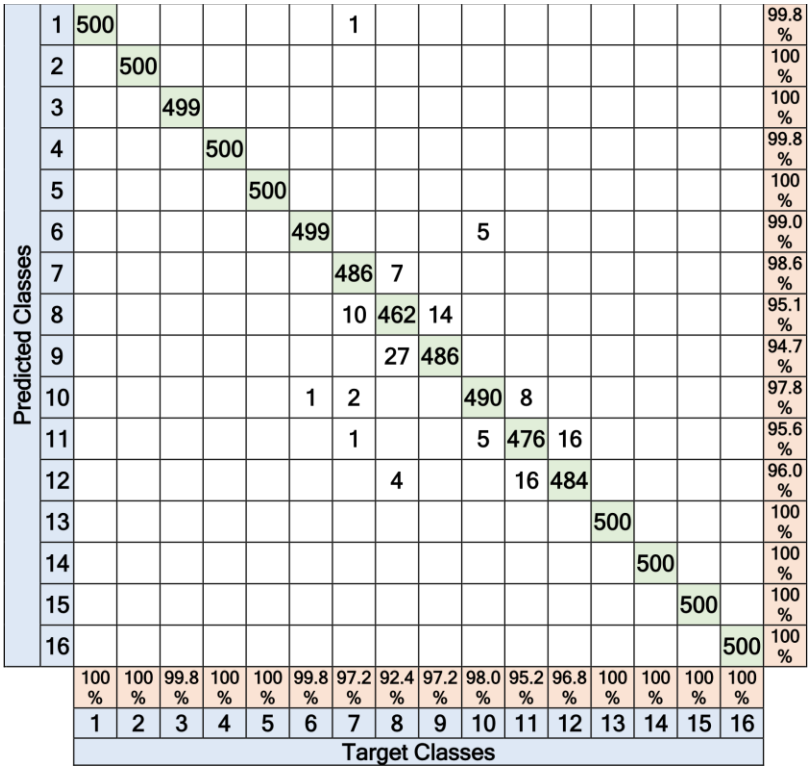

Fig. 5. Confusion matrix.

The experimental curves are fed into the ANN to check its classification ability on real cases. For each group, the number of curves, the value of the additional series resistance and its percentage increase with respect to the reference value (the internal series resistance of the module), as well as the target class and the percentage of correctly classified samples are listed in Table 7.

The best cases are related to the 1st group and the 3rd group, with a classification accuracy of 100%, while the results are slightly worse for the 2nd group, characterized by a $\Delta R = 0.3\ \Omega$. Indeed, in this case the 96.2% of these curves are correctly labelled as class F2 and the only exception is one curve labelled as class F7. This misclassification can be justified by the fact that the misclassified experimental *I-V* curve is characterized by a high level of noise due to the on-field measurements, as it is highlighted in Fig. 7.

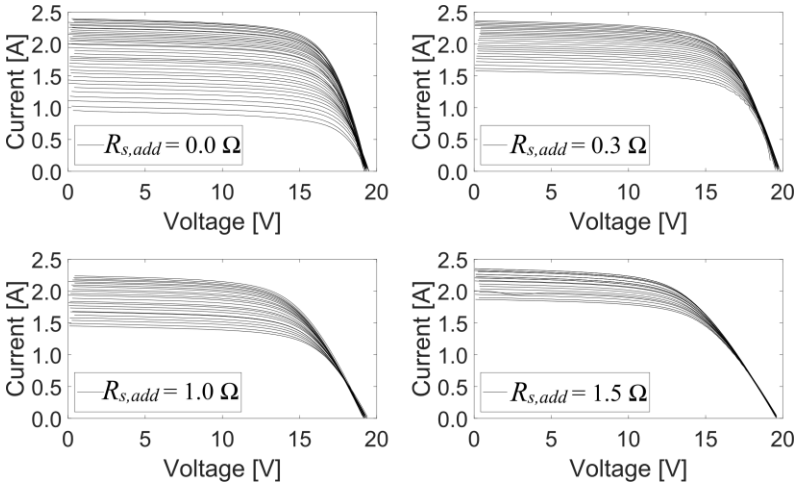

Fig. 6. Experimental *I-V* curves: 1st group (top left), 2nd group (top right), 3rd group (bottom left), 4th group (bottom right).

Table 7
Experimental test set and classification results

	Number of curves	**ΔR [Ω]**	**ΔR%** $R_{s,ref} = 0.348\ \Omega$	*Target class*	*True positive rate*
1st group	38	0.0	0%	1	100.0%
2ndgroup	26	0.3	86%	2	96.2%
3rd group	24	1.0	287%	3	100.0%
4th group	22	1.5	431%	4	90.9%

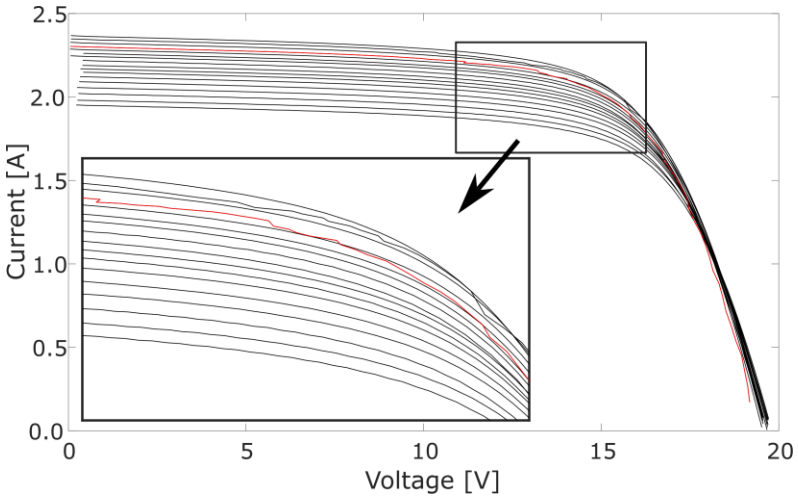

Fig. 7. Part of the experimental *I-V* curves belonging to the 2nd group. The black curve is misclassified due to its high noise.

Concerning the 4th group of experimental *I-V* curves, characterized by a $\Delta R = 1.5\ \Omega$, the 9% of them are wrongly labelled as class F3 instead of class F4, this meaning 2 out of 22 elements. In general, this misclassification is reasonable since curves belonging to class F3 and class F4 are simulated by varying the series resistance R_s of the SDM model in two adjacent variation ranges that are: $R_s \rightarrow +[200\%, 350\%]$ and $R_s \rightarrow +[350\%, 500\%]$.

It goes without saying that *I-V* curves belonging to class F3 and corresponding to the upper extreme value of R_s interval are very similar to *I-V* curves belonging to class F4 and corresponding to the lower extreme value of R_s interval. The difference between them is hardly noticeable visually (as it can be seen in Fig. 7), and for this reason the model sometimes misclassifies them. However, also in this case the misclassification of one of the two curves is due to the high intensity of the noise characterizing the on-field measurements.

These results clearly demonstrate that the developed classifier is able to effectively identify faults and to generalize on new samples.

6. Conclusions

The present work proposes an ANN-based method for the detection and classification of sixteen different operating status of a PV panel (healthy state, abnormal variation of R_s, abnormal variation of R_{sh}, combination of both, short-circuit fault and open-circuit fault) under different environmental conditions in terms of irradiance G and cell temperature T_c. The ANN is a four-layered feedforward neural network trained with the Scaled Conjugate Gradient backpropagation algorithm. A novel approach is proposed, since the classifier operates in the *I-V* plane and takes as input the duly sampled *I-V* curves related to healthy and faulty conditions of the PV module. The points are obtained by sampling the *I-V* curves with a normalization and denormalization procedure to ensure the most accurate description of *I-V* curves shape in different faulty conditions.

Two sensitivity analysis are conducted: the first aimed at establishing the best architecture of the network (i.e., the optimal number of hidden neurons), while the second at determining the optimal training set size (i.e., number of per-class samples in the training set). The results show that the network configuration with two hidden layers of 100 neurons each and trained with 500 training samples per class best fits the classification problem since it achieves a very high training accuracy (98.55%), as well as an acceptable training time.

Additionally, the model is validated with experimental *I-V* curves related to two of the selected operating conditions (healthy state, abnormal variation of R_s). The algorithm exhibits a minimum classification accuracy of 90.5% for the group of experimental curves affected by the highest entity of the fault. Indeed, some of these curves are misclassified due to the high intensity of noise affecting the experimental on-field measurement. For the groups with the minimum and intermediate entity of the fault, a maximum accuracy of 100% is reached, thus all the available curves are correctly identified.

Acknowledgments:

The authors would like to acknowledge the availability of the measurements provided by the research group of the *Laboratory of Photovoltaic Systems* of the *University of Málaga* led by Prof. Mariano Sidrach-de-Cardona.

Funding:

This work was supported by: *Gobierno de España. Ministerio de Ciencia, Innovación y Universidades* (Spain) [grant number RTI2018–095097–B–I0]; *Ministero dell'Istruzione, dell'Università e della Ricerca* (Italy) [grant 2017WA5ZT3003]; *Università degli Studi di Salerno* [FARB funds].

References

Ahmad, J., Ciocia, A., Fichera, S., Murtaza, A. F., Spertino, F., 2019. Detection of typical defects in silicon photovoltaic modules and application for plants with distributed MPPT configuration. *Energies, 12*(23), 4547. https://doi.org/10.3390/en12234547.

Anani, N., Ibrahim, H., 2020. Adjusting the Single-Diode Model Parameters of a Photovoltaic Module with Irradiance and Temperature. *Energies, 13*(12), 3226. https://doi.org/10.3390/en13123226

Aziz, F., Haq, A. U., Ahmad, S., Mahmoud, Y., Jalal, M., Ali, U., 2020. A novel convolutional neural network-based approach for fault classification in photovoltaic arrays. *IEEE Access, 8*, 41889-41904. https://doi.org/10.1109/ACCESS.2020.2977116.

Basnet, B., Chun, H., Bang, J., 2020. An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems. *J Sensors, 2020*, 6960328. https://doi.org/10.1155/2020/6960328.

Chine, W., Mellit, A., Lughi, V., Malek, A., Sulligoi, G., Pavan, A. M., 2016. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. *Renew Energ, 90*, 501-512. https://doi.org/10.1016/j.renene.2016.01.036.

De Soto, W., Klein, S. A., Beckman, W. A., 2006. Improvement and validation of a model for photovoltaic array performance. *Sol Energy, 80*(1), 78-88. https://doi.org/10.1016/j.solener.2005.06.010.

Dębska, B., Guzowska-Świder, B., 2011. Application of artificial neural network in food classification. *Anal Chim Acta, 705*(1-2), 283-291. https://doi.org/10.1016/j.aca.2011.06.033.

Driss, S. B., Soua, M., Kachouri, R., Akil, M., 2017. A comparison study between MLP and convolutional neural network models for character recognition. *Proc SPIE. Conference on Real-Time Image and Video Processing. International Society for Optics and Photonics. Anaheim (CA, USA), 10223*, 1022306. https://doi.org/10.1117/12.2262589.

Garoudja, E., Chouder, A., Kara, K., Silvestre, S., 2017. An enhanced machine learning based approach for failures detection and diagnosis of PV systems. *Energ Convers Manage, 151*, 496-513. https://doi.org/10.1016/j.enconman.2017.09.019.

Grandini, M., Bagli, E., Visani, G., 2020. Metrics for Multi-Class Classification: an Overview. *arXiv:2008.05756*. https://arxiv.org/pdf/2008.05756.pdf.

IEA., 2021. *Global Energy Review 2021. Assessing the effects of economic recoveries on global energy demand and CO_2 emissions in 2021*. Paris (France), France: International Energy Agency IEA. https://www.iea.org/reports/global-energy-review-2021.

IRENA., 2021. *Renewable Capacity Statistics 2021.* Abu Dhabi (UAE), EAU: International Renewable Energy Agency IRENA. https://irena.org/publications/2021/March/Renewable-Capacity-Statistics-2021.

Kratochvil, J. A., E., B. W., King, D. L., 2004. Photovoltaic array performance model. SAND2004-3535. *Sandia National Laboratories.* https://doi.org/10.2172/919131.

Petrone, G., Ramos-Paja, C. A., Spagnuolo, G., 2017. *Photovoltaic sources modeling.* John Wiley & Sons. ISBN: 978-1-118-67903-6. https://doi.org/10.1002/9781118755877.

Piliougine, M., Oukaja, A., Sánchez-Friera, P., Petrone, G., Sánchez-Pacheco, F. J., Spagnuolo, G., Sidrach-de-Cardona, M., 2021. Analysis of the degradation of single-crystalline silicon modules after 21 years of operation. *Prog Photovoltaics*. https://doi.org/10.1002/pip.3409.

Piliougine, M., Spagnuolo, G., Sidrach-de-Cardona, M., 2020. Series resistance temperature sensitivity in degraded mono–crystalline silicon modules. *Renew Energ, 162*, 677-684. https://doi.org/10.1016/j.renene.2020.08.026.

Sabbaghpur Arani, M., Hejazi, M. A., 2016. The comprehensive study of electrical faults in PV arrays. *J Electr Comput Eng, 2016*, 8712960. https://doi.org/10.1155/2016/8712960.

Sauer, K. J., Roessler, T., Hansen, C. W., 2014. Modeling the irradiance and temperature dependence of photovoltaic modules in PVsyst. *IEEE J Photovoltaics, 5*(1), 152-158. https://doi.org/10.1109/JPHOTOV.2014.2364133.

Sera, D., Teodorescu, R., Rodriguez, P., 2007. PV panel model based on datasheet values. *IEEE International Symposium on Industrial Electronics ISIE2007. Vigo (Spain)*, 2392-2396. http://doi.org/10.1109/ISIE.2007.4374981.

Spataru, S. V., Sera, D., Hacke, P., Kerekes, T., Teodorescu, R., 2016. Fault identification in crystalline silicon PV modules by complementary analysis of the light and dark current–voltage characteristics. *Prog Photovoltaics, 24*(4), 517-532. https://doi.org/10.1002/pip.2571.

Stein, J. S., Cameron, C. P., Bourne, B., Kimber, A., Posbic, J., Jester, T., 2010. A standardized approach to PV system performance model validation. *35th IEEE Phot Spec Conf. Honoluu (HI,USA)*, 001079-001084. https://doi.org/10.1109/PVSC.2010.5614696.

Sug, H., 2009. Performance Comparison of RBF networks and MLPs for Classification. *9th WSEAS Int C Appl Inform C. Stevens Point (WI, USA)*, 450-454. https://doi.org/10.5555/1628095.1628170.

Vallejo, A. N., Ruiz, F., Patiño, D., 2019. Characterization of electric faults in photovoltaic array systems. *DYNA, 86*(211), 54-63. https://doi.org/10.15446/dyna.v86n211.79085.

Wilcox, J. R., Haas, A. W., Gray, J. L., Schwartz, R. J., 2011. Estimating Saturation Current Based on Junction Temperature and Bandgap. *AIP Conf Proc, 1407*, 30. https://doi.org/10.1063/1.3658288.

Zhu, H., Lu, L., Yao, J., Dai, S., Hu, Y., 2018. Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model. *Sol Energy, 176*, 395-405. https://doi.org/10.1016/j.solener.2018.10.054.

Ziar, H., Nouri, M., Asaei, B., Farhangi, S., 2014. Analysis of Overcurrent Occurrence in Photovoltaic Modules With Overlapped By-Pass Diodes at Partial Shading. *IEEE J Photovoltaics, 4*(2), 713-721. https://doi.org/10.1109/JPHOTOV.2013.2292578.