



HAL
open science

Fault diagnosis of photovoltaic panels using full I–V characteristics and machine learning techniques

Baojie Li, Claude Delpha, Anne Migan-Dubois, Demba Diallo

► **To cite this version:**

Baojie Li, Claude Delpha, Anne Migan-Dubois, Demba Diallo. Fault diagnosis of photovoltaic panels using full I–V characteristics and machine learning techniques. *Energy Conversion and Management*, 2021, 248, pp.114785. 10.1016/j.enconman.2021.114785 . hal-03415367

HAL Id: hal-03415367

<https://hal.science/hal-03415367>

Submitted on 4 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fault Diagnosis of Photovoltaic Panels Using Full I-V Characteristics and Machine Learning Techniques

Baojie LI^{1,2}, Claude DELPHA², Anne MIGAN-DUBOIS¹, Demba DIALLO^{1*},

¹ Université Paris-Saclay, CentraleSupélec, CNRS, GeePs, Sorbonne Université, 3-11 Rue Joliot Curie, Gif Sur Yvette, 91192, France

² Université Paris-Saclay, CNRS, CentraleSupélec, L2S, 3 Rue Joliot Curie, Gif Sur Yvette, 91192, France

* Corresponding author: demba.diallo@centralesupelec.fr

Abstract: The current-voltage characteristics (I-V curves) of photovoltaic (PV) modules contain a lot of information about their health. In the literature, only partial information from the I-V curves is used for diagnosis. In this study, a methodology is developed to make full use of I-V curves for PV fault diagnosis. In the pre-processing step, the I-V curve is first corrected and resampled. Then fault features are extracted based on the direct use of the resampled vector of current or the transformation by *Gramian angular difference field* or *recurrence plot*. Six machine learning techniques, i.e., *artificial neural network*, *support vector machine*, *decision tree*, *random forest*, *k-nearest neighbors*, and *naive Bayesian classifier* are evaluated for the classification of the eight conditions (healthy and seven faulty conditions) of PV array. Special effort is paid to find out the best performance (accuracy and processing time) when using different input features combined with each of the classifier. Besides, the robustness to environmental noise and measurement errors is also addressed. It is found out that the best classifier achieves 100% classification accuracy with both simulation and field data. The dimension reduction of features, the robustness of classifiers to disturbance, and the impact of transformation are also analyzed.

Keywords: Photovoltaic; fault diagnosis; I-V curve; feature transformation; I-V curve correction; machine learning;

Nomenclature			
Terminology		Symbols	
1D, 2D	1 Dimension, 2 Dimension	a	Irradiance correction factor
ANN	Artificial Neural Network	α_{rel}	Relative temperature coefficient of I_{SC} (%/°C)
CNN	Convolutional Neural Network	β_{rel}	Relative temperature coefficient of V_{OC} (%/°C)
DT	Decision Tree	τ	Shift of phase space
FDD	Fault Detection and Diagnosis	E_{area}	Error of area (%)
FF	Fill Factor	D_{matrix}	Number of features of matrix
GADF	Gramian Angular Difference Field	D_{vector}	Number of features of vector
I-V curve	Current-voltage Characteristic	G	Global in-plane irradiance (W/m ²)
kNN	k-Nearest Neighbors	I	Current (A)
MLT	Machine Learning Technique	I_{MPP}	Current at MPP (A)
MPP	Maximum Power Point	I_{SC}	Short-circuit current (A)
NBC	Naive Bayesian Classifier	κ	Curve correction factor
OC	Open circuit	N	Number of resampling points
PCA	Principal Component Analysis	V	Voltage (V)
PS	Partial Shading	V_{MPP}	Voltage at MPP (V)
PV	Photovoltaic	V_{OC}	Open-circuit voltage (V)
RF	Random Forest	P_m	Maximum Power (W)
RMSE	Root Mean Squared Error	R_{OC}	Resistance used in simulation for OC (Ω)
RP	Recurrence Plot	R_s	Series resistance (Ω)
SC	Short Circuit	R_{s_degra}	Resistance used in simulation for R_s degradation (Ω)
STC	Standard Test Condition	R_{SC}	Resistance used in simulation for SC (Ω)
SVM	Support Vector Machine	R_{sh}	Shunt resistance (Ω)
TC	Temperature Coefficient	R_{sh_degra}	Resistance used in simulation for R_{sh} degradation (Ω)
t-SNE	t-Distributed Stochastic Neighbour Embedding		

$S_{original}$	Area size enclosed by the original I-V curve
$S_{Area\ of\ error}$	Area size enclosed by the difference between the original and the resampled I-V curve
T_m	Module temperature ($^{\circ}C$)

28
29

30 1 Introduction

31 The solar photovoltaic (PV) installed capacity has experienced rapid growth among all the main energy types
32 in recent years [1]. However, due to the environmental threats and the potential damages in the manufacturing,
33 transportation, installation, or other procedures, various PV faults may arise up [2,3]. These faults could not
34 only affect the power generation but also could cause severe safety hazards [4], e.g., fire risk, electrical shock,
35 physical danger. Therefore, it is necessary to perform effective detection and diagnosis of PV faults.

36 Common PV electrical data used for diagnosis include different types: output power, output voltage or current
37 at DC or AC side, and current-voltage characteristic (I-V curve) [5]. Since an I-V curve generally embeds rich
38 information about the health status of PV modules, I-V curve-based diagnosis is a popular topic [6]. As for
39 acquiring I-V curves, common I-V tracers already support the measurement for a single module or small-scale
40 string or array. In recent years, the hardware solutions (integrated at inverter level) have become commercially
41 available to measure I-V curves periodically at the power plant level [7,8]. In this sense, diagnosis approaches
42 based on the I-V curves could be applied to all common PV facilities.

43 In the literature, there are different methods of using I-V curves to extract information. The most common can
44 be grouped into three categories: (1) Extract key features from the curve (like V_{OC} , I_{SC} , V_{MPP} , I_{MPP} , FF, R_s , and
45 R_{sh}) and then use these features for diagnosis (by threshold analysis, statistical approaches, or machine learning
46 techniques (MLT), etc.). For example, Chine et al. [9] use V_{OC} , I_{SC} , V_{MPP} , and I_{MPP} features to detect partial
47 shading (PS), SC, open-circuit (OC), and bypass diode failures. Both threshold method and Artificial Neural
48 Network (ANN) are applied to classify the faults. Huang et al. [10] extract V_{OC} , V_{MPP} , I_{MPP} and R_s from I-V
49 curves to classify short-circuit (SC) fault, PS and degradation. CART decision tree is used as the diagnosis
50 tool. Fadhel et al. [11] adopt V_{MPP} , I_{MPP} and P_{MPP} as features to detect PS. Dhimish et al. [12] use thresholds to
51 analyze V_{MPP} and the voltage drop on the I-V curves to identify PS. Similar approaches can also be found in
52 [13–16]. It should be noted that these studies use only partial information from the I-V curves. (2) Calculate
53 the first or second-order derivative of the curve or identify steps on the I-V curves. For example, Ma et al. [17]
54 analyze the negative peaks on the derivative of the I-V curve to extract the steps. Then, a threshold is set to
55 classify PS and crack faults. Zhang et al [18] break down the I-V curves into low and high voltage domains.
56 With a statistical method, a detection line is used to identify hot spot, PS, and crack. Similar studies can also
57 be found in [19,20]. (3) Comparison of full I-V curves with simulated ones to generate residuals for analysis.
58 Mohamed et al. [21] compare the measured I-V curves with the simulated ones obtained from the double diode
59 model. They use threshold analysis to classify several types of PS and degradation. Bonsignore et al. [22]
60 generate residuals from I-V curve features to detect PS, SC and ground fault. The pros and cons of the three
61 categories are summarized in Table 1.

62

63

Table 1 Comparison of common methods using I-V curves

Ref.	Methods	Pros	Cons
[9–16]	Use extracted key curve features	<ul style="list-style-type: none"> • Simple feature extraction • Suitable for multiple faults diagnosis 	<ul style="list-style-type: none"> • Partial information • Sensible to noise and outliers • Hard to extract R_s, R_{sh} from curves with irregular shape

[17–20]	Calculate the 1 st or 2 nd derivative of the curve to identify steps on the curve	<ul style="list-style-type: none"> • Able to identify the number of reflection points 	<ul style="list-style-type: none"> • Sensible to noise and outliers • Only applied to detect mismatch faults (PS, crack, hot spot)
[21,22]	Generate residual between measured I-V curves and simulated ones	<ul style="list-style-type: none"> • Full information • Suitable for multiple faults diagnosis 	<ul style="list-style-type: none"> • Poor generalization capability • Slow decision making

64

65 Additionally, with the popularity of deep learning in PV diagnosis, in [23], I-V curves are integrated with
66 irradiance (G) and module temperature (T_m) into 4-column matrices. These feature matrices are then processed
67 by an improved *ResNet* model (one type of deep neural network) to classify 8 PV array fault conditions. The
68 classification accuracy reaches 98.8%, but the duplication of information is introduced in the input features.
69 Besides, the diagnosis model exhibits a high complexity of 18 layers compared with common shallow neural
70 networks adopted in PV diagnosis (1 to 3 hidden layers).

71 Based on the literature review several gaps have been identified; there is no effective use of all the information
72 contained in the I-V curves [9–16]; the detectable faults using the proposed methods are quite limited, like
73 only the mismatch faults in [17–20]; practical applicability is limited due to low generalizability or redundancy
74 of information. The methodology proposed below therefore aims to fill these gaps.

75 The proposed PV diagnosis methodology uses three different methods to integrate the full I-V curves into the
76 input features for diagnosis. Then, the performance of six classifiers based on MLT techniques [18] when using
77 different input features will be evaluated and compared for the fault classification.

78 The contribution of this work lies in the following points:

- 79 • A new methodology for photovoltaic panel fault diagnosis, based on the full use of I-V curves, is
80 proposed. It outperforms methodologies based on partial use of I-V curves;
- 81 • Based on the correction procedures of IEC 60891, a new procedure is proposed and applied to the I-V
82 curves of faulty photovoltaic panels, measured under different environmental conditions;
- 83 • In order to make the method independent of the number of points measured to obtain the I-V, a
84 resampling method is proposed. The number of resampling points is determined based on resampling
85 performance and computational cost;
- 86 • Two feature transformation techniques are applied and adapted to I-V curves for the first time to
87 improve defect discrimination.

88

89 The paper is organized as follows: Section 2 introduces the configuration of the simulated dataset of I-V curves;
90 Section 3 presents the procedures of pre-processing, including the correction and resampling of I-V curves;
91 Section 4 puts forwards the feature extraction methods; Section 5 performs the feature analysis using different
92 MLT classifiers for PV FDD, presents the diagnosis results using simulated and experimental data, and a
93 comparative study with methods from the literature. Section 6 concludes the paper.

94 2 Configuration of the simulated dataset

95 2.1 PV array model

96 A small-scale PV array model, which corresponds to the setup of the field test (presented in Section 5), is
97 constructed under Matlab Simulink®. The array consists of 6 Wiltec 62391-50W sc-Si modules, with two
98 strings in parallel and each string of three modules in series, as shown in Fig.1. Each module has 36 cells and
99 two bypass diodes. Detailed module and array parameters are listed in Table 2 and Table 3, respectively.

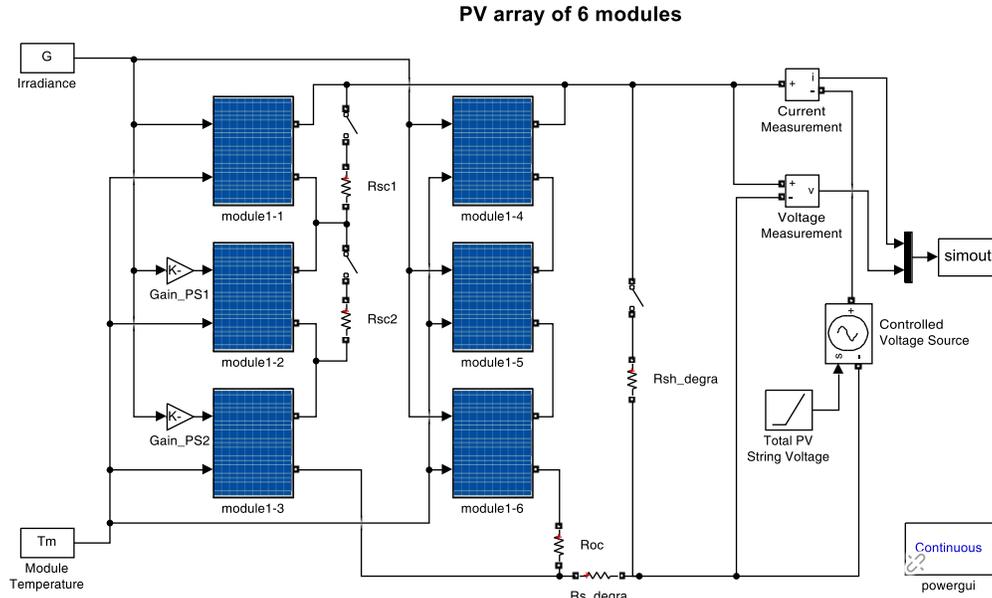


Fig. 1. Simulation model of the PV array

Table 2 Parameters of Wiltec 62391-50W PV module

Variable	Value	Variable	Value
I_{SC}	3.24 A	V_{MPP}	17.55 V
V_{OC}	21.58 V	α_{rel}	0.05 %/°C
I_{MPP}	2.85 A	β_{rel}	-0.3 %/°C

Table 3 Parameters of PV array

Variable	Value	Variable	Value
I_{SC}	6.48 A	V_{MPP}	52.65 V
V_{OC}	64.74 V	I_{MPP}	5.70 A

Where, V_{OC} is the open-circuit voltage, V_{MPP} and I_{MPP} refer to the voltage and current at maximum power point (MPP), α_{rel} and β_{rel} are the temperature coefficients (TC) of short-circuit current (I_{SC}), and the open-circuit voltage (V_{OC}), respectively.

In total, eight configurations (i.e., healthy and seven faulty conditions) are studied:

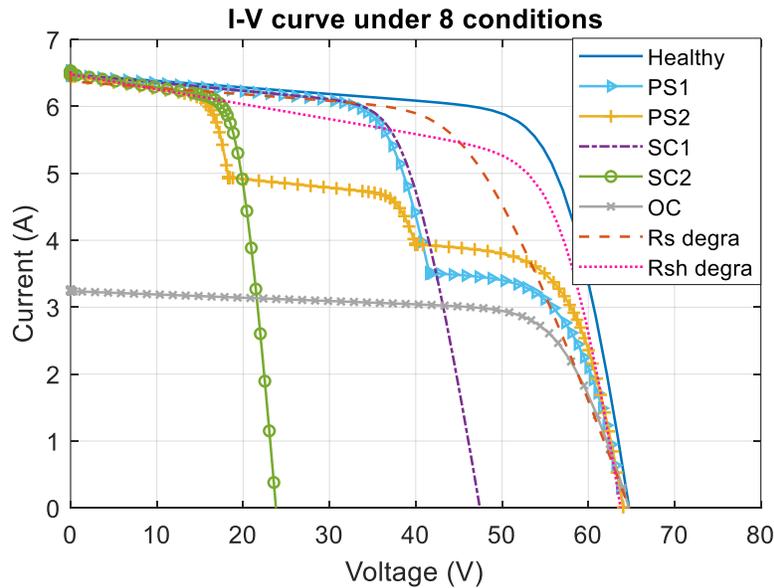
- 1) Two types of Partial Shading (PS): 1 (or 2) module(s) in one string is (are) shaded (abbreviated as PS1 and PS2, respectively) by adjusting the block gain value that controls the irradiance;
- 2) Two types of Short-Circuit (SC): 1 (or 2) module(s) in one string is (are) short-circuited (abbreviated as SC1 and SC2, respectively) by connecting a resistance R_{SC} in parallel;
- 3) Open-Circuit (OC): 1 string is open-circuited by connecting a resistance R_{OC} in series;
- 4) R_s degradation (R_s degra): increase of array equivalent series resistance, controlled by a resistance R_{s_degra} added in series;
- 5) R_{sh} degradation (R_{sh} degra): decrease of array equivalent shunt resistance, controlled by a resistance R_{sh_degra} added in parallel;

These faults are chosen mainly because of their significant impact on the PV power output and their high

120 frequency of occurrence [24]. Besides, the reproducibility in simulation and in real conditions is also
 121 considered.

122 Some examples of the I-V curves under these configurations are shown in Fig. 2. Typical I-V curves also
 123 simulated under Matlab Simulink® reported in the literature [10, 23] are presented in Fig. 3. It can be noted
 124 the similarity with those in Fig. 2.

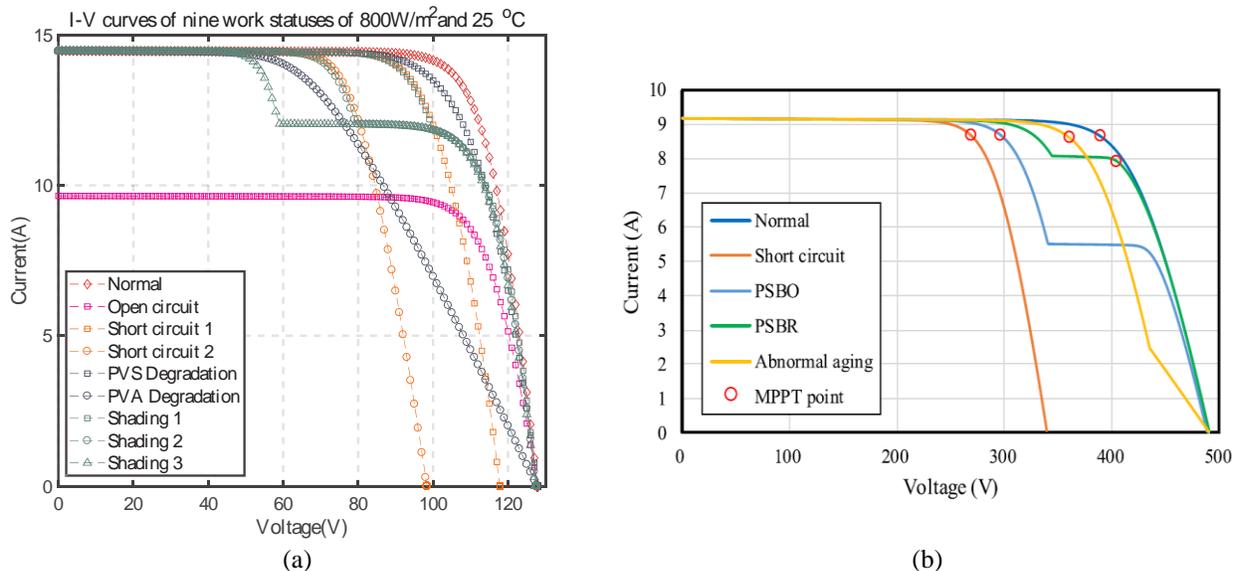
125 For PS1 and PS2, reflection points appear due to the activation of bypass diodes that bypass the shaded module.
 126 For SC1 and SC2 corresponding to one or two short-circuited modules out of three in one
 127 string, V_{OC} decreases by 1/3 or 2/3 compared to the healthy condition. For OC, the I_{SC} decreases by 1/2
 128 because one out of two strings is OC. For the degradation faults, the slope near the voltage or the current axis
 129 region changes for R_s and R_{sh} degra, respectively.



130

131

Fig. 2. Examples of simulated I-V curves under eight PV array conditions



132

133

Fig. 3 Simulated I-V curves reported in literature, (a): [23], (b): [10]

134 2.2 Generation of dataset

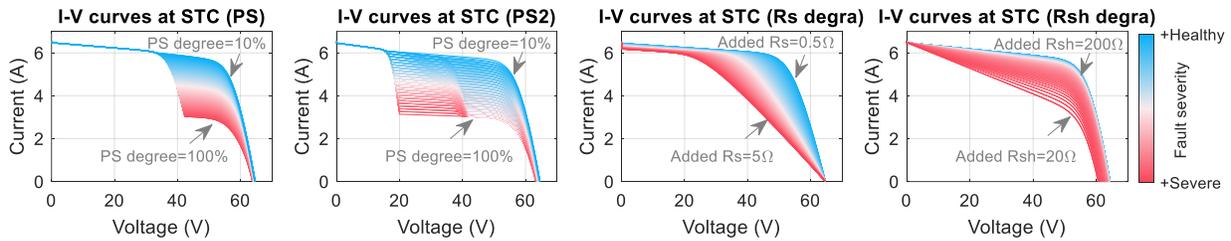
135 Based on the array model, two datasets are generated, one for training and one for testing. For each I-V curve
 136 in the data set, G varies between 400 and 1200 W/m² and T_m between 10 and 80°C. The corresponding
 137 parameters for the two datasets are identical and presented in Table 4. Compared to previous studies in
 138 literature [23], for PS, R_s , and R_{sh} degradation, the fault severity is not simply set as constant. Instead, the
 139 fault severity is also varied with the corresponding parameters covering uniformly relatively wide ranges listed
 140 in Table 4, which are also the effective range for FDD in this study. Examples of curves under several faulty
 141 conditions are presented in Fig. 4. For better illustration of the impact of fault severity, the given curves are
 142 generated under identical Standard Test Condition (STC, $G=1000$ W/m², $T_m=25^\circ\text{C}$).

143 For PS, the PS degree is set from 10% (low shaded) to 100% (total shaded). The lower limit of 10% is set to
 144 avoid the confusion between incipient PS (such as 1% PS degree) and the healthy condition [25]. For SC, the
 145 R_{SC} is set to $10^{-5}\Omega$ to emulate a full SC, similar for OC with R_{OC} equal to $10^5\Omega$. For R_s and R_{sh} degradation,
 146 the fault severity is tuned to introduce a wide range of shape distortion on the I-V curves, as shown from the
 147 slope change in Fig. 4.

148 Table 4 Parameter setting for the array under healthy and faulty conditions

Condition	PS degree (%)	R_{SC} (Ω)	R_{OC} (Ω)	R_{s_degra} (Ω)	R_{sh_degra} (Ω)
Healthy	100	-*	10^{-5}	10^{-5}	-
PS (1 or 2 modules)	10 - 100	-	10^{-5}	10^{-5}	-
SC (1 or 2 modules)	100	10^{-5}	10^{-5}	10^{-5}	-
OC (1 string)	100	-	10^5	10^{-5}	-
R_s degradation	100	-	10^{-5}	0.5 - 5	-
R_{sh} degradation	100	-	10^{-5}	10^{-5}	20 - 200

149 * '-' means that the related resistance is disconnected in the simulation model.



150

151 Fig. 4. I-V curves (STC) under PS1, PS2, R_s , and R_{sh} degradation (For PS2, the shading level for the two modules
 152 could be the same or different as shown in Fig. 2. For R_{sh} degradation, the shape of the curves does not vary linearly
 153 with the additional resistance in parallel to the array)

154 The simulation curves are obtained under ideal conditions, i.e. without any measurement errors or fluctuations.
 155 Therefore, in order to train the classifiers with data more representative of those measured in real conditions,
 156 environmental noise and random measurement errors are introduced. The environmental noise follows a
 157 normal distribution with a SNR of 35dB, and the random error is set at 0.5% of the variance of V and I . These
 158 values are determined by combing the datasheets of the measurement devices used in field tests to the
 159 uncertainties given in the literature and technical reports [26,27].

160 Finally, in total, the training dataset contains 12000 simulated I-V curves for the eight conditions (each
 161 condition corresponding to 1500 curves). The test dataset contains 2400 curves (300 curves for each condition).
 162 Both datasets will be subjected to the following pre-processing operations.

163 3 Pre-processing of I-V curves

164 The pre-processing of I-V curves consists of two main operations: correction and resampling. Irradiance or/and
 165 temperature variations can introduce differences among into I-V curves. These differences could be mis-

166 interpreted as fault signatures. Therefore, before exploiting I-V curves obtained by measurement or simulation
 167 under different environmental conditions, they must be corrected to an identical condition, usually the STC.
 168 The IEC 60891 standard is intended for this correction. However, the methods it proposes are not suitable for
 169 curves taken from defective photovoltaic panels. A new correction method is presented and applied in this
 170 study. Another difficulty is related to the difference in the number of points that may exist between two I-V
 171 curves. This may be due to the quantization step or to the measuring equipment. After the correction step, a
 172 resampling step is needed to ensure that all curves have the same number of points before being analyzed.

173
 174 Different from the double-resampling method adopted in [23], a single resampling method is proposed and
 175 employed in this study. A special procedure to determine the number of resampling points is also designed
 176 considering resampling performance and computational cost. These steps are now detailed in the following.

177 3.1 Correction of I-V curves

178 Under various environmental conditions (different G and T_m), measured or simulated I-V curves could have
 179 distinct shapes. Thus, to avoid any misinterpretation, the I-V curves are corrected to the same environmental
 180 condition. Here, STC is used as the target condition.

181 In [28,29], it is found that the usual I-V curve correction methods are the procedure 1 and procedure 2 from
 182 the IEC 60891 standard [30]. However, these methods have limited performance in the presence of faults.
 183 Therefore, an improved correction method has been proposed and proved able to achieve better overall
 184 correction performance under all the tested faulty conditions (same fault types studied in this paper).
 185 Accordingly, this proposed method, detailed as follows, will be applied to correct the collected I-V curves.

$$186 \quad I_2 = I_1(1 + \alpha_{rel}(T_{m2} - T_{m1}))G_2/G_1 \quad (1)$$

$$187 \quad V_2 = V_1 + c[\beta_{rel}(T_{m2} - T_{m1}) + a \cdot \ln(G_2/G_1)] \\ - R_s(I_2 - I_1) - \kappa \cdot I_2(T_{m2} - T_{m1}) \quad (2)$$

$$188 \quad c = V_{oc1}[1 + \beta_{rel}(25 - T_1)] \quad (3)$$

189 where, I_1 and I_2 , V_1 and V_2 , T_{m1} and T_{m2} , G_1 and G_2 are the current, voltage, T_m , and G before and after
 190 correction, respectively; R_s is the internal series resistance; a and κ are the curve correction factors; R_s , a , and
 191 κ are determined from the simulations.

192 3.2 Resampling of I-V curves

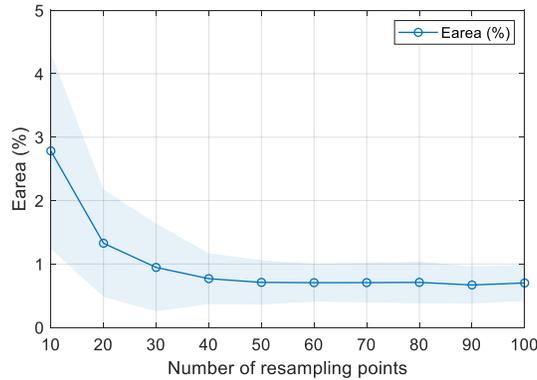
193 After the correction, the I-V curves are resampled. This is particularly necessary when dealing with new I-V
 194 curves with different number of points or distribution. This is the case, for example, for I-V curves measured
 195 by different tracers. This treatment not only ensures that each I-V curve has the same number of points, but
 196 also, and more importantly, that the points on the curve are uniformly distributed. The resampling to make the
 197 FDD methodology interoperable is described below in several steps:

- 198 - Construct a voltage vector with N points linearly distributed in the range $[0, V_{max}]$ with a constant step
 199 (V_{max} is a constant for all the conditions, which could be set a little higher than the array V_{OC} at STC in
 200 healthy condition to avoid the voltage of improperly-corrected curves to exceed this limit);
- 201 - For each voltage value, find the nearest point on the original curve and record the corresponding PV current
 202 value;
- 203 - Finally, construct a new current vector with the N values.

204 To determine N , the quality of the resampling must be quantified. Here, the area error (E_{area}) [31] is adopted
 205 as the metric, as defined in (4).

$$E_{area} = \frac{S_{Area\ of\ error}}{S_{original}} \times 100\% \quad (4)$$

206 Where, $S_{original}$ is the area enclosed by the original I-V curve, $S_{Area\ of\ error}$ is the area enclosed by the
 207 difference between the original and the resampled I-V curve. Considering the I-V curves for the eight cases
 208 studied (the original number of points varies between 298 and 363 due the variable simulation time step), the
 209 mean and standard deviation of E_{area} with varying N are presented in Fig. 5.



210

211 Fig. 5 E_{area} of resampling of I-V curves as a function of N (the dots represent the mean value, while the band
 212 represents the standard deviation)

213 It is observed that, when $N > 40$, E_{area} becomes relatively small ($\sim 0.75\%$) and stable. However, it should be
 214 noted that, beside the quality of the resampling, the complexity of the model (the computational burden of the
 215 classifiers increases with the number of points) also needs to be considered. Therefore, in this study, N is set
 216 to 50 to ensure good resampling performance and reduce computational cost. Thus, after resampling, each I-
 217 V curve will consist of 50 points.

218 4 Feature extraction

219 After the correction and the resampling of the I-V curves, three types of features are proposed: raw I-V data
 220 or transformed into 1D or 2D vectors to increase the discriminability of features. The transformations are based
 221 on two techniques that have not been so far, to the best of our knowledge, applied to I-V curves. A dimension
 222 reduction is also evaluated to reduce information redundancy and computational cost. These steps are now
 223 presented.

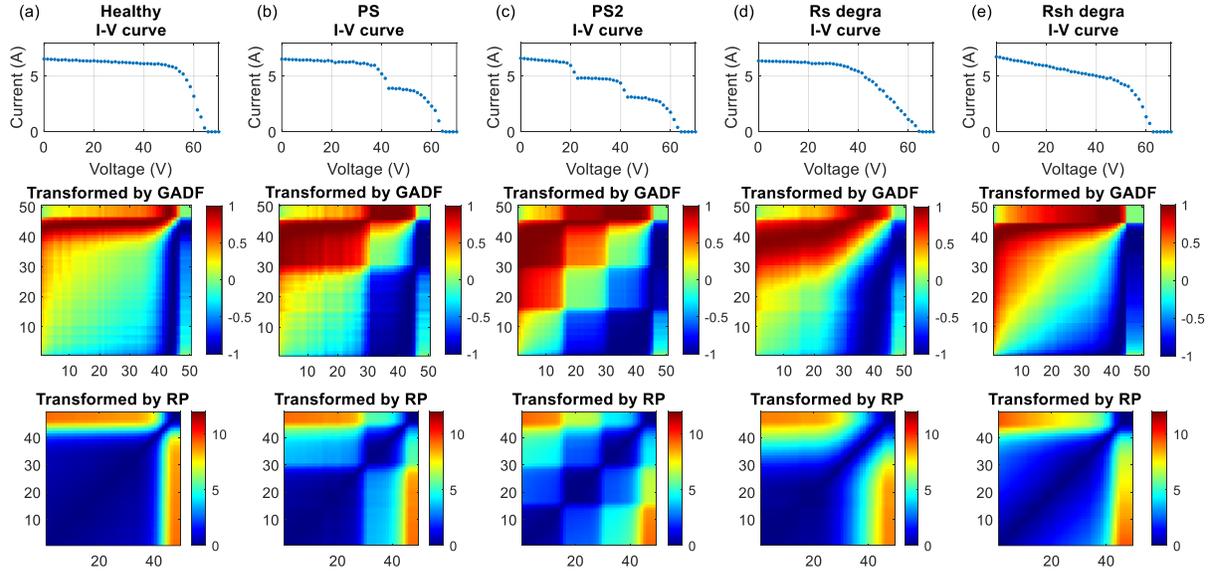
224 4.1 Feature transformation

225 After the resampling, there are two means to extract features from the resampled current vectors. The first one
 226 is to use these vectors directly as input to classifiers (this method is named ‘direct I-V’ hereafter), while the
 227 second transforms the 1-dimensional (1D) vector to a 2-dimensional (2D) matrix. Specifically, two
 228 transformation techniques, commonly practiced in the transformation of time-series signal will be used in this
 229 research:

- 230 1) *Gramian Angular Difference Field* (GADF) [32]: It calculates the inner product of the input vector via
 231 the characterization of angular difference and preserves the temporal dependency. The size of the
 232 generated matrix ($D_{matrix} = D_{vector}^2$), where D_{vector} is the number of features of the original vector. In this
 233 sense, when $D_{vector} = 50$, each current vector is transformed to one matrix with $D_{matrix} = 2500$.
- 234 2) *Recurrence Plot* (RP) [33]: It also reveals all the time dependency of vectors and constructs a square
 235 matrix by calculating the inner distance among all the states in the phase space. The phase space is

236 generated by embedding a shift (τ) in the input vector, where τ is determined by optimization. The D_{matrix}
 237 of RP equals $(D_{vector} - \tau)^2$. In this study, $\tau=1$, thus, $D_{matrix}=2401$.

238 Some examples of GADF and RP transformation are illustrated in Fig. 6. The reason for performing these
 239 operations will be explained in Section 5.1.4.



240

241 Fig. 6. Examples of matrices transformed by GADF and RP: (a) Healthy, (b) PS, (c) PS2, (d) R_s degradation,
 242 (e) R_{sh} degradation (the value of each component on the matrix is marked by colormap)

243 From the examples of transformation, it is observed that the values in both RP and GADF matrix stretches
 244 symmetrically around the main diagonal. When the value of the current vector exhibits a decline, a ‘square’
 245 will correspondingly appear, as seen from the PS and PS2 cases. When the slope near the V_{OC} or I_{SC} part of
 246 the I-V curve changes, the diagonal area in the matrix will also shrink along with different directions. In this
 247 way, all the changes in an I-V curve could be reflected in the transformed matrix.

248 All the three feature extraction methods (direct I-V, RP, and GADF) capture complete information from the
 249 vector of current resampled from original I-V curves. Their performance will be compared to the method that
 250 only uses partial information of the I-V curves.

251 4.2 Dimension reduction of features

252 When there is a slight variation in one I-V curve, the resampled current vector used in direct I-V may contain
 253 redundant points and the transformed features could also have information duplication. To reduce the
 254 complexity of the classifiers, and improve the efficiency of the training process, a reduction of the number of
 255 features is performed, using Principal Component Analysis (PCA). The number of principal components
 256 finally retained is a trade-off between the dimension reduction and the loss of information due to the eliminated
 257 components. The usual rule is to keep the minimum number of components that represent a certain amount of
 258 the original information (cumulated variance); 95% is usually adopted. The diagnosis performance without
 259 PCA will also be discussed in in Section 5.1.3.

260 5 Feature analysis for PV FDD

261 For the feature extraction methods, except the ones which capture complete I-V curve information (direct I-V,
 262 RP, GADF), one traditional method which uses partial I-V curve information will also be tested for comparison.
 263 This method extracts the eight key features (named ‘ $8paras$ ’ hereafter) from the original I-V curve and

264 environmental parameters, i.e., G , T_m , V_{MPP} , I_{MPP} , V_{OC} , I_{SC} , R_s and R_{sh} . The parameters R_s and R_{sh} are
 265 calculated by the reciprocal of the slopes of one I-V curve near the V_{OC} and I_{SC} , respectively. In the section
 266 devoted to the diagnosis, several MLT classifiers will be evaluated for the fault classification to find which
 267 one performs better when using the four different input features.

268 In total, six common MLT classifiers are studied: Artificial Neural Network (ANN), Support Vector Machine
 269 (SVM), k-Nearest Neighbors (kNN), Decision Tree (DT), Random Forest (RF), and Naive Bayesian Classifier
 270 (NBC). These classifiers all tuned and evaluated under Matlab® for the classification of the eight conditions
 271 of PV array. The configuration of these classifiers is listed in Table 5.

272 Table 5 Configuration of MLT classifiers

MLT	Configuration
All types	<i>Common setting</i> <ul style="list-style-type: none"> • Normalization: true (to [-1,1]) • K-fold validation: 5 • Hyperparameter optimization method: grid search or Bayesian method
ANN	<ul style="list-style-type: none"> • Structure: 1 or 2 hidden layers (5-35 neurons) • Loss function: Cross entropy • Train algorithm: Scaled conjugate gradient • Function: <i>tansig</i> (hidden layer), <i>softmax</i> (output layer)
SVM	<ul style="list-style-type: none"> • Multiclass method: one-vs-one • Kernel scale: optimized • Penalty: optimized • Kernel function: Linear, Quadratic, Cubic, Gaussian, or RBF
kNN	<ul style="list-style-type: none"> • Number of neighbors: optimized • Distance metric: Euclidean, City block, or Chebyshev
DT	<ul style="list-style-type: none"> • Max number of splits: optimized • Split criterion: Gini's diversity index • Minimum number of leaf nodes: optimized
RF	<ul style="list-style-type: none"> • Max number of splits: optimized • Minimum number of leaf nodes: optimized • Number of predictors selected for each split: optimized
NBC	<ul style="list-style-type: none"> • Distribution type: Gaussian, or Kernel • Kernel Type: Gaussian, Box, Epanechnikov, or Triangle • Kernel smoothing window width: optimized

273

274 5.1 Diagnosis results using simulated data

275 5.1.1 Performance of fault classification

276 The training database (presented in Section 2.2) is divided into 70% for training and 30% for validation. This
 277 partitioning ratio (2/3 for learning and 1/3 for testing) is a commonly adopted rule of thumb [34]. With a dataset
 278 of 12,000 samples (I-V curves), this partitioning ratio allows efficient learning and stable validation [35].
 279 Thanks to this validation the hyperparameters of the six MLT classifiers can be determined. The trained model
 280 will be evaluated with the test database (also presented in Section 2.2).

281 The test accuracy of the six classifiers using the four different input features is presented in Fig. 7. The details
 282 (including the macro precision, recall, F1 score, accuracy, hyperparameters) of the best classifier are listed in
 283 Table 6, and the corresponding confusion matrices are illustrated in Fig. 8.

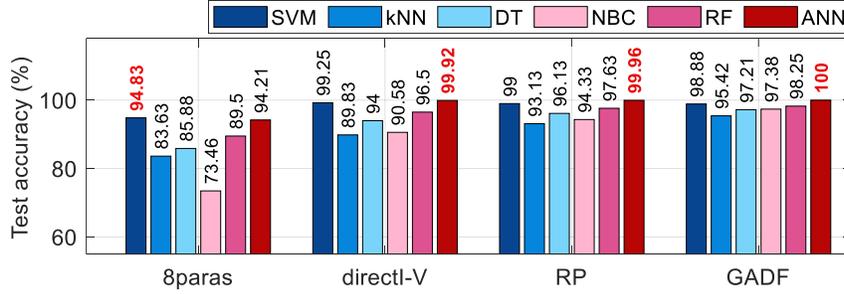


Fig. 7. Test accuracy of all classifiers (the best accuracy is marked in red bold)

Table 6 Details of the best classifiers

Type of input features	Best classifier	Precision	Recall	F1 score	Test accuracy	Classifier hyperparameters
8paras	SVM	0.9520	0.9483	0.9479	94.83%	Kernel: Polynomial 2, BoxConstraint=29.3
direct I-V	ANN	0.9992	0.9992	0.9992	99.92%	#Neurons in hidden layer =15
RP	ANN	0.9996	0.9996	0.9996	99.96%	#Neurons in hidden layer =10
GADF	ANN	1	1	1	100%	#Neurons in hidden layer =32

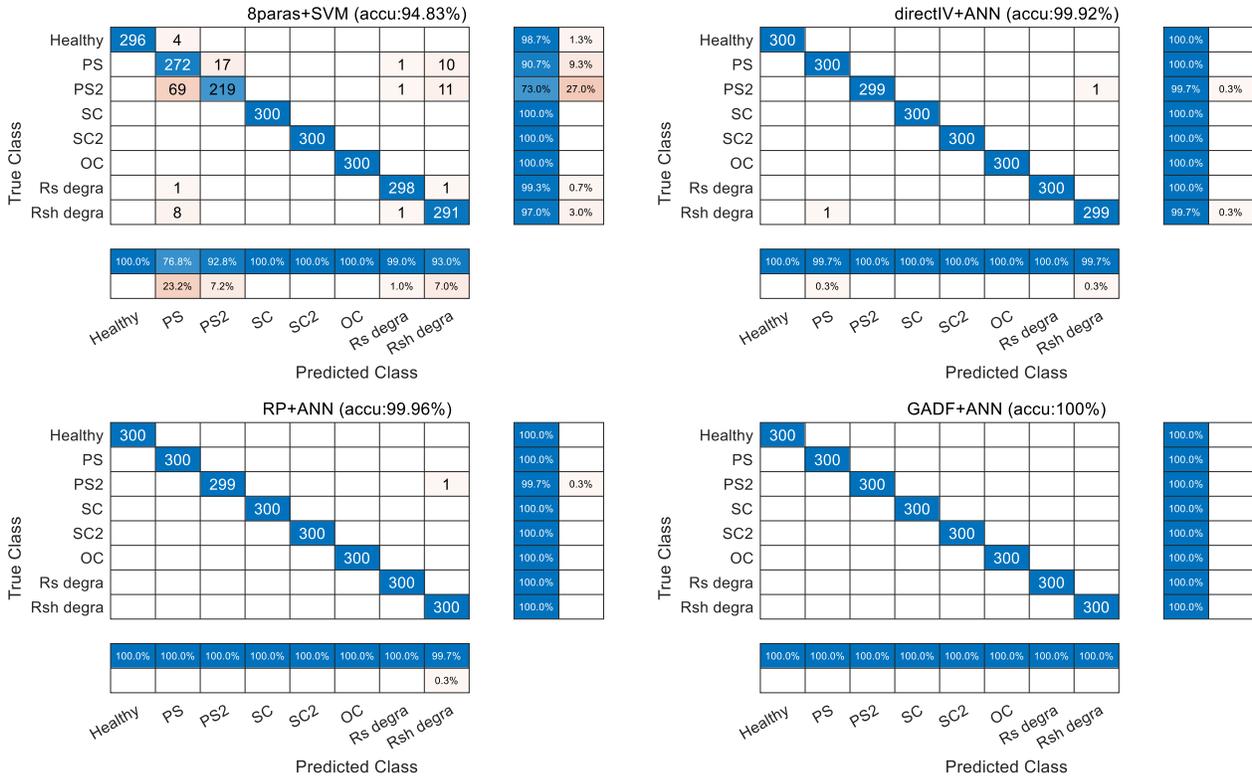


Fig. 8. Confusion matrix of the best classifier when using four types of input features

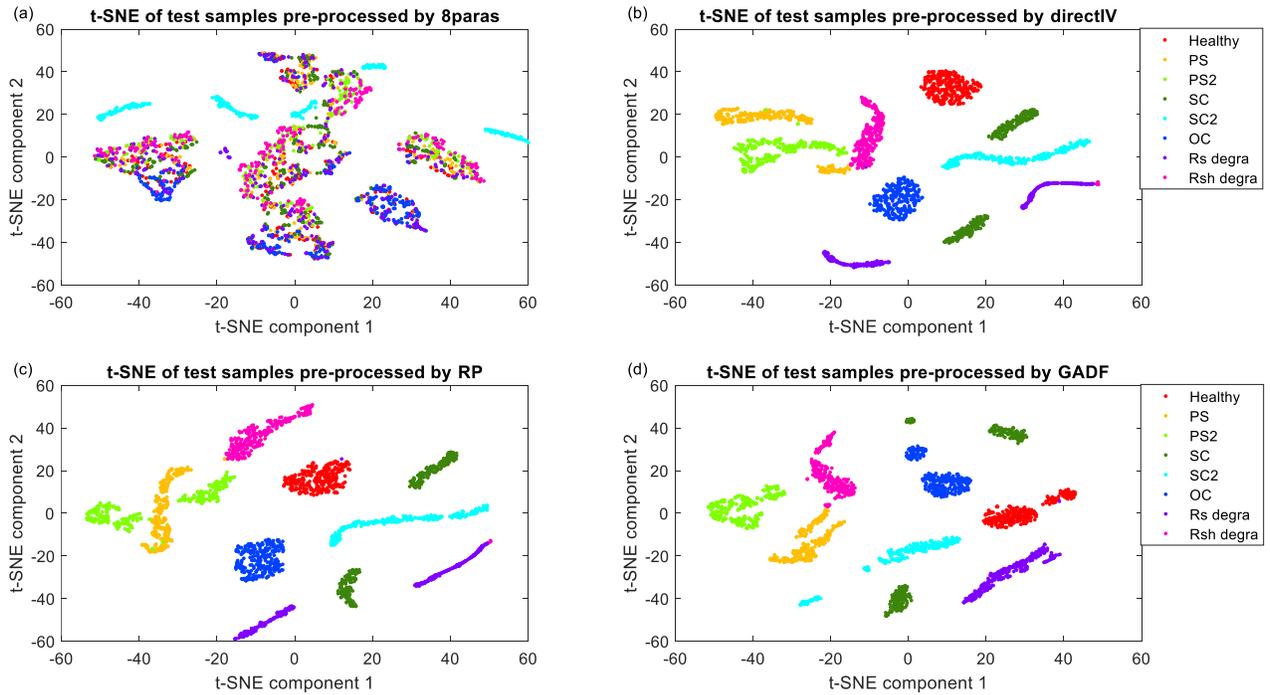
From an input perspective, it can be seen that using features based on the full I-V curve (i.e. direct I-V, RP and GADF) leads to higher classification accuracy and F1 score than using partial I-V curve information (i.e. 8paras). For the 8paras type, the highest accuracy is only 94.83%.

Large misclassification occurs for PS1, PS2, and Rs degradation. This was predictable because under these

295 conditions, the MPPs of some of the curves could have similar or even identical positions, especially in the
 296 presence of measurement noise. In conclusion, if only V_{MPP} , I_{MPP} , V_{OC} , I_{SC} , R_S and R_{Sh} are adopted as curve
 297 features, poor performance will be obtained.

298 Comparatively, for the full I-V curve-based methods (direct I-V, RP, and GADF), almost all the curve
 299 information is embedded in the input features. Thus, the classifiers can learn the variation trend of the curves
 300 and achieve high classification results, all of them higher than 99.9%. For GADF, using ANN as the classifier
 301 achieves 100% classification. For direct I-V and RP, there are one or two I-V curves wrongly classified.

302 In terms of average accuracy, RP and GADF perform better than direct I-V, with the six MLT classifiers. This
 303 proves that the transformation of features is relevant as it improves the classification performance. This could
 304 also be seen from the 2D t-SNE [36] graphs shown in Fig. 9, where the discriminability of the different features
 305 can be observed.



306
 307 Fig. 9. t-SNE graphs of 4 types of input features extracted from identical test database: (a) 8paras, (b) direct I-V, (c)
 308 RP, (d) GADF

309 From the t-SNE graphs, it can be observed that the clusters of features- 8paras are nearly all mixed up, which
 310 shows up its low capability of discriminating the different cases. For RP and GADF, the clusters are relatively
 311 more separated than that of direct I-V, like among PS1, PS2, and R_{sh} degradation. This is clearer for GADF.
 312 This means, using RP or GADF, the discriminability of input features is enforced, which therefore allows all
 313 the MLT classifiers to achieve better classification, as presented in Fig. 7.

314 5.1.2 Robustness to additional disturbance

315 In this part, the best MLT classifiers for the four types of input features (presented in Table 6) will be tested
 316 with new test data sets to assess their robustness to random measurement errors and environmental noise. In
 317 the following, the range of variation is larger than in the previous test data set. The corresponding best
 318 classification results when using the four types of input features are presented in Fig. 10.

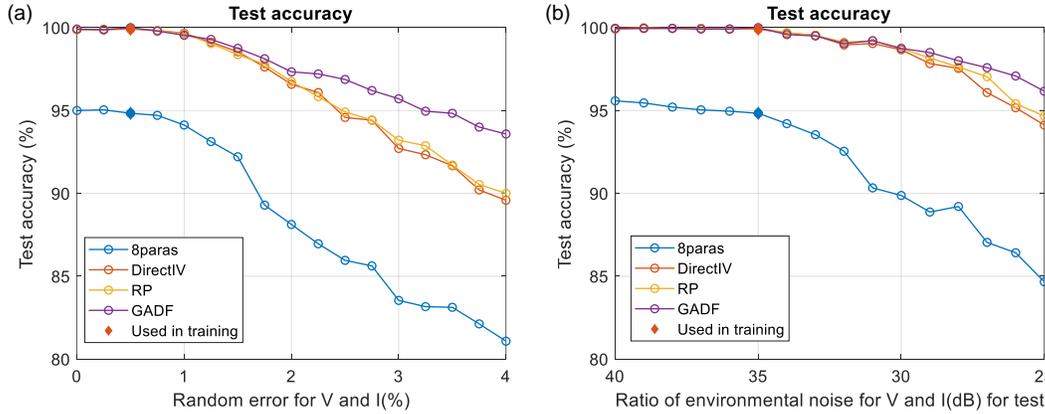


Fig. 10. Best classification accuracy using test dataset with varying level of added disturbance ((a): varying measurement random error of V and I ; (b): varying environmental noise of V and I . The disturbance level adopted in the training dataset is marked in 'diamond')

319
320
321
322
323

324 It is observed that, overall, the best accuracy, when using all four types of features, decreases with increasing
325 disturbance level. However, the use of features based on full I-V curves (direct I-V, RP and GADF) is more
326 robust than the partial use of curves (8paras).

327 As for direct I-V, RP, and GADF, the best accuracy has a similar trend for low disturbance level. However,
328 GADF experiences a slighter decrease of performance with increasing disturbance in V and I . Nevertheless,
329 GADF clearly outperforms direct I-V and RP, with fewer I-V curves under healthy and Rsh degradation
330 misclassified in other conditions. It can be concluded that the classifier using GADF to extract features has a
331 better overall robustness to additional disturbance than other types of feature extraction methods.

332 5.1.3 Impact of PCA operation

333 As presented in Section 4.2, for direct I-V, RP, or GADF, the dimension of extracted features has been reduced
334 with PCA before fed to the MLT classifiers. In the following, the impact of this operation on the diagnosis
335 performance will be discussed. Specifically, the processing time, and the test accuracy with and without PCA
336 are compared as shown in Table 7. In this study, the platform is Matlab® R2020b with CPU of Intel(R) Xeon(R)
337 E-2174G CPU @ 3.80GHz and RAM of 32G.

338

Table 7 Influence of PCA on time and accuracy

Input feature type	PCA	Pre-processing time (PCA included) (s)	Training time (s)	Preprocess+ train time (s)	Testing time (s)	Test accuracy
direct I-V	Y	0.39	41.57	41.96	0.01	99.92%
	N	0.37	58.65	59.02	0.01	99.88%
RP	Y	97.72	48.31	146.03	0.01	99.96%
	N	89.33	1209.23	1298.56	0.13	99.92%
GADF	Y	48.88	31.75	80.63	0.01	100%
	N	46.90	670.89	717.79	0.11	100%

339

340 From these results, the following conclusions can be drawn about the effects of PCA :

- 341 ■ Significant reduction of the total training time (pre-processing + training), especially for RP and GADF
342 for which there is a decrease of almost 90% when PCA is used,
- 343 ■ The accuracy of the classification is hardly affected.

344 Thus, it is favorable to conduct the dimension reduction with PCA, especially for high-dimensional features

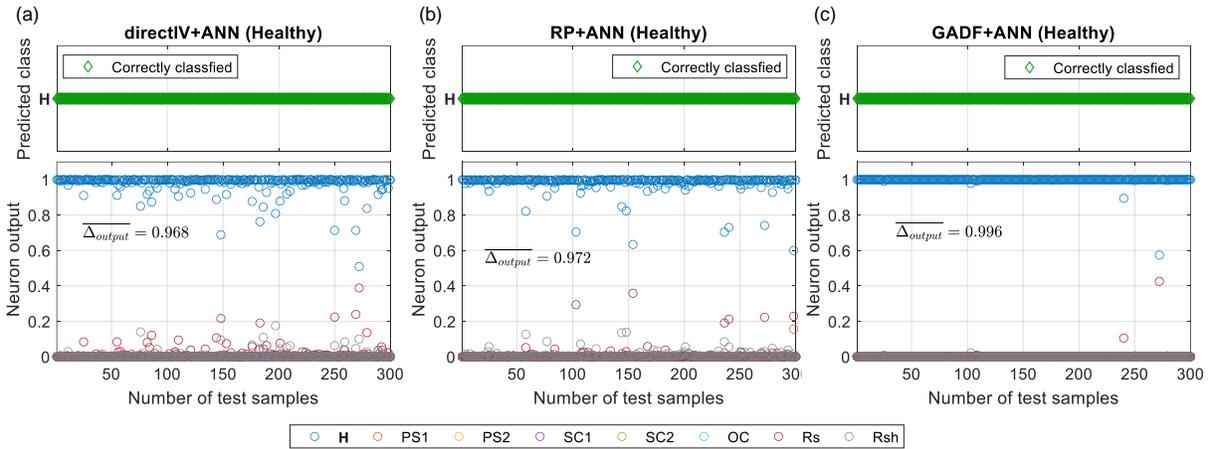
345 extraction, as for RP and GADF.

346

347 5.1.4 Impact of transformation

348 In Section 4.1, in addition to direct I-V, two feature transformation methods (RP and GADF) have been
 349 introduced. Clearly, the feature dimension after transformation greatly increases as the current vector is
 350 transformed into a square matrix. On one side, this could increase the complexity of FDD methodology, like
 351 the longer processing time needed to perform dimension reduction and training, as it can be observed in Table
 352 7.

353 On the other side, it should be noted that the transformation is able to increase the discriminability of features,
 354 as presented in the t-SNE graph in Fig. 9. This can also be noted from the observation of the outputs of the
 355 neurons. From the test results given in Fig. 8, for the I-V curves of the array under healthy condition, direct I-V
 356 V, RP and GADF could all achieve 100% classification. However, from the detailed results of the output of
 357 the best classifier, a difference in the confidence of the classification could be observed, for healthy conditions,
 358 as shown in Fig. 11.



359

360 Fig. 11 Neuron output of the best classifier under healthy condition using direct I-V, (a) direct I-V, (b)
 361 RP, (c) GADF

362 For these three types of features, ANN is the best classifier. The output values of the eight neurons in the output
 363 layer (corresponding to the eight cases) are presented. For each prediction of class, the sum of all the outputs
 364 equals to one. The predicted class is the one whose neuron has the highest output. In this sense, the greater
 365 the difference between the neurons' outputs, the more reliable the prediction. To quantify the output difference,
 366 Δ_{output} is defined as the difference between the largest neuron output minus the second largest output. The
 367 larger Δ_{output} , the more confidence can be placed in the classifier. In Fig. 12, the mean of the Δ_{output}
 368 ($\overline{\Delta_{output}}$) for the 300 test I-V curves under healthy condition is displayed. It can be observed that direct I-V
 369 and RP have almost the same performance, while GADF reaches the highest value with 0.996. This means,
 370 when using GADF for feature extraction, the decision made by the best classifier is more reliable, which once
 371 more confirms the merit of performing feature transformation.

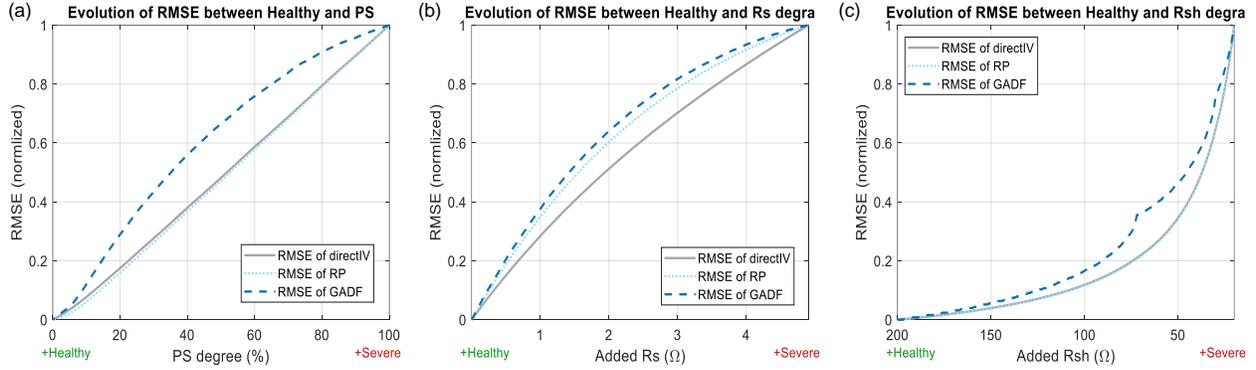
372 In addition, the t-SNE figure of features (already presented in Fig. 8) also proves that the robustness to the
 373 additional disturbance when using RP and GADF is enforced to that using direct I-V, especially for GADF.

374 To investigate why the transformation could increase the discriminability of features, the difference between
 375 features under different conditions after transformation is analyzed. In fact, the I-V curves of an array under
 376 incipient fault conditions (like low PS degree, low additional R_s , large additional R_{sh}) are prone to be confused

377 with those under healthy conditions. Therefore, taking the features of I-V curves under healthy condition as a
 378 reference, the difference between the reference and the features from I-V curves under PS, R_s , or R_{sh}
 379 degradation with different fault severities can be quantified. Here, the Root Mean Squared Error (RMSE) is
 380 used as the metric, defined as:

$$RMSE = \sqrt{\frac{\sum_1^n (X_i - X_{i_{ref}})^2}{n}} \quad (5)$$

381 where, X_i is the i^{th} element in the transformed feature X , n is the number of elements in the feature. X could
 382 be the current vector built by direct I-V method or the matrix built by RP or GADF method.



383

384 Fig. 12 Normalized RMSE of features between healthy and (a) PS, (b) R_s degradation, (c) R_{sh} degradation

385 From Fig. 12, it is observed that the normalized RSME of direct I-V and RP have almost a similar evolution
 386 under the three conditions, while that of GADF increases clearly faster, especially under PS and R_{sh}
 387 degradation. This means, with increasing fault severity, GADF features will be more sensitive. It can also be
 388 pointed out that incipient fault (variation lower than 10%) diagnosis is an opened challenge as the deviation is
 389 very low for all the transformations. However, the discriminability of features under the different cases is
 390 enforced with the feature transformation, especially with GADF.

391 5.2 Diagnosis results using experimental data

392 5.2.1 Field test setup

393 To evaluate the trained classifiers under real case, an experimental PV array of 6 sc-Si modules (Wiltec 62391-
 394 50W) is constructed, as shown in Fig. 13. The parameters of the PV modules in the simulation and field tests
 395 are identical. A reference cell (RG100) is used to measure G , and a Pt100 probe (class A) to measure the back-
 396 sheet T_m of one module. Two multimeters (Keithley 2440 and 175) record the measurements G and T_m . The I-
 397 V tracer (Chauvin-Arnoux FTV200) records the array's I-V curves.



Fig. 13 Field test setup

398
399

400 All the eight PV array cases are reproduced, and the corresponding I-V curves are recorded. Some fault setups
401 are shown in Fig. 14. Table 8 shows the severities of the faults, which are consistent with those used for
402 training and presented in Table 4. Fifteen curves are recorded for each of the eight cases. Therefore, there are
403 120 I-V curves, some of which after correction and extrapolation are shown in Fig. 15.

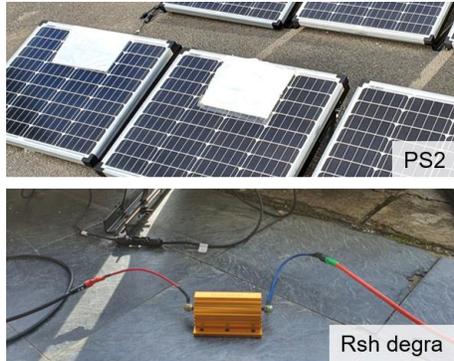
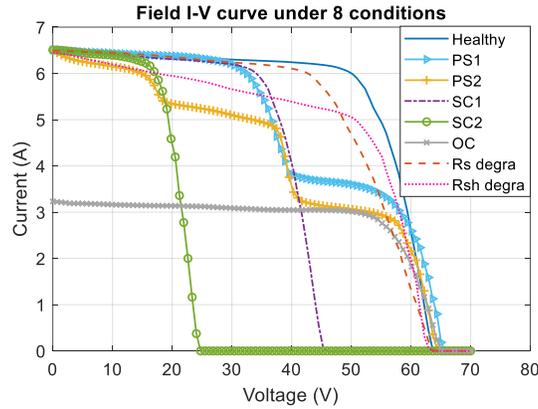


Fig. 14. Examples of fault setup

404
405



406
407
408

Fig. 15. Examples of measured I-V curves under 8 PV array conditions

Table 8 Setting of fault severity

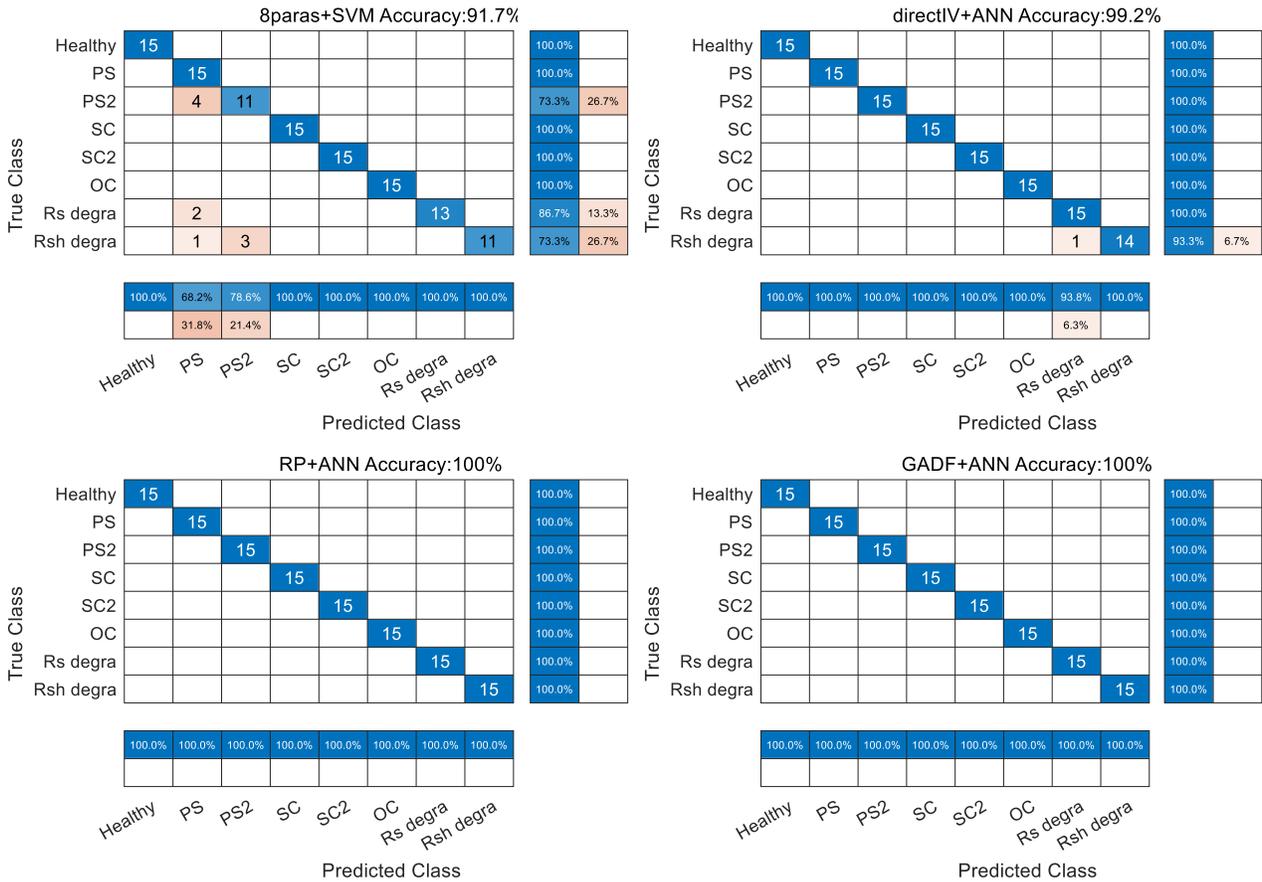
Condition	Varying fault severity
-----------	------------------------

Healthy, SC, SC2, OC	None
PS, PS2	PS degree controlled by 3 different distances between the shelter and the panels
<i>Rs</i> degradation	Additional resistance in series: 1Ω, 2Ω, 4Ω
<i>Rsh</i> degradation	Additional resistance in parallel: 30Ω, 50Ω, 150Ω

409

410 5.2.2 Classification results

411 After the pre-processing (correction, resampling) and feature extraction (with four feature extraction methods,
 412 namely, 8paras, direct I-V, RP, or GADF), the processed samples are then fed into the best tuned classifiers
 413 (trained with simulated data presented in Table 6) for classification. The results are presented in Fig. 16.



414

415 Fig. 16. Testing results using field-measured data

416 According to Fig. 15, similar to the previous observations for simulation results, the classifiers using the
 417 features based on full I-V curve information (direct I-V, RP, and GADF) outperform the one using partial
 418 information (8paras). For 8paras, the main misclassification occurs between PS and PS2, between PS/PS2 and
 419 *Rs/Rsh* degradation, which corresponds to Fig. 8. For the three types of features based on the full I-V curves,
 420 the classification performance reaches 100% with GADF and RP. There is one I-V curve wrongly classified
 421 when using direct I-V. This experimental test not only validates the effectiveness of the tuned classifiers using
 422 simulated data but also shows the benefit of using features based on full I-V curves for PV FDD.

423 5.3 Comparative study

424 There are several other methods in the literature to extract features from I-V curves for FDD. In the following,
 425 a comparative study will be conducted with the simulated and experimental database presented in Section 2.2.
 426 The extracted features will be processed by the MLTs presented in Section 错误!未找到引用源。 . For each
 427 type of new feature, the MLT classifiers will be re-tuned. Then, the best classifier will be validated with the
 428 same field-measured dataset. The accuracies and processing times are summarized in Table 9.

429 Table 9 Comparison of PV FDD results of different methods

Use of I-V curves	Method	Accuracy			Time		
		Train	Test	Field test	Pre-process +train	Test	Field test*
Partial	6paras [37]	95.41%	94.04%	90.83%	3.59s	0.05s	0.05s
	8paras	96.21%	94.83%	91.67%	3.08s	0.05s	0.09s
Complete	IVGT [23]	99.75%	98.83%	97.5%	38.62s	0.01s	2.13s
	IV image [38]	100%	99.58%	95.00%	20min	13.9s	29.4s
	direct I-V	100%	99.92%	99.17%	41.96s	0.01s	0.17s
	RP	100%	99.96%	100%	146.03s	0.01s	0.29s
	GADF	100%	100%	100%	80.63s	0.01s	0.21s

430 * Field test time includes the time of pre-processing and classification of 120 curves

431 The method based on partial use of an I-V curve requires less time for the pre-processing and the training than
 432 the others to the detriment of a limited accuracy. For the methods based on complete use of one I-V curve, the
 433 three proposed methods (i.e., direct I-V, RP, and GADF) show better classification accuracy than the other
 434 ones. Additionally, the decision-making time in field tests is also rapid, less than 0.3s. The IVGT method is
 435 also based on resampling but performs worse than the proposed three methods. This is supposedly due to a
 436 lack of efficient integration of G and T_m information into the features. As for the IV image method, its
 437 performance is similar to the IVGT method. However, due to the complexity of image data, the training and
 438 field test times are significantly longer than the other methods, making it less competitive for rapid deployment
 439 in PV FDD applications.

440 Compared with other I-V curve-based feature extraction methods, the three proposed methods (direct I-V, RP
 441 and GADF) has the following advantages:

- 442 ■ the classification performance is very good both with simulation and field data,
- 443 ■ the processing time is reasonable with low complexity in the pre-processing and learning stages, so as
 444 to ensure rapid decision making in field trials.

445 Therefore, the whole FDD methodology based on full I-V curves and these feature extraction methods is very
 446 promising for health monitoring of PV modules.

447 5.4 Discussion

448 From the diagnosis results, it is found that the performance of the classifier depends on the input features, i.e.,
 449 support vector machine when using 8paras, artificial neural network when using other features. This is why it
 450 is essential to tune several classifiers to find out which one will have the best performance depending on which
 451 features are being used as inputs. Besides, it is also noteworthy that the type of input features has a more
 452 significant impact on the classification performance (accuracy and robustness to disturbance) than the type of

453 classifiers. These results prove the utmost importance of the pre-processing stage in which the original samples
 454 are transformed to extract efficient features for higher discriminability. In essence, the objective of the pre-
 455 processing process is to improve the quality of the input features. This, if achieved, could not only ease the
 456 subsequent tuning process of classifiers but also improve the diagnosis performance.

457 **6 Conclusion**

458 In this paper, a methodology using full I-V curves and machine learning techniques for the fault diagnosis of
 459 PV array under eight conditions has been introduced. It has been shown that the three feature extraction
 460 methods that employ full I-V curve information perform better than the one using partial information of one I-
 461 V curve. Among the three complete information-based methods, through 1-dimension or 2-dimension
 462 transformation, Gramian angular difference field has the highest capability of discriminating the eight
 463 conditions and exhibits highest robustness to measurement errors and environmental disturbance. The
 464 classifiers tuned with simulated samples have been validated with field-measured I-V curves. The best
 465 classifier (artificial neural network) achieves 100% classification accuracy for both simulation and field tests.
 466 The methodology and the results may constitute a valuable experience for future researchers on how to take
 467 benefit from the complete information embedded in I-V curves, to perform efficient PV fault diagnosis.

468 There are still some challenges to be addressed in future works: (1) incipient fault (at their earliest stage) should
 469 be detected. I-V curves should be simulated and measured under these conditions to evaluate the proposed PV
 470 fault detection and diagnosis methodology; (2) other transformation and dimension reduction techniques could
 471 be evaluated to enhance discriminability and reduce computational cost and complexity; (3) for very large PV
 472 plants, measurement of I-V curves at string or module level should be addressed; (4) the fault detection and
 473 diagnosis methods should address different simultaneous faults that may affect the same module or string; (5)
 474 the performance of FDD methods should be analyzed for PV plants in which only inverter-level measurements
 475 are available.

476 **References**

- 477
- 478 [1] SolarPower-Europe. Global Market Outlook for Solar Power / 2020 - 2024. 2020.
 479 <https://www.solarpowereurope.org/global-market-outlook-2020-2024/>.
 - 480 [2] Lorenzo G Di, Araneo R, Mitolo M, Niccolai A, Grimaccia F. Review of O&M Practices in PV Plants: Failures,
 481 Solutions, Remote Control, and Monitoring Tools. *IEEE J Photovoltaics* 2020;10:914–26.
 482 <https://doi.org/10.1109/JPHOTOV.2020.2994531>.
 - 483 [3] Alam MK, Khan F, Johnson J, Flicker J. A Comprehensive Review of Catastrophic Faults in PV Arrays: Types,
 484 Detection, and Mitigation Techniques. *IEEE J Photovoltaics* 2015;5:982–97.
 485 <https://doi.org/10.1109/JPHOTOV.2015.2397599>.
 - 486 [4] Hernández-Callejo L, Gallardo-Saavedra S, Alonso-Gómez V. A review of photovoltaic systems: Design,
 487 operation and maintenance. *Sol Energy* 2019;188:426–40. <https://doi.org/10.1016/j.solener.2019.06.017>.
 - 488 [5] Livera A, Theristis M, Makrides G, Georghiou GE. Recent advances in failure diagnosis techniques based on
 489 performance data analysis for grid-connected photovoltaic systems. *Renew Energy* 2019;133:126–43.
 490 <https://doi.org/10.1016/j.renene.2018.09.101>.
 - 491 [6] Mellit A, Tina GM, Kalogirou SA. Fault detection and diagnosis methods for photovoltaic systems: A review.
 492 *Renew Sustain Energy Rev* 2018;91:1–17. <https://doi.org/10.1016/j.rser.2018.03.062>.
 - 493 [7] Huawei. Smart I-V Curve Diagnosis. 2020. [https://solar.huawei.com/en-](https://solar.huawei.com/en-GB/download?p=%2F-%2Fmedia%2FSolar%2Fattachment%2Fpdf%2F%2Fdatasheet%2FIV-Curve.pdf)
 494 [GB/download?p=%2F-%2Fmedia%2FSolar%2Fattachment%2Fpdf%2F%2Fdatasheet%2FIV-Curve.pdf](https://solar.huawei.com/en-GB/download?p=%2F-%2Fmedia%2FSolar%2Fattachment%2Fpdf%2F%2Fdatasheet%2FIV-Curve.pdf)
 495 (accessed August 17, 2020).
 - 496 [8] Spataru S, Sera D, Kerekes T, Teodorescu R. Monitoring and Fault Detection in Photovoltaic Systems Based
 497 On Inverter Measured String I-V Curves. 31st Eur. Photovolt. Sol. Energy Conf. Exhib. (EU PVSEC),
 498 Hamburg, Germany: 2015, p. 1667–74. <https://doi.org/10.4229/EUPVSEC20152015-5BO.12.2>.
 - 499 [9] Chine W, Mellit A, Lughy V, Malek A, Sulligoi G, Massi Pavan A. A novel fault diagnosis technique for
 500 photovoltaic systems based on artificial neural networks. *Renew Energy* 2016;90:501–12.
 501 <https://doi.org/10.1016/j.renene.2016.01.036>.

- 502 [10] Huang JM, Wai RJ, Gao W. Newly-designed fault diagnostic method for solar photovoltaic generation system
503 based on IV-Curve measurement. *IEEE Access* 2019;7:70919–32.
504 <https://doi.org/10.1109/ACCESS.2019.2919337>.
- 505 [11] Fadhel S, Delpha C, Diallo D, Bahri I, Migan A, Trabelsi M, Mimouni MF. PV shading fault detection and
506 classification based on I-V curve using principal component analysis: Application to isolated PV system. *Sol*
507 *Energy* 2019;179:1–10. <https://doi.org/10.1016/j.solener.2018.12.048>.
- 508 [12] Dhimish M, Chen Z. Novel Open-Circuit Photovoltaic Bypass Diode Fault Detection Algorithm. *IEEE J*
509 *Photovoltaics* 2019;9:1819–27. <https://doi.org/10.1109/JPHOTOV.2019.2940892>.
- 510 [13] Bouaichi A, Alami A, Hajjaj C, Messaoudi C, Ghennioui A, Benlarabi A, Ikken B, El A, Zitouni H. In-situ
511 evaluation of the early PV module degradation of various technologies under harsh climatic conditions : The
512 case of Morocco. *Renew Energy* 2019;143:1500–18. <https://doi.org/10.1016/j.renene.2019.05.091>.
- 513 [14] Singh R, Sharma M, Rawat R, Banerjee C. An assessment of series resistance estimation techniques for
514 different silicon based SPV modules. *Renew Sustain Energy Rev* 2018;98:199–216.
515 <https://doi.org/10.1016/j.rser.2018.09.020>.
- 516 [15] Chen Z, Wu L, Cheng S, Lin P, Wu Y, Lin W. Intelligent fault diagnosis of photovoltaic arrays based on
517 optimized kernel extreme learning machine and I-V characteristics. *Appl Energy* 2017;204:912–31.
518 <https://doi.org/10.1016/j.apenergy.2017.05.034>.
- 519 [16] Fadhel S, Diallo D, Delpha C, Migan A, Bahri I, Trabelsi M, Mimouni MF. Maximum power point analysis for
520 partial shading detection and identification in photovoltaic systems. *Energy Convers Manag* 2020;224:113374.
521 <https://doi.org/10.1016/j.enconman.2020.113374>.
- 522 [17] Ma M, Zhang Z, Xie Z, Yun P, Zhang X, Li F. Fault diagnosis of cracks in crystalline silicon photovoltaic
523 modules through I-V curve. *Microelectron Reliab* 2020;114:113848.
524 <https://doi.org/10.1016/J.MICROREL.2020.113848>.
- 525 [18] Zhang Z, Ma M, Wang H, Wang H, Ma W, Zhang X. A fault diagnosis method for photovoltaic module current
526 mismatch based on numerical analysis and statistics. *Sol Energy* 2021;225:221–36.
527 <https://doi.org/10.1016/J.SOLENER.2021.07.037>.
- 528 [19] Miwa M, Yamanaka S, Kawamura H, Ohno H, Kawamura H. Diagnosis of a power output lowering of PV
529 ARRAY with a (-dI/dV)-V characteristic. *Conf. Rec. 2006 IEEE 4th World Conf. Photovolt. Energy Convers.*,
530 vol. 2, Waikoloa, HI, USA: IEEE Computer Society; 2006, p. 2442–5.
531 <https://doi.org/10.1109/WCPEC.2006.279703>.
- 532 [20] Bressan M, El Basri Y, Galeano AG, Alonso C. A shadow fault detection method based on the standard error
533 analysis of I-V curves. *Renew Energy* 2016;99:1181–90. <https://doi.org/10.1016/j.renene.2016.08.028>.
- 534 [21] Ali MH, Rabhi A, Hajjaji A El, Tina GM. Real Time Fault Detection in Photovoltaic Systems. *Energy Procedia*
535 2017;111:914–23. <https://doi.org/10.1016/j.egypro.2017.03.254>.
- 536 [22] Bonsignore L, Davarifar M, Rabhi A, Tina GM, Elhajjaji A. Neuro-Fuzzy Fault Detection Method for
537 Photovoltaic Systems. *Energy Procedia* 2014;62:431–41.
538 <https://doi.org/https://doi.org/10.1016/j.egypro.2014.12.405>.
- 539 [23] Chen Z, Chen Y, Wu L, Cheng S, Lin P. Deep residual network based fault detection and diagnosis of
540 photovoltaic arrays using current-voltage curves and ambient conditions. *Energy Convers Manag*
541 2019;198:111793. <https://doi.org/10.1016/j.enconman.2019.111793>.
- 542 [24] Köntges M, Oreski G, Jahn U, Herz M, Hacke P, Weiss K-A, Razongles G, Paggi M, Parlevliet D, Tanahashi T,
543 French R. Assessment of photovoltaic module failures in the field. 2017. [http://www.iea-](http://www.iea-pvps.org/index.php?id=92&eID=dam_frontend_push&docID=4059)
544 [pvps.org/index.php?id=92&eID=dam_frontend_push&docID=4059](http://www.iea-pvps.org/index.php?id=92&eID=dam_frontend_push&docID=4059).
- 545 [25] Salem F, Awadallah MA. Detection and assessment of partial shading in photovoltaic arrays. *J Electr Syst Inf*
546 *Technol* 2016;3:23–32. <https://doi.org/10.1016/j.jesit.2015.10.003>.
- 547 [26] Dirnberger D, Kraling U. Uncertainty in PV module measurement-part I: Calibration of crystalline and thin-film
548 modules. *IEEE J Photovoltaics* 2013;3:1016–26. <https://doi.org/10.1109/JPHOTOV.2013.2260595>.
- 549 [27] Reise C, Müller B, Moser D, Belluardo G, Ingenhoven P. Report IEA-PVPS T13-12:2018 Uncertainties in PV
550 System Yield Predictions and Assessments. 2018. [https://iea-pvps.org/wp-](https://iea-pvps.org/wp-content/uploads/2020/01/Uncertainties_in_PV_System_Yield_Predictions_and_Assessments_by_Task_13.pdf)
551 [content/uploads/2020/01/Uncertainties_in_PV_System_Yield_Predictions_and_Assessments_by_Task_13.pdf](https://iea-pvps.org/wp-content/uploads/2020/01/Uncertainties_in_PV_System_Yield_Predictions_and_Assessments_by_Task_13.pdf)
552 (accessed May 18, 2020).
- 553 [28] Li B, Migan-dubois A, Delpha C, Diallo D. Analysis of the performance of I-V curve correction methods in the
554 presence of defects. 37th Eur. Photovolt. Sol. Energy Conf. Exhib. (EU PVSEC), Virtual Conference:
555 EUPVSEC; 2020, p. 1599–603. <https://doi.org/10.4229/EUPVSEC20202020-5CV.3.54>.
- 556 [29] Li B, Migan-Dubois A, Delpha C, Diallo D. Evaluation and improvement of IEC 60891 correction methods for
557 I-V curves of defective photovoltaic panels. *Sol Energy* 2021;216:225–37.
558 <https://doi.org/10.1016/j.solener.2021.01.010>.

- 559 [30] IEC 60891. Photovoltaic devices - Procedures for temperature and irradiance corrections to measured I-V
560 characteristics. 2009.
- 561 [31] Phang JCH, Chan DSH. A review of curve fitting error criteria for solar cell I-V characteristics. *Sol Cells*
562 1986;18:1–12. [https://doi.org/10.1016/0379-6787\(86\)90002-5](https://doi.org/10.1016/0379-6787(86)90002-5).
- 563 [32] Wang Z, Oates T. Imaging Time-Series to Improve Classification and Imputation. *IJCAI Int. Jt. Conf. Artif.*
564 *Intell.*, Buenos Aires, Argentina: International Joint Conferences on Artificial Intelligence; 2015, p. 3939–45.
- 565 [33] Marwan N, Carmen Romano M, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems. *Phys*
566 *Rep* 2007;438:237–329. <https://doi.org/10.1016/j.physrep.2006.11.001>.
- 567 [34] Khanna K, Martha TR, Roy P, Kumar KV. Effect of time and space partitioning strategies of samples on
568 regional landslide susceptibility modelling. *Landslides* 2021 186 2021;18:2281–94.
569 <https://doi.org/10.1007/S10346-021-01627-3>.
- 570 [35] Sarkar BK. A case study on partitioning data for classification. *Int J Inf Decis Sci* 2016;8:73–91.
571 <https://doi.org/10.1504/IJIDS.2016.075788>.
- 572 [36] Belkina AC, Ciccolella CO, Anno R, Halpert R, Spidlen J, Snyder-Cappione JE. Automated optimized
573 parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large
574 datasets. *Nat Commun* 2019;10:1–12. <https://doi.org/10.1038/s41467-019-13055-y>.
- 575 [37] Chine W, Mellit A. ANN-based fault diagnosis technique for photovoltaic stings. 2017 5th Int. Conf. Electr.
576 *Eng. - Boumerdes, Boumerdes, Algeria: IEEE; 2017, p. 1–4. https://doi.org/10.1109/ICEE-B.2017.8192078.*
- 577 [38] Lu X, Lin P, Cheng S, Lin Y, Chen Z, Wu L, Zheng Q. Fault diagnosis for photovoltaic array based on
578 convolutional neural network and electrical time series graph. *Energy Convers Manag* 2019;196:950–65.
579 <https://doi.org/10.1016/j.enconman.2019.06.062>.
- 580