Health monitoring of photovoltaic modules using electrical measurements

Baojie Li

université PARIS-SACLAY

Health monitoring of photovoltaic modules using electrical measurements

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 575, Electrical, Optical, Bio-Physics and Engineering (EOBE)
Spécialité de doctorat: Génie Electrique
Unités de recherche: Université Paris-Saclay, CentraleSupélec, CNRS, GeePs, L2S, Gif Sur Yvette, 91192, France
Référent : CentraleSupélec

Thèse présentée et soutenue à Paris-Saclay, le 08/10/2021, par

Baojie LI

Composition du Jury

Eric LABOURE Professeur des Universités, Université Paris-Saclay, GeePs	Président du jury
Maxime DARNON Chargé de Recherche – HDR, Université de Sherbrooke, LN2	Rapporteur & Examinateur
Bertrand RAISON Professeur des Universités, Université Grenoble Alpes, G2ELab	Rapporteur & Examinateur
Mohamed BENBOUZID Professeur des Universités, Université de Bretagne Occidentale, IRDL	Examinateur
Tianzhen WANG Professeure, Shanghai Maritime University	Examinatrice
Demba DIALLO Professeur des Universités, Université Paris-Saclay, GeePs	Directeur de thèse
Claude DELPHA Maitre de Conférence - HDR, Université Paris-Saclay, L2S	Co-Encadrant
Anne MIGAN-DUBOIS Professeure des Universités, Université Paris-Saclay, GeePs	Co-Encadrante

Thèse de doctorat

NNT : 2021UPAST087

Acknowledgement

The three years of my Ph.D. life flies quickly, and every bit of it left me with deep memories. In this land, there are many people and things worthy of my thanks.

First of all, I would like to express my sincere gratitude to my Ph.D. supervisors, Prof. Demba DIALLO, Prof. Claude DELPHA and Prof. Anne MIGAN-DUBOIS, for their instructive advice and useful suggestions on my thesis.

Secondly, I would also like to express my gratitude to the members of my jury, Prof. Mohamed BENBOUZID, M. Maxime DARNON, Prof. Bertrand RAISON, Prof., Eric LABOURE, Prof. Tianzhen WANG for my Ph.D. defense and their constructive comments on my work. I appreciate our fruitful discussion very much.

Besides, special thanks should go to my friends and colleagues in GeePs, L2S and other laboratories, especially M. Guanli HE, M. Junjie YANG, M. Haiteng SUN, M. Qihao GUO, Mme. Yao PEI, Mme. Junling QU, Mme. Sufang LI, Mme. Tong YU, M. Lingle Chen, M. Tanguy PHULPIN, M. Loïc QUEVAL. We have spent a long time together. This journey was interesting because of you.

Last but not least, I should express my gratitude to my family, especially my parents. Their support and love make me courageous.

At the end, a little poem in my native language may conclude my feelings:

秋冬春夏，三度轮回，
每个季节打上了独特的印记，
那些爱我的和我爱的人们，
谢谢你们。
与这一切告别的时刻，
我会不舍，但也注定会满载而去，
因为巴黎这场流动的盛宴，
已经流入了我的生命，
直到尽头。

保杰，
2021年10月4日于巴黎

Table of contents

List of figures

List of tables

List of abbreviations

Terminology

1D, 2D	1 Dimension, 2 Dimension
AC	Alternating Current
AF	Arc Fault
ANN	Artificial Neural Network
ART	Adaptive Resonance Theory
BP	Back Propagation
BPD	Bypass Diode
CCC	Current Carrying Conductor
CCD	Charged Coupled Device
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
c-Si	Crystalline Silicon
DBN	Deep Belief Network
DC	Direct Current
DnCNN	Denoising CNN
DHI	Diffuse Horizontal Irradiance
DNN	Deep Neural Network
DT	Decision Tree
DWT	Discrete Wavelet Transform
EL	Electroluminescence
ELM	Extreme Learning Machine
ENN	Extension Neural Network
EVA	Ethylene Vinyl Acetate
FDD	Fault Detection and Diagnosis
FF	Fill Factor
FL	Fuzzy Logic
FT	Fourier Transform
GADF	Gramian Angular Difference Field
GBSSL	Graph-Based Semi-Supervised Learning
GF	Ground Fault
GHI	Global Horizontal Irradiance
GK-FCM	Gaussian Kernel Function-based Fuzzy C-Means
GLCM	Grey Level Co-occurrence Matrix
ICA	Independent Component Analysis
IGBT	Insulated Gate Bipolar Transistor
IR	Infrared Thermography

kNN	k-Nearest Neighbors
LCOE	Levelized Cost of Electricity
LDA	Linear Discriminant Analysis
LID	Light Induced Degradation
LLF	Line-to-Line Fault
mc-Si	Multicrystalline Silicon
MLP	Multi-layer Perception
MLT	Machine Learning Technique
MPP(T)	Maximum Power Point (Tracking)
NBC	Naive Bayes Classifier
OC	Open Circuit
OVA	One-versus-all
OVO	One-versus-one
P1/2/3	Procedure 1/2/3
PCA	Principal Component Analysis
PID	Potential Induced Degradation
PID-s	Shunt PID
PL	Power Loss
PNN	Probabilistic Neural Network
PS	Partial Shading
PV	Photovoltaic
QDA	Quadratic Discriminant Analysis
RBF	Radical Basis Function
ReLU	Rectified Linear Unit
RDA	Regularized Discriminant Analysis
RF	Random Forest
RMS	Root Mean Square
RMSE	Root Mean Squared Error
RP	Recurrence Plot
SC	Short Circuit
sc-Si	Single Crystalline Silicon
SNR	Signal-to-Noise Ratio
SPE	Squared Prediction Error
SR	Security Risk
STC	Standard Test Condition
SVM	Support Vector Machine
T^2	T-squared
TC	Temperature Coefficient
TCO	Transparent Conductive Oxide
t-SNE	t-Distributed Stochastic Neighbor Embedding

UAV	Unmanned Aerial Vehicle
UV(F)	Ultraviolet (Fluorescence)
Vis.	Visible
WT	Wavelet Transform

Symbol

a	Irradiance correction factor
α	Absolute temperature coefficient of I_{SC} (A/°C)
α_{rel}	Relative temperature coefficient of I_{SC} (%/°C)
β	Absolute temperature coefficient of V_{OC} (V/°C)
β_{rel}	Relative temperature coefficient of V_{OC} (%/°C)
C_{E_g}	Temperature coefficient of bandgap energy (%/°C)
D_{matrix}	Size of matrix
Δ_{output}	Largest neuron output minus the second largest output for the neurons in the output layer of ANN
E_{curve}	Curve error (%)
E_g	Bandgap energy (W)
σ_{dis_env}	Standard deviation of environmental noise
σ_{dis_rand}	Standard deviation of random error
G	Irradiance (W/m^2)
G_{poa}	Global in-plane irradiance (W/m^2)
$GainPS$	Block gain used in simulation to control PS degree
I	Current (A)
I_0	Diode saturation current (A)
I_{AC}	AC current
I^c	Current of corrected curve (A)
I^{real}	Current of real curve at STC (A)
I_{MPP}	Current at MPP (A)
I_{rp}	Current at the inflection point under PS (A)
I_{SC}	Short-circuit current (A)
I_{SC}^{real}	I_{SC} of real curve at STC (A)
κ	Curve correction factor
k_B	Boltzmann constant
N	Number of points in a vector
V	Voltage (V)
V_{AC}	AC voltage
V_{MPP}	Voltage at MPP (V)
V_{OC}	Open-circuit voltage (V)

V_{OC}^{STC}	V_{OC} under STC in healthy condition (provided in datasheet) (V)
V_{OC}^{real}	V_{OC} of real curve at STC (V)
V_{rp}	Voltage of inflection point under PS (V)
P_m	Maximum Power (W)
PR	Performance Ratio
q	Absolute value of electron's charge
R	Recurrence plot matrix
E_{dis_env}	Error rate of environmental noise (%)
R_{dis_rand}	Ratio of random error (%)
RE	Relative error (%)
R_{OC}	Resistance used in simulation for OC (Ω)
R_s	Series resistance (Ω)
R_{s_degra}	Resistance used in simulation for Rs degradation (Ω)
R_{SC}	Resistance used in simulation for SC (Ω)
R_{sh}	Shunt resistance (Ω)
R_{sh_degra}	Resistance used in simulation for Rsh degradation (Ω)
std	Standard deviation
T_a	Ambient temperature (°C)
T_c	Cell temperature (°C)
T_m	Module temperature (°C)
γ	Interpolation constant for P3 correction method

List of publications

International journal papers:

1. LI B., DELPHA C., MIGAN A., DIALLO D. Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review. Renewable and Sustainable Energy Reviews (2020): 110512. doi: 10.1016/j.rser.2020.110512

2. LI B., MIGAN A., DELPHA C., DIALLO D. Evaluation and improvement of IEC 60891 correction methods for IV curves of defective photovoltaic panels. Solar Energy 216 (2021): 225-237. doi: 10.1016/j.solener.2021.01.010

3. LI B., DELPHA C., MIGAN A., DIALLO D. Fault Diagnosis of Photovoltaic Panels Using Full I-V Characteristics and Machine Learning Techniques. Energy Conversion and Management 248 (2021). doi: 10.1016/j.enconman.2021.114785

International Conference papers

1. LI B., MIGAN A., DELPHA C., DIALLO D. Irradiance Dependence of the Short-Circuit Current Temperature Coefficient of sc-Si PV Module. IEEE PVSC 2020, virtual, 15 Jun.-21 Aug. 2020. doi: 10.1109/PVSC45281.2020.9300416

2. LI B., MIGAN A., DELPHA C., DIALLO D. Analysis of the performance of I-V curve correction methods in the presence of defects. EUPVSEC 2020, virtual, 7-11 Sept. 2020. doi: 10.4229/EUPVSEC20202020-5CV.3.54

National Conference papers:

1. LI B., MIGAN A., DELPHA C., DIALLO D. Complexity Analysis of Convolutional Neural Network Applied to PV Fault Diagnosis via Image Processing. JNPV 2019, Dourdan France, 3-6 Dec. 2019

2. LI B., MIGAN A., DELPHA C., DIALLO D. Are the curve correction methods of IEC 60891 suitable for I-V curves measured under faulty conditions? JNPV 2020, Dourdan France (Oral)

3. LI B., DELPHA C., MIGAN A., DIALLO D. Diagnostic de défauts photovoltaïques par analyse des caractéristiques I-V complètes et utilisation de techniques d'apprentissage automatique, JNPV 2021, Dourdan France, 30 Nov.-3 Dec. 2021 (Oral)

General introduction

Background and motivation

Solar energy, as typical renewable energy, presents various merits [1]. It is pollution-free and causes no greenhouse gases (like carbon dioxide, methane, and nitrous oxide) to be emitted after installation [2]. Besides, this clean power is available all the year, even under cloudy days [3]. In recent years, the solar photovoltaic (PV) installed capacity has experienced rapid growth [4]. The global total cumulative installed capacity in 2020 has reached more than 758.9 GW [5]. China leads the cumulative capacity share with 253.4 GW, followed by the European Union (151.2 GW) and the USA (93.2 GW). Among the countries in the European Union, Germany is at the top with 53.9 GW, then followed by Italy (21.7 GW), Spain (12.7 GW), France (10.9 GW), and the Netherlands (10.2 GW).

In fact, for the year 2020, the COVID-19 pandemic has resulted in a severe economic downturn and social damage. However, this pandemic did not significantly impact the PV market development [5]. According to BP World Energy & Ember, global power generation from PV increased by 20.2% in 2020, to 844.4 TWh [6]. With this increase, the solar PV share in global electricity generation has reached about 3.3%. Thus, the resiliency of the PV market is remarkable and shows the potential to limit the impact brought by the pandemic. As for the future scope, according to the "Net Zero Emissions by 2050" scenario of the WEO 2020, the worldwide installed PV capacity would increase to 3929 GW by 2030 [7]. According to this scenario, solar electricity will reach about 5420 TWh or 15.9% of the global energy production [8].

In addition to the environmental goals, the prosperity of the PV industry is also due to the gradual falling prices of PV modules and the progress in plant engineering and construction. Actually, the competitiveness of PV projects measured by the Levelized Cost of Electricity (LCOE) has also continuously decreased in recent years [9]. In this context, the energy generation of a PV power plant plays a significant role in the market assessment of PV projects. In the fact sheets on PV, IEA has announced that "PV is the cheapest electricity source almost everywhere" [10].

With the rapid development of the PV industry and the increase in the installed capacity, efficient operation and maintenance strategies are increasingly required [8]. In fact, manufacturing, transportation, installation, and operating conditions can cause PV cell or module failures [11], [12]. In the case of a PV array or power plant, which is composed of several electrically connected modules, any fault in one cell or module affects the performance of the whole array or system. The PV faults could cause a severe safety hazard, e.g., fire risk, electrical shock, physical danger [13], or power loss

[14]. Therefore, to assure the reliability, availability and safety of the PV installations, their health status should be monitored regularly to prevent from failures and contribute to an efficient condition-based maintenance policy. To this end, efficient PV fault detection and diagnosis (FDD) strategies are required.

The cornerstone of a PV FDD strategy is the PV data, of which the electrical types are the most common ones because they are able to reflect the actual operating condition of a PV module or array [15]. Among the PV electrical data, the current-voltage characteristic (I-V curve) contains the most information about the health condition of a PV module or array, which should allow for a more accurate diagnosis [16]. As for the availability of I-V curves, common I-V tracers already support the measurement for a single module or small-scale string or array. In recent years, the hardware solutions (integrated at inverter level) have become commercially available to measure I-V curves periodically at the power plant level [17], [18]. Therefore, it is expected in the near future that I-V curve data will be measurable from almost all the common PV facilities. With this in mind, this thesis aims to propose a strategy based on I-V curve data for accurate and robust diagnosis of common faults in PV arrays.

Problem statement

- Lack of effective use of the full I-V curve information for the PV FDD

In the literature, different approaches are proposed to manipulate the I-V curve for PV diagnosis, like key parameters extraction [19]–[21], derivative calculation [22], [23], and construction of feature matrix based on resampling [24]. However, these approaches all have major limits. For example, the methods based on the extraction of key parameters [19]–[21] (like open circuit voltage, short circuit current, voltage and current at maximum power point) capture only partial information from the I-V curve, and thus sometimes fail to classify similar fault conditions, like when one or more modules are shaded in one PV string. The method based on the first or second-order derivative of the whole curve is only suitable to identify partial shading (PS) fault [22], [23]. Authors in [24] integrate resampled full I-V curves with irradiance (G) and module temperature (T_m) into 4-column matrices (named IVGT matrix) as features for FDD. However, the duplication of information compromises the diagnosis accuracy. Therefore, effective use of the full I-V curve for the diagnosis of multiple PV faults is still an opened research topic.

- Lack of an evaluation of correction procedures for I-V curves of faulty PV panels

Since field-measured I-V curves are recorded under varying environmental conditions, the correction of curves to a common specific environmental condition is generally performed to ensure a fair comparison and facilitate the FDD task. However, to the best

of our knowledge, there are no specific methods proposed for the correction of faulty I-V curves. Therefore, most of the researchers adopt the original or simplified correction procedures from the standard IEC 60891 [25]. It should be noted that these procedures are all initially designed for the correction of curves measured from healthy PV panels. The suitability of these procedures for the correction of faulty panels I-V curves is rarely investigated, and even less if the fault characteristics are distorted by the correction and thus lead to a diagnostic error. Yet, all these issues are decisive for the diagnosis of PV faults [26]. Besides, due to the difference between I-V curves from faulty and healthy PV panels, special correction procedures are expected to be designed for the curves of faulty PV panels.

- Lack of efficient feature transformation methods to improve the discriminability of I-V curve features under different faulty conditions

For the I-V curves showing similar shapes but from PV panels under different faulty conditions, the features extracted from these curves may only show little discrepancy and therefore are prone to be misclassified. The common I-V curve features in the literature (key parameters, derivatives, IVGT matrix) are based on direct extraction of features but without further processing. Therefore, the discriminability of these extracted features under different conditions is almost the same as that of original I-V curves. Under tricky fault classification cases, like when dealing with similar curves, the discriminability of features is expected to be increased with adequate transformations. This would help to improve the FDD performance. To the best of our knowledge, these methods have not been yet explored for PV FDD.

Thesis outline

Following these findings, this thesis proposes a new PV FDD methodology, which is based on an improved correction of the I-V curves of defective PV panels, achieves an efficient integration of the full I-V curves as features for diagnosis, and also deals with the transformation of the features to improve the FDD performance. To be specific, this thesis consists of the following four main chapters:

- *Chapter I*: A state of the art of the common PV array faults is presented. The fault causes, fault classification, impact of faults on the power loss and safety risk, and the frequency of fault occurrence will be all analyzed. A method for classifying the faults according to the level is proposed, i.e., at the cell, module, or array level. Through this classification method, the usual faults are presented based on a literature review. With joint consideration of fault impact, frequency of occurrence and reproducibility, the faults considered in this thesis are selected.

- *Chapter II*: The common FDD techniques for PV faults in literature are reviewed through two categories: visual inspection and automatic analysis. The automatic FDD methods will be described in a four-step scheme: modeling, pre-processing, feature extraction, and feature analysis. The proposed FDD strategy in our thesis will also be accordingly presented.

- *Chapter III*: The correction procedures in the IEC standard are evaluated for I-V curves from faulty PV arrays. The correction performance on the whole curve and the key curve parameters are addressed. Besides, the impact of fault severity on the correction performance is also analyzed. A new correction procedure will be proposed and compared with the existing ones in the IEC standard.

- *Chapter IV*: A novel PV FDD methodology based on full I-V curves is presented. This methodology will be detailed from the four-step FDD scheme. The modeling of a PV array and the preparation of the simulated dataset will be presented. Three methods to extract fault features from full I-V curves for diagnosis are proposed, where one is based on the resampled vectors of current, and the other two are based on feature transformation. Six common machine learning classifiers are tuned and evaluated for PV fault classification. The configuration of the classifiers, the diagnosis results, and the impact of various factors are analyzed. Finally, the best-trained classifier will be validated with field-measured dataset.

- Conclusions and perspectives of the thesis will be given. Besides, the summary in French is also presented at the end.

Contribution of thesis

The contribution of this thesis lies in the following points:

- A four-step FDD scheme is firstly applied in the PV FDD domain, which allows a systematic comprehension of the various PV FDD cases and facilitates the comparison from the modeling, pre-processing, feature extraction, and feature analysis steps;

- An improved I-V curve correction procedure is proposed and tested exhibiting overall better performance than the traditional single-curve-based correction procedures proposed in IEC 60891 standard;

- A new methodology based on full I-V curves is proposed and outperforms traditional methodology based on partial use of I-V curves;

- Two feature transformation techniques are applied to I-V curves, and proved able to improve the discriminability of features, which allows the MLT classifiers to

have higher robustness to additional disturbance and better classification performance with both simulation and field dataset.

Chapter I

Photovoltaic array faults: State of the art

I.1. Introduction

A photovoltaic (PV) array fault is a defect that either affects the output power or poses a safety risk, while the difference in appearance due to fabrication that is not inducing safety risk or power loss will not be considered as fault [27]. Due to various internal and external factors, outdoor PV arrays are subject to a great variety of faults, which could occur in nearly all the components of the PV array. These faults may lead to different levels of degradation, power loss or sometimes even fire risk. Before the design of effective detection and diagnosis strategies, it is essential to understand the common faults in PV array well.

This chapter aims to provide a comprehensive state of the art on common PV array faults. Section I.2 introduces the leading causes of PV faults, while in Section I.3, we present the faults in detail through a classification method. In Section I.4, we analyze the impact of faults in terms of power loss and safety. The frequency of occurrence of faults will be presented in Section I.5. Section I.6 introduces the faults studied in our work, and Section I.7 closes the chapter.

I.2. Faults causes

I.2.1. Defects due to human error

Manual operations, from transport to installation of PV modules, are one of the first sources of PV faults.

Transport is the first critical stage of the life cycle [28]. Shocks, vibrations, lack of a rigid support pallet during transport, or dropping during unstacking can lead to glass breakage or cell cracking [29].

During installation, clamping is a frequent cause of glass breakage [30], especially when mounting multiple PV modules. Screws that are too tight, clamps that are too short or too narrow can generate high mechanical stresses and lead to glass breakage. Another potential risk is the incorrect wiring of connectors. Poorly fitted or crimped connectors can lead to an open circuit, line fault, earth fault, or power loss [31]. Worse still, in most cases, connectors are usually placed near flammable materials, such as wooden roof beams or thermal insulation components. Under these conditions, arcing faults can lead to fires.

I.2.2. Environmental factors

Outdoor PV modules are exposed to long-term harsh environmental conditions, which poses a great risk to the normal operation of a PV system. The environmental factors could be classified into permanent and non-permanent types [32]. The permanent ones

result in irreversible fault of PV module or components, like lightning [33], heavy snow [27], or hail [34] , causing mechanical damages (detachment, breakage) to the PV module frames or glasses [15]. The glass breakage leads to the penetration of oxygen and moisture into the PV module and causes the corrosion of electrical circuits, which provokes the loss of performance.

Contrastingly, the non-permanent factors can be of short duration, e.g., dust, soiling, shadow of buildings or trees. They lead to shading fault and may cause localized heating or hot spot [35]. These causes can generally be removed automatically or manually by the maintenance team.

I.2.3. Material interaction

A PV module consists of multiple layers of different materials bonded adhesively. For the PV modules occupying the main commercial market share, i.e., crystalline silicon (c-Si) and thin film type, the common structure usually includes the layers of front glass, encapsulant layer, solar cells/ thin film substrate, and back sheet. The interactions at the various interfaces, e.g., front glass/ encapsulant, encapsulant/ solar cell, encapsulant/ back sheet, can be responsible for PV faults.

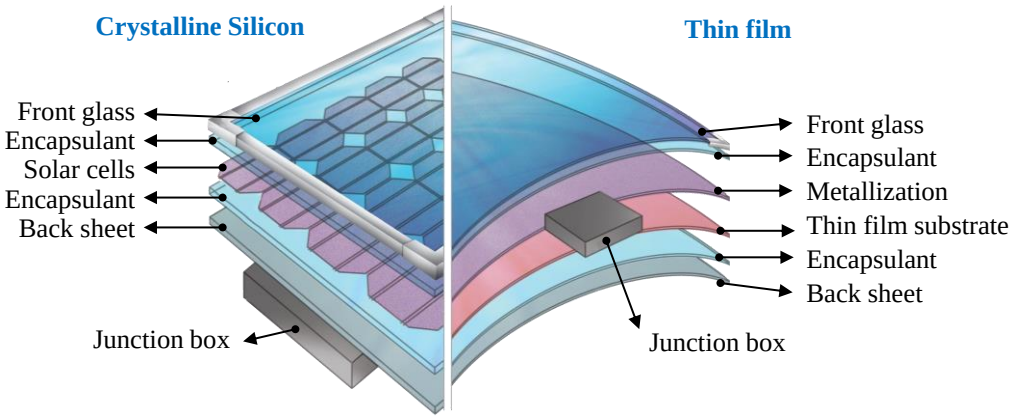


Figure I-1: Structure of standard crystalline silicon and thin film PV module [36]

The interactions can give rise to visible phenomena such as discoloration of the encapsulant (yellowing or browning), corrosion, cell crack or delamination. In addition, these interactions can also result in electrical degradations such as disconnection of cells, short circuits, and potential induced degradation (PID) which have no visual effects.

It should be noted that environmental factors (humidity, heat, UV radiation, etc.) can also accelerate the formation of these defects, which is also related to climatic conditions. A global study [37] of 1.9 million modules installed in different climates found that climatic conditions have a strong impact on the occurrence of defects due to

material interactions. These effects are more pronounced in hot arid climates than in tropical and temperate climates.

I.2.4. Cause-effect circle

In fact, the mechanism of most defects in photovoltaic modules can be quite complex. It is difficult, and sometimes impossible, to attribute a single origin to the formation of a defect. Indeed, the occurrence of a defect is often accompanied by a change in other properties of the PV module: mechanical, chemical or electrical, which will in turn aggravate the original defect or generate other defects.

Cause and effect relationships between faults are complex but need to be understood or otherwise identified to improve design and monitoring. It is also important to analyze and identify the faults impact. In [38], a causal circle is established between mismatch and degradation as shown in Figure I-2. The mismatch is recognized as the intermediary fault mode, which could be caused by different operating parameters (temperature, voltage, current) and external factors (environmental factors and human errors).

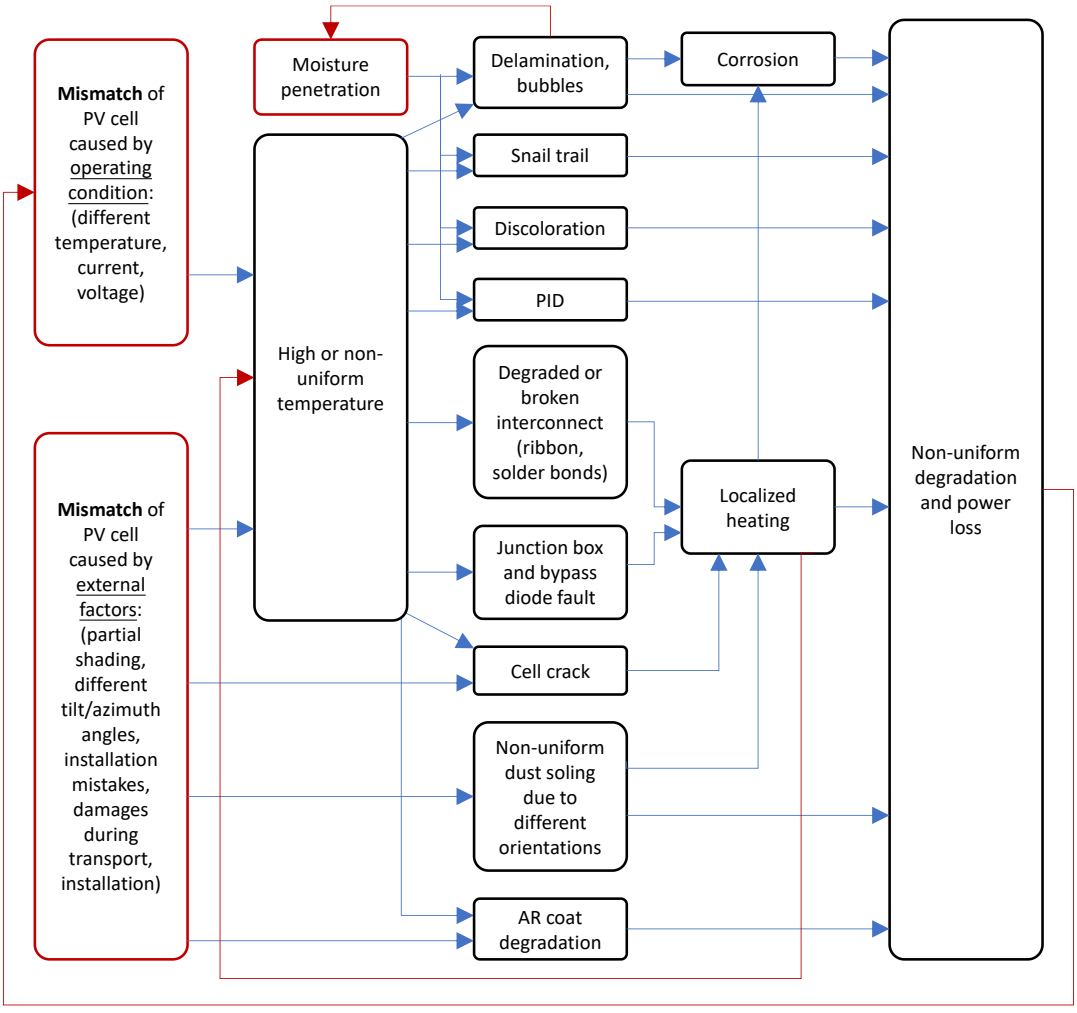


Figure I-2: Cause-effect circle between mismatch and degradation [38]

The causal circle reveals the relationship between mismatch and degradation. Besides, other intermediary fault modes have also been identified. For example, mismatch due to operating conditions or external factors can lead to the high or non-uniform distribution of module temperature, which, combined with other causes, induces the commonly encountered PV faults displayed in the central column. Then, these faults, directly or indirectly, are responsible for non-uniform degradation and power loss. The non-uniformity of the degradation, in turn, aggravates the mismatch level and closes the causal circle.

I.3. Presentation of faults

After identifying the causes, we present in this section the different PV panel defects in detail. First, we summarize the popular classifications of PV faults in the literature. Then, a new three-level classification is proposed, based on which the common PV panel faults are categorized and presented.

I.3.1. Fault classification

A well-designed fault classification permits a better understanding of the similarities and differences between the PV faults under different categories.

In the literature, several classifications have been proposed for common PV faults. They are based on different criteria, such as duration, degree of severity, location, cause, or nature. Table I-1 shows some classical classification approaches.

Table I-1: Common classification of PV faults

Ref.	Classification criteria	Categories	PV faults
[32]	Duration and degree of severity	Permanent	Cell crack, line-line fault, ground fault, arc fault, bypass diode fault, junction box fault, etc.
		Intermittent	Shading, soiling, heavy snow, hail, etc.
		Incipient	Abnormal degradation, corrosion, snail, delamination, etc.
[26]	Components of PV system	Module faults	Cell crack, hot spot, corrosion, PID, Light induced degradation (LID), back sheet adhesion loss, shading/soiling, etc.
		Inverter faults	Manufacturing and design faults, control faults, electrical components faults, etc.
		Others	Bypass diode fault, junction box fault, mismatch, line-line fault, ground fault, arc fault, etc.
[39]	Cause and nature of PV faults	Physical	Cell crack, bypass diode fault, temperature anomaly, abnormal degradation, etc.
		Environmental	Partial shading, bird drop, soiling, etc.
		Electrical	Line-line fault, ground fault, open-circuit, arc fault, maximum power point tracking (MPPT) fault, converter switch fault, battery bank fault, islanding operation fault, grid faults, etc.

It can be observed that classification criteria may be vague, not allowing for the proper classification of certain PV defects. For example, in [32], defects categorized as incipient are likely to develop into permanent defects such as delamination or corrosion. In [39], defects are classified according to their cause and nature. However, some defects may meet several criteria simultaneously. For example, a line fault, an earth fault, may also be caused by physical damage, and classified as a physical type. Similarly, some physical faults, like bypass diode faults and abnormal degradation, could also affect the PV electrical features and hinder output power.

The proposal based on the location or components of the PV system allows a more distinct classification of PV faults, as in [26]. However, the 3 proposed categories are still relatively coarse, and some PV faults are grouped in the "other" category. We propose a three-level fault classification: cell, module and array, as shown in Figure I-3.

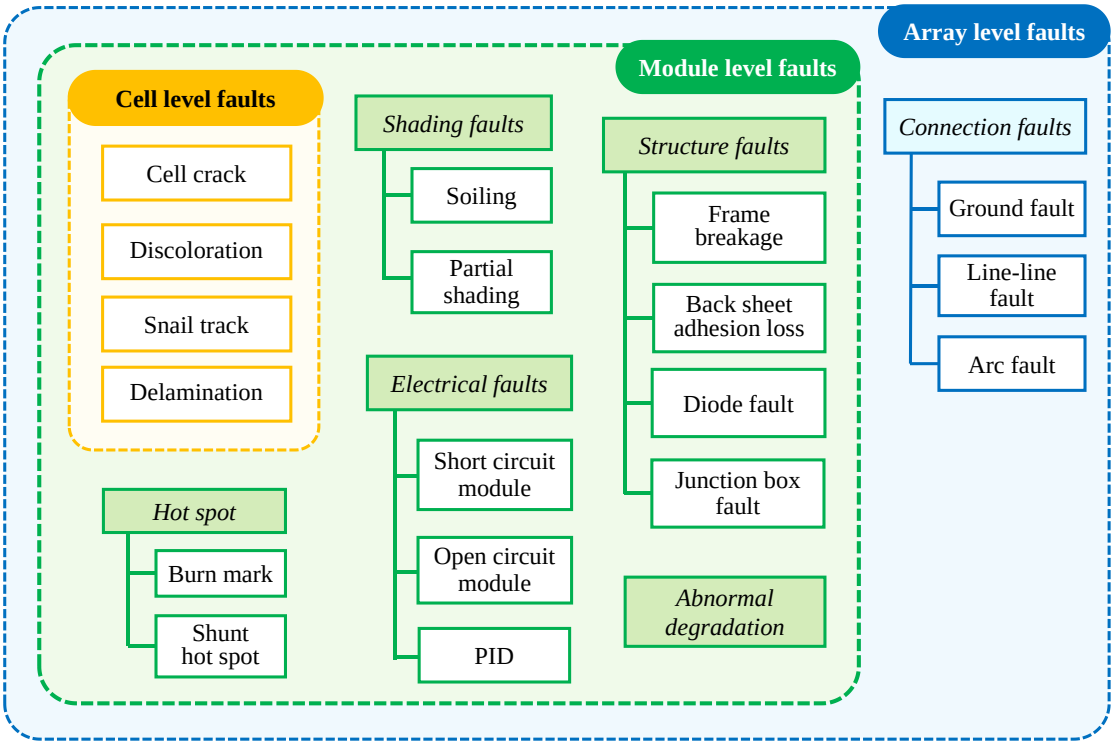


Figure I-3: Three-level PV fault tree classification

Most of the common PV faults presented in Table I-1 are categorized in the 'PV fault tree', except for LID, because LID has to be taken into account by manufacturers for the power rating as required in the standard EN 50380 [40]. Therefore, LID is generally not defined as a module fault [27]. In addition, as this thesis concentrates on the PV array faults, the faults related to the DC conditioning units (like converter, inverter, battery bank) are also excluded. Besides, faults occurring at the AC side of a grid-connected or standalone PV system are not considered. These faults can be detected and isolated using the protection standards specified in [41], [42].

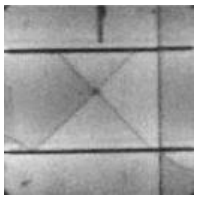
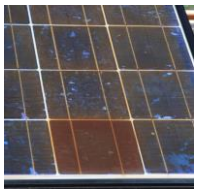
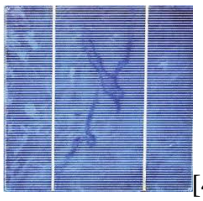
I.3.2. PV array faults

Based on the proposed 'PV fault tree' (Figure I-3), we will detail in the following the typical PV faults related to the three levels.

I.3.2.1. Cell-level faults

Cell-level faults refer to the PV faults affecting a single PV cell. These faults might spread over adjacent areas over time, but generally will not cause an overall impact on the whole surface of the PV module. These faults are introduced in Table I-2 with corresponding images.

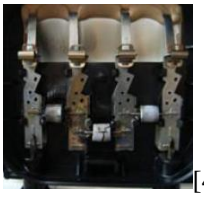
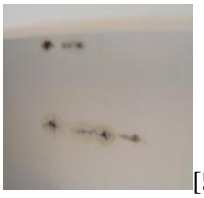
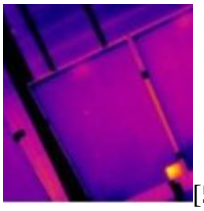
Table I-2: Cell-level faults

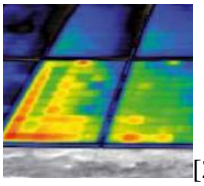
Fault	Description	Example
Cell crack	Cell cracks are the cracks induced by mechanical stress in the silicon substrate of the photovoltaic cells, which usually are invisible by naked eyes. It could rise from the production, transport, installation, and operation stages. The shape of crack has different lengths and orientation in one solar cell. Cell crack could lead to different levels of power loss, which depends on the 'inactive' area of the cell.	 [43]
Discoloration	Discoloration is generally related to the PV modules using EVA (Ethylene Vinyl Acetate) as the encapsulant material. Discoloration refers to the yellowing or browning of PV cells. It causes a change in the transmission of solar irradiance reaching the cell surface and consequently a reduction in production. Nowadays, this fault is greatly eased for the PV modules with new encapsulant material. For example, for thermoplastic polyolefin, the discoloration rate is reported around 9 times lower [44].	 [45]
Snail track	Snail track is grey/black discoloration of the silver paste of the front metallization of screen-printed solar cells. The discoloring typically occurs 3 months to 1 year after installation of the PV modules. The origin of snail track is not clear, may be due to silver particles which contains sulfur, phosphorus, or carbon. The growth speed of the snail track discoloration may be very slow [27], or it saturates directly after the first occurrence.	 [46]
Delamination	Delamination is the adhesion loss between the glass, encapsulant, active layer and back layer. For thin-film PV type, the Transparent conductive oxide (TCO) may as well delaminate from the adjacent glass layer. Bubble is also a form of delamination. Delamination will lead to optical reflection and therefore cause the decrease in power output. It also causes moisture penetration, which then leads to various chemical and physical degradation.	 [37]

I.3.2.2. Module-level faults

At the module level, the common PV faults could be categorized into shading, structure, and electrical faults. Further, hot spot, as a joint result of diode fault and partial shading or mismatch, is also a type of module-level fault. Detailed presentation of these faults is given in Table I-3.

Table I-3: Module-level faults

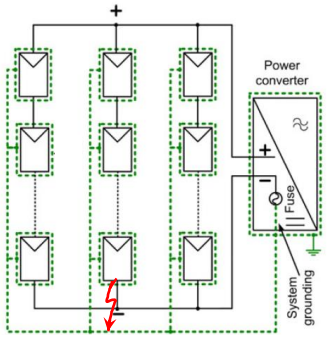
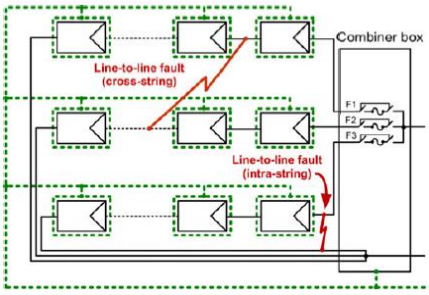
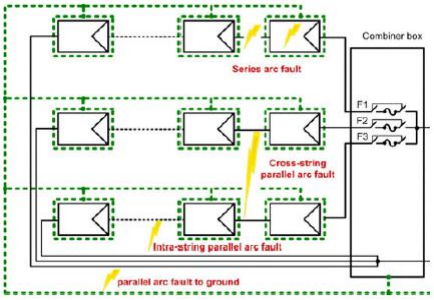
Fault	Description	Example
Shading/soiling	Shading and partial shading (PS) are usually caused by building, tree, moving cloud and so on; Soiling refers to snow, dirt, dust, and other particles that cover the surface of PV module. Shading/soiling could be classified into hard or soft type, or permanent/temporary type.	 [47]
Frame breakage	The main cause of frame breakage is heavy snow load, which will creep downhill and intrude into the space between frame and glass. It leads to deformation of module, detachment of frame from the PV glass, which will result in power loss.	 [27]
Back sheet adhesion loss	Depending on the material type of back sheet adopted in PV, causes for back sheet adhesion loss are similar to that of delamination, including temperature, moisture, mechanical stress, etc. It results in isolation default, therefore, increase the exposure to active electrical component, especially when happen near a junction box or edge of module.	 [48]
Junction box fault	Observed junction box fault includes poor fixing, bad wiring, broken connection, etc. They are mainly caused by energy overstress, rework cable during installation and rework connector, together with long heat exposure. It could result in moisture ingress, internal arcing and power loss.	 [27]
Diode fault	Common diode fault happens to the bypass diode (BPD), caused by excessive current level and improper or insufficient heat sinking. The lack of air flow in the junction box is also crucial to diode fault, particularly in the case of fast transitions shadow-sun-shadow. Burnt BPD could cause short or open circuit of the diode, and therefore different level of power loss.	 [49]
Burn mark	Partial shading + BPD fault or other mismatch fault (like low resistance defect in c-Si) could lead to energy consumption on the mismatched area instead of generation, therefore local high temperature of cell and induce burn mark. Besides, DC arc fault could also lead to burn mark. It may cause overheating, delamination or melting of material;	 [50]
Shunt hot spot	Partial shading could cause the cell turn to status of reverse biased voltage, to which thin film cells are extremely sensitive. Module current will concentrate on the shunt path and lead to hot spot. The behavior is quite different to c-Si hot spot as the BPD could not limit the reserved voltage. It is not likely to cause overheating but cause glass breakage and increase the risk of electrical shock.	 [51]

Short circuit (SC) / Open circuit (OC) module	Corrosion and structure damage are the main causes for open or short circuit of module. It leads to different level of power loss or shutdown of system and the risk of electrical shock or even arc fire.	 [52]
PID	Voltage difference between grounded framework and PV modules could establish path for leakage current and cause PID. Shunting type (PID-s) in p-type c-Si is the most frequently observed effect, which is mainly caused by migration of Na ions. Na ions drift anti-reflect coating under negatively biased condition, penetrate crystal defects, result in large shunting of cells and degrade the efficiency. It should be noted that PID is more common to the PV modules with EVA as encapsulant material. With a PID resistant material (e.g. polyolefin) gradually replacing EVA as an encapsulant, this defect has almost disappeared. [53].	 [27]
Abnormal degradation	Abnormal degradation is a comprehensive reflection of various faults, like delamination, bubbles, snail track, PID, and the related corrosion process of PV module [38]. This can be also seen via the cause-effect circle in Figure I-2. Its intuitive influence reflects in the power loss and the change of slope of the I-V curve of the module.	-

I.3.2.3. Array-level faults

At the array level, the main type of fault is the connection fault, which generally includes the earth fault, the line fault and the arc fault, as shown in Table I-4.

Table I-4: Array-level faults

Fault	Description	Illustration
Ground fault (GF)	It is caused by un unintentional low impedance path between Current Carrying Conductor (CCC) and the ground get established. For *grounded* PV system, GF causes a high current to flow through an intentional circulating path. For *ungrounded* PV system, it generates residual magnetic field between the forward and backward current flow. It causes a change in insulation resistance and a lasting loss of power.	 [54]
Line-line fault (LLF)	LLF rises from an unintentional low resistance path between two CCC with different electrical potential. It is caused by poor insulation of string connectors, accidental short-circuit between CCC, fault mounting, or external damage. It leads to high reverse current (depends on the potential difference of the location where the LLF happens) flowing to the faulty path and a lasting loss of power. LLF has 2 types, *intra-string*, and *cross-string* LLF.	 [54]
Arc fault (AF)	Several external factors could lead to the discontinuity or insulation fault of CCC and establish an air path for arc fault. Arc fault has two types: *series* and *parallel* AF (intra-string, cross-string and parallel to ground). It could occur in almost all the connection points or structures in the PV array, like cell, busbar, module, diode, string, safety devices, etc. It leads to transient but extremely high temperature that may burn the metal coating of the module. Besides, it generates high-frequency components causing serious nonlinear distortions in current and voltage.	 [54]

I.4. Impact of faults

A PV fault can result in both a safety hazard and a power loss. In this section, safety and power loss problems will first be defined with the specified impact categories. Then, the common PV faults presented in Section I.3 will be analyzed comparatively from these two aspects.

I.4.1. Safety hazard categorization

A safety hazard refers to the risks to personnel working on the installation or to bystanders. In Table I-5, three safety categories [27] are defined, which are useful in assessing the necessary action to be taken when the fault occurs.

Table I-5: Definition of safety risk categories

Safety level	Description
SR0: safe	Fault poses no clear threat to the safety.
SR1: medium	Medium possibility or potential threat to causing a fire, an electrical shock to working personnel or physical danger.
SR2: danger	High possibility to causing a fire, an electrical shock to working personnel or physical danger.

In addition, the action required at fault occurrence depends on the application class of the PV system, which is defined in the standard IEC 61730-1 [55].

I.4.2. Power loss categorization

Another important factor in assessing the impact of PV faults is the power loss level, which directly determines the functionality of the PV system. The power loss levels describe the evolution of the power variation between the initial value and the one produced over time. In most cases, this difference between the reference values can lead to inconsistent results, as the power indicated in the datasheet can deviate significantly from its actual initial value.

For legal application, to evaluate the power loss, the power printed on the PV module datasheet could be used as the reference value. For practical application, the initial power could be adopted as a reference value. For various PV array faults, the power loss level is not easy to evaluate as it could vary from case to case and is mainly determined by the fault severity and spread. However, the common trend of the power loss due to these faults can generally be categorized. In Table I-6, five classic trends of power loss over time are defined [27].

Table I-6: Definition of categories of power loss evolution over time

Power loss level	Description
PL0	Minimal power loss under detection threshold
PL-L	Linear-shaped degradation of power over time
PL-S	Saturated degradation over time (beyond a certain threshold, the power loss no longer evolves)
PL-H	Hybrid degradation over time (different evolutions at different stages or irregular evolution of the degradation)
PL-E	Exponential-shaped degradation of power over time

I.4.3. Summary of fault impact

Based on the definitions of the safety risk and loss of power loss categories, the impact of the common faults at cell, module and array levels are summarized in Table I-7.

Table I-7: Summary of the impact of common PV faults

PV fault		Safety risk			Power loss evolution				
Level	Fault	SR0	SR1	SR2	PL0	PL-L	PL-S	PL-H	PL-E
Cell level	Delamination		●				●	●	●
	Bubble		●		●				
	Discoloration		●				●	●	
	Cell crack	●	●		●	●	●		
	Snail track		●			●			
Module level	PID	●				●			
	Shading/soiling	●	●			●	●		
	Burn mark		●				●		
	Hot spot		●				●		
	Junction box fault		●		●				
	Diode fault	●	●		●		●		
	OC		●	●			●		
	SC		●	●			●		
	Frame breakage		●	●			●		
	Back sheet adhesion loss		●		●				
	Abnormal degradation	●	●			●	●		
Array level	Ground fault			●			●		
	Line-line fault			●			●		
	Arc fault			●	●				

(When a PV fault corresponds to several categories of risk for safety or power loss, this fault can induce all these effects depending on its severity level)

From the summary of fault impact, for the safety risk, most of the common PV faults result in SR1 level impact, i.e., medium safety risk of fire, electrical shock, or physical danger. However, it should be noted that some electrical faults (like open/short circuit module, ground fault, line-line fault, arc fault) and frame breakage may induce SR2 level impact, i.e., high-level danger. As for the power loss evolution, PL-S (saturated), PL-L (linear), and PL0 (negligible) are the more common. It should be noted that, for some faults of PL-S type like the electrical faults SC, OC, although the loss of power does not degrade at the first fault occurrence, the initial power loss level could still be quite high. Therefore, in order to analyze the impact of faults on power, it is necessary to examine both the initial level of loss and its evolution over time. Overall, faults with

a significant impact on safety and power loss are those that need to be detected in time as part of PV system health monitoring.

I.5. Frequency of fault occurrence

The frequency of fault occurrence is an important factor to design appropriate and relevant health monitoring and prevention strategies for PV system. In [56], a comprehensive survey on the PV system faults is carried out worldwide, covering different climate zones, and PV technologies. The results of this survey are presented in Figure I-4. PV modules from moderate climate areas or multi-crystalline silicon (mc-Si) technology constitute most of the objects surveyed. The frequency of occurrence of the different faults is presented in Figure I-5, where the most frequent faults and those that induce a measurable power loss are presented. The frequency of abrupt and gradual faults is also shown.

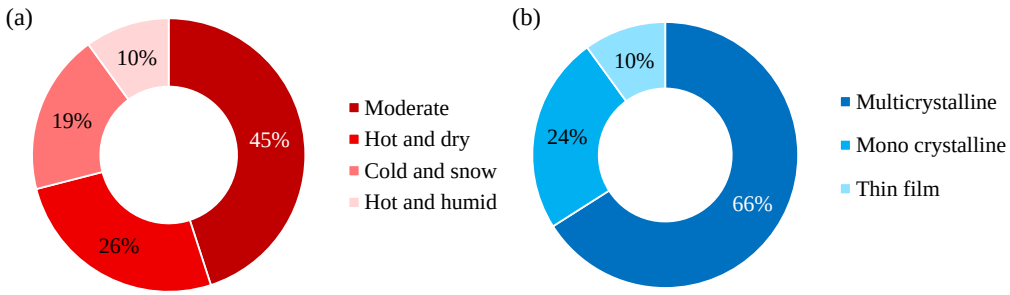


Figure I-4: Relative composition of survey (a) climate zones (b) PV technologies [56]

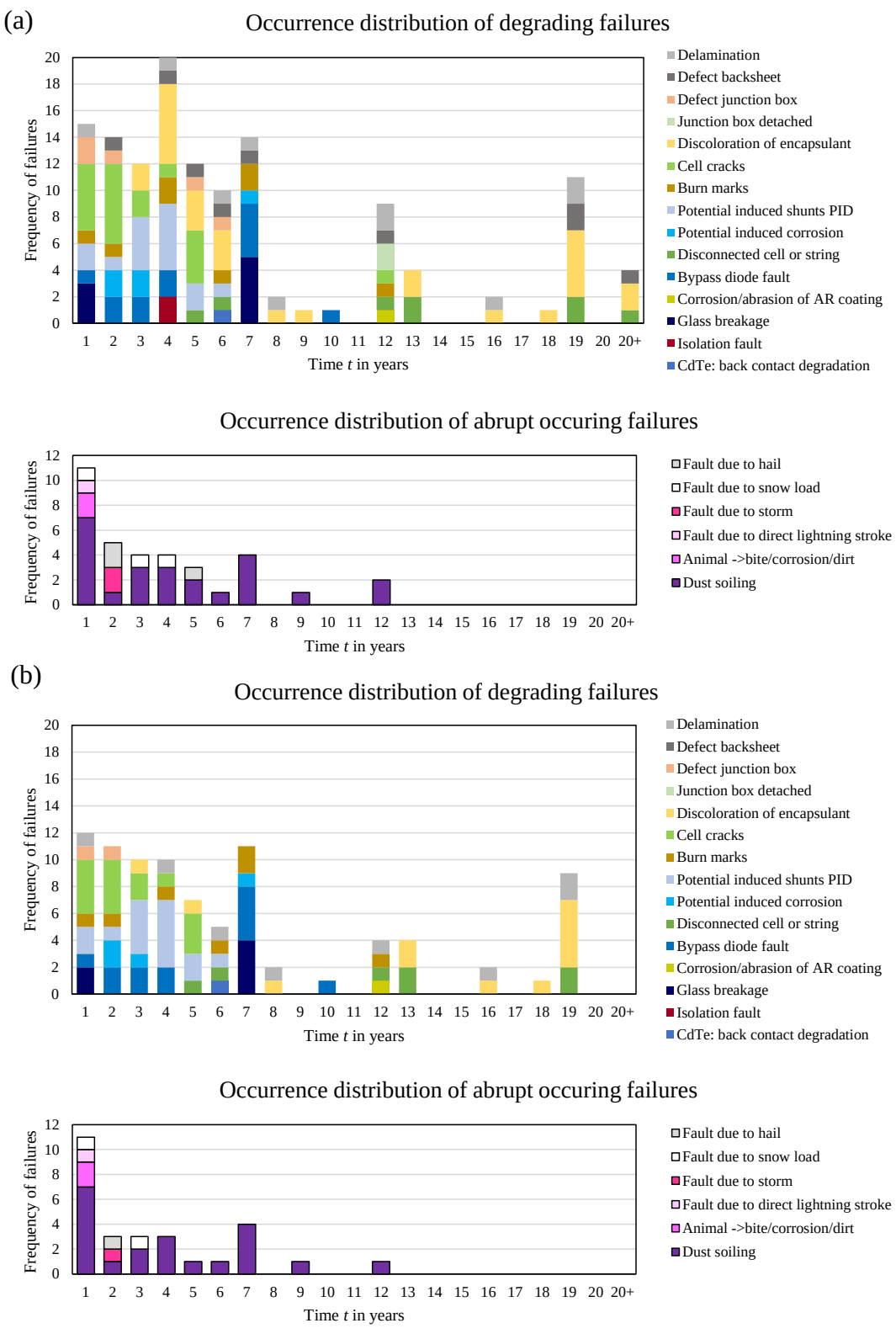


Figure I-5: Occurrence distribution of PV faults over years [56]

((a): Total occurrence of all types of faults, (b): Occurrence of detected faults causing measurable power loss)

For gradual PV faults, it is observed that the cell crack is usually reported at the very beginning (after the first or second year) of the PV system operation. The PID shunt (PID-s) appears mainly from year 3 or 4 onwards. Disconnection (for cells or strings in the module) starts from year 4 and covers the whole operating period. Discoloration of the encapsulant starts after year 3 and also extends over the years. In addition, the discoloration related to power loss reaches a high accumulation after 18 years of operation. The diode defect extends over the first 10 years of operation. The total occurrence of the other defects is too rare to allow a detailed analysis.

As far as abrupt PV failures are concerned, they are more related to environmental causes. It can be seen that dust souling is more frequently reported and extends over several years compared to the other types of defects.

Comparing the statistical results of the occurrence of all faults and faults leading to power loss, the results coincide with the study of the impact of faults on power loss in Section I.4. Some faults, e.g., junction box fault, delamination, although frequent, have a negligible impact on power output. Furthermore, some faults, e.g., delamination, although they may appear at the beginning of operation, the impact on power loss is not immediate but evolves and becomes measurable after years. Therefore, by combining both the impact and the distribution of occurrence of common PV faults, it is easier to indicate which PV faults should be prioritized for the different stages of operation of the PV array.

I.6. Studied fault cases

After having reviewed the state of the art on PV array faults, we have chosen in this work to study the faults taking into account their impact, their frequency of occurrence, the reproducibility in simulation and in real conditions. The faults chosen are PS, SC module, OC string, and abnormal degradation.

PS is chosen because of its universality under real conditions and ease of reproduction. SC and OC faults are the common electrical faults in PV panels that are likely to cause large power losses. Abnormal degradation reflects various defects, such as delamination, bubbles, snail marks and the associated corrosion process of PV modules. Moreover, it is easily reproducible in simulations and field tests. Thus, it is also considered in this thesis. The detailed configuration of these faults and the presentation of the PV array to be studied will be detailed in the following chapters.

I.7. Conclusion

In this section, we have presented a state of art on PV array faults. The three main origins of faults and the causal relationships between mismatch and abnormal degradation have been identified.

The main PV fault classification tools reported in the literature are analyzed and summarized. We have proposed a method for classifying faults according to their level: at the cell, module, or array level. Finally, through this approach, the usual faults are presented based on a literature review.

The impact of each fault is studied with regard to power loss and safety risk.

It should be noted that the different faults occur at different stages of the PV field operation. Thus, the frequency of occurrence of faults over time is also addressed. With a joint consideration of fault impact, frequency of occurrence and reproducibility, the faults PS, SC module, OC string, and abnormal degradation are chosen as the faults to be addressed in this thesis.

Chapter II

Fault detection and diagnosis of photovoltaic array: State of the art

II.1. Introduction

Fault detection and diagnosis (FDD) is a fundamental issue to ensure the normal operation of a PV system. It can prevent damage and eliminate potential fire risks [57]. The main task of detection for photovoltaic panels is to certify with the greatest certainty that faults have occurred and that the system is no longer operating in its healthy operating range. The fault diagnosis focuses on fault identification, fault severity estimation, and fault isolation. The whole process is based on a priori knowledge, estimation and field measurements [16].

The PV FDD methods can be classified into visual inspection and automatic analysis, which will be discussed in this section.

II.2. Visual inspection for fault diagnosis

Visual inspection is a quick and effective method to identify faults of PV modules. It could be performed before and after the module has experienced environmental, electrical or stress test in the laboratory or operation in the field. The common detectable PV faults by visual inspection are listed in Table II-1.

Table II-1 List of detectable PV module faults by visual inspection

PV module component	PV module faults
Front/Back of PV module	Bubbles, delamination, yellowing, browning
PV cells	Cracked cell, discolored anti reflection
Cell metallization	Burned, oxydized
Frame	Bent, broken, misaligned
Junction box	Broken, loose, oxydation, corrosion
Wires, connectors	Detachment, broken, exposed electrical parts
Bypass diode	Burned, broken connection

Besides, PV images captured by various PV imaging techniques (electroluminescence, infrared, etc.) can also be examined by humans. These types of images could reveal some mechanical or electrical failures inside PV modules. A detailed presentation of these imaging techniques will be given in Section II.3.1.2.1. Despite its effectiveness, visual inspection is time-consuming and requires expensive equipment and personnel. According to [57], visual inspection is more appropriate for small-scale facilities where they can be frequent and cost-effective.

Some PV module faults that lead to performance degradation (e.g., induced partial discharge, broken internal interconnection) are usually invisible to the naked eye and cameras or sensors adopted for visual inspection. Detection of these types of faults requires other information and the implementation of more advanced methods.

II.3. Automatic information analysis for fault diagnosis

Automatic information analysis is playing an increasingly important role in PV FDD techniques. On the one hand, this is due to the increasing availability of PV data with the rapid development of the PV industry and, on the other hand, to the growing demand for automated and accurate monitoring of the health status of PV systems.

Automatic information analysis for PV FDD relies on a great variety of methods and techniques. The methods can be decomposed into four steps: modelling, pre-processing, feature extraction and feature analysis [58], as shown in Figure II-1.

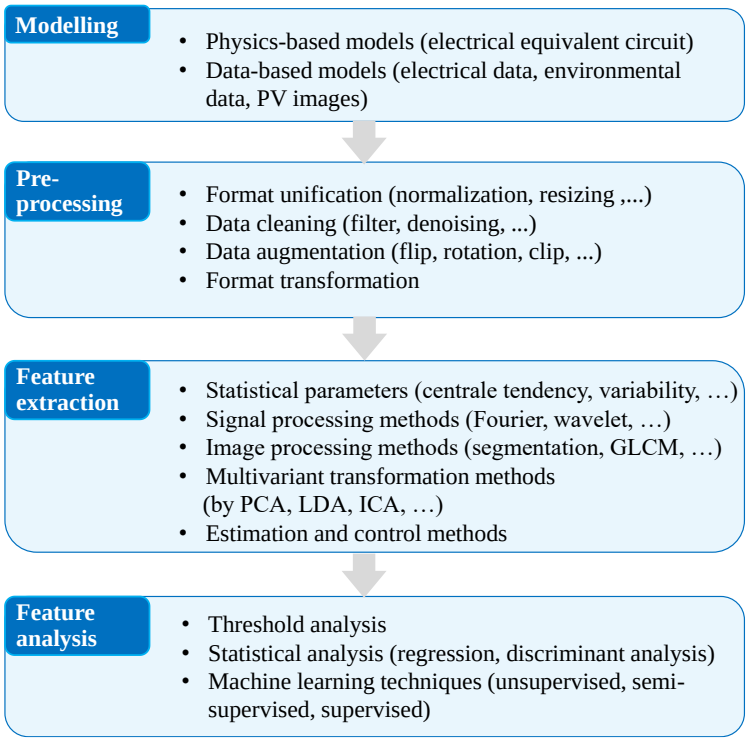


Figure II-1 Four-step PV FDD scheme

A feature refers to an individual measurable property or characteristic of a phenomenon. Generally, features are numeric, but structural types such as strings and graphs are also possible. In this four-step based scheme, under some circumstances, the pre-processing and feature extraction step could be skipped. In addition, it should be noted that the same techniques can be used in different steps. For example, Machine learning techniques (MLT) can be adopted for pre-processing, feature extraction, and feature

analysis. Based on this four-step scheme, the common automatic PV FDD methods will be analyzed in the following.

II.3.1. Modelling

A model is a representation of knowledge. It can be:

- Analytical: based on mathematical relationships between inputs and outputs, which describe the physical phenomena occurring in the system,
- Numerical: based on data measured in the system during operation,
- Descriptive: based on the expression of operating modes.

PV modelling plays an important role in the design, characterization, and maintenance of PV systems [59]. In order to use the PV module model for diagnosis, it must be able to accurately simulate the operating points of a PV system under varying environmental conditions. Therefore, modelling the PV module is an essential step for fault detection and diagnosis [60].

In the current state of knowledge, the model of a PV system is either analytical (also denoted as physics-based) or numerical (denoted as data-driven).

II.3.1.1. Physics-based modelling

The models of a PV module are based on the description of the photoelectricity at a cell level. The cell models are combined in series and parallel to obtain the model of a module. They represent the relation between environmental inputs (irradiance and temperature) and the electrical signals, output current and voltage. The most common models are electrical equivalent circuit (one or two diode model [61], Merten's model [62]), semi-empirical "point" model [63], Evans model or called simple efficiency model [64]). Thanks to their simplicity and efficacity, the one or two diode models are the most used ones to estimate the operational parameters and characterize PV array. The electrical equivalent circuits for the two models are displayed in Figure II-2 and Figure II-3, respectively. It should be noted that the accuracy of the models depends on the accuracy of the model parameters, which can be extracted by various estimation methods [65], [66].

In the single-diode model: I_{ph} is the photocurrent, I_d the diode current representing the diffusion phenomenon, R_{sh} the shunt resistance representing leakage current path caused by the distributed manufacturing defects inside the solar cell, R_s the series resistance representing the power dissipation caused by the thermal effect in the whole junction substrates, and the electrical contacts.

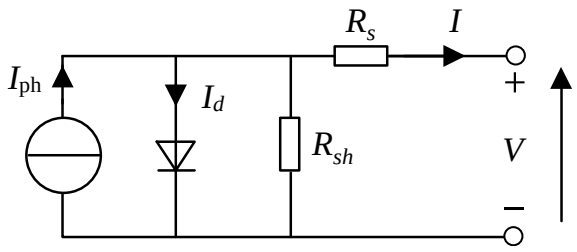


Figure II-2 Equivalent circuit of the single-diode model

The double-diode model has two diodes in parallel. It takes the recombination phenomena into consideration, and therefore could achieve a better characterization [67]. This model provides more a reliable estimation of the internal parameters, which can better reflect the phenomena occurring inside the solar wafer.

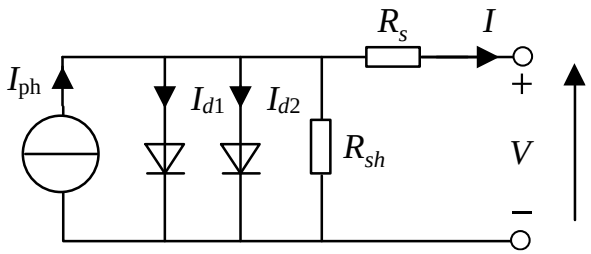


Figure II-3 Equivalent circuit of the two-diode model

II.3.1.2. Data-based modelling

Data-driven modelling is based on historical measurements made during the operation of the PV module (electrical and environmental measurements) or images captured by ground cameras or unmanned aerial vehicles (UAV)

II.3.1.2.1. PV images

The images of PV modules can be classified into four groups, which are detailed in the following, and compared in Table II-2.

- **Visible image**

Conventional imaging devices provide visible (Vis.) images similar to those obtained by human visual inspection (presented in Section I.2). In contrast, the images adopted for automatic fault diagnosis are usually obtained by UAVs, which are more suitable for large PV installations [68].These systems allow a large number of images to be available for analysis.

- **Infrared thermography image**

Infrared (IR) thermography imaging is a non-destructive measurement technique, which can provide rapid, real-time, and two-dimensional distributions of the characteristic features of PV modules [68]. The IR images permit the identification of

various PV faults causing temperature anomalies, e.g., PID, partial shading, invisible cell, crack, hot spots, bypass diode faults, OC or SC of cell/diode/module. IR imaging cameras can also be embedded on UAV devices.

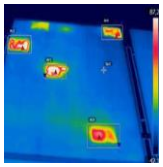
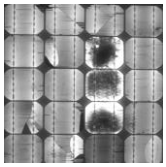
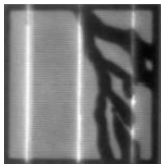
- **Electroluminescence image**

Electroluminescence (EL) images are captured by silicon charged coupled device (CCD) camera, which detects light emission caused by the radiative recombination of carriers when external DC current is fed into a PV module [69]. Therefore, EL imaging requires a dark environment. The light intensity increases with the local voltage, so that areas with poor contact appear dark [70].

- **Ultraviolet fluorescence image**

Ultraviolet fluorescence (UVF) [71] imaging is realized by exposing a PV cell or module to ultraviolet (UV) light to excite fluorophores in the material into higher energy states. The reemitted light has a longer wavelength than the absorbed radiation [72]. This light is visible to naked eyes and can be captured with a camera or analyzed with a UV/Visible spectrometer using a limited bandwidth source and long-pass filters.

Table II-2 The four types of PV images

	Visible	IR	EL	UVF
Examples	 [68]	 [68]	 [14]	 [27]
Environmental requirement	None	None	Darkness	Darkness
Disconnection of modules	No	No	Yes	No
External source	None	None	DC current	UV light
Typical detectable faults	Discoloration, burn mark, PS, delamination	Cell crack, Hot spot, PS, SC, OC, PID, diode fault	Cell crack, PID, diode fault, disconnection	Cell crack, disconnection

The PV image data from a variety of imaging techniques permits to detect the faults that have no significant impact on the electrical parameters of PV modules but still need to be noticed, like cell crack, discoloration, and delamination. However, as for the limits, the imaging devices are generally costly, and the image acquisition process could be complicated and time-consuming if external sources or specific light conditions are

required, especially for large-scale PV plants. The UAV-based imaging may provide solutions for the imaging of large-scale PV plants. However, an automatic identification and segmentation of PV modules from aerial images requires further study [73].

II.3.1.2.2. Environmental measurements

There are three environmental measurements commonly adopted for PV diagnosis, as specified in the following.

- **In plane irradiance**

Irradiance is the instantaneous measurement of the radiant power from the sun captured by a solar cell or module per unit area [74]. For performance analysis or health monitoring of PV modules, the global irradiance received in the plane (G_{poa}) of a PV module is commonly adopted [74]. The measurement of G_{poa} is generally done by a pyranometer or a reference cell. There are two types of pyranometer: thermopile and photodiode pyranometer. A thermopile pyranometer measures the irradiance in the range of 300 to 2800 nm with a flat spectral sensitivity, while the photodiode measures a portion of the solar spectrum between 400 nm and 1100 nm [75]. The reference cell provides measurement as the effective irradiance. It must be made of the same material as the PV module under study [76].

- **Module temperature**

Module temperature (T_m) is the second contributing factor to the output power of a PV module (primary factor is the irradiance), especially when the temperature is above the value of 25°C defined in the Standard Test Condition (STC). T_m is generally measured by two means:

1) Contact methods: sensor attached to the module back sheet.

2) Contactless methods: the temperature is estimated from ambient temperature (T_a) and irradiance [63], or extracted from the relationship between open circuit voltage (V_{OC}) and T_m [77]. Infrared cameras, if calibrated, can also measure the operating temperature [78].

- **Meteorological data**

Except G_{poa} and T_m, meteorological data are sometimes needed to understand the module performance better and evaluate its stability. The common data includes ambient temperature T_a, wind speed and direction, humidity, air pressure, Global Horizontal Irradiance (GHI) [79], Diffuse Horizontal Irradiance (DHI) [80], spectral irradiance [81].

II.3.1.2.3. Electrical measurements

Electrical measurements are also among the common features used for FDD. The typical ones can be classified depending on their locations, as presented in Table II-3.

Table II-3 Common electrical measurements in PV system

Location	Measurements
DC side	• Voltage and current adjusted by maximum power point trackers (MPPT) (V_{MPP}, I_{MPP}) • Open circuit voltage (V_{OC}) • Short-circuit current (I_{SC}) • Current-voltage characteristic (I-V curve)
DC/AC conditioning unit	• Voltage and current of the inverter • Voltage and current of the battery bank
AC side	• Voltage and current (V_{AC}, I_{AC})

For monitoring the health status of PV panels, electrical measurements on the DC side are more commonly adopted. Among these measurements, the I-V curve contains the most information: V_{MPP}, I_{MPP}, V_{OC} and I_{SC}, the fill factor (FF), and the slopes of the curve. I-V curves are usually obtained with an I-V plotter or by hardware solutions integrated in PV inverters, the latter allowing a periodic characterization of a PV module, string or plant in operation [18].

II.3.2. Pre-processing

The pre-processing of raw data is an essential step for automatic diagnosis of PV modules. Pre-processing techniques are applied to prepare the data for efficient features extraction. They can be categorized into format unification, data cleaning, data augmentation and format transformation.

II.3.2.1. Format unification

Measurements in the PV system are obtained by different sensors and acquisition systems. They may therefore have very different formats (variation intervals, length, sampling periods, etc.). In order to avoid biased analyses, it is recommended first to standardize the formats. For time or frequency-domain electrical signals, resampling or window cutting are the common operations [82]. Resampling is also required for I-V curves with a different number of points or current-voltage distribution [19]. In addition, when several non-commensurable PV system variables are used for FDD analysis, a normalization [83] is usually performed to standardize the range of variation

of the characteristics into [0, 1] or [-1, 1]. For images, the most common format unification operations are resizing [84], RGB separation [85], graying [86].

II.3.2.2. Data cleaning

Raw PV data may contain interference or invalid information that is removed by filters [87]. Advanced signal processing methods can also be used, such as wavelet denoising [88] or smoothing [89], to remove unwanted noise and fluctuations. For images, denoising is applied to remove external interference and restore the real image [90].

II.3.2.3. Data augmentation

Statistical and more generally machine learning techniques are more effective when they have a large amount of data that is sufficiently representative of all operating modes. However, this is a serious challenge for measurement systems, even more so for image measurements.

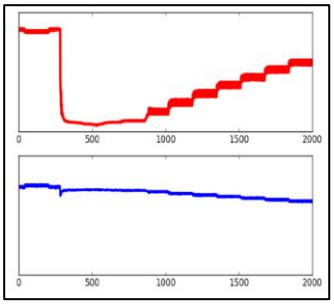
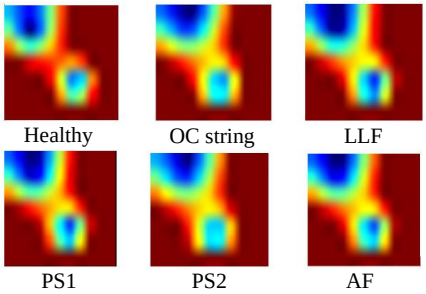
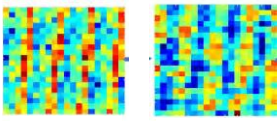
There are two main obstacles to construct a large PV image dataset: 1) insufficient quantity of images due to the limited number of PV modules or the complexity of imaging procedure; 2) unbalanced dataset, i.e., the number of images of healthy modules and faulty modules is different [91]. These two obstacles can significantly hinder the learning performance of machine learning models. Therefore, data augmentation [92] is usually adopted to increase the number of images or balance the number of images among the different classes (operating conditions). The most common techniques include rotation, flip, clip, blurring addition, and adjustment of the brightness [93].

II.3.2.4. Format transformation

The format transformation of PV data is sometimes performed before analysis. These transformations are mainly done for two reasons: 1) to find a more appropriate representation for FDD analysis, 2) to combine different PV data format into an identical one.

There are various transformations that are realized either by a simple rearrangement of data or signals into images or matrixes, or by special techniques. Some examples are presented in Table II-4.

Table II-4 Some format transformations for PV FDD

Ref.	Original features	Transformed features	Description	Example of transformed features
[82]	*V* and *I* signals	Image	Save sequential waveform of *V* and *I* as image	
[94]	G, T_m, V_{OC}, I_{SC}, V_{MPP}, I_{MPP}, P_m	Scalogram	Continuous Wavelet Transform (CWT) [95]	 Healthy, OC string, LLF, PS1, PS2, AF
[24]	I-V curve, G and T_m	Matrix	Construction of a feature matrix	T G I_1 V_1 T G I_2 V_2 T G I_3 V_3 … … … … T G I_{40} V_{40} 40×4 feature matrix
[96]	*I* signal	Matrix	Rearrange signal into a square matrix	 Healthy, Arc fault
[25]	I-V curve, G and T_m	Corrected I-V curve	Correction of I-V curve to a certain environmental condition	-

II.3.3. Feature extraction

After pre-processing the raw data, the feature extraction step aims to extract from the data the most significant features representing the defects. The extracted features or fault signatures should be highly informative and not redundant. For cost-effective and

reduced computation time, the extracted features should also be of reduced dimension. However, the extraction of the features should not degrade the overall FDD performance. Common feature extraction techniques for PV FDD can be classified into five categories as shown in Figure II-4: *statistical parameters*, *signal transformation*, *image processing*, *multivariate transformation*, and *estimation & control techniques*.

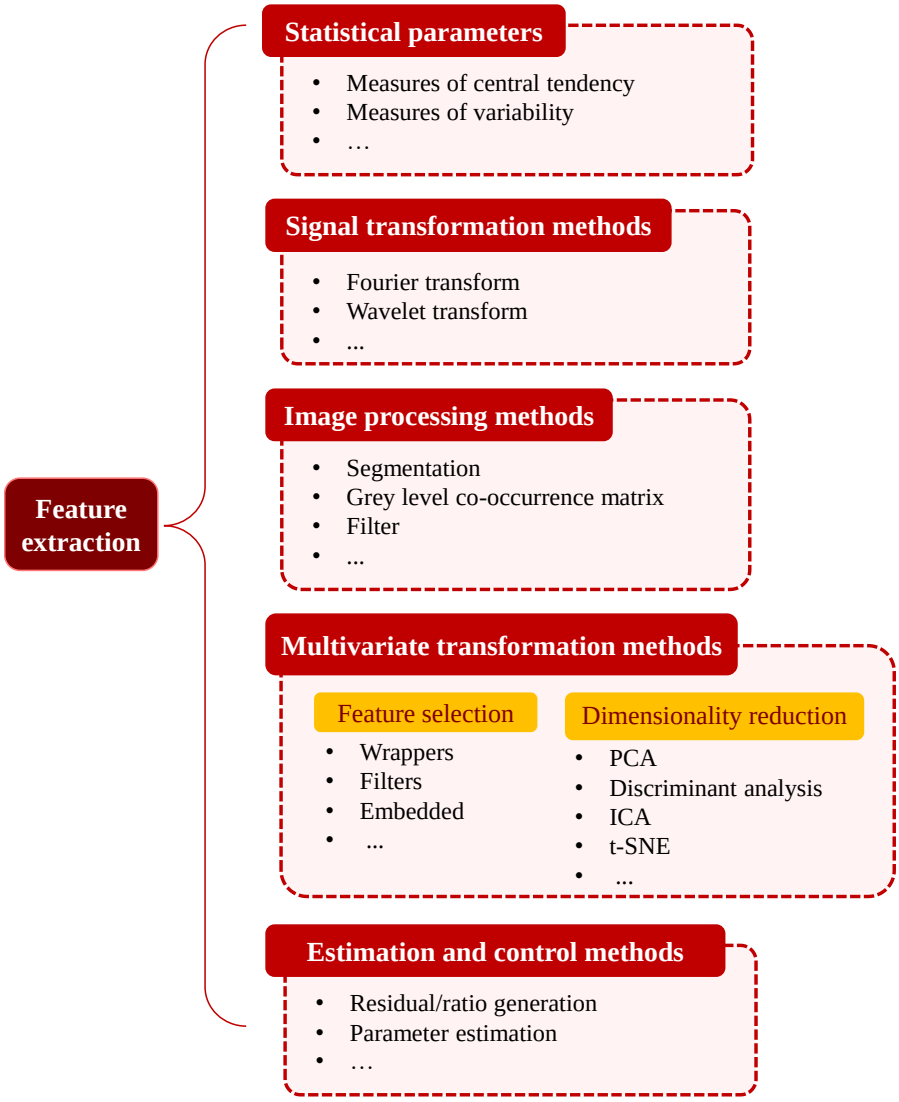


Figure II-4 Common techniques of feature extraction for FDD

II.3.3.1. Statistical parameters

Measures of central tendency, and measures of variability are used to extract relevant basic statistical properties from PV data [97].

II.3.3.1.1. Measures of central tendency

As the most common statistical descriptive measures, central tendency estimates the central location of one univariate PV data by calculating the mean, median or mode [98]. Each of these measures has its pros and limitations. The mean reflects each value

of the dataset but is particularly prone to outlier values. Comparatively, the median is a more suitable measure when the dataset includes outliers. The mode could represent the value that appears most often. It should be noted that there is no restriction on using only one of these measures of central tendency.

II.3.3.1.2. Measures of variability

Measures of variability or dispersion of univariate PV dataset can provide relevant information on the data distribution [99]. The three popular measures of variability are *range*, *variance*, and *standard deviation*. The selected measures depend on the data type, the shape of the distribution and the corresponding measure of central tendency used [100]. When combined with the measures of central tendency, the data exploration is significantly enhanced. For the application in PV FDD, authors in [101] calculates the mean and the standard deviation of *V*, *I* and P_m to detect PS, SC module and OC string fault.

II.3.3.2. Signal transformation methods

Signal transformation methods are used to extract local features from pre-processed raw measurements in the time domain, such as coving peak, crest factor, signal-to-noise ratio (SNR), Root Mean Square (RMS) level [102].

In the case of the I-V curve, which contains information on voltage and current signal, several parameters can be obtained: the open circuit voltage (V_{OC}), short circuit current (I_{SC}), voltage and current at maximum power point (V_{MPP}), (I_{MPP}), fill factor (FF), equivalent series resistance (R_s) and shunt resistance (R_{sh}) [103].

The signal processing techniques can also be used to transform the time-series data into the frequency domain for further analysis.

- **Fourier transform (FT)**

The FT [104] is used to determine the frequency components of a signal. The main variants of FT include Continuous Fourier Transform, Fourier Series, Discrete Fourier Transform, and Fast Fourier Transform [105]. For example, in [106], FFT is adopted to extract the frequency content of the current to detect arc fault (AF) in a PV array.

- **Wavelet transform (WT)**

WT is a transform, which decomposes an input signal into subsets. Each subset is constituted with time series of coefficients characterizing the evolution of the signal in the corresponding frequency band [107]. There are two types of WT: Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). CWT uses an infinite number of scales and locations, while DWT uses a finite set of wavelets [108].

WT can extract local spectral and temporal information simultaneously. The performance of the transformation depends on the choice of the mother wavelet function and the translation and expansion coefficients to tune time and frequency resolutions. For the application in PV FDD, authors in [109] have adopted DWT to extract features from the AC current (I_{AC}) to identify AF. Similarly, authors in [110] have used DWT with AC voltage (V_{AC}) and I_{AC} to classify LLF and GF.

II.3.3.3. Image processing methods

In the case of images of PV modules or arrays, the data can have a complex structure and a high dimension. Therefore, to extract the most relevant features, the images should be processed. Some examples are given below.

- **Segmentation**

Images in the PV domain are mainly recorded at array or power plant levels, less often at smaller scales for obvious technological reasons. However, for health monitoring, the analysis should be conducted at the cell or module level. The segmentation of large-scale images at the target scale can be realized manually or automatically. For example, authors in [111] perform an automatic segmentation to cell level via edge identification of module-level images obtained from EL. In [112], authors have used deep learning to segment aerial images to PV panel images for diagnosis.

- **Grey level co-occurrence matrix (GLCM)**

GLCM is used to analyze the spatial distribution of pixel intensity to calculate several image texture features [113]. In the case of PV, it is often exploited to extract 1D features such as contrast, homogeneity, or entropy parameters for fault diagnosis. For example, GLCM has been used to extract textural features from visual inspection images of PV modules to assess soiling [114]. GLCM has also been used to extract features like contrast and correlation from IR images of a PV module to detect hot spot [115].

- **Filter operation**

To extract 2D areas of interest from one PV image, filtering is a common operation. Several filters can be used, depending on the use case. For example, authors in [116], inspired by Sobel filter [117], have designed a specific filter to extract Hough and percentile regions from cell images as shown in Figure II-5.

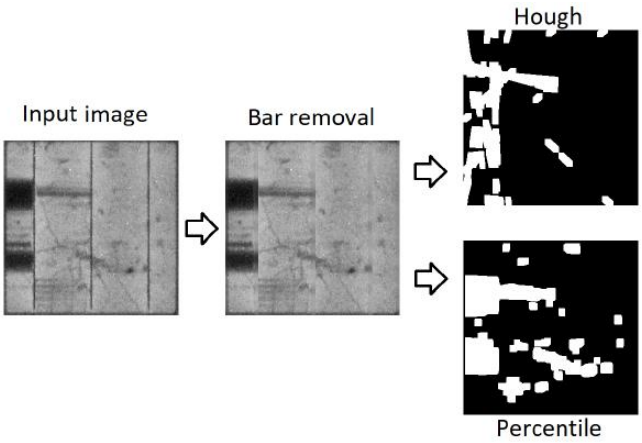


Figure II-5 Illustration of feature extraction from PV EL images [55]

Deep learning techniques can also be used for filtering. For example, authors in [118] have used denoising convolutional neural network (DnCNN) [119] to extract soiling layer from PV module images, as shown in Figure II-6.

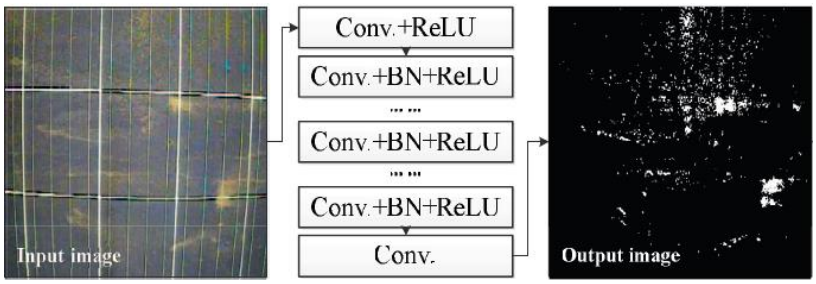


Figure II-6 Extraction of soiling layer of PV module images using denoising CNN [118]

II.3.3.4. Multivariate transformation techniques

When input data are diverse or low-separable in current space, it may be necessary to perform transformations or projections, especially for multivariate features. These allow the dimensionality to be reduced and the selection of features to be refined before the analysis stage. Therefore, the techniques can be broadly categorized into feature selection and dimensionality reduction type.

II.3.3.4.1. Feature selection techniques

A feature selection algorithm can be considered as the combination of a search technique to construct new feature subsets, combined with an evaluation measure which scores the different feature subsets. There are three main types of feature selection algorithms: wrappers, filters, and embedded methods. Wrapper method applies a predictive model to score feature subsets [120]. A model is trained for each new subset and tested on a hold-out set. Counting the error rate of the model on the hold-out set, the best subset is determined. Filter methods adopts a proxy measure to score a feature subset instead of the error rate [121]. The choice of the measure needs to be fast to

compute, while still conserve the usefulness of the feature set. Embedded methods combine the qualities of filter and wrapper methods on performing feature selection during the model construction [122]. An example for the application in the domain of PV FDD is given in [123], where the embedded method is used to select 5 features from originally 11 PV features to detect PS and SC module.

II.3.3.4.2. Dimensionality reduction techniques

- **Principal component analysis (PCA)**

PCA is the process of computing the principal components to perform a change of basis of the original data [124]. It is an efficient tool to reduce the dimension of the original features by projecting original data onto the first several principal components, which allows to obtain lower-dimensional features but can still preserve as much of the variation of original data as possible. Features of lower dimension permit to efficiently control the complexity of diagnosis models, especially for data-driven methods. Concerning PV FDD, [20] reduces three PV electrical features into two by selecting the first two components to detect PS. [125] adopts PCA to reduce the 280 wavelet coefficients into 20 for the detection of AF.

For the components extracted by PCA, two indices, the Hoteling's T-squared (T^2) statistic and the Q-statistic (also known as the squared prediction error (SPE)), can be further extracted [126]. For example, in [127], both T^2 and Q-statistic are extracted from the PCA model that is applied for a total of 9 original PV features. These two statistic features are then used to classify PS, connection fault and sensor fault.

- **Discriminant analysis**

Discriminant analysis is a category of techniques to analyze the data when the criterion or the dependent variable is categorical and the predictor or the independent variable is interval in nature [128]. Discriminant analysis encompasses methods that can be used for dimensionality reduction. Linear discriminant analysis (LDA) is particularly popular because it is both a classifier and a dimensionality reduction technique [129]. Quadratic discriminant analysis (QDA) is a variant of LDA that allows for non-linear separation of data [130]. Regularized discriminant analysis (RDA) is a compromise between LDA and QDA, it is particularly useful when there are many features that are potentially correlated [131]. An example of application in PV feature extraction is given in [132], where LDA is employed to extract two projected features from V_{MPP}, I_{MPP} and P_m for the diagnosis of PS.

- **Independent Component Analysis (ICA)**

ICA is a computational method for separating a multivariate data into additive subcomponents [133]. This is realized by assuming that all the subcomponents are non-

Gaussian features and are statistically independent from each other. ICA can be used for dimensionality reduction by extracting the directions that are as statistically independent as possible and projecting the process data onto the associated basis vectors. For example, in [134], ICA is applied to extract 2-3 features from 20 electrical signals in a PV array for the FDD of different PV faults, like OC string, PS, SC module, etc.

- **t-distributed Stochastic Neighbor Embedding (t-SNE)**

t-SNE is a non-linear dimensionality reduction technique for visualizing high-dimensional data by assigning each data a location in a two or three-dimensional space [135]. To be specific, t-SNE models each high-dimensional data by a two- or three-dimensional point. In this way, similar data are gathered by nearby points and dissimilar data are mapped by distant points with high probability. For PV FDD, [136] applied t-SNE to extract two features from originally 25 PV features to detect inverter fault. In addition, t-SNE is also commonly adopted for the visualization of high-dimensional features in PV FDD. For example, in [137], t-SNE is used to project the 4 PV features G, T_m, V_{MPP} and I_{MPP} into a two-dimensional space to analyze the discriminability of features for the classification of 8 PV faults, like SC, PS, degradation, etc.

II.3.3.5. Estimation and control techniques

The model of PV array, once established, can be used to estimate the operating parameters under different environmental conditions. The estimation process may require the input of real-time measurement: irradiance, temperature, meteorological data. The estimated parameters are generally the electrical parameters. For FDD analysis, these estimated parameters can either be used directly or used to generate residuals. The performance of the extracted features for analysis depends on the accuracy of the PV model. For example, authors in [21] generate a residual from measured and estimated P_{MPP} (via single diode model) to classify PS, OC string, and diode fault. Also, based on the single diode model, authors in [138] estimates intrinsic model parameters from dynamic I-V curves for the diagnosis of cell crack.

II.3.4. Feature analysis for FDD

After modelling, pre-processing and feature extraction, their analysis is the last step in the FDD methodology. As in the previous steps, several techniques can be used depending on the application domain, the nature of the data, the information domain and the desired performance. These techniques can be mono or multidimensional and rely on threshold, statistical analysis, or machine learning techniques (MLT), which are summarized in Figure II-7.

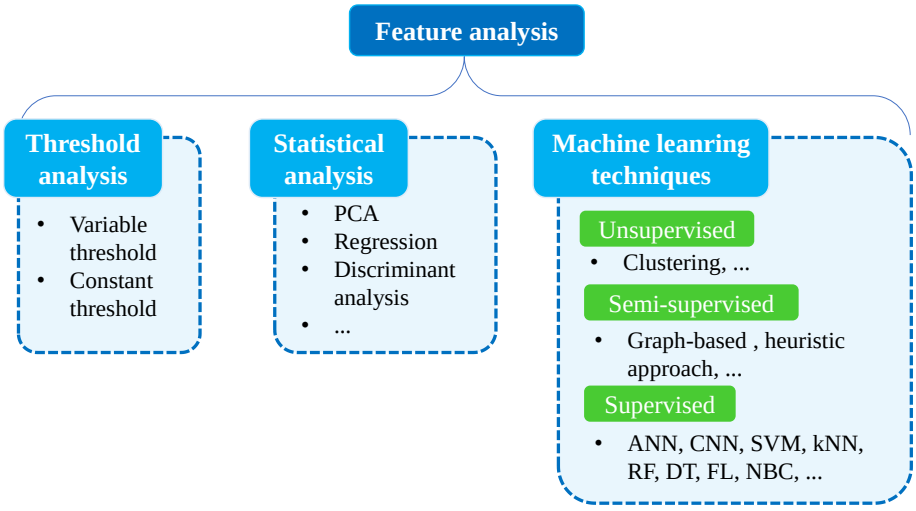


Figure II-7 Techniques for feature analysis in FDD

The common techniques or algorithms of these three categories will be detailed in the following. Application cases in the field of PV FDD will also be given after the presentation of each technique.

II.3.4.1. Threshold analysis

Threshold logic is the most basic tool for making a binary decision [139]. The crossing of a threshold, by the feature, gives an indication of whether the system is healthy or faulty. The threshold can be variable or constant:

- A variable threshold generally requires a model of the PV array, which parameters are updated in real-time. For the application in the PV field, the measured P_{MPP} is compared with its real-time estimated from the single-diode model to detect the presence of PS [140]. In [141], the performance ratio (PR) is compared with a simulated value obtained from the double-diode model to identify the GF.

- A constant threshold can be set based on: simulation of PV model, field experiments, or empirical knowledge. For example, in [142], from the past measurements, a threshold is set for the PR to detect power losses. In [143], a limit is determined for the array's I_{MPP} by empirical knowledge to identify hot spot.

II.3.4.2. Statistical analysis

As already mentioned for feature extraction, the statistical properties of the features can also be used for decision making. There are several tools available in the literature for statistical analysis [144], detailed in the following.

- **Regression analysis**

Regression analysis is a set of statistical processes for estimating the relationships between a feature or state and one or more independent features [145]. The objective of the regression is to find the coefficients of the curve (linear or non-linear) that most closely fits the data according to an optimization criterion. Regression analysis can be employed for prediction or classification. For the application in PV field, a regression model has been used to detect the power loss of a PV array by analyzing G_{poa}, T_m, V_{MPP} and I_{MPP} [146]. In [147], the Gaussian regression is applied to analyze V_{MPP}, I_{MPP} and P_{MPP} for the classification of PS and inverter fault.

- **PCA and discriminant analysis**

PCA and discriminant analysis methods (LDA, QDA and RDA), already presented in Section II.3.3.4.2 for dimensionality reduction, can also be used for classification [148]. To perform the classification, each new data sample is projected into the representation space previously built with one of these techniques, and then assigned to its class. Authors in [20] adopts PCA to detect array PS. QDA is used in [123] to analyze G_{poa}, T_m, T_a, V_{MPP}, I_{MPP} and P_m for the diagnosis of PS and SC module. Besides, in [127], a discriminant analysis model is developed to classify PS, connection fault and sensor fault in a GCPVs by analyzing the T^2 and Q-statistic features.

II.3.4.3. Machine learning techniques

Thanks to the development of PV power plants and their increased digitalization, operators have collected a huge amount of data over the past years. At the same time, data analysis tools, especially MLTs, have also grown tremendously. MLTs, are powerful tools that can deal with large amount of data [149]. MLTs can automatically analyze the relationships between variables and build an accurate model. Once the model has been established and evaluated during the training and validation processes, it can be used with new input samples for classification or prediction. This allows MLTs to be applied for the tedious task of FDD in PV field [147]. As a hot research topic for a long time, several MLTs have been developed, and applied to fault diagnosis in PV field. These techniques can be categorized into *unsupervised*, *semi-supervised* and *supervised learning techniques*.

II.3.4.3.1. Unsupervised learning technique

Unsupervised learning algorithms are designed to find the underlying structures in unlabeled input data [150]. The aim is then to assign any new sample to one of its structures. A typical unsupervised learning algorithm is clustering.

Clustering consists of forming groups (denoted clusters) in which objects with common properties are grouped together [151]. There are several clustering algorithms that differ according to how clusters are constructed and differentiated: k-means [152], DBSCAN

[153], OPTICS [154], hierarchical clustering [155]. For the application in PV FDD, k-means-based clustering has been applied in [156] by analyzing G_{poa}, T_M, V_{MPP}, I_{MPP}, V_{AC} and I_{AC} to detect PS and OC string, and in [157], [158] to evaluate the power loss. The density-peak type uses V_{MPP}, I_{MPP}, V_{OC}, I_{SC} in [159] to identify OC string, and LLF of a GCPVs.

II.3.4.3.2. Semi-supervised learning technique

Semi-supervised learning combines a small amount of labeled data with a large amount of unlabeled data during the training phase [160]. It is a special case of weak supervision. Common semi-supervised learning techniques encompass graph-based methods [161], heuristic approaches [162], generative models [163]. In [164], Graph-Based Semi-Supervised Learning (GBSSL) analyzes V_{MPP}, I_{MPP}, V_{OC}, I_{SC} at array level to classify PS, LLF, OC in a GCPVs. Similar application with GBSSL is conducted in [165] to detect LLF, SC module, OC string in a PV array.

II.3.4.3.3. Supervised learning techniques

Unlike unsupervised learning, supervised learning infers a function from labeled training data, that maps an input to an output via a model [166]. The common supervised learning techniques are detailed in the following.

- **Artificial Neural Network (ANN)**

ANN, inspired by biological neural networks, works as a hierarchical model [167]. It generally includes one input layer, several hidden layers, and one output layer, as depicted in Figure II-8. Each layer is composed of several connected units (named neurons), each one associated to an activation function [168]. It operates as parallelized processors to deal with complex systems. There are several variants of ANN, like Multilayer Perceptron neural network (MLP) [169], Radial basis function neural network (RBF) [170], Probabilistic neural network (PNN) [171], Extension neural network (ENN) [172], Extreme learning machine (ELM) [173].

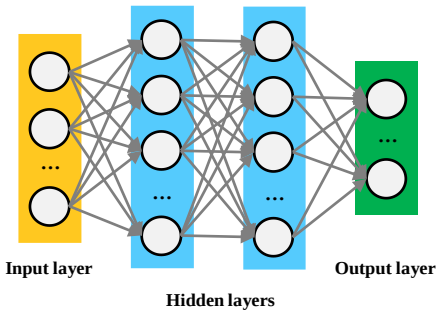


Figure II-8 Basic structure of ANN

Regarding the application to feature analysis for PV FDD, G_{poa}, T_m, V_{MPP} and I_{MPP} are used as input features to a MLP model to detect SC module, and OC string faults [174].

Using the same features, a PNN model is used to identify LLF and OC string faults [175]. A WNN model is used in [176] to analyze V_{MPP}, I_{MPP}, I_{SC} and V_{OC} for the classification of PS, SC module, OC string and degradation. In [177], features from PV EL cell images are analyzed by a MLP model to identify cell crack.

- **Convolutional Neural Network (CNN)**

Convolutional Neural Network (CNN) [178] belongs to the category of ANN but generally exhibits much more complicated network structure. It is the most representative deep learning technique, which is skilled at learning and classifying 2D features, i.e., PV images or transformed 2D features. A typical CNN model consists of one input layer, several groups of convolutional and pooling layers, several fully connected layers and one output layer as depicted in Figure II-9 [179] . Since the structure of the CNN can be pretty complex, various types of CNNs with a different number of layers and connections have been designed over time. The most classical CNN structures are: LeNet [180], GoogLeNet [181], VGG [182], R-CNN [183], ResNet [184], AlexNet [185], MobileNet [186], Attention U-Net [187], and YOLO [188].

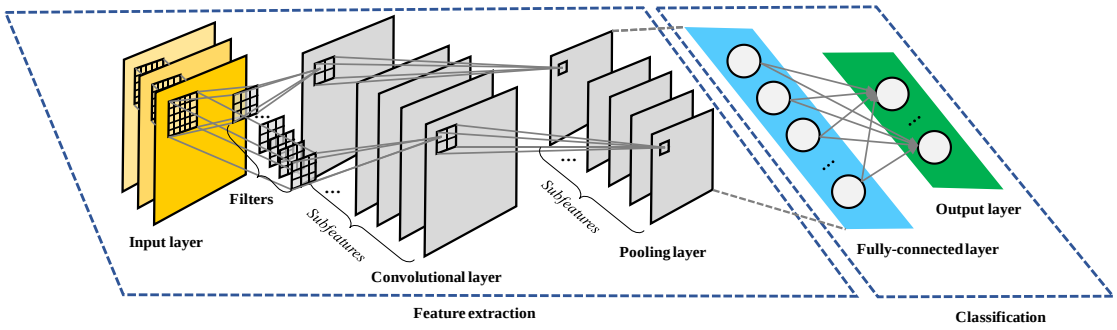


Figure II-9 Basic architecture of CNN

For PV FDD, LeNet model is used to analyze PV EL cell images to identify cell crack [189]. PV IR module images are used as input of MobileNet to detect hot spot [190]. A CNN with 9 layers is used to classify delamination, soiling, snail track and discoloration from UAV-measured PV visible module images [191]. Time series graphs of I and V at array level are taken as features for a ResNet to detect LLF**,** SC module**,** OC string faults [82].

- **Fuzzy logic (FL)**

FL is a form of many-valued logic in which the truth value of variables may be any real number between 0 and 1 both inclusive [192]. It is employed to handle the concept of partial truth, where the truth value may range between completely true and false. FL assigns the numerical input of a system to fuzzy sets with some degree of membership. The common membership functions include the singleton, Gaussian, trapezoidal, and

triangular types. In PV FDD, a FL with triangle as membership function is developed in [193] to study V_{MPP}, P_m to detect array PS. [194] proposes a FL with trapezoid function to classify PS, PID and degradation by analyzing features extracted from I-V curve and irradiance. In [195], a FL with combination of triangle and trapezoid functions is built to detect LLF with features extracted from array V_{MPP} and I_{MPP}.

- **Support Vector Machine (SVM)**

SVM constructs a hyperplane to achieve the largest separation or margin between classes [196]. For non-linear classification, SVM maps the inputs into high-dimensional feature spaces via a kernel function. Common kernels include polynomial (degree of 1, 2 or 3), Radial basis function (RBF). The use of kernel avoids the expensive calculations of the *dot* product in feature space. To achieve multiclass SVM, one-versus-all (OVA) or one-versus-one (OVO) strategies can be adopted. Concerning PV FDD, [197] developed a RBF-based OVO SVM to classify OC module, SC module, PS with array V_{MPP}, I_{MPP} , I_{SC} and V_{OC} as features. In [110], features extracted from PV module images are used to evaluate the module soiling via a polynomial SVM.

- **k-Nearest Neighbors (kNN)**

kNN is a non-parametric lazy classification method, all the training data is needed during the testing phase [198], [199]. An object is classified in the category to which its *k* nearest neighbors in the space of the characteristics identified during the learning process belong. The common distance metrics are Euclidean, City block, or Chebyshev. Regarding PV FDD, a kNN model with Euclidean distance to analyze G_{poa}, V_{MPP} and I_{MPP} for the diagnosis of PS, BPD fault, and inverter fault [200]. The same model has been used to classify PS, OC string and SC module with the features G_{poa} , T_m, V_{MPP}, I_{MPP} and P_{MPP} [201].

- **Decision Tree (DT)**

DT is a decision-making tool that applies a tree-like model [202]. It usually works from top to bottom, by choosing a variable at each step that best splits the set of items [203]. It generally consists of 3 types of nodes: root node, children node and leaf node. The key issue of designing a DT is to determine the best splits. To this end, different algorithms with different metrics can be used, which generally include ID3 [204], C4.5 [205], CART [206]. For PV FDD, [165] built a C4.5 DT with G, T_a, V_{MPP}, I_{MPP} as features to classify LLF, SC module, OC string faults. In [164], G, V_{MPP}, P_{MPP}, V_{OC} and FF are used to construct a DT to detect PS, LLF and OC string faults in a PV array.

- **Random Forest (RF)**

RF is an ensemble learning method for classification that constructs multiple decision trees during training phase and outputs the class by majority voting [207]. Each sub tree

is trained on different parts of the same training set, with the goal of reducing the variance [208]. RF mitigates the overfitting of DT during training and thus generally outperforms DT [209]. Concerning PV FDD, in [210], RF is applied to analyze features extracted from I-V curves to classify PS, OC, SC and degradation fault in a PV array . A similar approach is used in [211] for the classification of PS, LLF, OC string, and degradation fault. A RF model is used with V_{MPP}, I_{MPP}, and T_m as features to detect and analyze the soiling rate [212].

- **Naive Bayes classifier (NBC)**

NBC, based on Bayes' theorem [213], assumes that the variables are conditionally independent. NBC assigns observations to the most probable class based on the maximum a posteriori decision rule [214]. Explicitly, NBC estimates the densities of the variables within each class. Then, it models posterior probabilities according to the Bayes rule. Finally, it classifies a new sample by estimating the posterior probability for each class, and then assigns the sample to the class yielding the maximum posterior probability. In [215], a NBC is developed to analyze the texture features extracted from thermal PV module images to detect hot spot. NBC with kNN and SVM are used to identify the array LLF with 15 features reconstructed from I-V curves [216].

The main properties of these MLTs reported in literature are summarized in Table II-5. It should be noted that Table II-5 is not a technically rigorous comparison of these techniques, which, in fact, is quite tricky to perform due to the variability of each MLT and the application cases.

Table II-5 Main properties of common MLTs applied in PV FDD

MLTs	Pros	Cons
ANN [217], [218]	• Fast decision making • Good approximation of nonlinear relationships	• Appears as Black Box • Slow training • Overfitting risk
CNN [217], [218]	• Fast decision making • Efficient in learning 2D features • Transfer learning permits easier tuning of model	• Appears as Black Box • Large amount of labelled 2D features required • High computational complexity
FL [192], [219]	• Flexible structure and modifiable rules	• Low reliability • Completely depend on human knowledge and expertise

SVM [220], [221]	• Fast decision making • Outliers have less impact • Relatively memory efficient	• Slow training, especially for large dataset • Poor performance with overlapped classes • Overfitting risk
kNN [198], [221]	• Easy implementation	• Slow decision making • Must store all training data
DT [222], [223]	• Easy to understand and interpret • No need for normalization	• Prone to be non-robust • Overfitting risk
RF [207], [208]	• Reduce overfitting, higher accuracy than a single DT	• Lack of the intrinsic interpretability of DT
NBC [214], [215]	• Mitigate the curse of dimensionality	• Often fails to produce a good estimate of the correct probabilities

II.3.5. Illustration of the four-step automatic PV FDD scheme

The four-step PV FDD methodology (modelling, pre-processing, feature extraction and feature analysis) has been detailed. This scheme can be applied to analyze most of the application cases in the PV field. To better illustrate how this four-step automatic FDD scheme is performed, based on the abundant literature, some typical use cases have been selected and analyzed in the light of this approach, as described in Table II-6.

Table II-6 Illustration of the four-step automatic PV FDD scheme with several use cases

Ref.	Modelling		Pre-processing	Feature extraction	Feature analysis	Target fault	Accuracy
	Physics-based model	Measurands					
[21]	Yes	P_{MPP}	-	Residual generation	Threshold	PS, OC, BPD fault	90.3%
[224]	No	I-V curve	-	MSD	SVM	LLF	94.7%
[225]	Yes	V_{MPP}, I_{MPP}	Normalization	-	PNN	LLF, SC, OC	97%
[82]	No	*V*, *I* signal	Transformation	-	CNN	LLF, SC, OC	99.5%
[226]	No	EL images	Data augmentation	Segmentation	CNN	Cell crack	88.4%
[118]	No	Vis. images	Data augmentation	DnCNN	ResNet	Soiling	90%

[227]	No	IR images	Resizing	PCA	GoogleNet	Hot spot, cell crack	97.9%

From Table II-6, it is observed that not all the four steps in the scheme are clearly addressed. For example, authors in [21], [228] have no pre-processing step and the methods in [82], [225] have no feature extraction step. However, with this four-step scheme, every use case can be interpreted step by step, which significantly facilitates the understanding and comparison of PV FDD cases.

II.4. FDD proposal

In this part, based on the summary of the fault diagnosis methods in the literature, the proposal of our FDD strategy will be presented.

II.4.1. Summary of fault diagnosis methods

The two main categories of PV FDD methods, i.e., visual inspection and automatic methods, have been presented. From the aspect of application, the typical properties of these techniques for PV FDD are summarized in Table II-7 from the aspects of pros and limits. For the automatic FDD methods, the three main types of feature analysis are also addressed. It should be noted that Table II-7 is just a presentation of the properties reported in the literature or related to the application in PV FDD. It is not dedicated to conducting an elaborated comparison of these techniques, which is difficult to perform.

Table II-7 Main properties of FDD methods in PV field

FDD methods		Pros	Cons
Visual inspection [27], [57]		• Easy implementation • Practical	• Unsuitable for large-scale PV plants • Only detect visible PV faults • Completely depend on human knowledge and experts
Automatic analysis	Threshold analysis [139], [229]	• Easy implementation • Rapid decision making • Commercially effective	• Accuracy depends on the quality of threshold limits • Prone to noise interference
	Statistical analysis [16], [144]	• Rapid decision making • Clear pattern and correlations	• Assumptions may not be exact • Prone to misinterpret causation

Machine learning analysis [217], [218]	• Easily identifies trends and patterns • Variety of mature models available for application in different conditions • Good at handling multi-dimensional and multi-variate data • Continuous improvement of performance with increasing amount of data	• Large amount of data required • High computational complexity

From Table II-7, it is observed that automatic analysis is more suitable than visual inspection for the FDD of large-scale PV array or PV power plants. Among the three categories of feature analysis techniques, machine learning is more promising. The application of MLTs for FDD in the PV field will be less constrained by the rapid development of powerful processors and adequate instrumentation.

II.4.2. Description of the proposed FDD strategy

Among visual inspection and automatic information analysis, since visual inspection is unsuitable for monitoring large-scale PV array and detecting invisible fault, the automatic information analysis is adopted. This strategy will be discussed from the four-step PV FDD scheme in detail:

- *Modelling*: Due to the rapid development of PV power plants and their increased digitalization, operators have collected a huge amount of data for PV FDD. Thus, our FDD strategy focuses on the data-based modelling. As for the measurands, the I-V curve, as containing rich information for assessing the health of the PV modules and arrays, is chosen for analysis. Besides, G and T_m will also be used since they provide critical environmental information additional to the I-V curve.

- *Feature pre-processing*: To eliminate the impact of environmental condition on the I-V curve, the correction of I-V curve to an identical environmental condition is executed using G and T_m. Besides, as one key point of our FDD strategy is to employ complete I-V curve for PV FDD, a resampling of the corrected I-V curve will also be performed.

- *Feature extraction*: To improve the quality and the discriminability of features, special feature extraction methods, i.e., Recurrence Plot (RP) and Gramian Angular Difference Field (GADF), will be applied. These methods, efficient in the signal transformation, have not been employed for PV FDD in the literature. Then, PCA, a typical dimensionality reduction technique, will be applied to reduce the

dimension of the transformed features. Besides, t-SNE, a useful tool of visualizing high-dimensional data, will also be used to analyze the extracted features.

- *Feature analysis for PV FDD*: The MLTs, powerful in handling big data and multi-condition classification, are chosen as the analysis techniques. Since the I-V curves are obtained under known PV conditions, common supervised learning techniques like ANN, SVM, kNN will be practiced to analyze the features for the classification of PV faults (presented in Section I.6).

II.5. Conclusion

The fault detection and diagnosis methods have been discussed through two categories: visual inspection, and automatic analysis.

Visual inspection permits to conduct simple and fast examination. However, it is obviously not suitable for monitoring large-scale PV array and detecting invisible faults. In addition, it poses problems of reproducibility and reliability because observations may vary from one operator to another.

The availability of large amounts of operational data and the development of computing resources encourages automatic learning methods. The automatic FDD methodology can be described in a four-step scheme: modelling, pre-processing, feature extraction and feature analysis.

- *Modelling*: it includes physics-based and data-based modelling. Common PV data adopted for FDD task consists of 3 types, i.e., PV images, environmental measurements, and electrical measurements. The electrical ones, especially the I-V curve, contains useful information for assessing the health of the PV modules and arrays.

- *Feature pre-processing*: it is usually optionally applied with different purposes: format unification, data cleaning, data augmentation and format transformation.

- *Feature extraction*: typical techniques applied for PV FDD include statistical methods, signal processing methods, image processing methods, transformation methods and estimation methods. Although feature extraction may be optional, it may help to further process the PV features and improve the quality and discriminability of features for automatic FDD analysis.

- *Feature analysis for PV FDD*: these techniques are classified into threshold analysis, statistical analysis, and machine learning analysis. The MLTs, showing a

great feasibility in handling big data and exhibiting a good generalization capability, are very promising.

From the above analysis, we have decided to adopt the I-V curve, G and T_m as input data of automatic four-step FDD methodology for an application in the PV field. Specific pre-processing and extraction methods will be designed and evaluated to enhance the discriminability capability of fault features. Finally, MLTs will be used to analyze the features for classifying PV array faults.

Chapter III

Correction of PV I-V curve measured under faulty condition

III.1. Introduction

The I-V curves measured from a faulty PV module or array (hereinafter, termed as faulty I-V curve) contain valuable information on its health status [20], [230]. However, because field-measured I-V curves are recorded under varying environmental conditions, the comparison between different curves would not be relevant if they are not brought to a same condition. Thus, each I-V curve is corrected for a specific environmental condition (usually STC, 1000W/m^2 and 25°C) in order to make the curves measured under different environmental conditions comparable.

In literature, various parameters, like V_{OC}, I_{SC}, V_{MPP}, I_{MPP} and P_m are commonly extracted from corrected I-V curves for fault diagnosis [19], [231], [232] or power loss analysis [233]. In addition, equivalent R_s or R_{sh} are also commonly calculated from corrected I-V curves to analyze the degradation mode of PV networks [234], [235].

Therefore, if there are significant errors in the corrected curves or in the parameters extracted from the curves, the accuracy of the diagnosis will be seriously compromised. As a consequence, the performance of the I-V curve correction is an important step in the diagnostic process of PV modules.

To the best of our knowledge, there are no specific methods proposed for the correction of faulty I-V curves. Consequently, most of the researchers adopt the original or simplified IEC 60891 [25] correction procedures [236], [237]. For example, several authors have adopted the procedure 1 in IEC 60891 to correct I-V curves measured under partial shading (PS) [238], potential induced degradation (PID) [239], dust soiling [240], [241], or hot spot [242]. Besides, the procedure 2 of IEC 60891 is also used to correct the key parameters (like V_{OC}, I_{SC}, P_m, etc.) on I-V curves [19]. In [19], these parameters can then be used as inputs of a classifier to automatically identify the fault types (PS, OC, SC, or R_s degradation). Similar applications can also be found in [243].

It should be noted that these procedures based on the IEC 60891 standard have been all initially designed for the correction of curves measured from healthy PV panels. Furthermore, in the literature, the suitability of these procedures for the correction of the I-V curves of faulty panels is rarely investigated, and even less so if the fault characteristics are distorted by the correction and thus lead to a diagnostic error. Yet, all these issues are decisive for the diagnosis of PV faults [26].

With this in mind, this chapter aims to evaluate the performance of current procedures for correcting faulty I-V curves. A new procedure will then be proposed.

Section III.2 presents the preparation of I-V curves for correction with the modeling of a PV array detailed. Section III.3 introduces the common correction procedures and the new

proposal. Section III.4 defines the metrics for the evaluation of correction performance. Sections III.5 presents the correction performance using single or multiple-curves-based procedures under constant or varying fault severity. A discussion of the performance of these procedures is done in Section III.6.

III.2. Preparation of I-V curves for correction

Since our target is to evaluate the error caused by the correction procedures, it is essential to avoid the effects of other factors, like the measurement uncertainty on in-plane irradiance *G*, module temperature T_m and I-V curve, which could be up to 5 % according to [74]. Therefore, the curves to be corrected are generated in simulation rather than from field measurement to ensure all the parameters used in the correction formula have no bias.

III.2.1. PV array modeling

To address several fault types, a PV array is modeled under Simulink® to generate the I-V curves. The regulable 'PV array' block [244] provided in the 'Simscape' library is chosen as the basic model. It requires the following parameters:

- *PV scale parameters*: number of cells per module, number of series-connected modules per string, number of parallel strings,
- PV electrical parameters: V_{OC}, I_{SC}, V_{MPP}, I_{MPP} temperature coefficient (TC) of I_{SC} (α), temperature coefficient of V_{OC} (β).

From the parameter setting, it is easy to find out that this block can be easily configured into a PV cell, module, or array. In this study, a PV array is built based on the repetition and combination of the block. To introduce the array's model, the structure and principle will be detailed in the following starting from the cell and module level.

III.2.1.1. Cell-level modeling

To characterize a PV cell, the single-diode and double-diode models are the most common ones adopted in the literature. Among them, the single diode model is the most popular due to its simplicity and the capability to well reproduce the main characteristics of PV cells. The electrical equivalent circuit of this model is illustrated in Figure III-1.

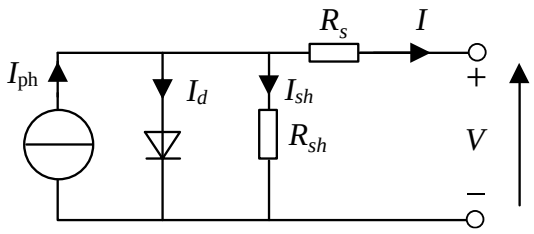


Figure III-1: Electrical equivalent circuit of single diode model

This model consists of a current source that produces the photocurrent I_{ph} proportional to G_{poa} (to facilitate the presentation, G_{poa} is hereafter showed simplified as G if without special notes) and a single diode in anti-parallel with a diode current I_d.

Under solar irradiation, the output current I is expressed by the photocurrent I_{ph}, the diode current I_d and the leakage one I_{sh} following the relationship as:

$$I = I_{ph} - I_d - I_{sh} \quad \text{(III-1)}$$

The photocurrent I_{ph} is determined by the characteristics of the semiconductor. It depends on the G and the cell temperature (T_c) as follows:

$$I_{ph} = \frac{G}{G_{STC}}\left[I_{sc} + \alpha(T_c - T_{c\_STC})\right] \quad \text{(III-2)}$$

where, α represents the absolute temperature coefficient of current (expressed as A/°C), which is usually provided in the datasheet. G_{STC} and T_{c_STC} are the irradiance and cell temperature at STC, equaling 1000W/m^2 and 25°C, respectively.

Based on the Shockley diode equation, the diode current I_d is expressed as:

$$I_d = I_0\left[\exp\left(\frac{V + R_s I}{nV_t}\right) - 1\right] \quad \text{(III-3)}$$

where,

- I_0 refers to the saturation current of the diode
- n: the ideality or the quality factor of the diode, generally between 1 and 2
- V_t is the thermal voltage, which is given by $V_t = k_B T_c/q$
- k_B: Boltzmann constant (k_B = 1.38 × 10^{-23} J/K)
- q: constant absolute value of electron's charge (1.6 × 10^{-19} C),

Regarding the diode saturation current I_0, there are various equivalent calculation methods in the literature [245]. In this model, I_0 is defined as in [246]:

$$I_0 = I_{0\_STC}\left(\frac{T_c}{T_{c\_STC}}\right)^3 \exp\left(\frac{E_{g\_STC}}{kT_{c\_STC}} - \frac{E_g}{kT_c}\right) \quad \text{(III-4)}$$

where, I_{0_STC} is the saturation current at STC, which can be measured by flash test [247] or estimated by parameter extraction method from module datasheet [65]. E_g is the material

bandgap energy, and E_{g_STC} is the STC value. For silicon PV cells, typical E_{g_STC} is 1.12eV. E_g shows a small temperature dependency with cell temperature and its relationship with E_{g_STC} can be expressed as $E_g = E_{g\_STC}\left[1 - C_{E_g}(T_c - T_{c\_STC})\right]$. C_{E_g} is the temperature coefficient for bandgap energy. It is commonly a small value, e.g., 2.67×10^{-4} for silicon cells [246].

The leakage current I_{sh} is determined by the Ohm's Law as:

$$I_{sh} = \frac{V + R_s I}{R_{sh}} \tag{III-5}$$

Substituting I_d and I_{sh} in (III-1), the output current of the single diode model becomes:

$$I = I_{ph} - I_0\left[\exp\left(\frac{V + R_s I}{nV_t}\right) - 1\right] - \frac{V + R_s I}{R_{sh}} \tag{III-6}$$

In Simulink, with the datasheet values given, the modelling of a PV cell is obtained through the resolution of equation (III-6) under different irradiances and temperatures.

III.2.1.2. Module-level modeling

Although the 'PV array' block provided in the Simulink library can be directly configured into a PV module, the structure of this type of PV module is different from common real modules used in the field. The main difference is that this type of PV module contains only series-connected PV cells with no bypass diodes, which are mandatory to ensure the safe operation of real PV modules.

Therefore, in this study, a PV module is built with a combination of PV cells and a bypass diode. Specifically, the module is based on the FranceWatt single-crystalline silicon (sc-Si) PV module -FL60-250MBP, which consists of 60 PV cells in series and 3 bypass diodes (each diode in parallel to 20 cells). In this study, as the faults to study do not include those that cause non-uniform distribution of *G* or T_m on one module, the module is thus built with a PV block containing 60 cells in series and a bypass diode in anti-parallel to simplify the simulation, as shown in Figure III-2. Detailed module parameters are listed in Table III-1.

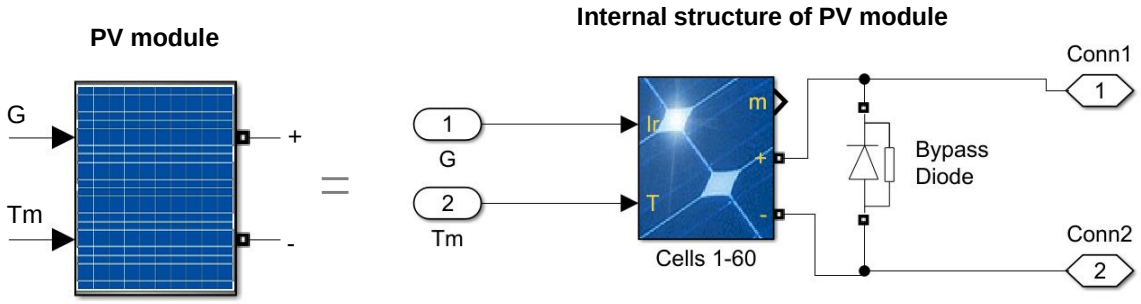


Figure III-2: Structure of PV module model

As seen from Figure III-2, G and T_m are the two inputs of the module. Under real conditions, T_c may vary slightly from cell to cell. And T_m is generally different with T_c due to the isolation between the cell layer and back sheet layer with a difference up to 2-3 °C depending on the structure and the mount of a PV module [63]. In this simulation study, to avoid introducing additional uncertainty factors from the aspect of temperature, all the cells are considered to have the same T_c and this value is equal to the input T_m. In healthy condition, all the cells share the same irradiance and temperature. Under irradiance mismatch conditions, the bypass diode provides a current path to prevent the faulty module from overheat.

Table III-1: Parameter of FL60-250MBP PV module

Variable	Value	Variable	Value
I_{SC}	8.64 A	V_{MPP}	30.51 V
V_{OC}	37.68 V	α_{rel}	0.02 %/°C
I_{MPP}	8.21 A	β_{rel}	-0.36 %/°C

(α_{rel} and β_{rel} are the relative TC of I_{SC} and V_{OC}, respectively)

III.2.1.3. Array-level modeling

In this study, we consider a PV array with two strings in parallel. Each string is composed of three FL60-250MBP modules in series, as shown in Figure III-3. To simulate the I-V curve, the array is connected to a controlled voltage source to simulate the I-V tracer. There are also additional resistances or gain blocks to generate specific faulty conditions, which will be detailed in Section III.2.3.

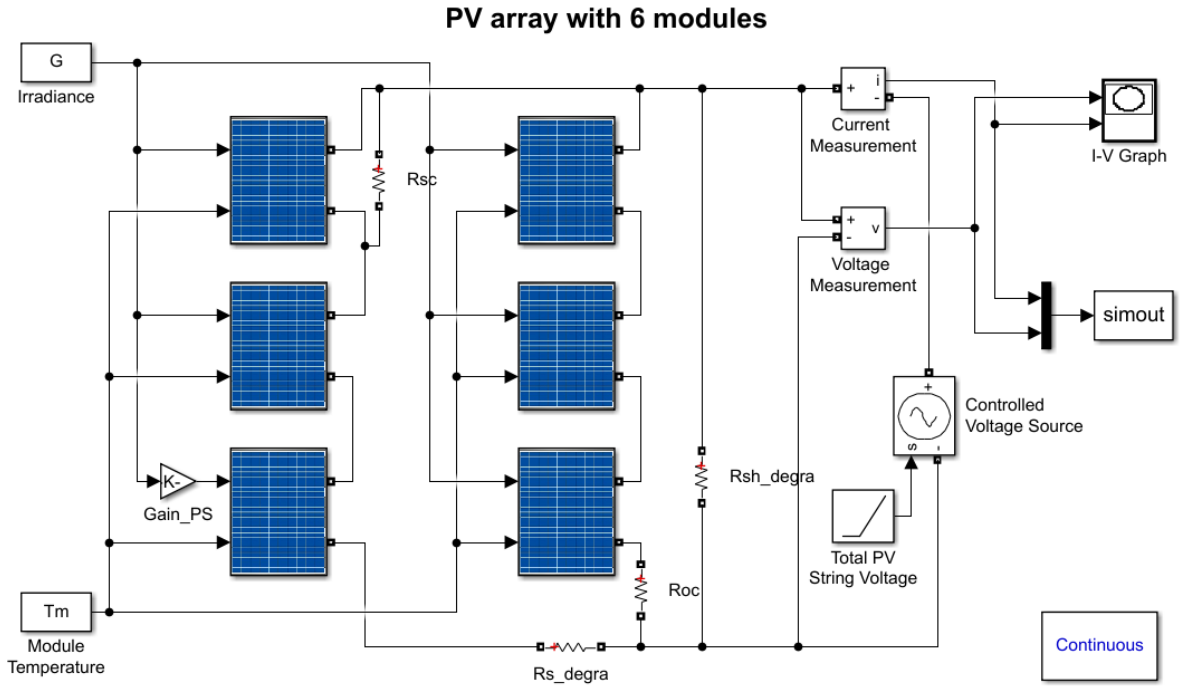


Figure III-3: Structure of PV array model

(The structure of each module is presented in Figure III-2)

The output voltage of the PV array is determined by the voltage of each string, which depends on the number of modules connected in series. In healthy condition, the numbers of functional modules of each string are identical and corresponding string voltages are also the same. However, when the output voltage of one module in one string decreases, e.g., become SC, the voltage of this string also decreases, which then forces the array voltage down to this value as all these strings are connected in parallel.

As for the array current, according to Kirchhoff laws, it is the sum of the current of each string. And within each string, the output current of each module is forced to be identical as they are connected in series. When the current of one module becomes different, e.g., under PS, the bypass diodes will be activated to bypass this faulty module. This could keep the string current at the normal level outputted by healthy modules but decreases the string voltage.

For the array presented in Figure III-3, the array parameters are summarized in Table III-2.

Table III-2: Parameter of the simulated PV array

Variable	Value	Variable	Value
I_{SC}	17.28 A	V_{MPP}	91.53 V
V_{OC}	113.04 V	I_{MPP}	16.42 A

III.2.2. Environmental settings

The form of the array's I-V curve is mainly determined by the G and T_m. The short-circuit current I_{SC} increases with G while the open-voltage V_{OC} decreases with T_m, as depicted in Figure III-4.

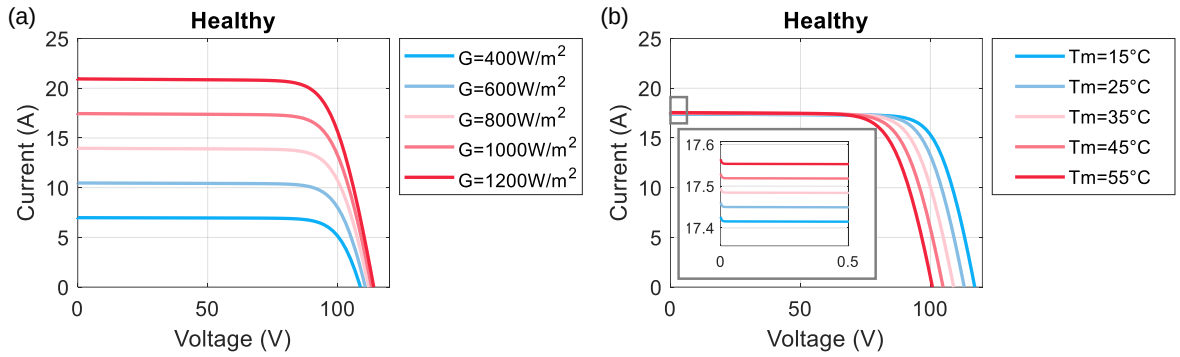


Figure III-4: Impact of G and T_m on I-V curve of healthy PV array

((a): impact of varying G with T_m =25°C, (b): impact of varying T_m with G=1000W/m^2)

The setting of G and T_m to prepare the dataset for correction varies with the correction procedure for evaluation. For example, for correction procedures based on a single I-V curve, the G and T_m will be selected from field-measured pairs of G and T_m. For correction procedures based on multiple I-V curves, i.e., the correction of one I-V curve requires multiples curves, the G and T_m of these curves to correct could be quite different. Detailed explanation will be given in Sections III.5.1.1 and III.5.2.1.

III.2.3. Configuration of faults

With the gain block and the additional resistances, the model shown in Figure III-3 can simulate PV array under healthy and faulty conditions. In this study, specifically, five typical faults are considered:

- Partial Shading (PS): one module is shaded by adjusting the value of the gain block (i.e., *GainPS*) in the range [0, 1] to control the irradiation of the module. The corresponding PS degree varies from 0 to 100%;
- Short-Circuit (SC): one module is short-circuited by connecting the resistance R_{SC} in parallel;
- Open-Circuit (OC): one string is open-circuited by connecting the resistance R_{OC} in series;
- R_s degradation: increases the equivalent series resistance of the array (R_{s_degra});

- R_{sh} degradation: decreases the equivalent shunt resistance of the array (R_{sh_degra});

Table III-3 summarizes the different fault parameters and their corresponding ranges of variation that will be used. The analyses of the impact of fault severity on the correction performance will be presented in Sections III.5.1 and III.5.2.

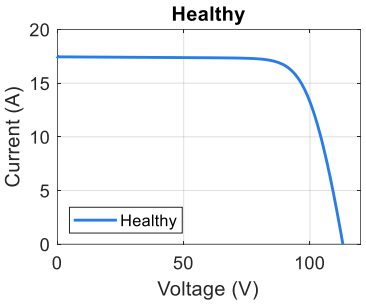
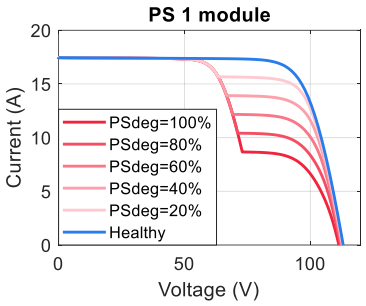
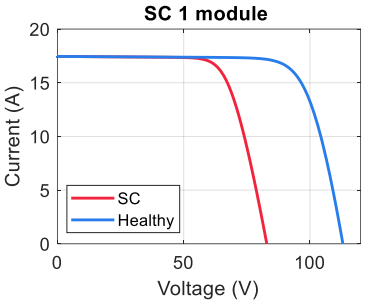
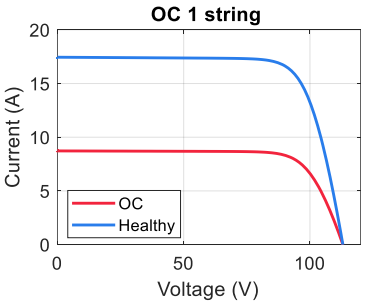
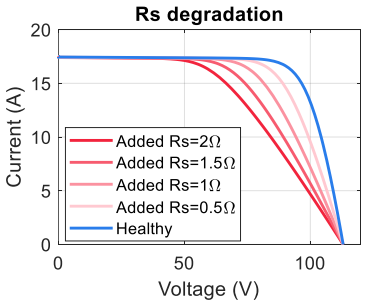
Table III-3: Parameter setting for the different operating conditions

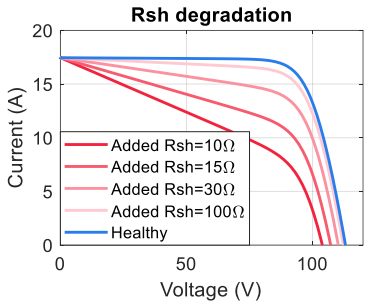
Condition	PS degree (%)	$\boldsymbol{R_{SC}}$ (Ω)	$\boldsymbol{R_{OC}}$ (Ω)	$\boldsymbol{R_{s\_degra}}$ (Ω)	$\boldsymbol{R_{sh\_degra}}$ (Ω)
Healthy	0	10^5	10^{-5}	10^{-5}	10^5
PS 1 module	[0 - 100]	10^5	10^{-5}	10^{-5}	10^5
SC 1 module	0	10^{-5}	10^{-5}	10^{-5}	10^5
OC 1 string	0	10^5	10^5	10^{-5}	10^5
R_s degradation	0	10^5	10^{-5}	[10^{-5} - 2]	10^5
R_{sh} degradation	0	10^5	10^{-5}	10^{-5}	[10^5 - 20]

III.2.4. Impact of faults on I-V curves

In fact, the five faulty conditions could lead to different distortions of the I-V curve. To illustrate their impact, I-V curves under STC are simulated. The I-V curves of healthy and five types of faulty array are compared in Table III-4. For PS, R_s and R_{sh} degradations, the severity of the fault is varied with several representative values to show its impact on the shape of the I-V curve.

Table III-4: Impact of faults on I-V curves

Condition	Impact on I-V curve	Examples
Healthy	None	
PS 1 module	Appearance of an inflection point The voltage change at the inflection point occurs at about 2/3 of V_{OC} under healthy condition since there are three modules in one string while one module is shaded; larger the PS degree, lower is the current of the inflection point.	
SC 1 module	Voltage decrease The V_{OC} on SC curve decreases by 1/3 of V_{OC} in healthy condition since one out of three modules is SC.	
OC 1 string	Current decrease The I_{SC} on OC curve decreases by 1/2 of I_{SC} in healthy condition since one out of two strings is OC.	
R_s degradation	Slope decreases in the voltage region The greater the value of the added resistance is, the smaller the slope in the voltage area is.	

R_{sh} degradation	Slope increases in the current region The smaller the value of the added resistance is, the greater the slope is.	

III.3. I-V curve correction procedures

In this section, the usual I-V curve correction procedures and a new proposal are presented.

III.3.1. Usual correction procedures

Through the literature review, the most common I-V curve correction procedures are based on the three procedures proposed in IEC 60891 [25]. These procedures are either based on single curve for correction, like procedures 1 and 2, or based on multiple curves, like procedure 3, all of which are detailed in the following. Since STC (G =1000 W/m^2 and T_m =25 °C) is more commonly adopted as the target condition for I-V curve-based diagnosis, in the following, if not stated, the correction of I-V curves refers to the correction to STC.

III.3.1.1. Procedure 1 (P1)

The P1 is based on the following equations:

$$I_2 = I_1 + I_{SC1}\left(\frac{G_2}{G_1} - 1\right) + \alpha(T_{m2} - T_{m1}) \quad \text{(III-7)}$$

$$V_2 = V_1 - R_s(I_2 - I_1) - \kappa I_2(T_{m2} - T_{m1}) + \beta(T_{m2} - T_{m1}) \quad \text{(III-8)}$$

where, I_1 and I_2, V_1 and V_2, T_{m1} and T_{m2}, G_1 and G_2 are the currents, voltages, module temperature, and irradiances before and after correction, respectively; I_{SC1} is the short-circuit current before correction; α and β are the PV module absolute TC of I_{SC} and V_{OC}, respectively; $\alpha = \alpha_{rel} \cdot I_{SC}^{STC}$, $\beta = \beta_{rel} \cdot V_{OC}^{STC}$, α_{rel} and β_{rel} are the relative TC of I_{SC} and V_{OC} (presented in Table III-1); R_s is the internal series resistance and κ is the curve correction factor. These coefficients can be determined using the methods in [25].

III.3.1.2. Procedure 2 (P2)

The P2 is based on the following equations:

$$I_2 = \frac{I_1\left(1 + \alpha_{rel}(T_{m2} - T_{m1})\right)G_2}{G_1} \tag{III-9}$$

$$V_2 = V_1 + V_{OC1}\left[\beta_{rel}(T_{m2} - T_{m1}) + a \cdot ln\left(\frac{G_2}{G_1}\right)\right] - R_s(I_2 - I_1) - \kappa \cdot I_2(T_{m2} - T_{m1}) \tag{III-10}$$

where, V_{OC1} is the open-circuit voltage before correction; a is the irradiance correction factor; R_s and κ may not be the same value used in P1, but determined by the procedure detailed in [25].

In fact, the determination of the coefficients a, R_s and κ for P1 and P2 requires a group of I-V curves at the same *G* or at the same T_m according to [25], which is troublesome and time-consuming to fulfill in real life. Therefore, in most applications of P1 and P2, these coefficients are either tuned via simulation under healthy condition [235], or neglected [238] but to the detriment of poorer performance. Therefore, in this study, these coefficients will not be neglected and will be determined following the standard routine, i.e., via the simulation under healthy condition. In this case study, the determined coefficient of P1 and P2 and the TC used are presented in Table III-5.

Table III-5: Correction coefficients of P1 and P2

Correction procedure	Correction coefficients			Temperature coefficients			
	a	R_s (Ω)	κ	α (A/°C)	α_{rel} (%/°C)	β (V/°C)	β_{rel} (%/°C)
P1	-	0.512	0.0026	0.0035	-	-0.41	-
P2	0.0413	0.473	0.0025	-	0.02	-	-0.36

The examples of the correction of an I-V curve (simulated using the array model in Figure III-3) to STC using P1 and P2 are illustrated in Figure III-5.

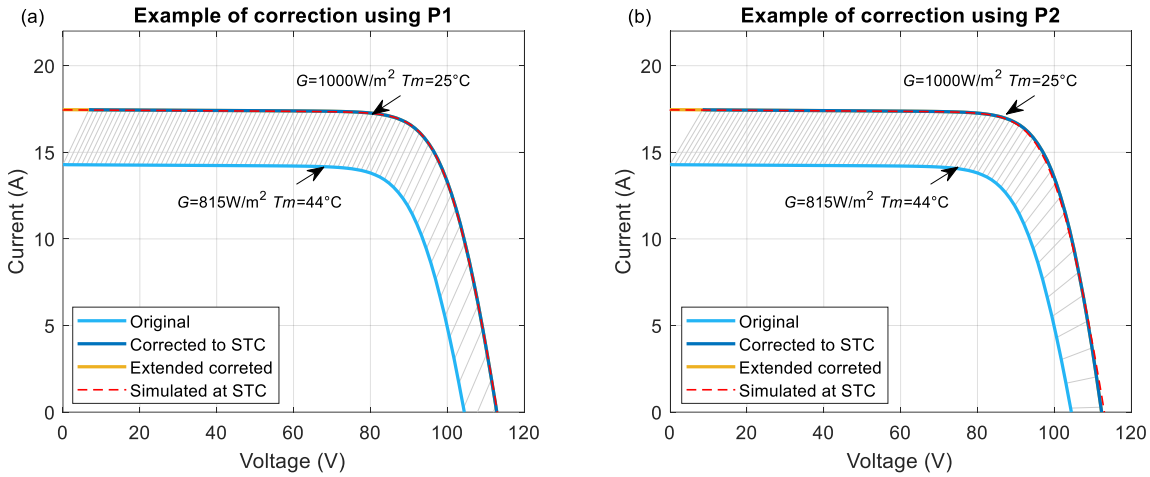


Figure III-5: Example of correction procedure P1 &P2 to STC using an I-V curve of a healthy array

((a): using P1, (b): using P2. Grey lines link the corresponding points on the original and corrected I-V curves, the voltage and current of which are calculated by (III-7), (III-8) for P1 and (III-9), (III-10)for P2)

From Figure III-5, it is observed that, sometimes, for both P1 and P2, the corrected I-V curve (before extended) shifts to the right, as seen from the part near V_{OC}. This is due to the corrected voltage is larger than the original one of each point on the original I-V curve at certain environmental conditions, which can be seen from the gray lines connecting each point before and after correction in Figure III-5. To obtain a complete I-V curve (voltage starting from 0), the corrected curve is extended by extrapolation using Phang's method [248], [249].

In Figure III-5, the I-V curve simulated at STC is also presented. It is observed that the corrected-to-STC curve is a little bit different with that simulated-at-STC (hereinafter called real curve), especially when using P2, which means that even after the correction, the I-V curve still cannot completely correspond to the real curve. To quantify this difference, different metrics will be adopted, which will be then presented in Section III.4.

III.3.1.3. Procedure 3 (P3)

Different from P1 and P2, P3 is free from correction coefficients but requires an interpolation constant γ:

$$I_3 = I_1 + \gamma(I_2 - I_1) \quad \text{(III-11)}$$

$$V_3 = V_1 + \gamma(V_2 - V_1) \quad \text{(III-12)}$$

According to [25], P3 requires at least two reference curves to obtain one corrected curve at specified G or T_m by calculating γ using (III-13) or (III-14). Thus, in order to correct to a condition with requirements on both G and T_m, at least three curves are necessary.

$$G_3 = G_1 + \gamma(G_2 - G_1) \quad \text{(III-13)}$$

$$T_3 = T_1 + \gamma(T_2 - T_1) \quad \text{(III-14)}$$

The three reference curves based P3 is the most common type of application of P3 due to its less requirement of reference curves. It is conducted by two steps according to [25]. An illustration of this correction methodology on $T_m\,(G)$ plot is shown in Figure III-6 and an example given in Figure III-7.

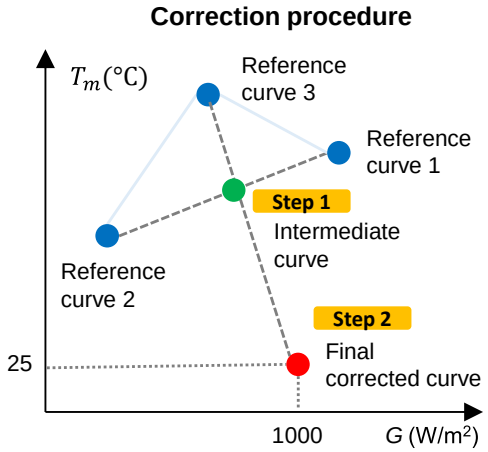


Figure III-6: Illustration of two-step correction procedure with three curves on $T_m\,(G)$ plot

(Step1: use reference curves 1 and 2 to generate the intermediate curve; Step 2: use reference curve 3 and the intermediate curve to get the final corrected curve)

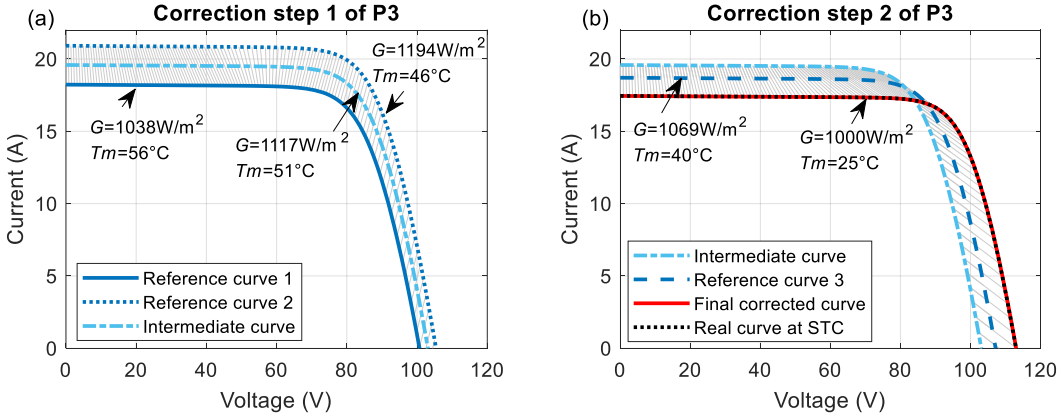


Figure III-7: Example of correction using P3 based on 3 reference curves

((a): step 1, (b): step 2. Grey lines link the corresponding points on the original and corrected I-V curves, the voltage and current of which are calculated by (III-11), (III-12))

In summary, the implementation of P1 and P2 requires only one single I-V curve, while P3 requires multiple curves. For P1 and P2, in fact, they do not perform well for the correction of I-V curves under all the tested faulty conditions regarding the correction of voltage or current. Thus, a new correction procedure is introduced.

III.3.2. New correction procedure

A P2-based procedure denoted NewP2 is proposed. It uses the same formula as P2 for current correction. But for the voltage correction, the term 'V_{OC1}' in (III-10) is replaced by '$V_{OC1}[1+\beta_{rel}(T_{m2}-T_{m1})]$'. The reason for this change is to improve the poor voltage correction of P2 by combining the merit of P1. The effect of this change on the correction performance will be presented in detail in Section III.5.1.2. In this way, the equations for NewP2 are:

$$I_2 = \frac{I_1\left(1+\alpha_{rel}(T_{m2}-T_{m1})\right)G_2}{G_1} \tag{III-15}$$

$$V_2 = V_1 + V_{OC1}[\mathbf{1}+\boldsymbol{\beta_{rel}}(\boldsymbol{T_{m2}}-\boldsymbol{T_{m1}})]\left[\beta_{rel}(T_{m2}-T_{m1}) + a\cdot ln\left(\frac{G_2}{G_1}\right)\right] - R_s(I_2-I_1) - \kappa\cdot I_2(T_{m2}-T_{m1}) \tag{III-16}$$

III.4. Metrics for the evaluation of correction performance

The evaluation of correction performance will be conducted from 2 aspects, i.e., from the entire I-V curve and from single extracted parameters (e.g., maximum power P_m, V_{OC} and I_{SC}). Two corresponding metrics are adopted to quantify the correction performance from these two aspects.

III.4.1.1. Metric for the evaluation of correction of the entire curve

Firstly, for the entire I-V curve, curve error (E_{curve}) is adopted as the metric. It is calculated by the normalized root-mean-square error between the corrected curve and the real curve. It should be noted that the real curve only means that G and T_m are at STC, but the array health status could be either healthy or faulty.

$$E_{curve} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_i^c - I_i^{real})^2}}{I_{SC}^{real}} \times 100 \tag{III-17}$$

where, I_i^c and I_i^{real} are the current values from the corrected and real curve for the same voltage V_i, respectively. V_i is the i^{th} element of a voltage vector with N points linearly distributed in $[0, V_{max}]$ range with a constant step (V_{max} is constant for all the conditions). V_{max} could be set a little larger than the array's V_{OC} at STC in healthy condition to avoid that the voltage of improperly corrected curve exceeds this range. In this study, V_{max} is set as 120V and N at 100. I_{SC}^{real} refers to the I_{SC} extracted from the real curve.

III.4.1.2. Metric for the evaluation of correction of single parameters

Then, regarding the single parameters, the relative error (*RE*) is adopted to evaluate the correction:

$$RE_X = \frac{X^c - X^{real}}{X^{real}} * 100 \quad \text{(III-18)}$$

where, X^c and X^{real} are the parameters from the corrected and real (i.e. simulated at STC) curves, respectively. X can be straightforward curve parameters (such as P_m, V_{OC} or I_{SC}) or the single-diode equivalent model's parameters (R_s or R_{sh}) considered as fault parameters.

III.5. Correction performance

In this section, the three correction procedures proposed in IEC 60891 (P1, P2, and P3) and the new one (NewP2) will be evaluated. However, it should be noted that P3 is based on multiple I-V curves for correction. Its principle is different from the others that are based on a single I-V curve. Therefore, the evaluation of these procedures will be conducted separately: Section III.5.1 compares the correction procedures using a single I-V curve (P1, P2, NewP2) and Section III.5.2 evaluates the correction procedure using multiple I-V curves (P3).

III.5.1. Performance of correction procedures using single I-V curve

In this part, we focus on the correction procedures based on a single I-V curve, i.e., P1, P2, and NewP2. Firstly, the selection of *G* and T_m to generate the curves is presented in Section III.5.1.1. Then, the correction performance (using the metrics E_{curve} and *RE*) will be evaluated using the curves simulated under constant or variable fault severity in Sections III.5.1.2 and III.5.1.3, respectively.

III.5.1.1. Selection of *G* and T_m based on field-measurements

To evaluate the correction procedures with real environmental conditions, the selection of *G* and T_m to generate the curves is based on field-measurements (in SIRTA meteorological and climate observatory [250], France) of the sc-Si FL60-250MBP PV module (identical to those used in the simulations in Section 2.1.1) during summer as illustrated in Figure III-8. The *G* is measured by a reference cell and T_m by a Pt100 probe attached to the back side of the module. To minimize the correction error, commonly, the irradiance of the curves for correction is selected in a higher range [251]. In our case studies, the lower bound of *G* is set as 800 W/m^2.

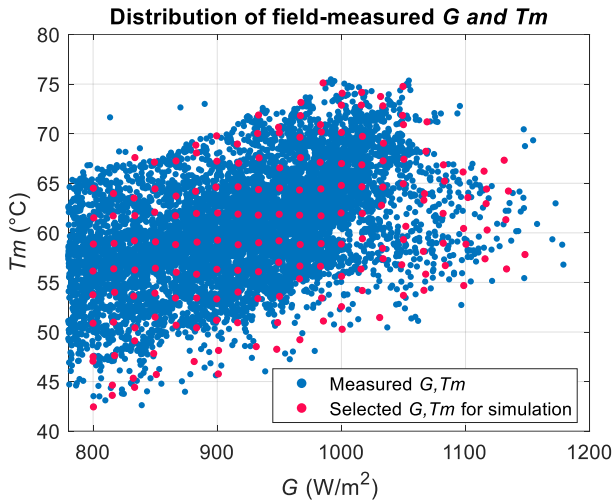


Figure III-8: Selected G and T_m based on field-measurements during summer for one sc-Si PV module

In Figure III-8, the blue points represent the measured G and T_m, while the red ones are those selected for the simulations. Totally, a group of 174 pairs of G and T_m is selected based on a quasi-uniform distribution within the area enclosed by the blue points.

III.5.1.2. Correction performance with constant fault severity

In order to investigate the impact of the selected G and T_m, the fault severity is firstly required to be settled. In this study, the severity for PS, R_s, and R_{sh} degradation needs to be set with values detailed in the following. These severities are the example values that could result in a relatively clear change to the I-V curves, as presented in Figure III-9. Other values of severity will be studied in Section III.5.1.3 using the same analysis approach.

- PS degree = 80% for partial shading,
- R_{s_degra}= 1 Ω for series resistance degradation,
- R_{sh_degra} = 30 Ω for shunt resistance degradation

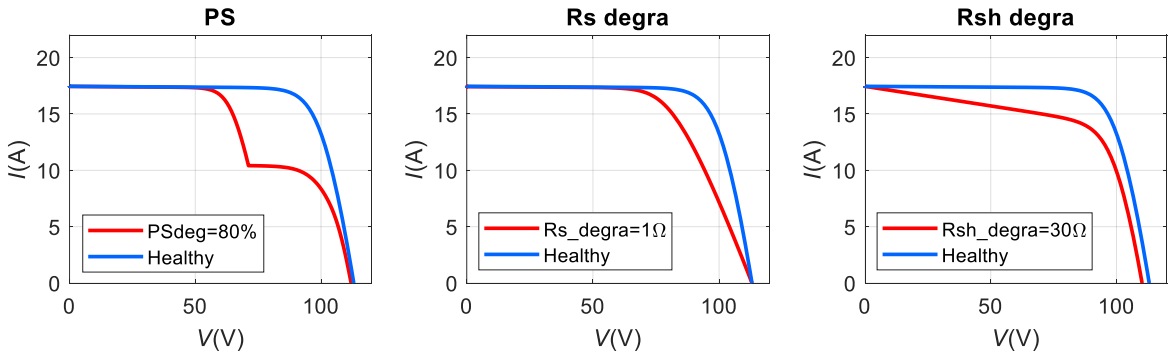


Figure III-9: I-V curves of PS, R_s and R_{sh} degradation under set fault severity

With the environmental setting and fault severity configured, the database of the curves to correct can be simulated. With this database, the correction performance is now evaluated from the shape of the corrected curves and the performance evaluation metrics.

- **Form of corrected I-V curve**

Using P1, P2, and NewP2, the curves before and after correction are displayed in Figure III-10.

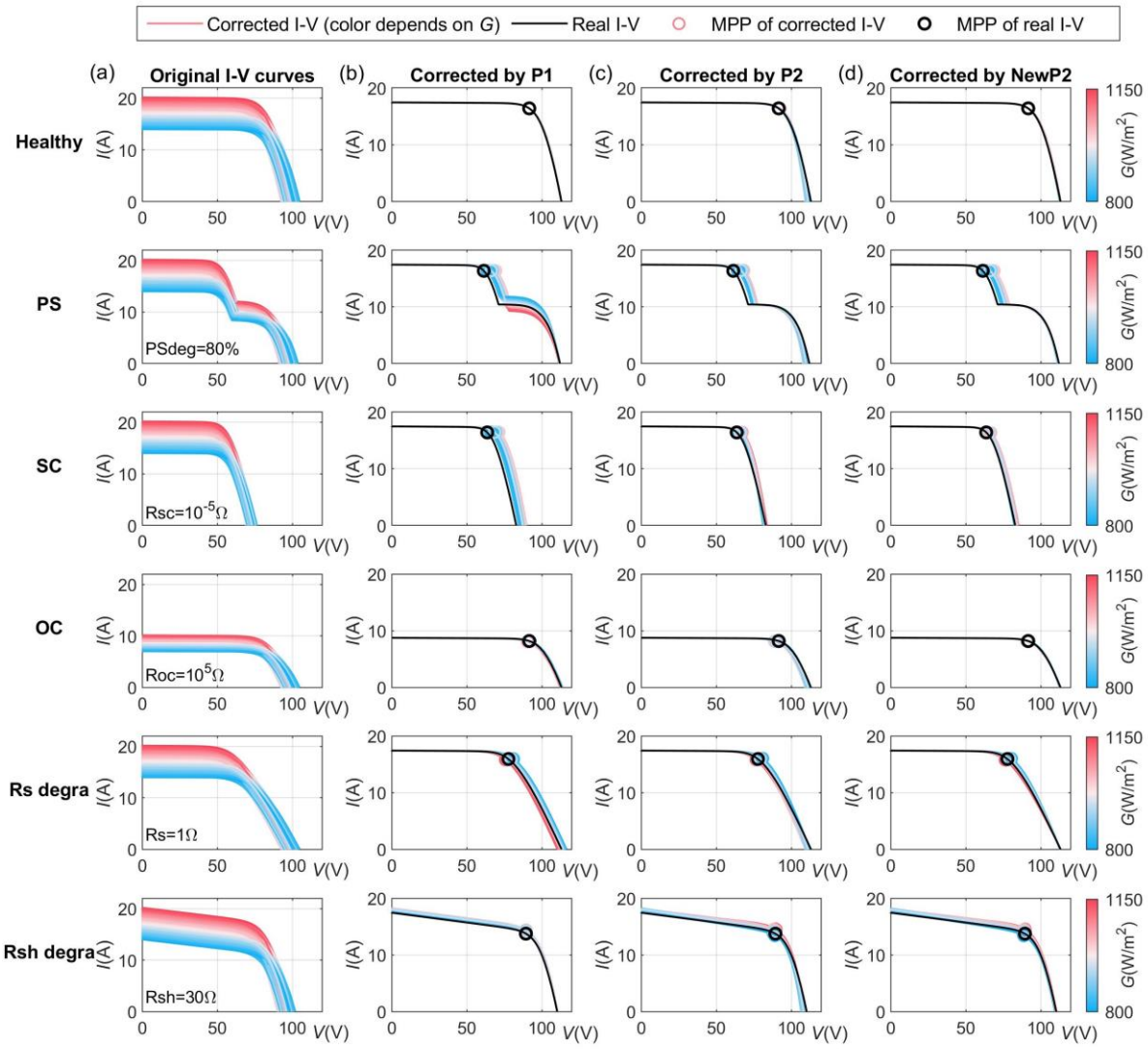


Figure III-10: Correction results using three procedures under eight conditions

((a): curves simulated for correction (each condition contains 174 curves with field-measured combinations of G and T_m at constant fault severity), (b): corrected curves using P1, (c): corrected curves using P2, (d): corrected curves using NewP2 (the displayed color of each curve is determined by the irradiance value with the colorbar on the right side of the figure, the circles on the curves represent the MPP))

For the corrected curves, using all three procedures, clear deviations from the real I-V curve can be observed for most faulty conditions. These deviations are due to the joint effect of

voltage and current corrections, reflected along the *x*-axis, and along the *y*-axis, respectively. They are now analyzed as follows:

- Voltage correction:

Seen from the V_{OC}, the deviations due to voltage corrections are observable in all the cases when using P2, but only for SC and R_s degradation when using P1 and for SC when using the NewP2. To analyze these results, we rearrange the voltage correction formulas (III-8), (III-10) and (III-16) in one single equation as it follows:

$$V_2 = k_1 \cdot V_1 + k_2 \cdot \left(T_{m_2} - T_{m1}\right) + \mathrm{k_3} \cdot \ln\left(\frac{G_2}{G_1}\right) + k_4 \cdot \left(I_2 - I_1\right) + k_5 \cdot I_2 \cdot \left(T_{m_2} - T_{m1}\right) \quad \text{(III-19)}$$

where the different coefficients k_i $(i = 1, \dots, 5)$ for each procedure, are given in Table III-6.

Table III-6: Coefficients for voltage correction in P1, P2, and NewP2

Coefficient	P1	P2	NewP2
k_1	1	1	1
k_2	$\beta_{rel} \cdot V_{OC}^{STC}$	$\beta_{rel} \cdot V_{OC1}$	$\beta_{rel} \cdot V_{OC1} \cdot [\mathbf{1} + \boldsymbol{\beta_{rel}}(\boldsymbol{T_{m2}^{*}} - \boldsymbol{T_{m1}})]$
k_3	0	$a \cdot V_{OC1}$	$a \cdot V_{OC1} \cdot [\mathbf{1} + \boldsymbol{\beta_{rel}}(\boldsymbol{T_{m2}^{*}} - \boldsymbol{T_{m1}})]$
k_4	$-R_s$	$-R_s$	$-R_s$
k_5	$-\kappa$	$-\kappa$	$-\kappa$

* In this study, we correct the curves to STC, thus T_{m2} =25 °C

At first, we compare P1 and P2. For the two procedures, k_4 and k_5 exhibit the same form of variables. Based on the tuning procedure in the standard IEC 60891, these coefficients are also obtained with similar values as presented in Table III-5. As for k_3, it is 0 in P1. In P2, since both coefficients a and '$ln(G_2/G_1)$' are small values, their product is a second-order term and its contribution is negligible (within ±0.3 %). As a consequence, the main difference between P1 and P2 lies in k_2. It is noteworthy that k_2 for P1 is constant, while for P2, it depends on V_{OC1} (V_{OC} of the I-V curve to correct). In Figure III-10 (b), there is almost no distortion of the corrected curves when using P1 under healthy condition, which demonstrates the voltage's correction efficiency. As for P2, since T_m of the curves to correct (T_{m1}) are all higher than the target T_m (T_{m2} = 25°C), V_{OC1} is thus always lower than V_{OC}^{STC} (impact of G on V_{OC1} is negligible compared to T_m). Accordingly, $k_2 \cdot (T_{m2} - T_{m1})$ in P2 is lower than in P1. Therefore, the corrected voltage in P2 is lower than in P1. That is the reason why in Figure III-10 (c), we can observe the slight distortions of the corrected curves shifted on the left side of the real curve near the open-circuit point.

Besides, the significant correction error under SC and PS for P1 and P2 also originates from the setting of k_2. For SC, to be specific, short-circuit is introduced in one module of one string. As a consequence, V_{OC}^{real} is different from V_{OC}^{STC} (value from datasheet, i.e., in healthy case), and is equal to around 2/3 of V_{OC}^{STC}. Therefore, in P1, the corrected voltage is higher than the real value. As for P2, V_{OC1}, although varying with T_m, is still closer to V_{OC}^{real}. Consequently, voltage correction is more efficient with P2. Regarding PS (one module shaded), during the process of I-V scan, when the bypass diodes get activated, the shaded module can be considered as SC. That explains why a similar deviation near the inflection point is also observed.

From the above analysis, P1 and P2 have their pros and cons in voltage correction. Accordingly, NewP2 is designed by combining their advantages. As observed before, in k_2, using V_{OC}^{STC} (like in P1) could generally lead to better voltage correction than using V_{OC1} (like in P2) except under SC. The exception is because V_{OC}^{STC} fails to reflect V_{OC}^{real}. Therefore, in NewP2, V_{OC1} is replaced by $V_{OC1} \cdot [1 + \beta_{rel}(T_{m2} - T_{m1})]$, which corrects the simulated V_{OC1} to the V_{OC} under STC of the real curve and, therefore, could better approximate V_{OC}^{real} under any condition. In this way, as shown in Figure III-10 (d), the correction errors near the open-circuit point using P1 (under SC) and using P2 (under other cases) are attenuated.

- Current correction:

For the current of corrected curves, noticeable dispersion along the *y*-axis is observed under PS near the inflection point using P1. In fact, the output current of PV module is mainly affected by *G*, while the impact of T_m is limited. Therefore, to analyze this phenomenon, for simplification, the contribution of '$(T_{m2} - T_{m1})$' is neglected. Then we can derive from (III-7) and (III-9):

For P1:

$$I_2 \approx I_1 + I_{SC1}(G_2/G_1 - 1) = I_{SC1} \cdot G_2/G_1 + I_1 - I_{SC1} \quad \text{(III-20)}$$

For P2:

$$I_2 \approx I_1 \cdot G_2/G_1 \quad \text{(III-21)}$$

Comparing these two expressions, for P2, the corrected current (I_2) is proportional to the current to correct (I_1). While for P1, there is always a bias that degrades correction's performance, particularly around the inflection point (where I_1 is only half of I_{SC1}), like in Figure III-10 (b) under PS. Therefore, P2 performs relatively better than P1 for current correction. That also explains why, for NewP2, the current formula of P2 is chosen, and the same performance is achieved in Figure III-10 (d). For R_{sh} degradation, it is observed a clear correction error near the I_{SC} part for all the three procedures. This phenomenon will be

investigated from the aspect of current extrapolation in the analysis of $RE_{I_{SC}}$, presented in the next part.

- **E_{curve} and RE of key curve parameters**

Based on the corrected curves, the performance evaluation metrics (E_{curve}, RE_{P_m}, $RE_{V_{OC}}$ and $RE_{I_{SC}}$) are calculated. Considering that G and T_m are both varying for the 174 curves to correct, the statistics of the metrics are presented in Figure III-11.

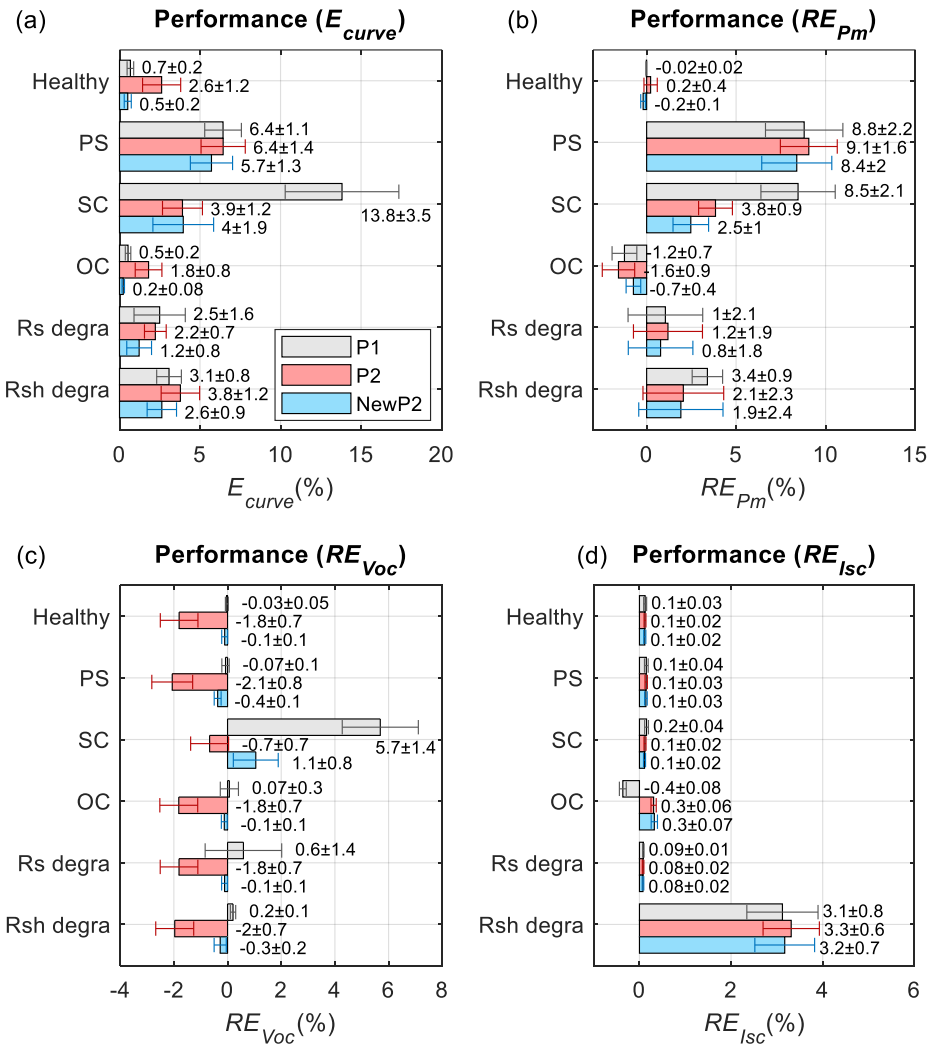


Figure III-11: Four metrics to present the correction performance of P1, P2, and NewP2

((a) E_{curve}, (b) RE of P_m, (c) RE of V_{OC}, (d) RE of I_{SC} (the bars represent the mean value for the correction of 174 curves, while the horizontal whiskers represent the standard deviation (std), these two values are marked as 'mean'±'std'))

These metrics are analyzed as follows:

- E_{curve}: The value of E_{curve}, which reflects the correction error on the entire I-V curve, corresponds to the observations in Figure III-11. As expected, high E_{curve} is found under SC using P1, and PS using all the methods. None of P1 and P2 outperforms in all the conditions, but the proposed NewP2 performs better with generally lower and more stable E_{curve}. The overall average E_{curve} (2.37 %) of NewP2 decreased by 31.3 % compared to the average E_{curve} of P2 (3.45 %) and by 47.3% to that of P1 (4.5%);

- RE_{P_m} : Similarly, NewP2 performs better correction of P_m except under heathy condition with the mean P_m of -0.2% higher than the P1 of -0.02%. However, it should be noticed that RE_{P_m} is positive under nearly all the conditions for all the methods. This indicates that the fault impact on P_m is underestimated. The mean maximum value is up to 9.1 %, which could hinder the detection of incipient PV fault if P_m is used as a fault indicator;

- $RE_{V_{OC}}$: The results are consistent with our previous observations, i.e., the corrected V_{OC} with P2 is always lower than the real value under all conditions, while the V_{OC} using P1 is relatively better corrected except under SC, where the maximum value of the mean $RE_{V_{OC}}$ could reach 5.7 %. And NewP2 effectively reduces $RE_{V_{OC}}$ compared with P1 (in SC), and P2 (in other cases);

- $RE_{I_{SC}}$: The performance of the three methods is similar. It should be noted that, based on (III-8) and (III-10), the voltage of the short-circuit point on the original curve (i.e., equals 0), after correction, is positive. This phenomenon is also observed in [248]. Therefore, to retrieve the I_{SC} (according to the definition, the corresponding voltage should be 0), the corrected curve must be extrapolated. Here, the extrapolation method in [248] is adopted. In this way, although the current value corrected from the original I_{SC} (I_{SC1}) is close to the real I_{SC}, there is still a difference that depends on the initial slope of the corrected curve. Indeed, for all the conditions except R_{sh} degradation, where the slope is mild, $RE_{I_{SC}}$ is low (within ±0.5 %). But for R_{sh} degradation, where the current at MPP is severely lowered down, the slope is steep, and $RE_{I_{SC}}$ is larger.

- ***RE* of fault parameters**

Now, we evaluate the correction performance for one or two typical curve parameters, selected as follows:

- PS: The voltage and current at the inflection point, named V_{rp} and I_{rp} respectively. I_{rp} could reflect the shading level, and V_{rp} the number of activated bypass diodes of PV modules in one string;

- SC: The open-circuit voltage V_{OC};

- OC: The short-circuit current I_{SC};
- R_s degradation: The extracted R_s calculated by the reciprocal of the slope of the I-V curve near the V_{OC} side [252];
- R_{sh} degradation: The extracted R_{sh} calculated by the reciprocal of the slope of the I-V curve near the I_{SC} side [252];

For each fault parameter, the *RE* is calculated from the values obtained from the corrected and real curves. The results are summarized in Figure III-12 with the type of misestimation of parameters also marked.

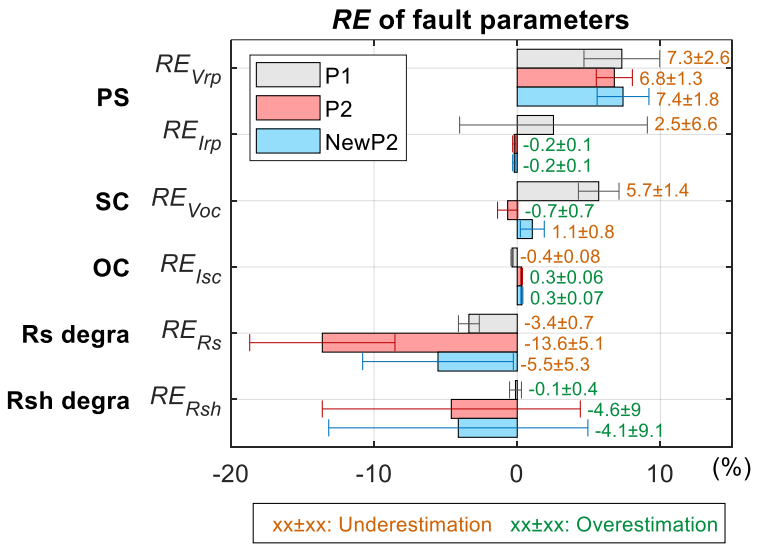


Figure III-12: Correction performance of fault parameters using P1, P2 and NewP2

(Values are presented in mean ± std)

From Figure III-12, We can observe large misestimation (absolute mean *RE* >2 %) with V_{rp}, I_{rp}, V_{OC} and R_s when using P1, and V_{rp}, R_s and R_{sh} when using P2 and NewP2. Worse yet, nearly all these large errors (except R_{sh} using P2 and NewP2) lead to underestimating fault parameters, which will affect the detection of incipient faults. Similarly, none of the three methods outperforms for all the fault parameters. Except for R_s and R_{sh} degradation, in all other fault conditions, P2 and NewP2 have higher or similar performance than P1. Compared to P2, NewP2 achieves almost the same accuracy but greatly decreases the RE_{R_s}. Thus, overall, NewP2 has a more stable performance with an average of the absolute *RE* of 3.1% compared to 4.37% for P2.

III.5.1.3. Correction performance with varying fault severity

In this subsection, the impact of each fault severity on correction performance is investigated. The severities of PS, R_s, and R_{sh} degradations are varied in the following ranges one separately:

PS degree = 0:10:100 %, R_{s_degra} =10^{-5}:0.2:2 Ω or R_{sh_degra}=$10^{1:1:5}$ Ω. For each fault level, the curves are simulated for the selected 174 pairs of G and T_m (presented in Section III.5.1.1) and corrected using the three single curve-based methods.

- **E_{curve} and RE of key curve parameters**

The performance metrics extracted from these corrected curves are now presented in Figure III-13 as a function of fault severity.

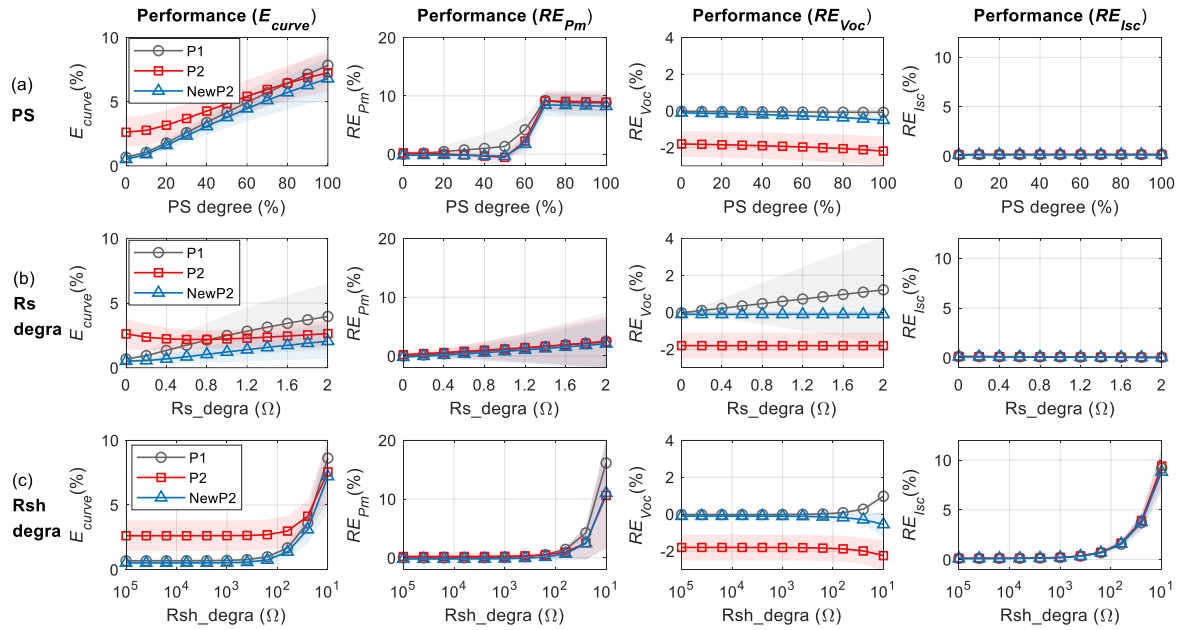


Figure III-13: Correction on the curve error and key curve parameters using the three procedures with varying fault severity

((a) PS, (b) R_s degradation, (c) R_{sh} degradation (the marked line represents the mean of 174 values obtained from the correction of 174 curves under each fault severity, while the band area behind represents the standard deviation; the x-axis for R_{sh} degradation is set as log scale; the degree of severity is presented from healthy to severe on the x-axis from left to right))

From Figure III-13, several remarks can be drawn:

- Nearly all the metrics have a monotonic variation with the fault severities (except the relatively stable variation of E_{curve} under R_s degradation using P2 and the surge on the RE_{P_m} under PS using P2 and new P2, which is analyzed from the remark on RE_{P_m} in the following);

- E_{curve} is mainly sensitive to PS and severe R_{sh} degradation;

- RE_{P_m} is more sensitive to severe R_{sh} degradation and PS. The surge observed for PS is due to the shifting of MPP. Examples are presented in Figure III-14. In Figure III-14 (a), where PS degree=40%, the MPP of both real and corrected curves are at the 'lower

stair' (lower flat part of I-V curve). When PS degree=60%, due to the correction error, part of the MPP on the corrected curves shift to the 'upper stair' (upper flat part of I-V curve). When PS degree=80%, all the MPP have arrived at the 'upper stair'. Thus, during the unsynchronized shifting process between the MPP of corrected and real curves, a surge on the RE_{P_m} under PS is resulted in;

- $RE_{V_{OC}}$ and $RE_{I_{SC}}$ are almost insensitive to all faults regardless of their level of severity, except for $RE_{I_{SC}}$ under severe R_s degradation;
- The overall performance of NewP2 is better than P1 and P2 with relatively lower E_{curve} and *RE* for key curve parameters.

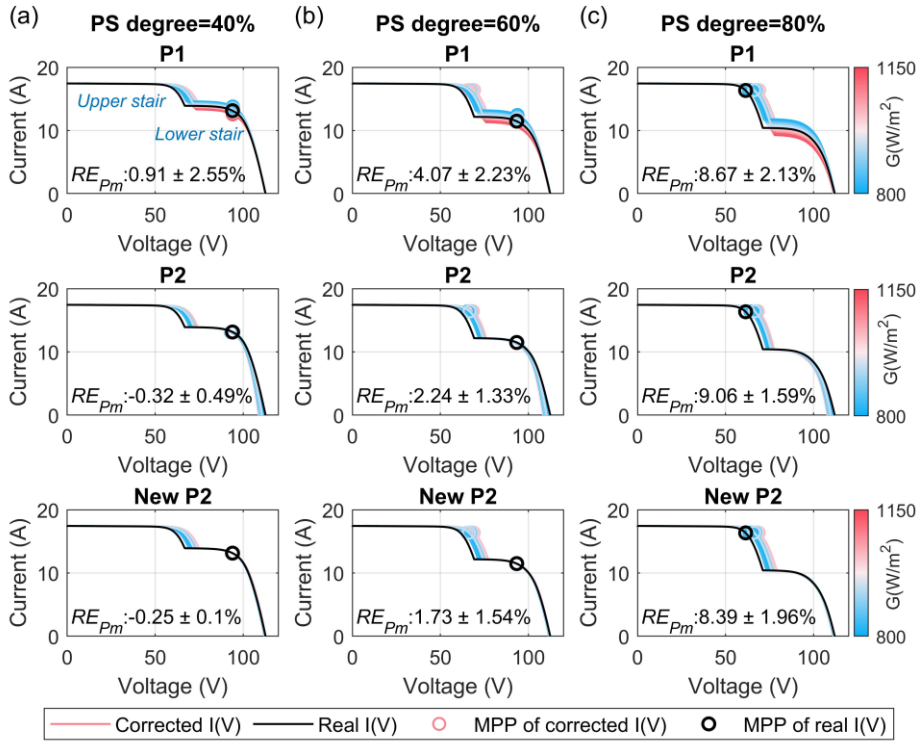


Figure III-14: Corrected curves using P1, P2, and NewP2 under three fault severities of PS

((a) PS degree=40%, (b) PS degree=60%, (c) PS degree=80% (the displayed color of corrected curves and MPP depends on the *G* of the curve to correct, the circles on the curves represent the MPP, RE_{P_m} are marked as 'mean'±'std'))

Moreover, it should be noted that, for P_m, which serves as an essential indicator for the health status of PV devices, its *RE* is nearly always positive (except low *RE* values at low fault severity) and gradually increases with the fault severity. This means that using these correction methods could result in a significant underestimation of the fault impact on P_m. Worse, the underestimation increases with the fault severity.

- ***RE* of fault parameters**

The impact of fault severity on fault parameters is also investigated, as illustrated in Figure III-15.

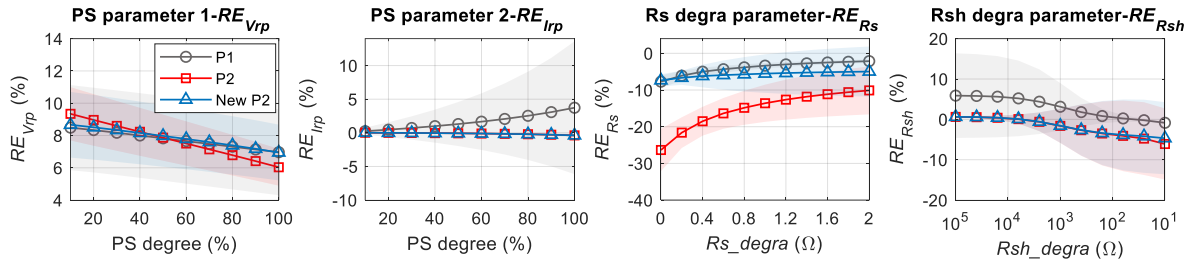


Figure III-15: *RE* of fault parameters under PS, R_s, and R_{sh} degradations with variable fault severity

(the marked line represents the mean of 174 values obtained from the correction of 174 curves under each fault severity, while the band area behind represents the standard deviation; the x-axis for R_{sh} degradation is set as log scale; the degree of severity is presented from healthy (except for V_{rp} and I_{rp} as these parameters do not exist under healthy condition) to severe on the x-axis from left to right)

With regard to the fault parameters as a function of fault severity, different trends are observed: for the parameters V_{rp} and R_s using all the three methods and for R_{sh} using P1, the *RE* decreases as the severity of the defect increases, while the opposite trend is observed for the other cases. This is due to the difference between the changing rate of the absolute error of one parameter and the changing rate of the real value, as defined in (III-18). Overall, NewP2 performs relatively better with lower and more stable *RE*.

It should be noted that within the range of severity of the defects tested, all the parameters are underestimated with the three methods. Therefore, if they are used as defect signatures, which is common for PV module degradation analyses that typically use R_s or R_{sh}. This could hamper the detection and diagnosis of these defects.

III.5.2. Performance of correction methods using multiple I-V curves

In this section, the method based on multiple I-V curves (P3) is evaluated. Section III.5.2.1 presents the selection of G and T_m for the reference curves. Sections III.5.2.2 and III.5.2.3 present the correction performance using simulated curves with the selected G and T_m, for constant and variable fault severity.

III.5.2.1. Selection of G and T_m for reference curves

Before the simulation of these curves, G and T_m need to be determined. In order to cover as many real situations as possible, the field-measured distributions of G and T_m of the same sc-Si module employed in the simulation are used. As shown in Figure III-6, three reference curves form one group of curves to get one corrected-to-STC curve. The G and T_m of these reference curves are determined by one-day measurement, covering both sunny and cloudy conditions, where the latter ones exhibit larger fluctuations of T_m. Besides, for both sunny and

cloudy conditions, three possible conditions are also considered for the irradiance: 1) lower than 1000 W/m^2, 2) around 1000 W/m^2, and 3) higher than 1000 W/m^2. In total, 120 groups of G and T_m are manually selected. Some examples are shown in Figure III-16.

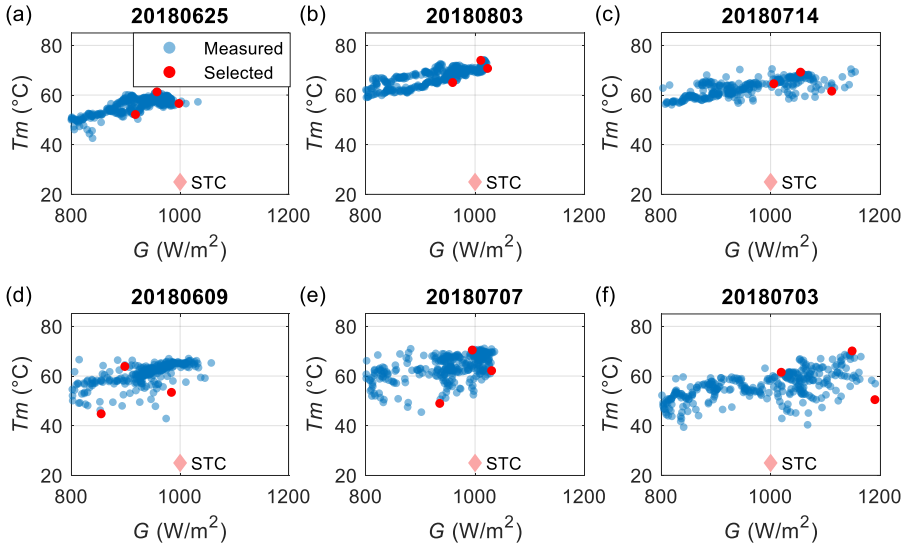


Figure III-16: Examples of selected G and T_m from different summer days

((a, b, c) clear sunny days, (d, e, f) cloudy days ((a, d): the 3 selected G all <1000 W/m^2, (b, e): selected G around 1000 W/m^2, (c, f): selected G all > 1000 W/m^2))

III.5.2.2. Correction performance with constant fault severity

Regarding the fault severity setting, firstly, we use the same fault configurations presented in Section III.5.1.2 and simulate 120 groups of reference curves. Then, P3 is applied to obtain 120 corrected curves, as shown in Figure III-17. The corresponding performance of key curve parameters and fault parameters are summarized in Figure III-18 and Figure III-19, respectively.

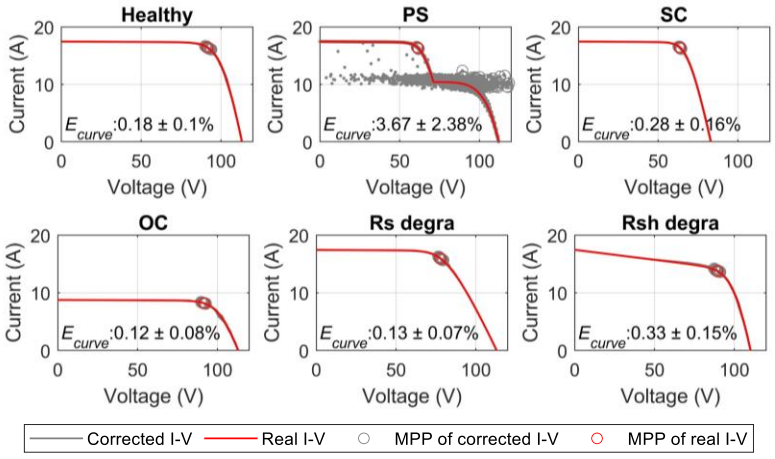


Figure III-17: Corrected I-V curves using P3

(For PS, due to the large fluctuations, the corrected curves are plotted with dot line for better presentation)

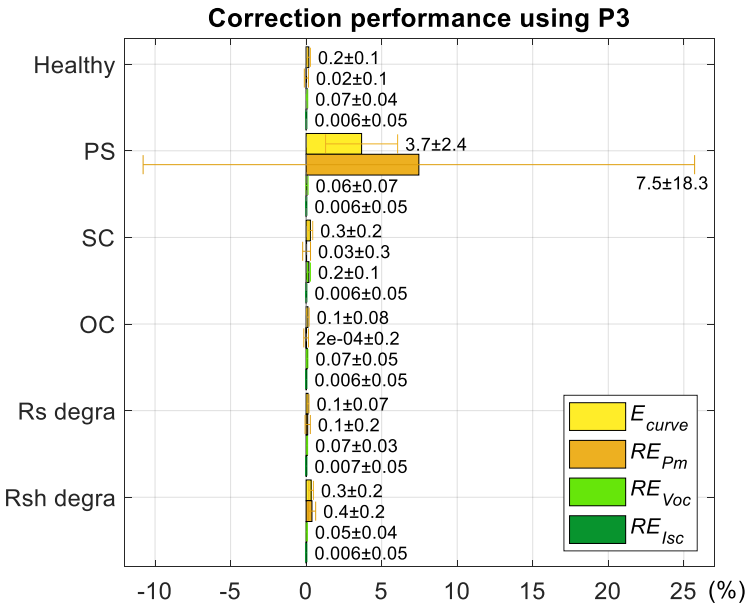


Figure III-18: Four metrics to present the correction performance using P3

(Values are presented in mean ± std)

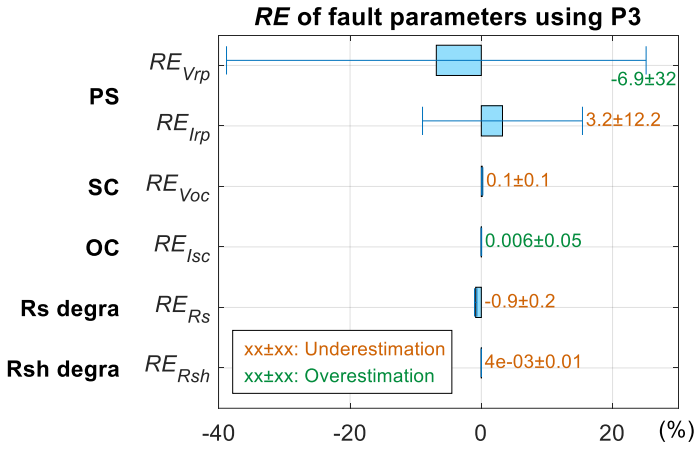


Figure III-19: Impact of correction on fault parameters using P3

(Values are presented in mean ± std)

From the results, clearly, except for PS, P3 achieves good correction (mean of all metrics within ±1 %) for all the types of PV array conditions (both healthy and faulty).

In the following, we analyze the 'catastrophic' correction error near the inflection point under PS. As illustrated in Figure III-6, the correction procedure using P3 is in two steps. An example of these steps under PS is shown in Figure III-20. In step 1, due to the unsynchronized appearance of inflection points on reference curves 1 and 2 (i.e., V_{rp} are not identical), another inflexion point is created on the intermediate curve, so that in step 2, a strange curve shape is

introduced. This phenomenon could lead to a significant error on P_m (15.75%) but with no significant impact on V_{oc} and I_{SC}.

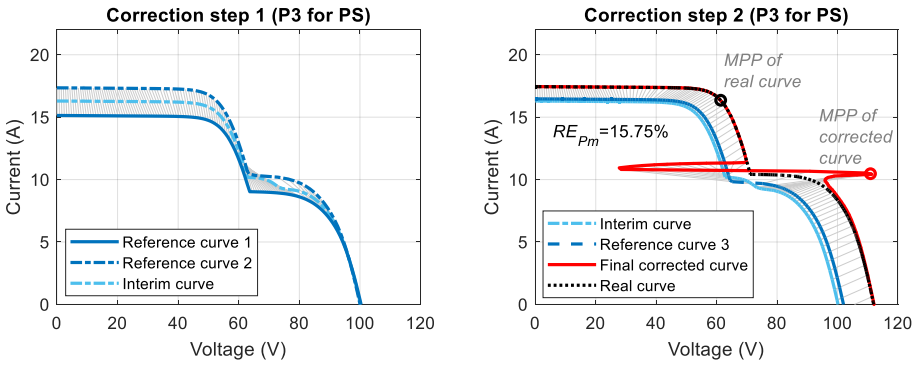


Figure III-20: Correction procedure under PS using three reference curves based on P3

((a) correction step 1, (b) correction step 2)

In addition, it should be noted that the fault impact on all the fault parameters is underestimated. As expected, the most significant error occurs under PS. Under the other conditions, the *RE* is within ±0.5 %.

III.5.2.3. Correction performance with variable fault severity

In this subsection, the impact of fault severity on correction performance using P3 is investigated. The severity for PS, R_s, and R_{sh} degradations is varied using the same settings as in Section III.5.1.3. The performance metrics as a function of fault severity are presented in Figure III-21.

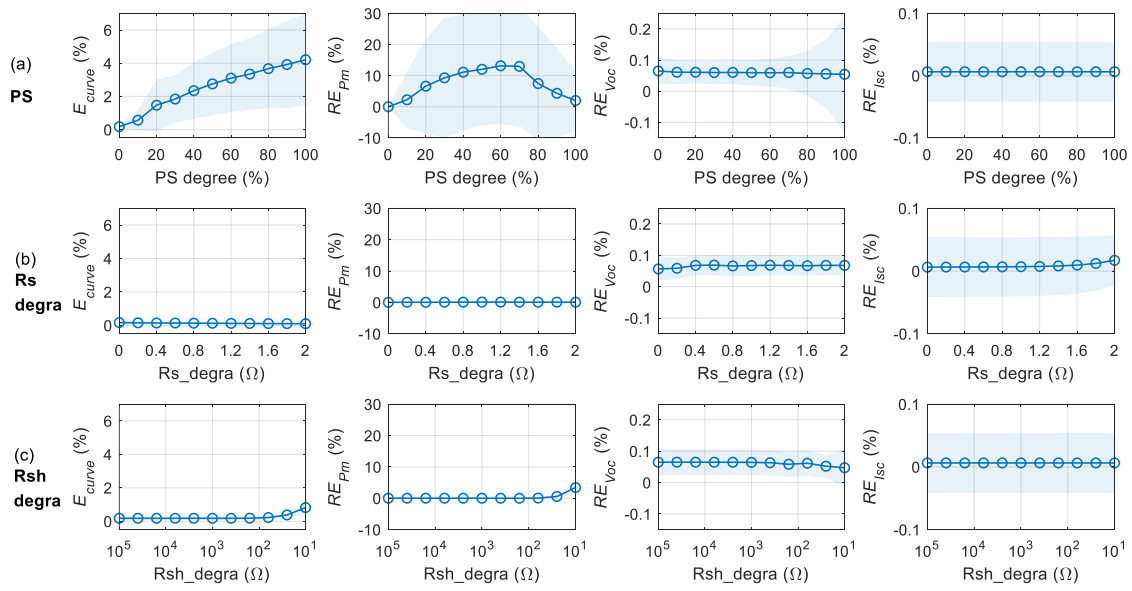


Figure III-21: Correction of curve error and key parameters for P3 under 3 faulty cases with variable fault severity

((a) PS, (b) R_s degradation, (c) R_{sh} degradation (the circled line represents the mean value for the 120 corrected curves at each fault severity, while the band area behind represents the standard deviation; the x-axis for R_{sh} degradation is set as log scale; the degree of severity is presented from healthy to severe on the x-axis from left to right))

From these results, the following remarks could be drawn:

- Regarding R_s and R_{sh} degradations, all the metrics are insensitive to the fault level;
- Regarding PS, only E_{curve} and RE_{P_m} vary according to the fault severity. E_{curve} has a monotonic variation, while RE_{P_m} exhibits a non-monotonic variation. This phenomenon is due to the displacement of the maximum power point (MPP) as illustrated in Figure III-22.

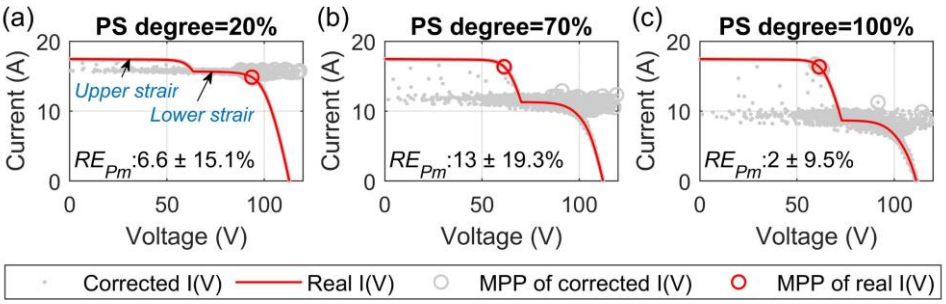


Figure III-22: Corrected curves using P3 under three fault severities of PS

((a) PS degree=20%, where MPP of real and corrected curves are all at 'lower stair,' (b) PS degree=70%, where MPP of real is at 'upper stair' while MPP of most corrected curves at 'lower stair' (c) PS degree=100%, MPP of real and most corrected curves are all at 'upper stair' (for each case, RE_{P_m} is marked as 'mean'±'std'))

The behaviors of fault parameters are also investigated with the comparison of identified parameters from both corrected and real I-V curves presented in Figure III-23 and the corresponding relative errors in Figure III-24.

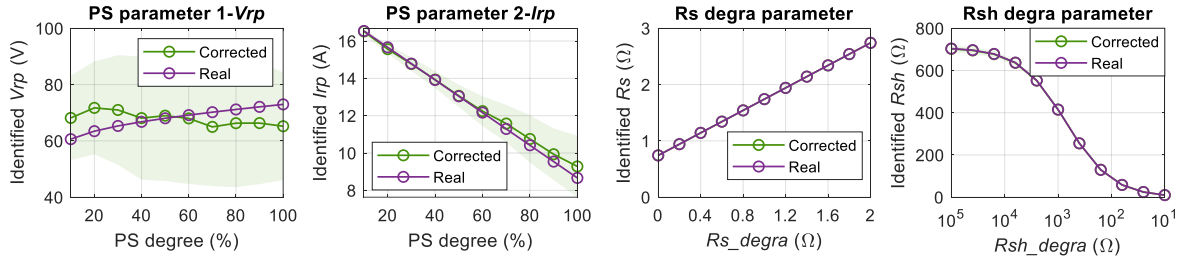


Figure III-23: Identified parameter from corrected (using P3) and original I-V curve for the fault of variable severity

(the circled line represents the mean, while the shadow band area represents the standard deviation; the values of R_s or R_{sh} identified from I-V curve are not equal to the additional resistance (illustrated in Figure III-3) due to the existence of inherent equivalent R_s (0.74 Ω) or R_{sh} (708.36 Ω) of PV arrays)

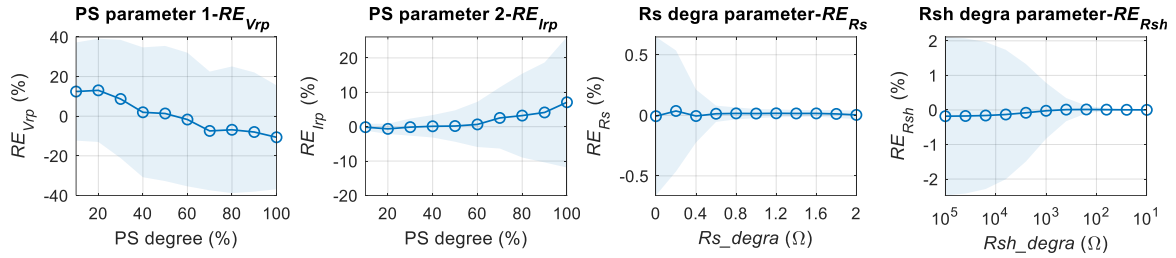


Figure III-24: *RE* of fault parameters under PS, R_s, and R_{sh} degradations for the fault of variable severity using P3

(the circled line represents the mean, while the shadow band area represents the standard deviation)

The large dispersion observed for RE_{Vrp} and RE_{Irp} confirms the poor performance of the correction near the inflection point under PS. The values of R_s and R_{sh} extracted from the original and corrected curves are very close, as confirmed in Figure III-23. This is also reflected in the low values of RE_{R_s} and $RE_{R_{sh}}$ displayed in Figure III-24. The mean value of RE_{R_s} lies within ±0.1 % and that of $RE_{R_{sh}}$ within ±0.2 %, and both are relatively insensible to the varying fault severity. It is also noted that the standard deviations of RE_{R_s} and $RE_{R_{sh}}$ exhibit a decreasing trend when the fault severity increases. In all, with varying faults, P3 achieves quite good and robust correction under R_s and R_{sh} degradation.

III.6. Discussion

The correction performance using three single and one multiple curves-based methods has been evaluated. Each method has its own pros and cons listed in Table III-7.

Table III-7: Pros and Cons of correction methods

	Single curve-based methods			Multiple curves-based method (P3)
	P1	P2	NewP2	
Pros	• Best overall voltage correction (except under OC) (For P1, P2, New P2) • Suitable for rapid field diagnosis	• Best overall current correction	• Best overall correction on entire I-V curve • Improve the voltage correction under OC compared to P1 • Similar performance as P2 for current correction but with more stable accuracy	• High-precision correction performance for most PV array conditions (except PS) • Free of correction coefficients
Cons	(For P1, P2, New P2) • Needs to determine the correction coefficients • Prone to large underestimation of P_m and fault parameters			• Not suitable for rapid field diagnosis • Needs at least three well-chosen reference curves • Poor correction near inflection point under PS with high effect on P_m

All these single curve-based correction methods, once established (i.e., correction coefficients determined), can conduct rapid correction of measured I-V curves. This allows their integration in real-time health monitoring of PV devices. However, as discussed in Section III.3.1, the determination of correction coefficients is a troublesome task. The coefficients determined by using simulated data do not always fit the real values, especially for the installed PV panels after a long-time operation. The proposed NewP2, which combines the advantages of P1 and P2, leads to better average correction performance on the entire I-V curve and key parameters for the tested faults. In fact, NewP2 only modifies the equation of voltage correction based on P2. Due to the inherent drawback of P2 on voltage correction and the independent relationship between *V* and *I*, NewP2 still does not outperform with all the faults. It introduces large correction error in the case of PS of one PV module, degradation of R_s and R_{sh} compared to healthy conditions, SC of one module and OC of one string.

As for the multiple-curves-based method (P3), except under PS, excellent correction is obtained. However, the selection of reference curves is still a manual process [25], [253], which could hinder its application in rapid or on-line field correction.

III.7. Conclusion

In this section, the 3 procedures proposed in IEC 60891, originally designed for the correction of I-V curves of healthy PV module or array, have been evaluated for the correction of I-V curves simulated on the defective PV array. It has been shown that all the procedures introduce significant errors due to G, T_m, and the severity of the defects. We have pointed out that using the Procedure 1 (P1) from the standard, a distortion of the curve's shape is commonly introduced with a relative error up to 13.8%. For P_m and the fault parameters extracted from the corrected curves, estimation errors also occur frequently. And worse still, most significant errors result in an underestimation of the parameters, up to 9.1% for P_m using Procedure 2 (P2). This can affect the detection of incipient PV defects if these parameters are used as signatures of alarm.

Among the single curve-based methods (P1 and P2), none of them could outperform for the correction of I-V curves simulated on all the PV array faulty conditions. Therefore, a P2-based improved method (NewP2) has been proposed. NewP2 modifies a coefficient in the equation of the voltage correction of P2 to reduce its correction error of voltage near V_{OC} part. At the same time, it keeps the original equation for the current correction of P2. NewP2 has exhibited more robust average performance than P1 and P2 with the decrease of average curve error from 3.45 % (using P2), 4.5 % (using P1) to 2.37% under constant fault severity. With varying fault severity, NewP2 also shows an overall better performance than P1 and P2 on the curve error, most of the key curve and fault parameters.

As for the procedure based on multiple curves, i.e., Procedure 3 (P3), it generally has higher performance than those based on a single curve except for the correction of I-V curves of PV array under PS. Under PS, due to the unsynchronized appearance of inflection points on reference curves, an additional inflection point will be created and then cause great distortion to the curve near the inflection point. With varying fault severity, similarly, except under PS, P3 is robust in the curve error (within ±1 %) and estimated fault parameters (within ±0.2 %). Besides, due to the requirement of at least three reference curves at different G and T_m for the correction of one curve, P3 is thus not suitable for the application of rapid field diagnosis.

PV health monitoring using I-V curves is a promising approach. In the next chapter, a complete PV FDD strategy using I-V curves will be investigated using the new proposed correction procedure (New P2).

Chapter IV

PV fault diagnosis using I-V curves and machine learning classifiers

IV.1. Introduction

Among the various types of PV electrical data (like output power, output voltage or current at DC or AC side as presented in Section II.3.1.2.3), the I-V curve generally embeds rich information about the health status of a PV module or array. Thus, I-V curve-based PV FDD is a promising topic [254]. As for acquiring I-V curves, common I-V tracers already support the measurement for a single module or small-scale string or array. In recent years, the hardware solutions (integrated at inverter level) have become commercially available to measure I-V curves periodically at the power plant level [17], [18]. In this sense, the I-V curve-based diagnosis approaches could be applied to all the common PV facilities.

One I-V curve, essentially, is determined by one vector of voltage and one of current. In the literature, there are different methods to manipulate the I-V curve to extract features for PV FDD. The typical practices include: (1) extract directly key features from the curve (like V_{OC}, I_{SC}, V_{MPP}, I_{MPP}, FF [19]–[21], or R_s, R_{sh} [234], [255]) for analysis; (2) calculate the first or second-order derivative of a curve as features to identify partial shading (PS) fault [22], [23]; (3) use complete I-V curves and compare it with simulated ones to generate residuals for analysis [256]. Their pros and cons are summarized in Table IV-1.

Table IV-1: Typical properties of common practices of I-V curves for PV FDD in literature

Methodologies	Pros	Cons
Use extracted key curve features	• Simple feature extraction • Suitable for multi-faults diagnosis	• Partial information • Sensible to noise and outliers • Hard to extract Rs, Rsh from curves of irregular shape
Calculate the 1st or 2nd derivative of the curve	• Able to identify the number of inflection points on the I-V curve under PS	• Sensible to noise and outliers • Only application to detect PS
Generate residual between measured complete I-V curve and simulated one	• Complete information • Suitable for multi-faults diagnosis (like PS , OC, SC, degradation)	• Require simulation model for real time simulation • Sensible to measurement error and environmental noise

Based on the comparison, it is observed that the common I-V curve-based diagnosis approaches all have their limits, especially lack an efficient usage of the complete I-V

curve information. For a given I-V curve, a complete usage is expected to capture richer information and achieve better diagnosis performance than a partial usage (like using only several extracted features).

Seen in this research gap, a novel PV diagnosis methodology is proposed in this section, which applies three different methods to embed complete I-V curves into input features for PV FDD. Since MLT commonly acts as a powerful tool for intelligent classification tasks [257], in this study, based on the popularity and applicability (like the dimension of input features), six common MLT classifiers are tuned and evaluated for PV fault classification. It is noticeable that we do not limit ourselves to one certain type of MLT but try to tune and test different MLTs simultaneously. This permits us to find out the best possible performance that could be achieved when using different types of input features.

This section is principally organized based on the four-step FDD scheme presented in Section II.3: Section IV.2 firstly presents the configuration of the I-V curve database; Section IV.3 then introduces the pre-processing of I-V curves; Section IV.4 presents the feature extraction, where the three methods to embed complete I-V curve information will be addressed; Section IV.5 performs the feature analysis results using different MLTs, where the FDD results using simulated data and validation using experimental data will be presented; Section IV.6 compares the proposed methodology with other common I-V curve based FDD techniques; Section IV.7 concludes the section.

IV.2. Configuration of the simulated datasets

IV.2.1. PV array model configuration

A PV array model similar to that used in Section III.2.1 to study the correction of the I-V curve is constructed as shown in Figure IV-1. The module parameters are based on Wiltec 62391-50W sc-Si module. This type of module is used in the experimental test and its power is within the measurement range of the I-V tracer in the laboratory. The array structure and parameters (detailed in Table IV-2) also fully correspond to that used in the experimental test (presented in Section IV.5.3). This permits the classifier tuned by simulated data could be applied to the diagnosis of the real PV array with the same configuration.

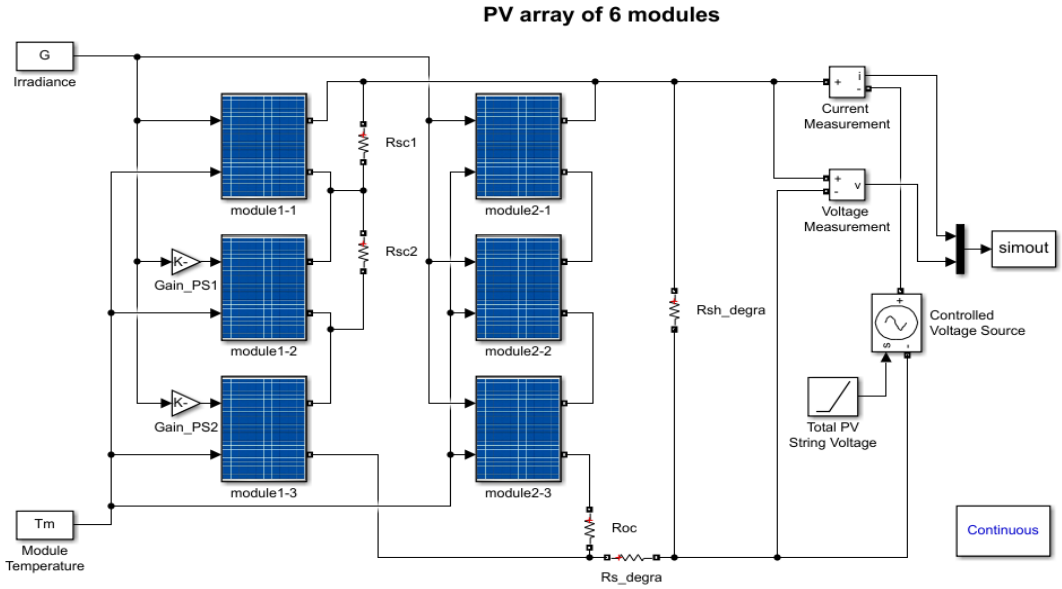


Figure IV-1: Simulation model of the PV array

Table IV-2: Parameter setting 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

In total, eight configurations of PV array are studied, i.e., healthy and seven faulty conditions:

- PS: 2 types, 1 or 2 modules in 1 string are shaded (abbreviated as PS1 and PS2, respectively) by adjusting the block gain value (Gain_PS) that controls the module's irradiance;
- SC: 2 types, 1 or 2 modules in 1 string is short-circuited (abbreviated as SC1 and SC2, respectively) by connecting a resistance R_{sc} in parallel;
- OC: 1 string is open-circuited by connecting a resistance R_{oc} in series;
- Rs degradation (Rs degra): increase of array equivalent series resistance, controlled by a resistance R_{s_degra} added in series;
- Rsh degradation (Rsh degra): decrease of array equivalent shunt resistance, controlled by a resistance R_{sh_degra} added in parallel;

Some examples of the I-V curves under these conditions are shown in Figure IV-2.

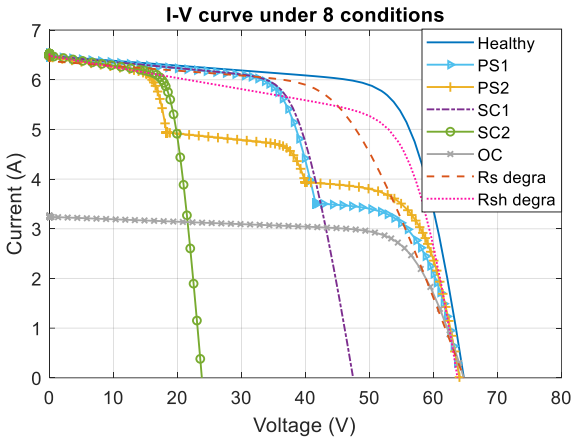


Figure IV-2: Examples of simulated I-V curves under 8 PV array conditions

From Figure IV-2, it is observed that these conditions could induce different changes on the I-V curves and these changes are useful for the following diagnosis.

IV.2.2. Generation of datasets

Based on the array model, two datasets are generated, one for training and another for the test. For each curve in both datasets, G is randomly varied in [400, 1200] W/m^2 and T_m in [20, 80] °C. Globally, T_m is set linear with G but varies in a range of 30 °C to cover more randomness of the real environment. An example of a generation of 300 pairs of (G, T_m) is shown in Figure IV-3.

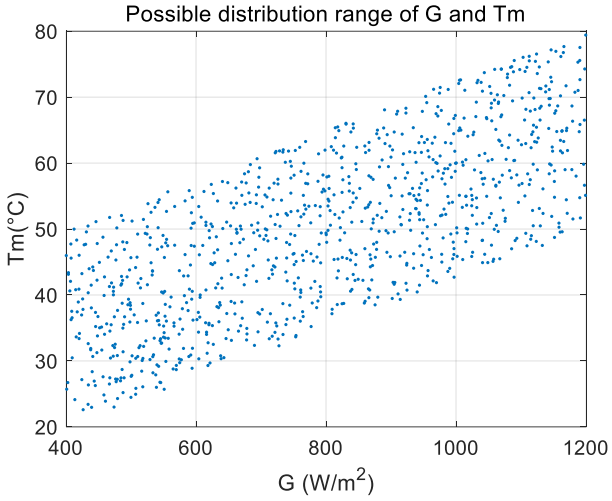


Figure IV-3: Examples of the distribution of G and T_m

The corresponding controlling parameters of I-V curves in these two datasets are identical as presented in Table IV-3. Besides, for PS, Rs, and Rsh degradation, the fault severity is also varied with the corresponding parameters covering uniformly the range listed in Table IV-3.

Table IV-3: Parameter setting for the healthy and faulty conditions

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

It is noted that the variation range for PS and degradation faults in Table IV-3 and Table III-3 are not identical. The variation ranges used in this chapter are relatively smaller than Chapter III. This setting is to avoid samples at too incipient fault severity, which may be confused with the ones at healthy condition. Examples of the curves of these conditions are presented in Figure IV-4. To better illustrate the impact of fault severity, the presented curves are generated under STC.

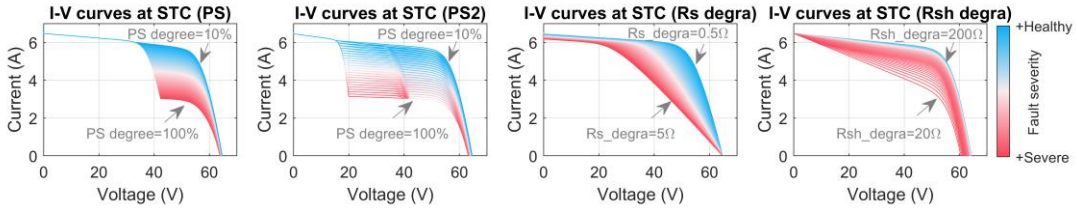


Figure IV-4: I-V curves (STC) under PS1, PS2, Rs, and Rsh degradation

(For PS2, the shading level for the two modules could be the same or different as shown in Figure IV-1. For Rsh degradation, the shape of the curve does not vary linearly with the Rsh_degra as this additional resistance is added in parallel to the array)

In fact, the curves obtained from simulation are all under ideal conditions, i.e., without any measurement error or fluctuation. Therefore, to train the classifiers with the samples with greater variability and closer to the field-measured ones, an estimated disturbance is added to the simulated I-V curve, specifically, to the vector of *V* and *I* that compose an I-V curve. Environmental noise and measurement random error are emulated. Additional disturbance vectors are firstly generated following a normal distribution and then added to the original *V* and *I*. The disturbance vector of *V* or *I* is generated randomly with the zero mean and a standard deviation value denoted σ_{dis_rand} for the random error and σ_{dis_env} for the environmental noise. They are determined by (IV-1) and (IV-2) , respectively.

$$E_{dis\_rand} = \sigma_{dis_{rand}}/\sigma_S \times 100\% \quad \text{(IV-1)}$$

$$R_{dis\_env} = 10\log_{10}(\sigma_S^2/\sigma_{dis\_env}^2) \quad \text{(IV-2)}$$

where, σ_S is the standard deviation of the signal *S*, which refers to the vector of either *V* or *I*. E_{dis_rand} is the error rate of random error (presented in percentage), set as 0.5% for both *V* and *I*. R_{dis_env} is the ratio of environmental noise, presented in decibel of 35dB for *V* and *I*. These values are determined by combing the datasheets of our measurement equipment used in field tests and the commonly reported measurement uncertainty in literature [258], [259].

For each simulated I-V curve (without additional disturbance), the above process is repeated five times. The original training dataset contains 2400 I-V curves simulated under all eight conditions (each condition has 300). With the additive disturbance, in total, the training dataset contains 12000 I-V curves (each condition has 1500 curves). The test dataset, with disturbance added, contains a total of 2400 curves where each condition is described with 300 curves. Both these two datasets will undergo the following pre-processing operations.

IV.3. Pre-processing of I-V curves

The pre-processing of I-V curves contains two principal steps: correction and resampling. These operations will be detailed in succession.

IV.3.1. Correction of I-V curves

Under various environmental conditions (different *G* and T_m), simulated or measured I-V curves could have distinct shape forms. Thus, to allow the classifiers to learn the curve shape better, the I-V curves are corrected to the same environmental condition, here, STC. The correction method (New P2) proposed in Section III.3.2 is adopted to correct the I-V curves.

IV.3.2. Resampling of I-V curves

After the correction, resampling of I-V curves is performed. This operation not only permits to make every I-V curve have an identical number of points, but also, which is more important, to guarantee uniform distribution of voltage curves points. This operation is particularly indispensable when treating new I-V curves with a different number of points or distribution, e.g., real array I-V curves measured by different I-V tracers. Thus, resampling is conducted here to ensure the general applicability of the proposed FDD methodology. Specifically, it is performed via the following steps:

1) Construct a voltage vector with N points uniformly distributed in [0, V_{max}] with a constant step (V_{max} is a constant for all the conditions, which could be set a little larger than the array V_{oc} at STC in healthy condition to avoid the voltage of improperly-corrected curve exceeding this range). An illustration of the voltage vector is marked in Figure IV-5 as markers of 'triangle';

2) At each voltage point, find the nearest point on the original curve and record its current value, as shown in Figure IV-5. It should be noted that, for the I-V curve after correction, the curves are generally shifted towards the right side on the I-V figure, i.e., the voltage of the far-left point not starting from 0, as discussed in Figure III-5. Thus, similarly, the extrapolation of I-V curve by Phang's method [249] is performed to obtain a complete I-V curve (voltage starting from 0). This permits the resampling based on the voltage on [0, V_{max}]. The extrapolation step is set at 0.2 V.

3) Form up a current vector of N points, where the N must be lower than the number of points on the original I-V curve to allow a down resampling.

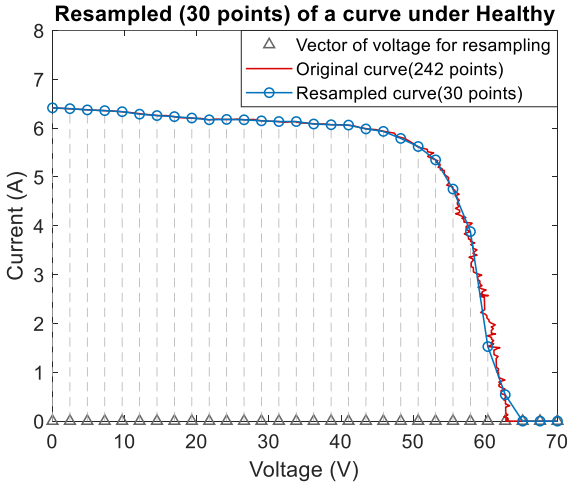


Figure IV-5: Illustration of resampling of an I-V curve from healthy array

It is easy to find out that the more points there are, the more information is captured. To determine N, the efficiency of resampling needs to be quantified. Here, the area error (E_{area}) [260] is adopted as the metric, as defined in (IV-3).

$$E_{area} = \frac{S_{Area\ of\ error}}{S_{original}} \times 100\% \qquad \text{(IV-3)}$$

where, $S_{original}$ is the size of the area enclosed by the original I-V curve and coordinate axis, $S_{Area\ of\ error}$ is the size of the area enclosed by the original and resampled I-V curve. An example of the illustration of the resampling (N =10 or 50) of a real I-V curve

for a PV array under PS2 (the original number of points is 363) is presented in Figure IV-6.

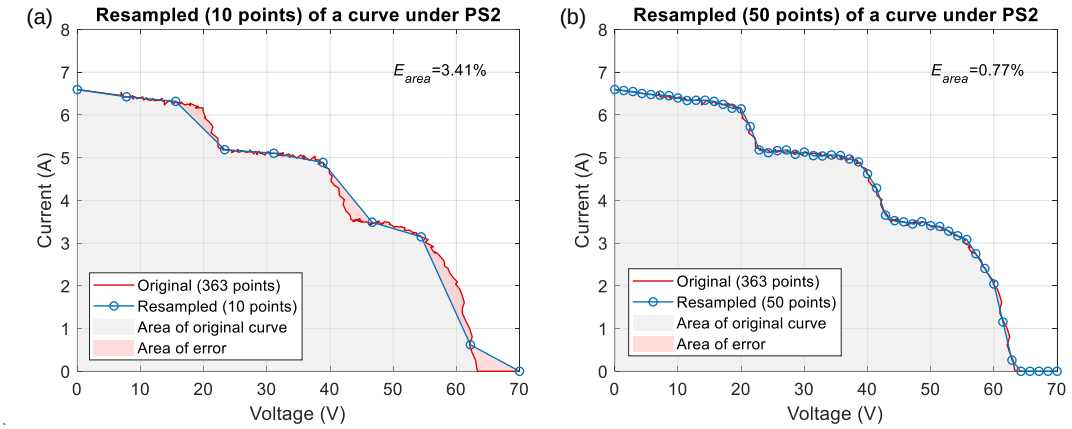


Figure IV-6: Example of resampling of an I-V curve from the array under PS2

Taking the I-V curves under the eight conditions, the mean and standard variation of E_{area} with the varying N are presented in Figure IV-7. The number of points of these original curves is not constant but varies between 298 to 363, which is due to the setting of the variable time step in the simulation.

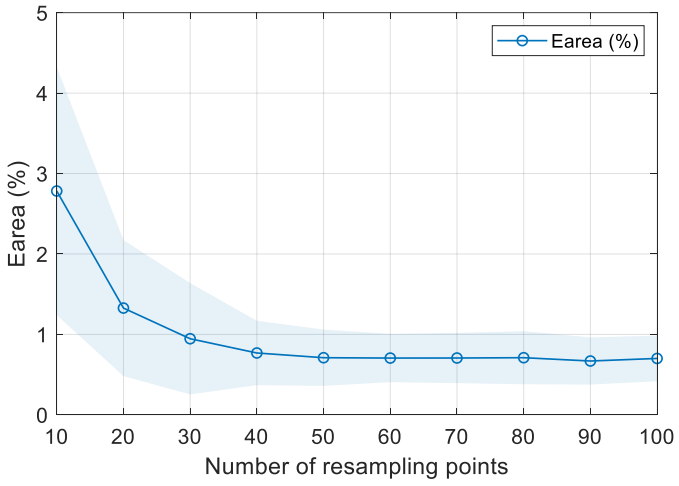


Figure IV-7: E_{area} of resampling of I-V curves as a function of N

(the blue line refers to the mean value while the band represents the standard deviation)

It is observed that, when N surpasses 40, the E_{area} becomes relatively small (~0.75%) and stable. However, it should be noted that, except for the efficiency of resampling, the complexity of the model (more points, larger complexity of classifier) also needs to be considered. Therefore, in this study, N is taken as a trade-off (N=50) to guarantee both good resampling performance and low methodology complexity. In this way, after resampling, for each I-V curve, one vector of current with 50 features is obtained.

IV.4. Feature extraction

After the pre-processing, feature extraction is performed to prepare the features with higher discriminability and lower redundancy. To this end, two operations are conducted, feature transformation and dimensionality reduction. Both are described in the following.

IV.4.1. Feature transformation

For the resampled current vectors, two types of approaches are proposed to further extract features. The first type is to use these vectors directly as input for classifiers (this method is named hereafter '*direct I-V*'). The second one is to transform the 1D vector of the current values to 2D matrix. Specifically, two 1D-to-2D transformation techniques, commonly practiced in the transformation of time-series signal, are adopted in this research: Recurrence plot (RP) and Gramian Angular Difference Field (GADF)

IV.4.1.1. Recurrence Plot (RP)

RP is a visualization tool that aims to explore the time series data via a 2D representation graph of its recurrences [261]. It reveals all the time dependency of vectors and constructs a square matrix by calculating the inner distance among all the points in the phase space [262]. The phase space is generated by embedding a delay (τ) of the input vector. In this study, it is noted that, in fact, the I-V curve is not a time series signal. However, as the voltage vector is uniformly resampled, the resampled current information then can be considered as a timely discretized information (as presented in Section IV.3.2). This allows us to use the time series assumption and apply RP to transform the resampled current vector.

Given a vector $\boldsymbol{x}\ (x_1, x_2, \dots, x_N)$ (N is the number of points), here, referring to the resampled current vector, the extracted trajectories in phase space are:

$$\overrightarrow{x_i} = (x_i, x_{i+\tau}), i \in \{1, \dots, N-\tau\} \quad \text{(IV-4)}$$

The RP matrix, denoted R, is the pairwise distance between the trajectories

$$R_{i,j} = \Theta\left(\epsilon - \left\|\overrightarrow{x_i} - \overrightarrow{x_j}\right\|\right), i, j \in \{1, \dots, N-\tau\} \quad \text{(IV-5)}$$

where; $\|.\|$ is the norm function, $\Theta(.)$ is the Heaviside function, ϵ is a threshold distance, x_i and x_j refers to i^{th} and j^{th} value in the resampled current vector $\boldsymbol{x}$, respectively. Based on ((IV-5), a RP matrix with binary values could be obtained. However, it should be noted that the operation of a threshold-based Heaviside function could lead to a loss of information [263]. Therefore, in this study, we directly take the pairwise distance to construct the RP matrix with $R_{i,j} = \left\|\overrightarrow{x_i} - \overrightarrow{x_j}\right\|, i, j \in \{1, \dots, N-\tau\}$.

The *R*-matrix reveals textures which are single dots, diagonal lines as well as vertical and horizontal lines. For example, fading to the upper left and lower right corners demonstrates that the original signal exhibits a trend or drift; or vertical and horizontal lines/clusters mean that some states do not change or change slowly for some time and this can be interpreted as laminar states [261]. The size of feature D_{matrix} of RP equals $(N-\tau)^2$. In this study, by finding the local minimum of the average mutual information of input vectors in phase space, τ is optimized to be 1, thus, D_{matrix}=2401. Some examples of RP transformations are illustrated in Figure IV-9

IV.4.1.2. Gramian Angular Difference Field (GADF)

GADF calculates the inner product of the input vector via the characterization of angular difference and preserves the temporal dependency [262]. Similar to the application case of RP, as the voltage vector is uniformly resampled, the current information can be considered as a timely discretized information. Therefore, GADF can be applied to the resampled current vector. GADF creates a matrix of temporal correlations for each pair (x_i, x_j) $(i, j \in \{1, \dots, N\}$, x_i and x_j refers to i^{th} and j^{th} value in the resampled current vector $\boldsymbol{x}$, respectively). To be specific, firstly it normalizes the vector into a range [-1, 1] with x_i normalized to x_i' :

$$x_i' = 2 \times \frac{x_i - \min(\boldsymbol{x})}{\max(\boldsymbol{x}) - \min(\boldsymbol{x})} - 1, \quad \forall i \in \{1, \dots, N\} \tag{IV-6}$$

Then it computes the polar coordinates of the scaled vector by *arccosine* function.

$$\phi_i = \arccos(x_i'), \quad \forall i \in \{1, \dots, N\} \tag{IV-7}$$

Finally, it computes the sine of the difference of the angles to form a GADF matrix.

$$GADF_i = \sin(\phi_i - \phi_j), \quad \forall i, j \in \{1, \dots, N\} \tag{IV-8}$$

The size of the generated matrix $D_{matrix} = N^2$. In this sense, when N=50, each current vector is transformed to one squared matrix with D_{matrix} =2500. An illustration of the transformation is shown in Figure IV-8.

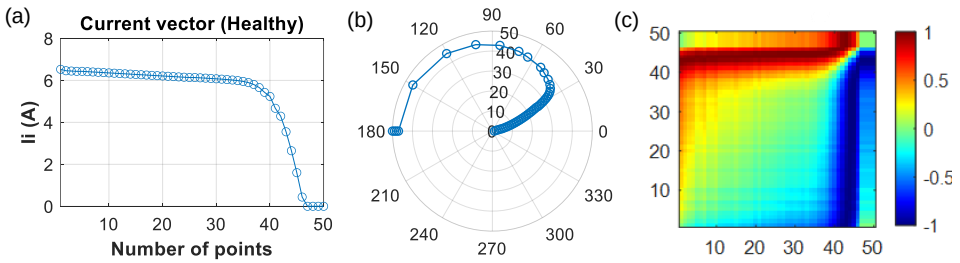


Figure IV-8: Process of the GADF transformation based on a resampled current vector
((a) resampled current vector, (b) polar coordinate graph, (c) GADF matrix)

Examples of GADF for other conditions are also illustrated in Figure IV-9.

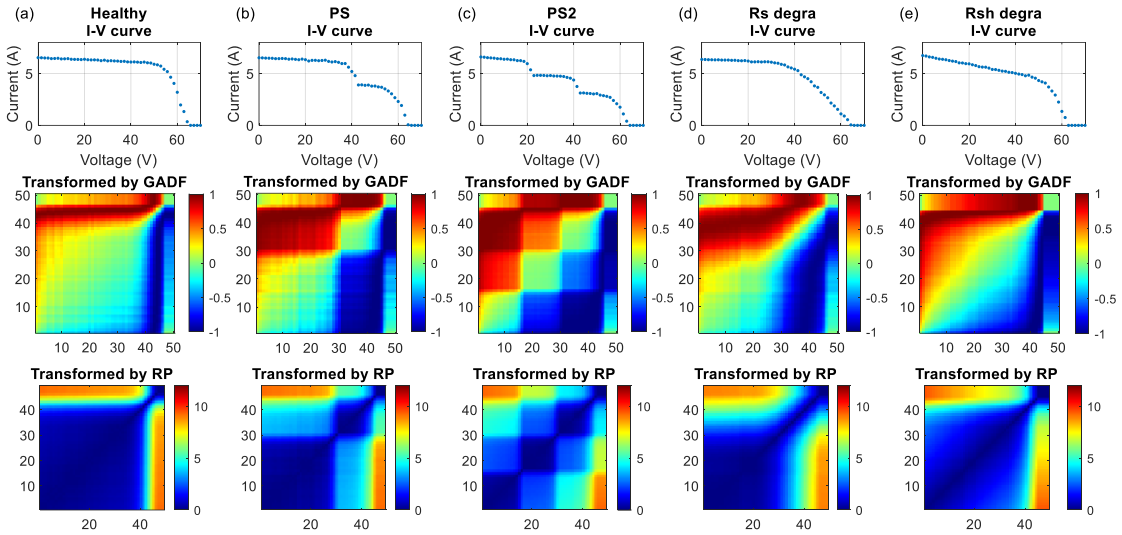


Figure IV-9: Examples of matrices transformed by GADF and RP

((a) Healthy, (b) PS, (c) PS2, (d) Rs degradation, (e) Rsh degradation (the value of each component on the matrix is marked by colormap))

From the examples of transformation, it is observed that the values in both RP and GADF matrix are symmetric, and the diagonal values are 0. When the value of current vector exhibits a decline, a 'square' will correspondingly appear, as seen from the PS and PS2 cases. When the slope near the V_{OC} or I_{SC} part of the I-V curve changes, the diagonal area in the matrix will also shrink along with different directions. In this way, all the changes in the I-V curve could be reflected in the transformed matrix.

These transformation operations increase the complexity of features for analysis. However, the discriminability of features under different conditions could be improved, which will be reflected in the diagnosis results and further discussed in Section IV.5.2.4.

All the presented three feature extraction methods (direct I-V, RP, and GADF) capture complete information from the current vectors resampled from original I-V curves. In addition to the proposed three feature extraction methods, one typical traditional feature extraction method for PV FDD, which uses partial I-V curve information by extracting key parameters, will also be evaluated for comparison. This method, denoted as '*8paras*', uses G, T_m, V_{MPP}, I_{MPP}, V_{OC}, I_{SC}, R_s and R_{sh} as features for analysis. The R_s and R_{sh} are calculated by the reciprocal of the slope of the I-V curve near the V_{OC} side and I_{SC} side, respectively. In fact, R_s and R_{sh} are not easily obtained due to the manipulation of the whole I-V curve, but they can present the variation trends of the curve, which are also the key information.

IV.4.2. Dimensionality reduction of features

For the features extracted by direct I-V, RP, or GADF, it is easy to recognize the large redundancy of information. For example, the current vector used in direct I-V may contain points of similar current value, especially for the points resampled from the part of the I-V curve with a gentle variation. Each of these points, in fact, all corresponds to a feature. Thus, redundancy of information is involved in the original feature. To reduce the complexity of classifiers and speed up the training process, a reduction in the number of features is performed. One commonly adopted technique, PCA [124], is applied. The number of principal components is determined by explaining at least 98% of the variance of the original data. This permits the final samples used for classification to preserve quasi-complete information of original I-V curves. In addition, the diagnosis performance without PCA will also be discussed in Section IV.5.2.3.

IV.5. Feature analysis and FDD results

After the feature extraction, the four types of features (8paras, direct I-V, RP, or GADF) will be analyzed for PV FDD. The analysis procedure is presented in Figure IV-10.

MLTs are adopted as feature analysis tools. For each type of feature, six common types of MLTs (presented in Section IV.5.1) will be employed. For each type of MLT, candidate models with different hyperparameters will be built. Specifically, the training database (presented in Section IV.2.2) is divided into 80% for training and 20% for validation. The validation dataset is used to select the trained candidate models of one type of MLT with the best performance. Then, the test database (presented in Section IV.2.2) is used to evaluate the 6 selected MLT models and find the best one. The corresponding diagnosis results will be presented in Section IV.5.2. This model will then be further evaluated by the field test dataset, which will be detailed in Section IV.5.3.

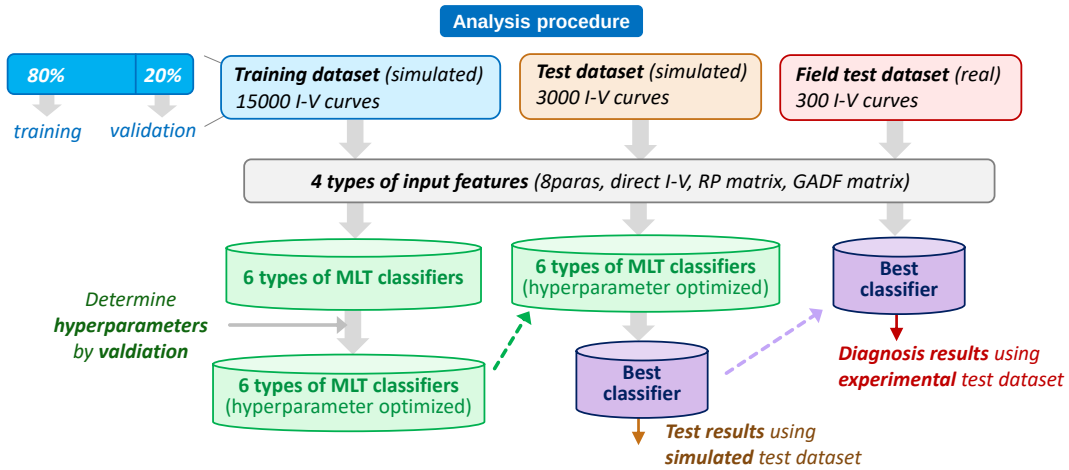


Figure IV-10: Feature analysis procedure

IV.5.1. Analysis techniques-machine learning classifiers

Based on the literature review of applications of MLT in PV FDD (presented in Section II.3.4.3), in this study, 6 common MLT classifiers are applied: ANN, SVM, kNN, DT, RF, and NBC. To build up each type of MLT, various hyperparameters need to be determined, e.g., the neuron numbers and the layer numbers for ANN, kernel function, kernel scale and penalty for SVM. The hyperparameters of these MLT classifiers to be tuned are listed in Table IV-4.

Table IV-4: Configuration of MLT classifiers

MLT	Configuration
All types	Common setting • Normalization: [-1,1] • 5-fold validation • Hyperparameter optimization method: grid search or Bayesian method
ANN	• Structure: 1 or 2 hidden layers (5-50 neurons) • Loss function: Cross entropy • Train algorithm: Scaled conjugate gradient • Function: tansig (hidden layer), softmax (output layer)
SVM	• Multiclass method: one-vs-one • Kernel scale: optimized • Penalty: optimized • Kernel function: Linear, Quadratic, Cubic, Gaussian, or RBF
kNN	• Number of neighbors: optimized • Distance metric: Euclidean, City block, or Chebyshev
DT	• Max number of splits: optimized • Split criterion: Gini's diversity index • Minimum number of leaf nodes: optimized
RF	• Max number of splits: optimized • Minimum number of leaf nodes: optimized • Number of predictors selected for each split: optimized
NBC	• Distribution type: Gaussian, or Kernel • Kernel Type: Gaussian, Box, Epanechnikov, or Triangle • Kernel smoothing window width: optimized

Relying on the toolbox ‘Classifier learning’ in Matlab, for each type of MLT, classifiers with different hyperparameter settings are built up as candidate models. These models

will then be validated by the validation dataset. Specifically, the 5-fold cross-validation is applied. The original dataset is randomly partitioned into 5 equal-sized subsamples. Of these subsamples, a single subsample is used as the validation dataset, and the remaining 4 are used as the training dataset. This cross-validation process is repeated 5 times, with each of these subsamples used exactly once as the validation dataset. Then, the average classification performance is used to select the best model among the candidate models of different hyperparameters. In this way, using each type of input feature, 6 MLT models (ANN, SVM, kNN, DT, RF, and NBC) are built up.

IV.5.2. Diagnosis results using simulated data

For each type of input feature, the 6 MLT models are evaluated using the test dataset. The diagnosis results, robustness to additional disturbance, discussion on the impact of PCA and the feature transformation (RP and GADF) will be addressed in the section.

IV.5.2.1. Performance of fault classification

The test accuracy of the 6 classifiers using the 4 types of input features (8paras, direct I-V, RP, or GADF) is presented in Figure IV-11. The details of the best classifier are listed in Table IV-5 and the corresponding confusion matrices are illustrated in Figure IV-12-Figure IV-15.

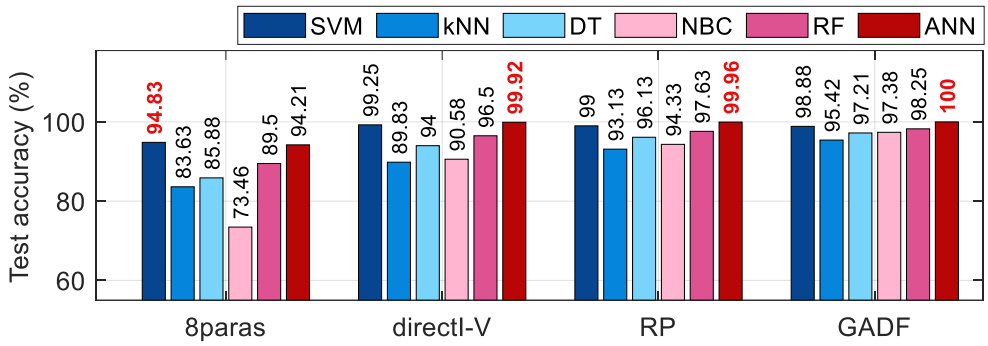


Figure IV-11: Test accuracy of all classifiers

(the best accuracy is marked in red bold)

Table IV-5: Details of the best classifiers

Type of input features	Best classifier	Test accuracy	Classifier hyperparameters
8paras	SVM	94.83%	Kernel: Polynomial 2, BoxConst=29.3
direct I-V	ANN	99.92%	#Neurons of hidden layer =15
RP	ANN	99.96%	#Neurons of hidden layer =10
GADF	ANN	100%	#Neurons of hidden layer =32

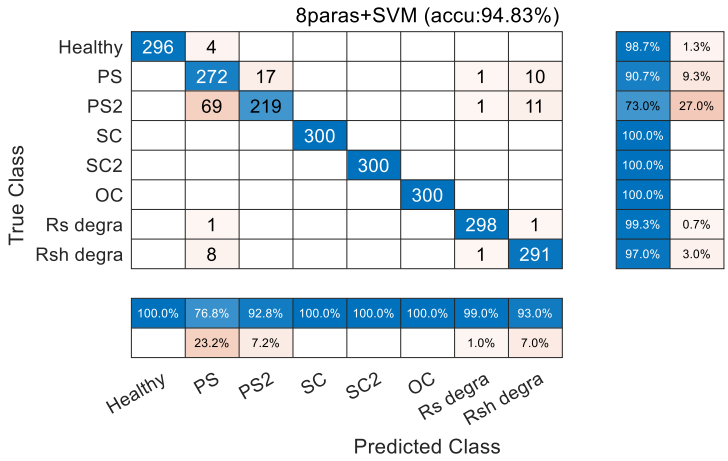


Figure IV-12: Confusion matrix of the best classifier when using 8paras as input feature

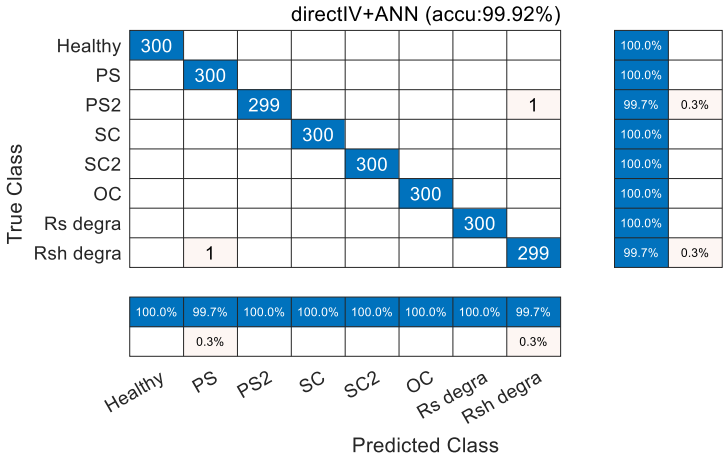


Figure IV-13: Confusion matrix of the best classifier when using direct I-V as the input feature

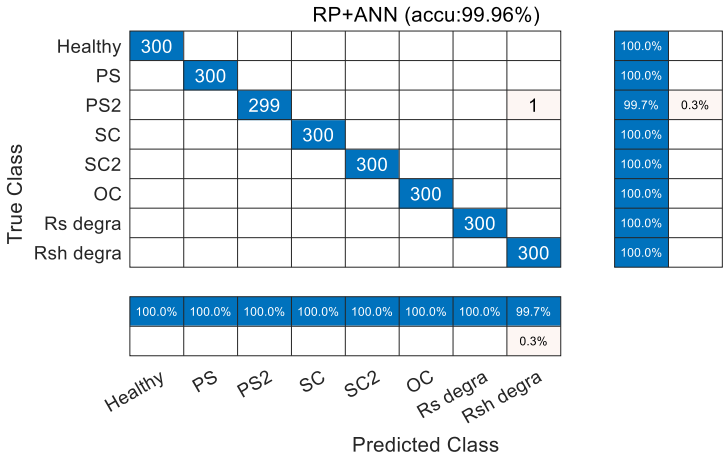


Figure IV-14: Confusion matrix of the best classifier when using RP as the input feature

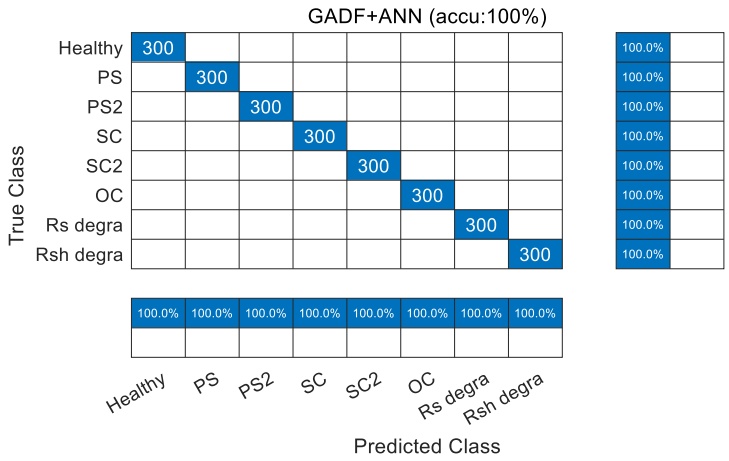


Figure IV-15: Confusion matrix of the best classifier when using GADF as the input feature

From the view of input feature types, it is found that the complete I-V curve-based features (i.e., direct I-V, RP, and GADF) permit the classifiers to achieve globally higher classification accuracy than that based on partial I-V curve information (i.e., 8paras). For 8paras type, the best test accuracy only achieves 94.83%. Large misclassification occurs to PS1, PS2, and Rs degradation. This is logical since, under these conditions, the MPPs of some of the curves could have similar or even identical positions, especially with the existence of measurement noise. Therefore, if only V_{MPP}, I_{MPP}, V_{OC}, I_{SC}, R_s and R_{sh} are adopted as curve features, less satisfying performance will be resulted in.

Comparatively, for the complete I-V curve-based methods (direct I-V, RP, and GADF), the quasi-whole curve information is embedded in the input features. Thus, the classifiers can learn the variation trend of the curves and achieve better classification results with the best performance all higher than 99.9%. For GADF, using ANN as the classifier could achieve 100% classification. For direct I-V and RP, there are 1 or 2 I-V curves wrongly classified.

To further understand these results, the output values of the neurons in the output layer are studied. A zoom on the output layer of the ANN model is presented in Figure IV-16, where there are eight neurons in the output layer. Each neuron corresponds to one condition. The sum of all the output values equals 1 as the SoftMax activation function is applied in the output layer. The predicted class is assigned to the condition whose corresponding neuron outputs the largest value. In this way, by analyzing the neuron output values, the confidence of decision-making could be understood.

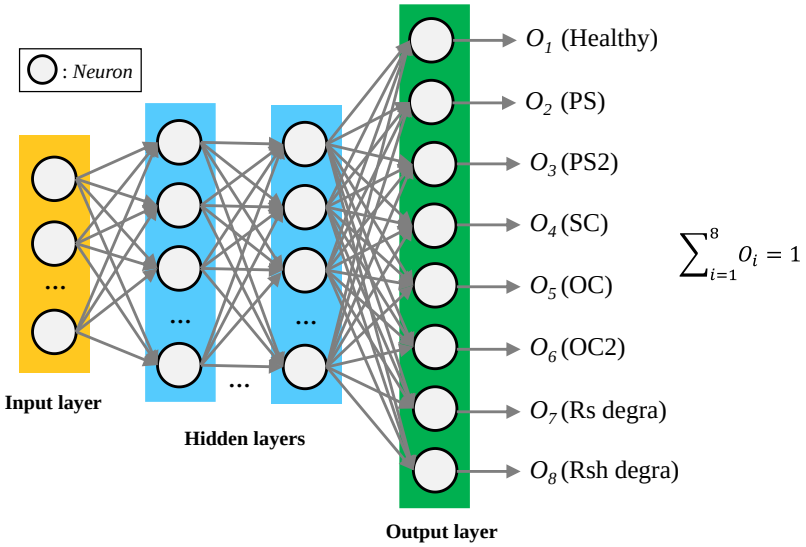


Figure IV-16: Illustration of the output neurons in ANN models

For these wrongly-classified cases presented in the confusion matrices, the neuron outputs, the wrongly identified I-V curves and the original I-V curves without disturbance added are illustrated together in Figure IV-17- Figure IV-19.

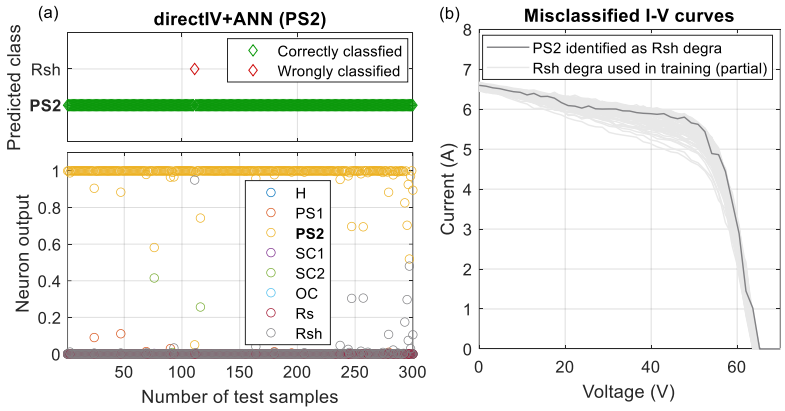


Figure IV-17: Detailed test results using direct I-V +ANN under PS2

((a): upper figure presents the predicted class as a function of number of samples, lower figure presents the value of 8 neurons in the output layer of ANN classifier, each neuron corresponds to one condition, the predicted class is assigned to the condition whose corresponding neuron outputs the largest value; (b): misclassified I-V curve and the curves of the wrong class used in training)

Seen from Figure IV-17 (a), there is one red 'diamond' marked in upper graph and in the lower neuron output, the corresponding gray point (presenting the neuron output of Rsh degradation) surpasses the yellow one (that of PS2). These correspond to the result that there is one I-V curve of PS2 misclassified as Rsh degradation. This misclassified I-V curve is shown in Figure IV-17 (b), in which it is clearly within the area surrounded by the I-V curves under Rsh degradation used for training. This means that this PS2

curve is similar to the ones under Rsh degradation, especially the ones with low fault severity. This suggests that this curve is prone to be misclassified as Rsh degradation.

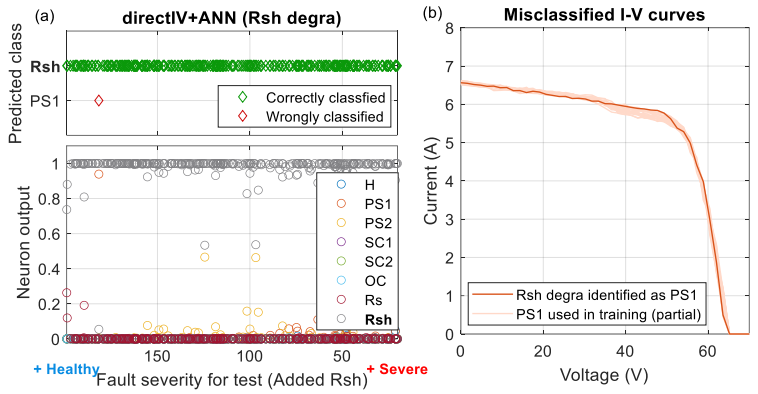


Figure IV-18: Detailed test results using direct I-V +ANN under Rsh degradation

When using direct I-V +ANN, there is one I-V curve under Rsh degradation misclassified as PS, as seen from Figure IV-18(a). In Figure IV-18 (b), similarly, the shape of the misclassified I-V curve is like those under PS used in training (at low PS degree).

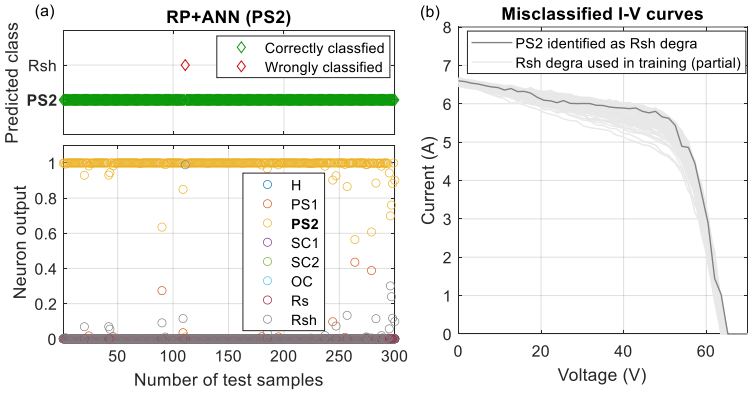


Figure IV-19: Detailed test results using RP +ANN under PS2

When using RP as features for FDD, there is only one curve misclassified, i.e., the same one in Figure IV-17 when using direct I-V as the feature. This suggests, when using direct I-V and RP to extract features from this curve, these features are easily confused with those from curves under Rsh degradation. Nevertheless, it could be noted from Figure IV-15 that, when using GADF to extract features, all the I-V curves are correctly classified, even for this PS2 curve. This means GADF is capable to better extract the fault features from this curve and distinguish it from the others, as can be observed from the neuron outputs under PS2 shown in Figure IV-20.

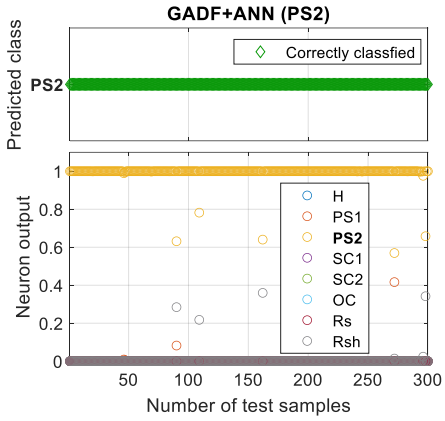


Figure IV-20: Detailed test results using GADF +ANN under PS2

Among direct I-V, RP, and GADF, if we count the average accuracy of all the 6 types of MLT classifiers, it is the RP and GADF that achieve higher global accuracy. This demonstrates that the transformation of features permits all the types of MLT classifiers easier to achieve better classification performance. This could also be seen from the t-SNE [264] graphs shown in Figure IV-21, which visualize the discriminability of these features in 2D graphs.

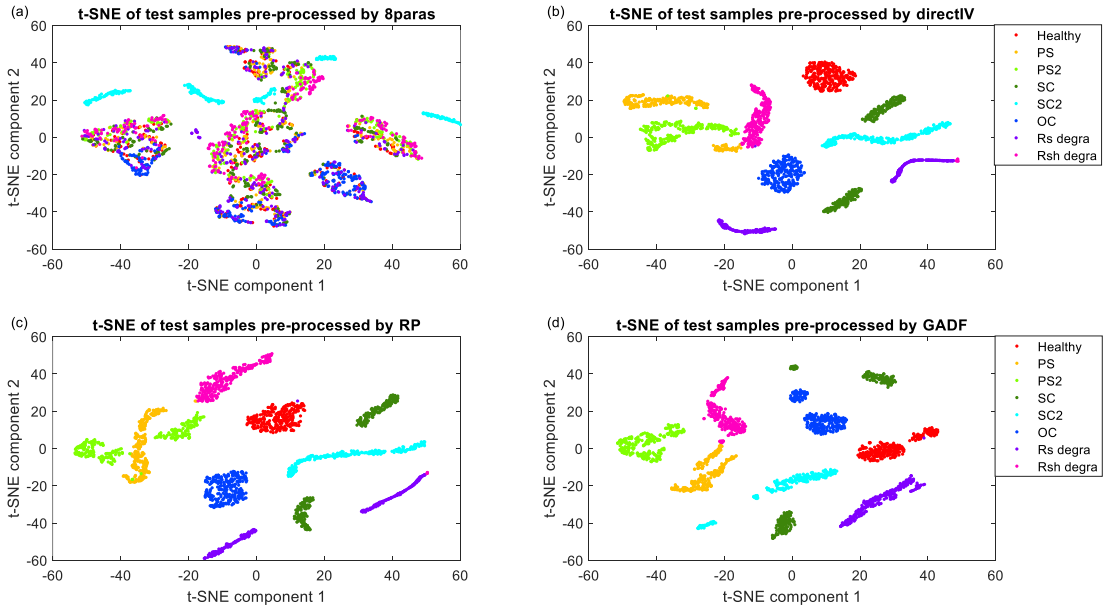


Figure IV-21: t-SNE graphs of 4 types of input features extracted from the test database

((a) 8paras, (b) direct I-V, (c) RP, (d) GADF)

From the t-SNE graphs, it is easy to find out that the clusters of features- 8paras are nearly all mixed up, which shows up its low discriminability and corresponds to its less satisfying performance. For RP and GADF, the clusters are relatively more separated than that of direct I-V, like among PS1, PS2, and Rsh degradation, which is clearer for

GADF. This means, using RP or GADF, the discriminability of input features is enforced, which therefore allow all the MLT classifiers easier to achieve better classification, as presented in Figure IV-11.

IV.5.2.2. Robustness to additional disturbance

In this part, the best MLT classifiers for the 4 types of input features (presented in Table IV-5) will be tested by new test datasets to evaluate the robustness of the classifiers. In these datasets, the added disturbance level of random measurement error and environmental noise will be varied independently from 0 to a higher level than that used in previous test dataset. The corresponding best classification results when using the 4 types of input features are presented in Figure IV-22.

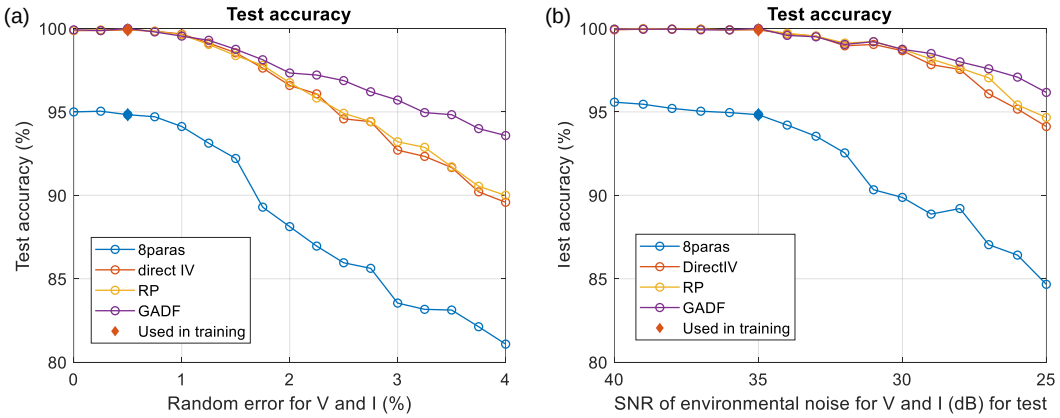


Figure IV-22: Best classification accuracy using test dataset with varying level of added disturbance

((a): varying 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')

It is observed that, globally, the best test accuracy when using the 4 types of features all gradually decreases with the increase of disturbance level. The complete I-V curve-based features (direct I-V, RP, and GADF) once more show superiority than the partial curve-based feature (8paras).

As for direct I-V, RP, and GADF, the best test accuracy shows a similar variation trend at a low disturbance level. However, GADF experiences a slighter decrease of performance with the increase of disturbance, especially with the increase of random error on *V* and *I*, e.g., from the confusion matrix of test results depicted in Figure IV-23 when the random error is at 4%.

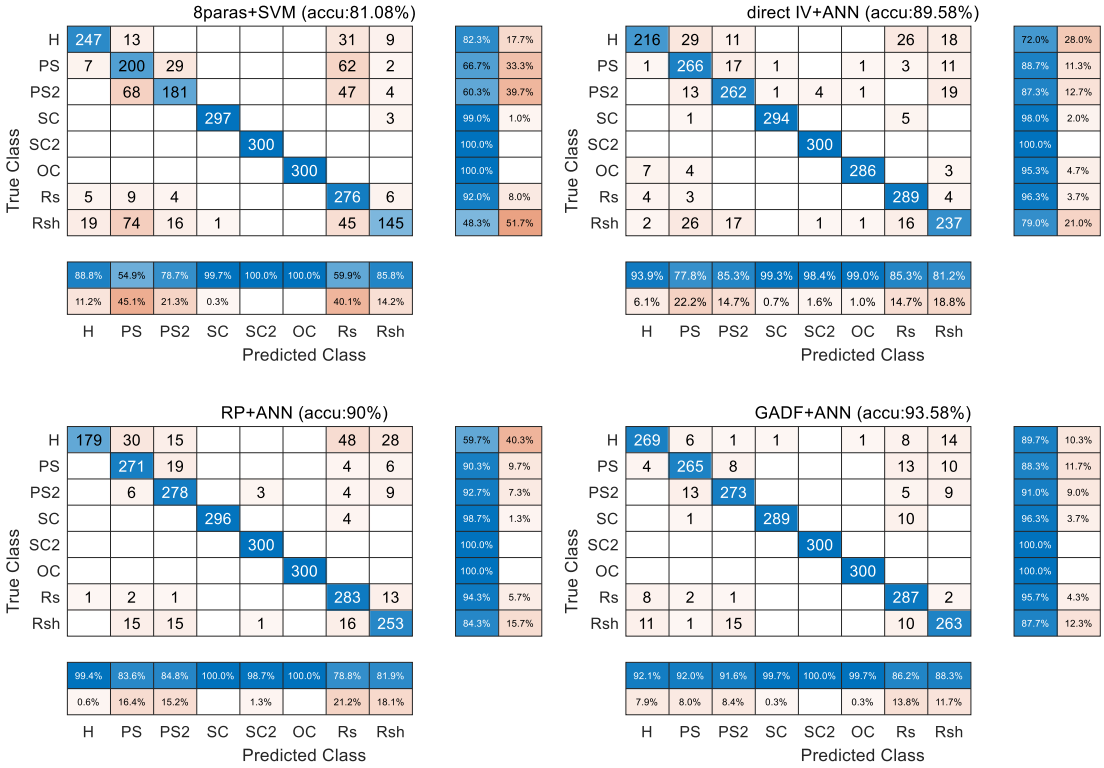


Figure IV-23: Confusion matrix of test results using the best classifier with the random error of V and I at 4% in the test dataset

From Figure IV-18, when the random error of V and I in the test dataset reaches a high level of 4% (compared to the 0.5% used in training), the classification performance when using all the types of complete I-V curve-based features decreases, which can also be seen from Figure IV-23. Nevertheless, it is observed that GADF clearly outperforms direct I-V and RP, with fewer I-V curves under heathy and Rsh degradation misclassified into other conditions. This demonstrates that the classifier using GADF to extract features has overall better robustness to the additional disturbance than using other types of feature extraction methods.

IV.5.2.3. Influence of PCA

As presented in Section IV.4.2, for direct I-V, RP, or GADF, the dimension of features will be reduced by PCA before fed into MLT classifiers. In this scenario, the impact of this operation on the diagnosis performance will be discussed. Specifically, the processing time and the test accuracy with and without PCA are compared as shown in Table IV-6. In this study, the platform is Matlab R2020b with a CPU of Intel(R) Xeon(R) E-2174G CPU @ 3.80GHz and RAM of 32Go.

Table IV-6: Influence of PCA on time and accuracy

Input feature type	PCA	Computation duration (s)	Test duration (s)	Test accuracy
direct I-V	Yes	41.96	0.01	99.92%
	No	59.02	0.01	99.88%
RP	Yes	146.03	0.01	99.96%
	No	1298.56	0.13	99.92%
GADF	Yes	80.63	0.01	100%
	No	717.79	0.11	100%

From the comparison, it is observed that PCA operation allows the following merits:

- No compromise on the diagnosis accuracy
- Reduce the total computation duration (pre-processing + training), especially for RP and GADF with the time with PCA taking only about 11% of that without PCA.
- Increase the prediction speed (less test duration)

Thus, it is favorable to conduct the dimension reduction operation by PCA, especially for high-dimensional features, like RP and GADF.

IV.5.2.4. Influence of transformation

In Section IV.4.1, in addition to direct I-V, we have introduced two feature transformation methods, RP and GADF. Clearly, the feature dimension after transformation greatly increases as the current vector is transformed into a square matrix. This, of course, could increase the complexity of FDD methodology, like the longer time needed to perform dimensionality reduction and training, which can also be seen from Table IV-6.

On the other side, it should be noted that the transformation is able to increase the discriminability of features, as presented in the t-SNE graph in Figure IV-17. This can also be noted from the neuron output. From the test results given in Figure IV-13 - Figure IV-15, for the I-V curves of the array under healthy condition, direct I-V, RP and GADF all could achieve 100% classification. However, if we focus on the detailed output results of the best classifier, the difference in the confidence of classification could be observed, e.g., from the results under healthy condition, as shown in Figure IV-24.

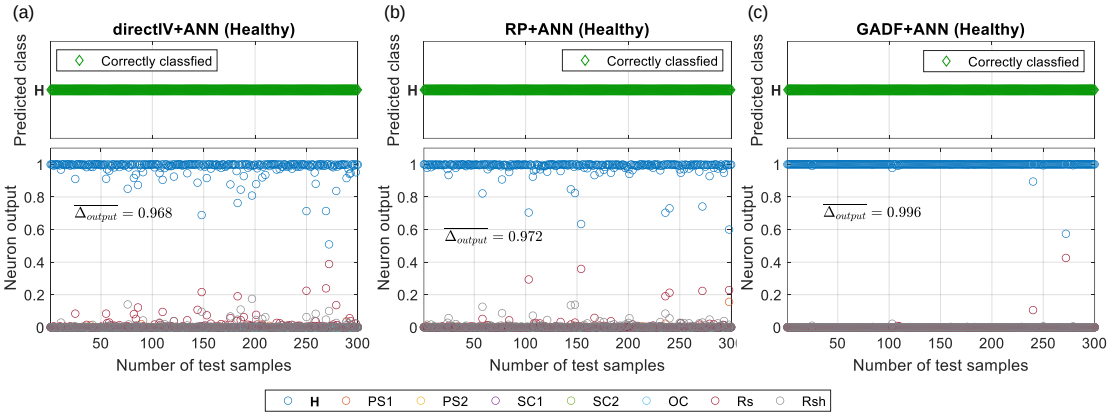


Figure IV-24: Neuron output of the best classifier under heathy condition using direct I-V, RP and GADF

((a): direct I-V, (b): RP, (c): GADF)

For these 3 types of features, ANN is the best classifier. The output values of the 8 neurons in the output layer (corresponding to the 8 conditions) are presented. For each prediction of class, the sum of all the 8 neuron outputs equals 1, the predicted class is assigned to the class whose neuron outputs the largest. In this way, the larger the difference between the neuron outputs is, the more confidence in prediction making. To quantify the neuron output difference, Δ_{output} is defined, which equals the largest neuron output minus the second largest output. The larger the Δ_{output} is, the more confident the classifier is to make this decision. In Figure IV-24, the mean of the Δ_{output} ($\overline{\Delta_{output}}$) for the 300 test I-V curves under healthy condition is also presented. From the results, $\overline{\Delta_{output}}$ of direct I-V and RP are similar with the one of RP a little bit larger, while the $\overline{\Delta_{output}}$ of GADF is the largest reaching 0.996. This means, using GADF as features, the best classifier is more confident to make correct classification, which once more reflects the merit of performing feature transformation.

In addition, it is also a manifestation that the robustness to the additional disturbance when using RP and GADF is enforced to that using direct I-V, as seen from Figure IV-18, especially for GADF.

To investigate why the transformation could increase the discriminability of features, we analyze the difference between features under different conditions after transformation. In fact, the I-V curves of the array under incipient fault conditions (like low PS degree, low additional R_s, large additional R_{sh}) are prone to be confused with those under healthy conditions, as seen from the wrongly-classified I-V curves shown in Figure IV-17 to Figure IV-19. Therefore, taking the features of the I-V curve under healthy condition as a reference, we quantify the difference between the reference and the features from the I-V curve under PS, Rs or Rsh degradation with different fault

severity as shown in Figure IV-25 -Figure IV-27. Here, we use the RMSE as the metric, defined as:

$$RMSE = \sqrt{\frac{\sum_1^n \left(X_i - X_{i_{ref}}\right)^2}{n}} \quad \text{(IV-9)}$$

where, X_i is the i^{th} element in the transformed feature X, n is the number of elements in the feature. X could be the current vector built by the direct I-V method or the matrix built by RP or GADF method. Since the values of the three types of features are different, in order to fairly compare their RMSE and the variation trend with fault severity, it is the normalized value (normalized in the range [0,1] [99]) that is presented in Figure IV-25 -Figure IV-27.

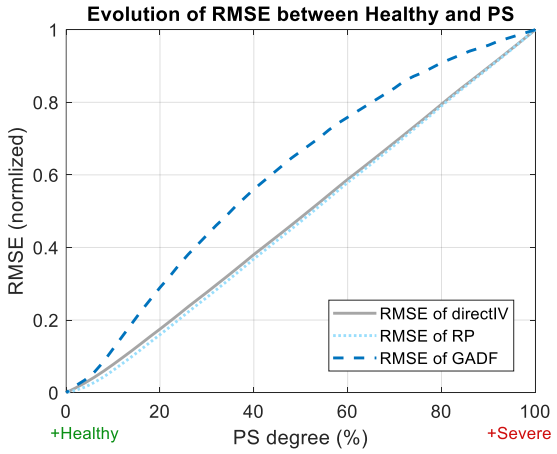


Figure IV-25: RMSE of features between healthy and PS condition

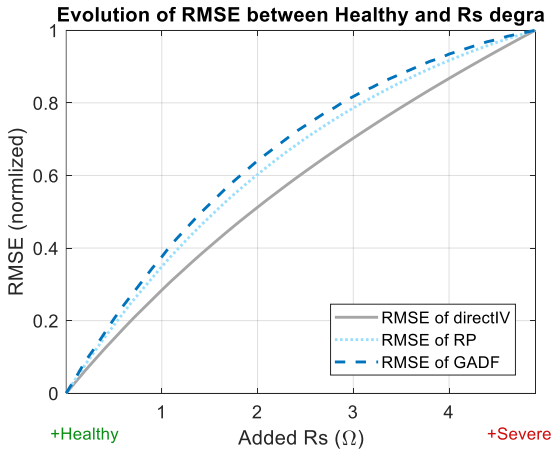


Figure IV-26: RMSE of features between healthy and Rs degradation condition

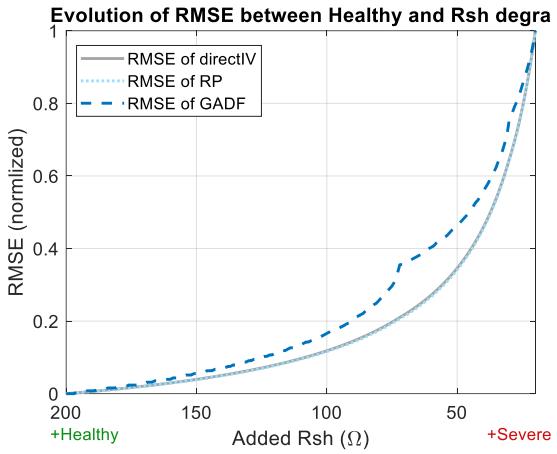


Figure IV-27: RMSE of features between healthy and Rsh degradation condition

From Figure IV-25 -Figure IV-27, it is observed that the normalized RSME of direct I-V and RP exhibit a similar evolution trend under the three conditions, while that of GADF increases clearly faster than RP and direct I-V, especially under PS and Rsh degradation. This means, with the aggravation of fault severity, the difference of GADF feature between the current condition and the healthy condition becomes more significant than that of direct I-V and RP. This allows the classifiers easier to distinguish the current condition from the healthy one, especially under incipient fault conditions. In other words, the discriminability of features under different conditions is enforced with the feature transformation, especially GADF.

IV.5.3. Diagnosis results using experimental data

IV.5.3.1. Description of experimental platform

To evaluate the trained classifiers under real case, an experimental PV array of 6 sc-Si modules (Wiltec 62391-50W) is constructed, as shown in Figure IV-28. The parameters of PV modules in the simulation and field tests are identical. A reference cell (RG100) measures G, and a Pt100 probe (class A) measures the back-sheet T_m of one module continuously. 2 multimeters (Keithley 2440 and 175) record the measured G and T_m. The I-V tracer (Chauvin-Arnoux FTV200) records the array I-V curve.

All the 8 PV array conditions are produced to record the corresponding I-V curves. Some fault setups are shown in Figure IV-29. The setting of fault severity is detailed in Table IV-7 on covering the range used for training presented in Table IV-3. For PS, the module is not fully covered by shelters but it could have same impact as in the simulation. In total, 120 I-V curves are recorded for the test with each condition of 15 curves. Some examples of the measured I-V curves under these conditions are shown in Figure IV-30.

To quantify the matching degree between simulated and real I-V curves, the E_{curve} (defined in (III-17) and used in Chapter III) is used. The average E_{curve} between the measured 120 curves and simulated ones is 1.9%, which shows a good match between the simulation model and the real array.

Figure IV-28: Field test setup

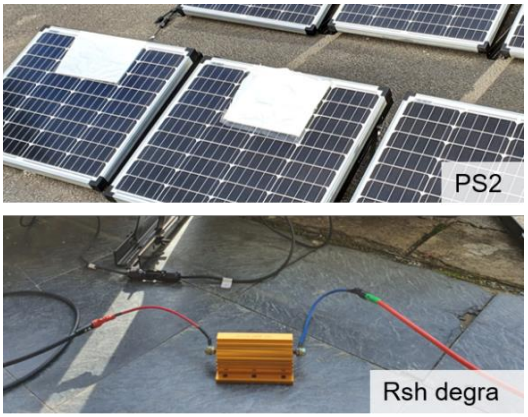


Figure IV-29: Examples of fault setup

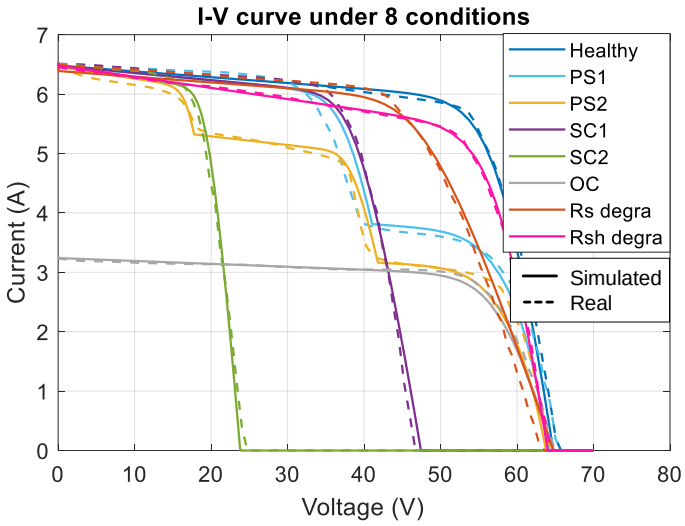


Figure IV-30: Examples of corrected real and simulated I-V curves

Table IV-7: Setting of fault severity

Condition	Varying fault severity
Healthy, SC, SC2, OC	None
PS, PS2	PS degree controlled by 3 different distance between the shelter and panels
Rs degradation	Added resistance in series: 1Ω, 2Ω, 4Ω
Rsh degradation	Added resistance in parallel: 30Ω, 50Ω, 150Ω

IV.5.3.2. Experimental test result

After the pre-processing (correction, resampling) and feature extraction (using the 4 feature extraction selected methods: 8paras, direct I-V, RP, or GADF), the processed samples are then fed into the best-tuned classifiers (trained by simulated data, presented in Table IV-5) for classification. The diagnosis results are presented in Figure IV-31.

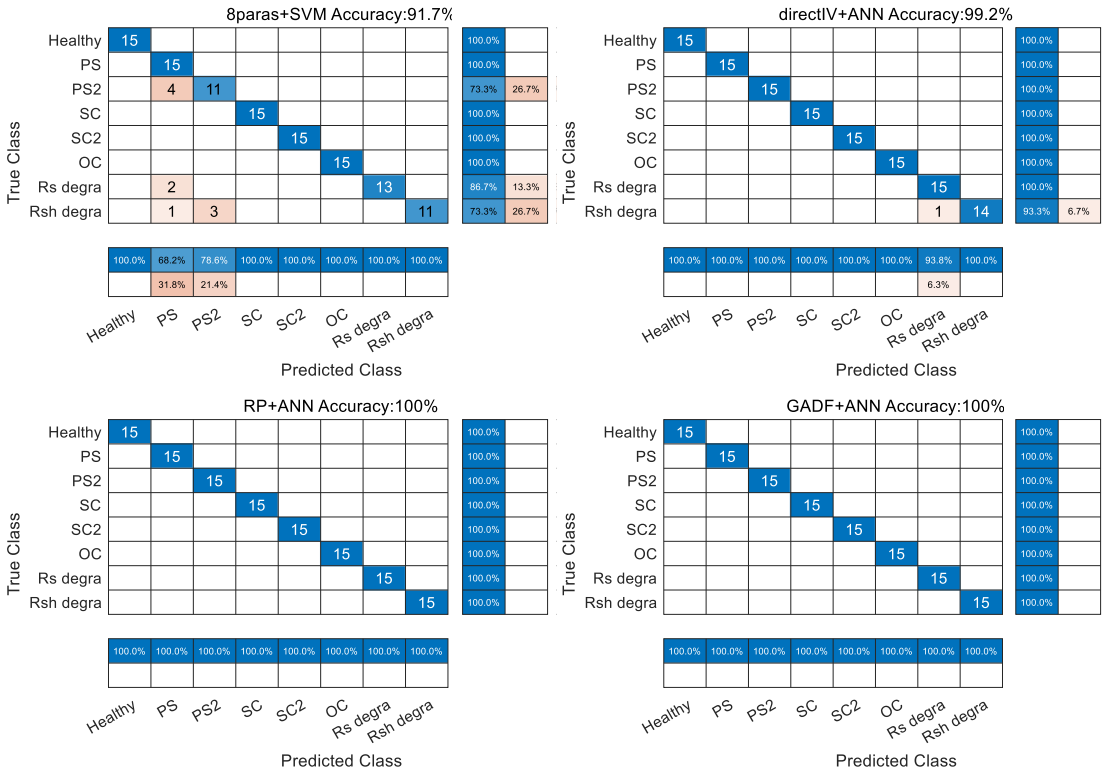


Figure IV-31: Test results using field-measured data

From the results, similar to the previous observation in simulation, the classifiers using the features based on complete I-V curve in direct or transformed feature space (direct I-V, RP, and GADF) outperform that using partial information (8 paras). For 8 paras, the main misclassification happens between PS/PS2 and Rs/Rsh degradation, which corresponds to the result in Figure IV-12. For the 3 types of features based on the complete I-V curve, except one I-V curve wrongly classified when using direct I-V, the classification performance could reach 100% (when using GADF and RP). This experimental test validates the effectiveness of the tuned classifiers using simulated data and also once demonstrates the superiority of using features based on complete I-V curve information for PV FDD.

IV.6. Comparative studies based on the same benchmark

From one I-V curve, there are other different methods, either proposed in the literature, or possible to be adopted, to extract features for PV FDD. Based on the established simulated and experimental database adopted in Section IV.5, these methods will also be evaluated to perform a comparative study.

IV.6.1. Methods for comparison

On the whole, the common methods are categorized into partial usage and complete usage of I-V curve information as illustrated in Figure IV-32. These methods [19]–[21], [24], [82], [234], [265] all rely on I-V curves to extract features for PV FDD. Detailed presentation of each method is given in the following.

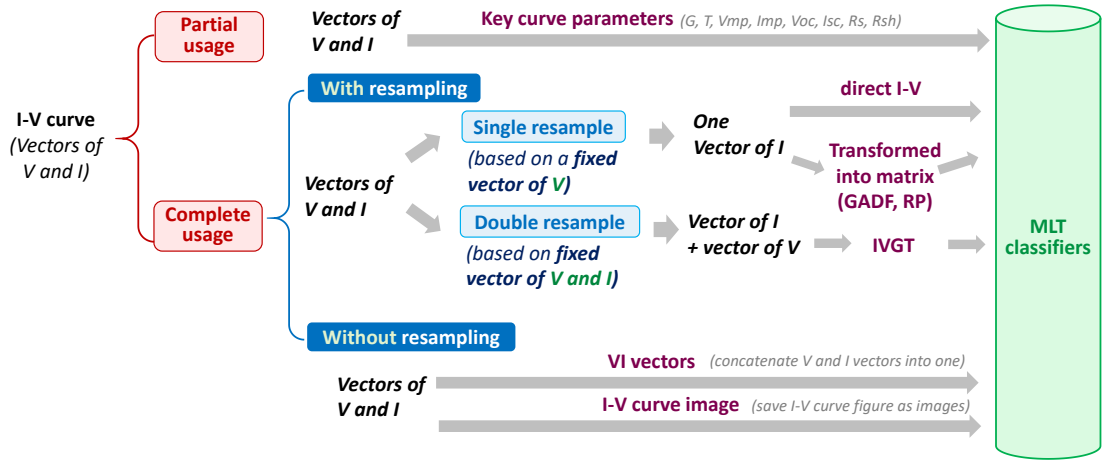


Figure IV-32: Feature extraction methods of I-V curve for comparison study

IV.6.1.1. Methods based on partial usage of I-V curves

The methods of partial usage of I-V curve mainly refers to the ones that adopt key I-V curve parameters as features for PV FDD [19]–[21], [234], [265]. The '8paras' evaluated in Section IV.5 is an example. For comparative study, other common combinations of key curve parameters for PV FDD will also be tested as listed in Table IV-8.

Table IV-8: Parameters used in 4 methods based on partial usage of I-V curve

Technique name	Correction of I-V curve	Parameter used							
		G	T_m	V_{OC}	I_{SC}	V_{MPP}	I_{MPP}	R_s*	R_{sh}*
8paras	No	✓	✓	✓	✓	✓	✓	✓	✓
6parasA	Yes			✓	✓	✓	✓	✓	✓
6parasB	No	✓	✓	✓	✓	✓	✓		
4paras	Yes			✓	✓	✓	✓		

*R_s and R_{sh} are calculated from the slope of I-V curve [252]

For 6parasA and 4paras, the parameters are extracted from the I-V curve after the correction to STC, in which the G and T_m information is already involved. Thus, these two techniques do not require to include G and T_m in the features. For 8paras and 6parasA, the estimated R_s and R_{sh} of I-V curves are also taken as the features for FDD.

IV.6.1.2. Methods based on full usage of I-V curves

IV.6.1.2.1. Methods with resampling of I-V curves

- **Single resampling**

The proposed PV FDD methodology in this thesis contains the step of resampling. Specifically, it is of the single resampling type, i.e., based on a fixed vector of voltage to construct a resampled vector of current as presented in Section IV.3.2.

- **Double resampling**

In literature, there is one double resampling of I-V curves for PV FDD proposed in [24]. Firstly, 20 data points are resampled from one I-V curve based on a voltage vector with the range of [0, V_{OC}], while other 20 points are resampled along the current axis with a current vector [0, I_{SC}]. Then, the 40 new resampled data points are combined and sorted following descending order of the voltage. To consider the impact of G and T_m, the G and T_m are repeated to construct a vector with an identical length of the I-V curve, respectively. Then, the G vector, T_m vector and I-V curve matrix are combined to finally construct the feature of a 40×4 matrix, as illustrated in Figure IV-33.

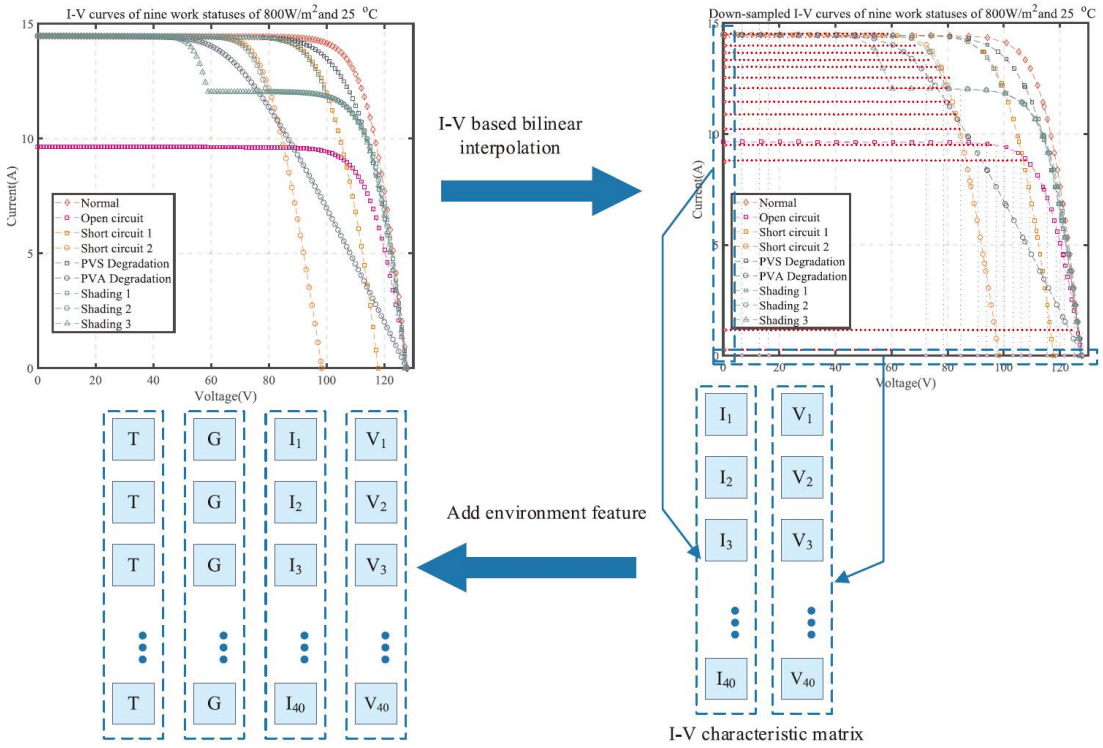


Figure IV-33: Feature extraction based double resampling of I-V curve [24]

IV.6.1.2.2. Methods without resampling of I-V curves

With a full usage of the I-V curve, it is also possible to avoid performing resampling on keeping all the raw I-V curve information. The typical techniques include:

- **VI vector**

As one I-V curve is formed by a vector of *V* and a vector *I*, this method thus directly concatenates them into one vector as features. However, there is one thing that should be paid attention to. That is, the number of points in vectors should always be identical, whatever the ones used for training by simulation or the ones measured in the field test. In this study, the number of points in *V* and *I* vector is set as 300 (larger than that of the measured I-V tracer (FTV200)). The concatenated vector is thus of 600 points. For the curves not reaching 300 points, the vectors will be padded by 0. An illustration for the construction of VI vector feature is shown in Figure IV-34.

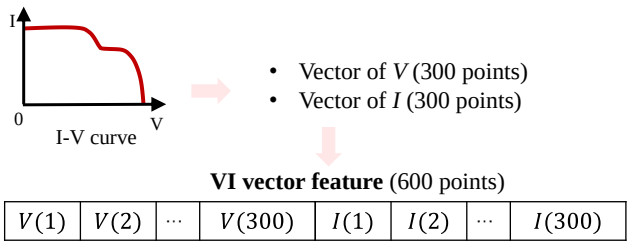


Figure IV-34: Illustration of the VI vector feature

- **I-V curve image**

Another resampling-free usage of the complete I-V curve as the feature is to directly save the I-V curve figures as images. Then, these images will be processed by deep neural networks that are proficient in pattern recognition via images. A similar application can be found in [82]. In this comparison study, I-V curves are saved as images with voltage range in [0, 70] V and current [0, 10] A. The images are square and grayscale. Examples of I-V curve images are given in Figure IV-35.

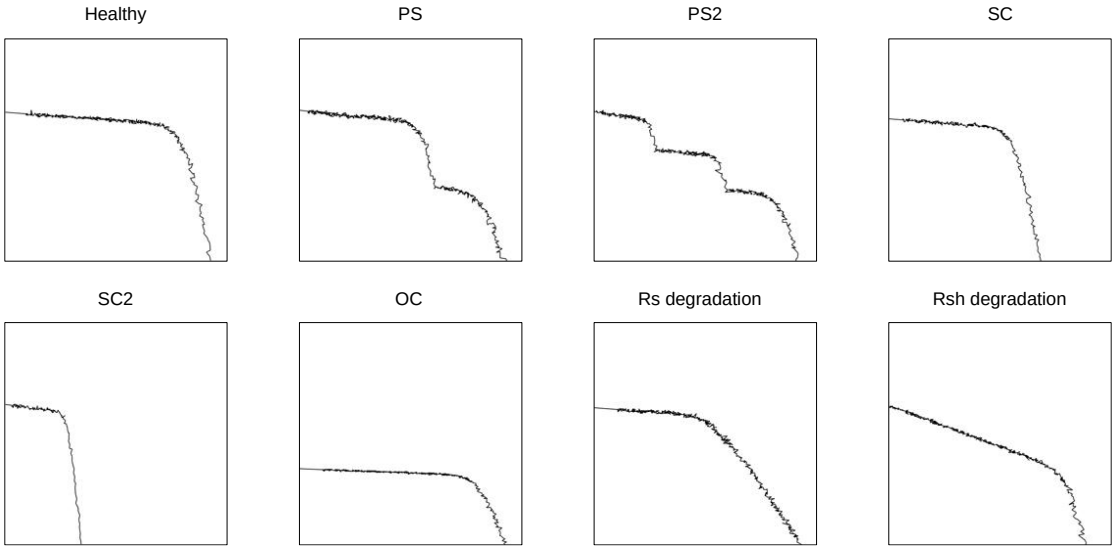


Figure IV-35: Examples of I-V curve images (each image is 100*100 pixels)

IV.6.2. Comparison results

The features extracted by the aforementioned methods will be processed by the MLTs presented in Section IV.5.1 for PV FDD. For each type of new feature, the MLT classifiers will all be re-tuned. Then, the best classifier will be validated by the same field-measured dataset. The accuracy and time are now summarized in Table IV-9.

Table IV-9: Comparison of PV FDD results of different methods

Usage of I-V curve	Resample	Method	Accuracy			Time		
			Train	Test	Field test	Pre-process +train	Test	Field test*
Partial usage	-	8 paras	96.21%	94.83%	91.67%	3.08s	0.05s	0.09s
		6 parasA	95.41%	94.04%	90.83%	3.59s	0.05s	0.05s
		6 parasB	95.08%	93.54%	89.17%	5.86s	0.03s	0.05s
		4 paras	94.59%	93.00%	87.50%	4.16s	0.04s	0.04s
Complete usage	Yes (single)	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
	Yes (double)	IVGT	99.75%	98.83%	97.5%	38.62s	0.01s	2.13s
	No	VI vector	100%	99.96%	15.00%	69.21s	0.01s	0.15s
		I-V image	100%	99.58%	95.00%	1214.84s	13.9s	29.4s

* Time includes pre-processing, extraction and analysis

The methods based on partial usage of the I-V curve present similar performance and require less time in the pre-processing and training steps than the others. Among these partial usage methods, the 8 paras, free of I-V curve correction and containing more curve information than 6 parasB and 4 paras, exhibits higher accuracy in training, test, and field test. Nevertheless, it should be noted that the calculation of R_s and R_{sh} requires the measurement of a full I-V curve, unlike V_{OC} and I_{SC}, which can be more easily obtained by setting the array at OC or SC condition, respectively. Therefore, the choose of methods based on the partial usage of I-V curve will be a trade-off between the FDD performance and complexity of data acquisition.

Concerning the methods based on complete usage of the I-V curve, the proposed 3 methods in this thesis (i.e., direct I-V, RP and GADF) show better classification accuracy than the other methods. Additionally, the decision-making time in the field test is also rapid, all within 0.3s. The IVGT method is also based on resampling but

performs worse than the proposed 3 methods. This is supposedly due to a lack of efficient integration of G and T_m information into the features. As seen from Figure IV-33, the resampled V and I vectors in the IVGT matrix feature are from the I-V curves without correction, the shape of which greatly depends on G and T_m. The G and T_m information are integrated into the features by adding two vectors with repeated values of G and $T_{m,}$ respectively, which inevitably results in information duplication.

For the methods free of resampling, VI vector also achieves quite good performance in training and simulation tests but has an inferior performance in the field test with the classification accuracy of 15%. This is due to the different spacing of points in the I-V curve used in training and field test, for example, under the healthy condition shown in Figure IV-36. As no resampling is performed for VI vector, this different spacing of points leads to the fact that the classifiers trained by simulated I-V curves do not fit the real I-V curves recorded by the I-V tracer. Thus, poor classification result is led to as observed from the field test accuracy in Table IV-9. In fact, the spacing of points of real I-V curves could vary from case to case, depending on the environmental condition, I-V tracer type, I-V characterization technology, etc. Therefore, to ensure the adaptability of one trained classifier, a pre-processing step is needed to bring about an identical spacing of points. That is why the resampling is performed in our PV FDD methodology. As for the I-V image method, its performance is similar with the IVGT method, but still inferior to the proposed methods. Besides, due to the complexity of dealing with image data, the training and field test time are extremely longer than the other methods. This renders this method less competitive for the application in rapid field PV FDD tasks.

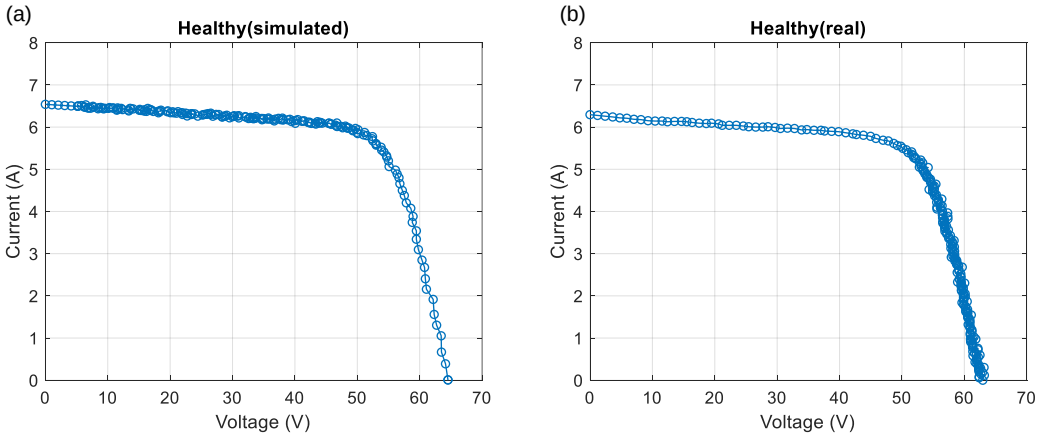


Figure IV-36: Examples of simulated and measured I-V curves of healthy array

To summarize, compared with other I-V curve-based feature extraction methods, the proposed 3 methods (direct I-V, RP and GADF), on the one hand, achieve the best classification performance in both simulation and field test; on the other hand, they do not introduce large time complexity in the pre-processing and training steps, and guarantee a rapid decision making in the field test. Therefore, these feature extraction

methods together with the whole FDD methodology, are promising to be applied for PV FDD.

IV.7. Conclusion

A PV FDD methodology based on complete I-V curve information and MLT classifiers has been presented in this section based on the four-step FDD scheme, i.e., modeling (configuration of I-V curve dataset), pre-processing, feature extraction, and feature analysis. Besides, a comparative study is also performed.

- *Configuration of dataset*: I-V curves of a PV array under healthy and 7 fault conditions are simulated to construct training and test datasets. In addition, an experimental test is also conducted to collect the real I-V curves for the field test.

- *Pre-processing of I-V curves*: two operations are performed, i.e., correction and resampling. The proposed NewP2 correction method is adopted to correct I-V curves to STC. Then, the I-V curve is resampled based on a fixed voltage vector to generate a current vector. The number of points is selected as 50 on considering both the goodness of resampling and the complexity of the feature.

- *Feature extraction*: 3 methods, i.e., direct I-V, RP and GADF, are proposed to embed complete I-V curve information into features. Then, PCA is applied to reduce the dimensionality of extracted features.

- *Feature analysis for FDD*: six common MLT classifiers are all tuned to process the extracted features. The classification results using features based on complete I-V curve outperform that using partial curve information. The best test accuracy (using simulated test data) could reach 100% using GADF as features and ANN as the classifier. The tuned model also achieves 100% classification accuracy via the field test. Besides, GADF exhibits higher robustness to the additional disturbance in the test data than the other features. It is discussed that performing PCA could reduce the total training time and increase the prediction speed. It is also demonstrated that the transformation of features into 2D features, especially into GADF, could augment the discriminability of features, which helps to ameliorate the performance and increase the reliability of classifiers.

- *Comparative study*: the proposed 3 feature extraction methods have been compared with other methods based on partial and complete I-V curve information. The proposed ones achieve the best classification performance in both simulation and field tests on also guaranteeing rapid decision making.

This proposed FDD methodology (including the correction, resampling, feature transformation with GADF) is promising to be applied and further explored for future fault detection and diagnosis of PV modules.

Conclusion and perspectives

Conclusion

Based on a comprehensive state-of-the-art of PV faults and FDD techniques, in this work we have developed a new FDD strategy for PV systems using full I-V curves. This strategy proposes an improved method for correcting the I-V curves of faulty PV panels and uses full I-V curves as features for diagnosis. Besides, the feature transformation is also studied to improve FDD performance.

In Chapter I, a state-of-the-art for PV array faults is presented. The fault causes, impact (on power loss and safety risk), frequency of occurrence, and the causal relationships between mismatch and abnormal degradations have been identified. The main PV fault classification tools reported in the literature have been summarized. We have also proposed a method for classifying the faults according to their level, i.e., at the cell, module, or array level. Through this approach, the usual faults have been presented. Finally, with joint consideration of fault impact, frequency of occurrence, and reproducibility, partial shading, module short-circuit, string open-circuit, and abnormal degradations have been selected as study cases.

In Chapter II, the common FDD techniques for PV faults are reviewed through two categories: visual inspection and automatic analysis. The automatic FDD methods are described in a four-step scheme: modeling, pre-processing, feature extraction, and feature analysis. Our FDD strategy is also determined to follow this four-step scheme. The I-V curve, as containing rich information for assessing the health of the PV modules and arrays, is chosen as the data for analysis. The Machine Learning Techniques, decisive in handling big data and multi-condition classification, are chosen to analyze the fault features. Specific pre-processing and extraction methods are selected to embed more efficiently complete I-V curve information and enhance the discriminability of features.

In Chapter III, the three procedures (P1, P2, P3) proposed in IEC 60891 have been evaluated to correct simulated I-V curves for defective PV arrays. All three procedures introduce significant errors to the estimation of P_m and the fault parameters extracted from the corrected curves due to irradiance G, temperature $T_{m,}$ and the severity of the defects. Moreover, worse still, most significant errors result in underestimating the parameters, up to 9.1% for P_m using the method denoted P2. Therefore, a P2-based improved method (NewP2) has been proposed and has exhibited better average performance than methods P1 and P2 with the decrease of the average curve error from

3.45 % (using P2), 4.5 % (using P1) to 2.37% under constant fault severity. NewP2 also shows an overall better performance with varying fault severity than P1 and P2 on the curve error and most of the key curve and fault parameters. As for P3, it generally has higher performance but is less suitable for application to rapid field diagnosis due to the requirement of multiple reference curves.

In Chapter IV, the proposed PV FDD methodology using full I-V curve information and MLT classifiers is detailed. This methodology is based on the four-step FDD scheme. Correction (using NewP2) and resampling operations are performed to pre-process the original curves. Direct I-V, RP, and GADF, are proposed to transform the complete I-V curve into relevant features. PCA is also applied to reduce the dimensionality of the extracted features. For the analysis of fault features, six common MLT classifiers are tuned and evaluated. The classification results have shown that using the features based on the complete I-V curve outperform those using only partially the information from the I-V curve. The best accuracy could reach 100% using GADF to extract the features and ANN as the classifier for simulation and field data. Besides, GADF exhibits higher robustness to the additional disturbance than the other transformation techniques. It has been noted from the results that performing PCA could increase the speed of the prediction and reduce the total training time. It has also been shown that the transformation of features into 2D features, especially with GADF, could increase the discriminability of the features. Thus, it helps to ameliorate the performance and the reliability of the classifiers. Finally, the proposed methodology has been compared with other methods from the literature. Our proposal achieves the best classification performance for both simulation and field data with rapid decision making.

This proposed FDD strategy (including the correction, resampling, and feature transformation into GADF) is promising to be applied and further explored. This study may also provide valuable experience for future researchers on how to make full use of I-V curves for PV fault diagnosis with higher performance.

Perspectives

- Concerning the correction of I-V curves of defective PV panels:

For the single curve correction methods (P1, P2, and NewP2), the determination of the correction coefficients in the field remains a challenging task. It is necessary to develop a strategy for the determination of these coefficients on the basis of field data, which differs from the procedure proposed by the IEC 60891 standard. The latter requires controlled environmental conditions, which can only be achieved in well-equipped laboratories. Solutions to reduce dependence on correction coefficients, such as [266], [267], should also be further developed and validated with I-V curves of faulty panels.

For the correction method based on multiple curves, two improvements are expected: on the one hand, efforts are needed to improve the correction performance in the case of partial shading; on the other hand, the development of a methodology for the automatic and efficient selection of reference curves would facilitate the deployment of the technique in the field.

Finally, the proposed new correction procedure (NewP2) and the existing procedures still need to be evaluated with more types of PV faults and with field-measured data.

- Concerning the PV FDD strategy based on complete I-V curves:

In terms of modelling, the I-V curve database should be enriched with multiple defect cases and defects of varying severity. It would also be interesting to investigate a combination of physics-based and data-driven models.

In the pre-processing step, the current FDD strategy relies on the correction and resampling of I-V curves. These two operations are relatively complex and may introduce correction errors or information loss. Thus, these operations require further improvement, and alternative methods are expected to be developed with the aim to efficiently embed the environmental information into features and make better use of I-V curves.

Feature transformation techniques (RP and GADF) square the initial dimension of the workspace. This can become a heavy burden for data processing. Especially if the number of modules to be monitored increases as well as the number of I-V curves. Although dimension reduction techniques could be adopted, they increase the complexity of the FDD strategy. Thus, feature transformation techniques should be selected as a trade-off between limited feature-space dimension, and an enhanced capability of discriminating fault features. The combination with analytical models can help to shorten the duration of the calculations and allow faster decision making.

For feature analysis, a deeper analysis is required to better understand the discrepancy of classification performance when using different combinations of features as inputs of the machine learning-based classifiers. Special attention should be paid to the configuration of the hyperparameters. If the amount of available data or the feature dimension increases, the use of deep learning classifiers should be investigated.

Finally, the FDD strategy should combine measurements on the DC side with those on the AC side to monitor not only PV modules but also power converters, sensors, and cables.

Résumé en francais

L'énergie solaire photovoltaïque (PV), en tant qu'énergie renouvelable présente divers avantages [1]. En exploitation elle est non polluante et n'émet aucun gaz à effet de serre (comme le dioxyde de carbone, le méthane et l'oxyde nitreux) [2]. De plus, cette énergie propre est disponible toute l'année. Même par temps nuageux, la production n'est pas interrompue [3]. Ces dernières années, la capacité cumulée installée à l'échelle mondiale a connu une croissance rapide [4], qui a atteint en 2020 plus de 758,9 GW [5].

Pour l'année 2020, la pandémie de COVID-19 a entraîné un ralentissement économique et causé des dommages sociaux. Cependant, cette pandémie n'a pas eu d'impact significatif sur le développement du marché [5]. Selon le BP World Energy & Ember, la part du PV dans la production mondiale d'électricité a augmenté de 20,2% en 2020, s'établissant à 844,4 TWh [6]. Avec cette augmentation, la part du solaire PV dans la production mondiale d'électricité a atteint environ 3,3%. Quant au périmètre futur, selon le scénario « Net Zero Emissions by 2050 » du WEO 2020, la capacité PV installée dans le monde passerait à 3929 GW d'ici 2030 [7], et l'électricité solaire atteindra environ 15,9 % de la production énergétique mondiale [8].

Outre les objectifs environnementaux, la prospérité de l'industrie PV est également due à la baisse progressive des prix des modules photovoltaïques et aux progrès de l'ingénierie et de la construction des installations. En fait, la compétitivité des projets PV mesurée par le coût nivelé de l'électricité s'est également continuellement améliorée ces dernières années [9]. L'AIE a annoncé que « le PV est la source d'électricité la moins chère presque partout » [10].

Avec le développement rapide de l'industrie PV et l'augmentation de la capacité installée, des stratégies d'exploitation et de maintenance efficaces sont de plus en plus nécessaires [11]. En fait, les conditions environnementales, la fabrication, le transport, l'installation ou de mauvaises conditions de fonctionnement peuvent provoquer des pannes de cellules ou de modules PV [12]. Dans le cas d'un générateur PV ou d'une centrale électrique, qui est composé de plusieurs modules connectés électriquement, tout défaut dans une cellule ou un module affecte les performances de l'ensemble du générateur ou du système. Les défauts PV pourraient entraîner un grave danger pour la sécurité , par exemple, risque d'incendie, choc électrique, danger physique [13] ou perte de puissance [14]. Par conséquent, pour assurer la fiabilité, la disponibilité et la sécurité des installations photovoltaïques, leur état de santé doit être surveillé régulièrement pour éviter les pannes et contribuer à une politique de maintenance conditionnelle efficace.

La pierre angulaire d'une stratégie de détection et de diagnostic des défauts PV (FDD) sont les données PV, dont les types électriques sont les plus courants car ils sont capables de refléter l'état de fonctionnement réel d'un module ou d'un générateur PV [15]. Parmi les données électriques, la caractéristique courant-tension (courbe I-V) contient le plus d'information sur l'état de santé d'un module ou d'un générateur PV, ce qui devrait permettre un diagnostic plus précis [16]. Les traceurs I-V courants permettent déjà la mesure pour un seul module ou un string à petite échelle. Ces dernières années, les solutions matérielles (intégrées au niveau de l'onduleur) sont devenues disponibles dans le commerce pour mesurer périodiquement les courbes I-V au niveau de la centrale [17], [18]. Par conséquent, il est prévu dans un proche avenir que les courbes I-V seront mesurables dans presque toutes les installations PV. Dans cette perspective, cette thèse vise à proposer une stratégie basée sur les informations contenues dans la courbe I-V pour le diagnostic des défauts courants dans les panneaux PV. Pour être précis, cette thèse se compose des quatre principaux chapitres suivants:

Dans le Chapitre I, un état de l'art des défauts des générateurs PV est présenté. Les origines des défauts, leur impact (sur la puissance et le risque pour la sécurité), leur fréquence d'occurrence et les relations causales entre l'inadéquation et les dégradations anormales ont été identifiés. Les principaux outils de classification des défauts PV rapportés dans la littérature ont été résumés. Nous avons également proposé une méthode pour classer les défauts en fonction de leur emplacement, c'est-à-dire au niveau de la cellule, du module ou du champ. Enfin, en tenant compte conjointement de l'impact des défauts, de la fréquence d'occurrence et de la reproductibilité, l'ombrage partiel (PS), le court-circuit d'un module (SC), le circuit ouvert d'un string (OC) et les dégradations anormales ont été retenus comme cas d'étude.

Dans le Chapitre II, les techniques courantes de détection et de diagnostic des défauts PV sont analysée à travers deux catégories : l'inspection visuelle et l'analyse automatique. Les méthodes automatiques sont décrites au travers d'un schéma en quatre étapes : modélisation, pré-traitement, extraction de caractéristiques et analyse de caractéristiques. Notre stratégie est également conçue sur la base de ce schéma en quatre étapes. La courbe I-V contient des informations riches pour évaluer la santé des modules et des panneaux PV. Elle est donc ici retenue pour modéliser le système étudié. Suite à l'analyse bibliographique, nous avons noté que les courbes I-V ne sont que très partiellement exploitées pour le diagnostic. De plus, nous avons noté qu'il y avait très peu d'études sur les techniques de transformation des caractéristiques pour améliorer la séparabilité des défauts dans différentes conditions défectueuses. Par conséquent, nous proposons dans ces travaux de développer un pré-traitement adapté et des méthodes d'extraction pour exploiter plus efficacement les informations issues de la courbe I-V complète et améliorer la séparabilité des signatures de défauts. En ce qui concerne l'analyse des signatures, les techniques d'apprentissage automatique (MLT),

décisives dans le traitement du *big data* et de la classification multi-conditions, sont retenues.

Le chapitre III se concentre sur les méthodes de correction des courbes I-V des panneaux PV défectueux, ce qui est une étape essentielle dans le prétraitement. En effet, les courbes I-V mesurées sur le terrain sont enregistrées dans des conditions environnementales variables. Pour éviter de fausses interprétations, il faut ramener les courbes I-V mesurées dans des conditions différentes à la même condition environnementale, généralement la condition STC. Cependant, à notre connaissance, aucune méthode n'est proposée pour la correction des courbes I-V des panneaux PV défectueux. Dans la plupart des travaux, les chercheurs adoptent les procédures de correction originales ou simplifiées de la norme IEC 60891 [25]. Or ces procédures sont toutes initialement conçues pour la correction de courbes mesurées à partir de panneaux PV sains. L'adéquation de ces procédures pour la correction des courbes I-V des panneaux défectueux est rarement étudiée, et encore moins si les caractéristiques du défaut sont faussées par la correction et conduisent ainsi à une erreur de diagnostic [26].

Ainsi, le chapitre III évalue les trois procédures (P1, P2, P3) proposées dans la norme IEC 60891 pour la correction de courbes I-V de panneaux PV défectueux. Puisque notre objectif est d'évaluer l'erreur causée par les procédures de correction, il est essentiel d'éviter les effets d'autres facteurs, comme les incertitudes de mesure. Par conséquent, les courbes à corriger sont générées en simulation pour s'assurer que tous les paramètres utilisés dans la formule de correction n'ont pas de biais. Un système PV, basé sur le modèle à diode unique, est modélisé avec Matlab Simulink® pour générer les courbes I-V à corriger. Plusieurs conditions sont simulées; système sain, ombrage partiel (PS), circuit ouvert (OC) dans un string, module en court-circuit (SC) et dégradation des résistances Rs et Rsh. Les performances de correction sont évaluées à partir de la correction sur l'ensemble de la courbe et des paramètres clés de la courbe. En outre, l'impact de la gravité des défauts sur les performances de correction est également analysé. Les résultats montrent que les trois procédures introduisent des erreurs significatives dues à l'éclairement G, à la température T_m et à la gravité des défauts. Nous avons montré qu'en utilisant P1, une distorsion est introduite dans la forme de la courbe avec une erreur relative jusqu'à 13,8%. On observe également des erreurs d'estimation de la puissance maximale P_m et des paramètres de défaut extraits des courbes corrigées. Les erreurs les plus significatives se traduisent par une sous-estimation des paramètres, jusqu'à 9,1% pour la P_m en utilisant P2. Ces erreurs peuvent affecter la détection des défauts naissants si ces paramètres sont utilisés comme signatures. Quant à P3, elle a généralement des performances plus élevées, mais elle induit une distorsion de correction importante pour les courbes dans le cas de l'ombrage partiel, et elle est moins adaptée à une application au diagnostic de terrain rapide du fait qu'elle nécessite l'usage de plusieurs courbes de référence.

Par conséquent, une nouvelle méthode basée sur la procédure P2 (NewP2) a été proposée. Elle consiste à modifier un coefficient dans l'équation de la correction de tension de P2 pour réduire son erreur de correction de tension à proximité de la tension de circuit ouvert V_{OC}. En même temps, il conserve l'équation d'origine pour la correction actuelle de P2. NewP2 permet d'obtenir en moyenne des performances plus robustes que P1 et P2 avec une diminution de l'erreur de 3,45 % (en utilisant P2), 4,5 % (en utilisant P1) à 2,37 % pour un défaut de sévérité constante. Pour un défaut de gravité variable, NewP2 affiche également une meilleure performance globale que P1 et P2 sur l'erreur de courbe et la plupart des paramètres clés de la courbe.

Dans le Chapitre IV, une méthodologie de détection et de diagnostic qui s'appuie sur l'utilisation complète des courbes I-V et des classifieurs basés sur des techniques d'apprentissage automatique est proposée. Elle est évaluée sur une installation constituée de six modules sc-Si Wiltec 62391-50W organisés en deux strings (présenté dans Figure 1). Le même système est simulé sous Matlab Simulink®. Au total, huit cas sont étudiés: le cas sain, deux types dePS, OC dans un string, deux types de SC, et la dégradation des résistances Rs et Rsh. Deux ensembles de données sont générés, un pour l'apprentissage et un autre pour le test. L'ensemble des données d'apprentissage contient 12000 courbes I-V (1500 courbes pour chaque cas), dont 2400 sont destinés au test. Le bruit environnemental et l'erreur aléatoire de mesure sont également ajoutés aux échantillons simulés pour générer des courbes I-V plus proches de celles mesurées.

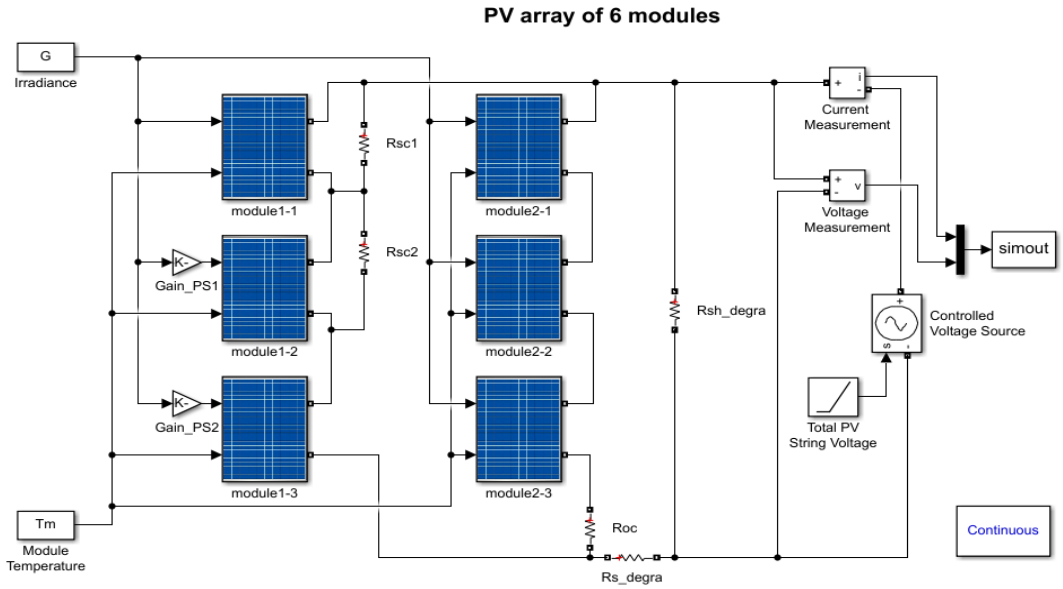


Figure 1: Modèle de simulation pour l'installation PV

La phase de prétraitement consiste à corriger les courbes par la procédure NewP2 puis à les ré-échantillonner. Le nombre de points de ré-échantillonnage est fixé à 50 pour garantir à la fois de bonnes performances et limiter la charge de calcul. Pour l'extraction des caractéristiques, trois méthodes dites, direct I-V, Recurrence Plot (RP) [261] et

Gramian Angular Difference Field (GADF) [262] sont utilisées pour transformer la courbe I-V complete (examples présentés dans Figure 2). Pour réduire la dimension des caractéristiques extraites et améliorer l'efficacité du processus d'analyse, l'Analyse en Composantes Principales (ACP) [124] est appliquée. Le nombre de composantes principales est déterminé en conservant au moins 98 % de la variance cumulée des données originales.

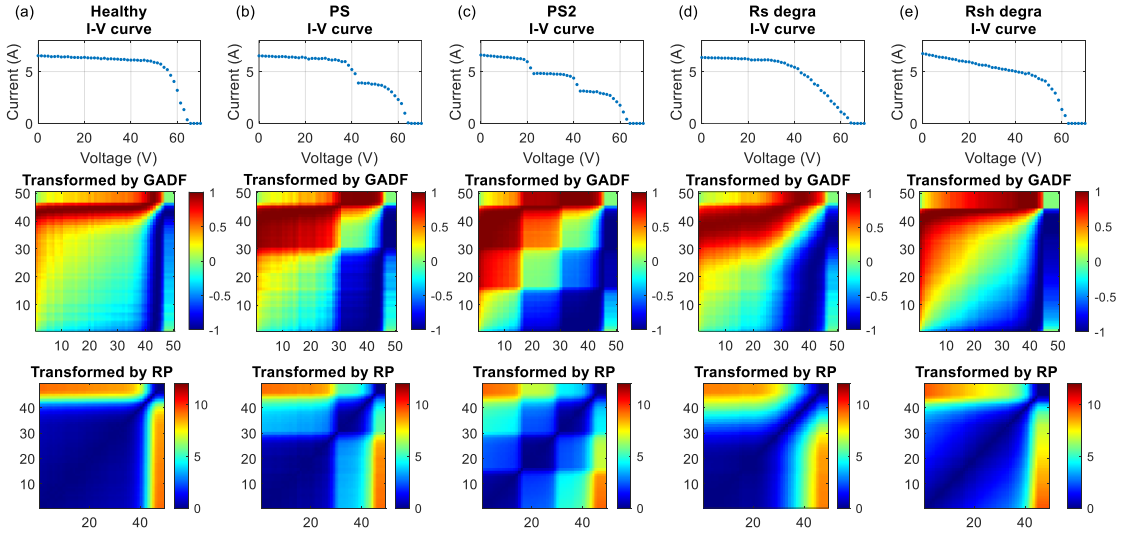


Figure 2 : Examples de caractéristiques transformés par GADF et RP

((a) Cas sain, (b) PS, (c) PS2, (d) Rs degradation, (e) Rsh degradation (la valeur de chaque composant sur la matrice de caractéristique *est marquée par colormap))*

Pour l'analyse des caractéristiques des défauts, six classificateurs MLT courants sont appliqués, y compris le réseau neuronal artificiel (ANN), la machine à vecteurs de support (SVM), l'arbre de décision (DT), la forêt aléatoire (RF), les voisins les plus proches (kNN) et Classificateur naïf bayésien (NBC). Une attention particulière est accordée à la recherche des meilleures performances possibles lors de l'utilisation de différents types de caractéristiques d'entrée en réglant bien les six classificateurs MLT. Les résultats de la classification ont montré que l'utilisation des caractéristiques basées sur la courbe I-V complète surpasse celles qui n'utilisent que partiellement les informations de la courbe I-V(cette méthode notée comme '8paras') avec les résusltats présentées dans Figure 3. La meilleure précision pourrait atteindre 100 % en utilisant GADF pour extraire les caractéristiques et ANN comme classificateur pour la simulation et les données de terrain (la plateforme d'expérimentation présenté dans Figure 4). De plus, GADF présente une plus grande robustesse à la perturbation supplémentaire que les autres techniques de transformation (présenté dans Figure 5).

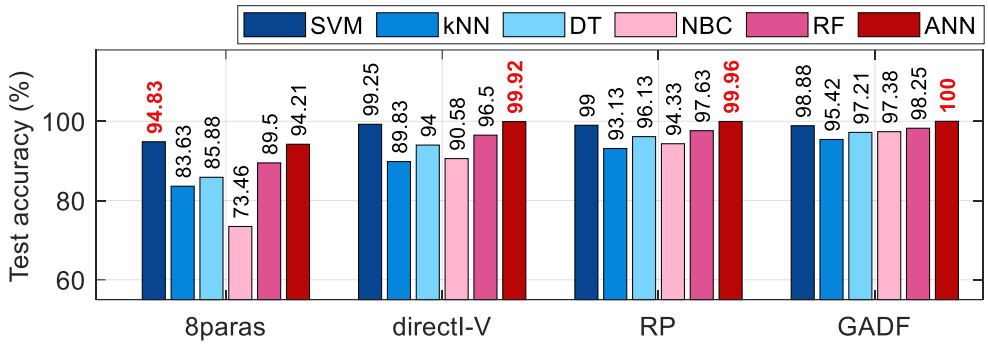


Figure 3: Précision de classification utilisant les courbes I-V simulées pour tous les classificateurs et les caractéristiques

Figure 4: Configuration des tests sur le terrain

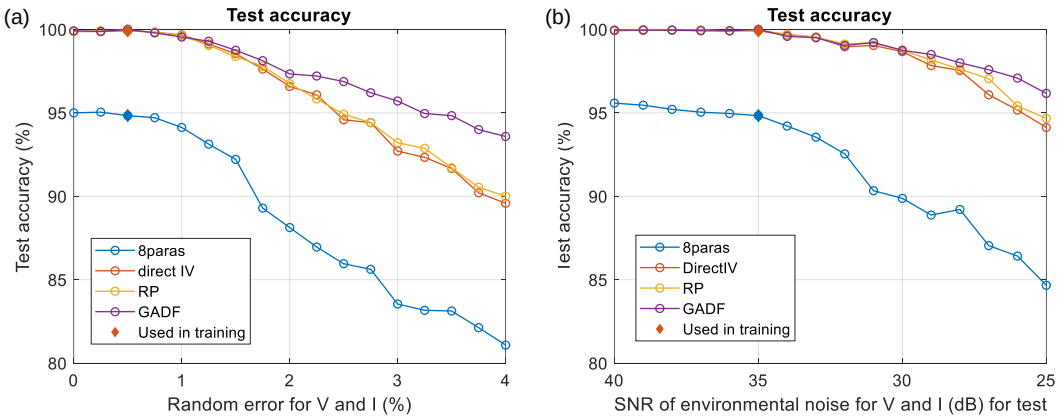


Figure 5: Meilleure précision de classification utilisant les données de test avec un niveau variable de perturbation supplémentaire

((a): Erreur aléatoire variable de V et I ; (b) : Bruit environnemental variable de V et I. Le niveau de perturbation adopté dans les données d'apprentissage est marqué en « losange »)

Il a été aussi noté à partir des résultats que l'exécution de l'ACP pouvait augmenter la vitesse de la prédiction et ainsi réduire la durée de l'apprentissage. Il a également été

montré que la transformation en 2D, notamment avec GADF, pouvait augmenter la discriminabilité des caractéristiques. Par conséquent, cela contribue à améliorer les performances et la fiabilité des classifieurs. Enfin, la méthodologie proposée a été comparée à d'autres méthodes de la littérature. Notre proposition obtient les meilleures performances de classification pour les données de simulation et expérimentales avec une prise de décision rapide.

L'apport de cette thèse réside dans les points suivants :

- Un schéma de détection et de diagnostic de défaut en quatre étapes est d'abord appliqué au domaine du PV. Cette décomposition a permis une meilleure analyse et compréhension des différentes approches présentées dans la littérature en termes de modélisation, de prétraitement, d'extraction et d'analyse de caractéristiques ;

- Une procédure de correction améliorée des courbes I-V de modules en défaut a été proposée et testée. Elle a conduit à de meilleures performances globales que les procédures de correction traditionnelles proposées dans la norme IEC 60891 et basées sur une seule courbe;

- L’exploitation de la courbe I-V complète conduit à de meilleures performances de classification que l'utilisation partielle des courbes I-V ;

- Deux techniques de transformation de caractéristiques sont appliquées aux courbes I-V. Elles ont permis d'améliorer la discriminabilité des caractéristiques, ce qui permet aux classifieurs MLT d'avoir une plus grande robustesse aux perturbations et de meilleures performances de classification aussi bien avec les données de simulation que les mesures.

Quant aux perspectives, elles peuvent être envisagées dans les deux directions suivantes :

1) Concernant la correction des courbes I-V des panneaux PV défectueux :

- La détermination des coefficients de correction sur le terrain reste une tâche difficile. Il est nécessaire de développer une stratégie efficace pour la détermination de ces coefficients à partir des données de terrain.

- Des solutions pour réduire la dépendance aux coefficients de correction devraient également être développées et validées avec des courbes I-V de panneaux défectueux.

- Pour la procédure P3, des efforts sont nécessaires pour améliorer les performances de correction en cas d'ombrage partiel. Le développement d'une méthodologie de sélection automatique et efficace des courbes de référence faciliterait son déploiement sur le terrain.

- La nouvelle procédure de correction proposée (NewP2) et les procédures existantes doivent encore être évaluées avec davantage de types de défauts et avec des données mesurées sur le terrain.

2) Concernant la stratégie PV FDD basée sur des courbes I-V complètes :

- En termes de modélisation, la base de données des courbes I-V devrait être enrichie de plusieurs cas de défauts à sévérité variable. Il serait également intéressant d'étudier une approche hybride qui combinerait les modèles analytiques et les modèles pilotés par les données.

- Dans l'étape de prétraitement, la stratégie FDD actuelle repose sur la correction et le ré-échantillonnage des courbes I-V. Ces deux opérations sont relativement complexes et peuvent introduire des erreurs de correction ou des pertes d'informations. Ainsi, ces opérations doivent encore être améliorées et des méthodes alternatives devraient être développées dans le but d'intégrer efficacement les informations environnementales dans les caractéristiques et de mieux utiliser les courbes I-V.

- Les techniques de transformation des caractéristiques (RP et GADF) conduisent à une augmentation de la dimension de l'espace de travail. Cela peut devenir une lourde charge pour le traitement des données. Ainsi, les techniques de transformation de caractéristiques à retenir doivent être le résultat d’un compromis entre une dimension limitée de l'espace et une capacité améliorée de discrimination des caractéristiques. La combinaison avec des modèles analytiques pourrait raccourcir la durée des calculs et permettre une prise de décision plus rapide.

- Pour l'analyse des caractéristiques, une étude plus approfondie est nécessaire pour mieux comprendre l'écart des performances de classification lors de l'utilisation de différentes combinaisons de caractéristiques comme entrées des classifieurs. Une attention particulière doit être portée à la configuration des hyperparamètres.

- La stratégie FDD pourrait combiner les mesures du côté courant continu avec celles du côté courant alternatif pour surveiller non seulement les modules PV mais aussi les convertisseurs de puissance, les capteurs et les câbles.

La stratégie proposée (y compris la correction, le ré-échantillonnage, la transformation des caractéristiques par GADF) s'est avérée prometteuse et mériterait d'être explorée plus avant. Ces travaux sont une contribution au diagnostic des modules PV en s'appuyant sur l'exploitation des courbes I-V complètes.

Bibliography

[1] P. G. V. Sampaio and M. O. A. González, "Photovoltaic solar energy: Conceptual framework," *Renewable and Sustainable Energy Reviews*, vol. 74. Elsevier Ltd, pp. 590–601, Jul. 01, 2017.

[2] A. Bahadori and C. Nwaoha, "A review on solar energy utilisation in Australia," *Renewable and Sustainable Energy Reviews*, vol. 18, pp. 1–5, Feb. 2013.

[3] J. Ni, L. Zhao, Z. Zhang, Y. Zhang, J. Zhang, S. Deng, and M. Ma, "Dynamic performance investigation of organic Rankine cycle driven by solar energy under cloudy condition," *Energy*, vol. 147, pp. 122–141, Mar. 2018.

[4] SolarPower-Europe, "Global Market Outlook for Solar Power / 2020 - 2024," Industrial Report, 2020. [Online]. Available: https://www.solarpowereurope.org/global-market-outlook-2020-2024/.

[5] K. Izumi, A. Jäger-Waldau, D. Jose, A. Detollenaere, and G. Masson, "IEA PVPS- Snapshot of Global PV Markets - 2020," Technical Report, 2021. [Online]. Available: https://iea-pvps.org/wp-content/uploads/2021/04/IEA_PVPS_Snapshot_2021-V3.pdf.

[6] BP, "Statistical Review of World Energy 2020," Industrial Report. [Online]. Available: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2020-full-report.pdf.

[7] IEA, "World Energy Outlook 2020," Technical Report, 2020. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2020.

[8] A. Jäger-Waldau, "Snapshot of photovoltaics− March 2021," *EPJ Photovoltaics*, vol. 12, p. 2, 2021.

[9] W. Herrmann, G. Eder, B. Farnung, F. Gabi, M. Köntges, B. Kubicek, O. Kunz, and H. Liu, "Report IEA-PVPS T13-24:2021 Qualification of Photovoltaic Power Plants using Mobile Test Equipment," Technical Report, 2021. [Online]. Available: https://iea-pvps.org/research-tasks/performance-operation-and-reliability-of-pho-.

[10] B. Heymi and B. Piotr, "IEA Solar PV tracking report 2020," Technical Report, 2020. [Online]. Available: https://www.iea.org/reports/solar-pv.

[11] G. Di Lorenzo, R. Araneo, M. Mitolo, A. Niccolai, and F. Grimaccia, "Review of O&M Practices in PV Plants: Failures, Solutions, Remote Control, and Monitoring Tools," *IEEE Journal of Photovoltaics*, vol. 10, no. 4, pp. 914–926, Jul. 2020.

[12] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A Comprehensive Review of Catastrophic Faults in PV Arrays: Types, Detection, and Mitigation Techniques," *IEEE Journal of Photovoltaics*, vol. 5, no. 3, pp. 982–997, May 2015.

[13] L. Hernández-Callejo, S. Gallardo-Saavedra, and V. Alonso-Gómez, “A review of photovoltaic systems: Design, operation and maintenance,” *Solar Energy*, vol. 188, pp. 426–440, Aug. 2019.

[14] S. K. Firth, K. J. Lomas, and S. J. Rees, “A simple model of PV system performance and its use in fault detection,” *Solar Energy*, vol. 84, no. 4, pp. 624–635, Apr. 2010.

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

[16] A. Livera, M. Theristis, G. Makrides, and G. E. Georghiou, “Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems,” *Renewable Energy*, vol. 133, pp. 126–143, Apr. 2019.

[17] Huawei, “Smart I-V Curve Diagnosis,” Technical Report, 2020. Accessed: Aug. 17, 2020. [Online]. Available: https://solar.huawei.com/en-GB/download?p=%2F-%2Fmedia%2FSolar%2Fattachment%2Fpdf%2Feu%2Fdatasheet%2FIV-Curve.pdf.

[18] S. Spataru, D. Sera, T. Kerekes, and R. Teodorescu, “Monitoring and Fault Detection in Photovoltaic Systems Based On Inverter Measured String I-V Curves,” in *31st European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC)*, Hamburg, Germany, 14-18 Sept. 2015, pp. 1667–1674.

[19] Z. Chen, L. Wu, S. Cheng, P. Lin, Y. Wu, and W. Lin, “Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics,” *Applied Energy*, vol. 204, pp. 912–931, 2017.

[20] S. Fadhel, C. Delpha, D. Diallo, I. Bahri, A. Migan, M. Trabelsi, and M. F. Mimouni, “PV shading fault detection and classification based on I-V curve using principal component analysis: Application to isolated PV system,” *Solar Energy*, vol. 179, pp. 1–10, 2019.

[21] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. Massi Pavan, “A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks,” *Renewable Energy*, vol. 90, pp. 501–512, May 2016.

[22] M. Miwa, S. Yamanaka, H. Kawamura, H. Ohno, and H. Kawamura, “Diagnosis of a power output lowering of PV ARRAY with a (-dI/dV)-V characteristic,” in *Conference Record of the 2006 IEEE 4th World Conference on Photovoltaic Energy Conversion (WCPEC-4)*, Waikoloa, HI, USA, 7-12 May 2006, vol. 2, pp. 2442–2445.

[23] M. Bressan, Y. El Basri, A. G. Galeano, and C. Alonso, “A shadow fault detection method based on the standard error analysis of I-V curves,” *Renewable Energy*, vol. 99, pp. 1181–1190, Dec. 2016.

[24] Z. Chen, Y. Chen, L. Wu, S. Cheng, and P. Lin, “Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions,” *Energy Conversion and Management*, vol. 198, p. 111793, Oct. 2019.

[25] IEC 60891, “Photovoltaic devices - Procedures for temperature and irradiance corrections to measured I-V characteristics,” International Standard, 2009.

[26] A. Triki-Lahiani, A. Bennani-Ben Abdelghani, and I. Slama-Belkhodja, “Fault detection and monitoring systems for photovoltaic installations: A review,” *Renewable and Sustainable Energy Reviews*, vol. 82, no. March 2017, pp. 2680–2692, 2018.

[27] M. Koentges, S. Kurtz, C. E. Packard, U. Jahn, K. A. Berger, K. Kato, T. Friesen, H. Liu, M. Van Iseghem, and J. Wohlgemuth, “T13-01:2014 Review of failures of photovoltaic modules,” Technical Report, IEA International Energy Agency, 2014. [Online]. Available: http://iea-pvps.org/index.php?id=275&eID=dam_frontend_push&docID=2064.

[28] K. Strohkendl, W. Herrmann, W. Vaassen, J. Althaus, and F. Reil, “The Effect of Transportation Impacts and Dynamic Load Tests on the Mechanical and Electrical Behaviour of Crystalline PV Modules,” in *25th European Photovoltaic Solar Energy Conference and Exhibition / 5th World Conference on Photovoltaic Energy Conversion*, Valencia, Spain, 6-10 Sept. 2010, pp. 3989–3992.

[29] M. Köntges, S. Kajari-Schröder, I. Kunze, and U. Jahn, “Crack Statistic of Crystalline Silicon Photovoltaic Modules,” in *26th European Photovoltaic Solar Energy Conference and Exhibition*, Hamburg, Germany, 5-9 Sept. 2011, pp. 3290–3294.

[30] S. Dietrich, M. Pander, M. Ebert, and J. Bagdahn, “Mechanical Assessment of Large Photovoltaic Modules by Test and Finite Element Analysis,” in *23rd European Photovoltaic Solar Energy Conference and Exhibition*, Valencia, Spain, 1-5 Sept. 2008, pp. 2889–2892.

[31] S. Gallardo-Saavedra, L. Hernández-Callejo, and O. Duque-Pérez, “Quantitative failure rates and modes analysis in photovoltaic plants,” *Energy*, vol. 183, pp. 825–836, Sep. 2019.

[32] K. Abdulmawjood, S. S. Refaat, and W. G. Morsi, “Detection and prediction of faults in photovoltaic arrays: A review,” in *Proceedings - 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering, CPE-POWERENG 2018*, 10-12 April 2018, pp. 1–8.

[33] M. C. Falvo and S. Capparella, “Safety issues in PV systems: Design choices for a secure fault detection and for preventing fire risk,” *Case Studies in Fire Safety*, vol. 3, pp. 1–16, May 2015.

[34] V. Makarskas, M. Jurevičius, J. Zakis, A. Kilikevičius, S. Borodinas, J. Matijošius, and K. Kilikevičienė, “Investigation of the influence of hail mechanical impact parameters on photovoltaic modules,” *Engineering Failure Analysis*, vol. 124, p. 105309, Jun. 2021.

[35] E. Molenbroek, D. W. Waddington, and K. A. Emery, “Hot spot susceptibility and testing of PV modules,” in *The Conference Record of the Twenty-Second IEEE Photovoltaic Specialists Conference - 1991*, Las Vegas, NV, USA, 7-11 Oct. 1991, pp. 547–552.

[36] Dupont, “Dupont photovoltaic solutions,” Industrial report, 2019. [Online]. Available: https://www.dupont.com/solar-photovoltaic-materials.html.

[37] T. Felder, H. Hu, W. Gambogi, K. R. Choudhury, S. MacMaster, L. G. Lles, and T. J. Trout, “Field study and analysis of backsheet degradation in 450MW+ PV installations,” Technical Report, 2017. [Online]. Available: https://www.nist.gov/system/files/documents/2018/01/09/felder-dupont_worldwide_field_inspection_results.pdf.

[38] P. Manganiello, M. Balato, and M. Vitelli, “A Survey on Mismatching and Aging of PV Modules: The Closed Loop,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 11, pp. 7276–7286, Nov. 2015.

[39] D. S. Pillai and N. Rajasekar, “A comprehensive review on protection challenges and fault diagnosis in PV systems,” *Renewable and Sustainable Energy Reviews*, vol. 91, no. July 2017, pp. 18–40, 2018.

[40] “EN 50380 ed. 2 Marking and documentation requirements for Photovoltaic Modules,” European Standard, 2017. [Online]. Available: https://www.en-standard.eu/csn-en-50380-ed.-2-marking-and-documentation-requirements-for-photovoltaic-modules/.

[41] “IEC 60364-7-712:2017 Requirements for special installations or locations - Solar photovoltaic (PV) power supply systems,” International Standard, 2017. [Online]. Available: https://webstore.iec.ch/publication/28213.

[42] “IEC 62548:2016 Photovoltaic (PV) arrays - Design requirements,” International Standard, 2016. [Online]. Available: https://webstore.iec.ch/publication/25949.

[43] M. Sun, S. Lv, X. Zhao, R. Li, W. Zhang, and X. Zhang, “Defect Detection of Photovoltaic Modules Based on Convolutional Neural Network,” in *International Conference on Machine Learning and Intelligent Communications*, Weihai, China, 6-8 July 2018, pp. 122–132.

[44] B. Adothu, P. Bhatt, S. Chattopadhyay, S. Zele, J. Oderkerk, H. P. Sagar, F. R. Costa, and S. Mallick, “Newly developed thermoplastic polyolefin encapsulant–A potential candidate for crystalline silicon photovoltaic modules encapsulation,” *Solar Energy*, vol. 194, pp. 581–588, Dec. 2019.

[45] C. E. Packard, J. H. Wohlgemuth, and S. R. Kurtz, “Development of a visual inspection data collection tool for evaluation of fielded PV module condition,” Technical Report, National Renewable Energy Lab.(NREL), Golden, CO, USA, 2012. [Online]. Available: http://www.osti.gov/servlets/purl/1050110/.

[46] I. Duerr, J. Bierbaum, J. Metzger, J. Richter, and D. Philipp, “Silver Grid Finger Corrosion on Snail Track affected PV Modules - Investigation on Degradation Products and Mechanisms,” *Energy Procedia*, vol. 98, pp. 74–85, 2016.

[47] Y. Sun, S. Chen, L. Xie, R. Hong, and H. Shen, “Investigating the Impact of Shading Effect on the Characteristics of a Large-Scale Grid-Connected PV Power Plant in Northwest China,” *International Journal of Photoenergy*, vol. 2014, 2014.

[48] F. Novoa, “Adhesion and Reliability of Solar Module Materials,” PhD Thesis, Stanford University, 2015.

[49] M. Chang, C. Chen, C. H. Hsueh, W. J. Hsieh, E. Yen, K. L. Ho, H. P. Chuang, C. Y. Lee, and H. Chen, “The reliability investigation of PV junction box based

on 1GW worldwide field database," in *2015 IEEE 42nd Photovoltaic Specialist Conference, PVSC 2015*, New Orleans, LA, USA, 14-19 June 2015.

[50] A. Omazic, G. Oreski, M. Halwachs, G. C. Eder, C. Hirschl, L. Neumaier, G. Pinter, and M. Erceg, "Relation between degradation of polymeric components in crystalline silicon PV module and climatic conditions: A literature review," *Solar Energy Materials and Solar Cells*, vol. 192, pp. 123–133, Apr. 2019.

[51] M. Aghaei, A. Gandelli, F. Grimaccia, S. Leva, and R. E. Zich, "IR real-time analyses for PV system monitoring by digital image processing techniques," in *2015 International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, Krakow, Poland, 17-19 June 2015, pp. 1–6.

[52] V. Engineering, "Solar Arc Fault Troubleshooting," Industrial report, 2019. [Online]. Available: https://www.veloengineering.com/news/solar-arc-fault-troubleshooting.

[53] M. C. López-Escalante, L. J. Caballero, F. Martín, M. Gabás, A. Cuevas, and J. R. Ramos-Barrado, "Polyolefin as PID-resistant encapsulant material in PV modules," *Solar Energy Materials and Solar Cells*, vol. 144, pp. 691–699, Jan. 2016.

[54] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A Comprehensive Review of Catastrophic Faults in PV Arrays: Types, Detection, and Mitigation Techniques," *IEEE Journal of Photovoltaics*, vol. 5, no. 3, pp. 982–997, 2015.

[55] IEC 61730-1, "Photovoltaic (PV) module safety qualification - Part 1: Requirements for construction," International Standard, 2016. [Online]. Available: https://webstore.iec.ch/publication/25674.

[56] M. Köntges, G. Oreski, U. Jahn, M. Herz, P. Hacke, K.-A. Weiss, G. Razongles, M. Paggi, D. Parlevliet, T. Tanahashi, and R. French, "Assessment of photovoltaic module failures in the field," Technique report, 2017. [Online]. Available: http://www.iea-pvps.org/index.php?id=92&eID=dam_frontend_push&docID=4059.

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

[58] J. Harmouche, C. Delpha, D. Diallo, and Y. Le Bihan, "Statistical Approach for Nondestructive Incipient Crack Detection and Characterization Using Kullback-Leibler Divergence," *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1360–1368, Sep. 2016.

[59] D. Stellbogen, "Use of PV circuit simulation for fault detection in PV array fields," in *Conference Record of the IEEE Photovoltaic Specialists Conference*, Louisville, KY, USA, 10-14 May 1993, pp. 1302–1307.

[60] K. H. Chao, S. H. Ho, and M. H. Wang, "Modeling and fault diagnosis of a photovoltaic system," *Electric Power Systems Research*, vol. 78, no. 1, pp. 97–105, Jan. 2008.

[61] N. Barth, R. Jovanovic, S. Ahzi, and M. A. Khaleel, "PV panel single and double diode models: Optimization of the parameters and temperature dependence," *Solar Energy Materials and Solar Cells*, vol. 148, pp. 87–98, Apr. 2016.

[62] J. Merten, J. M. Asensi, C. Voz, A. V. Shah, R. Platz, and J. Andreu, "Improved equivalent circuit and analytical model for amorphous silicon solar cells and modules," *IEEE Transactions on Electron Devices*, vol. 45, no. 2, pp. 423–429, 1998.

[63] D. L. King, J. A. Kratochvil, and W. E. Boyson, "Photovoltaic array performance model," Technical Report, United States. Department of Energy, 2004. [Online]. Available: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online.

[64] D. L. Evans, "Simplified method for predicting photovoltaic array output," *Solar Energy*, vol. 27, no. 6, pp. 555–560, Jan. 1981.

[65] A. R. Jordehi, "Parameter estimation of solar photovoltaic (PV) cells: A review," *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 354–371, Aug. 2016.

[66] J. P. Ram, T. S. Babu, T. Dragicevic, and N. Rajasekar, "A new hybrid bee pollinator flower pollination algorithm for solar PV parameter estimation," *Energy Conversion and Management*, vol. 135, pp. 463–476, Mar. 2017.

[67] F. Masmoudi, F. Ben Salem, and N. Derbel, "Single and double diode models for conventional mono-crystalline solar cell with extraction of internal parameters," in *13th International Multi-Conference on Systems, Signals and Devices*, Leipzig, Germany, 21-24 March 2016, pp. 720–728.

[68] U. Jahn, M. Herz, M. Köntges, D. Parlevliet, M. Paggi, and I. Tsanakas, "Report IEA-PVPS T13-10:2018- Review on Infrared and Electroluminescence Imaging for PV Field Applications," Technical Report, International Energy Agency, 2018. [Online]. Available: https://iea-pvps.org/wp-content/uploads/2020/01/Review_on_IR_and_EL_Imaging_for_PV_Field_Applications_by_Task_13.pdf.

[69] K. Bedrich, M. Bokalic, M. Bliss, M. Topic, T. R. Betts, and R. Gottschalg, "Electroluminescence Imaging of PV Devices: Advanced Vignetting Calibration," *IEEE Journal of Photovoltaics*, vol. 8, no. 5, pp. 1297–1304, Sep. 2018.

[70] K. G. Bedrich, W. Luo, M. Pravettoni, D. Chen, Y. Chen, Z. Wang, P. J. Verlinden, P. Hacke, Z. Feng, J. Chai, Y. Wang, A. G. Aberle, and Y. S. Khoo, "Quantitative Electroluminescence Imaging Analysis for Performance Estimation of PID-Influenced PV Modules," *IEEE Journal of Photovoltaics*, vol. 8, no. 5, pp. 1281–1288, Sep. 2018.

[71] M. Kontges, A. Morlier, G. Eder, E. Fleis, B. Kubicek, and J. Lin, "Review: Ultraviolet Fluorescence as Assessment Tool for Photovoltaic Modules," *IEEE Journal of Photovoltaics*, vol. 10, no. 2, pp. 616–633, Mar. 2020.

[72] A. Morlier, M. Siebert, I. Kunze, G. Mathiak, and M. Kontges, "Detecting Photovoltaic Module Failures in the Field during Daytime with Ultraviolet Fluorescence Module Inspection," *IEEE Journal of Photovoltaics*, vol. 7, no. 6, pp. 1710–1716, Nov. 2017.

[73] J. de Hoog, S. Maetschke, P. Ilfrich, and R. R. Kolluri, "Using satellite and aerial imagery for identification of solar pv: State of the art and research

opportunities," in *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, Virtual event, 22-26 June 2020, pp. 308–313.

[74] G. Friesen, W. Herrmann, G. Belluardo, and B. Herteleer, "Report IEA-PVPS T13-11:2018 Photovoltaic module energy yield measurements: Existing approaches and best practice," Technical Report, 2018. [Online]. Available: https://iea-pvps.org/wp-content/uploads/2020/01/Photovoltaic_Module_Energy_Yield_Measurements_Existing_Approaches_and_Best_Practice_by_Task_13.pdf.

[75] T. Stoffel, P. Gotseff, and M. Sengupta, "Evaluation of Photodiode and Thermopile Pyranometers for Photovoltaic Applications," *27th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 3705–3708, Oct. 2012.

[76] L. Dunn, M. Gostein, and K. Emery, "Comparison of pyranometers vs. PV reference cells for evaluation of PV array performance," in *2012 38th IEEE Photovoltaic Specialists Conference*, Austin, TX, USA, 3-8 June 2012, pp. 2899–2904.

[77] "IEC 60904-5:2011 Photovoltaic devices - Part 5: Determination of the equivalent cell temperature (ECT) of photovoltaic (PV) devices by the open-circuit voltage method," International Standard, 2011.

[78] "IEC TS 62446-3:2017 Photovoltaic (PV) systems - Requirements for testing, documentation and maintenance - Part 3: Photovoltaic modules and plants - Outdoor infrared thermography," International Standard, 2017.

[79] M. Lave, W. Hayes, A. Pohl, and C. W. Hansen, "Evaluation of global horizontal irradiance to plane-of-array irradiance models at locations across the United States," *IEEE Journal of Photovoltaics*, vol. 5, no. 2, pp. 597–606, Mar. 2015.

[80] B. Marion, "A model for deriving the direct normal and diffuse horizontal irradiance from the global tilted irradiance," *Solar Energy*, vol. 122, pp. 1037–1046, Dec. 2015.

[81] D. Dirnberger, G. Blackburn, B. Müller, and C. Reise, "On the impact of solar spectral irradiance on the yield of different PV technologies," *Solar Energy Materials and Solar Cells*, vol. 132, pp. 431–442, Jan. 2015.

[82] X. Lu, P. Lin, S. Cheng, Y. Lin, Z. Chen, L. Wu, and Q. Zheng, "Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph," *Energy Conversion and Management*, vol. 196, pp. 950–965, Sep. 2019.

[83] S. G. K. Patro and K. K. Sahu, "Normalization: A Preprocessing Stage," *IARJSET*, pp. 20–22, Mar. 2015.

[84] A. M. Karimi, J. S. Fada, N. A. Parrilla, B. G. Pierce, M. Koyuturk, R. H. French, and J. L. Braid, "Generalized and Mechanistic PV Module Performance Prediction from Computer Vision and Machine Learning on Electroluminescence Images," *IEEE Journal of Photovoltaics*, vol. 10, no. 3, pp. 878–887, May 2020.

[85] M. Aghaei, S. Leva, and F. Grimaccia, "PV power plant inspection by image mosaicing techniques for IR real-time images," in *Conference Record of the*

IEEE Photovoltaic Specialists Conference, Portland, OR, USA, 5-10 June 2016, vol. 2016-Novem, pp. 3100–3105.

[86] Y. Meng, Z. Zhang, H. Yin, and T. Ma, "Automatic detection of particle size distribution by image analysis based on local adaptive canny edge detection and modified circular Hough transform," *Micron*, vol. 106, pp. 34–41, Mar. 2018.

[87] "IEC 61724-1:2017 Photovoltaic system performance - Part 1: Monitoring," International Standard, 2017.

[88] A. Chikh and A. Chandra, "An Optimal Maximum Power Point Tracking Algorithm for PV Systems with Climatic Parameters Estimation," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 644–652, Apr. 2015.

[89] W. Juxing, W. Ziyu, Q. Junfeng, Z. Qi, and S. Xu, "A Novel Multichannel Online Denoising Method for the Three Phase Electrical Signals of PV System," in *Proceedings of the 32nd Chinese Control and Decision Conference, CCDC 2020*, Hefei, China, 22-24 Aug. 2020, pp. 5021–5026.

[90] L. Fan, F. Zhang, H. Fan, and C. Zhang, "Brief review of image denoising techniques," *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 1, p. 7, Dec. 2019.

[91] J. Wang and L. Perez, "The effectiveness of data augmentation in image classification using deep learning," *Convolutional Neural Networks Vis. Recognit*, vol. 11, 2017.

[92] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[93] A. M. Karimi, J. S. Fada, M. A. Hossain, S. Yang, T. J. Peshek, J. L. Braid, and R. H. French, "Automated Pipeline for Photovoltaic Module Electroluminescence Image Processing and Degradation Feature Classification," *IEEE Journal of Photovoltaics*, vol. 9, no. 5, pp. 1324–1335, 2019.

[94] F. Aziz, A. Ul Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays," *IEEE Access*, vol. 8, pp. 41889–41904, 2020.

[95] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 569–586, Mar. 1992.

[96] S. Lu, T. Sirojan, B. T. Phung, D. Zhang, and E. Ambikairajah, "DA-DCGAN: An Effective Methodology for DC Series Arc Fault Diagnosis in Photovoltaic Systems," *IEEE Access*, vol. 7, pp. 45831–45840, 2019.

[97] W. M. K. Trochim, *Descriptive statistics*, Book. Research Methods of Knowledge, 2006.

[98] N. O'Rourke, L. Hatcher, and E. J. Stepanski, *A step-by-step approach to using SAS for univariate & multivariate statistics*, Book. SAS institute, 2005.

[99] D. Freedman, R. Pisani, R. Purves, and A. Adhikari, *Statistics*, Book. WW Norton & Company, New York, 2007.

[100] H. A. David, “Early sample measures of variability,” *Statistical Science*, pp. 368–377, 1998.

[101] H. Wang, J. Zhao, Q. Sun, and H. Zhu, “Probability modeling for PV array output interval and its application in fault diagnosis,” *Energy*, vol. 189, p. 116248, Dec. 2019.

[102] L. L. Scharf and C. Demeure, *Statistical signal processing: detection, estimation, and time series analysis*, Book. Prentice Hall, 1991.

[103] C. L. Garrido-Alzar, “Algorithm for extraction of solar cell parameters from I-V curve using double exponential model,” *Renewable Energy*, vol. 10, no. 2-3 SPEC. ISS., pp. 125–128, Feb. 1997.

[104] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*, Book, vol. 31999. McGraw-Hill New York, 1986.

[105] H. J. Nussbaumer, “The fast Fourier transform,” in *Fast Fourier Transform and Convolution Algorithms*, Springer, 1981, pp. 80–111.

[106] M. I. Fitrianto, E. Wahjono, D. O. Anggriawan, E. Prasetyono, R. H. Mubarok, and A. Tjahjono, “Identification and Protection of Series DC Arc Fault for Photovoltaic Systems Based on Fast Fourier Transform,” in *IES 2019 - International Electronics Symposium: The Role of Techno-Intelligence in Creating an Open Energy System Towards Energy Democracy, Proceedings*, Surabaya, Indonesia, 27-28 Sept. 2019, pp. 159–163.

[107] C. E. Heil and D. F. Walnut, “Continuous and Discrete Wavelet Transforms,” *SIAM Review*, vol. 31, no. 4, pp. 628–666, Dec. 1989.

[108] A. Graps, “An introduction to wavelets,” *IEEE computational science and engineering*, vol. 2, no. 2, pp. 50–61, 1995.

[109] Z. Wang and R. S. Balog, “Arc fault and flash detection in photovoltaic systems using wavelet transform and support vector machines,” in *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, 2016, vol. 2016-Novem, pp. 3275–3280.

[110] M. Manohar and E. Koley, “SVM based protection scheme for microgrid,” in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2017*, 2018, vol. 2018-Janua, pp. 429–432.

[111] S. Deitsch, C. Buerhop-Lutz, E. Sovetkin, A. Steland, A. Maier, F. Gallwitz, and C. Riess, “Segmentation of Photovoltaic Module Cells in Electroluminescence Images,” Jun. 2018. [Online]. Available: http://arxiv.org/abs/1806.06530.

[112] L. Bokkers, L. Ambrogioni, and U. Güçlü, “Segmentation of photovoltaic panels in aerial photography using group equivariant FCNs,” in *CEUR Workshop Proceedings*, 2019, vol. 2491.

[113] R. M. Haralick, I. Dinstein, and K. Shanmugam, “Textural Features for Image Classification,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[114] W. A. Hanafy, A. Pina, and S. A. Salem, “Machine Learning Approach for Photovoltaic Panels Cleanliness Detection,” in *2019 15th International*

Computer Engineering Conference (ICENCO), Giza, Egypt, 29-30 Dec. 2019, pp. 72–77.

[115] V. S. B. Kurukuru, A. Haque, M. A. Khan, and A. K. Tripathy, "Fault classification for Photovoltaic Modules Using Thermography and Machine Learning Techniques," in *2019 International Conference on Computer and Information Sciences (ICCIS)*, Aljouf, Saudi Arabia, 10–11 Apr. 2019, pp. 1–6.

[116] C. Mantel, F. Villebro, G. Alves dos Reis Benatto, H. Rajesh Parikh, S. Wendlandt, K. Hossain, P. B. Poulsen, S. Spataru, D. Séra, and S. Forchhammer, "Machine learning prediction of defect types for electroluminescence images of photovoltaic panels," in *Applications of Machine Learning*, 2019, vol. 11139, p. 1.

[117] J. Monson, M. Wirthlin, and B. L. Hutchings, "Optimization techniques for a high level synthesis implementation of the Sobel filter," in *2013 International Conference on Reconfigurable Computing and FPGAs (ReConFig 2013)*, Cancun, Mexico, 9-11 Dec. 2013.

[118] Y. Tan, K. Liao, X. Bai, C. Deng, Z. Zhao, and B. Zhao, "Denoising Convolutional Neural Networks Based Dust Accumulation Status Evaluation of Photovoltaic Panel," in *2019 IEEE International Conference on Energy Internet (ICEI)*, Nanjing, China, 20-24 May 2019, pp. 560–566.

[119] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[120] S. Maldonado and R. Weber, "A wrapper method for feature selection using Support Vector Machines," *Information Sciences*, vol. 179, no. 13, pp. 2208–2217, Jun. 2009.

[121] N. Sánchez-Maroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán, "Filter methods for feature selection - A comparative study," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, vol. 4881 LNCS, pp. 178–187.

[122] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014.

[123] C. Kapucu and M. Cubukcu, "A supervised ensemble learning method for fault diagnosis in photovoltaic strings," *Energy*, vol. 227, p. 120463, Jul. 2021.

[124] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, Jul. 2010.

[125] K. Xia, S. He, Y. Tan, Q. Jiang, J. Xu, and W. Yu, "Wavelet packet and support vector machine analysis of series DC ARC fault detection in photovoltaic system," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 14, no. 2, pp. 192–200, Feb. 2019.

[126] Z. Chen, K. Zhang, Y. A. W. Shardt, S. X. Ding, X. Yang, C. Yang, and T. Peng, "Comparison of Two Basic Statistics for Fault Detection and Process Monitoring," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14776–14781, Jul. 2017.

[127] M. Hajji, M. F. Harkat, A. Kouadri, K. Abodayeh, M. Mansouri, H. Nounou, and M. Nounou, “Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems,” *European Journal of Control*, vol. 59, pp. 313–321, May 2020.

[128] W. R. Klecka, G. R. Iversen, and W. R. Klecka, *Discriminant analysis*, Book, vol. 19. Sage, 1980.

[129] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, [Book]. John Wiley & Sons, 2004.

[130] A. Tharwat, “Linear vs. quadratic discriminant analysis classifier: a tutorial,” *International Journal of Applied Pattern Recognition*, vol. 3, no. 2, p. 145, 2016.

[131] J. H. Friedman, “Regularized discriminant analysis,” *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.

[132] S. Fadhel, D. Diallo, C. Delpha, A. Migan, I. Bahri, M. Trabelsi, and M. F. Mimouni, “Maximum power point analysis for partial shading detection and identification in photovoltaic systems,” *Energy Conversion and Management*, vol. 224, p. 113374, Nov. 2020.

[133] P. Comon, “Independent component analysis, A new concept?,” *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[134] F. A. Qureshi, Z. Uddin, M. B. Satti, and M. Ali, “ICA-based solar photovoltaic fault diagnosis,” *International Transactions on Electrical Energy Systems*, vol. 30, no. 8, p. e12456, Aug. 2020.

[135] G. Hinton and S. T. Roweis, “Stochastic neighbor embedding,” in *NIPS*, 2002, vol. 15, pp. 833–840.

[136] Z. He, X. Zhang, C. Liu, and T. Han, “Fault Prognostics for Photovoltaic Inverter Based on Fast Clustering Algorithm and Gaussian Mixture Model,” *Energies*, vol. 13, no. 18, p. 4901, Sep. 2020.

[137] W. Gao and R. J. Wai, “A Novel Fault Identification Method for Photovoltaic Array via Convolutional Neural Network and Residual Gated Recurrent Unit,” *IEEE Access*, vol. 8, pp. 159493–159510, 2020.

[138] W. Wang, A. C. F. Liu, H. S. H. Chung, R. W. H. Lau, J. Zhang, and A. W. L. Lo, “Fault Diagnosis of Photovoltaic Panels Using Dynamic Current-Voltage Characteristics,” *IEEE Transactions on Power Electronics*, vol. 31, no. 2, pp. 1588–1599, Feb. 2016.

[139] D. Gianola, “Theory and analysis of threshold characters,” *Journal of animal Science*, vol. 54, no. 5, pp. 1079–1096, 1982.

[140] M. Hosseinzadeh and F. Rajaei Salmasi, “Determination of maximum solar power under shading and converter faults - A prerequisite for failure-tolerant power management systems,” *Simulation Modelling Practice and Theory*, vol. 62, pp. 14–30, Mar. 2016.

[141] A. Umana and A. P. S. Meliopoulos, “Detection of cell-level fault conditions within a photovoltaic array system,” in *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, Dallas, TX, USA, 3-5 May 2016.

[142] T. Shimakage, K. Nishioka, H. Yamane, M. Nagura, and M. Kudo, "Development of fault detection system in PV system," in *INTELEC, International Telecommunications Energy Conference (Proceedings)*, Amsterdam, Netherlands, 9-13 Oct. 2011.

[143] M. Alajmi and I. Abdel-Qader, "Fault detection and localization in solar photovoltaic arrays using the current-voltage sensing framework," in *IEEE International Conference on Electro Information Technology*, Grand Forks, ND, USA, 19-21 May 2016, vol. 2016-Augus, pp. 307–312.

[144] W. J. Dixon and F. J. Massey Jr, *Introduction to statistical analysis*, Book. McGraw-Hill, 1951.

[145] N. R. Draper and H. Smith, *Applied regression analysis*, Book, vol. 326. John Wiley & Sons, 1998.

[146] S. Spataru, D. Sera, T. Kerekes, and R. Teodorescu, "Photovoltaic array condition monitoring based on online regression of performance model," in *Conference Record of the IEEE Photovoltaic Specialists Conference*, Tampa, FL, USA, 16-21 June 2013, pp. 815–820.

[147] R. Fazai, K. Abodayeh, M. Mansouri, M. Trabelsi, H. Nounou, M. Nounou, and G. E. Georghiou, "Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems," *Solar Energy*, vol. 190, pp. 405–413, Sep. 2019.

[148] T. Li, S. Zhu, and M. Ogihara, "Using discriminant analysis for multi-class classification: An experimental investigation," *Knowledge and Information Systems*, vol. 10, no. 4, pp. 453–472, Nov. 2006.

[149] C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, May 2017.

[150] H. B. Barlow, "Unsupervised learning," *Neural computation*, vol. 1, no. 3, pp. 295–311, 1989.

[151] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," in *ACM Computing Surveys*, 1999, vol. 31, no. 3, pp. 264–323.

[152] K. Krishna and M. N. Murty, "Genetic K-means algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 3, pp. 433–439, 1999.

[153] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN," *ACM Transactions on Database Systems*, vol. 42, no. 3, pp. 1–21, Jul. 2017.

[154] M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. Sander, "OPTICS: Ordering Points to Identify the Clustering Structure," *SIGMOD Record (ACM Special Interest Group on Management of Data)*, vol. 28, no. 2, pp. 49–60, Jun. 1999.

[155] L. Rokach and O. Maimon, *Data Mining and Knowledge Discovery Handbook*, [Book]. Springer-Verlag, 2006.

[156] G. Liu, L. Zhu, X. Wu, and J. Wang, "Time series clustering and physical implication for photovoltaic array systems with unknown working conditions," *Solar Energy*, vol. 180, pp. 401–411, Mar. 2019.

[157] S. Wang, F. Gao, J. Wu, C. Zheng, X. Fu, and F. Duan, “Online Clustering based Fault Data Detection Method for Distributed PV Sites,” in *Chinese Control Conference, CCC*, Shenyang, China, 27-29 July 2020, pp. 4341–4346.

[158] S. Park, S. Park, M. Kim, and E. Hwang, “Clustering-Based Self-Imputation of Unlabeled Fault Data in a Fleet of Photovoltaic Generation Systems,” *Energies*, vol. 13, no. 3, p. 737, Feb. 2020.

[159] P. Lin, Y. Lin, Z. Chen, L. Wu, L. Chen, and S. Cheng, “A Density Peak-Based Clustering Approach for Fault Diagnosis of Photovoltaic Arrays,” *International Journal of Photoenergy*, vol. 2017, 2017.

[160] J. E. van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” *Machine Learning*, vol. 109, no. 2, pp. 373–440, Feb. 2020.

[161] T. Washio and H. Motoda, “State of the art of graph-based data mining,” *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 1, pp. 59–68, Jul. 2003.

[162] A. Demiriz, K. P. Bennett, and M. J. Embrechts, “Semi-supervised clustering using genetic algorithms,” *Artificial neural networks in engineering (ANNIE-99)*, pp. 809–814, 1999.

[163] C. Li, J. Zhu, and B. Zhang, “Max-Margin Deep Generative Models for (Semi-)Supervised Learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 11, pp. 2762–2775, Nov. 2018.

[164] Y. Zhao, L. Yang, B. Lehman, J. F. De Palma, J. Mosesian, and R. Lyons, “Decision tree-based fault detection and classification in solar photovoltaic arrays,” in *Conference Proceedings - IEEE Applied Power Electronics Conference and Exposition - APEC*, Orlando, FL, USA, 5-9 Feb. 2012, pp. 93–99.

[165] R. Benkercha and S. Moulahoum, “Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system,” *Solar Energy*, vol. 173, pp. 610–634, Oct. 2018.

[166] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*, Book. MIT press, 2018.

[167] A. H. Elsheikh, S. W. Sharshir, M. Abd Elaziz, A. E. Kabeel, W. Guilan, and Z. Haiou, “Modeling of solar energy systems using artificial neural network: A comprehensive review,” *Solar Energy*, vol. 180, pp. 622–639, Mar. 2019.

[168] D. Nelson and J. Wang, “Introduction to artificial neural systems,” *Neurocomputing*, vol. 4, no. 6, pp. 328–330, Dec. 1992.

[169] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[170] J. Park and I. W. Sandberg, “Universal Approximation Using Radial-Basis-Function Networks,” *Neural Computation*, vol. 3, no. 2, pp. 246–257, Jun. 1991.

[171] D. F. Specht, “Probabilistic neural networks,” *Neural Networks*, vol. 3, no. 1, pp. 109–118, Jan. 1990.

[172] M. H. Wang and C. P. Hung, “Extension neural network and its applications,” *Neural Networks*, vol. 16, no. 5–6, pp. 779–784, Jun. 2003.

[173] G. Bin Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.

[174] C. K. Khelil, K. Kara, and A. Chouder, “Fault detection of the photovoltaic system by artificial neural networks,” in *4th International Conference on Green Energy and Environmental Engineering (GEEE-2017)*, Sousse, Tunisia, 22-24 Apr. 2017.

[175] B. Basnet, H. Chun, and J. Bang, “An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems,” *Journal of Sensors*, vol. 2020, pp. 1–11, Jun. 2020.

[176] X. Li, P. Yang, J. Ni, and J. Zhao, “Fault diagnostic method for PV array based on improved wavelet neural network algorithm,” in *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, Shenyang, China, 2-5 Mar. 2015, pp. 1171–1175.

[177] N. Mathias, F. Shaikh, C. Thakur, S. Shetty, P. Dumane, and D. S. Chavan, “Detection of Micro-Cracks in Electroluminescence Images of Photovoltaic Modules,” *SSRN Electronic Journal*, Apr. 2020.

[178] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*, Antalya, Turkey, 21-23 Aug. 2018, pp. 1–6.

[179] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” Nov. 2015, [Online]. Available: http://arxiv.org/abs/1511.08458.

[180] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[181] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 8-10 Jun. 2015, pp. 1–9.

[182] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *International Conference on Learning Representations*, San Diego, CA, USA, 7-9 May 2015.

[183] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, USA, 24-27 June 2014, pp. 580–587.

[184] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27-30 Jun. 2016, pp. 770–778.

[185] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[186] H. Y. Chen and C. Y. Su, “An Enhanced Hybrid MobileNet,” in *2018 9th International Conference on Awareness Science and Technology, iCAST 2018*, Fukuoka, Japan, 19-21 Sept. 2018, pp. 308–312.

[187] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention U-Net: Learning Where to Look for the Pancreas,” in *Medical Imaging with Deep Learning*, Amsterdam, Netherlands, 4-6 July 2018, pp. 1–10.

[188] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767.

[189] P. Banda and L. Barnard, “A deep learning approach to photovoltaic cell defect classification,” in *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists on (SAICSIT)*, Port Elizabeth, South Africa, 26-28 Sept. 2018, pp. 215–221.

[190] C. Dunderdale, W. Brettenny, C. Clohessy, and E. E. van Dyk, “Photovoltaic defect classification through thermal infrared imaging using a machine learning approach,” *Progress in Photovoltaics: Research and Applications*, vol. 28, no. 3, pp. 177–188, 2020.

[191] X. Li, Q. Yang, Z. Lou, and W. Yan, “Deep Learning Based Module Defect Analysis for Large-Scale Photovoltaic Farms,” *IEEE Transactions on Energy Conversion*, vol. 34, no. 1, pp. 520–529, Mar. 2019.

[192] L. A. Zadeh, “Fuzzy Logic,” *Computer*, vol. 21, no. 4, pp. 83–93, 1988.

[193] E. Natsheh and S. Samara, “Tree Search Fuzzy NARX Neural Network Fault Detection Technique for PV Systems with IoT Support,” *Electronics*, vol. 9, no. 7, p. 1087, Jul. 2020.

[194] S. Spataru, D. Sera, T. Kerekes, and R. Teodorescu, “Diagnostic method for photovoltaic systems based on light I-V measurements,” *Solar Energy*, vol. 119, pp. 29–44, Sep. 2015.

[195] Z. Yi and A. H. Etemadi, “Fault Detection for Photovoltaic Systems Based on Multi-Resolution Signal Decomposition and Fuzzy Inference Systems,” *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1274–1283, May 2017.

[196] V. Cherkassky and Y. Ma, “Practical selection of SVM parameters and noise estimation for SVM regression,” *Neural Networks*, vol. 17, no. 1, pp. 113–126, Jan. 2004.

[197] J. Wang, D. Gao, S. Zhu, S. Wang, and H. Liu, “Fault diagnosis method of photovoltaic array based on support vector machine,” *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 2019.

[198] T. Denœux, “A k-nearest neighbor classification rule based on Dempster-Shafer theory,” *Studies in Fuzziness and Soft Computing*, vol. 219, pp. 737–760, 2008.

[199] S. A. Dudani, “The Distance-Weighted k-Nearest-Neighbor Rule,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-6, no. 4, pp. 325–327, 1976.

[200] A. Livera, G. Makrides, J. Sutterlueti, and G. E. Georghiou, “Advanced Failure Detection Algorithms and Performance Outlier Decision Classification for Grid-connected PV Systems,” in *33rd European Photovoltaic Solar Energy Conference and Exhibition*, Amsterdam, Netherlands, 25-29 Sept. 2017, pp. 2358–2363.

[201] F. Harrou, B. Taghezouit, and Y. Sun, “Improved κnN-Based monitoring schemes for detecting faults in PV systems,” *IEEE Journal of Photovoltaics*, vol. 9, no. 3, pp. 811–821, May 2019.

[202] S. R. Safavian and D. Landgrebe, “A Survey of Decision Tree Classifier Methodology,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[203] B. Kamiński, M. Jakubczyk, and P. Szufel, “A framework for sensitivity analysis of decision trees,” *Central European Journal of Operations Research*, vol. 26, no. 1, pp. 135–159, Mar. 2018.

[204] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[205] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, “A comparative study of decision tree ID3 and C4.5,” *International Journal of Advanced Computer Science and Applications, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications*, vol. 13, 2014.

[206] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda, “The CART decision tree for mining data streams,” *Information Sciences*, vol. 266, pp. 1–15, May 2014.

[207] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[208] T. K. Ho, “Random decision forests,” in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, Montreal, QC, Canada, 14-16 Aug. 1995, vol. 1, pp. 278–282.

[209] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, “Random forests and decision trees,” *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 5, p. 272, 2012.

[210] L. Hu, J. Ye, S. Chang, H. Li, and H. Chen, “A novel fault diagnostic technique for photovoltaic systems based on cascaded forest,” in *SmartIoT 2017 - Proceedings of the Workshop on Smart Internet of Things*, New York, New York, USA, 2017, pp. 1–5.

[211] Z. Chen, F. Han, L. Wu, J. Yu, S. Cheng, P. Lin, and H. Chen, “Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents,” *Energy Conversion and Management*, vol. 178, pp. 250–264, Dec. 2018.

[212] M. Heinrich, S. Meunier, A. Samé, L. Quéval, A. Darga, L. Oukhellou, and B. Multon, “Detection of cleaning interventions on photovoltaic modules with machine learning,” *Applied Energy*, vol. 263, p. 114642, Apr. 2020.

[213] D. V Lindley, “Fiducial distributions and Bayes’ theorem,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 20, no. 1, pp. 102–107, 1958.

[214] I. Rish, “An empirical study of the naive Bayes classifier,” in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 2001, vol. 3, no. 22, pp. 41–46.

[215] K. A. K. Niazi, W. Akhtar, H. A. Khan, Y. Yang, and S. Athar, “Hotspot diagnosis for solar photovoltaic modules using a Naive Bayes classifier,” *Solar Energy*, vol. 190, pp. 34–43, Sep. 2019.

[216] A. Eskandari, J. Milimonfared, and M. Aghaei, “Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on I-V characteristics,” *Solar Energy*, vol. 211, pp. 354–365, Nov. 2020.

[217] A. Mellit and S. Kalogirou, “Artificial intelligence and internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: Challenges, recommendations and future directions,” *Renewable and Sustainable Energy Reviews*, vol. 143. Elsevier Ltd, p. 110889, Jun. 01, 2021.

[218] B. Li, C. Delpha, D. Diallo, and A. Migan-Dubois, “Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review,” *Renewable and Sustainable Energy Reviews*, p. 110512, Nov. 2020.

[219] L. Suganthi, S. Iniyan, and A. A. Samuel, “Applications of fuzzy logic in renewable energy systems - A review,” *Renewable and Sustainable Energy Reviews*, vol. 48, pp. 585–607, Aug. 2015.

[220] Z. Yin and J. Hou, “Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes,” *Neurocomputing*, vol. 174, pp. 643–650, Jan. 2016.

[221] M. N. A. H. Sha’abani, N. Fuad, N. Jamal, and M. F. Ismail, “kNN and SVM Classification for EEG: A Review,” in *Lecture Notes in Electrical Engineering*, Baltimore, MD, USA, Sept. 29 – Oct. 3 2020, vol. 632, pp. 555–565.

[222] I. D. Mienye, Y. Sun, and Z. Wang, “Prediction performance of improved decision tree-based algorithms: A review,” in *Procedia Manufacturing*, 2019, vol. 35, pp. 698–703.

[223] S. Gavankar and S. Sawarkar, “Decision tree: Review of techniques for missing values at training, testing and compatibility,” in *Proceedings - AIMS 2015, 3rd International Conference on Artificial Intelligence, Modelling and Simulation*, Kota Kinabalu, Malaysia, 2-4 Dec. 2016, pp. 122–126.

[224] Z. Yi and A. H. Etemadi, “Line-to-Line Fault Detection for Photovoltaic Arrays Based on Multiresolution Signal Decomposition and Two-Stage Support Vector Machine,” *IEEE Transactions on Industrial Electronics*, vol. 64, no. 11, pp. 8546–8556, 2017.

[225] X. X. Wang, L. Dong, S. Y. Liu, Y. Hao, and B. Wang, “A Fault Classification Method of Photovoltaic Array Based on Probabilistic Neural Network,” in *Proceedings of the 31st Chinese Control and Decision Conference (CCDC)*, Nanchang, China, 3-5 Jun. 2019, pp. 5260–5265.

[226] J. Balzategui, L. Eciolaza, N. Arana-Arexolaleiba, J. Altube, J. P. Aguerre, I. Legarda-Ereño, and A. Apraiz, "Semi-automatic quality inspection of solar cell based on Convolutional Neural Networks," in *24th IEEE Conference on Emerging Technologies and Factory Automation*, Zaragoza, Spain, 10-13 Sept. 2019, vol. 2019-Septe, pp. 529–535.

[227] B. Du, Y. He, J. Duan, and Y. Zhang, "Intelligent classification of silicon photovoltaic cell defects based on eddy current thermography and convolution neural network," *IEEE Transactions on Industrial Informatics*, pp. 1–1, Nov. 2019.

[228] Z. Yi and A. H. Etemadi, "Line-to-line fault detection for photovoltaic arrays based on multi-resolution signal decomposition and two-stage support vector machine," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 11, 2017.

[229] D. S. Pillai, F. Blaabjerg, and N. Rajasekar, "A Comparative Evaluation of Advanced Fault Detection Approaches for PV Systems," *IEEE Journal of Photovoltaics*, vol. 9, no. 2, pp. 513–527, Mar. 2019.

[230] Y. Li, K. Ding, J. Zhang, F. Chen, X. Chen, and J. Wu, "A fault diagnosis method for photovoltaic arrays based on fault parameters identification," *Renewable Energy*, vol. 143, pp. 52–63, 2019.

[231] J. M. Huang, R. J. Wai, and W. Gao, "Newly-designed fault diagnostic method for solar photovoltaic generation system based on IV-Curve measurement," *IEEE Access*, vol. 7, pp. 70919–70932, 2019.

[232] N. Agrawal, B. Bora, and A. Kapoor, "Experimental investigations of fault tolerance due to shading in photovoltaic modules with different interconnected solar cell networks," *Solar Energy*, vol. 211, pp. 1239–1254, Nov. 2020.

[233] M. Wang, J. Liu, T. J. Burleyson, E. J. Schneller, K. O. Davis, R. H. French, and J. L. Braid, "Analytic Isc–Voc Method and Power Loss Modes From Outdoor Time-Series I–V Curves," *IEEE Journal of Photovoltaics*, vol. 10, no. 5, pp. 1379–1388, Sep. 2020.

[234] A. Bouaichi, A. Alami, C. Hajjaj, C. Messaoudi, A. Ghennioui, A. Benlarabi, B. Ikken, A. El, and H. Zitouni, "In-situ evaluation of the early PV module degradation of various technologies under harsh climatic conditions : The case of Morocco," *Renewable Energy*, vol. 143, pp. 1500–1518, 2019.

[235] A. M. Silva, F. C. Melo, J. H. Reis, and L. C. G. Freitas, "The study and application of evaluation methods for photovoltaic modules under real operational conditions, in a region of the Brazilian Southeast," *Renewable Energy*, vol. 138, pp. 1189–1204, 2019.

[236] P. Rajput, G. N. Tiwari, O. S. Sastry, B. Bora, and V. Sharma, "Degradation of mono-crystalline photovoltaic modules after 22 years of outdoor exposure in the composite climate of India," *Solar Energy*, vol. 135, pp. 786–795, 2016.

[237] Y. R. Golive, H. K. Singh, A. Kottantharayil, J. Vasi, and N. Shiradkar, "Investigation of Accuracy of various STC Correction Procedures for I-V Characteristics of PV Modules Measured at Different Temperature and Irradiances," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, Chicago, IL, USA, 16-21 June 2019, pp. 2743–2748.

[238] A. Dolara, G. C. Lazaroiu, S. Leva, and G. Manzolini, “Experimental investigation of partial shading scenarios on PV (photovoltaic) modules,” *Energy*, vol. 55, pp. 466–475, 2013.

[239] F. Martínez-Moreno, G. Figueiredo, and E. Lorenzo, “In-the-field PID related experiences,” *Solar Energy Materials and Solar Cells*, vol. 174, no. September 2017, pp. 485–493, 2018.

[240] J. Tanesab, D. Parlevliet, J. Whale, T. Urmee, and T. Pryor, “The contribution of dust to performance degradation of PV modules in a temperate climate zone,” *Solar Energy*, vol. 120, pp. 147–157, 2015.

[241] J. Tanesab, D. Parlevliet, J. Whale, and T. Urmee, “Seasonal effect of dust on the degradation of PV modules performance deployed in different climate areas,” *Renewable Energy*, vol. 111, pp. 105–115, 2017.

[242] M. Ma, H. Liu, Z. Zhang, P. Yun, and F. Liu, “Rapid diagnosis of hot spot failure of crystalline silicon PV module based on I-V curve,” *Microelectronics Reliability*, vol. 100–101, 2019.

[243] H. Zhu, L. Lu, J. Yao, S. Dai, and Y. Hu, “Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model,” *Solar Energy*, vol. 176, pp. 395–405, 2018.

[244] J. M. Freeman, N. A. DiOrio, N. J. Blair, T. W. Neises, M. J. Wagner, P. Gilman, and S. Janzou, “System Advisor Model (SAM) general description (version 2017.9. 5),” Technical report, National Renewable Energy Lab.(NREL), Golden, CO, United States, 2018. [Online]. Available: https://www.nrel.gov/docs/fy18osti/70414.pdf.

[245] J. Ramos-hernanz, J. M. Lopez-guede, E. Zulueta, and U. Fernandez-Gamiz, “Reverse saturation current analysis in photovoltaic cell models,” *WSEAS Transactions on Power Systems*, vol. 12, pp. 231–237, 2017.

[246] W. De Soto, S. A. Klein, and W. A. Beckman, “Improvement and validation of a model for photovoltaic array performance,” *Solar Energy*, vol. 80, no. 1, pp. 78–88, Jan. 2006.

[247] J. Summhammer, “Short flash and constant load PV-module tester,” in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Vienna, Austria, 10-13 Nov. 2013, pp. 8116–8120.

[248] A. J. Bühler, F. Perin Gasparin, and A. Krenzinger, “Post-processing data of measured I–V curves of photovoltaic devices,” *Renewable Energy*, vol. 68, pp. 602–610, Aug. 2014.

[249] J. C. H. Phang, D. S. H. Chan, and J. R. Phillips, “Accurate analytical method for the extraction of solar cell model parameters,” *Electronics Letters*, vol. 20, no. 10, pp. 406–408, 1984.

[250] A. Migan, T. Mambrini, V. Bourdin, and J. Badosa, “Deployment of a multi-technology photovoltaic module test bench on the SIRTA meteorological and climate observatory,” in *31st European PV Solar Energy Conference and Exhibition (Eu-PVSEC)*, Hamburg, Germany, 14-18 Sept. 2015.

[251] K. Ding, J. Zhang, X. Bian, and J. Xu, “A simplified model for photovoltaic modules based on improved translation equations,” *Solar Energy*, vol. 101, pp. 40–52, 2014.

[252] S. Kaplanis and E. Kaplani, “Energy performance and degradation over 20 years performance of BP c-Si PV modules,” *Simulation Modelling Practice and Theory*, vol. 19, no. 4, pp. 1201–1211, Apr. 2011.

[253] Y. Hishikawa, H. Ohshima, M. Higa, K. Yamagoe, and T. Doi, “Precise Determination of the STC I-V Curves by Wide-Range Linear Extrapolation of Outdoor I-V Curves on Partly Sunny Days,” in *32nd European Photovoltaic Solar Energy Conference and Exhibition (Eu-PVSEC)*, München, Germany, 21-24 June 2016, pp. 1716–1719.

[254] A. Mellit, G. M. Tina, and S. A. Kalogirou, “Fault detection and diagnosis methods for photovoltaic systems: A review,” *Renewable and Sustainable Energy Reviews*, vol. 91, no. April, pp. 1–17, Aug. 2018.

[255] R. Singh, M. Sharma, R. Rawat, and C. Banerjee, “An assessment of series resistance estimation techniques for different silicon based SPV modules,” *Renewable and Sustainable Energy Reviews*, vol. 98, no. September, pp. 199–216, 2018.

[256] M. H. Ali, A. Rabhi, A. El Hajjaji, and G. M. Tina, “Real Time Fault Detection in Photovoltaic Systems,” *Energy Procedia*, vol. 111, pp. 914–923, Mar. 2017.

[257] A. Peinado Gonzalo, A. Pliego Marugán, and F. P. García Márquez, “Survey of maintenance management for photovoltaic power systems,” *Renewable and Sustainable Energy Reviews*, vol. 134, p. 110347, Dec. 2020.

[258] D. Dirnberger and U. Kraling, “Uncertainty in PV module measurement-part I: Calibration of crystalline and thin-film modules,” *IEEE Journal of Photovoltaics*, vol. 3, no. 3, pp. 1016–1026, 2013.

[259] C. Reise, B. Müller, D. Moser, G. Belluardo, and P. Ingenhoven, “Report IEA-PVPS T13-12:2018 Uncertainties in PV System Yield Predictions and Assessments,” Technical report, 2018. Accessed: May 18, 2020. [Online]. Available: https://iea-pvps.org/wp-content/uploads/2020/01/Uncertainties_in_PV_System_Yield_Predictions_and_Assessments_by_Task_13.pdf.

[260] J. C. H. Phang and D. S. H. Chan, “A review of curve fitting error criteria for solar cell I-V characteristics,” *Solar Cells*, vol. 18, no. 1, pp. 1–12, Jul. 1986.

[261] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Physics Reports*, vol. 438, no. 5–6, pp. 237–329, Jan. 2007.

[262] Z. Wang and T. Oates, “Imaging Time-Series to Improve Classification and Imputation,” in *IJCAI International Joint Conference on Artificial Intelligence*, Buenos Aires, Argentina, 25-31 July 2015, pp. 3939–3945.

[263] S. Schinkel, O. Dimigen, and N. Marwan, “Selection of recurrence threshold for signal detection,” *European Physical Journal: Special Topics*, vol. 164, no. 1, pp. 45–53, Nov. 2008.

[264] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Spidlen, and J. E. Snyder-Cappione, “Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets,” *Nature Communications*, vol. 10, no. 1, pp. 1–12, Dec. 2019.

[265] S. R. Madeti and S. N. Singh, “Modeling of PV system based on experimental data for fault detection using kNN method,” *Solar Energy*, vol. 173, no. June, pp. 139–151, 2018.

[266] Y. Hishikawa, T. Takenouchi, M. Higa, K. Yamagoe, H. Ohshima, and M. Yoshita, “Translation of Solar Cell Performance for Irradiance and Temperature from a Single I-V Curve Without Advance Information of Translation Parameters,” *IEEE Journal of Photovoltaics*, vol. 9, no. 5, pp. 1195–1201, Sep. 2019.

[267] Y. Hishikawa, T. Doi, M. Higa, K. Yamagoe, H. Ohshima, T. Takenouchi, and M. Yoshita, “Voltage-dependent temperature coefficient of the I-V curves of crystalline silicon photovoltaic modules,” *IEEE Journal of Photovoltaics*, vol. 8, no. 1, pp. 48–53, 2018.

ÉCOLE DOCTORALE
Physique et ingénierie: Electrons, Photons, Sciences du vivant (EOBE)

Titre : Surveillance de l'état de santé des modules photovoltaïques à partir des mesures électriques

Mots clés : Défauts PV, courbe I-V, détection de défaut et diagnostic, méthodes d'apprentissage automatique

Résumé : La détection et le diagnostic des défauts sont des éléments essentiels pour la maintenance conditionnelle des panneaux photovoltaïques (PV). Cette thèse propose une nouvelle stratégie en quatre étapes (modélisation, prétraitement, extraction et analyse des signatures) basée sur l'utilisation des caractéristiques courant-tension (courbes I-V) complètes. La modélisation s'appuie sur une approche pilotée par les données simulées ou mesurées. Pour le prétraitement, afin d'atténuer les effets des différentes conditions de mesure, nous avons proposé une procédure de correction améliorée des courbes I-V qui est mieux adaptée aux panneaux défectueux que celles de la norme, ainsi qu'un ré-échantillonnage du vecteur courant pour toujours disposer du même nombre de points. Pour l'extraction des caractéristiques après ce prétraitement, trois méthodes sont développées : l'utilisation directe de la courbe I-V ou sa transformation par la technique Gramian Angular Difference Field (GADF) ou celle de Recurrence Plot (RP). L'analyse en composantes principales (ACP) est également appliquée pour réduire la dimension de la matrice des caractéristiques.

Pour l'analyse des caractéristiques, six techniques courantes d'apprentissage automatique sont évaluées: le réseau de neurones artificiels (ANN), la machine à vecteurs de support (SVM), l'arbre de décision (DT), la forêt aléatoire (RF), les k-plus proches voisins (kNN) , et le classifieur naïf bayésien (NBC). Pour évaluer les différentes combinaisons entre les signatures et les classifieurs, les critères de performances retenues sont la précision de la classification et la complexité du calcul. Huit conditions (une saine et sept défectueuses) des panneaux PV sont étudiées à partir de courbes I-V simulées et mesurées pour constituer la base de données. Les résultats montrent que l'utilisation des caractéristiques issues de la transformation GADF des courbes I(V) comme entrées du classifieur ANN permet d'obtenir une précision de classification de 100 %, aussi bien pour les données simulées que celles mesurées sur um banc de test développé au laboratoire. La robustesse aux perturbations, l'impact de l'ACP et de la transformation des caractéristiques sont également traités. La stratégie proposée est également comparée à celles qui n'utilisent que partiellement les informations de la courbe I-V et les techniques dans la litérature.

Title : Health monitoring of photovoltaic modules using electrical measurements

Keywords : PV faults, I-V curve, fault detection and diagnosis, machine learning

Fault detection and diagnosis are essential elements for the condition monitoring of photovoltaic (PV) panels. This thesis proposes a new four-step strategy (modelling, pre-processing, extraction, and analysis of signatures) using full current-voltage characteristics (I-V curves). The modelling is based on an approach driven by simulated or measured data. For the pre-processing, to mitigate the effects of the different measurement conditions, we proposed an improved I-V curve correction procedure that is better adapted to defective panels than the standard ones. Besides, the current vector is resampled to have the same number of points. For feature extraction after this pre-processing, three methods are developed: direct use of the I-V feature or its transformation by the Gramian Angular Difference Field (GADF) or Recurrence Plot (RP) technique. Principal component analysis (PCA) is also applied to reduce the dimension of the feature matrix.

For feature analysis, six common machine learning techniques are evaluated: artificial neural network (ANN), support vector machine (SVM), decision tree (DT), random forest (RF), k-nearest neighbour (kNN), and Bayesian naive classifier (NBC). To evaluate the different combinations of features and classifiers, the performance criteria used are classification accuracy and computational complexity. Eight conditions (one healthy and seven defective) of the PV panels are studied using simulated and measured I-V curves to build the database. The results show that using the features from the GADF transformation of the I-V curves as inputs to the ANN classifier achieves 100% classification accuracy for both simulated and measured data on a test bench developed in the laboratory. The robustness to perturbations, the impact of PCA and the feature transformation are also addressed. The proposed strategy is also compared to those that only partially use the I-V curve information and techniques in literature.

Maison du doctorat de l'Université Paris-Saclay
2ème étage aile ouest, Ecole normale supérieure Paris-Saclay
4 avenue des Sciences,
91190 Gif sur Yvette, France