



HAL
open science

A Fault Diagnosis Strategy Based on Multilevel Classification for a Cascaded Photovoltaic Grid-Connected Inverter

Wenyi Yuan, Tianzhen Wang, Demba Diallo, Claude Delpha

► **To cite this version:**

Wenyi Yuan, Tianzhen Wang, Demba Diallo, Claude Delpha. A Fault Diagnosis Strategy Based on Multilevel Classification for a Cascaded Photovoltaic Grid-Connected Inverter. *Electronics*, 2020, 9 (3), pp.429. 10.3390/electronics9030429 . hal-02903345

HAL Id: hal-02903345

<https://centralesupelec.hal.science/hal-02903345v1>

Submitted on 6 Jan 2025

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Article

A Fault Diagnosis Strategy Based on Multilevel Classification for a Cascaded Photovoltaic Grid-Connected Inverter

Wenyi Yuan ¹, Tianzhen Wang ^{1,*} , Demba Diallo ^{1,2}  and Claude Delpha ³ 

¹ College of Logistics Engineering, Shanghai Maritime University, Shanghai 201306, China; ywy_95@163.com (W.Y.); demba.diallo@geeps.centralesupelec.fr (D.D.)

² Université Paris-Saclay, CentraleSupélec, CNRS, Group of Electrical Engineering of Paris, Sorbonne Université, 3 & 11 rue Joliot-Curie, 91192 Gif-sur-Yvette, France

³ Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France; claude.delpha@l2s.centralesupelec.fr

* Correspondence: tzwang@shmtu.edu.cn; Tel.: +86-21-3828-2639

Received: 12 February 2020; Accepted: 29 February 2020; Published: 4 March 2020



Abstract: In this paper, an effective strategy is presented to realize IGBT open-circuit fault diagnosis for closed-loop cascaded photovoltaic (PV) grid-connected inverters. The approach is based on the analysis of the inverter output voltage time waveforms in healthy and faulty conditions. It is mainly composed of two parts. The first part is to select the similar faults based on Euclidean distance and set the specific labels. The second part is the classification based on Principal Component Analysis and Support Vector Machine. The classification is done in two steps. In the first, similar faults are grouped to do the preliminary diagnosis of all fault types. In the second step the similar faults are discriminated. Compared with existing fault diagnosis strategies for several fundamental periods and under different external environments, the proposed strategy has better robustness and higher fault diagnosis accuracy. The effectiveness of the proposed fault diagnosis strategy is assessed through simulation results.

Keywords: fault diagnosis; closed-loop photovoltaic system; cascaded multilevel inverter; principal components analysis; support vector machine

1. Introduction

Among the renewable energies promoted worldwide due to the environmental issues, photovoltaic (PV) energy systems are one of the most promising due to its lower environmental impact and abundance [1]. However, the connection of PV plant to the power utility grid was limited because of voltage mismatch and grid code requirements that could not be met. Thanks to the development of power converters and their control, PV plants can be connected without degrading the energy conversion efficiency thanks to the low switching frequency of cascaded multilevel inverters [2]. One key aspect in power electronic system is reliability [3], for those applications that consider availability as a critical parameter, it is important that the application continues to operate even under faulty conditions. For PV grid-connected system, the performance of the inverter is one of the key factors that determine whether the system can continue to operate. Open-circuit and short-circuit faults are the most common faults affecting inverters. Since most modern gate-drivers are equipped with short-circuit protection unit, open-circuit fault attracts more attention [4]. Figure 1a shows the application of PV grid-connected system and Figure 1b shows the consequence of photovoltaic inverter fires. Once the fault occurs, the output voltage is distorted and the produced power is degraded. If it

cannot be diagnosed and repaired in time, a derivative fault will occur, which may seriously lead to system crash.



Figure 1. Photovoltaic grid-connected system. (a) Photovoltaic power generation devices; (b) Fire in photovoltaic systems. [5].

The literature on fault diagnosis methods is abundant [6] but for each system, an appropriate strategy is required. For PV grid-connected systems there are many studies on the closed-loop control. However, for the purpose of health monitoring, most of the studies are conducted considering that the system is in open loop, which is not the usual case [7–9]. Moreover, fault diagnosis of PV systems cannot ignore the variability of the irradiance and the temperature induced by the environmental conditions. Indeed, this variability influences the inverter output voltage. Therefore, the results presented in [10–13] which only consider one environmental condition for PV inverter fault diagnosis, are limited in scope. Fault diagnosis methods can be decomposed in four steps: modelling, pre-processing, feature extraction and feature analysis for fault detection, fault classification and fault estimation [14]. In the following, only fault detection and fault classification will be discussed. Fault features can be extracted from different signals obtained from raw measurements in the time domain or transformed into another domain that can be time-frequency, time-scale or frequency. Different techniques can be used to extract and analyze the fault features ranging, e.g., from signal or information processing tools or machine-learning tools.

Here are some examples of signal-processing-based methods. Authors in [15] have proposed a relative weighting operator of principal component analysis (PCA) to extract the fault information of a cascaded inverter. In Reference [16], a multilevel signal decomposition and coefficients reconstruction method is used to generate the multiscale features for fault feature extraction. In [17], authors adopted a second low frequency processing (SLFP) method to obtain the small low-frequency data from the feedback controller. Authors in [18] have used the average bridge arm pole-to-pole (PTP) voltage and error-adaptive thresholds of the inverter to extract the fault information. In [19], an adaptive confidence limit (ACL) fault detection method is proposed to process the changing signals. The main drawbacks of these methods are their sensitivity to frequency resolution and environmental nuisances.

Machine-learning methods are becoming more and more attractive in engineering applications. Authors in [20] have designed a new generator and discriminator of Generative Adversarial Network (GAN) to extract more fault features from Auto Encoder (AE). In Reference [21], authors have proposed a multiclass Relevance Vector Machine (mRVM) to achieve higher model sparsity and shorter diagnosis time. In Reference [22], intuitionistic fuzzy logic is integrated to original spiking neural P systems for dealing with the uncertain knowledge of the power system. Authors in [23] have adopted Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Cuckoo Search Algorithm (CSA) methods to optimize the neural network in order to have the lowest mean square error.

On one hand, machine learning methods are highly adaptable and do not rely on accurate mathematical models [24]. On the other hand, they need a large amount of data (representing several operating conditions) for training the network, a significant experience to set a large

number of parameters, and the effect of each algorithm is very different for different types of input. The computational cost may also constitute an obstacle to its implementation in real engineering applications.

From the above discussion, we can conclude that time-domain analysis using signal and information processing tools may be more suitable for developing an inverter fault diagnosis method for PV grid-connected inverter system. In addition, the method should be able to cope with the closed-loop behavior and be robust to the variations of the environmental conditions (irradiance and temperature). The fault diagnosis strategy proposed in this paper is based on principle component analysis (PCA) and support vector machine (SVM). It consists of three parts. The first part is devoted to group the similar faults based on Euclidean distance and set the specific labels. The second part is the first classification level based on PCA-SVM. PCA is known as one of the most common multivariate statistical process control (MSPC) methods for dimensionality reduction while retaining the meaningful information [25]. After the feature extraction, the fault classification is performed with SVM, a classical algorithm for pattern classification. It has better generalization capability than artificial neural networks (ANN), and guarantees that local and global optimal solutions are identical [26]. The third part is the second classification level. PCA and two-class SVM are used to discriminate the similar faults. The performances of the overall method are evaluated for different environmental conditions.

The fault diagnosis process consists of four steps: modeling, pre-processing, features extraction and features analysis.

The first step is devoted to knowledge building. It can be done through physics-based equations, language-based models or data-driven. In the second step, the input data is pre-processed. The data can be filtered to reduce the nuisances or transformed from time domain to frequency domain or time-frequency domain or projected into another reference frame. The objective of this step is to prepare the information from which the best features will be extracted in the third step. In this third step the fault signatures can be extracted with different techniques ranging from signal processing, information processing and control theory for example. In the last step, the features are analyzed to decide whether a fault has occurred, to classify the different fault types, isolate the faults and eventually estimate the fault severities.

In our study we take benefit of the measured output voltage historical data to model our system. Depending on the applications one can use different signals like vibration, acoustic, phase current or electromagnetic field. Vibration and acoustic signals are most usually used to diagnose mechanical faults. In energy conversion systems the phase currents are very popular. However, in our application the current depends on the load requirement, which varies continuously during daytime. There are many papers [27–29] that have already proposed multi-level inverter fault diagnosis using voltage spectral analysis. However, in order to avoid any additional transformation such as Fast Fourier Transform (FFT) or Wavelet Transform, we have decided to exploit the voltage characteristics in the time domain.

Moreover, in multi-level converters [30], the shape of the output voltage depends on the states of the power switches. Therefore, any fault affecting the power converter will directly modify the shape of the output voltage.

This paper is outlined as follows. In Section 2, the open-circuit fault features of a cascaded five-level inverter in closed-loop PV grid-connected system are analyzed under different external environments. In Section 3, the proposed fault diagnosis strategy is presented. In Section 4, the effectiveness of the proposed strategy is evaluated for several operating time durations and under different environmental conditions through numerical simulations. Finally, the conclusion is provided in Section 5.

2. Problem Description

The cascaded five-level inverter for a single-phase PV grid-connected system is shown in Figure 2 [31], which is mainly composed of PV sources, two H-bridge inverters connected in series, inductive filter and the public grid. PV voltages, PV currents, grid current and grid voltage are required

to generate control signals for the closed-loop control strategies of Maximum Power Point Tracking (MPPT), voltage and current control loops and power balance control.

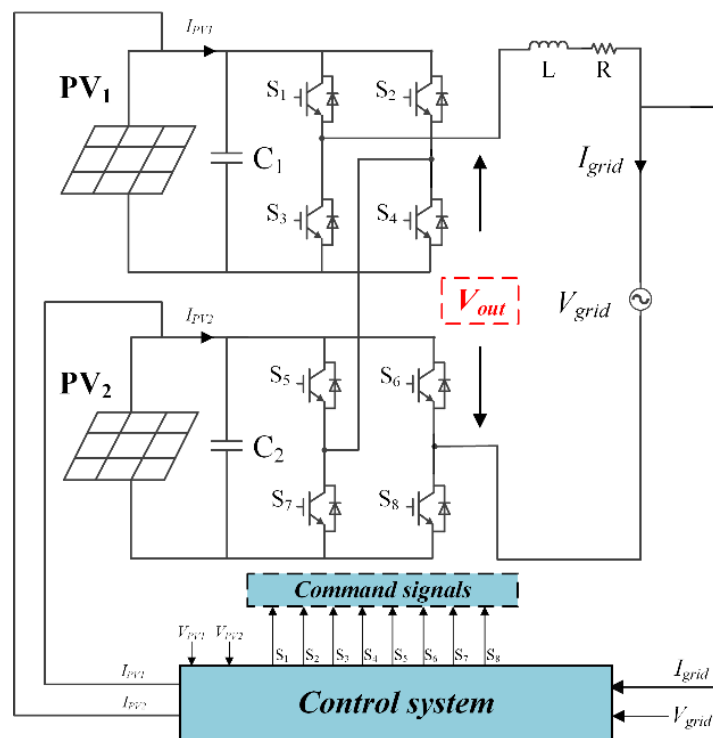


Figure 2. Cascaded photovoltaic grid-connected system topology.

In this PV system, the two H-bridges are composed of eight IGBTs. Since the most common faults in the industry are single IGBT open-circuit faults [32]; in this paper, the healthy state of nine conditions will be analyzed, alongside eight single IGBT open-circuit faults.

2.1. Similar Faults

Figure 3 shows the healthy state and eight IGBTs ($S_1 \sim S_8$) open-circuit output voltage waveforms. At fault occurrence ($t = 0.2$ s) the output voltage waveform is distorted and after a transient it gradually stabilizes due to the closed-loop adjustment. When analyzing these eight IGBTs open-circuit waveforms, we have found a fault diagnosis accuracy of 80% in [33], 85% in [34] and 90% in [35]. We can also deduce from Figure 3 that there are two groups of similar output voltage waveforms as shown in Table 1. The high similarity makes them difficult to distinguish. The S_2 and S_3 open-circuit faults are in group 1 while S_5 and S_8 open-circuit faults are in group 2. In the following, due to page limitation, we will focus only on feature analysis of data for the similar faults. Taking group 1 as an example, the inverter output voltage waveforms over several periods and under different environmental conditions will be analyzed in detail.

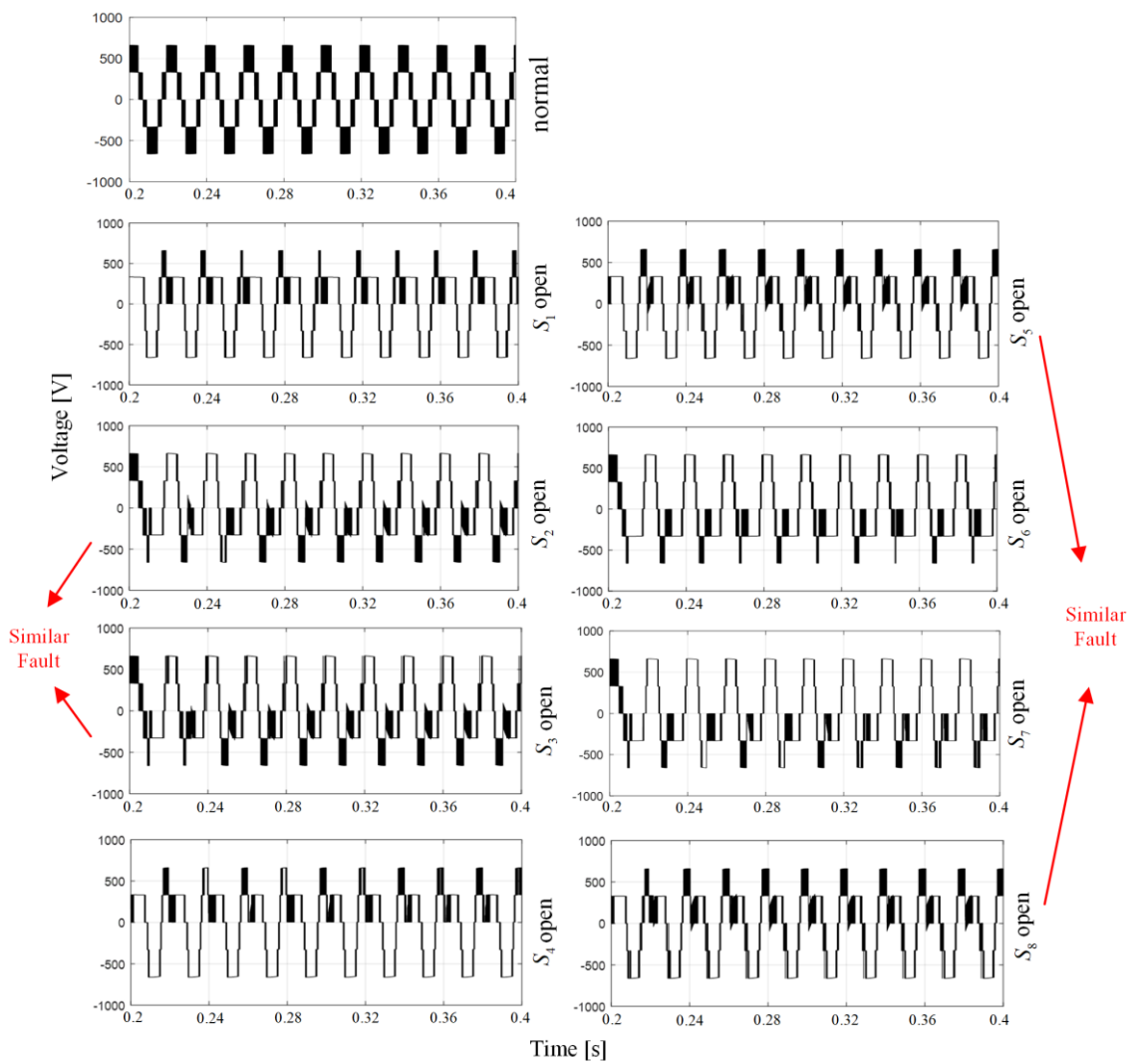


Figure 3. Healthy and eight IGBTs open-circuit output voltage waveforms.

Table 1. Similar faults.

Group Number	Fault Type
1	S_2 open-circuit and S_3 open-circuit
2	S_5 open-circuit and S_8 open-circuit

2.2. The Impact of Different Fundamental Periods and Different Environmental Conditions

In order to illustrate the effect of different fundamental periods and different environmental conditions on the fault waveforms, two conditions are chosen for S_2 and S_3 open-circuit faults for 10 periods, as shown in Figures 4 and 5. The environmental condition 1 is 9:00 a.m. on February 18th: the solar irradiation intensity is 224 W/m^2 , temperature is $4.5 \text{ }^\circ\text{C}$; the environmental condition 2 is 13:00 p.m. on August 18th: the solar irradiation intensity is 698 W/m^2 , temperature is $28.9 \text{ }^\circ\text{C}$ (the data is acquired from Harnhill and Diddington in the U.K [36]).

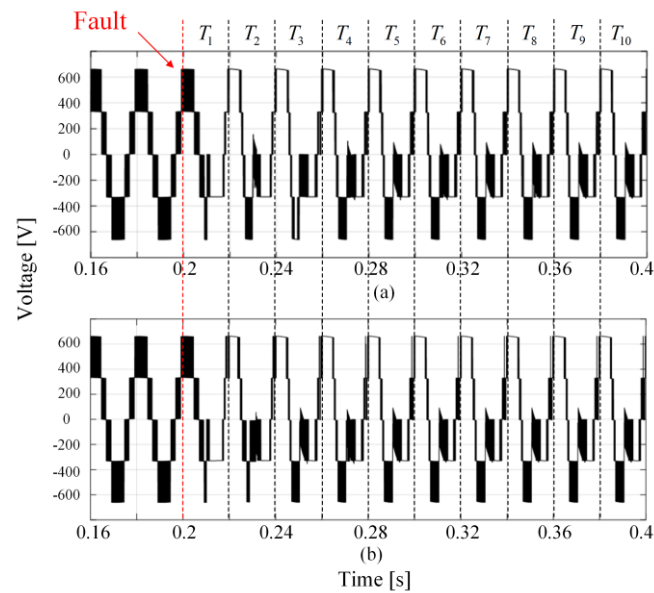


Figure 4. Inverter output voltage waveforms over 10 periods for S_2 and S_3 open-circuit faults at 0.2s under condition 1. (a) Switch S_2 open-circuit fault; (b) Switch S_3 open-circuit fault.

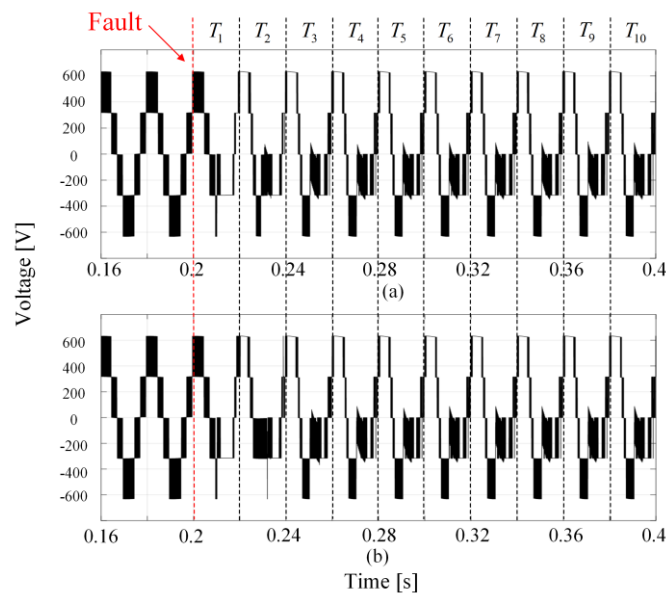


Figure 5. Inverter output voltage waveforms over 10 periods for S_2 and S_3 open-circuit faults at 0.2 s under condition 2. (a) Switch S_2 open-circuit fault; (b) Switch S_3 open-circuit fault.

Figure 4 shows the inverter output voltage waveform over 10 periods under condition 1. At fault occurrence at 0.2 s, the output voltage waveforms fluctuate in a period of time due to the self-regulating effect of the closed-loop system. We take the moment when the fault occurs as the beginning of the first period T_1 ; T_n ($n = 1, 2, \dots, 10$) represents a sequence of fundamental periods—each fundamental period is equal to 0.02 s.

Figure 5 shows the inverter output voltage waveforms over 10 periods when S_2 and S_3 open-circuit faults occur at 0.2 s under condition 2. We can notice from Figures 4 and 5 that despite the same fault there are some differences due to the different environmental conditions. To show the differences more clearly, Euclidean distance is calculated between S_2 and S_3 open-circuit faults in each period under each environmental condition (after standardization). The results are plotted in Figure 6. The first two periods correspond to the healthy state (denoted as H), so the Euclidean distances are close to

zero. At fault occurrence, the Euclidean distances clearly change. The distances increase during the periods T_1 and T_2 , before decreasing gradually due to the control loops actions. We can also notice slight differences for the two environmental conditions.

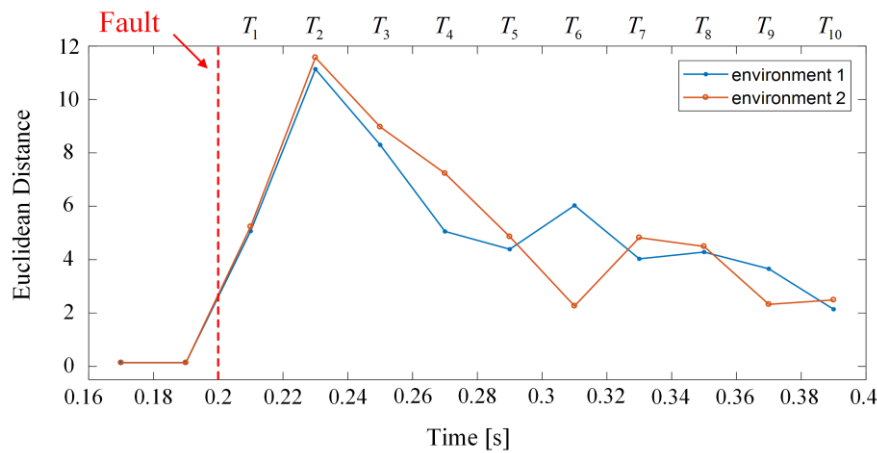


Figure 6. Euclidean distance over 10 periods for S_2 and S_3 open-circuit fault waveforms under two environmental conditions.

The former analysis has pointed out that the output voltage waveform is sensitive to the inverter IGBT open-circuit fault. However, the results also show that the waveform is affected by the dynamics of closed-loop action and the variations of the environmental conditions (the irradiance and the temperature). The results also show that some faults have similar signatures. All these issues should be addressed using the fault diagnosis method detailed in the following section.

3. Fault Diagnosis Strategy Based on Multilevel Classification

As shown in Figure 7, a block diagram of grid-connected PV plant fault diagnosis is illustrated. The DC supply of 5-level inverter is from PV modules, which is influenced by solar irradiance and temperature. The output of 5-level inverter is connected to the grid by control strategies. The output voltage of the inverter is collected as the fault diagnosis signals and through the proposed fault diagnosis strategy, the health status of the 5-level inverter is monitored. In this section, the fault diagnosis strategy is focused on described in detail, which is contained three parts: data standardization and faults labeling, the first classification level for all fault types and the second classification level for the faults with similar signatures.

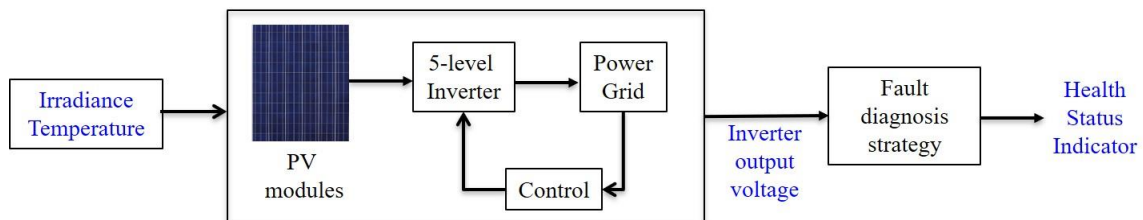


Figure 7. Grid-connected PV plant fault diagnosis block diagram.

3.1. Data Standardization and Faults Labeling

In order to reduce the influence of the dimension and the wide range of variation of the inverter output voltage on the fault diagnosis, the first step is to standardize the input signals using the Z-score

method. Let $X_{[N \times m]}$ be the original data matrix, where m is the number of variables and N is the number of samples. The matrix is given by:

$$X_{[N \times m]} = [x_1, x_2, \dots, x_j, \dots, x_m], \tag{1}$$

where $x_j(j = 1, 2, \dots, m)$ is the j th observation. The Z-score formula is expressed as:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \text{ for } i = 1, 2, \dots, N, j = 1, 2, \dots, m \tag{2}$$

where x_{ij} is the i th sample of the j th observation, \bar{x}_j and σ_j are respectively the mean value and the standard deviation of the j th observation.

Hence, the standardized matrix after Z-score is given by:

$$Z_{[N \times m]} = [z_1, z_2, \dots, z_j, \dots, z_m], \tag{3}$$

The second step is to add category labels for the different fault types. In the previous section we have shown that some faults have similar signature. Therefore, in our approach, we will develop a multi-level fault classification. In the first level, faults with similar signatures are merged in the same group and distinguished from other faults. In the second level, they will be discriminated. In this paper we introduce Euclidean distance to group similar faults. Assume that there are h kinds of faults, denoted as F_1, F_2, \dots, F_h , each kind of fault containing p features. Considering two faults F_v and F_w their Euclidean distance $dist(F_v, F_w)$ is computed and compared to a threshold. If equation (4) is verified, the two faults F_v and F_w are assumed to be similar and classified in the same group.

$$dist(F_v, F_w) = \sqrt{\sum_{q=1}^p [F_v(q) - F_w(q)]^2} \leq \alpha, \tag{4}$$

where α is a similarity threshold adaptively set according to the different systems. Based on the similarity threshold, we will obtain d groups of similar faults.

In the first classification level, the similar faults of each group are regarded as one fault and then all fault types are labeled. In the second classification level, the labels of similar faults in each group are updated. Therefore, in the end each fault has its own and unique label.

3.2. The First Classification Level for all Fault Types

The objective of this first classification level using PCA-SVM is to make a preliminary diagnosis of the faults having distinctive signatures.

PCA [37] is one of the most widely used data dimensionality reduction methods. It maps the original data to a new coordinate system through linear transformations. It retains the main features and removes noise and outliers to achieve data dimensionality reduction. Starting with the standardized matrix Z given in Equation (3), the covariance matrix s calculated as

$$Cov = \frac{1}{N-1} Z^T Z, \tag{5}$$

where $(.)^T$ is the transpose operation. The Cumulative Percentage of Variance for the eigenvalue in descending order is given by:

$$CPV(k) = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^m \lambda_j} \times 100\%, \tag{6}$$

where $CPV(k)$ is k th cumulative percentage of variance, $\lambda_j (j = 1, 2, \dots, m)$ are the descending eigenvalues of the covariance matrix. The retained number l of principal components:

$$l = \arg \min(CPV(k) \geq \beta), \quad (7)$$

where β is a threshold set to minimize the loss information due to the dimension reduction. Finally, the projection of matrix Z into the principal subspace is the matrix of principal components denoted as:

$$Y = Z\bar{P}, \quad (8)$$

where $\bar{P}_{[m \times l]} = [p_1, p_2, \dots, p_l]$ is the matrix of eigenvectors spanning the principal subspace.

Support vector machine (SVM) will be used for fault classification. SVM [38,39] has been originally designed for classifying a dataset in two groups. The main idea consists in finding the linear classifier (hyperplane) in a higher dimensional space that will allow to maximizing the distance between the two classes. Currently, to address multi-classification, the original problem is converted into several two-class problems that can be directly solved by multiple SVMs [40]. In this paper, the one-versus-one method is used to do the preliminary classification for all fault types.

One-versus-one SVM uses the majority voting mechanism to classify the unknown samples. The classification result is determined by the largest number of votes. In this study, we have used the LIBSVM tool. Y and labels of the first classification level are used to train the SVM multi-classifier.

3.3. The Second Classification Level for the Faults with Similar Signatures

The goal of this second classification level is to discriminate the faults within the d groups of faults with similar signatures. Indeed, after the first classification these faults share the same label. PCA-SVM is also applied in this part and as the methodology is the same, in the following we use group 1 as an example. The classification is organized in three steps:

Step 1. Select the observations that belong to group 1.

Denote $Z_{g[N_g \times m]}$ as the selected data matrix of group 1 with N_g observations of m feature variables, the selected data matrix is given by:

$$Z_{g[N_g \times m]} = [(z_1)_g, (z_2)_g, \dots, (z_m)_g], \quad (9)$$

Step 2. Feature extraction for the selected data matrix $Z_{g[N_g \times m]}$ by using PCA. The matrix of principal components $Y_{g[N_g \times l_g]}$ is obtained, where l_g is the number of principal components of group 1.

Step 3. Fault classification for the selected observations using Y_g and the second level classification labels as input data to SVM.

The flowchart of the multi-level classification fault diagnosis strategy based on PCA-SVM is shown in Figure 8 and the working process can be described as in the following.

The proposed fault diagnosis strategy is divided into two parts, offline process and online process. The offline process includes data standardization, grouping the similar faults based on the similarity threshold, labeling of all faults, then building the proposed classification model, including training the first classification level model for all fault types, and training the second classification level model. For the online process, after the data standardization, the first classification level is performed based on the trained model. Then the similar faults based on the first classification results are processed through the second classification level. Finally, the fault diagnosis results are obtained.

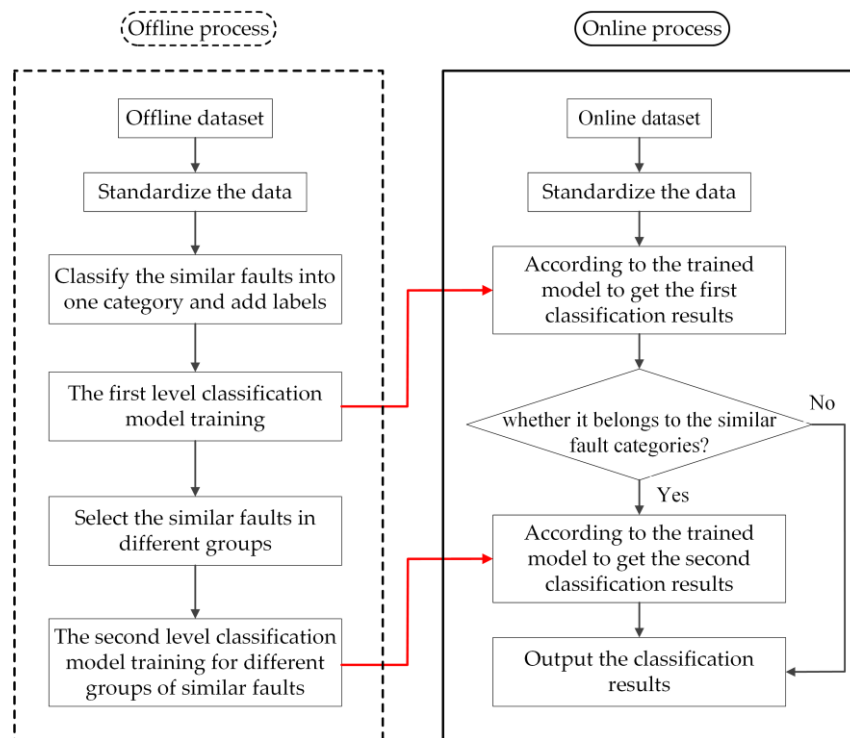


Figure 8. Flowchart of the proposed fault diagnosis strategy.

4. Simulation Results and Analysis

In this section, the simulation results of the proposed fault diagnosis strategy are presented along with its performances. The single-phase cascaded five-level photovoltaic grid-connected system is modeled under Matlab-Simulink[®]. The output voltage of each PV array is 330 V, the inductance filter is 380 mH, the resistance is 10 Ω , and the voltage frequency of the public grid is 50 Hz. The switching frequency of the inverter is set as 5 kHz, and for data acquisition the sampling frequency is 50 kHz. The corresponding parameters of the fault diagnosis strategy are given in Table 2. Open-circuit fault is achieved by disconnecting the IGBTs gate drive signals in steady state, and the output voltage of the inverter is used as fault signature.

For the hardware, the system is designed for health monitoring and does not need to be triggered continuously. Considering conventional centralized PV plants, a judicious partitioning could be envisaged between software and hardware. For the electronic hardware, one solution could be to have a dedicated PCB for data acquisition using FPGA (e.g., Altera EP3C16F484C6) at a high sampling rate and another PCB with a microcontroller or a DSP (e.g., TMS320F28335) for data processing. For decentralized PV plants (meaning small DC-DC and DC-AC converters for 2 PV modules) the control and monitoring are embedded within the box attached with the power converters and their sensors. We can take benefit of the rapid development in electronic equipment to include more computational and data acquisition capability for monitoring purposes.

The environmental data for PV panels such as solar irradiation and temperature is acquired from Harnhill and Diddington in United Kingdom [36]. In order to have variability, we retain the data of every three months in a year (February 18, May 18, August 18 and November 18) and several time ranges in each day at 9:00 a.m., 11:00 a.m., 13:00 p.m. and 15:00 p.m. However, because [41] have shown that the PV panels output voltage remain fairly constant below 200 W/m², we have removed the data with an irradiance lower than 200 W/m². Finally, we have worked with 13 different environmental conditions, denoted as E_c ($c = 1, 2, \dots, 13$).

Under these conditions, we have collected the output voltage for the healthy state and the eight faulty conditions. Each voltage time-series is composed of 10 fundamental periods after fault occurrence and 1000 samples per period.

Table 2. Main parameters of the fault diagnosis strategy.

Notation	Description	Value
N	Total number of samples	58,500
m	Number of feature variables	1000
h	Number of fault categories	9
α	The similarity threshold	5
d	Number of group with similar faults	2
β_1, β_2	Percentage of the information retained	95%, 99%
f_{switch}	Switching frequency	5 kHz
f_{sample}	Sampling frequency	50 kHz
f_{grid}	Voltage frequency of the public grid	50 Hz
V_{PV1}, V_{PV2}	Output voltage of each PV array	330 V
L	Inductance filter	380 mH
R	Resistance	10 Ω
V_{grid}	Voltage of the public grid	220 V

For training the proposed fault diagnosis model, Table 3 shows the fault labels for the different classification levels. In the first level, S_2 and S_3 open-circuit faults of group 1 are labeled as 3, S_5 and S_8 open-circuit faults of group 2 are labeled as 6. The other faults are labeled in order. In the second classification level designed to discriminating faults with similar signatures, the labels of group 1 are changed from 3 to 3 and 4 for S_2 and S_3 open-circuit faults respectively, and for group 2 are changed from 6 to 6 and 9 for S_5 and S_8 open-circuit faults respectively. Therefore, the final output labels have a one-to-one correspondence with all the conditions. The CPV (Cumulative Percentage of Variance) is set to 95% for the first classification level and 99% for the second one. The PCA output will be used as input data for the SVM classifier. In the first classification level, we have used the LIBSVM module that adopts the “one-versus-one” method to do the multi-classification. A Radial Basis Function (RBF) is selected as kernel function and its parameter and the error cost coefficient are both set to 2. In the second classification level, linear kernel function is selected, but for group 1 of similar faults, the parameter and the error cost coefficient are set respectively to 2 and 0.5. For group 2 of similar faults, the parameter and the error cost coefficient are set respectively to 3.1 and 0.4.

Table 3. Fault labels in different level classification models.

Fault Type	Label for The First Classification	Label for The Second Classification	Output Label
Healthy state	1		1
S_1 open-circuit	2		2
S_2 open-circuit	3	3	3
S_3 open-circuit	3	4	4
S_4 open-circuit	5		5
S_5 open-circuit	6	6	6
S_6 open-circuit	7		7
S_7 open-circuit	8		8
S_8 open-circuit	6	9	9

Finally, its performance is analyzed with regard to the stability of its results over different periods and its robustness against variations in environmental conditions.

4.1. Stability over Different Periods of The Proposed Strategy

In order to demonstrate that the proposed fault diagnosis strategy is still effectiveness for all types of faults over different periods, we use 10 periods of faulty samples as 10 different test sets for

evaluation. Each of the tests set contains samples representing the different environmental conditions. Denote the first period after the fault occurrence as T_1 , the second period is T_2 and so on. The accuracy is introduced as an evaluation index of the performance of the fault diagnosis strategy, and its formula is given by:

$$\text{Accuracy} = \frac{\text{Predict the correct samples in the test set}}{\text{Samples of the test set}} \times 100\%, \tag{10}$$

Table 4 shows the accuracy of the strategy over the 10 periods and for comparison, three other classical fault diagnosis strategies are chosen, PCA-SVM [35], PCA-ELM (Extreme Learning Machine) [33] and PCA-DT (Decision Tree) [34]. It can be seen from Table 4 that the accuracy of the proposed strategy is always above 90% and the average accuracy is 95.13%. PCA-SVM is the first part of the proposed strategy but the output labels have a one-to-one correspondence with all types of faults. The accuracy of PCA-SVM is around 90% and the average accuracy is 92.31%, which is lower than the proposed strategy. In the diagnostic strategy of PCA-ELM, the hidden layer nodes are set to 40, and the activation function of the hidden layer neuron is ‘sig’. The accuracy of PCA-ELM is around 80% and the average accuracy is 79.40%, which is much lower than the proposed strategy. PCA-DT is used with the C4.5 algorithm. The accuracy of PCA-DT is around 87%, and the average accuracy is 87.61%.

Table 4. Accuracy in percentage of different fault diagnosis strategies over 10 periods.

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
[35]	89.74	96.58	95.73	88.03	94.02	90.60	93.16	88.89	95.73	90.60
[33]	83.76	76.07	81.20	79.49	82.05	78.63	77.78	81.20	77.78	76.07
[34]	85.47	88.03	89.74	87.18	88.03	89.74	88.89	84.62	87.18	87.18
Our proposal	98.29	94.02	94.87	91.45	98.29	92.31	96.58	91.45	98.29	95.73

In order to show the performance of each fault diagnosis strategy more intuitively, we have drawn the results in Table 4 into a line chart, as shown in Figure 9. The red line is the accuracy of proposed strategy, and the blue yellow and green lines represent PCA-SVM, PCA-ELM and PCA-DT, respectively. From Figure 9, we can observe that the accuracy of the proposed strategy is higher than that of the other three strategies over all periods except for T_2 and T_3 where PCA-SVM performs better. Taking period T_2 under the environment E_1 as an example, Table 5 shows the Euclidean distance for every two faults over period T_2 under E_1 . From Table 5 we can see that the Euclidean distance between S_4 and S_5 open-circuit faults is smaller than that between S_2 and S_3 open-circuit faults (group 1); meaning that the proposed fault diagnosis strategy with its two classification levels has no advantage over PCA-SVM. The same results are observed over period T_3 .

Table 5. Euclidean distance for every two faults over period T_2 under E_1 .

Open-Circuit	Healthy	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
Healthy	0								
S_1	22.32	0							
S_2	23.23	34.83	0						
S_3	21.03	32.24	10.95	0					
S_4	22.49	12.97	35.73	32.93	0				
S_5	22.54	12.86	35.36	32.85	7.98	0			
S_6	23.81	34.47	11.71	11.25	35.42	35.05	0		
S_7	23.95	36.23	9.59	14.14	36.93	36.47	13.10	0	
S_8	21.75	12.74	34.33	31.75	10.89	7.98	33.90	35.47	0

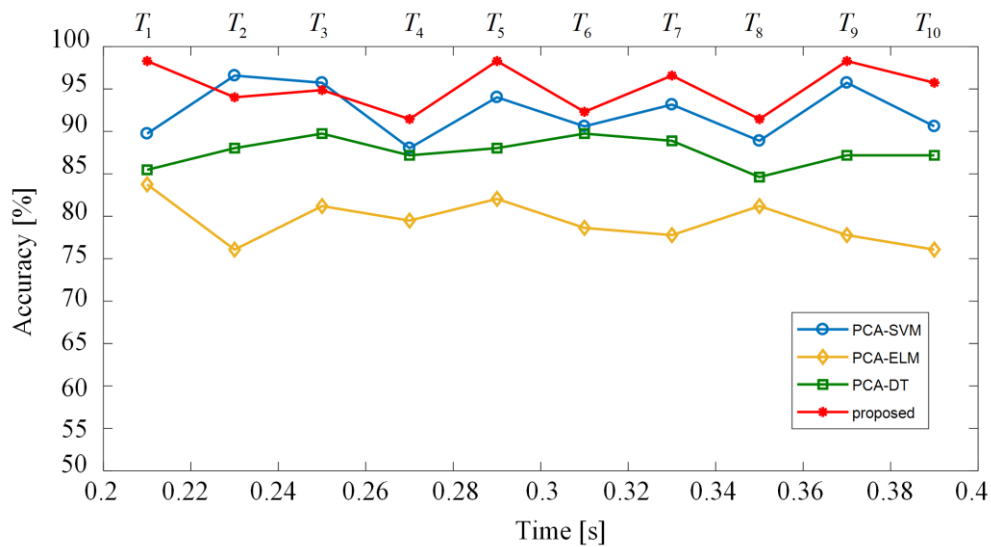


Figure 9. Accuracy of different fault diagnosis strategies for 10 periods.

4.2. Robustness Against Different Environmental Conditions

We have used different kinds of fault samples as test sets to evaluate the robustness of the proposed strategy against the variation of the environmental conditions; irradiance and temperature. Table 6 shows the accuracy of the different fault diagnosis strategies under 13 environmental conditions. The corresponding line chart is shown in Figure 10. It can be seen from Table 6 that the accuracy of the proposed strategy is around 95%, and the average accuracy is 95.81%. Its accuracy is higher than PCA-SVM (85.56%), PCA-ELM (77.10%) and PCA-DT (79.15%). From Figure 10, we can see that the proposed strategy has a higher accuracy in most cases. The accuracy of PCA-SVM in E₅ is a little bit higher than the proposed strategy, as a whole. The accuracy of PCA-SVM oscillates too much compared to the proposed strategy. That is to say, PCA-SVM has good fault diagnosis performance for constant environmental condition, but the proposed fault diagnosis strategy is more suitable, stable and robust for variable environmental conditions.

Table 6. Accuracy of different fault diagnosis strategies under 13 different environmental conditions (E₁ to E₁₃).

13 Different Environmental Conditions: E ₁ ~E ₁₃						
	E ₁	E ₂	E ₃	E ₄	E ₅	E ₆
[35]	68.89%	96.67%	93.33%	97.78%	92.22%	87.78%
[33]	73.33%	70%	85.56%	76.67%	77.78%	75.56%
[34]	67.78%	82.22%	90%	86.67%	87.78%	73.33%
proposal	98.89%	97.78%	97.78%	98.89%	90%	96.67%
	E ₇	E ₈	E ₉	E ₁₀	E ₁₁	E ₁₂
	83.33%	77.78%	82.22%	83.33%	81.11%	80%
	77.78%	78.89%	74.44%	78.89%	76.67%	75.56%
	81.11%	74.44%	72.22%	80%	77.78%	75.56%
	96.67%	96.67%	92.22%	93.33%	96.67%	93.33%
	E ₁₃					
	87.78%					
	81.11%					
	80%					
	96.67%					

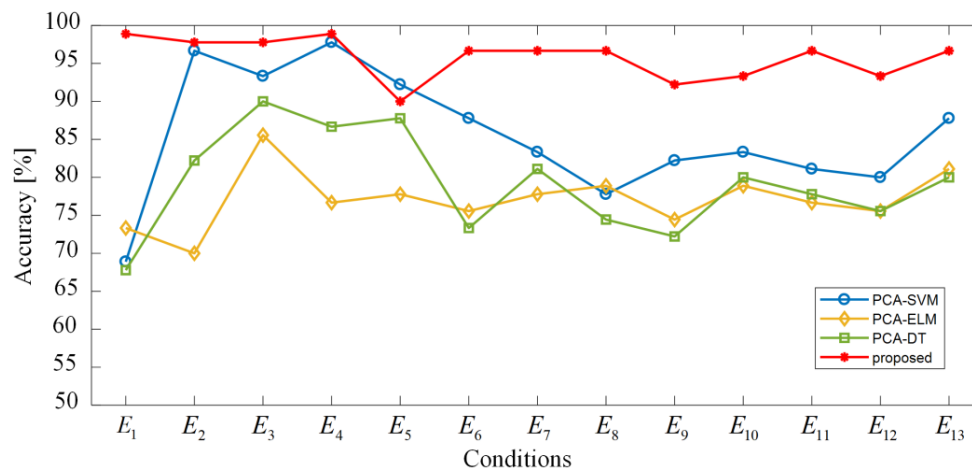


Figure 10. Accuracy of different fault diagnosis strategies under 13 variable conditions.

5. Conclusions

In this paper, a fault diagnosis strategy for a cascaded PV grid-connected inverter has been proposed. Open-circuit faults are addressed. The output inverter voltage waveform in the time domain is used as input signal for features extraction. Unfortunately, the analysis has shown that different faults have similar signatures, for which a Euclidean distance has been found lower than the preset threshold. Therefore, the method is based on a two-level classification approach using PCA-SVM. In the first level, the classification is done among faults having distinctive signatures while the similar ones having the same label are grouped. In the second classification level, those with similar signatures are discriminated with updated labels. The method has been evaluated with a closed-loop PV system and under different environmental conditions with changing irradiance and temperature. The simulation results have shown the effectiveness of the proposed strategy over several fundamental periods and under different irradiances and temperatures. The comparison with classical fault diagnosis strategies such as PCA-SVM, PCA-ELM and PCA-DT has shown an improvement in fault diagnosis performances.

Author Contributions: Conceptualization, W.Y. and T.W.; methodology, W.Y., T.W., D.D. and C.D.; formal analysis, W.Y.; investigation, W.Y.; writing—original draft preparation, W.Y., T.W., D.D. and C.D.; supervision, T.W., D.D. and C.D.; writing—review and editing, W.Y., T.W., D.D. and C.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 61673260.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Manai, L.; Dabboussi, M.; Armi, F.; Besbes, M. Cascaded multilevel inverter control considering low harmonic content based on comparison study between firefly and Newton Raphson Algorithm. In Proceedings of the 4th International Conference on Control Engineering and Information Technology, Hammamet, Tunisia, 16–18 December 2016.
- Pires, V.F.; Amaral, T.G.; Foito, D.; Pires, A.J. Cascaded H-bridge multilevel inverter with a fault detection scheme based on the statistic moments indexes. In Proceedings of the 2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering, CPE-POWERENG, Cadiz, Spain, 4–6 April 2017; pp. 193–198.
- Olalla, C.; Maksimovic, D.; Deline, C.; Martinez-Salamero, L. Impact of distributed power electronics on the lifetime and reliability of PV systems. *Prog. Photovolt. Res. Appl.* **2017**, *25*, 821–835. [[CrossRef](#)]
- Wang, H.; Pei, X.; Wu, Y.; Xiang, Y.; Kang, Y. Switch Fault Diagnosis Method for Series-Parallel Forward DC-DC Converter System. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4684–4695. [[CrossRef](#)]

5. Hidden Dangers in PV Power Stations. Available online: https://www.sohu.com/a/216141961_289078 (accessed on 11 January 2018).
6. Gao, Z.; Cecati, C.; Ding, S.X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767. [[CrossRef](#)]
7. Boutasseta, N.; Ramdani, M.; Mekhilef, S. Fault-tolerant power extraction strategy for photovoltaic energy systems. *Sol. Energy* **2018**, *169*, 594–606. [[CrossRef](#)]
8. Ben Youssef, F.; Sbita, L. Sensors fault diagnosis and fault tolerant control for grid connected PV system. *Int. J. Hydrog. Energy* **2017**, *42*, 8962–8971. [[CrossRef](#)]
9. Espinoza-Trejo, D.R.; Diez, E.; Bárcenas, E.; Verde, C.; Espinoza-Pérez, G.; Bossio, G. Model-based Fault Detection and Isolation in a MPPT BOOST converter for photovoltaic systems. In Proceedings of the IECON Proceedings (Industrial Electronics Conference), Florence, Italy, 23–26 October 2016; pp. 2189–2194.
10. Chine, W.; Mellit, A.; Lugh, V.; Malek, A.; Sulligoi, G.; Massi Pavan, A. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. *Renew. Energy* **2016**, *90*, 501–512. [[CrossRef](#)]
11. Harrou, F.; Sun, Y.; Taghezouit, B.; Saidi, A.; Hamlati, M.E. Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches. *Renew. Energy* **2018**, *116*, 22–37. [[CrossRef](#)]
12. Kumar, B.P.; Ilango, G.S.; Reddy, M.J.B.; Chilakapati, N. Online fault detection and diagnosis in photovoltaic systems using wavelet packets. *IEEE J. Photovolt.* **2018**, *8*, 257–265. [[CrossRef](#)]
13. Lu, S.; Phung, B.T.; Zhang, D. A comprehensive review on DC arc faults and their diagnosis methods in photovoltaic systems. *Renew. Sustain. Energy Rev.* **2018**, *89*, 88–98. [[CrossRef](#)]
14. Zhang, X.; Ding, F.; Yang, E. State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors. *Int. J. Adapt. Control Signal Process.* **2019**, *33*, 1157–1173. [[CrossRef](#)]
15. Wang, T.; Qi, J.; Xu, H.; Wang, Y.; Liu, L.; Gao, D. Fault diagnosis method based on FFT-RPCA-SVM for Cascaded-Multilevel Inverter. *ISA Trans.* **2016**, *60*, 156–163. [[CrossRef](#)] [[PubMed](#)]
16. Wang, Z.; Huang, Z.; Song, C.; Zhang, H. Multiscale adaptive fault diagnosis based on signal symmetry reconstitution preprocessing for microgrid inverter under changing load condition. *IEEE Trans. Smart Grid* **2018**, *9*, 797–806. [[CrossRef](#)]
17. Huang, Z.; Wang, Z.; Yao, X.; Zhang, H. Multi-switches fault diagnosis based on small low-frequency data for voltage-source inverters of PMSM drives. *IEEE Trans. Power Electron.* **2019**, *34*, 6845–6857. [[CrossRef](#)]
18. Li, Z.; Ma, H.; Bai, Z.; Wang, Y.; Wang, B. Fast Transistor Open-Circuit Faults Diagnosis in Grid-Tied Three-Phase VSIs Based on Average Bridge Arm Pole-to-Pole Voltages and Error-Adaptive Thresholds. *IEEE Trans. Power Electron.* **2018**, *33*, 8040–8051. [[CrossRef](#)]
19. Wang, T.; Wu, H.; Ni, M.; Zhang, M.; Dong, J.; Benbouzid, M.E.H.; Hu, X. An adaptive confidence limit for periodic non-steady conditions fault detection. *Mech. Syst. Signal Process.* **2016**, *72–73*, 328–345. [[CrossRef](#)]
20. Zhou, F.; Yang, S.; Fujita, H.; Chen, D.; Wen, C. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowl. Based Syst.* **2020**, *187*, 104837.
21. Wang, T.; Xu, H.; Han, J.; Elbouchikhi, E.; Benbouzid, M.E.H. Cascaded H-Bridge Multilevel Inverter System Fault Diagnosis Using a PCA and Multiclass Relevance Vector Machine Approach. *IEEE Trans. Power Electron.* **2015**, *30*, 7006–7018. [[CrossRef](#)]
22. Peng, H.; Wang, J.; Ming, J.; Shi, P.; Perez-Jimenez, M.J.; Yu, W.; Tao, C. Fault diagnosis of power systems using intuitionistic fuzzy spiking neural P systems. *IEEE Trans. Smart Grid* **2018**, *9*, 4777–4784. [[CrossRef](#)]
23. Manjunath, T.G.; Kusagur, A. Multilevel inverter fault diagnosis using optimised radial basis neural network—A novel performance enhancement. In Proceedings of the 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, ICEECCOT, Mysuru, India, 9–10 December 2016; pp. 102–105.
24. Gorzalczy, M.B.; Rudzinski, F. Generalized self-organizing maps for automatic determination of the number of clusters and their multiprototypes in cluster analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2833–2845. [[CrossRef](#)]
25. Liu, Y.J.; André, S.; Saint Cristau, L.; Lagresle, S.; Hannas, Z.; Calvosa, É.; Devos, O.; Duponchel, L. Multivariate statistical process control (MSPC) using Raman spectroscopy for in-line culture cell monitoring considering time-varying batches synchronized with correlation optimized warping (COW). *Anal. Chim. Acta* **2017**, *952*, 9–17. [[CrossRef](#)]
26. An, Y.; Ding, S.; Shi, S.; Li, J. Discrete space reinforcement learning algorithm based on support vector machine classification. *Pattern Recognit. Lett.* **2018**, *111*, 30–35. [[CrossRef](#)]

27. Lezana, P.; Aguilera, R.; Rodríguez, J. Fault detection on multicell converter based on output voltage frequency analysis. *IEEE Trans. Ind. Electron.* **2009**, *56*, 2275–2283. [[CrossRef](#)]
28. Rodríguez, J.R.; Hammond, P.W.; Pontt, J.; Musalem, R.; Lezana, P.; Escobar, M.J. Operation of a medium-voltage drive under faulty conditions. *IEEE Trans. Ind. Electron.* **2005**, *52*, 1080–1085. [[CrossRef](#)]
29. Lamb, J.; Mirafzal, B. Open-Circuit IGBT Fault Detection and Location Isolation for Cascaded Multilevel Converters. *IEEE Trans. Ind. Electron.* **2017**, *64*, 4846–4856. [[CrossRef](#)]
30. Yu, Y.; Konstantinou, G.; Hredzak, B.; Agelidis, V.G. Operation of Cascaded H-Bridge Multilevel Converters for Large-Scale Photovoltaic Power Plants under Bridge Failures. *IEEE Trans. Ind. Electron.* **2015**, *62*, 7228–7236. [[CrossRef](#)]
31. Chaudhary, R.; Bhatia, R.S. A Grid Connected Five Level Cascaded Multilevel Inverter for PV Systems. In Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOE, Tirunelveli, India, 11–12 May 2018; pp. 488–492.
32. Zheng, H.; Wang, R.; Xu, W.; Wang, Y.; Zhu, W. Combining a HMM with a genetic algorithm for the fault diagnosis of photovoltaic inverters. *J. Power Electron.* **2017**, *17*, 1014–1026.
33. Wang, L.L.; Pei, F.; Zhu, Y.L. Transformer Fault Diagnosis Based on Online Sequential Extreme Learning Machine. *Appl. Mech. Mater.* **2014**, *721*, 360–365. [[CrossRef](#)]
34. Krishnakumari, A.; Elayaperumal, A.; Saravanan, M.; Arvindan, C. Fault diagnostics of spur gear using decision tree and fuzzy classifier. *Int. J. Adv. Manuf. Technol.* **2017**, *89*, 3487–3494. [[CrossRef](#)]
35. Miao, B.; Shen, Y.; Wu, D.; Zhao, Z. Three level inverter fault diagnosis using EMD and support vector machine approach. In Proceedings of the 2017 12th IEEE Conference on Industrial Electronics and Applications, ICIEA, Siem Reap, Cambodia, 18–20 June 2017; pp. 1595–1598.
36. The Detection of Archaeological Residues Using Remote-Sensing Techniques (DART) Project. DART Weather Data. Available online: http://dartportal.leeds.ac.uk/dataset/dart_monitoring_weather_data (accessed on 8 March 2018).
37. Deng, X.; Tian, X.; Chen, S.; Harris, C.J. Nonlinear Process Fault Diagnosis Based on Serial Principal Component Analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 560–572. [[CrossRef](#)]
38. Deng, F.; Yang, S.; Liu, Y.; Liao, Y.; Ren, B. Fault Diagnosis of Rolling Bearing Using the Hermitian Wavelet Analysis, KPCA and SVM. In Proceedings of the 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control, SDPC, Shanghai, China, 16–18 August 2017; pp. 632–637.
39. Ramesh Babu, N.; Jagan Mohan, B. Fault classification in power systems using EMD and SVM. *Ain Shams Eng. J.* **2017**, *8*, 103–111. [[CrossRef](#)]
40. Miao, Q.; Zhang, X.; Liu, Z.; Zhang, H. Condition multi-classification and evaluation of system degradation process using an improved support vector machine. *Microelectron. Reliab.* **2017**, *75*, 223–232. [[CrossRef](#)]
41. Njok, A.O.; Ogbulezie, J.C. The Effect of Relative Humidity and Temperature on Polycrystalline Solar Panels Installed Close to a River. *Phys. Sci. Int. J.* **2019**, *20*, 1–11. [[CrossRef](#)]

