

Article

A Method Based on NGO-HKELM for the Autonomous Diagnosis of Semiconductor Power Switch Open-Circuit Faults in Three-Phase Grid-Connected Photovoltaic Inverters

Xiao Xu , Teng Zhang, Ziwen Qiu, Hui Gao and Sanli Zhu *

College of Automation, Nanjing University of Posts and Telecommunications, Wenyuan Road 9, Nanjing 210023, China; xu.xiao@njupt.edu.cn (X.X.); teng_zhang_sg@sina.com (T.Z.); ziwen_qiu@sina.com (Z.Q.); hui_ga@163.com (H.G.)

* Correspondence: xux2013@gmail.com

Abstract: With accelerating grid decarbonization and technological breakthroughs, grid-connected photovoltaic (PV) systems are continuously connected to distribution networks at all voltage levels. As the grid interaction interfaces between PV panels and the distribution network, PV inverters must operate flawlessly to avoid energy and financial losses. As the failure of semiconductor switches is the leading cause of abnormal operation of PV inverters and typically cannot be detected by internal protection circuits, this paper aims to develop a method for the autonomous diagnosis of semiconductor power switch open-circuit faults in three-phase grid-connected PV inverters. In this study, a ReliefF-mRMR-based multi-domain feature selection method is designed to ensure the completeness of the fault characteristics. An NGO-HKELM-based classification method is proposed to guarantee the desired balance between generalization and exploration capability. The proposed method overcomes the common problems of poor training efficiency and imbalances between generalization and exploration capabilities. The performance of the proposed method is verified with the detection of switch OC faults in a three-phase H-bridge inverter and neutral-point-clamped inverter, with diagnostic accuracy of 100% and 99.46% respectively.

Keywords: photovoltaic inverters; fault detection; fault localization; maximum relevance-minimum redundancy; hybrid kernel extreme learning machine; northern goshawk optimization



Citation: Xu, X.; Zhang, T.; Qiu, Z.; Gao, H.; Zhu, S. A Method Based on NGO-HKELM for the Autonomous Diagnosis of Semiconductor Power Switch Open-Circuit Faults in Three-Phase Grid-Connected Photovoltaic Inverters. *Sustainability* **2023**, *15*, 9588. <https://doi.org/10.3390/su15129588>

Academic Editor: Cristina Ventura

Received: 12 April 2023

Revised: 22 May 2023

Accepted: 12 June 2023

Published: 14 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid growth of the global economy, human society has demonstrated drastically increasing energy demand. Using conventional fossil energy sources results in high CO₂ emissions, which not only accelerates global warming but also has detrimental impacts on biodiversity, ecosystems, and a sustainable human living environment [1,2]. To alleviate the severe climate problem, countries worldwide are accelerating the replacement of fossil fuels with inexhaustible renewable energies such as wind and solar [3,4]. As opposed to wind power, whose exploration requires an open area rich in wind all year around, distributed solar energy systems pose less requirements on the installation place. They can be mounted on rooftops/walls of buildings including residential houses, factories, shopping malls, etc. [5–7]. A grid-tied solar PV system is connected to the power grid via a dedicated PV inverter that can track the maximum power point of PV arrays. As the grid-interaction interface, the stable operation of PV inverters is crucial for avoiding potential profit losses and potential interference in the power grid [8]. According to a survey of a PV plant, PV inverter failures can account for 21% of unscheduled maintenance events [9]. The failure of a PV inverter is mostly due to the failure of its subsystems or components (e.g., power switches, diodes, capacitors, etc.) [10]. Based on a questionnaire survey conducted with different semiconductor manufacturers, it was reported that the leading causes of power converter failure are harsh environmental conditions, system transients, heavy load or

overload, and others (e.g., manufacturing design defects and abnormal power cycles). In contrast, the most vulnerable component inside a power converter is the semiconductor device, including the high-frequency power switches and diodes, as shown in Figure 1 [11]. Therefore, appropriate fault diagnosis methods have to be proposed to ensure the safe and reliable operation of PV power systems.

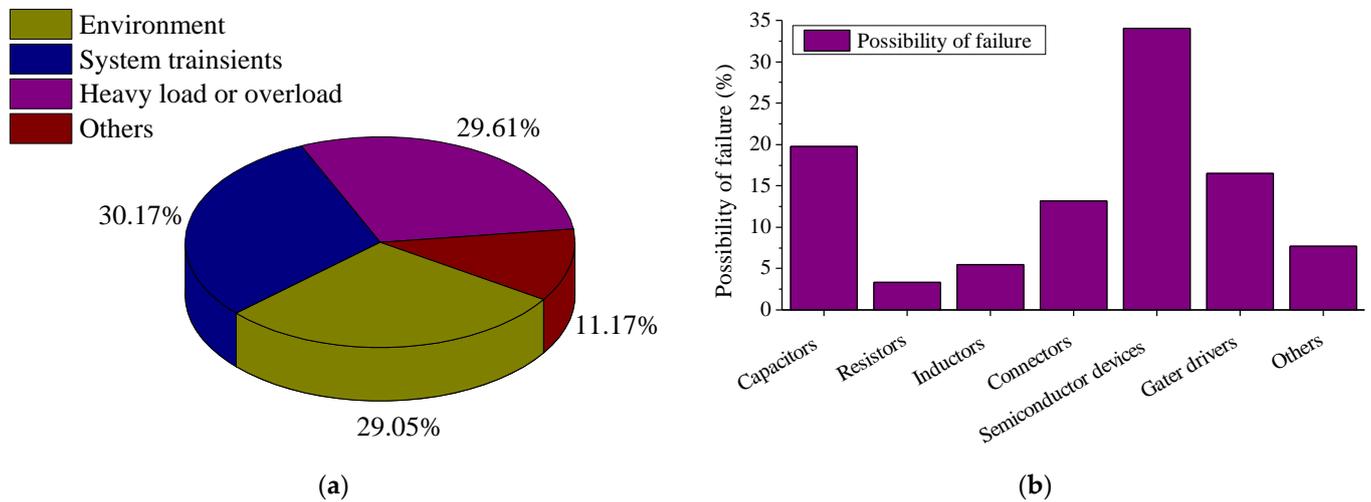


Figure 1. Potential causes of power converter failure and the possibility of failure for vulnerable power converter components: (a) potential causes of power converter failure; (b) the possibility of failure for vulnerable power converter components.

As the failure of semiconductor power switches causes most inverter faults, this paper focuses on investigating methods for the detection of power switch failures. When a grid-tied PV inverter has a power switch failure, its grid-side voltage waveform is unaffected. At the same time, its output AC current typically exhibits a unique time-domain waveform shape that can be extracted as the label of the particular fault [12,13]. The PV inverter fault diagnosis process can be divided into two steps: fault feature extraction and fault diagnosis. Fault feature extraction uses sensor devices to obtain the fault signals that can be monitored, such as current, voltage, power, etc., record the fault waveforms, and extract the fault signals' multi-domain features via signal processing techniques. The fault feature extraction techniques usually include the state observer method, parameter estimation method, vector analysis method, spectrum analysis method, wavelet analysis method, principal element analysis method, empirical modal decomposition method, etc. For example, Clarke transformation is applied in ref. [14] to extract the voltage pattern of a three-phase faulty inverter on the Alpha-Beta plane. In ref. [15], a high-frequency voltage signal is injected into the Alpha-Beta plane, with the induced current response applied to locate the fault of a three-phase H-bridge inverter. In refs. [16,17], the wavelet transform technique is applied to identify, separate and group the low-frequency and high-frequency components of a PV inverter's fault signals. Ref. [18] proposed a fault diagnosis strategy for three-phase H-bridge inverters that applies empirical mode decomposition to the inverter output current signals, then calculates the variance contribution of each eigenmode and selects the dominant eigenmodes as the open-circuit fault features. The selection of fault features requires the consideration of diagnostic time, accuracy, the extraction difficulty, etc., and meeting the target device's fault diagnosis needs.

Fault diagnosis is often combined with machine learning (ML) algorithms to automatically classify and identify faults. Typical methods include neural networks, Bayesian networks, support vector machines, fuzzy logic reasoning, data mining, and expert experience. In ref. [19], a back propagation neural network (BPNN) is applied to diagnose the fault types of a PV inverter based on its output current characteristics. Ref. [20] offers a Bayesian classifier-based fault diagnosis method for a three-phase inverter where the

grid-side voltages measured for different fault modes are used as the training data. In ref. [21], a secondary classification fault diagnosis strategy based on PCA-SVM has been proposed to distinguish two groups of similar faults that occur in cascaded PV inverters. To improve the learning speed of feedforward neural networks, extreme learning machine (ELM) is proposed with proven performance in multi-class classification [22]. To further enhance the non-linear handling capability of ELM, kernels of different types are integrated into ELM to form the hybrid kernel ELM (HKELM). However, the generalization and exploration capability of HKELM is strongly affected by the kernel types and HKELM parameters [23]. To further enhance the performance and stability of HKELM in anomaly detection, this paper combines HKELM with an advanced swarm-based algorithm named northern goshawk optimization (NGO). In this way, the parameters of kernels and ELM will be optimized to pursue the best performance of HKELM. The advantages and disadvantages of the ML-based fault diagnosis techniques commonly applied to PV inverters are tabulated in Table 1. As summarized in Table 1, the abovementioned methods can capture the complex relationship between fault features and fault modes to some extent. However, their fault diagnosis performance is highly dependent on the selected feature sets, the applied model structure, and hyperparameters. The widely applied experience-based model construction approach with suboptimal model settings tends to underperform the fault diagnosis model. When compared to the above-mentioned fault diagnosis methods, the main improvements of the proposed NGO-HKELM-based fault diagnosis method are as follows: (1) the completeness of fault information is captured by the proposed ReliefF-mRMR-based multi-domain fault feature selection method; (2) the balance between the generalization and exploration capabilities is ensured by the hybrid kernels; (3) optimal hyperparameters are guaranteed by the NGO optimizer.

Table 1. Advantages and disadvantages of the ML-based fault diagnosis techniques commonly applied to PV inverters.

ML-Based Fault Diagnosis Technique	Advantages	Disadvantages
BP neural network (e.g., [19])	(1) simple network structure; (2) easy to implement; (3) no requirement for parameter tuning;	(1) slow convergence speed; (2) local minimum; (3) sensitive to initial values.
Bayesian networks (e.g., [20])	(1) simple network structure; (2) can handle multi-class problems; (3) excellent performance for small-scale problems.	(1) high computational cost; (2) difficult to implement; (3) performance deteriorates with large-scale problems.
SVM (e.g., [24])	(1) can deal with nonlinear data via kernel; (2) good generalization capability; (3) can handle high-dimensional data.	(1) computationally intensive; (2) performance affected by the choice of kernel; (3) limited to two-class problems (but can be indirectly applied to multi-class problems).
Decision trees (e.g., [25])	(1) simple structure; (2) easy to interpret and implement; (3) can handle both outliers and missing values.	(1) overfitting can occur; (2) biased decisions can occur; (3) structure of decision trees is sensitive to the data.
Proposed NGO-HKELM	(1) improved global and local optimization ability in anomaly detection; (2) good generalization capability; (3) fast and efficient learning speed; (4) optimized HKELM parameters; (5) can handle large multi-class problems.	(1) increased computational costs; (2) performance is affected by the predefined hybrid kernels.

Inspired by the existing research, this paper proposes a novel fault diagnosis method for reliably and precisely identifying faults in three-phase grid-connected PV inverters due to failures of their semiconductor power switches. The main contributions of this research are summarized below:

- (i) As PV inverter faults are mainly caused by the failure of semiconductor power switches such as IGBTs and MOSFETs, this paper thoroughly investigates the fault operation modes and features of conventional three-phase grid-connected PV inverters due to the failure of semiconductor power switches.
- (ii) As the characteristics of PV inverter faults are weak, multi-domain fault feature indices are considered, and a novel ReliefF-mRMR-based feature set extraction method is proposed to find the most appropriate fault feature set.
- (iii) A novel NGO-HKELM-based fault diagnosis technique is proposed where the hyperparameters of HKELM can be internally optimized without relying on the experience-based hyperparameter tuning process. The results show that the proposed fault diagnosis technique performs outstandingly in the fault diagnosis of three-phase PV inverters.

The paper is structured as follows: Section 2 starts with a review of conventional topologies and control strategies of three-phase grid-connected PV inverters, followed by an analysis of the fault operation modes and features. Afterwards, a ReliefF-mRMR-based multi-domain fault feature extraction method and an NGO-HKELM-based fault diagnosis technique are proposed. The performance of the proposed fault diagnosis method is evaluated in Section 3, and the main conclusions are summarized in Section 4.

2. Materials and Methods

2.1. Conventional Topologies and Control Circuits of Grid-Connected PV Inverters

As the interaction interfaces between PV arrays and the power grid, PV inverters convert the DC power generated from PV panels to an AC form that the power grid can adopt. It is imperative to ensure that PV inverters operate without interruption under cyclic power fluctuations and temperature deviations. As the topology and control circuits have a deterministic impact on the fault operation modes and multi-domain fault features of PV inverters, analysis of the conventional topologies and control circuits of three-phase grid-connected PV inverters is required. As the most widely used and well-proven PV inverter topologies, H-bridge inverters and neutral-point clamped (NPC) PV inverters are the basis for many improved and advanced PV inverter topologies, such as the cascaded H-bridge multilevel inverter [26,27], five-level NPC with self-balanced switched-capacitor [28], and seven-level active boost NPC inverter [29,30]. The general topologies of H-bridge inverters and NPC inverters are shown in Figure 2, and they are made up of DC-link capacitors, diodes and insulated gate bipolar transistor (IGBT) based power switching devices. H-bridge inverters are featured by simple structure and control circuit, easy implementation, and low investment cost. However, due to the low utilization rate of DC-link voltage, the AC output current of H-bridge inverters is rich in high-frequency harmonics, resulting in higher requirements for grid-side EMI filters [31]. Accordingly, H-bridge inverter topology is mainly for low-power PV applications (e.g., residential grid-connected PV systems) that do not have stringent power quality requirements. NPC inverter topology is introduced to improve the DC-link voltage utilization ratio and alleviate the voltage stress on power switches, for which its output voltage has multiple levels [32]. Although this configuration is capable of reducing harmonics in both the output voltage and current, the NPC inverter topology is usually applied for medium-power PV applications due to the doubled number of power switches, additional clamping diodes (as opposed to H-bridge inverter topology), complicated control circuit design, and potential voltage deviation at the neutral point [33].

Apart from the topology of the inverter, its control circuit directly affects the fault operation modes and features. As illustrated in Figure 3, the control circuit of a PV inverter can be divided into four parts: the DC-link voltage control, grid synchronization, current control and pulse width modulation (PWM). The DC-link voltage control maintains the DC-link voltage at its reference value, while the grid synchronization transforms the three-phase (ABC) signal into a dq0 rotating frame (to facilitate the design of the control system). The current control is to inject a pure sinusoidal current into the grid while maintaining the system operating at a unity power factor. Finally, the gate signals for the

power switches are obtained from the PWM technique. The four parts are correlated, as in Figure 4, together constituting a three-phase PV inverter's outer voltage and inner current closed-loop control [31,34]. As the gate signals for high-frequency power switches are generated from the control circuit, the diversity of control strategies can result in different fault operation modes and features. A review of the state-of-the-art control strategies applied to three-phase PV inverters can be found in [34].

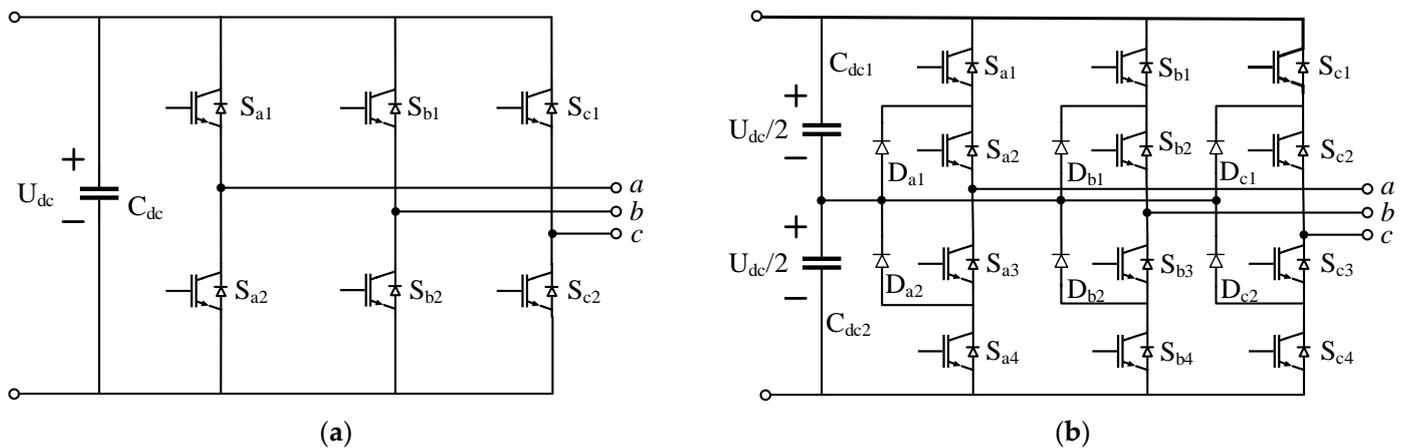


Figure 2. Typical topologies of three-phase grid-connected PV inverters: (a) H-bridge PV inverters; (b) neutral-point clamped PV inverters.

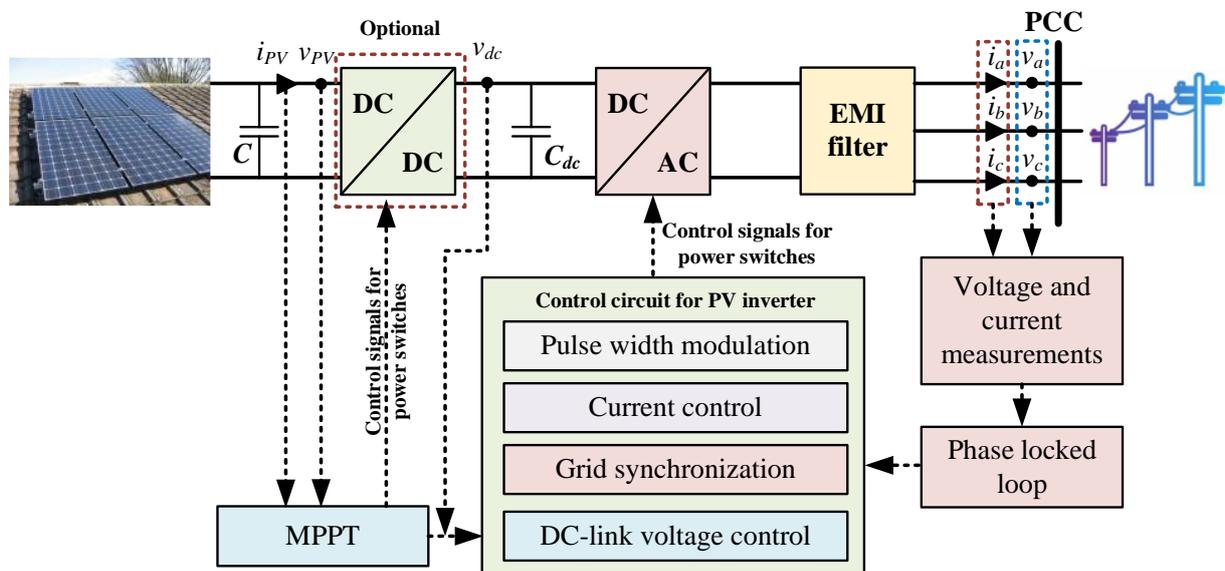


Figure 3. Illustration of the control circuit structure applied to a three-phase PV inverter.

2.2. Analysis of the Fault Operation Modes and Features of Grid-Connected PV Inverters

An analysis of the fault operation modes and multi-domain characteristics is required to select appropriate fault features for three-phase grid-connected PV inverters. As the failures of semiconductor switches are the most likely causes of PV inverter faults, this section focuses on investigating the fault operation modes and features of three-phase PV inverters with failures of semiconductor switches. The faults of semiconductor switches can be divided into short-circuit and open-circuit faults. When a short-circuit fault of a semiconductor switch occurs, it is often difficult to diagnose since the fault exists for a very short time (within $10 \mu\text{s}$). In most PV inverter circuit designs, the short-circuit detection and overcurrent protection circuits are generally integrated into the driver circuit. Moreover,

the short-circuit faults of semiconductor switches cannot last for an extended period and are usually converted into open-circuit faults in the form of permanent damage to switch modules. Unlike short-circuit faults, when an open-circuit fault occurs in a power switch module, the system does not immediately fail. Still, it works in an abnormal operating condition that is likely to cause secondary faults in connected devices, resulting in more significant losses. Therefore, only the open-circuit (OC) faults of semiconductor switches are considered in this paper. In addition, as the DC-link voltage is ideally maintained at its reference value with negligible high-frequency ripples superimposed, the DC-link voltage is represented as a controllable DC voltage source in the following subsections.

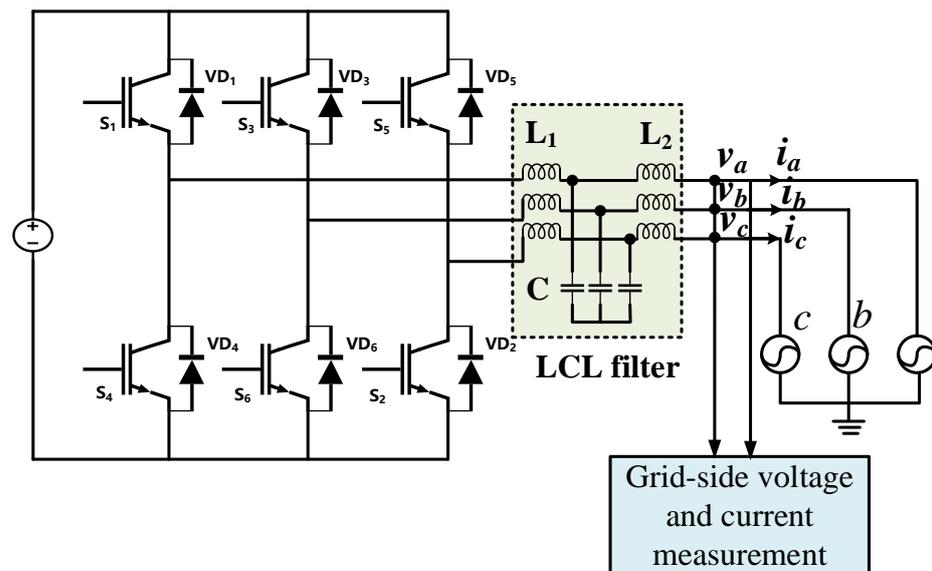


Figure 4. Illustration of a simplified three-phase H-bridge inverter.

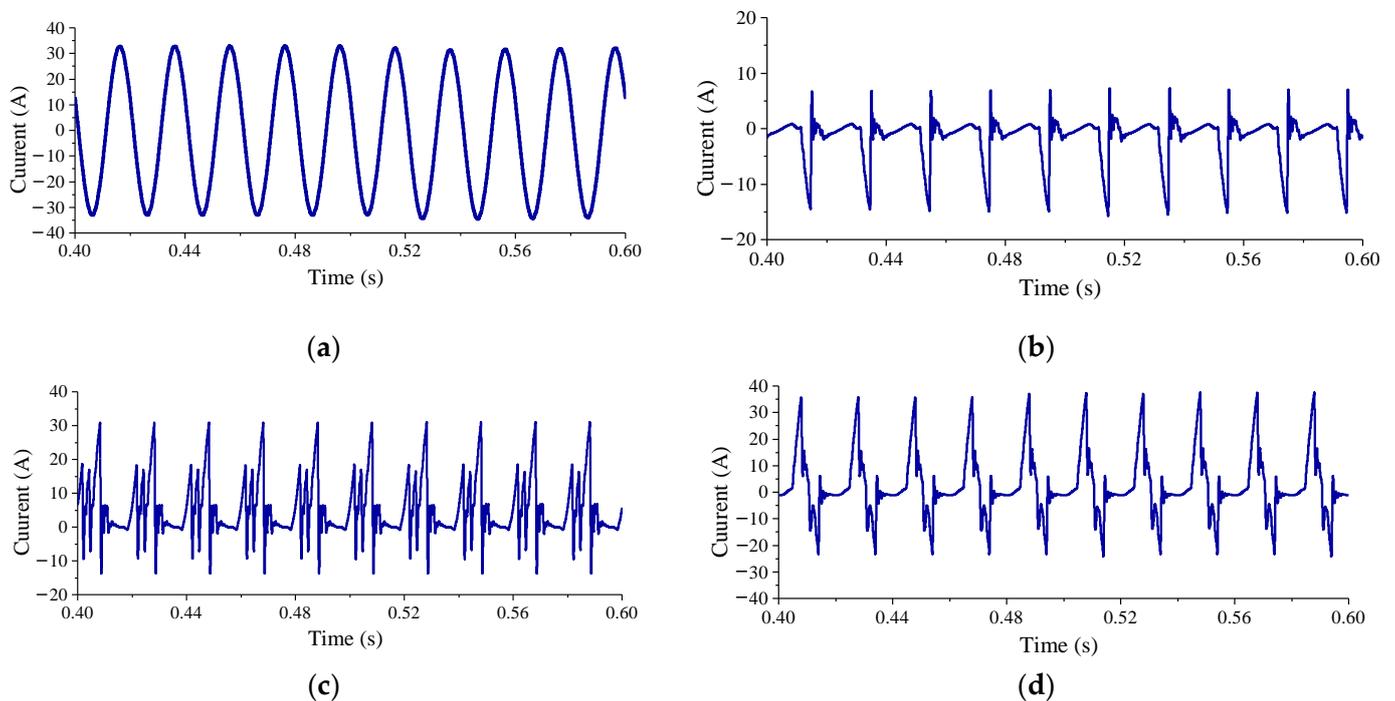
2.2.1. Fault Operation Modes and Features of a Three-Phase H-Bridge Inverter

As a conventional power conversion module in distributed PV applications, the three-phase H-bridge inverter is vital in ensuring the regular operation of the whole solar PV system. The general topology of a three-phase H-bridge inverter circuit using IGBTs as power tubes is shown in Figure 4 and can be regarded as a system made up of three identical single-phase half-bridge inverter circuits. In Figure 4, S_1 – S_6 are IGBTs and VD_1 – VD_6 are current-continuing diodes.

In the operation of a three-phase H-bridge inverter with IGBTs as power tubes, the probability of more than two IGBTs failing simultaneously is very low and can result in the immediate collapse of the whole system. Accordingly, it is assumed that at most two IGBTs may fail at the same time. Such fault cases can be divided into four types: (i) normal operation state, (ii) single IGBT failure, (iii) the two failed IGBTs located on the same bridge arm, and (iv) the two failed IGBTs located on different bridge arms, as tabulated in Table 2. In Table 2, S_x indicates that the IGBT labeled S_x alone has an OC fault, while $S_x S_y$ suggests that the two IGBTs labeled S_x and S_y have OC faults simultaneously. It is observed from Table 2 that there are a total of 22 different fault modes. The grid-side current waveforms of a three-phase H-bridge PV inverter under four types of OC faults are illustrated in Figure 5. It is observed that the three-phase H-bridge inverter can still operate under the considered OC faults of the IGBTs and exhibit a diversified waveform shape of the inverter grid-side current. As the fault currents are rich in high-frequency harmonics, using frequency-domain feature indices may facilitate fault diagnosis.

Table 2. Fault classification of three-phase H-bridge inverter.

Types of OC Faults	Failure Situation	Faulty Device Name	Number of Fault Modes
Type I	Normal operation state	/	1
Type II	Single IGBT failure	$S_1, S_2, S_3, S_4, S_5, S_6$	6
Type III	The two failed IGBTs located on the same bridge arm	S_1S_4, S_3S_6, S_2S_5	3
Type IV	The two failed IGBTs located on different bridge arms	$S_1S_2, S_1S_3, S_1S_5, S_1S_6, S_2S_3, S_2S_4, S_2S_6, S_3S_4, S_3S_5, S_4S_5, S_4S_6, S_5S_6$	12

**Figure 5.** Illustration of the grid-side current waveform of a three-phase H-bridge inverter under different types of OC faults: (a) Type I; (b) Type II; (c) Type III; (d) Type IV.

2.2.2. Fault Operation Modes and Features of a Three-Phase NPC Inverter

A distributed solar PV system based on a three-phase neutral point clamped (NPC) inverter is shown in Figure 6. As opposed to the three-phase H-bridge inverter, the number of IGBTs used by an NPC inverter is doubled, resulting in higher fault risks and increased complexity of switch fault diagnosis.

As the NPC inverter is three-phase symmetrical, its operating states can be analyzed on a single inverter bridge arm. As illustrated in Figure 7, the NPC inverter has three operating states (the P state, O state and N state in Table 3). The solid blue line represents the positive direction of the current, and the dashed red line represents the negative direction of the current. By neglecting the voltage drop across the on-state switches, S_{a1} and S_{a2} are turned on in the P state, and S_{a3} and S_{a4} are turned off. The potential of point A is always equal to that of point P, and the output voltage is $U_d/2$. In the O state, S_{a2} and S_{a3} are turned on, and S_{a1} and S_{a4} are turned off. The potential of point A is always equal to point O's, and the output voltage is 0. In the N state, S_{a3} and S_{a4} are turned on, and S_{a1} and S_{a2} are turned off. The potential of point A is always equal to point N's, and the output voltage is $-U_d/2$. The output voltage and the corresponding operating states are given in Table 3.

The inverter has three levels of output voltage: $U_d/2$, 0, and $-U_d/2$, and it is called a three-level inverter. Moreover, S_{a2} and S_{a3} switch on and off more frequently and for longer periods than the other switches, resulting in higher fault risks. The switching states of S_{a1} and S_{a3} and S_{a2} and S_{a4} are always complementary, and S_{a1} and S_{a4} cannot be turned on simultaneously. In addition, from the perspective of control methods, there will never be a direct switch between the P and N states, and there must be an O state for transition. Therefore, the devices' voltage is only half of the DC voltage, introducing a high safety margin.

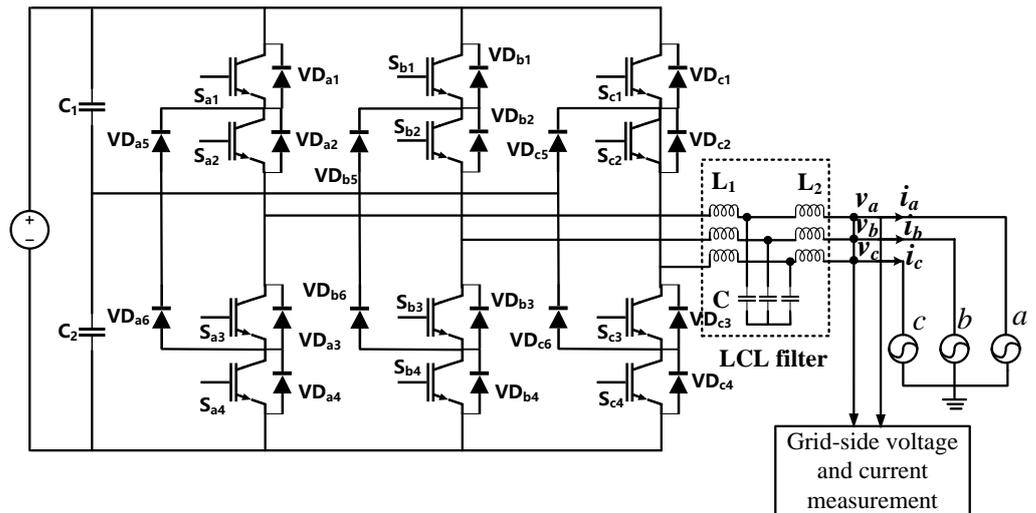


Figure 6. Illustration of a simplified three-phase NPC inverter.

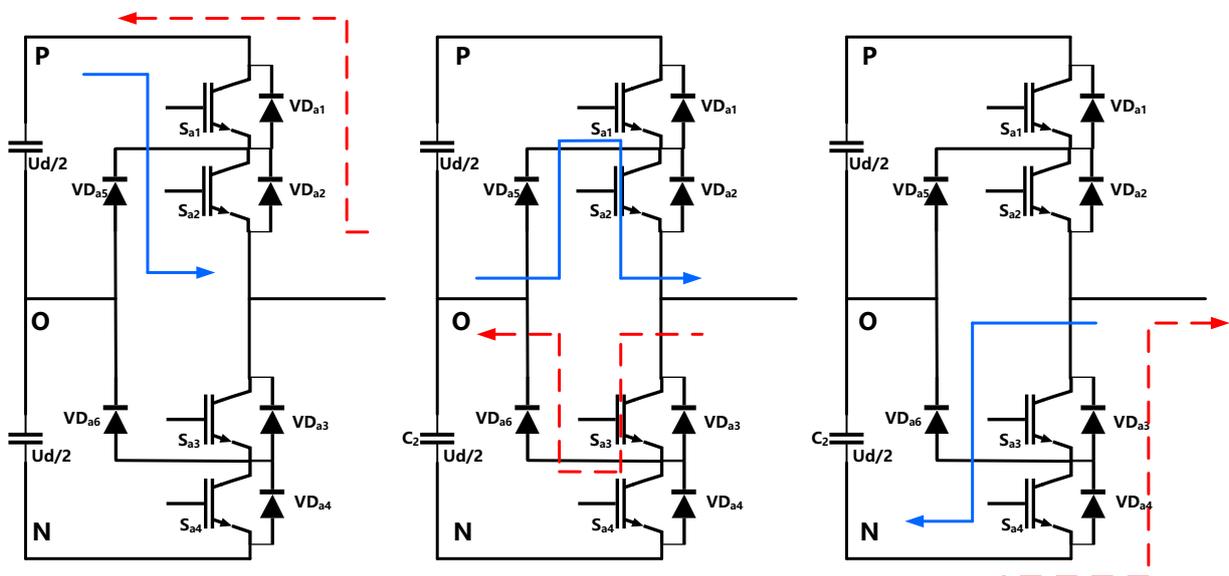


Figure 7. Illustration of the three operating states of an NPC inverter (red and blue arrows indicate the positive and negative current flow directions, respectively).

Table 3. Output voltage and operating states of a three-phase NPC inverter.

Operating State	S_{a1}	S_{a2}	S_{a3}	S_{a4}	Output Voltage
P state	ON	ON	OFF	OFF	$U_d/2$
O state	OFF	ON	ON	OFF	0
N state	OFF	OFF	ON	ON	$-U_d/2$

As a three-phase NPC inverter has twice the number of IGBTs compared to a three-phase H-bridge inverter, its potential fault cases also become more complicated. When the number of IGBT device faults on a particular bridge arm of the NPC inverter exceeds two, the bridge arm has no output at all, and the PV system immediately stops working due to internal protection. Therefore, only the fault cases where the number of failed IGBTs does not exceed two are considered here. As in Table 4, the OC faults of the PV system can be divided into four types: the first type is the normal operation of the NPC inverter, in which the IGBT devices are operating smoothly. The second type is the single IGBT failure, resulting in 12 fault cases depending on where the faulty IGBT is located. The single OC fault of an IGBT can be represented as S_{an} ($n = 1, 2, 3, 4$). The third type refers to the scenario of two failed IGBTs located on the same bridge arm, which make up 12 fault modes. The fourth type is the scenario of two failed IGBTs located on different bridge arms, constituting 48 fault modes. As there are 73 fault modes, appropriate feature indices and fault classifiers are required. The grid-side current waveforms of a conventional three-phase NPC inverter under four types of OC faults are illustrated in Figure 8. It is noticed that Type III and Type IV OC faults have very similar time-domain waveform shapes, indicating the requirement for proper frequency-domain fault feature indices.

Table 4. Fault classification of three-phase NPC inverter.

Types of OC Faults	Failure Situation	Faulty Device Name	Number of Fault Modes
Type I	Normal operation state	/	1
Type II	Single IGBT failure	$S_{a1}, S_{a2}, S_{a3}, S_{a4},$ $S_{b1}, S_{b2}, S_{b3}, S_{b4},$ $S_{c1}, S_{c2}, S_{c3}, S_{c4}$	12
Type III	Two failed IGBTs located on the same bridge arm	$S_{a1}S_{a3}, S_{a1}S_{a4}, S_{a2}S_{a3}, S_{a2}S_{a4},$ $S_{b1}S_{b3}, S_{b1}S_{b4}, S_{b2}S_{b3}, S_{b2}S_{b4},$ $S_{c1}S_{c3}, S_{c1}S_{c4}, S_{c2}S_{c3}, S_{c2}S_{c4},$	12
Type IV	Two failed IGBTs located on different bridge arms	$S_{a1}S_{b1}, S_{a1}S_{b2}, S_{a1}S_{b3}, S_{a1}S_{b4},$ $S_{a1}S_{c1}, S_{a1}S_{c2}, S_{a1}S_{c3}, S_{a1}S_{c4},$ $S_{a2}S_{b1}, S_{a2}S_{b2}, S_{a2}S_{b3}, S_{a2}S_{b4},$ $S_{a2}S_{c1}, S_{a2}S_{c2}, S_{a2}S_{c3}, S_{a2}S_{c4},$ $S_{a3}S_{b1}, S_{a3}S_{b2}, S_{a3}S_{b3}, S_{a3}S_{b4},$ $S_{a3}S_{c1}, S_{a3}S_{c2}, S_{a3}S_{c3}, S_{a3}S_{c4},$ $S_{a4}S_{b1}, S_{a4}S_{b2}, S_{a4}S_{b3}, S_{a4}S_{b4},$ $S_{a4}S_{c1}, S_{a4}S_{c2}, S_{a4}S_{c3}, S_{a4}S_{c4},$ $S_{b1}S_{c1}, S_{b1}S_{c2}, S_{b1}S_{c3}, S_{b1}S_{c4},$ $S_{b2}S_{c1}, S_{b2}S_{c2}, S_{b2}S_{c3}, S_{b2}S_{c4},$ $S_{b3}S_{c1}, S_{b3}S_{c2}, S_{b3}S_{c3}, S_{b3}S_{c4},$ $S_{b4}S_{c1}, S_{b4}S_{c2}, S_{b4}S_{c3}, S_{b4}S_{c4}$	48

2.3. ReliefF-mRMR-Based Multi-Domain Fault Feature Selection Method

This section introduces the ReliefF-mRMR-based multi-domain feature selection method to define the optimal feature set that will be applied for the next-stage fault classification and detection. It starts with an introduction of the considered multi-domain features, followed by a comparative analysis of the proposed ReliefF-mRMR-based feature selection method.

2.3.1. Fault Feature Indices

Before applying artificial intelligence-based fault detection algorithms, it is crucial to identify and extract appropriate fault feature indices. Fault feature indices can be acquired either in the time domain or in the frequency domain [9]. Specifically, considered time-domain feature indices include impulse indicator, crest factor, shape indicator, Kurtosis, skewness and clearance factor. The considered frequency-domain feature indices are further divided into type I indices describing the featured frequency and type II indices

describing the distribution patterns of frequency components. Type I frequency-domain feature indices include spectral centroid, mean square (MS) frequency, root-mean-square (RMS) frequency, variance frequency and root variance (RV) frequency. Type II frequency-domain feature indices include the mean of spectral kurtosis (SK), standard deviation (SD) of SK, skewness of SK and kurtosis of SK. The formulas for the abovementioned feature indices are tabulated in Table 5. The next-step fault feature set selection will consider all of the above-mentioned feature indices. It should be noted that for a three-phase signal, the feature indices will be calculated for each phase.

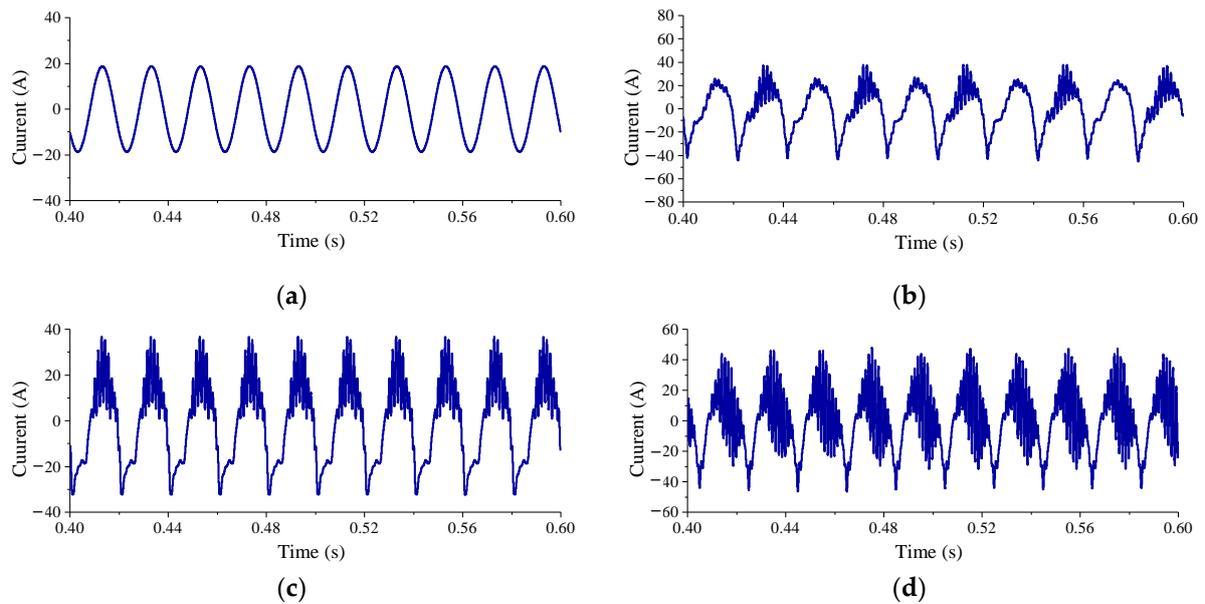


Figure 8. Illustration of the grid-side current waveform of a three-phase NPC inverter under different types of OC faults: (a) Type I; (b) Type II; (c) Type III; (d) Type IV.

Table 5. Considered multi-domain fault feature indices.

Feature Name	Formula	Feature Name	Formula
Time-domain features	impulse factor $IF = \frac{\max(S_k)}{\mu}$	Frequency-domain features-Type I	spectral centroid $SC = \frac{\sum_{m=1}^M f_m E_m}{\sum_{m=1}^M E_m}$
	crest factor $CF = \frac{\max(S_k)}{\sqrt{\frac{1}{n} \sum_{k=1}^n S_k^2}}$		MS frequency $MSF = \frac{\sum_{m=1}^M f_m^2 E_m}{\sum_{m=1}^M E_m}$
	shape factor $SF = \frac{\sqrt{\frac{1}{n} \sum_{k=1}^n S_k^2}}{\mu}$		RMS frequency $RMSF = \sqrt{MSF}$
	kurtosis $KT = \frac{\frac{1}{n} \sum_{k=1}^n (S_k - \mu)^4}{\left(\sqrt{\frac{1}{n} \sum_{k=1}^n (S_k - \mu)^2}\right)^4}$		variance frequency $VF = \frac{\sum_{m=1}^M (f_m - SC)^2 E_m}{\sum_{m=1}^M E_m}$
	skewness $SW = \frac{\frac{1}{n} \sum_{k=1}^n (S_k - \mu)^3}{\left(\sqrt{\frac{1}{n} \sum_{k=1}^n (S_k - \mu)^2}\right)^3}$		RV frequency $RVF = \sqrt{VF}$
	clearance factor $CRF = \frac{\max(S_k)}{\left(\frac{1}{n} \sum_{k=1}^n \sqrt{ S_k }\right)^2}$		Mean of SK $\mu_{SK} = \frac{1}{M} \sum_{m=1}^M K_m$
		SD of SK $\sigma_{SK} = \sqrt{\frac{1}{M} \sum_{m=1}^M (K_m - \mu_{SK})^2}$	
		skewness of SK $SW_{SK} = \frac{\frac{1}{n} \sum_{m=1}^M (K_m - \mu_{SK})^3}{\left(\sqrt{\frac{1}{n} \sum_{m=1}^M (K_m - \mu_{SK})^2}\right)^3}$	
		kurtosis of SK $KT_{SK} = \frac{\frac{1}{n} \sum_{m=1}^M (K_m - \mu_{SK})^4}{\left(\sqrt{\frac{1}{n} \sum_{m=1}^M (K_m - \mu_{SK})^2}\right)^4}$	

2.3.2. ReliefF-mRMR-Based Fault Feature Selection

To find the unique pattern of a specific fault signal (so that the fault classifier can detect it), an appropriate feature set has to be defined from the diversified time-domain and frequency-domain feature indices introduced in Section 2.3.1. To acquire the most suitable feature set from the numerous random combinations of features, feature selection techniques are required. Various feature selection techniques have been proposed in the existing literature, including principal component analysis (PCA) [35], the ReliefF [36] and the minimum redundancy maximum relevance (mRMR) [37], Pearson correlation, Euclidean distance, the cosine coefficient [38], etc. Among all these techniques, the ReliefF and mRMR methods are still the most widely used feature selection approaches. Specifically, the ReliefF method ranks the features based on their calculated weights (indicating the importance of features) [39,40]. The ReliefF method first sets the weights of all features to zero and iteratively selects a random sample S_i . For each sample S_i , the ReliefF method rewards features that give different values to nearby samples of other classes and penalizes features that give different values to nearby samples of the same class [40]. The main drawback of the ReliefF method is that it only considers the correlations between features and the sample classes without considering the correlations among the features, resulting in the redundancy of selected features and a suboptimal feature set. Unlike the ReliefF method, the mRMR method ranks features according to the mutual information (MI) of features. Feature with the highest relevance (i.e., largest MI) to the target sample class and the lowest correlations (i.e., smallest MI) with the other features will be ranked in priority order [40]. Although the mRMR method considers the relationships among features, it treats all features equally without considering the fact that different features can have distinct contributions to the target sample class.

To tackle the inherent disadvantages of the ReliefF and mRMR methods and achieve their complementary advantages, this paper applies a ReliefF-mRMR-based fault feature selection method. It first applies the ReliefF method to calculate the weights of all available features and ranks features based on their weights. Afterwards, features with weights below predefined threshold values are discarded. The remaining features are then reranked with the mRMR method, and the first m features are selected from the optimal feature set (m is the required number of features). The whole procedure of the ReliefF-mRMR-based fault feature technique is illustrated in Figure 9.

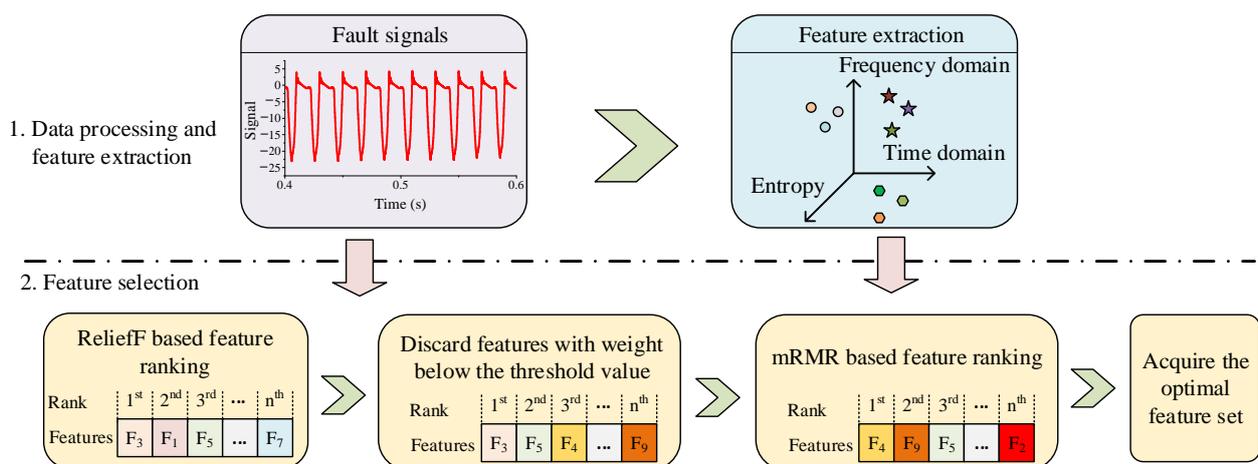


Figure 9. Illustration of the ReliefF-mRMR-based fault feature selection approach.

Figure 10 compares the accuracy among the ReliefF-mRMR, ReliefF, and mRMR methods in a BPNN-based fault diagnosis of a three-phase H-bridge inverter. The number of features gradually increases from 3 to 36, with a step of 3. It turns out that ReliefF has better accuracy than mRMR when the number of selected features is small. When the

number of selected features is above a certain value, mRMR has better accuracy than ReliefF. ReliefF-mRMR combines the merits of both methods and achieves the highest accuracy.

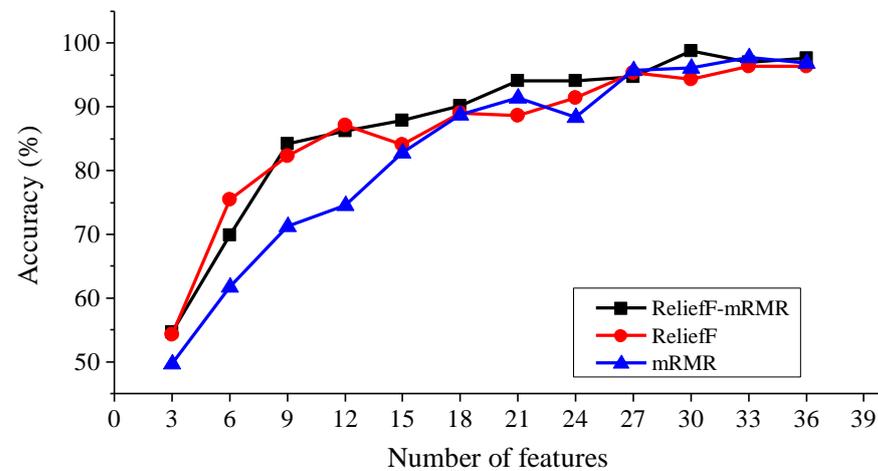


Figure 10. Comparison of the accuracy of the ReliefF-mRMR method, the ReliefF method and the mRMR method in a BPNN-based fault diagnosis of a three-phase H-bridge inverter.

2.4. NGO-HKELM-Based Fault Diagnosis Method

Once the fault feature set is defined, the next step is to select an appropriate fault diagnosis method to detect any abnormal operation patterns of PV inverters and locate the faulty power electronic components. This section proposes an NGO-HKELM-based fault diagnosis method to enhance the stability and classification capability of the conventional KELEM-based fault diagnosis method.

2.4.1. NGO-Based Optimization Technique

The NGO-based optimization technique is a novel swarm intelligence-based algorithm proposed by Mohammad Dehghani, with competitive performance compared to other well-known metaheuristic algorithms such as PSO, GA, GWO, etc. [41,42]. The main idea behind the algorithm is to emulate the behavior of a northern goshawk while hunting and catching prey [42]. The hunting behavior of the northern goshawk can be further divided into two phases: the prey identification phase and the chase and escape phase. Different mathematical models are formulated for the two phases.

(a) Phase one: prey identification phase

During the first phase of its hunting behavior, the northern goshawk will randomly select a prey and attack it, reflecting the global exploration capability of the NGO algorithm. First, the positions of all northern goshawks (representing the populations) are initiated under the constraints of position boundaries. Then, for each northern goshawk, its position value X_i (representing the solution of the control variable) is updated by comparing the objective functions between the current position of the northern goshawk X_i and its randomly selected prey position PR_i . The behavior of phase one can be represented by (1)–(3).

$$PR_i = X_k, i = 1, 2, \dots, N; k = 1, 2, \dots, i - 1, i + 1, \dots, N \quad (1)$$

$$x_{i,j}^{up,s1} = \begin{cases} x_{i,j} + r(pr_{i,j} - Ix_{i,j}), & F_{PR_i} < F_i \\ x_{i,j} + r(x_{i,j} - pr_{i,j}), & F_{PR_i} \geq F_i \end{cases} \quad (2)$$

$$X_i = \begin{cases} X_i^{up,s1}, & F_i^{up,s1} < F_i \\ X_i, & F_i^{up,s1} \geq F_i \end{cases} \quad (3)$$

where: the prey position for i th northern goshawk, PR_i , is randomly selected initially, and its position value at j th dimension is represented by $pr_{i,j}$; k is a random integer number within $[1, N]$; $x_{i,j}$ is the position value at j th dimension, and its updated value is represented by $x_{i,j}^{up,s1}$; X_i refers to the position of i th northern goshawk, and its updated solution is represented by $X_i^{up,s1}$; the objective function values of PR_i , X_i and $X_i^{up,s1}$ are represented by F_{PR_i} , F_i and $F_i^{up,s1}$, respectively; r is a random number in $[0, 1]$ and I is a random number equaling 1 or 2.

(b) Phase two: the chase and escape phase

Once the target prey is located and attacked by the northern goshawk, it will start to escape under the chase by the northern goshawk. The basic assumptions made for the chase and the escape stage are: (1) the northern goshawk can eventually capture the target prey in almost any situation; (2) the chasing range of the northern goshawk is a cycle with center and radius of $x_{i,j}$ and R , respectively; (3) the chasing area is constantly shrinking with the increasing count of chasing behaviors (representing the increasing iteration number). In this stage, the exploitation capability of the algorithm is enhanced via the local search. The process of the chase and escape phase can be represented by (4)–(6).

$$x_{i,j}^{up,s2} = x_{i,j} + R(2r - 1)x_{i,j} \quad (4)$$

$$R = 0.02 \left(1 - \frac{it}{T} \right) \quad (5)$$

$$X_i = \begin{cases} X_i^{up,s2}, & F_i^{up,s2} < F_i \\ X_i, & F_i^{up,s2} \geq F_i \end{cases} \quad (6)$$

where: R is the radius of the search cycle; $X_i^{up,s2}$ is the updated position of i th northern goshawk, and its value at j th dimension is represented by $x_{i,j}^{up,s2}$; the objective function value of $X_i^{up,s2}$ is represented by $F_i^{up,s2}$; it and T refer to the current iteration number and the maximum iteration number, respectively.

2.4.2. HKELM-Based Fault Classification

As the randomly generated input weights and hidden bias of extreme learning machine (ELM) can result in strong fluctuation of classification accuracy, kernel extreme learning machine (KELM) is applied where the original sigmoid and Gaussian function-based mapping function are replaced by kernel functions with outstanding non-linear mapping capability [43,44]. The added kernel functions bring KELM improved performance on classification tasks [44]. KELM's typically applied kernel functions include polynomial kernel function, radial basis function (RBF) based kernel function, Gaussian kernel function, linear kernel function, etc. As the selection of kernel functions directly affects the classification accuracy of KELM, it is crucial to select the appropriate kernel function type [45]. For example, the polynomial kernel function features good generalization capability but relatively poor learning capability. In contrast, the linear kernel function features good learning capability but poor generalization capability. Therefore, to further improve the performance stability of KELM, HKELM is introduced, and the single kernel function of KELM is replaced by a hybrid kernel function that combines several different kernel functions. For example, for a hybrid kernel function with a mixture of polynomial and RBF kernel functions, its mathematical expression is represented by Equations (7)–(10). It should be noted that different selections of kernel parameters and the weighting coefficients of kernel functions directly affect the performance of the final HKELM-based fault classification. Specifically, the polynomial kernel function is a type of global kernel and is featured by prominent generalization capability and poor learning capability. In contrast, the RBF kernel is a type of local kernel with superior local exploration capability and poor

generalization capability. Therefore, fine-tuning with a weighing coefficient is crucial for achieving the desired balance between stability and classification performance.

$$K_{HKELM}(x, x_i) = c_1 K_{poly}(x, x_i) + c_2 K_{RBF}(x, x_i) \tag{7}$$

$$K_{poly}(x, x_i) = (x \cdot x_i + c_3)^d \tag{8}$$

$$K_{RBF}(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right) \tag{9}$$

$$c_1 + c_2 = 1 \tag{10}$$

where: $K_{HKELM}(x, x_i)$, $K_{poly}(x, x_i)$ and $K_{RBF}(x, x_i)$ refer to the hybrid kernel function, the polynomial kernel function and the RBF kernel function, respectively; c_3 and d are the kernel parameters of $K_{poly}(x, x_i)$, while σ is the kernel parameter of $K_{RBF}(x, x_i)$; c_1 and c_2 are the weighting coefficients of $K_{poly}(x, x_i)$ and $K_{RBF}(x, x_i)$, respectively, and are in the range of [0, 1].

The general structure of the HKELM-based fault classification algorithm is illustrated in Figure 11. For a sample set (X, Y) where X refers to the input feature sets of samples and Y refers to the output label sets of samples, the acquisition of output weights $\beta_{HKELM}(x)$ under HKELM can be expressed as (11), which will decide the output label. Apart from the kernel parameters and the weighting coefficients of kernel functions, the regularization coefficient, C, also affects the classification performance of HKELM. A large value of C can improve the accuracy of HKELM, whereas a small value of C can enhance the generalization capability of HKELM. Accordingly, it is imperative to define the optimal value set of kernel parameters, the kernel function weighting coefficients and the regularization coefficient to achieve harmonization among the generalization capability, performance stability, and accuracy. More details on the implementation procedure for HKELM can be found in [45–47].

$$\beta_{HKELM}(x) = \begin{bmatrix} K_{HKELM}(x, x_1) \\ K_{HKELM}(x, x_2) \\ \vdots \\ K_{HKELM}(x, x_n) \end{bmatrix} \left(\Omega_{HKELM} + \frac{I}{C} \right)^{-1} Y \tag{11}$$

where: Ω_{HKELM} refers to the hybrid kernel matrix, and I and C refer to the diagonal matrix and the regularization coefficient, respectively.

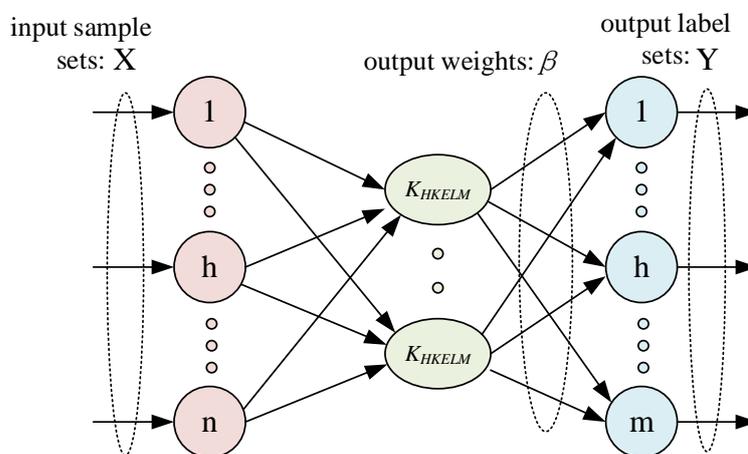


Figure 11. The general structure of HKELM.

2.4.3. An NGO-HKELM-Based Fault Diagnosis Technique

As discussed in Section 2.4.2, the performance of HKELM is mainly determined by kernel parameters, weighting coefficients of selected kernel functions and the regularization coefficient, assuming the considered kernel function types are predefined. To pursue the optimal parameter set of HKELM, the NGO-based optimization technique is integrated into the HKELM algorithm, and the resultant NGO-HKELM-based classification algorithm has its flowchart illustrated in Figure 12. Specifically, the basic parameters of the NGO algorithm are first defined, and the HKELM parameter sets are randomly initiated within their considered boundaries. Then, the generated HKELM parameter sets are fed to the HKELM-based classification algorithm, and their corresponding classification accuracy values (for both the training dataset and the test dataset) are taken into the objective function (12) to create the initial populations of the northern goshawk. Afterwards, the positions (corresponding to the parameter sets of HKELM) of northern goshawks are updated via the prey identification phase (using (1)–(3)) and the ensuing chase and escape phase (using (4)–(6)). During the two-stage optimization process, the parameter sets of HKELM are fed to the HKELM-based classification model, with the updated model tested on the multi-domain feature sets and the corresponding label sets. After the performance evaluation of the revised classification model, the retrieved accuracy values for the training and test datasets are again returned to the NGO algorithm for further optimization of the HKELM parameter set.

$$F_i = 1 - \frac{acc_{train,i} + acc_{test,i}}{2} \quad (12)$$

where: $acc_{train,i}$ and $acc_{test,i}$ refer to the classification accuracy values of the training dataset and the test dataset, respectively, while F_i is the objective function of i th northern goshawk.

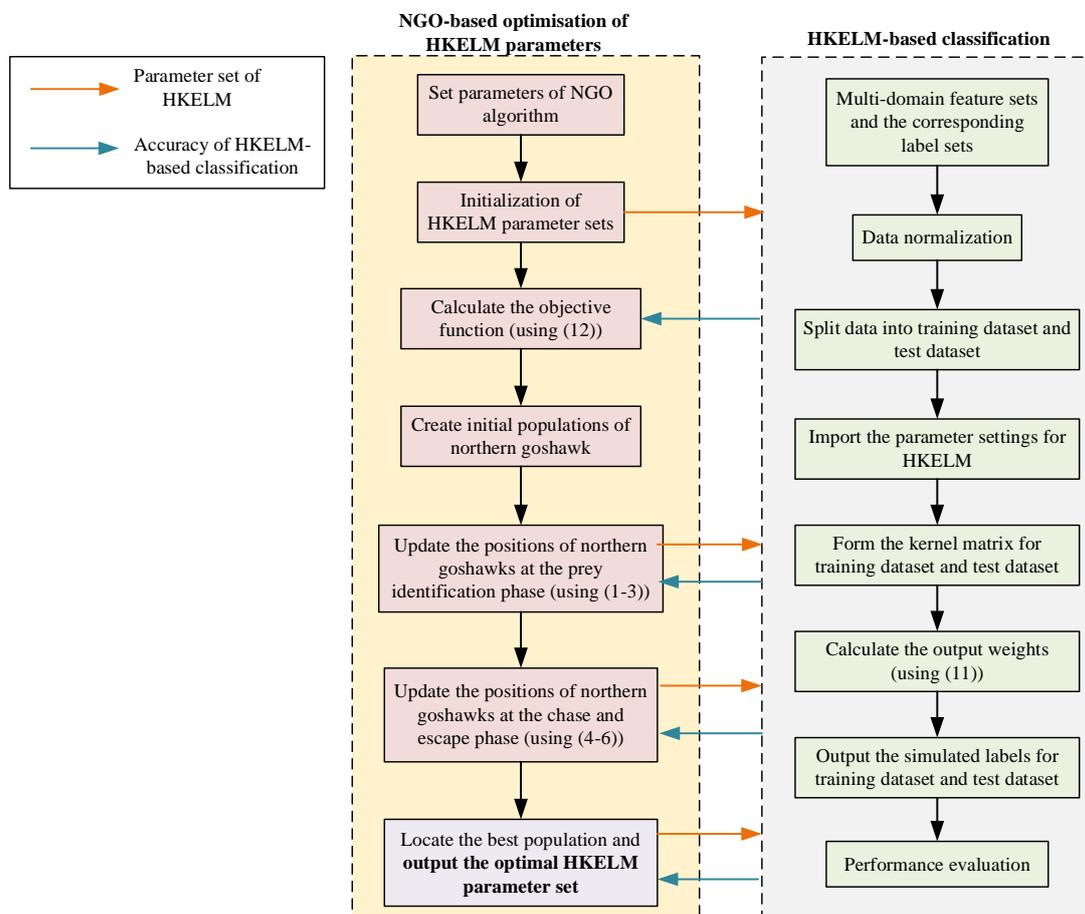


Figure 12. The flowchart of the NGO-HKELM-based classification algorithm.

Based on the selected multi-domain feature set, the ReliefF-mRMR-based feature selection technique and the proposed NGO-HKELM-based classification algorithm, a novel fault diagnosis technique for PV inverters is derived, with its implementation procedure shown in Figure 13. The main steps are summarized as follows:

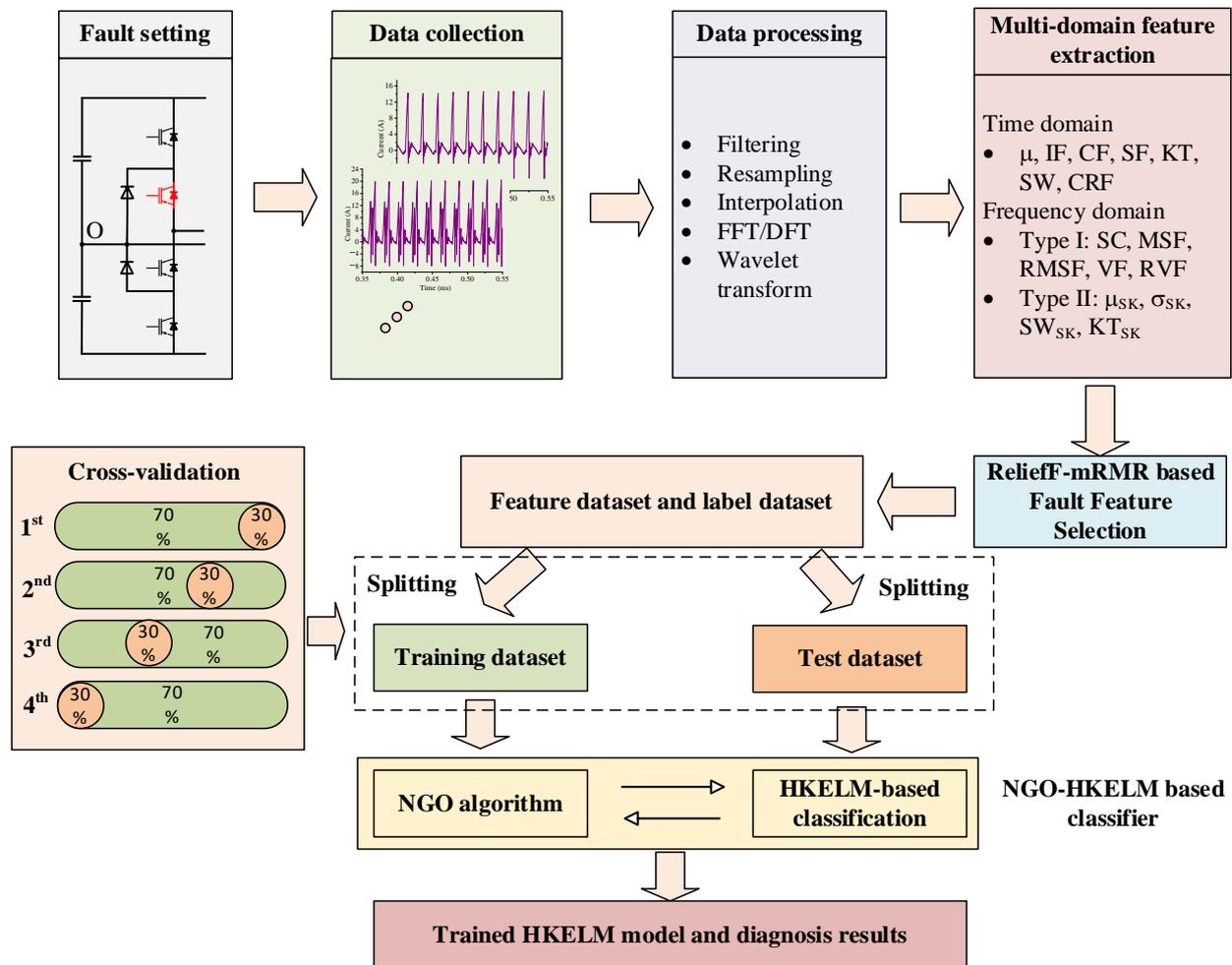


Figure 13. The general procedure of the proposed NGO-HKELM-based fault diagnosis technique for PV inverters.

Step 1: Analyze the fault operation modes of the three-phase PV inverter under test and define the switch fault types (as in Section 2.2). Conduct the fault tests either via simulations or laboratory experiments and record the current signals measured at the grid integration point of the PV inverter.

Step 2: Apply appropriate data processing techniques (e.g., filtering, resampling, interpolation, FFT/DFT and wavelet transform) to the collected signal set and extract the considered multi-domain features (i.e., the fault feature indices presented in Table 5). Apply the ReliefF-mRMR-based fault selection technique to define the optimal feature set and remove the redundant features. Create a dataset for the training of the fault classifier.

Step 3: Split the dataset into training and test datasets based on cross-validation. Train the NGO-HKELM-based classifier for which the kernel parameters, the kernel function weighting coefficients and the regularization coefficient are internally optimized by the NGO optimizer. The output weights of the NGO-HKELM classifier are then applied to locate the fault type. The accuracy of the final NGO-HKELM classifier is validated on the test dataset.

Step 4: Retrieve the finely trained NGO-HKELM classifier and apply it to the fault detection and diagnosis of real-time signals of PV inverters.

3. Performance Evaluation

To evaluate the performance of the proposed fault diagnosis technique for PV inverters, two commercially available grid-connected PV inverters are tested on a laboratory testbed, followed by developing their representative component-based model (CBM). The developed CBMs will be further used to perform comprehensive fault simulations and generate the required dataset to train the NGO-HKELM classifier. The performance of the proposed fault diagnosis technique will be further compared to conventional classifiers, including BP neural network (BPNN), decision tree (DT) and support vector machine (SVM).

3.1. Testbed Setup and Component-Based PV Inverter Model Build-Up

To train the proposed NGO-HKELM classifier, a collection of PV inverter fault signals is required. For a grid-connected PV inverter, obtaining fault signals of all types is challenging and will significantly affect the accuracy of the trained NGO-HKELM classifier. Although fault data collection can be conducted in a laboratory environment, irreversible damage to the PV inverters may occur during fault testing. Therefore, component-based models (CBMs) are built-up for two commercially available three-phase grid-connected PV inverters. Then, comprehensive fault simulations of semiconductor power switches are conducted to collect the fault signals that will be applied for the next-stage fault classifier training.

The testbed setup for three-phase grid-connected PV inverters is illustrated in Figure 14. It consists of a three-phase programmable power source, impedance section module, data acquisition system, breakers, three-phase PV inverter under test, the PV panel emulator (programmable DC source) and a load for dissipating the power generated from the PV system. The testbed is fully automated via a central control PC, and the SCPI commands are exchanged among several programmable devices via GPIB or RS-232 port. With the aid of the laboratory testbed, the two PV inverters are tested at different operating powers (in a range from 70% P_{rated} to 100% P_{rated} , where P_{rated} is the rated power of the PV inverter), with their input and output voltage and current waveforms recorded. The basic information on the two PV inverters is tabulated in Table 6. Specifically, PVI-A is a three-phase H-bridge inverter rated at 10 kW, while PVI-B is a three-phase NPC inverter rated at 15 kW.

Table 6. Basic information on the two grid-connected PV inverters under test.

	PVI-A	PVI-B
Topology	H-bridge	NPC
Rated power (kW)	10	15
Rated current (A)	14.5	21
Phase connection	Three-phase	Three-phase

Based on the pre-knowledge of the circuit topologies of the two inverters under test, their generalized CBMs are first developed, followed by tuning the circuit and control parameters to match the measurements and simulation results. The circuit schematics of developed CBMs for the three-phase H-bridge inverter and the three-phase NPC inverter are shown in Figure 15, with their finely tuned model parameters in Table 7. To consider the possible parameter deviations between the developed CBM model and the actual PV inverter device, a standard deviation of 20% is applied to the parameter value of the electrical components (i.e., parameter values of filter and DC-link capacitor). Moreover, to consider the impact of the fluctuating power output from the PV system on the fault diagnosis, different operating power points of PV inverters are considered (ranging from 70% P_{rated} to 100% P_{rated} , with a step of 10% P_{rated}). For each combination of the above impact factors, fault tests are performed by using developed CBMs for PVI-A and PVI-B.

The numbers of simulated fault tests for PVI-A and PVI-B are listed in Table 8. The simulated grid-side voltage and current waveforms are recorded for each test point to form the fault database.

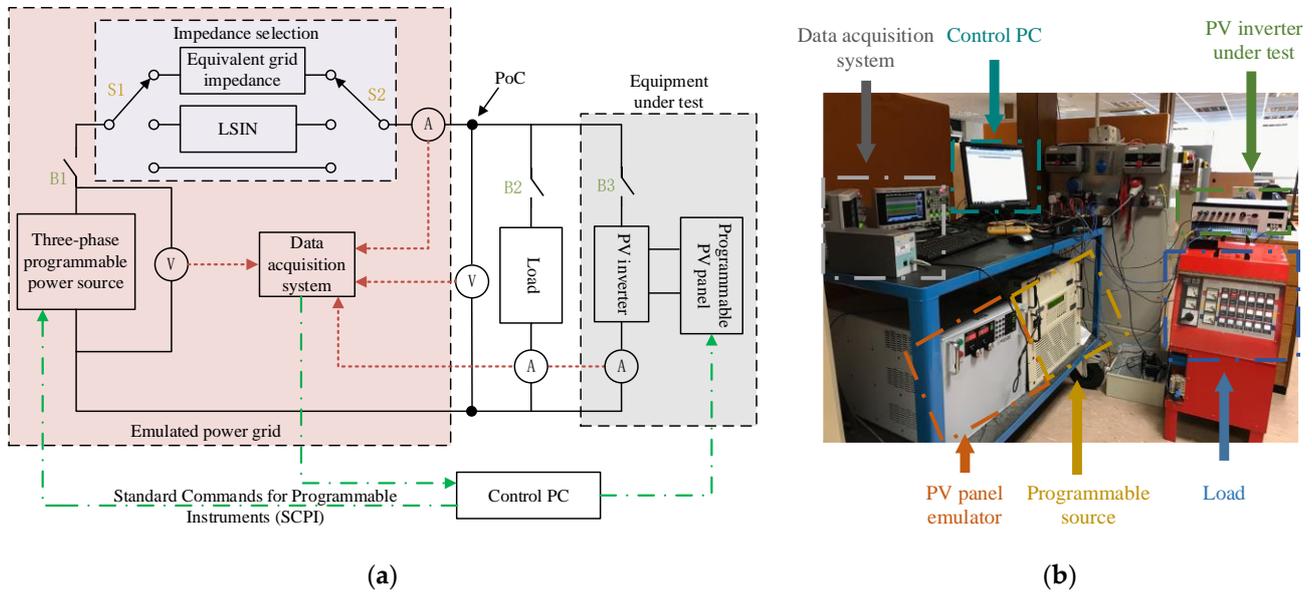


Figure 14. The testbed setup for three-phase PV inverters: (a) block diagram of the testbed setup; (b) actual laboratory testbed.

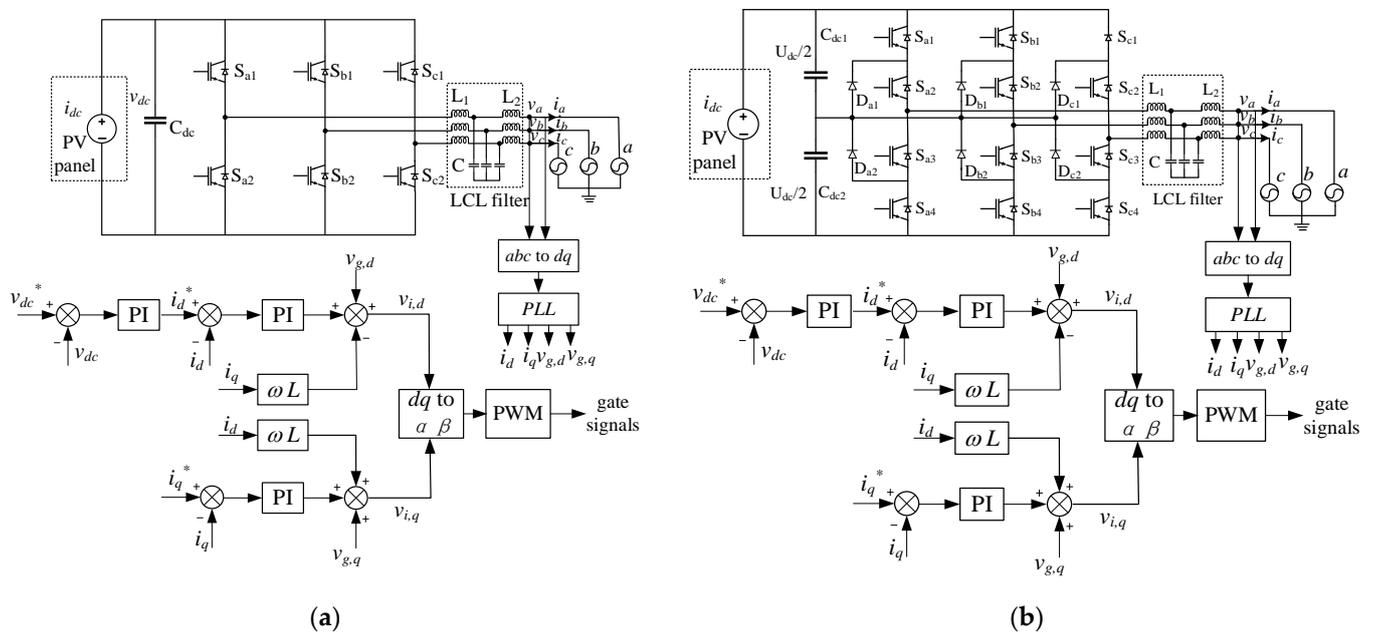


Figure 15. Developed CBMs for PVI-A and PVI-B: (a) CBM for PVI-A; (b) CBM for PVI-B.

Table 7. The model parameters for PVI-A and PVI-B.

	CBM Parameters for PVI-A	CBM Parameters for PVI-B
Input filter	$L_1 = 0.85 \text{ mH}, L_2 = 0.55 \text{ mH}, C = 12 \text{ }\mu\text{F}$	$L_1 = 0.5 \text{ mH}, L_2 = 0.5 \text{ mH}, C = 100 \text{ }\mu\text{F}$
DC-link	$v_{dc}^* = 690 \text{ V}, C_{dc} = 600 \text{ }\mu\text{F}$	$v_{dc}^* = 800 \text{ V}, C_{dc1} = C_{dc2} = 1100 \text{ }\mu\text{F}$
DC-link PI controller	$K_P = 1.1, K_I = 1$	$K_P = 1.1, K_I = 1$
Current PI controller	$K_P = 4, K_I = 90$	$K_P = 5, K_I = 5000$

Table 8. The number of simulated fault tests for PVI-A and PVI-B.

	Number of Simulated Fault Tests for PVI-A	Number of Simulated Fault Tests for PVI-B
Different operating powers	4 (70%, 80%, 90% and 100% P_{rated})	4 (70%, 80%, 90% and 100% P_{rated})
Parameter sets	11	11
Fault types	22	73
Total	968 ($4 \times 11 \times 22$)	3212 ($4 \times 11 \times 73$)

3.2. Comparative Performance Evaluation Results

The multi-domain fault features are extracted and applied with the collected fault signals to train the proposed NGO-HKELM classifier by following the procedure presented in Figure 14. The acquired optimal parameter sets of the proposed NGO-HKELM are tabulated in Table 9. Here, the RBF and polynomial functions are selected to form the hybrid kernel applied in the NGO-HKELM classifier. In this way, the advantages of both the local kernel (e.g., the RBF kernel) and global kernel (e.g., the polynomial kernel) can be complementary to achieve a good balance between the learning capability and the generalization capability of the classifier. The performance of the proposed NGO-HKELM classifier is compared to the other three well-known fault classification techniques, including the BPNN classifier, the SVM classifier and the DT classifier. For the processed fault datasets for PVI-A and PVI-B, 70% of the data sets are used for classifier training, while the other 30% are applied for model validation. The classification accuracy will be evaluated using (13).

$$Accuracy = \frac{N_{TP,T} + N_{TP,V}}{N_{TP,T} + N_{FP,T} + N_{TP,V} + N_{FP,V}} \times 100\% \quad (13)$$

where: $N_{TP,T}$ and $N_{TP,V}$ refer to the total number of correct predictions in the training dataset and test dataset, respectively; $N_{FP,T}$ and $N_{FP,V}$ refer to the total number of wrong predictions in the training dataset and test dataset, respectively.

The acquired actual class and predicted class for the fault diagnosis of PVI-A and PVI-B are illustrated in Figures 16 and 17, respectively, with the cross-validation accuracy tabulated in Table 10. For the fault diagnosis of PVI-A, it is observed from Figure 16 that the proposed NGO-HKELM classifier has the best accuracy, as opposed to the other three classifiers, and can correctly identify all 22 OC fault modes (those listed in Table 2) of PVI-A. The BPNN classifier can accurately identify the 1st~16th and 18th~22nd OC fault modes but often mistakes the 17th fault mode type for the 22nd fault mode. It indicates that the BPNN classifier cannot precisely capture the slight difference between the 17th and the 22nd fault modes via the considered feature indices. It is mainly due to the occurrence of overfitting problems in the process of the BPNN classifier training. The SVM and the DT classifiers have similar accuracy, and the misprediction only occurs for a few cases (less than 1% of the total 968 datasets). The prediction error results of the SVM model are owing to the fact that the default linear kernel cannot perfectly separate the large-scale data that are highly linearly inseparable. Compared to the other three classifiers, the DT classifier has the poorest performance in discriminating the different fault modes of the Type IV OC fault of PVI-A, where the misprediction occurs for the 2nd, 12th~14th and 17th~20th fault modes. For PVI-A operating under Type IV fault modes, the grid-side AC current

becomes less distinguishable in both time and frequency domains, and the DT classifier is incapable of capturing the slight difference between different fault modes due to the occurrence of overfitting. Similarly, for the fault diagnosis of PVI-B, the NGO-HKELM classifier can achieve an accuracy above 99%, which is better than that of the other three classifiers. The conspicuous performance of the NGO-HKELM classifier is attributed to the superior generalization and exploration capability arising from the utilization of hybrid kernels and the NGO-based model parameter optimization. Regarding the distribution patterns of the prediction error results, it turns out that the NGO-HKELM classifier has its prediction errors roughly evenly distributed across the four types of faults, while the other three classifiers have their prediction errors occur mostly in the Type IV fault. It indicates that the NGO-HKELM classifier has more prominent performance in capturing the almost indistinguishable differences between various fault patterns. In contrast, the BPNN and SVM classifiers have competitive performance in detecting the Types I~III OC faults due to their inherent nonlinear mapping capability between fault features and fault modes. The DT classifier has the worst accuracy compared to the other three, and the misprediction mostly occurs for the Type IV OC fault of PVI-B, as illustrated in Figure 17d. This is due to the fact that the DT classifier can easily overfit the training samples when the tree structure is over-complex.

Table 9. Optimal parameter sets for the proposed NGO-HKELM classifier applied to PVI-A and PVI-B.

	Parameter Set for the NGO-HKELM Classifier of PVI-A	Parameter Set for the NGO-HKELM Classifier of PVI-B
Types of kernel functions	RBF kernel and polynomial kernel	RBF kernel and polynomial kernel
Kernel parameters in Equations (8) and (9)	$\sigma = 14.54, c_3 = 0.001, d = 1$	$\sigma = 12.30, c_3 = 243.791, d = 2.0038$
Weighting coefficients of kernel functions, c_1 and c_2 , in Equations (7) and (11)	$c_1 = 0.0112, c_2 = 0.9888$	$c_1 = 0.2728, c_2 = 0.7272$
Regularization coefficient, C in Equation (11)	19.3252	13.3914
Parameters for NGO optimizer (population number N_p , number of dimensions N_{dim} , and maximum iteration number $iter_{max}$)	$N_p = 50, N_{dim} = 5, iter_{max} = 20$	$N_p = 50, N_{dim} = 5, iter_{max} = 20$

Table 10. Comparison of the accuracy of the proposed NGO-HKELM classifier, BPNN classifier, the SVM classifier and DT classifier when applied to the fault diagnosis of PVI-A and PVI-B.

Type of Classifier	Fault Diagnosis of PVI-A	Fault Diagnosis of PVI-B
NGO-HKELM	100.00%	99.46%
BPNN	98.87%	97.64%
SVM	99.48%	94.98%
DT	99.50%	81.52%

Finally, a comparison of the accuracy of the proposed NGO-HKELM-based fault diagnosis method and the relevant ML-based fault diagnosis methods in the literature is conducted, with the outcomes given in Table 11. The considered ML-based classifiers include BPNN, BN, SVM, DT and random forest (RF). It turns out that the proposed NGO-

HKELM-based fault diagnosis method achieves the highest accuracy in detecting the switch OC faults of a three-phase H-bridge inverter and NPC inverter.

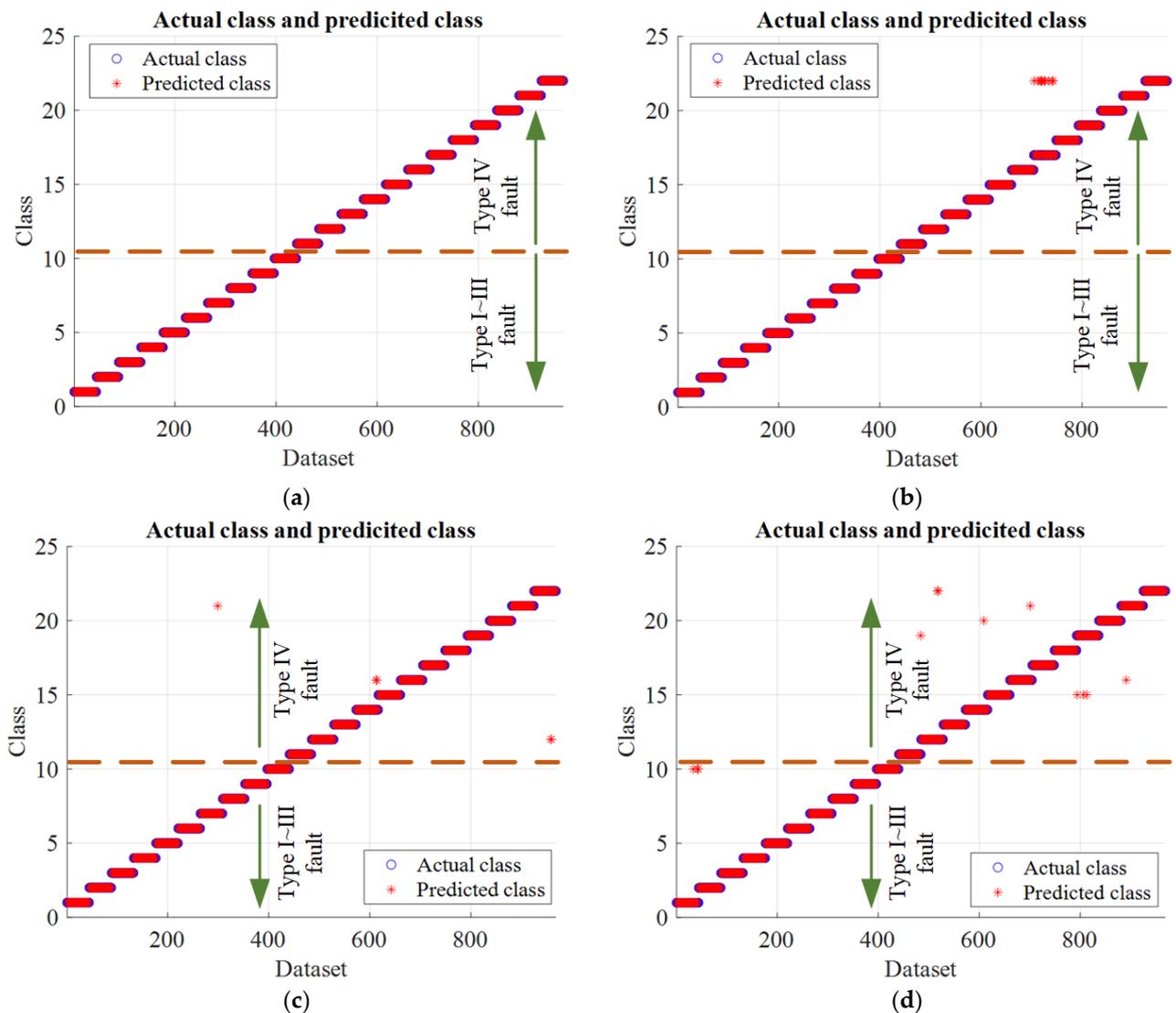


Figure 16. The acquired actual class and predicted class for the fault diagnosis of PVI-A: (a) the proposed NGO-HKELM classifier; (b) the BPNN classifier; (c) the SVM classifier; (d) the DT classifier.

Table 11. Comparison of the accuracy of the proposed NGO-HKELM-based fault diagnosis method and the relevant ML-based fault diagnosis methods in the literature.

Type of Classifier	Inverter Topology	Measured Signals	Accuracy
NGO-HKELM	Three-phase H-bridge inverter	Inverter output current	100%
	Three-phase NPC inverter	Inverter output current	99.46%
BPNN [19]	Three-phase H-bridge inverter	Inverter output current	/
BN [20]	Three-phase H-bridge inverter	Inverter output voltage	98.99%
SVM [24]	Three-phase H-bridge inverter	Inverter output current	95.6%
DT [25]	Three-phase H-bridge inverter	Inverter output current	96.4%
RF [48]	Three-phase NPC inverter	Inverter output current	97.27%

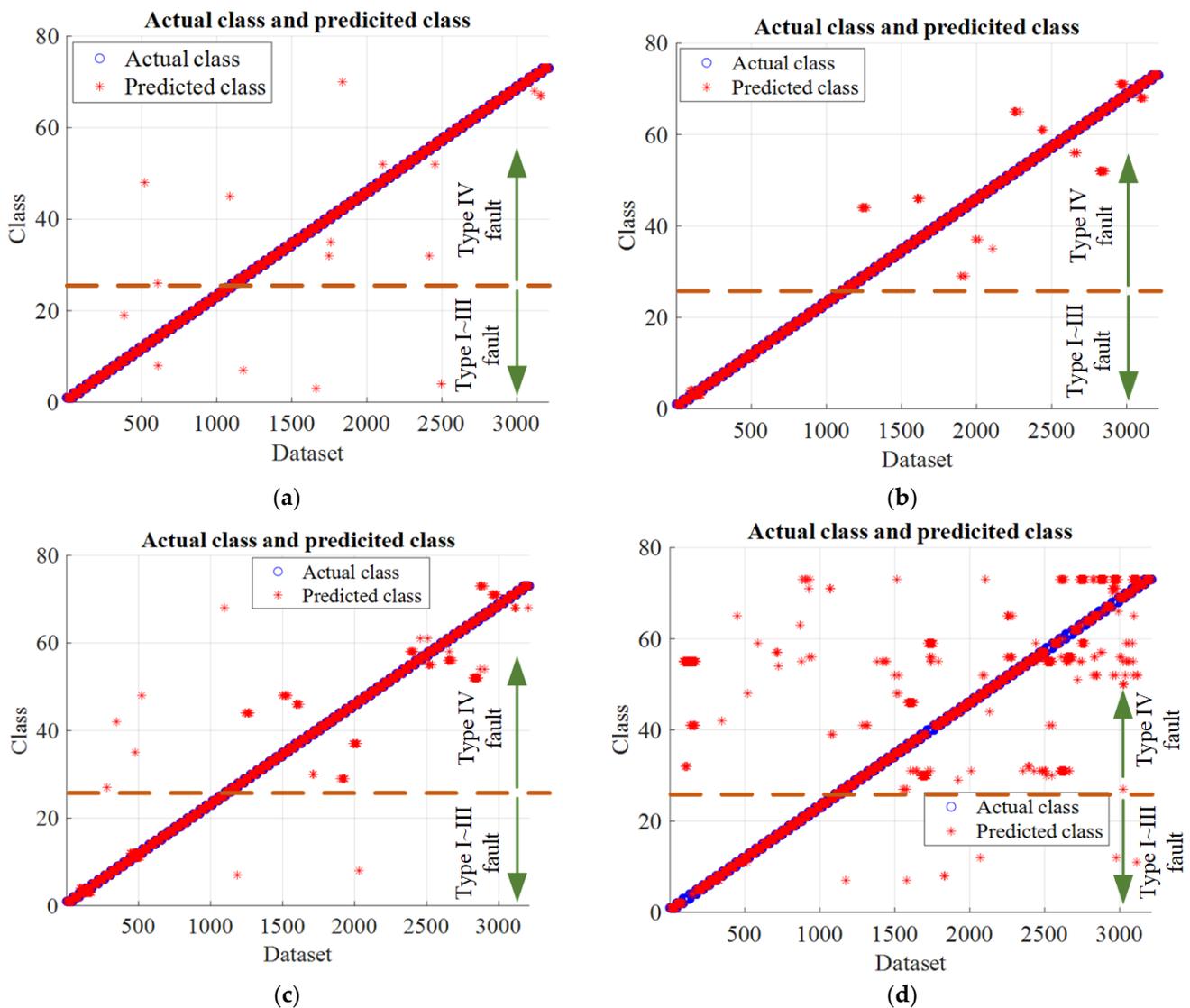


Figure 17. The acquired actual class and predicted class for the fault diagnosis of PVI-B: (a) the proposed NGO-HKELM classifier; (b) the BPNN classifier; (c) the SVM classifier; (d) the DT classifier.

4. Conclusions

In this paper, we proposed an NGO-HKELM-based fault diagnosis method to precisely and reliably identify three-phase PV inverter faults caused by the OC failure of semiconductor switches. We investigated the fault operation modes and features of the two most widely used three-phase PV inverter topologies—the H-bridge inverter topology and NPC inverter topology. Then, diversified multi-domain fault feature indices were reviewed, and a ReliefF-mRMR-based fault feature selection technique was introduced, which can select the optimal feature set for the following classifier training. As the HKELM classifier has a mixture of different kernel functions, it can achieve a balance between the learning capability and the generalization capability if only the model parameters are finely tuned. Therefore, we proposed an NGO-HKELM classifier where the NGO optimizer is applied to retrieve the optimal parameter set from the training dataset automatically. As the fault classifier training relies on extensive historical fault datasets that are hard to obtain practically, we used simulation-generated training datasets instead. Specifically, we developed two CBMs for a three-phase H-bridge inverter and a three-phase NPC inverter, where the model parameters were finely tuned via laboratory measurements. Afterwards, the two CBMs emulated diversified OC faults of semiconductor switches, with the generated

fault database applied for the fault classifier training. To evaluate the performance of the proposed NGO-HKELM-based fault diagnosis technique, we compared the fault diagnosis accuracy of the NGO-HKELM classifier with the other three widely used classifiers, namely the BPNN classifier, the SVM classifier and the DT classifier. It turned out that the proposed NGO-HKELM classifier achieved the best accuracy for all of the considered fault cases. It could accurately predict all of the considered faults of a three-phase H-bridge inverter. When it was applied to the fault diagnosis of a three-phase NPC inverter, its prediction accuracy was maintained above 99%. As opposed to the mainstream ML-based fault diagnosis methods in PV inverters, the proposed NGO-HKELM-based fault diagnosis method featured the following superiorities:

(i) The proposed ReliefF-mRMR-based multi-domain fault feature selection method can capture the most appropriate feature set that ensures the completeness of fault information.

(ii) The proposed NGO-HKELM classifier is proven to have remarkable classification capability in the fault diagnosis stage. This is attributed to applying hybrid kernels and the model parameter optimization via the novel NGO algorithm.

(iii) The proposed fault diagnosis method only needs the inverter output current signal that is captured by the current sensor already installed within the PV inverter (for control and protection purposes). Therefore, no additional sensors are required for measurements.

(iv) The proposed fault diagnosis method can be implemented in a PV inverter by modifying the microcontroller embedded within the inverter control circuit.

From the viewpoint of practical application, there are still many aspects that are not covered in this paper, including the measurement noise and outliers, the sampling rate of the inverter output current, the applied time-frequency signal processing technique, etc. In our future work, we are further motivated to include those aspects and promote the proposed fault diagnosis method toward practical implementation.

Author Contributions: Conceptualization, X.X. and H.G.; methodology, X.X.; software, T.Z.; validation, X.X. and T.Z.; formal analysis, T.Z.; investigation, X.X.; resources, H.G.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X. and S.Z.; visualization, Z.Q.; supervision, H.G.; project administration, X.X.; funding acquisition, X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Innovation and Entrepreneurship Program of Jiangsu Province, Grant No. JSSCBS20210538, and sponsored by NUPTSE, Grant No. NY220082.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

- Dhar, A.; Naeth, M.A.; Jennings, P.D.; El-Din, M.G. Perspectives on Environmental Impacts and a Land Reclamation Strategy for Solar and Wind Energy Systems. *Sci. Total Environ.* **2020**, *718*, 134602. [[CrossRef](#)]
- Mahalik, A.K.; Mallick, H.; Padhan, H. Do Educational Levels Influence the Environmental Quality? The Role of Renewable and Non-Renewable Energy Demand in Selected BRICS Countries with a New Policy Perspective. *Renew. Energy* **2021**, *164*, 419–432. [[CrossRef](#)]
- Dahl, M.; McMahon, K.; Lavery, P.S.; Hamilton, S.H.; Lovelock, C.E.; Serrano, O. Ranking the Risk of CO₂ Emissions from Seagrass Soil Carbon Stocks under Global Change Threats. *Glob. Environ. Chang.* **2023**, *78*, 102632. [[CrossRef](#)]
- Kruitwagen, L.; Story, K.T.; Friedrich, J.; Fthenakis, V. A Global Inventory of Photovoltaic Solar Energy Generating Units. *Nature* **2021**, *598*, 604–610. [[CrossRef](#)] [[PubMed](#)]
- Hossain, M.S.; Abboodi Madlool, N.; Al-Fatlwi, A.W.; El Haj Assad, M. High Penetration of Solar Photovoltaic Structure on the Grid System Disruption: An Overview of Technology Advancement. *Sustainability* **2023**, *15*, 1174. [[CrossRef](#)]
- Chen, H.; Chen, W. Status, trend, economic and environmental impacts of household solar photovoltaic development in China: Modelling from subnational perspective. *Appl. Energy* **2021**, *303*, 117616. [[CrossRef](#)]
- Panagiotidou, M.; Brito, M.C.; Hamza, K.; Jasieniak, J.J.; Zhou, J. Prospects of photovoltaic rooftops, walls and windows at a city to building scale. *Sol. Energy* **2021**, *230*, 675–687. [[CrossRef](#)]

8. Lee, C.G.; Shin, W.G.; Lim, J.R.; Kang, G.H.; Ju, Y.C.; Hwang, H.M.; Chang, H.S.; Ko, S.W. Analysis of Electrical and Thermal Characteristics of PV Array under Mismatching Conditions Caused by Partial Shading and Short Circuit Failure of Bypass Diodes. *Energy* **2021**, *218*, 119480. [[CrossRef](#)]
9. Malik, A.; Haque, A.; Kurukuru, V.B.; Khan, M.A.; Blaabjerg, F. Overview of Fault Detection Approaches for Grid Connected Photovoltaic Inverters. *e-Prime-Adv. Electr. Eng. Electron. Energy* **2022**, *2*, 100035.
10. Rahimpour, S.; Tarzamni, H.; Kurdkandi, N.V.; Husev, O.; Vinnikov, D.; Tahami, F. An Overview of Lifetime Management of Power Electronic Converters. *IEEE Access* **2022**, *10*, 109688–109711. [[CrossRef](#)]
11. Yang, S.; Bryant, A.; Mawby, P.; Xiang, D.; Ran, L.; Tavner, P. An Industry-Based Survey of Reliability in Power Electronic Converters. *IEEE Trans. Ind. Appl.* **2011**, *47*, 1441–1451. [[CrossRef](#)]
12. Stonier, A.A.; Lehman, B. An Intelligent-Based Fault-Tolerant System for Solar-Fed Cascaded Multilevel Inverters. *IEEE Trans. Energy Convers.* **2018**, *33*, 1047–1057. [[CrossRef](#)]
13. Choi, U.-M.; Lee, J.-S.; Blaabjerg, F.; Lee, K.-B. Open-Circuit Fault Diagnosis and Fault-Tolerant Control for a Grid-Connected NPC Inverter. *IEEE Trans. Power Electron.* **2016**, *31*, 7234–7247. [[CrossRef](#)]
14. Xu, X.; Yu, F. A switch open-circuit fault diagnosis method for three-phase inverters based on voltage patterns. In Proceedings of the 2022 2nd International Conference on Electrical Engineering and Control Science (IC2ECS), Nanjing, China, 16–18 December 2022; pp. 901–908.
15. Sun, J.; Zhu, Q.; Zhou, Y.; Zhao, J. Open-Circuit Fault Diagnosis of Voltage Source Inverters for PMSM Drive System using Sine-Wave Injection Method. In Proceedings of the 2021 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 22–24 May 2021; pp. 4598–4603.
16. Gong, X.; Wang, N.; Zhang, Y.; Yin, S.; Wang, M.; Wu, G. Fault Diagnosis of Micro Grid Inverter Based on Wavelet Transform and Probabilistic Neural Network. In Proceedings of the 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020; pp. 4078–4082.
17. Abari, I.; Hamouda, M.; Ben Hadj Slama, J. Three-Level NPC Inverter Fault Diagnosis using Wavelet Packet Transform and Statistical Analysis of the Emitted Near-Field. In Proceedings of the 2021 12th International Renewable Energy Congress (IREC), Hammamet, Tunisia, 26–28 October 2021; pp. 1–5.
18. Long, H.; Ma, M.; Guo, W.; Li, F.; Zhang, X. Fault Diagnosis for IGBTs Open-Circuit Faults in Photovoltaic Grid-Connected Inverters Based on Statistical Analysis and Machine Learning. In Proceedings of the 2020 IEEE 1st China International Youth Conference on Electrical Engineering (CIYCEE), Wuhan, China, 1–4 November 2020; pp. 1–6.
19. Ni, X.; Yang, Z. Inverter fault diagnosis based on BP neural network. In Proceedings of the 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), Dalian, China, 15–17 July 2021; pp. 185–190.
20. Cai, B.; Zhao, Y.; Liu, H.; Xie, M. A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems. *IEEE Trans. Power Electron.* **2017**, *32*, 5590–5600. [[CrossRef](#)]
21. Yuan, W.; Wang, T.; Diallo, D. A Secondary Classification Fault Diagnosis Strategy Based on PCA-SVM for Cascaded Photovoltaic Grid-connected Inverter. In Proceedings of the IECON 2019—45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 14–17 October 2019; pp. 5986–5991.
22. Chen, Z.; Jiang, C.; Xie, L. A novel ensemble ELM for human activity recognition using smartphone sensors. *IEEE Trans. Ind. Inf.* **2019**, *15*, 2691–2699. [[CrossRef](#)]
23. Yang, H.; Peng, Z.; Xu, Q.; Huang, T.; Zhu, X. Inverter fault diagnosis based on Fourier transform and evolutionary neural network. *Front. Energy Res.* **2023**, *10*, 1090209. [[CrossRef](#)]
24. Bowen, C.; Wei, T. Switch open-circuit faults diagnosis of inverter based on wavelet and support vector machine. In Proceedings of the 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Changsha, China, 1–3 November 2019; pp. 1178–1184.
25. Nguyen, N.T.; Nguyen, H.P. Fault diagnosis of voltage source inverter for induction motor drives using decision tree. *Lect. Notes Electr. Eng.* **2017**, *398*, 819–826.
26. Katir, H.; Abouloifa, A.; Lachkar, I.; Noussi, K.; Giri, F.; Guerrero, J.M. Advanced Nonlinear Control of A Grid-Connected Photovoltaic Power System Using N-Cascaded H-Bridge Multilevel Inverters. *IFAC-PapersOnLine* **2020**, *53*, 12859–12864. [[CrossRef](#)]
27. Jayaprakash, S.; Balamurugan, R.; Gopinath, S.; Kokilavani, T.; Maheswaran, S. 3-Phase multi-inverter with cascaded H-bridge inverter designing and implementation for renewable system. *Sustain. Energy Technol. Assess.* **2022**, *52 Pt B*, 102088. [[CrossRef](#)]
28. Ye, Y.; Hua, T.; Chen, S.; Wang, X. Neutral-point-clamped five-level inverter with self-balanced switched-capacitor. *IEEE Trans. Ind. Electron.* **2022**, *69*, 2202–2215. [[CrossRef](#)]
29. Ye, Y.; Zhang, G.; Huang, J.; Chen, S.; Wang, X. Comparative analysis of hybrid NPP and NPC seven-level inverter with switched-capacitor. *IEEE Access* **2021**, *9*, 85852–85863. [[CrossRef](#)]
30. Siwakoti, Y.P.; Liese, S.; Mahajan, A.; Palanisamy, A.; Rogers, D.; Blaabjerg, F. A New Seven-Level Active Boost Neutral Point Clamped (7L-ABNPC). In Proceedings of the IEEE Energy Conversion Congress and Exposition (ECCE) 2018, Portland, OR, USA, 23–27 September 2018; pp. 5636–5642.
31. Teodorescu, R.; Liserre, M.; Rodriguez, P. Photovoltaic Inverter Structures. In *Grid Converters for Photovoltaic and Wind Power Systems*; IEEE: Piscataway, NJ, USA, 2007; pp. 5–29.
32. Zhang, Z.; Thomsen, O.C.; Andersen, M.A.E. Discontinuous PWM Modulation Strategy With Circuit-Level Decoupling Concept of Three-Level Neutral-Point-Clamped (NPC) Inverter. *IEEE Trans. Ind. Electron.* **2013**, *60*, 1897–1906. [[CrossRef](#)]

33. Stala, R. Application of balancing circuit for DC-link voltages balance in a single-phase diode-clamped inverter with two three-level legs. *IEEE Trans. Ind. Electron.* **2011**, *58*, 4185–4195. [[CrossRef](#)]
34. Sinha, A.; Jana, K.C.; Das, M.K. An inclusive review on different multi-level inverter topologies, their modulation and control strategies for a grid connected photovoltaic system. *Sol. Energy* **2018**, *170*, 633–657. [[CrossRef](#)]
35. Yu, X.; Chum, P.; Sim, K.-B. Analysis the Effect of PCA for Feature Reduction in Nonstationary EEG Based Motor Imagery of BCI System. *Opt.—Int. J. Light Electron Opt.* **2014**, *125*, 1498–1502. [[CrossRef](#)]
36. Kononenko, I. Estimating Attributes: Analysis and Extensions of RELIEF. In Proceedings of the European Conference on Machine Learning, Catania, Italy, 6–8 April 1994; pp. 171–182.
37. Jeyachidra, J.; Punithavalli, M. Distinguish Ability Based Weighted Feature Selection Algorithms for Classification of Gene Microarray Data Set. *Int. J. Eng. Sci. Innov. Technol.* **2013**, *2*, 6.
38. Singh, V.P.; Arvind, S.G.; Mahapatra, A.G. Hybrid Correlation Based Gene Selection for Accurate Cancer Classification of Gene Expression Data. *Int. J. Comput. Appl.* **2012**, *43*, 13–18.
39. Robnik-Šikonja, M.; Kononenko, I. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Mach. Learn.* **2003**, *53*, 23–69. [[CrossRef](#)]
40. Ebrahimzadeh, A.; Ghafari, M.; Moshkbar-Bakhshayesh, K. Detection and Estimation of Faulty Sensors in NPPs Based on Thermal-Hydraulic Simulation and Feed-Forward Neural Network. *Ann. Nucl. Energy* **2022**, *166*, 108726. [[CrossRef](#)]
41. Dehghani, M.; Hubálovský, H.; Trojovský, A. Northern Goshawk Optimization: A New Swarm-Based Algorithm for Solving Optimization Problems. *IEEE Access* **2021**, *9*, 162059–162080. [[CrossRef](#)]
42. Liang, Y.; Hu, X.; Hu, G.; Dou, W. An Enhanced Northern Goshawk Optimization Algorithm and Its Application in Practical Optimization Problems. *Mathematics* **2022**, *10*, 4383. [[CrossRef](#)]
43. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme Learning Machine: Theory and Applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
44. Huang, G.B.; Siew, C.K. Extreme Learning Machine with Randomly Assigned RBF Kernels. *Int. J. Inf. Technol.* **2005**, *11*, 16–24.
45. Wang, S.; Liu, Z.; Jia, Z.; Li, Z. Incipient Fault Diagnosis of Analog Circuit with Ensemble HKELM Based on Fused Multi-Channel and Multi-Scale Features. *Eng. Appl. Artif. Intell.* **2023**, *117 Pt B*, 105633. [[CrossRef](#)]
46. Lv, L.; Wang, W.; Zhang, Z.; Liu, X. A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. *Knowl.-Based Syst.* **2020**, *195*, 105648. [[CrossRef](#)]
47. Gaspar, A.; Oliva, D.; Hinojosa, S.; Aranguren, I.; Zaldivar, D. An optimized Kernel Extreme Learning Machine for the classification of the autism spectrum disorder by using gaze tracking images. *Appl. Soft Comput.* **2022**, *120*, 108654. [[CrossRef](#)]
48. Kou, L.; Liu, C.; Cai, G.W.; Zhou, J.N.; Yuan, Q.D.; Pang, S.M. Fault diagnosis for open-circuit faults in NPC inverter based on knowledge-driven and data driven approaches. *IET Power Electron.* **2020**, *13*, 1236–1245. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.