

## Article

# Arc Fault Detection Algorithm Based on Variational Mode Decomposition and Improved Multi-Scale Fuzzy Entropy

Lina Wang <sup>1</sup>, Hongcheng Qiu <sup>1</sup>, Pu Yang <sup>1,\*</sup> and Longhua Mu <sup>2</sup>

<sup>1</sup> School of Automation Science and Electrical Engineering, Beihang University, 37 Xueyuan Road, Haidian District, Beijing 100191, China; wangln@buaa.edu.cn (L.W.); MICATITAN@BUAA.EDU.CN (H.Q.)

<sup>2</sup> Department of Electrical Engineering, Tongji University, 4800 Cao'an Highway, Shanghai 201804, China; lhm@tongji.edu.cn

\* Correspondence: yangpu1996@buaa.edu.cn

**Abstract:** Arc fault diagnosis is necessary for the safety and efficiency of PV stations. This study proposed an arc fault diagnosis algorithm formed by combining variational mode decomposition (VMD), improved multi-scale fuzzy entropy (IMFE), and support vector machine (SVM). This method first uses VMD to decompose the current into intrinsic mode functions (IMFs) in the time-frequency domain, then calculates the IMFE according to the IMFs associated with the arc fault. Finally, it uses SVM to detect arc faults according to IMFEs. Arc fault data gathered from a PV arc generation experiment platform are used to validate the proposed method. The results indicated the proposed method can classify arc fault data and normal data effectively.

**Keywords:** arc fault diagnosis; variational mode decomposition; multi-scale fuzzy entropy; intrinsic mode functions; support vector machine



**Citation:** Wang, L.; Qiu, H.; Yang, P.; Mu, L. Arc Fault Detection Algorithm Based on Variational Mode Decomposition and Improved Multi-Scale Fuzzy Entropy. *Energies* **2021**, *14*, 4137. <https://doi.org/10.3390/en14144137>

Academic Editor: Teuvo Suntio

Received: 26 May 2021  
Accepted: 5 July 2021  
Published: 8 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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 shortage of fossil fuels and increasing concern for the environment, demand for clean energies is increasing. Among clean energies, solar energy has been found to have the advantages of sustainability and cleanliness, attracting more and more attention. According to the reports of the International Energy Agency (IEA) and the International Renewable Energy Agency (IRENA) [1,2], the total installed capacity of the global photovoltaic industry has maintained a stable upward trend from 2011 to 2019.

At present, photovoltaic systems usually operate at a high voltage to save on the investment in cables [3–8]. For example, the DC voltage of a photovoltaic system in some places has reached 1000 V [9]. High voltage makes air more easily ionized and greatly increases the possibility of a photovoltaic DC arc [10]. In addition, a PV system is often placed in harsh conditions. Small creatures, high air humidity, and so on can also trigger photovoltaic DC arcs. If an arc is not detected in time, a continuous arc can generate high temperature plasma, which can severely damage the components of the photovoltaic system or even cause a fire. This kind of damage can severely affect the efficiency and profitability of a PV system. Therefore, arc fault diagnosis plays an important role in keeping photovoltaic systems operating with reasonable efficiency, profitability, and reliability.

DC arcs can be divided into parallel arcs and series arcs. Parallel arcs will cause over-current and can be easily detected and prevented in time by the over-current protection device [11]. However, series arcs will not cause over-current. It is necessary to conduct deeper research on the diagnosis of DC-side series arcs [12–15].

At present, series arcs can be detected by using sensors to collect several types of signals or by analyzing electrical data in the time domain or the frequency domain [16–22]. Sensors can effectively detect arcs and complete the task of circuit protection. However, sensors are expensive and hard to use widely.

Analyzing data in the time domain has critical requirements for the input data. However, the output of the DC side of a photovoltaic power generation system will be affected by the external environment, which results in nonlinearity and instability in the data. It is usually difficult to make arc detections by using this method alone.

Analyzing electrical data in the frequency domain avoids these disadvantages. Frequency domain signal analysis methods such as the fast Fourier transformation (FFT) [23,24], the short-time Fourier transformation (STFT) [25,26], the wavelet transformation (WT) [27,28], and the empirical mode decomposition (EMD) [29,30] have been applied to the detection of DC series arcs with relatively good detection effects. However, the FFT is not suitable for nonstationary signals and is not capable of detecting when the arc occurs. The STFT relies heavily on the choice of window function. Narrow windows lead to low frequency resolution. Wide windows lead to low temporal resolution. The WT depends heavily on the choice of wavelet basis. An improper wavelet basis affects the detection. The EMD suffers from the issue of mode mixing of the IMFs. When mode mixing occurs, the decomposed signals have overlapping parts in the frequency domain. This phenomenon makes decomposition meaningless. However, the variational mode decomposition (VMD) [31–36] avoids almost all of the disadvantages mentioned above. It is a completely non-recursive signal decomposition method based on the frequency domain, which has been widely used in bearing fault detection, EEG signal detection, and so on. Due to its advantages, this method is used to decompose the current data for further analysis in this paper.

After decomposition is completed, some nonlinear parameter estimation methods should be used to furtherly analyze the results of decomposition. Approximate entropy [37], sample entropy [38], and fuzzy entropy [39,40] are used to extract the characteristic parameters of the signals obtained by analysis in the frequency domain. However, approximate entropy often produces illegal values. Sample entropy lacks stability. Fuzzy entropy only describes the characteristics of the signal in a single scale. However, the multi-scale entropy proposed by Costa makes fuzzy entropy capable of application in multi-scale situations. An improved version of multi-scale entropy is used to further analyze the decomposed data.

In this paper, an improved diagnosis algorithm consisting of VMD, IMFE, and the support vector machine (SVM) was proposed to diagnose PV DC series arc faults based on the current data. The block diagram of this algorithm is shown in Figure 1.

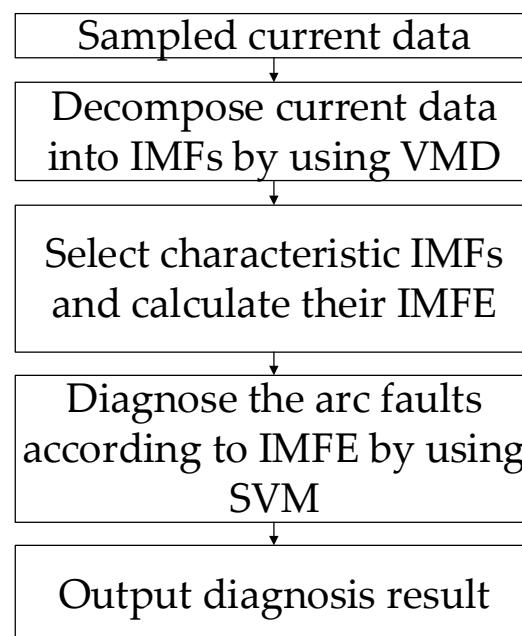


Figure 1. Block diagram of the proposed diagram.

This algorithm consists of VMD, IMFE, and SVM. The advantages of the proposed algorithm are as follows:

- VMD avoids mode mixing, which makes sure that decomposition is meaningful;
- IMFE is stable and can be used to describe signals from the multi-scale;
- SVM has great classification effects on nonlinear problems.

## 2. Arc Fault Diagnosis Algorithm

### 2.1. Arc Fault Data Feature Analysis

The proposed method of our research focused on its application in power stations located in Hubei Enshi Badong. One of them was the Songjialiangzi power plant. Photos of it are shown in Figure 2a,b.



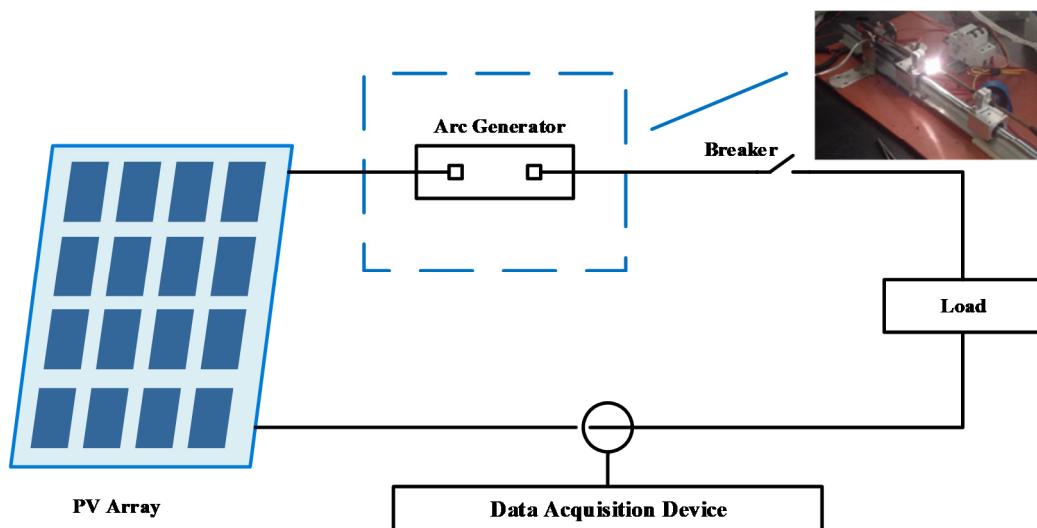
(a)



(b)

**Figure 2.** Photos of the Songjialiangzi power plant: (a) Frontal view; (b) Side view. (Writings in the figure are not relevant to the meaning of the figure).

In order to analyze the features of arc fault data, a photovoltaic DC arc generation experiment platform was built. The system chart is shown in Figure 3.



**Figure 3.** The experimental platform.

This system consisted of a PV array, a breaker, a load resistance, a current and voltage sampling device, and an arc generator. The sampling device consisted of a Hall current sensor and an oscilloscope. The sampling frequency was 500 kHz. The arc generator was established according to the UL1699B standard, whose structure is shown in Figure 4. The photos of sampling devices, the arc generator, and the arc are shown in Figures 5–7.

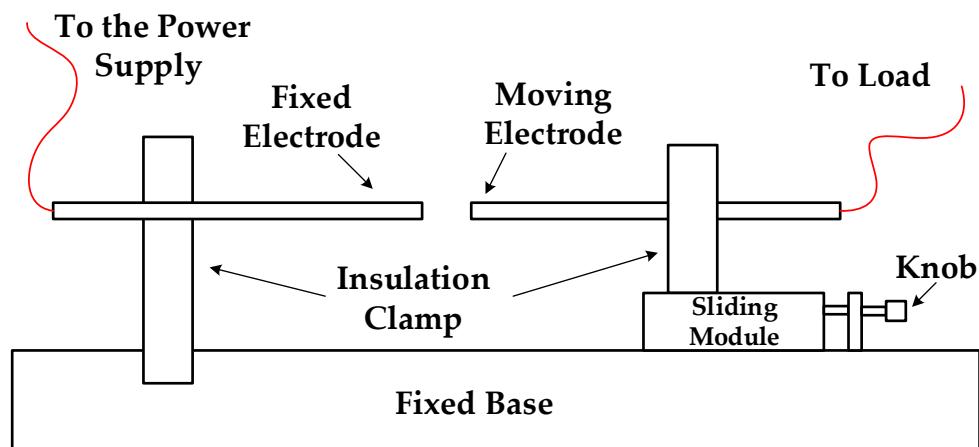


Figure 4. Structure of the arc generator.

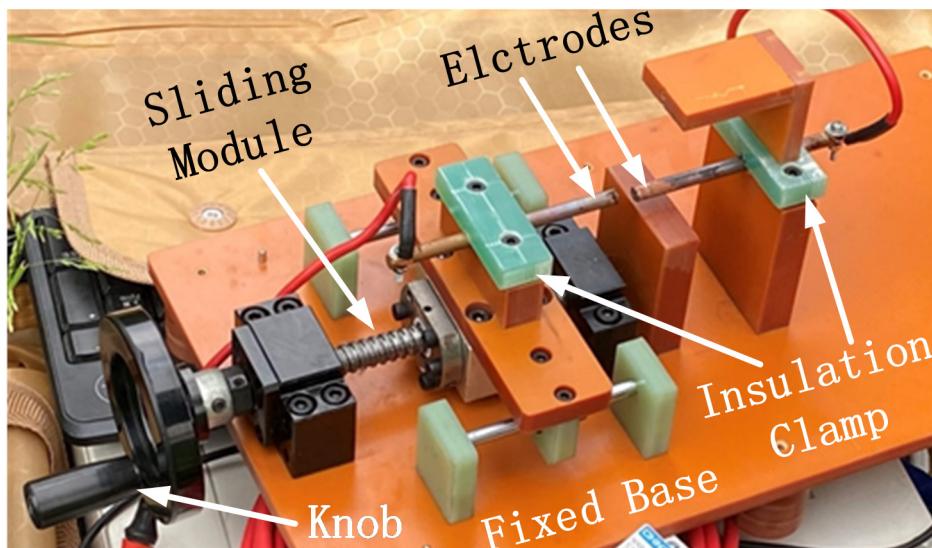


Figure 5. Photo of the arc generator.

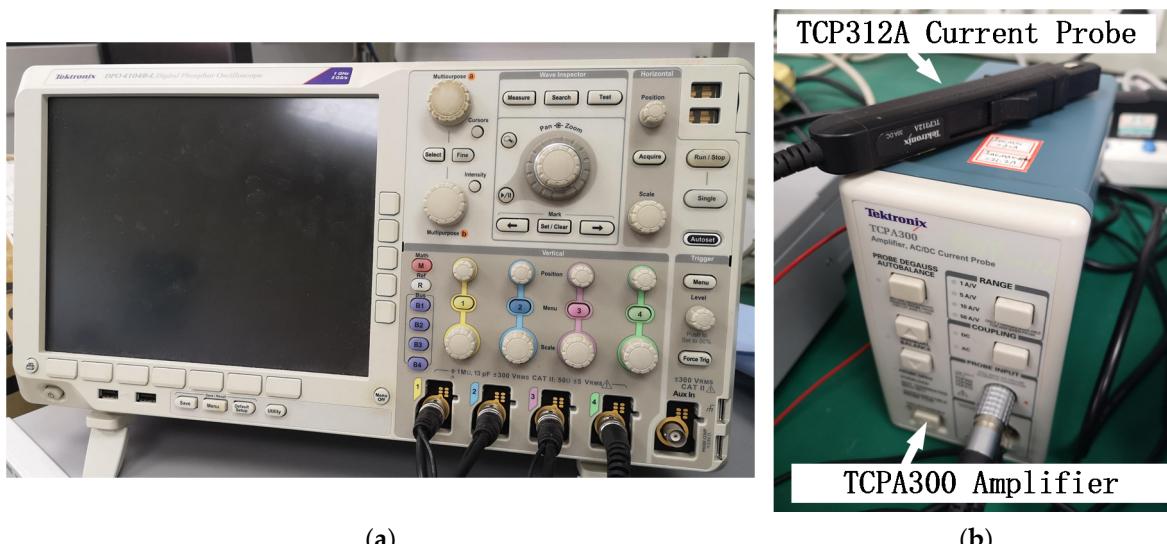
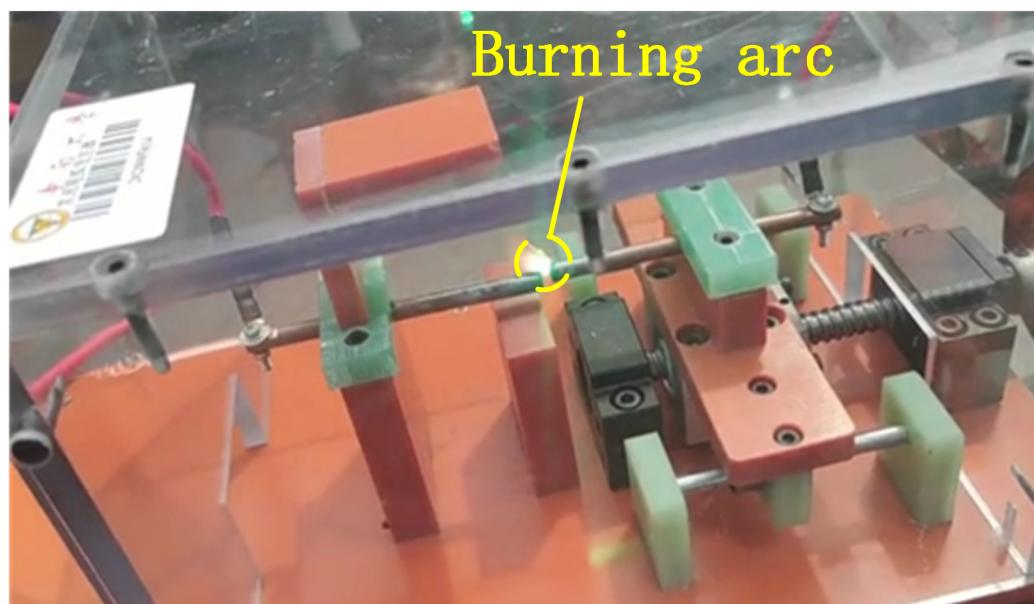


Figure 6. Photos of the sampling devices: (a) oscilloscope; (b) Hall current sensor.

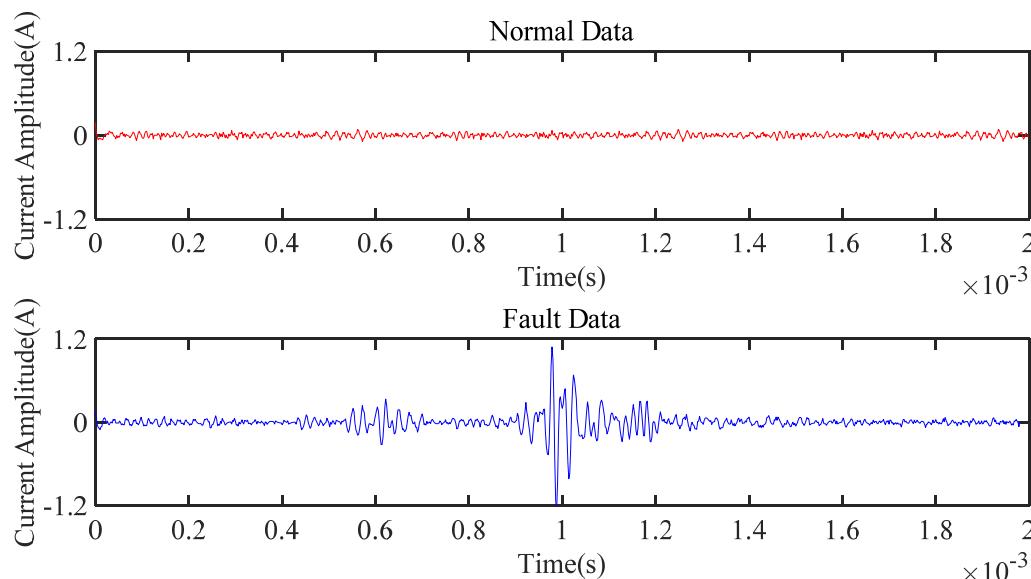


**Figure 7.** Photo of the arc.

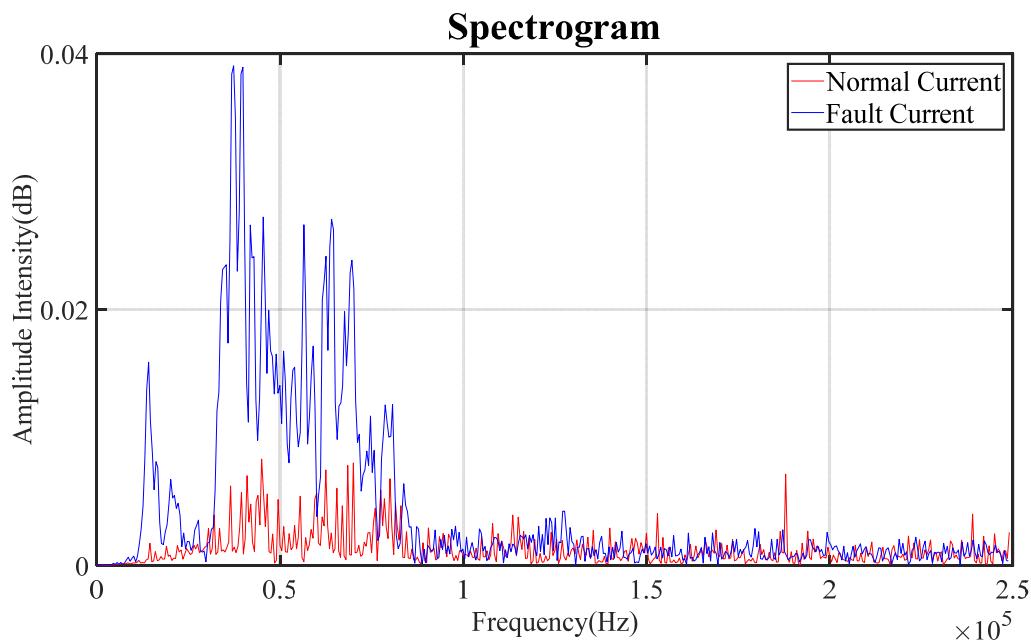
In the experiment, the knob of the arc generator was slowly rotated to shorten the distance between the electrodes. As the distance became smaller, an arc was generated between the electrodes. Then the oscilloscope collected the current data. The current data of an arc fault under different load conditions was obtained by changing the value of the load resistance. By removing the arc generator, normal current data were also obtained.

The unprocessed signal showed that the low frequency AC noise existed in both normal and arc fault situations. This type of noise does not help to detect the fault and makes the problem complicated. Therefore, a high-pass filter with a cut-off frequency of 30 kHz was used to remove the low frequency noise. The comparison of processed current data is shown in Figures 8 and 9:

### Comparison Between Normal Data and Fault Data



**Figure 8.** The comparison of current waveforms in the time-domain.



**Figure 9.** The comparison of current waveforms in the frequency-domain.

Figure 8 shows that in the time domain, the current of the arc fault oscillated wildly compared to normal situations. Figure 9 indicates that the fault current was significantly different from the normal one in the 30–100 kHz band [10]. VMD was used to process the signal with the aim of analyzing the frequency components related to the arc accurately.

## 2.2. Variational Model Decomposition

VMD decomposes a time-series signal  $f(t)$  into  $K$  IMFs. Each IMF has a central frequency. VMD is an optimization problem based on IMF bandwidth minimization. It can be described by Formula (1). In Formula (1),  $K$  is the number of IMFs,  $k$ -th IMF is presented by  $u_k(t)$ , the central frequency of  $u_k(t)$  is presented by  $\omega_k$ ,  $(a * b)$  represents a convolved with  $b$ .  $\|a\|_2$  represents the L2-norm of  $a$ , which can be calculated by Formula (2), and  $\partial_t X$  represents the partial derivative of  $X$  with respect to  $t$ .

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \cdot u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{subject to: } \sum_{k=1}^K u_k(t) = f(t) \end{cases} \quad (1)$$

$$\|a\|_2 = \sqrt{(\text{Re}(a))^2 + (\text{Im}(a))^2} \quad (2)$$

Quadratic penalty term  $a$  and Lagrangian multiplier  $\lambda(t)$  are used to transform this problem into a non-constrained problem as shown in Formula (3).  $\langle x, y \rangle$  represents the inner product of  $x$  and  $y$ .

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \cdot u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \quad (3)$$

This problem can be solved by using alternating the direction multiplier method (ADMM) until an augmented Lagrange saddle point is found.  $u_k$ ,  $\omega_k$ , and  $\lambda$  can be updated by using Formulas (4)–(6). For the convenience of subsequent analysis, Formulas (4)–(6)

are given in the frequency domain. In Formulas (4)–(6),  $\hat{u}_k(\omega)$  is the  $k$ -th IMF, and  $\hat{f}(\omega)$  is the Fourier transformation result of the  $k$ -th IMF.  $\theta$  is the noise tolerance parameter.

$$u_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \hat{\lambda}(\omega)}{1 + 2\alpha(\omega - \omega_k)^2} \quad (4)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (5)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \theta \left[ \hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right] \quad (6)$$

The iteration of  $u_k$ ,  $\omega_k$ , and  $\lambda$  can be stopped when an augmented Lagrange saddle point is found, which can be judged by Formula (7):

$$\frac{\sum_k \| \hat{u}_k^{n+1} - \hat{u}_k^n \|_2^2}{\| \hat{u}_k^n \|_2^2} < \varepsilon \quad (7)$$

When Formula (7) is satisfied, the original signal  $f(t)$  has been decomposed into  $K$  IMFs  $\{u_k\}$ . The central frequency of  $u_k$  is  $\omega_k$ .

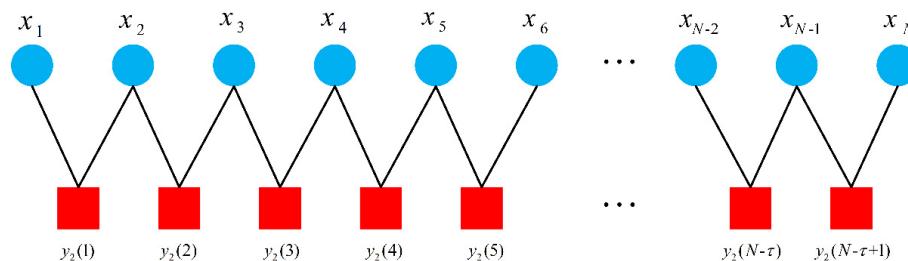
### 2.3. Multi-Scale Fuzzy Entropy

After  $K$  IMFs are obtained by using VMD, the parameter calculated from the IMFs must be chosen as the feature of the arc. This parameter should basically maintain the same value when an arc fault does not occur. Once an arc fault occurs, this parameter should change noticeably. In this research, IMFE was chosen as the parameter. IMFE is used to measure the complexity of the time-series signal under different scales. It can be calculated through following steps.

Firstly, some sampling points in the original time-series signal were selected by using the sliding window method. Then the selected sampling points sequence  $\{u(i) : 1 \leq i \leq N\}$  needed to be coarsened. The coarsened sequence was established by using Formula (8):

$$y_\tau(j) = \frac{1}{\tau} \sum_{i=j}^{j+\tau-1} u(i), 1 \leq j \leq N - \tau + 1 \quad (8)$$

In Formula (8),  $\tau$  represents the length of a factor in the coarsened sequence. This process is illustrated in Figure 10.



**Figure 10.** Schematic diagram of the improved multi-scale fuzzy entropy algorithm.

Then the coarsened sequence  $\{y_\tau(j) : 1 \leq j \leq (N - \tau + 1)\}$  was converted into a set of vectors  $\{X_l^m : 1 \leq l \leq (N - \tau - m + 2)\}$ . Formula (9) indicates the specific calculation:

$$X_j^m = \{y_\tau(j), y_\tau(j+1), \dots, y_\tau(j+m-1)\} - \overline{y_\tau(j)}, j = 1, 2, \dots, N-m-\tau+2 \quad (9)$$

In Formula (9),  $\overline{y_\tau(j)}$  can be calculated by using Formula (10):

$$\overline{y_\tau(j)} = \frac{1}{m} \sum_{j'=0}^{m-1} y_\tau(j+j') \quad (10)$$

The distance  $d[X_i^m, X_j^m]$  between  $X_i^m$  and  $X_j^m$  is defined as the maximum value of the difference between the two corresponding elements:

$$d_{ij}^m = d[X_i^m, X_j^m] = \max_{k \in (0, m-1)} \left\{ \left| y_\tau(i+k) - \overline{y_\tau(i)} - (y_\tau(j+k) - \overline{y_\tau(j)}) \right| \right\} \quad (11)$$

$i, j = 1, 2, \dots, N-m-\tau+2, i \neq j$

The similarity degree  $D_{ij}^m$  of the vectors  $X_i^m$  and  $X_j^m$  is defined by a fuzzy function  $\mu(d_{ij}^m, \beta, r)$ . In Formula (12),  $r$  is calculated according to the original data and  $\beta$  is an adjustable constant.  $\rho$  is an adjustment factor used to enhance the ability of noise tolerance.

$$D_{ij}^m = \mu(d_{ij}^m, \beta, r) = \begin{cases} 1, 0 \leq d_{ij}^m \leq \rho r \\ e^{-\ln 2(\frac{d_{ij}^m}{r})^\beta}, d_{ij}^m > \rho r \end{cases} \quad (12)$$

Once a similarity degree is defined, IMFE can be calculated according to Formula (13):

$$\text{FuzzyEn}(m, \beta, r) = \lim_{N \rightarrow \infty} [\ln \phi^m(\beta, r) - \ln \phi^{m+1}(\beta, r)] \quad (13)$$

In Formula (13):

$$\phi^m(\beta, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left( \frac{1}{N-m-1} \sum_{\substack{j=1 \\ j \neq i}}^{N-m} D_{ij}^m \right) \quad (14)$$

#### 2.4. The Steps of Arc Fault Diagnosis Algorithm

Aiming at the non-linear and non-stationary characteristics of PV DC current, arc diagnosis is particularly suitable for combining VMD and IMFE together. The specific steps of the arc fault diagnosis algorithm are described below.

For the first step, DC current data were decomposed into several IMFs by VMD. Firstly, we initialized the number of IMFs,  $K$  IMF signal set  $\{\hat{u}_k^1\}$ , central frequency set  $\{\hat{\omega}_k^1\}$ , and Lagrangian  $\hat{\lambda}^1$ . We set  $n$  as 0. We set the value of the quadratic penalty term

a. We set the noise tolerance parameter  $\theta$ . Then we updated  $\{\hat{u}_k^1\}$ ,  $\{\hat{\omega}_k^1\}$ , and  $\hat{\lambda}^1$  by using Formulas (4)–(6) until Formula (7) was satisfied. Eventually the  $K$  IMFs  $\{u_k\}$  and the corresponding central frequency  $\{\omega_k\}$  were obtained.

In the second step, based on the wave form of IMFs  $\{u_k\}$ , IMFs related to arc faults were selected.

For the third step, we used IMFE to analyze the effective intrinsic mode function. Firstly, we initialized the length of the sliding window  $N$ , the length of a factor in the coarsened sequence  $\tau$ , the dimension of the converted vector  $m$ , the values of the adjustable constants  $\rho$  and  $\beta$ , and the formula for calculating  $r$  according to the original data. Then  $N$  sampling points from an IMF signal  $u_k$  were selected by using a sliding window.

The selected sampling points  $\{u(i) : 1 \leq i \leq N\}$  were coarsened by using Formula (8). Then the coarsened sequences  $\{y_\tau(j) : 1 \leq j \leq (N - \tau + 1)\}$  were converted into vectors by using Formulas (9) and (10). At last, by defining distance and the similarity degree function of these vectors  $\{X_l^m : 1 \leq l \leq (N - \tau - m + 2)\}$ , the IMFE could be calculated by using Formulas (13) and (14). By sliding the window forward, IMFE based on IMF could be calculated eventually.

In the final step, we took the IMFE obtained in step 3 and the corresponding labels (normal data were marked as 1, and fault data were marked as 0) of them as the training objectives. These data were used to train the support vector machine. The remaining untrained data were used to verify the training effect of the support vector machine.

### 3. The Validation of the Arc Fault Diagnosis Algorithm

#### 3.1. VMD Parameter Selection

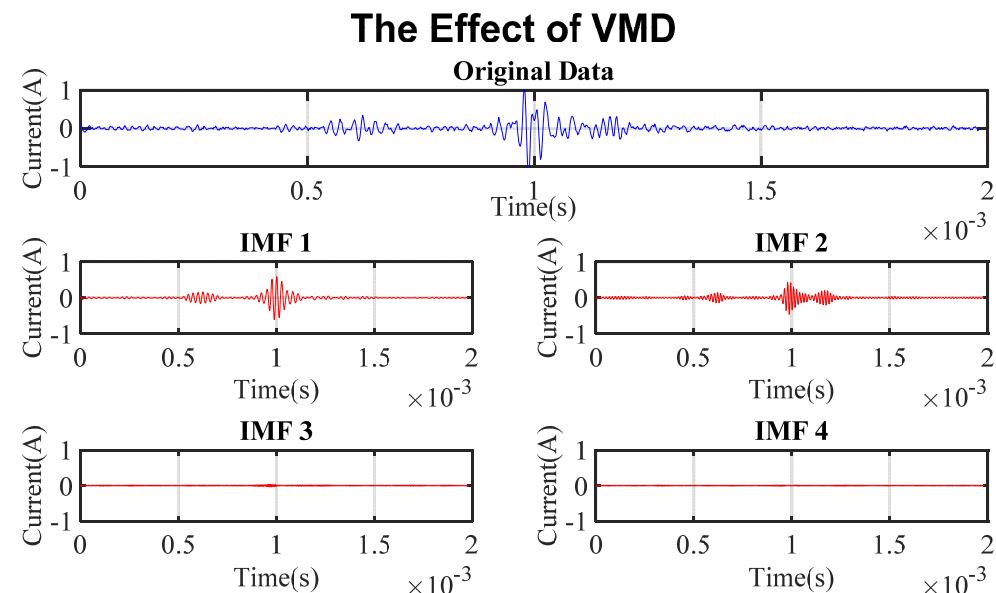
Before performing VMD, the number of decomposed IMFs must be selected. When the K value is small, the decomposed IMFs cannot contain all the information of the original data. When the K value is large, mode mixing will occur between the IMFs, which makes these IMFs distorted signals. Therefore, the selection of the K value is very important.

In this research, the K value was chosen by decomposing the signal several times with different K values. After a set number of iterations with N as 0, set  $\{\hat{u}_k^1\}$ ,  $\{\hat{\omega}_k^1\}$ , and  $\hat{\lambda}^1$  as suitable size zero matrix, we set the noise tolerance  $\theta$  as 0.5 and set the quadratic penalty term a as 2000. Table 1 shows the results of decomposition when K varies from 1 to 5.

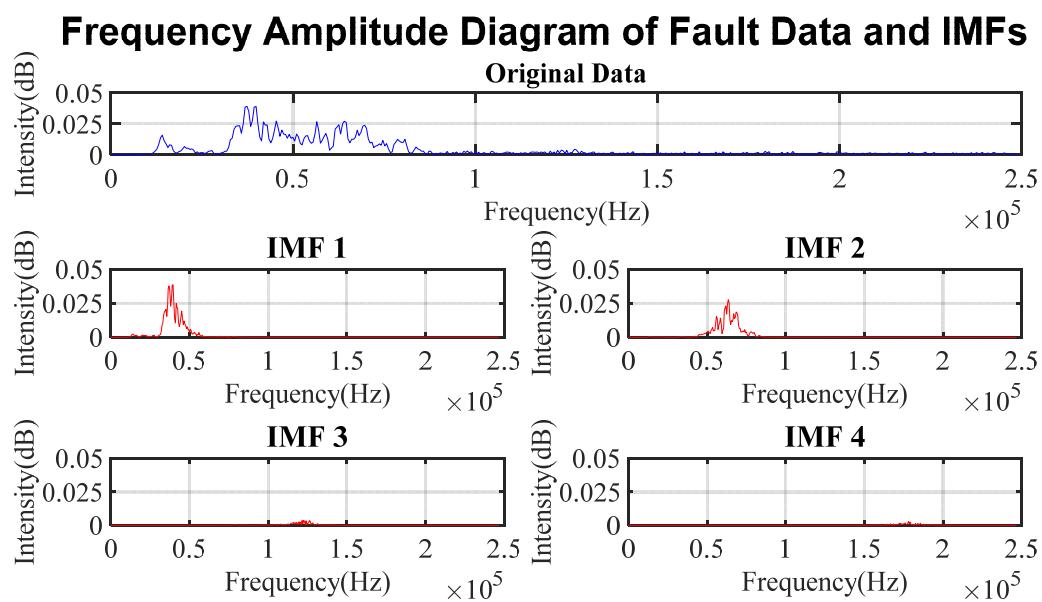
**Table 1.** K from 1 to 5, the center frequency of IMFs after VMD.

K Value	IMF1	IMF2	IMF3	IMF4	IMF5
K = 1	40.7 kHz	N/A	N/A	N/A	N/A
K = 2	39.3 kHz	64.1 kHz	N/A	N/A	N/A
K = 3	39.3 kHz	64.1 kHz	122.4 kHz	N/A	N/A
K = 4	39.3 kHz	64.1 kHz	121.9 kHz	179.3 kHz	N/A
K = 5	37.9 kHz	48.8 kHz	65.8 kHz	123.3 kHz	193.9 kHz

As shown in Table 1, IMF1 and IMF2 have relatively close central frequencies when K = 5. Mode mixing may occur in this situation. In order to decide whether 4 or 5 is the appropriate K value, the IMFs when K is 4 and 5 are shown in Figures 11–14.

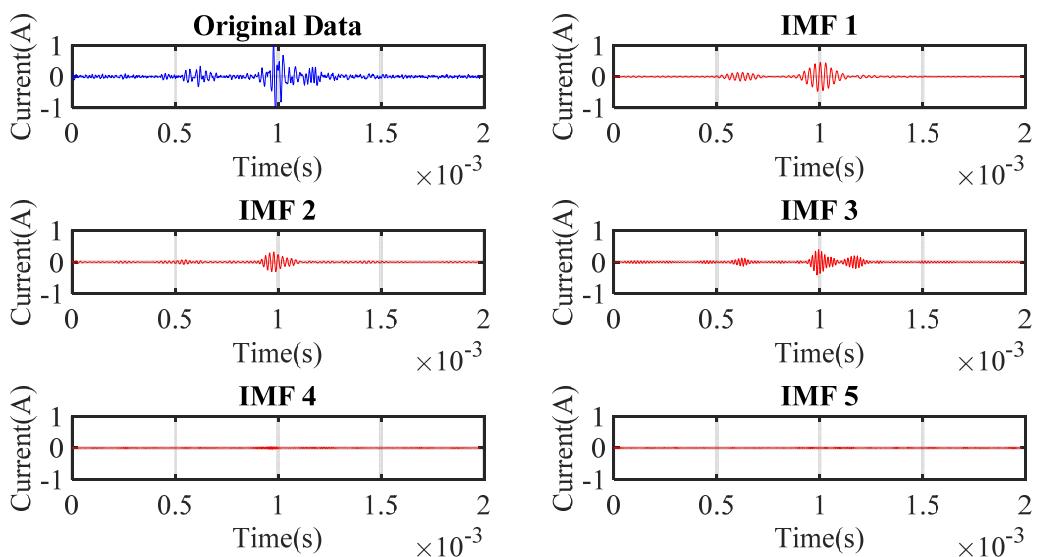


**Figure 11.** The waveform of IMFs when K = 4.



**Figure 12.** Spectrogram of IMFs when  $K = 4$ .

### The Effect of VMD



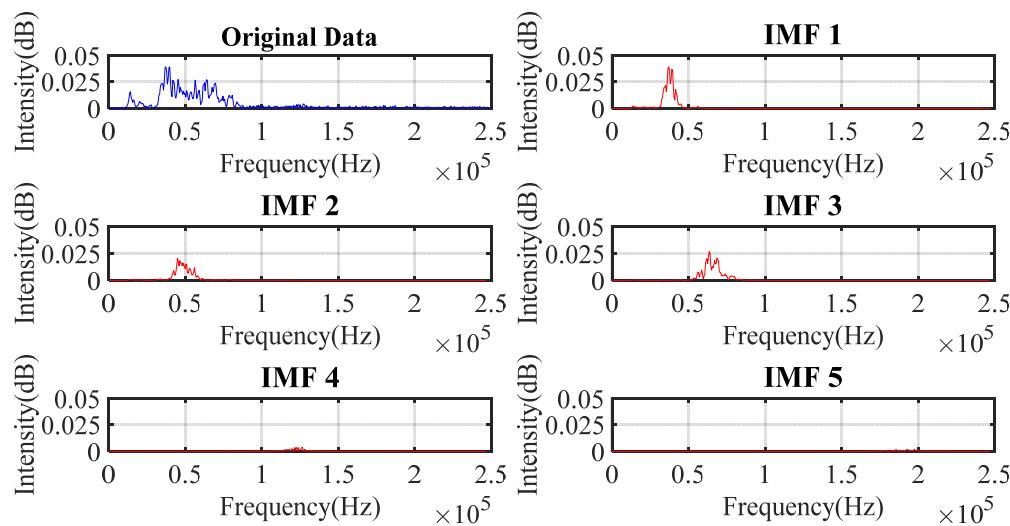
**Figure 13.** The waveform of IMFs when  $K = 5$ .

Figure 14 indicates that IMF1, IMF2, and IMF3 had an obvious overlapping part in the frequency domain. This phenomenon indicates that from the perspective of avoiding mode mixing,  $K = 4$  was a more appropriate choice than  $K = 5$ .

Except for mode mixing, the stability of the results can also be the criterion for choosing the  $K$  value. Tables 2 and 3 show the central frequencies of the IMFs decomposed from several sets of data when  $K$  was set as either 4 or 5.

Tables 2 and 3 indicate that the stability when  $K = 4$  was better than the stability when  $K = 5$ . Therefore, from the perspectives of stability and mode mixing,  $K$  should be set as 4 in this research.

## Frequency Amplitude Diagram of Fault Data and IMFs



**Figure 14.** Spectrogram of IMFs when  $K = 5$ .

**Table 2.**  $K = 4$ , the center frequency of IMFs after VMD of multiple sets of data.

Type	Label	IMF1	IMF2	IMF3	IMF4
normal	1	42.2 kHz	64.5 kHz	125.6 kHz	187.1 kHz
normal	2	41.7 kHz	65.0 kHz	124.3 kHz	190.4 kHz
normal	3	42.8 kHz	64.1 kHz	125.2 kHz	179.5 kHz
fault	1	42.2 kHz	64.7 kHz	126.7 kHz	188.1 kHz
fault	2	42.1 kHz	66.8 kHz	124.1 kHz	186.6 kHz
fault	3	42.6 kHz	64.3 kHz	123.8 kHz	183.9 kHz

**Table 3.**  $K = 5$ , the center frequency of IMFs after VMD of multiple sets of data.

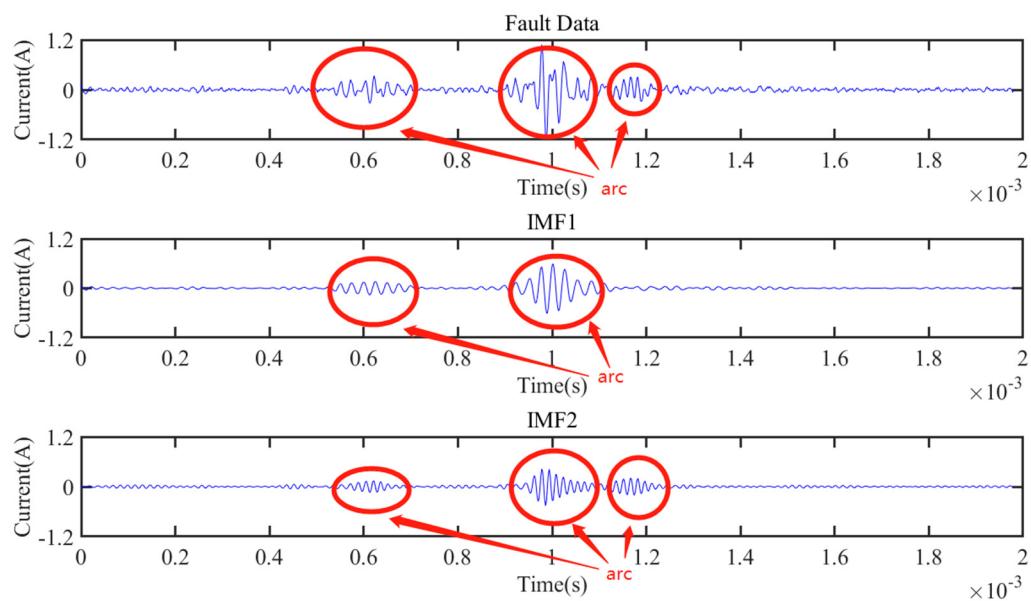
Type	Label	IMF1	IMF2	IMF3	IMF4	IMF5
normal	1	42.2 kHz	64.1 kHz	95.7 kHz	127.5 kHz	187.4 kHz
normal	2	41.7 kHz	64.8 kHz	105.2 kHz	156.1 kHz	203.0 kHz
normal	3	42.9 kHz	63.9 kHz	110.9 kHz	160.7 kHz	210.0 kHz
fault	1	40.6 kHz	55.4 kHz	66.9 kHz	128.9 kHz	188.2 kHz
fault	2	40.9 kHz	57.6 kHz	68.1 kHz	124.5 kHz	186.7 kHz
fault	3	41.8 kHz	60.7 kHz	69.1 kHz	123.8 kHz	163.8 kHz

### 3.2. Selection of IMF and Calculation of IMFE

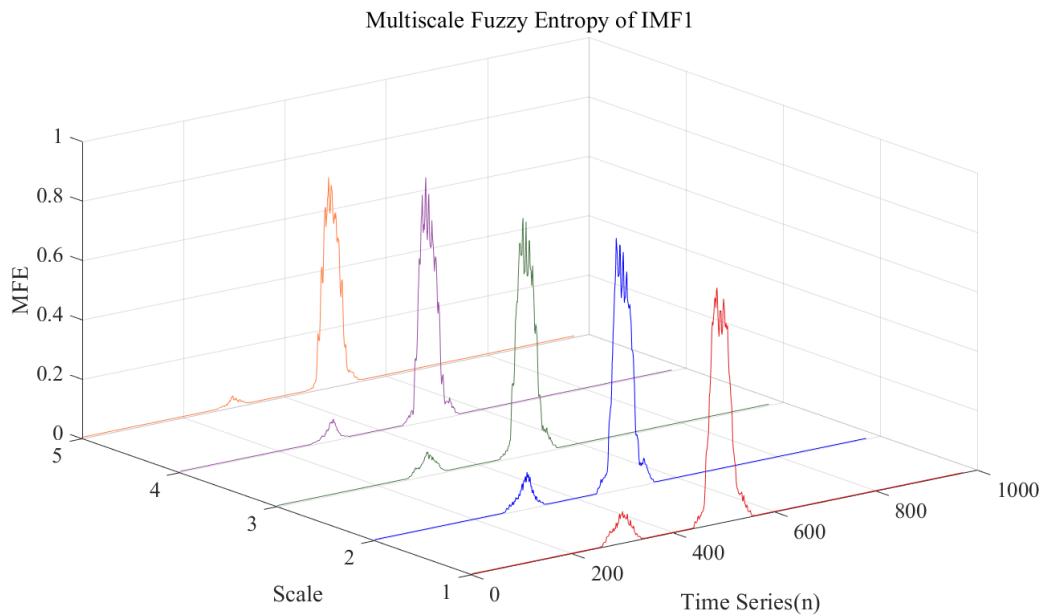
Figure 9 indicates that the frequency bandwidth related to arc faults ranges from 30 kHz to 100 kHz. Figure 12 illustrate that when  $K = 4$ , IMF1 and IMF2 are in this bandwidth. In addition, Figure 15 shows that once the current fluctuates wildly because an arc fault occurs, IMF1 and IMF2 will fluctuate synchronously. Therefore, IMF1 and IMF2 should be chosen as the IMFs related to the arc fault.

As proposed in Section 2.3, the IMFEs of these 2 IMFs were calculated. The length of the sliding window, which is represented by  $N$ , was chosen as 20. The length of a factor in the coarsened sequence, which is represented by  $\tau$ , was less than or equal to 5. The dimension of vectors converted from the coarsened sequence, which is represented by  $m$ , was 3. As for parameters in the similarity degree function,  $\rho = 1$ ,  $\beta = 2$ ,  $r$  is calculated by Formula (15), and  $S$  is the standard deviation of all the original data. The IMFEs calculated from IMF1 and IMF2 are shown in Figures 16 and 17.

$$r = 0.15 \cdot S \quad (15)$$



**Figure 15.** The observed arc part of original data, IMF1 and IMF2.



**Figure 16.** IMFE under different scales of IMF1.

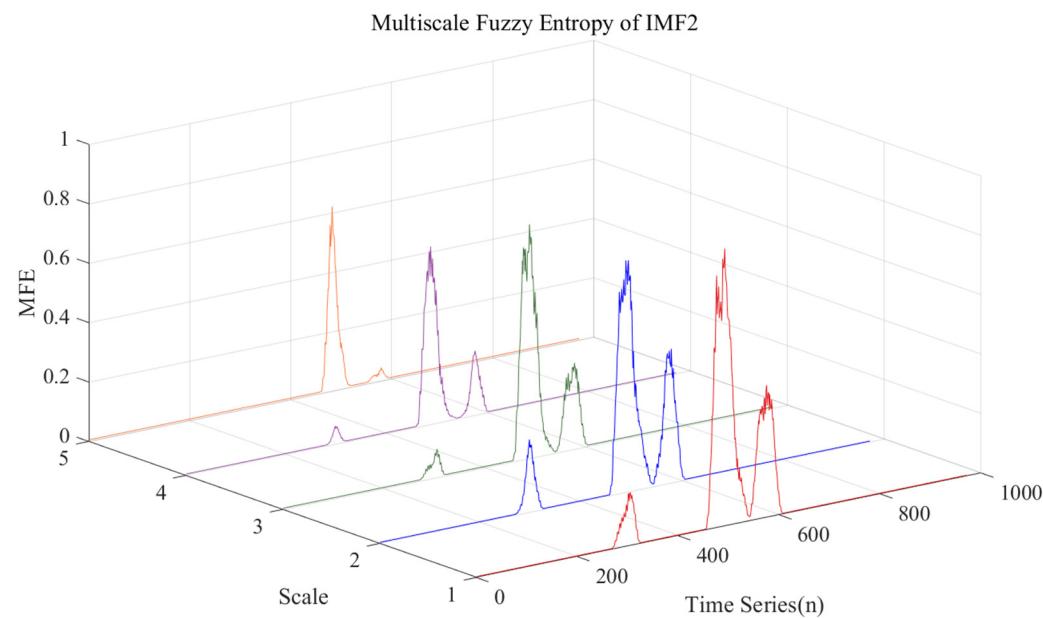
By combining Figures 15–17, the IMFE successfully met the requirement of the features of arc diagnosis. The value of IMFE was close to 0 when an arc fault did not happen. Once an arc fault occurred, the value of IMFE increased noticeably, and the magnitude was proportional to the severity of the arc fault.

Before conducting validation, the choices of  $\tau$  and  $N$  need to be discussed. The IMFE calculation above was conducted when  $\tau$  was less than 5. When the length of the sliding window and the length of the factor are chosen differently, the result could be different.

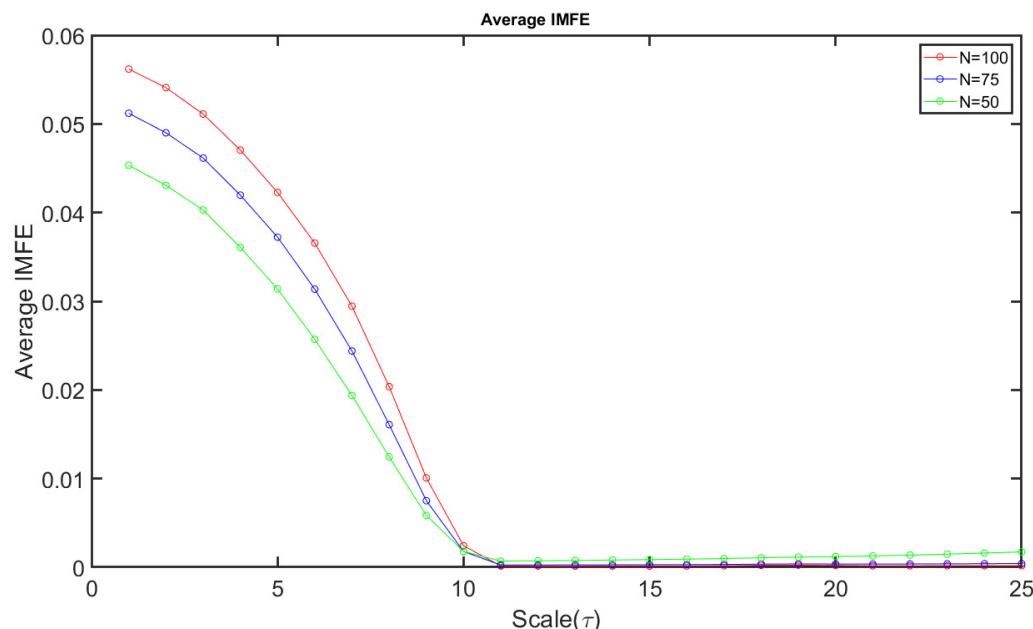
Figure 18 shows the average IMFE-scale curve when the length of the sliding window is chosen as 100, 75, or 50. The IMFE is calculated from IMF1 in Figure 18.

Figure 18 indicates that when the  $N$  chosen is larger, the average IMFE is larger. This means that the fluctuation in IMFE caused by an arc is stronger when the chosen  $N$  is larger. However, a larger length of the sliding window will make the calculation more time-consuming. In Figures 16 and 17,  $N$  is set as 20. The fluctuation is still strong

enough for arc detection. Therefore,  $N$  should be chosen based on computing power and speed needs.



**Figure 17.** IMFE under different scales of IMF2.



**Figure 18.** Average IMFE-scale curve when  $N = 100, 75$ , or  $50$ .

Regardless of  $N$ , the average IMFE always decays to a small value when  $\tau$  is equal to 10. A possible explanation can be stated as the following: in the IMF1 which IMFE is calculated from, the arc portion counts for the same number of data points regardless of the value of  $N$ . When the chosen  $\tau$  is larger than the number of arc fault data points, the IMFEs which contain arc data points are reduced by the presence of normal data points, which can cause the arc to go undetected based on IMFE. Therefore, the IMFE of a small  $\tau$  value will make sure that an arc with short bandwidth can be detected. If fluctuations on the IMFE exists, the larger the  $\tau$  of IMFE is, and the wider the bandwidth of the arc is.

### 3.3. Fault Diagnosis and Validation

In order to validate the algorithm, 5100 current sampling points were obtained from the experimental platform proposed in Section 2.1. In this process, 24 experiments, including 12 normal experiments and 12 arc fault experiments with different load resistances, were carried out. SVM was used to detect the arc fault based on the IMFEs of IMF1 and IMF2. Therefore, these data were divided into a training set and a verification set. The training set contains 70% of the data while the verification set contains the rest.

The SVM program used in this paper was the LibSVM program provided by Chang Chih-Chung and Lin Chih-Jen [41]. The radial basis function was selected as the kernel function of the SVM, and the particle swarm optimization (PSO) algorithm was used to optimize the penalty parameter  $c$  and the kernel function parameter  $g$  of the SVM.

The diagnosis result type can be divided into TP (true positive, an arc fault sample was diagnosed as an arc fault sample), FP (false positive, a normal sample was diagnosed as an arc fault sample), TN (true negative, a normal sample was diagnosed as a normal sample) and FN (false negative, an arc fault sample was diagnosed as a normal sample). From the definitions of these result types, FP and FN were the wrong diagnosis results and the others were the right diagnosis results.

The 1530 samples in the validation set were diagnosed by SVM and the results are shown in Table 4. Table 4 illustrates that the proposed arc fault diagnosis algorithm had a great effect on detecting arc faults. Of the 1530 samples, 1514 were diagnosed correctly. The accuracy of this algorithm was 99.0% on this validation data set.

**Table 4.** SVM Training Effects.

Result Type	TP	FP	TN	FN	Total
Number	985	11	529	5	1530
Percentage	64.4%	0.7%	34.6%	0.3%	100%

### 4. Conclusions

This research proposed a fault diagnosis method based on VMD, IMFE, and SVM for PV arc diagnosis. An experimental platform was established to obtain the training and validation current data. During the arc diagnosis, the current data were first decomposed into several IMFs by using VMD. Then the IMFs related to an arc were picked for IMFE calculation. After IMFEs of these IMFs were obtained, they were used as the input of SVM to detect the arc fault. According to the results of validation, the proposed arc fault diagnosis method could be used to detect an arc fault based on current data and was relatively accurate.

The advantages of this method can be stated as follows: using VMD makes sure that the decomposed signals have a central frequency and limited frequency bandwidth; by choosing an appropriate  $K$  value, mode mixing can be avoided, and IMFs are guaranteed to make sense; compared to other decomposition methods, the result of VMD contains relatively rich information, which is beneficial to the subsequent analysis. Using IMFE successfully converts fluctuation in IMFs into amplitude change in IMFEs, which makes using SVM possible. Compared to developed methods, the advantages of the proposed method can be stated as follows: compared to using sensors to detect the arc [42], the proposed method only requires users to add sampling devices to the scene. High sampling rate devices can be chosen with same investment. Fewer devices needing to be installed also makes the proposed method easy to apply in different situations. Compared to using thermal imaging [43], the proposed method can significantly save on the cost of equipment installation. The proposed method is also more universal than thermal imaging. Compared to other frequency-domain methods [44], this method avoids the choice of functions such as wavelet basis, window functions, etc. By decreasing the length of the sliding window, the calculation speed of the proposed method can be improved, though the significance of the results may be weakened at the same time. Users can find a proper balance between

them based on the requirements of the actual application. Furthermore, this algorithm has potential application in bearing fault diagnosis, mechanical fault analysis of circuit breakers, multi-level inverter analysis, and so on.

However, compared to using sensors and thermal imaging, accuracy may be a disadvantage of the proposed method. The rapidity of the proposed algorithm was not validated in this research. It is possible that an integrated online arc fault diagnostic device could be established based on this algorithm.

**Author Contributions:** L.W.’s main contributions are conceptualization, project administration and supervision. H.Q.’s main contributions are writing original draft and writing review & editing. P.Y.’s main contributions are investigation, methodology, data curation and software. L.M.’s main contribution is resources. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is funded by the National Key Research and Development Program of China (No. 2018YFB1500800).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** According to the requirement of funding program, the data should not be shared in 5 years.

**Acknowledgments:** Thanks to Mu Longhua and his experimental team from Tongji University for providing technical support for the experiments in this article.

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

## References

1. PVPS; Kaizuka, I.; Jäger, W.A. Snapshot of Global PV Markets. 2020. Available online: [https://www.researchgate.net/publication/324728482\\_2017\\_SNAPSHOT\\_OF\\_GLOBAL\\_PHOTOVOLTAIC\\_MARKETS](https://www.researchgate.net/publication/324728482_2017_SNAPSHOT_OF_GLOBAL_PHOTOVOLTAIC_MARKETS) (accessed on 8 July 2021).
2. Taylor, M.; Ralon, P.; Ilas, A. The Power to Change: Solar and Wind Cost Reduction Potential to 2025. International Renewable Energy Agency, 2016; p. 3. Available online: <https://www.irena.org/publications/2016/Jun/The-Power-to-Change-Solar-and-Wind-Cost-Reduction-Potential-to-2025> (accessed on 8 July 2021).
3. Hossam, A.A.; Ahmed, E.E.; Ibrahim, B.M.T. A new monitoring technique for fault detection and classification in PV systems based on rate of change of voltage-current trajectory. *Int. J. Electr. Power Energy Syst.* **2021**, *133*. [[CrossRef](#)]
4. Shi, Y.; Sun, Y.; Liu, J.; Du, X. Model and stability analysis of grid-connected PV system considering the variation of solar irradiance and cell temperature. *Int. J. Electr. Power Energy Syst.* **2021**, *132*. [[CrossRef](#)]
5. Jiang, Y.; Kang, L.; Liu, Y. The coordinated optimal design of a PV-battery system with multiple types of PV arrays and batteries: A case study of power smoothing. *J. Clean. Prod.* **2021**, *310*. [[CrossRef](#)]
6. Haffaf, A.; Lakdja, F.; Abdeslam, D.O.; Meziane, R. Monitoring, measured and simulated performance analysis of a 2.4 kWp grid-connected PV system installed on the Mulhouse campus, France. *Energy Sustain. Dev.* **2021**, *62*, 44–55. [[CrossRef](#)]
7. Houssein, E.H.; Mahdy, M.A.; Fathy, A.; Rezk, H. A modified Marine Predator Algorithm based on opposition based learning for tracking the global MPP of shaded PV system. *Expert Syst. Appl.* **2021**, *183*. [[CrossRef](#)]
8. Yurtseven, K.; Karatepe, E.; Deniz, E. Sensorless fault detection method for photovoltaic systems through mapping the inherent characteristics of PV plant site: Simple and practical. *Sol. Energy* **2021**, *216*, 96–110. [[CrossRef](#)]
9. Shapiro, F.R.; Radibratovic, B. DC arc flash hazards and protection in photovoltaic systems. In Proceedings of the 2013 IEEE 39th Photovoltaic Specialists Conference (PVSC), Tampa Bay, FL, USA, 16–21 June 2013.
10. 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](#)]
11. Dhar, S.; Patnaik, R.K.; Dash, P.K. Fault detection and location of photovoltaic based DC microgrid using differential protection strategy. *IEEE Trans Smart Grid* **2018**, *9*, 4303–4312. [[CrossRef](#)]
12. Mu, L.; Wang, Y.; Jiang, W.; Zhang, F. Study on characteristics and detection method of DC arc fault for photovoltaic system. *Proc. CSEE* **2016**, *36*, 5236–5244.
13. Lu, S.; Ma, R.; Sirojan, T.; Phung, B.T.; Zhang, D. Lightweight transfer nets and adversarial data augmentation for photovoltaic series arc fault detection with limited fault data. *Int. J. Electr. Power Energy Syst.* **2021**, *130*. [[CrossRef](#)]
14. Humbert, J.B.; Schweitzer, P.; Weber, S. Serial-arc detection by use of Spectral Dispersion Index (SDI) analysis in a low-voltage network (270V HVDC). *Electr. Power Syst. Res.* **2021**, *196*. [[CrossRef](#)]
15. Chen, S.; Li, X.; Meng, Y.; Xie, Z. Wavelet-based protection strategy for series arc faults interfered by multicomponent noise signals in grid-connected photovoltaic systems. *Sol. Energy* **2019**, *183*, 327–336. [[CrossRef](#)]

16. Goodman, M. How ultrasound can detect electrical discharge non-invasively and help eliminate arc flash incidents. In Proceedings of the 2007 Electrical Insulation Conference and Electrical Manufacturing Expo, Nashville, TN, USA, 22–24 October 2007.
17. Panetta, S. Design of arc Flash protection system using solid state switch, photo detection, with parallel impedance. In Proceedings of the 2013 IEEE IAS Electrical Safety Workshop, Dallas, TX, USA, 11–15 March 2013.
18. Restrepo, C.E. Arc Fault detection and discrimination methods. In Electrical Contacts-2007. In Proceedings of the 53rd IEEE Holm Conference on Electrical Contacts, Pittsburgh, PA, USA, 16–19 September 2007.
19. Xiong, Q.; Ji, S.; Zhu, L.; Zhong, L.; Liu, Y. A novel DC arc fault detection method based on electromagnetic radiation signal. *IEEE Trans. Plasma Sci.* **2017**, *45*, 472–478. [\[CrossRef\]](#)
20. Guo, S.Y.; Jones III, J.L.; Dooley, A.S. DC arc Detection and Prevention Circuit and Method. U.S. Patent 6,683,766 B1, 27 January 2004.
21. Zhen, M.; Li, W.; Qiang, G.S. The characteristics of DC arc faults current. In Proceedings of the 2013 15th European Conference on Power Electronics and Applications (EPE), Lille, France, 2–6 September 2013.
22. Yao, X.; Herrera, L.; Ji, S.; Zou, K.; Wang, J. Characteristic study and time-domain discrete-wavelet-transform based hybrid detection of series DC arc faults. *IEEE Trans. Power Electron.* **2014**, *29*, 3103–3115. [\[CrossRef\]](#)
23. Chae, S.; Park, J.; Oh, S. Series DC Arc fault detection algorithm for DC microgrids using relative magnitude comparison. *IEEE Trans. Emerg. Sel. Top. Power Electron.* **2016**, *4*, 1270–1278. [\[CrossRef\]](#)
24. Aarstad, C.; Taufik, T.; Kean., A. Development of arc fault interrupter laboratory testing for low voltage DC electricity. In Proceedings of the 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA), Lombok, Indonesia, 28–30 July 2016.
25. Chen, S.; Li, X.; Xiong, J. Series arc fault identification for photovoltaic system based on time-domain and time-frequency-domain analysis. *IEEE J. Photovolt.* **2017**, *7*, 1105–1114. [\[CrossRef\]](#)
26. Chen, S.; Li, X. PV series arc fault recognition under different working conditions with joint detection method. In Proceedings of the 2016 IEEE 62nd Holm Conference on Electrical Contacts (Holm), Clearwater Beach, FL, USA, 9–12 October 2016.
27. He, C.; Mu, L.; Wang, Y. The detection of parallel arc fault in photovoltaic systems based on a mixed criterion. *IEEE J. Photovolt.* **2017**, *7*, 1717–1724. [\[CrossRef\]](#)
28. Zhu, H.; Wang, Z.; Balog, R.S. Real time arc fault detection in PV systems using wavelet decomposition. In Proceedings of the 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC), Portland, OR, USA, 5–10 June 2016.
29. Yusof MF, M.; Kamaruzaman, M.A.; Ishak, M.; Ghazali, M.F. Porosity detection by analyzing arc sound signal acquired during the welding process of gas pipeline steel. *Int. J. Adv. Manuf. Technol.* **2017**, *89*, 3661–3670. [\[CrossRef\]](#)
30. Thomas, S.R.; Kurupath, V.; Nair, U. A passive islanding detection method based on K-means clustering and EMD of reactive power signal. *Sustain. Energy Grids Netw.* **2020**, *23*. [\[CrossRef\]](#)
31. Dibaj, A.; Hassannejad, R.; Ettefagh, M.M.; Ehghaghi, M.B. Incipient fault diagnosis of bearings based on parameter-optimized VMD and envelope spectrum weighted kurtosis index with a new sensitivity assessment threshold. *ISA Trans.* **2021**, *114*, 413–433. [\[CrossRef\]](#)
32. Li, H.; Liu, T.; Wu, X.; Chen, Q. An optimized VMD method and its applications in bearing fault diagnosis. *Measurement* **2020**, *166*. [\[CrossRef\]](#)
33. Sukriti; Chakraborty, M.; Mitra, D. Epilepsy seizure detection using kurtosis based VMD’s parameters selection and bandwidth features. *Biomed. Signal Process. Control* **2021**, *64*.
34. Zhu, S.; Xia, H.; Peng, B.; Zio, E.; Wang, Z.; Jiang, Y. Feature extraction for early fault detection in rotating machinery of nuclear power plants based on adaptive VMD and Teager energy operator. *Ann. Nucl. Energy* **2021**, *160*. [\[CrossRef\]](#)
35. Wang, Y.; Kiziltas, A.; Blanchard, P.; Walsh, T.R. Calculation of 1D and 2D densities in VMD: A flexible and easy-to-use code. *Comput. Phys. Commun.* **2021**, *266*. [\[CrossRef\]](#)
36. Shi, G.; Qin, C.; Tao, J.; Liu, C. A VMD-EWT-LSTM-based multi-step prediction approach for shield tunneling machine cutterhead torque. *Knowl. Based Syst.* **2021**. [\[CrossRef\]](#)
37. Pincus, S. Approximate entropy (ApEn) as a complexity measure. *Chaos* **1995**, *5*, 110–117. [\[CrossRef\]](#)
38. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ.* **2000**, *278*, 2039–2049. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Xie, H.B.; He, W.X.; Liu, H. Measuring time series regularity using nonlinear similarity-based sample entropy. *Mod. Phys. Lett.* **2008**, *372*, 7140–7146. [\[CrossRef\]](#)
40. Chen, W.; Zhuang, J.; Yu, W.; Wang, Z. Measuring complexity using FuzzyEn, ApEn, and SampEn. *Med. Eng. Phys.* **2008**, *31*, 61–68. [\[CrossRef\]](#)
41. Chang, C.C.; Lin, C.J. Libsvm. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [\[CrossRef\]](#)
42. Luna, B.V.; Monroy-Morales, J.L.; Madrigal Martínez, M.; Torres-Lucio, D.; Weber, S.; Schweitzer, P. Analysis of Internal Signal Perturbations in DC/DC and DC/AC Converters under Arc Fault. *Energies* **2021**, *14*, 3005. [\[CrossRef\]](#)
43. Glowacz, A. Ventilation Diagnosis of Angle Grinder Using Thermal Imaging. *Sensors* **2021**, *21*, 2853. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Kim, J.-C.; Kwak, S.-S. Frequency-Domain Characteristics of Series DC Arcs in Photovoltaic Systems with Voltage-Source Inverters. *Appl. Sci.* **2020**, *10*, 8042. [\[CrossRef\]](#)