International Journal of Smart Grid and Clean Energy

PV arc-fault feature extraction and detection based on bayesian support vector machines

Yuan Gao[abc], Jianfei Dong[abc], Yaojie Sun[d*], Yandan Lin[d], Rui Zhang[d]

[a] *Changzhou Institute of Technology Research for Solid State Lighting, Changzhou, 213161, China*
[b] *State Key Laboratory of Solid State Lighting, Changzhou, 213161, China*
[c] *Beijing Research Center, Delft University of Technology, Beijing, 100083, China*
[d] *Department of Light Sources and Illuminating Engineering, Fudan University, Shanghai, 200433, China*

Abstract

In a PV system, DC arc is regarded as a serious fault, which might cause circuit damage and trigger fires. The arc fault, however, is hard to detect due to the special fields of photovoltaic systems: constant direct current without zero-crossing point, sophisticated components leading to noise interruption, and usually occupying large area. Therefore, detectable characteristics are of great importance to diagnosis and alarm of fault arcs in PV systems. In this paper, we presented a classification method of separating arcing and non-arcing in the feature space. First, data sets of current signal were sampled by designing field experiments with “pull apart” method for arc ignition. Then seven features in both time and frequency domains were defined and two of them in each domain were selected to train BSVM. In order to simplify the computation, the trained BSVM network was replaced by a separating line, which was proved to have a better performance of classification. Testing results showed that this method could diagnose fault arcs with high accuracy. But whether this method is suitable for other PV systems needs to be verified in further work.

Keywords: Photovoltaic systems, arc fault detection, bayesian support vector machines, PV testing

1. Introduction

Electrical fires in photovoltaic systems have been reported worldwide over the last decade, causing significant damage to surrounding facilities and threatening the personal safety of workers or occupants. Investigation found that most of photovoltaic fires were triggered by DC arcs accidentally [1]. DC arc is an electrical breakdown of air that produces a plasma discharge in a direct current circuit, giving out great heat and light. Besides a possible electric shock, it will ignite flammable materials and thus start a fire. The PV fault arcs usually arise from accidental disconnection of the DC circuit due to line corrosion or abrasion, loose connectors, solder disjoint, etc. Such fire accidents caused by DC arcs in PV systems pose a rigorous challenge to firemen because PV panels will keep generating electricity, which cannot be de-energized as long as they are exposed to the sunlight. As a result, National Electrical Code® (NEC) 2011 added the requirement of DC arc-fault circuit protection in the section of 690.11 that PV system with maximum system voltage of 80 volts or above shall be protected by a DC Arc-Fault Circuit Interrupter (AFCI) [2]. Since DC arcs hold different characters from AC arcs, previous detection methods for AC arcs are no longer applicable. Also, PV systems consist of sophisticated components, which enhance the difficulty of arc detection. Therefore, it is a crucial technology to find detectable characteristics of DC arcs in PV systems for AFCIs or components requiring the function of DC-arc detection.

Generally, DC fault arcs in PV systems can be classified into series arcs and parallel arcs [3]. Series arc faults are in the majority of arc accidents currently. Therefore, NEC 2011 only requires PV DC series arc-fault protection, without mentioning parallel arc-fault protection. In order to develop required AFCIs,

* Manuscript received June 21, 2015; revised September 17, 2015.
Corresponding author. Tel.: +86-21-65642778; *E-mail address*: yjsun@fudan.edu.cn.
doi: 10.12720/sgce.4.4.283-290

relevant experiments have been conducted by research organizations and manufacturers. In field experiments, researchers found that obvious increase of AC noise could be detected when a DC arc was triggered. Sandia National Laboratories (SNL) studied the electromagnetic propagation features of arcing signals through PV arrays [4]. The investigation on RF effects in PV systems indicates that RF noise can be transmitted and inducted by PV cells, modules, DC cables, and inverters, affecting the frequency response above 100 kHz. High frequency will be attenuated by weak or unshielded connections. Moreover, high-frequency components of DC arc are not obvious for detection. Thus, it is suggested that the detection frequencies should be set below 100 kHz. Meanwhile, frequencies below 100-1000 Hz are sensitive to incident irradiance oscillations, 120 Hz inverter noise and 60 Hz mains noise. Therefore, favorable frequency bandwidths between 1-100 kHz are recommended for DC arc fault detection. Additionally, switching frequencies for most inverters are between 1-100 kHz, so selecting a single frequency within this range is not advised. Instead, multiple frequencies or broadband noise would be a better choice for DC arc detection in PV systems [4]. Different detection frequencies are also mentioned in other articles. Korean scholars narrow this range down between 50-100 kHz to avoid inverter operating frequency [5]. Engineers from TI draw this range between 40-100 kHz directly from the comparison of arcing signatures with non-arcing signatures [6]. As mentioned in [4], detectable frequency ranges above are just suggestions for manufacturers. Multiple bandwidths and complex algorithms are necessary for accurate and robust arc detection.

Most of research on PV DC arc characters [7]-[11] were conducted according to the arc initiation platform and method using fine steel wood and tube in the test regulation of UL (Underwriter Laboratories Inc.) 1699B [12]. It is notable that UL 1699B also regulated the reaction time of AFCIs under test within two seconds, otherwise the accumulated heat will cause fire. It is telling that the initial features of arcs are the critical factor to detect arcs and extinguish fires in time. It is claimed that there are several uncontrollable parameters affecting arc characteristics in the UL method [13]. The fusion of steel wood can be regarded as a random change of electrode movements, shape, and gap distance. Meanwhile, the tube affects the discharge gas and even blocks the arc path when it melts. Therefore, a more reliable arc ignition method of separating electrodes, which is called “pull apart” method, is proposed in [13]. The velocity and acceleration of the moving electrode can be controlled by the stepping motor. By this means, the arc characteristics, especially the frequency-domain related criteria, are independent without the influence of random parameters.

In this paper, field experiments were designed and conducted with “pull apart” method instead of UL method by using a customized platform in order to obtain reliable data. Four time-domain features and three frequency-domain features were defined through analyzing collected data sets. Then two features in each domain were selected to train BSVM for classification. In order to simplify the computation, we took a separating line in the feature instead of a trained BSVM network. Testing results revealed that the separating line held a better performance of classification.

2. Experiment & Data

In [13], experimental results proved that “pull apart” method held many advantages against UL method. Thus, subsequent experiments and data acquisition are based on this method. The field experiment was conducted in a small PV power station. The PV array consisted of four strings and each string contains 16 series multicrystalline silicon modules (STP140-24/AC, 140Wp). The maximum output current is 19.12 A, but it cannot reach this number in reality, usually around 10 A. It can be calculated that the installed capacity of this power station is 8.96 kW. The PV system also contained a combiner box, a 10 kw inverter (Eversol-TLC), and an AC switch.

In order to create a DC arc in this PV system, we disconnected the line between combiner box and inverter, where arcs occurred frequently, then inserted an arc generator, as shown in Fig. 1. The switch S1 and S2 were used to protect against electrical shock during the installation. The data acquisition system consisted of a bench-top oscilloscope (YOKOGAWA DLM2024) with maximum sampling rate of 2.5 GS/s, and a pull-through current sensor (HIOKI CT6863 AC/DC). In this experiment, the sampling rate was set as 1.25 M/s. We collected 8 data sets in a sunny day. The DC current of 8 sets varies with the solar irradiance. Each data set contains sampled data in 5 seconds, thus the number of total sampling points is 6250000. The collected data sets will be used for feature extraction in the next section.

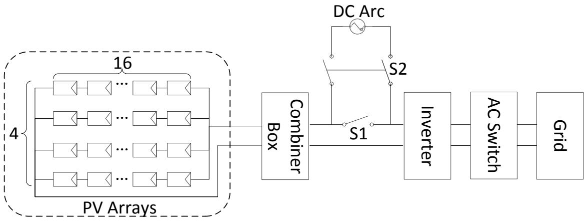


Fig. 1. Schematic diagram of PV system and arc generator.

3. Feature Extraction

DC arc characteristics and thresholds are already discussed in [13]. Here, we mainly focus on the definition of each feature. As shown in Fig. 2, the obvious difference between non-arcing (blue) and arcing (red) current lies in both time and frequency domain. In time domain, the waveform fluctuates strongly when the arc is ignited and the average current amplitude falls down a little during arcing. The Fast Fourier Transform (FFT) data as shown in Fig. 2 (b) comes from the color-marked windows in (a). The size of windows is enlarged 10 times visually because it is too narrow to see. It is obvious that the arcing current contains more noise than the non-arcing current in the low-frequency band. Therefore, we will define arc characteristics in both domains in this paper.

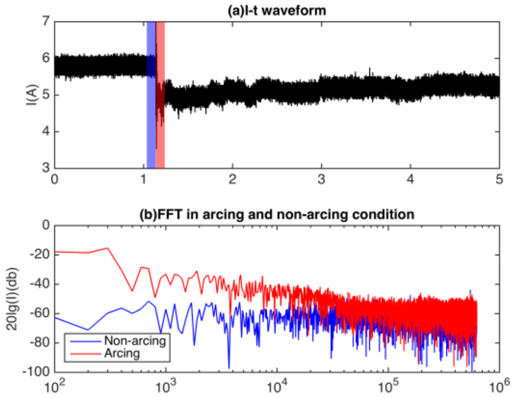


Fig. 2. Comparison of non-arcing and arcing: (a) time domain (b) frequency domain.

First, for all characteristics, we define a 10 ms (12500 sampling points) moving window along the I-t waveform, with the same size of the windows in Fig. 2. Then four Time-Domain Features (TDF) and three Frequency-Domain Features (FDF) are defined as listed below.

TDF1: the mean value of the current within 10 ms;

TDF2: the difference between the minimum and maximum current values in 10 ms;

TDF3: the difference between arcing and non-arcing current values, where the non-arcing current value is defined as the current value 2 s ago because the arc shall be interrupted within 2 s as regulated in UL 1699B;

TDF4: the mean absolute value of current gradient;

FDF1: the current component in certain frequency band;

FDF2: the ratio of the current component in certain frequency band to the DC current component;

FDF3: the ratio of the current component in certain frequency band to the AC current component.

Within the scope, “certain frequency band” has been proved to be 0.1-4 kHz [13]. Fig. 3 (a) shows the same I-t waveform as shown in Fig. 2 (a). When a moving window with a 1 ms step is running along the I-t waveform, defined features present various changes as shown in Fig. 3 (b)-(h).

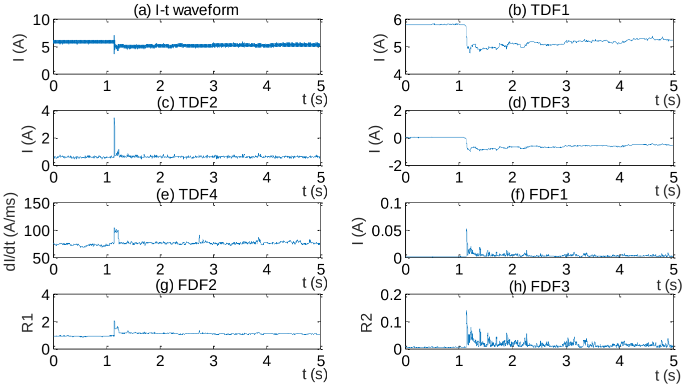


Fig. 3. (a) typical sampled DC current, (b)-(e) potential features in time domain, (f)-(h) potential features in frequency domain.

In [13], the threshold of each feature was obtained by analyzing 10 data sets, except TDF1, which varies under different conditions of solar irradiance. However, none of them is a unique arc characteristic. For example, the distinct jump of DC current also exists when some PV modules are short circuited. The external interference also leads to the increase of noise in certain frequency band. Thus, it is reasonable to choose features in both domains for the classification to separate arcing from non-arcing conditions.

4. Classification

4.1. Introduction of BSVM

Support Vector Machines (SVM) for classification, as described by Vapnik [14], exploit the idea of mapping input data into a high dimensional (often infinite) Hilbert space. The SVM methods have many advantages, including a global minimum solution as the minimization of a convex programming problem and sparseness in solution representation. Studies on classification can be found in [15]-[17]. In this work, the algorithm proposed in [17] is applied, which satisfies the following characteristics:

- Naturally normalized in likelihood evaluation;
- Possessing a flat zero region that results in sparseness property;
- Smooth and continuous up to the first order derivative.

Classification problems aim to infer a latent function that maps received patterns/features to prescribed classes. In the Bayesian approach, the latent function $f_{\boldsymbol{x}}$ with input $\boldsymbol{x}$ can be inferred by maximizing the posterior probability, which is defined by Bayes’ theorem as:

$$P(\boldsymbol{f}\mid\boldsymbol{D})=\frac{P(\boldsymbol{D}\mid\boldsymbol{f})P(\boldsymbol{f})}{P(\boldsymbol{D})} \tag{1}$$

where $\boldsymbol{D}$ is the training data set.

The latent functions $\boldsymbol{f}$ can be modelled as a multivariate Gaussian random variable with zero mean and a $n\times n$ covariance matrix Σ [3], given by (2) and (3).

$$P(\boldsymbol{f})=\frac{1}{Z_{\boldsymbol{f}}}\exp\left(-\frac{1}{2}\boldsymbol{f}^{T}\sum{}^{-1}\boldsymbol{f}\right) \tag{2}$$

$$\sum_{ij} = Cov\left[f_{x_i}, f_{x_j} \right] = k_0 \exp\left[-\frac{1}{2} \sum_{i=1}^{d} k_l \left(x_i^l - x_j^l \right)^2 \right] + k_b \tag{3}$$

where $Z_f = (2\pi)^{\frac{n}{2}} \sqrt{|\Sigma|}$; $k_0 > 0$ denotes the average power of f_x ; $k_l > 0$, l = 1, 2, …, d is the parameter that determines the relevance of the $l-th$ input dimension to the prediction of the output variables; d is the dimension of input vectors; $k_b > 0$ denotes the variance of the offset to the function f_x ; and $\boldsymbol{x}^l$ denotes the $l-th$ element of the input vector $\boldsymbol{x}$.

The trigonometric loss function, proposed in [17], is used in evaluating the likelihood. It takes the following form:

$$l_t\left(y_x \cdot f_x\right) = \begin{cases} +\infty & if\, y_x \cdot f_x \in (-\infty, -1] \\ 2 \ln \sec\left(\frac{\pi}{4}(1 - y_x \cdot f_x) \right) & if\, y_x \cdot f_x \in (-1, +1) \\ 0 & if\, y_x \cdot f_x \in [+1, +\infty) \end{cases} \tag{4}$$

The likelihood function can therefore be evaluated by:

$$P\left(y_x \mid f_x\right) = \exp\left(-l_t\left(y_x \cdot f_x\right)\right) = \begin{cases} 0 & if\, y_x \cdot f_x \in (-\infty, -1] \\ \cos^2\left(\frac{\pi}{4}(1 - y_x \cdot f_x) \right) & if\, y_x \cdot f_x \in (-1, +1) \\ 1 & if\, y_x \cdot f_x \in [+1, +\infty) \end{cases} \tag{5}$$

The distribution of $P\left(f\left(\boldsymbol{x}\right) \mid \boldsymbol{D}, \theta \right)$ can be evaluated as the Gaussian distribution,

$$P\left(f\left(\boldsymbol{x}\right) \mid \boldsymbol{D}, \theta \right) \sim N\left(\mu_t, \sigma_t^2\right) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{\left(f(x) - \mu_t \right)^2}{2\sigma_t^2} \right) \tag{6}$$

where the mean is $\mu_t = \boldsymbol{\upsilon}_M^T \boldsymbol{k}_M$, the variance is $\sigma_t^2 = Cov\left[f\left(\boldsymbol{x}\right), f\left(\boldsymbol{x}\right) \right] - \boldsymbol{k}_M^T \left(\Lambda_M^{-1} + \Sigma_M \right)^{-1} \boldsymbol{k}_M$, and $\boldsymbol{k}_M$ is the sub-vector of $\boldsymbol{k}$ by keeping the entries associated with SVs.

Given the hyper parameters θ , the probability of the binary class label $y_x = \pm 1$ is evaluated by the trigonometric likelihood function in (5) and $P\left(f\left(\boldsymbol{x}\right) \mid \boldsymbol{D}, \theta \right)$ in (6):

$$\begin{aligned} P\left(y_x \mid \boldsymbol{D}, \theta\right) &= \int_{-\infty}^{+\infty} P\left(y_x \mid f\left(\boldsymbol{x}\right), \boldsymbol{D}, \theta\right) P\left(f\left(\boldsymbol{x}\right) \mid \boldsymbol{D}, \theta \right) df\left(\boldsymbol{x}\right) \\ &= \int_{-1}^{+1} \cos^2\left(\frac{\pi}{4}\left(1 - y_x \cdot f\left(\boldsymbol{x}\right)\right) \right) N\left(\mu_t, \sigma_t^2\right) df\left(\boldsymbol{x}\right) + \frac{1}{2} erfc\left(\frac{1 - y_x \mu_t}{\sqrt{2}\sigma_t} \right) \end{aligned} \tag{7}$$

where $erfc(\upsilon) = \frac{2}{\sqrt{\pi}} \int_{\upsilon}^{+\infty} \exp\left(-z^2\right) dz$. Setting $y_x = +1$, we can predict the probability of the testing case to have the label of +1.

4.2. Data preprocessing

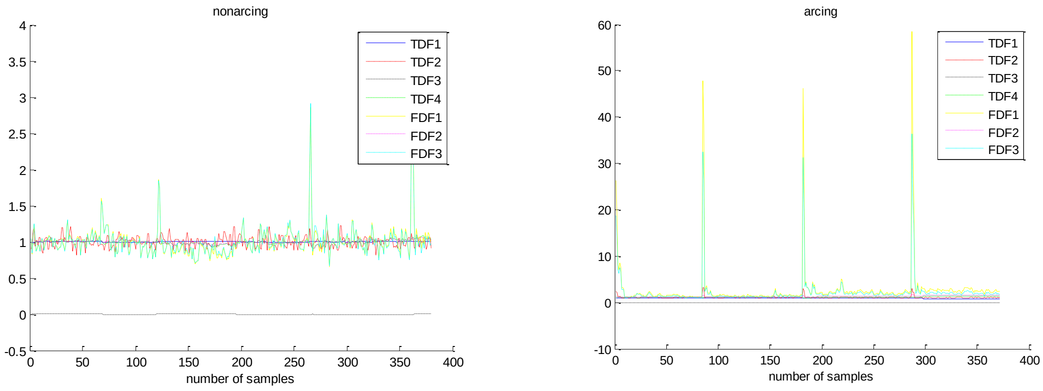
All features except TDF3 are first normalized as follows.

- Compute the mean value of each feature from the first 20 numbers, when there is no arcing;
- Compute the ratio of each feature to its corresponding mean value.

Since TDF3 is a relative changing signal. The mean value is trivial when arcing does not occur. Therefore, it does not need to be normalized.

We divide the 8 collected data sets into two groups. One is to train the BSVM, and the other is to test it.

From Fig. 4 (a) and (b), TDF3 and FDF2 are selected to train the BSVM. The other features are either too noisy at the non-arcing stage, or the changes in them are not as evident as those in TDF3 and FDF2.



(a) (b)

Fig. 4. (a) seven features in the training data set when arcing does not occur (b) seven features in the training data set when arcing occurs.

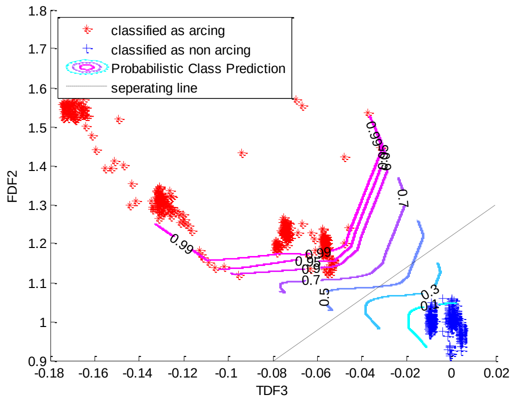


Fig. 5. Probabilistic class prediction of the training feature set.

4.3. *Training*

The two selected features are used in training BSVM. The probability of classifying the features at arcing stage is shown in Fig. 5. The hyperparameters, k_b and k_0, are optimized, from the initial values of respectively 1 and 1. After the training, $k_b = 1$ and $k_0 = 11.01$. The number of support vectors is 174 selected from totally 752 training vectors, i.e. 23.1%. But still, the size of the hidden layer in the BSVM network is large, resulting in a considerable computation burden.

In order to simplify the computation, we take a separating line in the feature space as shown in Fig. 5. According to the probability contours, the 50% probability is roughly in the middle of the two clusters of feature vectors. A line is drawn to roughly cover this contour. The separating line equation is as follows,

$$y = 4x + 1.22 \qquad (8)$$

where x and y respectively stand for TDF3 and FDF2.

According to Fig. 5, the non-arcing cluster is below the line; while the arcing cluster is above it. Therefore, the classification according to this line can be done as follows.

$$\begin{cases} \textit{classified as non} - \textit{arcing}, & \textit{if } y < 4x + 1.22 \\ \textit{classified as rcing}, & \textit{if } y \geq 4x + 1.22 \end{cases} \tag{9}$$

4.4. Testing

In order to test the classification performance, the following three values are computed.

- Correct classification rate, is the ratio between the number of correctly classified samples and the total number of the samples in the feature set.
- False alarm rate, denoted as $f_a = \mathrm{P}\left(arcing ? detected \mid non\, arcing\right)$, is the ratio between the number of the detected arcing samples and the number of the samples when arcing does not occur in the feature set.
- Missed detection rate, denoted as $f_m = \mathrm{P}\left(no ? arcing\, detected \mid arcing\right)$, is the ratio between the number of the missed arcing detections and the number of the samples when arcing occurs in the feature set.

The results are listed in Table 1. The classification results with the separating line is slightly better, because the BSVM is highly nonlinear with as many as 174 support vectors, and can hence be over fitted to the training data. However, the separating line is found based on the probabilistic property of BSVM. Its performance rooted from BSVM.

Table 1. Testing results

Test data set	BSVM			Line		
	Correct classification rate	False alarm rate	Missed detection rate	Correct classification rate	False alarm rate	Missed detection rate
1	99.90%	0.18%	0	100%	0	0
2	99.90%	0.18%	0	100%	0	0
3	99.90%	0.14%	0	99.90%	0	0.3%
4	96.92%	0.21%	5.60%	100%	0	0

It is noticeable that the separating line is gained by sampling data from a certain PV power station at a certain testing point. Hence, the line may vary slightly when applied to other testing locations and power stations. It needs to be demonstrated by further experiments and data analysis.

5. Conclusion & Discussion

In PV systems, DC fault arcs with high voltage have caused severe hazards, which lead to great casualties and enormous property loss. In order to prevent such catastrophic events, NEC requires the AFCI installation in PV systems and UL releases a standard to test a PV DC AFCI. However, in China, and other developing or undeveloped countries, such regulations and standards are still unavailable. Though there are some AFCIs and PV products integrated with AFCI functions in the market currently, it is still necessary to advertise the hidden risk of DC fault arcs in PV systems and promote relevant regulations and standards. Also, research on arc features, detection and classification methods, prevention mechanisms, and system robustness shall be conducted based on large data sets and demonstrated by repeatable field experiments.

Previous research has proved that the “pull apart” method revealed a stable performance of arc generation. Thus, our field experiments are designed based on this arc-ignition method and platform. The main contributions of this paper are: (1) sampling current signal data when DC arc occurs through field experiments in a certain PV power station, (2) analyzing collected data and extract 7 features in time and frequency domains, (3) training BSVM with two selected features and achieving the classification of arcing and non-arcing by a separating line in the feature space, and (4) testing the performance of classification by defining and calculating the rate of correct classification, false alarm and missed detection. Results show that the separating line can diagnose the arc fault in this testing condition with high accuracy.

Testing results in this study, however, don't necessarily prove this classification method feasible for all PV systems. Thus, further study is needed in evaluating other influences of DC arc fault detection, such as various arcing locations, PV systems with different components and structures, noise interference, line attenuation, etc. Data collection work shall be carried by international cooperation and data sharing platform.

Acknowledgements

This work was supported in part by a grant from National Natural Science Foundation of China (Project No. 61374129); Natural Science Foundation of Jiangsu (Project No. BK20130258); Science and Technology Commission of Shanghai Municipality (Project No. 12195811000); Project of Shanghai Municipal Science and Technology Achievements Transformation and Industrialization (Project No. 14441900400); Project of "Research on Detection and Evaluation of High-Capacity Movable PV Grid for High-Altitude" (Project No. NY71-13-036) cooperated with China Electric Power Research Institute.

References

[1] Bubble, Bubble, Toil, and Trouble. PHOTON International. 2006:130.

[2] National Electric Code. NFPA 70. National Fire Protection Association. Quincy, MA, 2011.

[3] Johnson J, Montoya M, McCalmont S, *et al*. Differentiating series and parallel photovoltaic arc-faults. In: *Proc. 38th Photovoltaic Specialists Conference*, 2012:000720-000726.

[4] Johnson J, Kuszmaul S, Bower W, Schoenwald D. Using PV module and line frequency response data to create robust arc fault detectors. In: *Proc. 26th European Photovoltaic Solar Energy Conference and Exhibition Hamburg*, Germany, 2011:3745-50.

[5] Seo GS, Cho BH, Lee KC. Photovoltaic module -level DC-DC converter with arc fault protection scheme for DC distribution system. In: *Proc*. *ECCE Asia Downunder*, 2013:917-923.

[6] Novak B. Implementing arc detection in solar applications: achieving compliance with the new UL 1699B Standard. Texas Instruments, 2012.

[7] Dini DA, Brazis PW, Yen KH. Development of arc-fault circuit-interrupter requirements for photovoltaic systems. In: *Proc. 37th Photovoltaic Specialists Conference*, 2011:1790-1794.

[8] McCalmont S. Low cost arc fault detection and protection for PV systems. National Renewable Energy Laboratory Report, 2013.

[9] Vaaßen W, Zornikau J. Failure mechanismen of contact faults in the DC-circuit of the PV arrays. In: *Proc. International Workshop, Arcing in Photovoltaic DC-Arrays*, Burgdorf, 2007.

[10] England B. An investigation into arc detection and fire safety aspects of photovoltaic installations: school of engineering and energy. Murdoch University, 2012.

[11] Johnson J, Blemel KD, Peter F. Preliminary photovoltaic arc-fault prognostic tests using sacrificial fiber optic cabling. Sandia Technical Report, SAND2013-1185, 2013.

[12] *Photovoltaic (PV) DC Arc-fault Circuit Protection*, UL 1699B (2011).

[13] Gao Y, Zhang J, Lin Y, Sun Y. An innovative photovoltaic DC arc fault detection method through multiple criteria algorithm based on a new arc initiation method. In: *Proc. 40th Photovoltaic Specialist Conference*, 2014:3188-3192.

[14] Vapnik V. *The Nature of Statistical Learning Theory*. Springer Science & Business Media; 2013.

[15] Seeger M. Bayesian model selection for support vector machines, Gaussian processes and other kernel classifiers. In: *Proc. 13th Annual Conference on Neural Information Processing Systems*, 2000:603-609.

[16] Kwok JTY. The evidence framework applied to support vector machines. *IEEE Transactions on Neural Networks*, 2000; 11:1162-1173.

[17] Chong WCSSK, Ong J. Bayesian inference in trigonometric support vector classifier. *Neural Computation*, 2003; 15(6):2227-254.