IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 2, 2019, accepted May 24, 2019, date of publication May 27, 2019, date of current version June 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919337

Newly-Designed Fault Diagnostic Method for Solar Photovoltaic Generation System Based on IV-Curve Measurement

JUN-MING HUANG, RONG-JONG WAI, (Senior Member, IEEE), AND WEI GAO

College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China
Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan

Corresponding author: Rong-Jong Wai (rjwai@mail.ntust.edu.tw)

This work was financially supported in part by the Ministry of Science and Technology of Taiwan, under Grant MOST 106-2221-E-011-028-MY2.

ABSTRACT Fault diagnosis of photovoltaic (PV) arrays is an essential task for improving the reliability and safety of a photovoltaic system (PVS). The PVS faults at the DC side are difficult to detect by traditional protective devices, which may reduce power conversion efficiency and even lead to safety matters and fire disaster. This study investigates a newly-designed fault diagnostic method for a PVS according to the following three steps. First, optimal fault features are extracted by analyzing I-V curves from different faults, including hybrid faults of the PVS under the standard test condition (STC). Moreover, the trust-region-reflective (TRR) deterministic algorithm combined with the particle-swarm-optimization (PSO) metaheuristic algorithm is proposed to standardize fault features into the ones under the STC. In addition, a multi-class adaptive boosting (AdaBoost) algorithm, which is the stage-wise additive modeling using multi-class exponential (SAMME) loss function based on the classification and regression tree (CART) as the weak classifier, is utilized to establish the fault diagnostic model. The effectiveness of the fault diagnostic model could long-term maintain by periodically updating the feature standardization equations to standardize the fault features into the ones under the STC. Various types of the PV modules are used to validate the generalization of the fault diagnostic method. Both the numerical simulations and experimental results show the accuracy and reliability of the proposed fault diagnostic method.

INDEX TERMS Photovoltaic system (PVS), fault diagnosis, I-V curve, trust-region-reflective (TRR), particle swarm optimization (PSO), SAMME-CART.

I. INTRODUCTION

With the exponential growth in global photovoltaic (PV) power capacity, the operating maintenance and protection have received great importance in the last few decades. PV arrays are vulnerable to suffer various unexpected faults due to the uncertain outdoor operating environment, which result in massive power loss, irreversible module damage, or even fire disaster [1]–[2]. Faults in the DC side of a PV system (PVS) are difficult to detect and distinguish by conventional protection devices merely based on NEC, IEC, and UL Standards [3], mainly due to the lower fault current magnitudes, the presence of maximum power point tracking (MPPT), non-linear PV characteristics and the dependency on the irradiance levels [4]. Therefore, numerous advanced PV fault detection approaches have been investigated in recent years [5]. These methods based on different PV fault features, which can be roughly divided into four categories to be based on materials physical properties, power loss, measured voltage and current, and current-voltage (I-V) curve.

Physical properties of faulty PV modules are studied and analyzed by physic means such as radiation detection [6]–[7], electric current injection [8] or incident signal transmission [9]–[10]. Infrared thermal imaging [6]–[7] is an effective method to detect hot spot in PV modules, while electro-luminescence technology [8] can be used to accurately detect micro-cracks, breaks and finger interruptions. However, the two techniques rely on costly sophisticated sensors and require particular test conditions. Recently, the spread spectrum time-domain reflectometry (SSTDR) [9] is utilized to determine the PVS impedance variation rather than depending on the fault current magnitudes. This method

The associate editor coordinating the review of this manuscript and approving it for publication was B. Chitti Babu.

can effectively detect the PVS ground fault compared with time-domain reflectometry [10]. However, the SSTDR needs a specific external signal function generator and only could be performed offline. Although the above diagnostic techniques could accurately identify even locate PV faults, the cost of maintenance is expensive and difficult to apply on arbitrary operational PVS.

As for the diagnostic method based on the analysis of power loss, the theoretical output power of PVS is estimated compared with the measured one. Chouder and Silvestre [11] proposed an automatic supervision and fault detection system by defining new power loss indicators and establishing theoretical boundaries. Hariharan *et al.* [12] developed a method to detect the mismatch and shading by calculating the power loss and the sudden change of power and irradiance. Fractional-order color relation classifier is investigated to quantify the output power degradation and separated normal conditions from fault events [13]. Dhimish *et al.* [14] adopted a 3rd-order polynomial function to describe the behavior of the faulty region, and then combined with fuzzy logic classification systems to enhance PV fault detection accuracy. Harrou *et al.* [15] proposed a framework with nonparametric thresholds by combining the benefit of k-nearest neighbors with univariate monitoring approaches. However, the method in [15] cannot discriminate partial shading from faults occurring at the dc side of a PV array. These power-loss-based methods depend on the accuracy and generalization of the estimated output power model. The corresponding estimated results may deviate from the theoretical values while changing to the other module specifications, especially under low irradiance condition. As a result, the accuracy of the fault identification will degenerate.

By employing sensors to measure voltages and currents of a PVS for further characteristic analysis, various diagnostic algorithms were presented to identify PV faults. Pillai and Rajasekar [16] used peculiar operating characteristics by perturbation and observation (P& O) MPPT, and proposed threshold-based detection rules to detect line-line and line-ground faults with any level of mismatches. Unfortunately, the performance of faults detection in [16] is sensitive to manual setting thresholds. Hu *et al.* [17] analyzed the optimal locations of voltage sensors and omitted current sensors, and developed a two-section PV fault detection method. But this method is unsuitable to implemented in the large scale PVS for the cost of an increased voltage sensors and the computational complexity. Aiming to detect line-to-line faults under low irradiation with an active MPPT algorithm, Yi and Etemadi extracted fault features based on multi-resolution signal decomposition, and proposed a fuzzy inference system in [18] and a two-stage support vector machine in [19]. A semi-supervised learning algorithm based on the graph theory is used to identify PV faults, which can still perform good results by only using a few labeled training data [20]–[21]. Graph-based semi-supervised learning (GBSSL) algorithms have been adopted to identify PV faults, which can perform good results by only using a few labeled training data [20]–[21]. However, the stability of GBSSL methods is susceptible to noise, and the speed of sample testing slows down with the accumulation of historical data. Moreover, these methods [18]–[21] based on the total voltage and current of PV arrays are limited to identify specific PV faults, such as partial shading and abnormal aging to be more common failures in a PVS.

I-V curve measurement of a PVS inherently contains wealth information. The diagnostic features are extracted from I-V curves for fault analyses and identifications. Hachana *et al.* [22] extracted parameters via the artificial bee colony optimization and the differential evolution, and distinguished fault types according to threshold values in the form of a look-up table. However, manual selection of the threshold value easily leads to the limitation of higher accuracy and generalization. Principal component analysis based statistical method is used to identify shading faults [23], but it can only diagnose single fault. Recently, artificial intelligence (AI) technologies have received great attention [24]–[27]. Fuzzy classifiers, artificial neural network, kernel-based extreme learning machine algorithm (KELM) and multiclass adaptive neuro-fuzzy classifier (MC-NFC) are used to establish PV fault diagnostic models respectively. Although AI technologies can improve the fault classification rate to some extent, few researchers attempt to verify the generalization ability of AI fault diagnostic algorithms or AI training models. The reasons are that the sensitivity of PV modules to irradiances and temperatures makes the output voltages and currents to be greatly varied, and the installation environments of a PVS are diverse. Moreover, the values of extracted indicators without effective normalization may greatly differ due to different datasheet STC parameters of PV modules.

To address the issues discussed previously, this study proposes a newly-designed method based on I-V curve measurement to diagnose PVS faults including short circuit, abnormal aging, two types of partial shading and hybrid faults. According to the varying degrees of the four faults and hybrid faults, the fault types are classified into 25 kinds in numerical simulations, and sorted into 11 categories in the experimentations. Four specifications of PV module in numerical simulations and two different modules on the real PVS are applied to verify the generalization of the proposed fault diagnostic method. This study is organized as five sections. Following the Introduction, Section II introduces the research procedure and analyzes optimal fault features. In Section III, the proposed PSO-TRR-based feature normalization process and the SAMME-CART algorism for the PV fault diagnostic model are expressed in detail. In Section IV, the performance of the fault diagnostic method is verified by numerical simulations and experimental results. Conclusions are given in Section V.

II. RESEARCH PROCEDURE AND FAULT ANALYSIS OF PVS

A. INTRODUCTION OF RESEARCH PROCEDURE

Since various faults exist in a photovoltaic system (PVS), it is difficult to distinguish simply by the voltage and current

measurement at the corresponding maximum-power-point-tracking (MPPT) point. Thanks to the integration of online I-V tracers into new-type smart photovoltaic (PV) inverters [23], the measurement of I-V curves is available to reflect the state of a PVS truly without additional hardware. In this study, a solar system analyzer (PROVA1011) manufactured by TES Electrical Electronic Corp. is used to extract I-V curves for practical verification as in [26]. Based on I-V curve measurement, the flowchart of the proposed research procedures is depicted in Fig. 1. The output voltage and current vary widely as PV modules are sensitive to the external environment, which makes it hard to identify possible faults. However, in a specific environment such as the standard test condition (STC), faults have significant impact on I-V curves and even can identify hybrid faults. Therefore, electrical characteristics and I-V curves in different faults are analyzed under STC for the optimal features extraction. Moreover, selected features under different temperatures and irradiances convert to the ones under STC could obtain higher recognition and generalization.

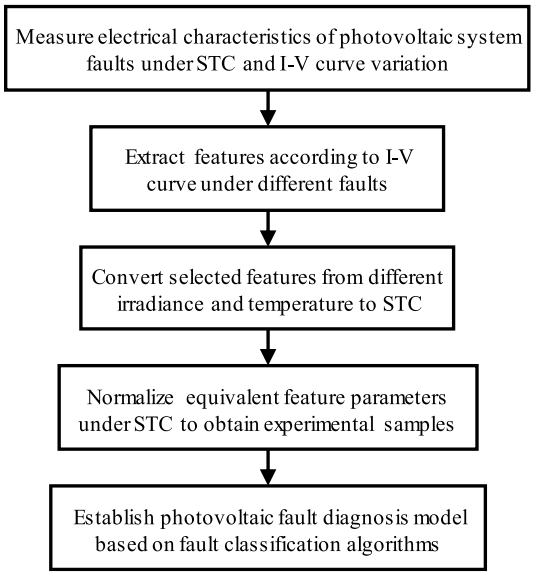

FIGURE 1. Flowchart of research procedure.

By simplifying the equivalent PV model, the STC parameters of I-V curves can be transformed into the ones under given irradiance and temperature via approximate formula [28]–[31]. Then shifting the formula, parameters under different temperatures and irradiances can be converted into the ones under STC. This idea is also used in the per-unit method [26] to eliminate the influence of irradiations and temperatures. However, this method suffers several problems in real operating PVS. For example, it will result in a low accuracy, especially under low irradiances. Moreover, various effects among different PV modules. In addition, the inherent deviation in the same module due to the difference in the installation of PV and the placement of the sensors.

Therefore, fault detection method needs to be separately tuned within a particular PVS. Based on low-cost and easily obtained normal samples, the PV normalization equations of fault features are built by a nonlinear least-square method based on the hybrid PSO-TRR algorithm. Finally, a variant of the basic AdaBoost, called stage-wise additive modeling via a multi-class exponential (SAMME) loss function based on the classification and regression tree (CART) as the weak classifier, is utilized to establish the PV fault diagnosis model.

B. FAULT ANALYSIS AND FEATURE EXTRACTION

I-V testing circuit and PV module modeling via MATLAB/Simulink simulation are depicted in Fig. 2. In this study, the configuration of a PVS is formed by a string with 13 series PV modules. Each module consists of 60 cells connected in series, which evenly gathered into three sub-strings with three bypass diodes. By controlling the output value of the voltage source to linearly increase, the output current and voltage of the PV string are recorded. Then input the corresponding data into the MATLAB workspace to obtain final I-V curves. A rectifier diode is employed in the output of the PV string to avoid the occurrence of negative currents. Irradiances and temperatures are set by the gain amplifier of each sub-string.

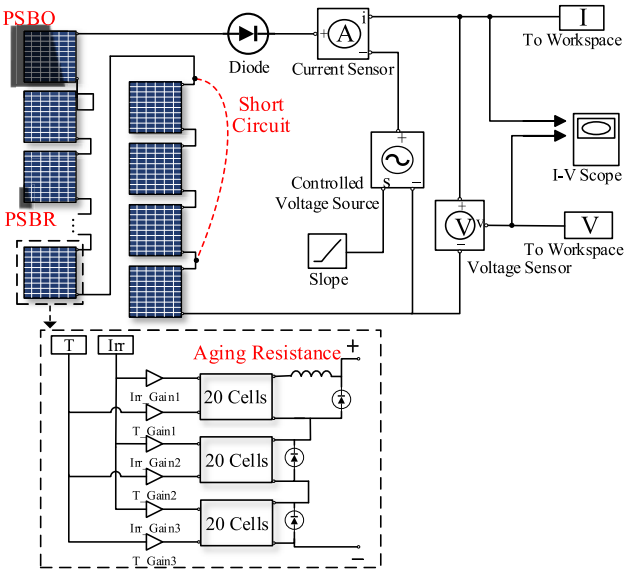

FIGURE 2. I-V testing circuit and PV module modeling via MATLAB/Simulink.

Faults of short-circuit, abnormal-aging, partial-shading and the corresponding hybrid faults of PV strings are considered. Typical I-V curves at single fault condition under STC is depicted in Fig. 3. Note that, in partial shading conditions, different from quantitative analyses according to the voltage at the peak power on the right-hand side in [16], this study analyses qualitatively from another perspective based on the activation of bypass diodes inside shaded PV modules at the global MPPT (GMPPT) point. Shading faults are divided into two types, which contain the partial shading with the bypass-diode reversed (PSBR) and the partial shading with

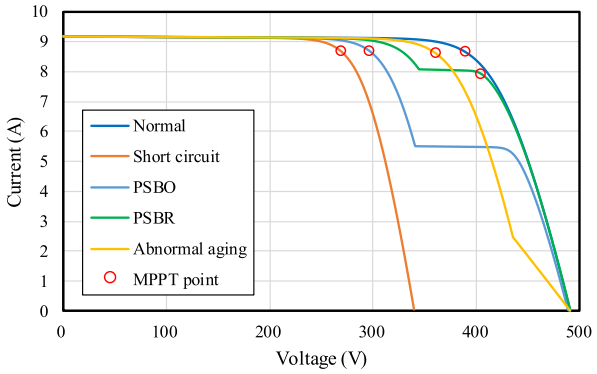

FIGURE 3. Typical I-V curves at single fault condition under STC.

the bypass-diode on (PSBO). Owing to the relatively small shading area in the PSBR condition, the GMPPT point at the right-hand side of the peak power still works in all modules together with a low current drop, which leads to being overheated and results in hot spots. In the PSBO condition, the considerable decrease of the current in the PSBO module leads to the activation of the bypass diode, which makes the voltage to be negative in shaded modules, and the GMPPT point is moved to the left-hand side of the peak power. As for the abnormal aging fault, it refers to a sudden increase in the series resistance and causes the bypass on under STC. The equivalent series resistance R_s of a PV module [30]–[31] can be represented as

$$R_s = -\frac{dV}{dI}\bigg|_{V\cong V_{oc}} = \frac{V_{oc}-V_1}{I_1} \tag{1}$$

$$R_s = \frac{1}{3}\left(\frac{V_{oc}-V_1}{I_1}+\frac{V_{oc}-V_2}{I_2}+\frac{V_{oc}-V_3}{I_3}\right) \tag{2}$$

where (I_1,V_1), (I_2,V_2) and (I_3,V_3) are three closest I-V points to (0, V_{oc}). In order to suppress external interference and measurement noise in experimentations, one can modify (1) to (2) by averaging three estimated values of R_s.

Since hybrid faults are superposition of single faults, electrical parameters in individual fault I-V curves are summarized in Table 1. Due to the value of I_{sc} just has a little change in the aforementioned faults, only the open circuit voltage V_{oc}, the equivalent series resistance R_s, and V_m and I_m at the MPPT point are considered as the diagnostic features of a PVS. These features with different variations perform the characteristics of different faults under the STC. Therefore, the fault type of the PV array can be accurately discriminated when selected features of the PVS can be converted into the ones under the STC.

III. NEWLY-DESIGNED FAULT DIAGNOSTIC TECHNIQUES FOR PVS

The proposed fault diagnostic method for a photovoltaic system (PVS) in this study includes the data pre-processing and pattern-recognition theory. The output parameters of the I-V curve are normalized by a nonlinear least-squares method based on the hybrid particle-swarm-optimization

TABLE 1. Variations of fault features at STC.

Fault Type	V_{oc}	I_{sc}	V_m	I_m	R_s
Short Circuit	↓	—	↓	—	—
PSBR	—	—	↑	↓	—
PSBO	—	—	↓	—	—
Abnormal Aging	—	—	↓	—	↑

trust-region-reflective (PSO-TRR) algorithm. The stage-wise additive modeling with a multi-class exponential loss function (SAMME) based on the classification and regression tree (CART) is trained by using normalized features and labels. The detection procedure is elaborated in the following subsections.

A. PROPOSED FEATURE NORMALIZATION PROCESS

In this study, the characteristic functions are extracted partly according to the traditional approximation formula of output parameters [28]–[31]. Coefficients in the characteristic function can be represented by unknown coefficients (a, b, c, and d). The characteristic equations for I-V curves can be rewritten as

$$V_{oc} = V_{oc\cdot stc} + a_1 \cdot \ln\frac{G}{G_{stc}} + a_2 \cdot dT + a_3 \cdot \frac{G}{G_{stc}}dT \tag{3}$$

$$V_m = V_{m\cdot stc} + b_1 \cdot \ln\frac{G}{G_{stc}} + b_2 \cdot dT + b_3 \cdot \frac{G}{G_{stc}}dT \tag{4}$$

$$I_m = c_1 \cdot I_{m\cdot stc}\frac{G}{G_{stc}} + c_2 \cdot dT + c_3 \cdot \frac{G}{G_{stc}}dT \tag{5}$$

$$R_s = R_{s\cdot stc}\left(\frac{G}{G_{stc}}\right)^{d_1} + d_2 \cdot dT + d_3 \cdot \frac{G}{G_{stc}}dT \tag{6}$$

where the third item in the right-hand side of (3)-(6) is the error reduction factor, and the corresponding fitting result can be further optimized. G is the measured irradiance, G_{stc} is a constant of 1000W/m ^{2}; dT is the measured temperature minus the STC temperature.$V_{oc.stc}$, $V_{m.stc}$, $I_{m.stc}$, and $R_{s.stc}$ represent the open-circuit voltage, the voltage and current at the MPPT point, and the equivalent series resistance of the PVS at the STC, respectively.

Due to the nonlinearity of the characteristic formula, the nonlinear least-squares method based on the PSO-TRR algorithm is proposed to calculate unknown coefficients. The nonlinear least-squares fitting is a parameter searching method for estimating the parameters of nonlinear static models by minimizing the following objective error function.

$$\min_x f(x, gt_i) = \min_x \sum_{i=0}^{n} [(F(x, gt_i) - y(gt_i))^2] \tag{7}$$

where gt_i represents the vector consist of measured temperatures and irradiances, $F(x, gt_i)$ is the PV characteristic equations from (3) to (6), x is the unknown coefficients of $F(x, gt_i)$, $y(gt_i)$ is the measured value from a PVS, and n is the number of measured samples.

Numerical optimization methods have been widely used to solve non-linear minimization problems as well as overcome

the difficulty of parameter determinations. In this study, the TRR algorism plus the PSO algorism is proposed to search the optimal solution. In the hybrid PSO-TRR algorithm, the PSO is used to globally explore the initial value of x_0, since the local optimization is accomplished using the TRR. The hybrid algorithm can enhance the convergence without sacrificing the accuracy and stability, and are briefly introduced later.

The trust-region approach incorporated with the interior-reflective Newton algorithm and the subspace algorism [32]–[35] is simple and efficient to solve the bound constrained minimization problems. It can be used to solve nonlinear least-squares problems. The error vector function $f(x)$ in (7) is the minimization object of the TRR algorithm, in which x is a vector to be optimized with upper and lower bounds ($l < x < u$). Coleman and Li [32] used a new affine scaling transformation to construct a standard scaled unconstrained trust region sub-problem. As for calculating the sub-problem, the TRR algorithm restricts the sub-problem to a two-dimensional subspace S [34]–[35] to accelerate the convergence for large-scale problems.

The TRR algorithm is implemented and coded in the MATLAB software in this study. The termination tolerance includes the tolerance on the function value, the tolerance on variables, maximum function evaluation numbers and maximum iteration numbers. Note that, unreasonable values can be eliminated by setting the upper and lower bounds of variables in the initialization of the TRR algorithm. For example, a_1 is used as V_t in the conventional formula, and the lower constraint of a_1 is set to $a_1 > 0$. The constrained bounds for unknown coefficients x_i are summarized in Table 2.

TABLE 2. Constrained bounds for unknown coefficients.

Coefficient subscript	a_i	b_i	c_i	d_i
i=1	(0,30)	(0,30)	(0,30)	(-30,0)
i=2	(-30,0)	(-30,0)	(-30,30)	(-30,30)
i=3	(-30,30)	(-30,30)	(-30,30)	(-30,30)

The powerful local optimization algorithm of the TRR can quickly and accurately calculate the optimal solution while given a reasonable initial value of x_0. However, when the initial value is irrational, the TRR algorithm is easily trapped into the local optima. In order to enhance the stability of the algorithm, the PSO is utilized to search the moderate initial values of the characteristic coefficients. Although the PSO does not always guarantee to discover the globally optimal solution during a finite time, it often quickly provides a sub-optimal near globally optimal. The PSO used in this study is briefly introduced as follows.

The PSO is one of the swarm intelligence algorithms introduced by Kennedy and Eberhart [36]–[37], which inspired by swarm behavior in birds flocking and fish schooling to solve optimization problems. This study uses the *global* variant of PSO, which learns from the personal best position and the best position attained so far by the whole swarm. The inertia weight used to balance between the global and local search abilities [38], which is linearly decreased with the iterative generations. In this study, the linearly weight is set to [0.8, 1] to enhance the global search ability. Since the fitting problem is to minimize the root-mean-square error (RMSE) objective function as

$$RMSE = \left(\sum_{i=1}^{n} (F(x, gt_i) - y(gt_i))^2 / n\right)^{\frac{1}{2}} \tag{8}$$

The fitness function of the PSO is defined as

$$Fitness = \frac{1}{1 + RMSE} \tag{9}$$

Since the PSO can converge quickly to the optimal positions at the beginning of the run, it may slowly converge with a low speed near a local solution [39]. In this study, the number particles and evolutions are only set to be 30 for obtaining a good initial value of x_0. The constrained position is the same as the TRR algorithm in Table 2.

When characteristic coefficients are calculated by the nonlinear PSO-TRR least-squares method, the complete characteristic equations for PV parameters can be obtained. Then, parameter normalization equations (10)-(13) can be obtained by shifting and normalizing the formula, which convert the output parameters of I-V curves under different temperatures and irradiances into the per-unit ones at STC.

$$V_{oc.norm} = \frac{1}{V_{oc.stc}}\left(V_{oc} - a_1 \cdot \ln\frac{G}{G_{stc}} - a_2 \cdot dT - a_3 \cdot \frac{G}{G_{stc}} dT\right) \tag{10}$$

$$V_{m.norm} = \frac{1}{V_{m.stc}}\left(V_m - b_1 \cdot \ln\frac{G}{G_{stc}} - b_2 \cdot dT - b_3 \cdot \frac{G}{G_{stc}} dT\right) \tag{11}$$

$$I_{m.norm} = \frac{1}{I_{m.stc}}\left(\frac{G_{stc}}{c_1 \cdot G}\left(I_{m.} - c_2 \cdot dT - c_3 \cdot \frac{G}{G_{stc}} dT\right)\right) \tag{12}$$

$$R_{s.norm} = \frac{1}{R_{s.stc}}\left(\left(\frac{G}{G_{stc}}\right)^{-d_1}\left(R_s - d_2 \cdot dT - d_3 \cdot \frac{G}{G_{stc}} dT\right)\right) \tag{13}$$

where V_{oc}, V_m, I_m, and R_s respectively represent the open-circuit voltage, the voltage and current at the MPPT point, and the equivalent series resistance. $V_{oc.norm}$, $V_{m.norm}$, $I_{m.norm}$, and $R_{s.norm}$ are the corresponding normalized parameters.

The premise of the parameter normalization is to measure I-V curves of the normal operating PVS under different irradiances and temperatures. Since the irradiance variable has a great influence on PV power generation compared to the temperature variable, the irradiance variable is used as the reference of measurement. Normal I-V curves are measured in a clear weather with different irradiances for the parameter normalization. Since the natural aging of PV arrays is over time, periodically re-extracting new I-V curves and reconstructing normalization equations can improve the long-term effectiveness of PV fault diagnostic techniques.

B. PVS FAULT CLASSIFIED METHOD BASED ON SAMME-CART

The adaptive boosting (AdaBoost) algorithm [40]–[41] is a typical case of the boosting algorithm, which can improve the performance of any given learning algorithm [42]. The AdaBoost algorithm combines weak classifiers generated by iterations to form a stronger final classifier. Compared with most learning algorithms, ensemble learning is less prone to over-fitting [43]. This study uses the multi-class Adaboost algorithm proposed by Zhu *et al*. [44], referred as the stage-wise additive modeling using a multi-class exponential loss function (SAMME), which extends the Adaboost from two-class to multi-class without reducing it to multiple two-class problems. In the k-class problem, the SAMME algorithm can be theoretically proven that it only requires the performance of each weak classifier to be better than the one with random guessing. In this study, the decision tree (DT) grown by the classification and regression (CART) algorithm is utilized as the weak classifier. 10-folder cross-validation is utilized to optimize the parameters of the SAMME algorithm.

1) CART ALGORITHM

The CART algorithm is characterized by constructing binary DTs in a top-down manner [45]. The DT begins with a root node derived from whichever variable in the feature space and minimizes a measure of the impurity of the two sibling nodes by a defined splitting rule. The Gini splitting rule to be most broadly used is employed as the splitting algorithm in this study. To avoid the over-fitting problem, limiting the splits number smaller than the classification category is utilized as the pre-prune method of the CART. The DTs based on the CART algorithm is used as the weak classifiers of the SAMME owing to its simple binary structure and easily handle outliers in a separate node.

2) SAMME-CART ALGORITHM

Let $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ be entire training set, where n is the total number of the training set, x_i is the training vectors containing four PV diagnostic features, $Y_i \in Y = \{1, 2, \ldots, k\}$ is the k-class label with respect to x_i. The initial sample weight $D_{1.i}$ and the number of iteration T are set. In the *t-th* iteration, the DT based on the CART algorism is used to classify the weighted training samples, and the weak hypothesis h_t:$X \rightarrow Y$ can be obtained. Based on the weak hypothesis results, the current DT is implemented with a calculated certain weight α_t, and the sample weight $D_{t.i}$ is further updated by the exponential loss function. After the T round, all the weak hypotheses h_1,…,h_T are weighted to get the final strong classifier H. Note that, the SAMME algorism is the same as the original AdaBoost as $k = 2$. The executed steps are expressed as follows.

1) Initialize sample weights as

$$D_{1\cdot i} = \frac{1}{n}, i = 1, 2, \ldots, n \tag{14}$$

2) For $t = 1$ to T :

   a. Train DT on the current D_i and compute the classification error of DT by

$$\varepsilon_t = \sum_{i=1}^{n} \left(D_{ti} \cdot \mathrm{I}\left(h_t\left(x_i\right) \neq y_t\right)\right) / \sum_{i=1}^{n} D_{ti} \tag{15}$$

where I(∗) equals to unity when the condition (∗) is satisfied. Otherwise, it is zero.

   b. Compute the classification weight of the DT as

$$\alpha_t = \log \frac{1 - \varepsilon_t}{\varepsilon_t} + \log(k - 1) \tag{16}$$

where k denotes the number of class.

   c. The sample weight is updated according to the following equation:

$$D_{(t+1)i} = \frac{D_{ti} \cdot \exp(\alpha_t \cdot \mathrm{I}(h_t(x_i) \neq y_t))}{Z}, \quad i = 1, \ldots, n \tag{17}$$

where Z_t is the normalization factor.

1) Assigning the weights to the classification of DTs provides the final strong classifier as follows:

$$H(x) = \arg\max_k \sum_{t=1}^{T} \alpha_t \cdot \mathrm{I}\left(h_t(x) = k\right) \tag{18}$$

3) SAMME PARAMETER OPTIMIZATION

The training error of the SAMME will be decreased with the increase of DTs. In order to get the optimal model, the number of weak classifiers DTs (i.e., the number of iterations) needs to determine. Since only one hyper-parameter should be tuned, this study uses 10-folds cross-validation for parameter optimization, and the number of weak classifiers increases from 1 to 100 with the increment step of 1. Based on the minimum average of ten verification errors, the number of weak classifiers can be obtained as the optimal number of the SAMME iterations. By using 10-folds cross-validation to fully take advantage of the training data, the PV fault diagnostic model can further avoid the over-fitting problem and has higher generalization ability.

IV. VERIFICATION OF PVS DIAGNOSTIC TECHNIQUES

In this study, the performance of the proposed PV fault diagnostic method is verified by numerical simulations and experimental results. Five specifications of PV module referred to PVM1-PVM5 are given in Table 3, and used to verify the generalization ability of the proposed method.

A. NUMERICAL SIMULATIONS OF PVS DIAGNOSES

Four common photovoltaic (PV) faults including the short-circuit fault (SCF), the partial shading with the bypass-diode on (PSBO), the partial shading with the bypass-diode reversed (PSBR), and the abnormal aging fault (AAF) are considered here. The type of partial shading depends on the number and shading level of PV modules. In this study, only the partial shading within three modules is considered. The value of the gain amplifier is randomly set during the range of [0.85, 0.92] to simulate the PSBR

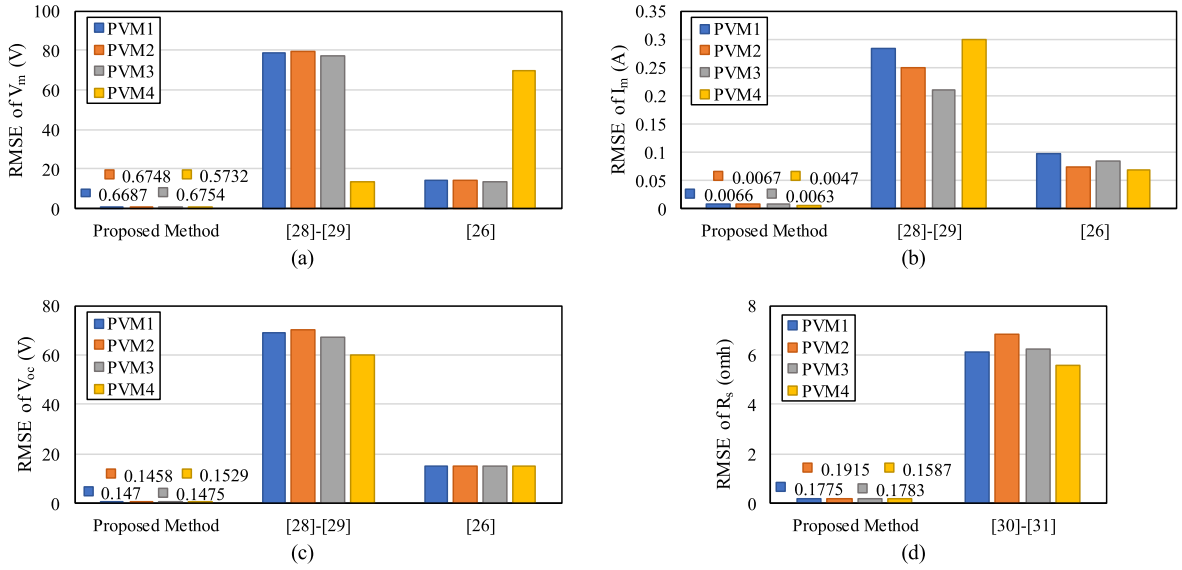

FIGURE 4. RMSE records of four module parameters for PVM1-PVM4 in comparison with other approximate methods in [28]–[31] and per unit method in [26]: (a) Parameter V_m; (b) Parameter I_m;(c) Parameter V_{oc};(d) Parameter R_s.

TABLE 3. Variations of fault features at STC.

PV Module	PVM1	PVM2	PVM3	PVM4	PVM5
P_{max}/Wp	260	235	250	290	300
V_m/V	30.02	29.34	29.67	32.3	31.51
I_m/A	8.66	8.01	8.51	8.99	9.52
V_{oc}/V	37.78	37.06	37.31	39.7	39.24
I_{sc}/A	9.12	8.45	9.02	9.57	9.93
k_p/%/k	-0.4631	-0.4631	-0.42	-0.42	-0.4003
k_v/%/k	-0.3315	-0.3315	-0.32	-0.3	-0.2906
k_i/%/k	0.0443	0.0443	0.03	0.05	0.053
N_s	60	60	60	60	60

condition, while set randomly during the range of [0.3, 0.75] to simulate the PSBO condition. As for simulating the AAF condition, the series aging resistor is randomly set during the range of [2Ω, 10Ω]. Because the proposed diagnostic basis in this study is I-V curves, the presence of MPPT will not affect the performance of the proposed method. In order to cover a wide range of operating environmental conditions, the irradiance range is from 100 W/m ^{2} to 1200 W/m ^{2}, while the temperature range varies from 35° C to 65° C. The varied steps are determined by a certain value (A) and a randomly varying value (random(B)) in (19) to truly reflect the real environment. In order to verify the accuracy of the proposed fault diagnostic method for a PVS, 25 types of PV fault situations including the normal condition are considered. The total amount of 10,560 simulated data can be obtained according to the datasheet parameter of a module at the STC. Different types including PVM1-PVM4 modules are used for simulated verification.

$$Step = A + random(B) \tag{19}$$

1) PERFORMANCE OF PARAMETER NORMALIZATION

The characteristic parameters (V_{oc}, V_m, I_m, and R_s) are extracted from I-V curves of the normal operating PV string under irradiances varied from 100W/m^2 to 1200W/m^2. One-quarter of normal samples are used to calculate the coefficients of the PV characteristic formula in Section III. Then, the complete PV output characteristic formula can be obtained via the nonlinear particle-swarm-optimization trust-region-reflective (PSO-TRR) least-squares method. The fitting characteristic equations of four parameters are evaluated by the rest three-quarter normal samples and the corresponding RMSE values. The fitting results are compared with the ones of traditional approximation formulas in [28]–[31] and per unit method in [26] as shown in Fig. 4. As can be seen from Fig. 4, the proposed method yields superior approximation accuracy than other methods in [26], [28]–[31]. All RMSE values of four parameters via the proposed method are smaller than 0.68, which indicate the accuracy of the fitting output characteristic equations. Note that, the errors of the approximated formulas in [26], [28]–[31] are different among different module parameters, which will affect the generalization ability of the diagnostic algorithm.

In order to compare the statistical distribution of the original simulated data and the normalized ones, the boxplot is formed by setting the single fault to show the performance of the parameter normalization visually. By taking PVM1 as an example, the corresponding boxplot is depicted in Fig. 5. The original features overlap and highly skewed among different faults, while the normalized ones in Fig. 5(b) show better data clustering and unification.

2) PERFORMANCE OF PV DIAGNOSTIC TECHNIQUES

According to the normalization equations, the samples composed of four PV diagnostic features can be obtained. All samples are randomly divided into a training set and a testing set by a ratio of 3:1. By taking PVM1 as an example, the PV fault diagnostic results are given in Table 4, where the symbol

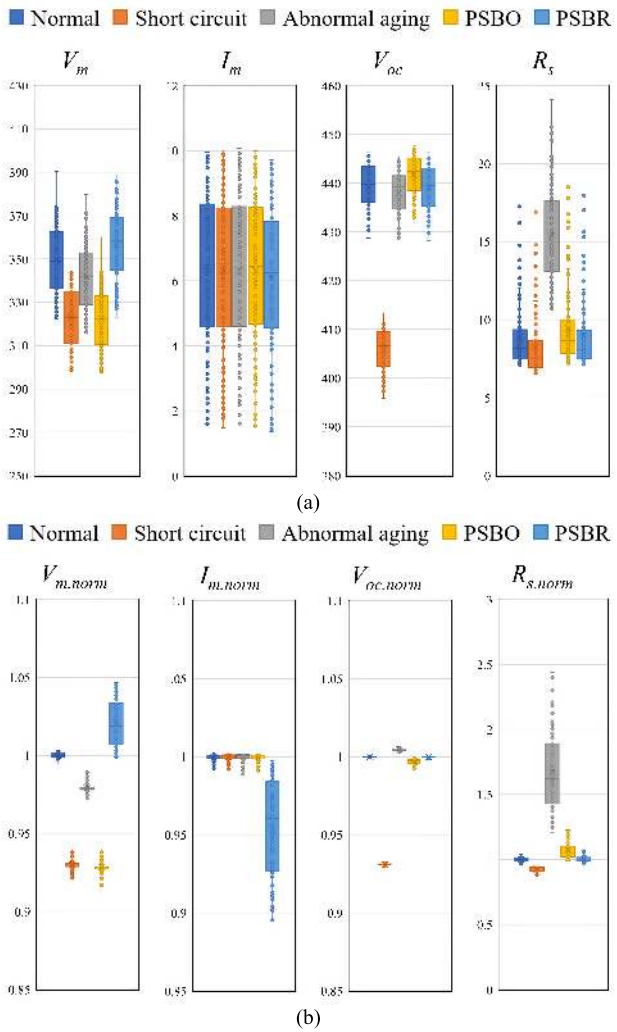

FIGURE 5. Boxplot of four simulated feature variables: (a) Original data samples distribution of PVM1; (b) Normalized data samples distribution of PVM1.

TABLE 4. Detailed classification accuracy of PVM1 simulated data samples.

Types of PVS faults	Training Accuracy (%)	Testing Accuracy (%)
Normal	100.00	100.00
SCF 1	100.00	100.00
SCF 2	100.00	100.00
SCF 3	100.00	100.00
SCF 4	100.00	100.00
PSBO 1	100.00	100.00
PSBO 2	100.00	98.33
PSBO 3	100.00	98.33
PSBO 1-sub	100.00	100.00
PSBO 2-sub	100.00	98.33
PSBR 1-3	100.00	100.00
PSBR 1-2-sub	100.00	98.33
AAF 1-sub	100.00	100.00
AAF 2-sub	100.00	99.17
PSBO 1 with PSBR 1-2	100.00	100.00
PSBO 2 with PSBR 1-2	100.00	99.17
SCF 1 with PSBO 1-2	100.00	100.00
SCF 2 with PSBO 1-2	100.00	100.00
SCF 1 with PSBR 1-2	100.00	99.17
SCF 2 with PSBR 1-2	100.00	100.00
SCF 1 with AAF 1-2-sub	100.00	100.00
SCF 2 with AAF 1-2-sub	100.00	100.00
PSBO 1 with AAF 1-2-sub	100.00	100.00
PSBO 2 with AAF 1-2-sub	100.00	100.00
PSBR 1-2 with AAF 1-2-sub	100.00	99.58
Average Accuracy	**100.00**	**99.70**

TABLE 5. Classified accuracy of simulation data samples from different modules.

PV Module	Training Accuracy (%)	Testing Accuracy (%)
PVM1	100.00	99.70
PVM2	100.00	99.73
PVM3	100.00	99.77
PVM4	100.00	99.81

"*-sub*" represents the extent to which the faults occurred in sub-string. As can be seen from Table 4, the training accuracy of each type of fault is 100%. However, the accuracy of some fault testing results is smaller than 100%, which are caused by the occurrence of some faults in the sub-string. As for sub-string faults mixed with different types of fault, the response with hybrid faults are similar to the one with single fault, which is difficult to correctly classify.

As for the generalization ability of the proposed diagnostic algorithm, this study uses four datasheet parameters at the STC for verification. The classified accuracy of simulation data samples from different modules are summarized in Table 5. As shown in Table 5, all the total fault classified accuracy at four cases is above 99.70%. These simulated results conclude that the high precision and generalization ability by the proposed PV fault diagnostic method can be obtained.

In order to verify the superiority of the SAMME-CART classification algorithm, this study uses three other popular machine-learning algorithms including the Naive Bayesian algorithm, the extreme learning machine (ELM) algorithm, and the random forest algorithm to compare the corresponding performances. Comparative results are depicted in Fig. 6. The results show that the SEMME-CART algorithm has the highest accuracy, and the corresponding performance is slightly better than the one of the random forest algorithm. Note that, the classification accuracy of each algorithm under different module parameters is higher than 95%. These results show that the per-unit normalization of equivalent parameters at the STC has extremely high recognition.

This study also verifies the generalization ability of the proposed PV fault diagnostic model. The trained PV fault diagnostic model is used to test the samples under different module parameters in Table 6. As shown in Table 6, the powerful generalization of the proposed PV fault diagnostic model can be verified. The result is attributed to the high fitness of the diagnostic equations of the PV output

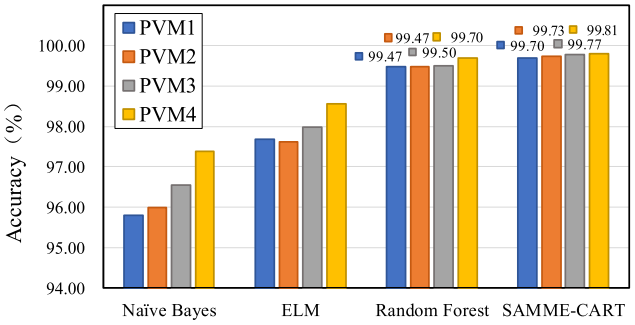

FIGURE 6. Classified result of SAMME-CART in comparison with other machine learning methods.

TABLE 6. Classification accuracy of testing data from different modules.

Training Model	Testing data	Testing Accuracy (%)
PVM1	PVM2	99.58
	PVM3	99.39
	PVM4	99.66
PVM2	PVM1	99.62
	PVM3	99.66
	PVM4	99.70
PVM3	PVM1	99.43
	PVM2	99.55
	PVM4	99.66
PVM4	PVM1	99.17
	PVM2	98.98
	PVM3	99.28

parameters. The difference of characteristic parameters in PV modules is negligible after the data pre-processing. As for the absence of data, it is only necessary to obtain data on the normal operation at different irradiances, and use the trained PV fault diagnostic model from other cases to discriminate PV faults. The cost for acquiring a large amount of fault data can be saved.

In practice, especially for utility-scale systems, PV modules are connected in a parallel-series configuration to expand the output power and voltage as desired. Therefore, the extension from a single PV string to multiple PV strings is also examined to verify the scalability of the proposed method in numerical simulations. The data are captured by expanding the capacity to a 6.1kWp PVS with two PV strings and a 9.2kWp PVS with three PV strings via the PVM1 module modeling in Fig. 2. The number of samples in each system is 2,640 to be randomly divided into a training set and a testing set by a ratio of 3:1. Only faults within one PV module occurred in a string of the PV array with multiple strings are considered under the same condition setup in Table 4. The classification accuracies with respect to two PV strings and three PV strings are summarized in Table 7. With the increase of PV capacity, the classification accuracy of some types of faults reduces, especially in the hybrid faults. Fortunately, the average accuracy also can be over 99.23%, and the scalability of the proposed method can be verified.

TABLE 7. Classification accuracy of PVM1 simulated data samples with multiple PV strings.

Types of PVS faults	Testing accuracy of two strings (%)	Testing accuracy of three strings (%)
Normal	100	98.33
SCF	100	100
PSBO	100	100
PSBR	100	99.65
AAF	100	100
SCF with PSBO	100	100
SCF with PABR	100	100
SCF with AAF	100	100
PSBO with PSBR	100	95.67
AAF with PSBO	98.33	97.88
AAF with PSBR	100	100
Average Accuracy	**99.85**	**99.23**

B. EXPERIMENTAL VERIFICATION OF PVS DIAGNOSES

In the real PV fault cases, two types of modules including the PVM1 manufactured by the polycrystalline silicon and the PVM5 manufactured by the monocrystalline silicon as shown in Fig. 7 are used to form the PVS with 13 modules in series for experimental verifications. The specifications PVM1 and PVM5 are given in Table 3. The experimental sites for the two PVS (3.51kWp and 3.9kWp) are located at National Taiwan University of Science and Technology, Taiwan. The photograph of the experimental hardware platform is depicted in Fig. 8. The short-circuit fault is created by Y-branch connectors. Small pieces, such as small bricks or discarded cigarette boxes, are used to simulate the PSBR condition. Thin plastic sheeting or paper sheets are employed to simulate the PSBO condition. The abnormal aging fault utilizes a sliding varistor as the aging resistor to be connected in series with the PV sub-string. I-V curves of the PVS, real-time irradiances and temperatures of solar panels are collected via the solar system analyzer (PROVA1011). The solar system analyzer measures the temperature of solar cells of back panels instead of the ones of front panels in the experiment. Besides, the installation angle of the irradiance sensor is not likely to be the same as the PV module, which will result in the measurement error. Fortunately, temperatures and irradiances are only taken as reference quantities in this study. As long as the sensor position is fixed, the corresponding experimental results will not be affected. Irradiances in experimental environment are range from 100 W/m ^{2} to 1000 W/m ^{2}, and experimental I-V curves under the occurrence of single fault at the STC in high and low irradiances are depicted in Fig. 9(a) and 9(b), respectively. The characteristics in Fig. 9 are similar to the ones in Fig. 3.

1) PERFORMANCE OF PARAMETER NORMALIZATION

Due to complex environmental factors and deviations in the signal conversion of the acquisition device, the noise inside the measured data is inevitable. In the data pre-processing of fitting the characteristic coefficients, the result depends on given numbers of data. The insufficient amount of data

FIGURE 7. Two type of experimental modules for verification.

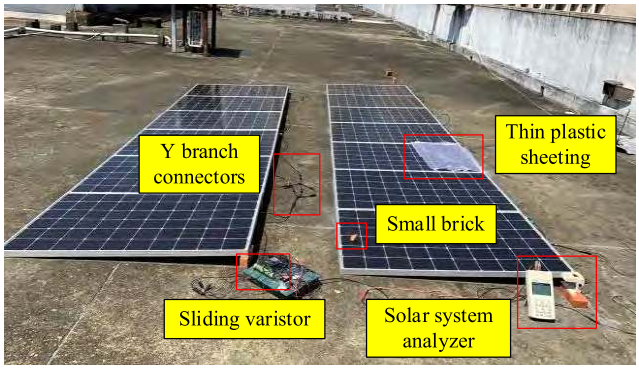

FIGURE 8. Experimental hardware platform.

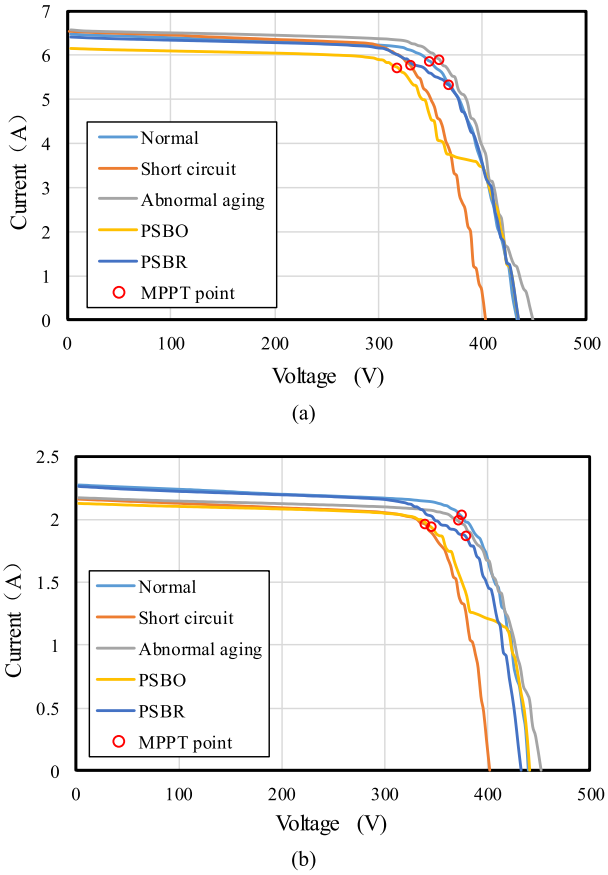

FIGURE 9. Experimental I-V curves under the occurrence of single fault at STC: (a) High irradiance condition; (b) Low irradiance condition.

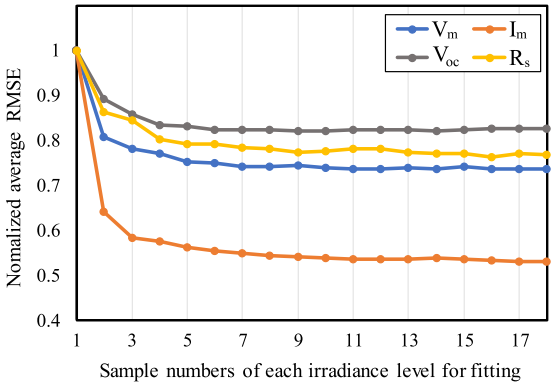

FIGURE 10. Optimal fitting number selection of PVM1.

may affect the fitting result. In order to find the minimum limit of the required normal data, the total measured normal data are divided into four levels of irradiances, including G <300W/m ^{2}, 300W/m ^{2} < G <500W/m ^{2}, 500W/m ^{2} < G <700W/m ^{2}, and G >700W/m ^{2}. Then, one randomly selects the same number (n) from each level to respectively perform 100 times fitting, and uses the remaining normal samples to evaluate the fitting results. Each parameter fitting is evaluated via the average of 100 RMSEs, and the largest average one is used as a reference for the fitting evaluation. By taking 389 experimental normal samples of PVM1 as an example, the result in Fig. 10 tends to be stable after $n = 7$. Therefore, the parameter normalization only needs 28 normal samples under different irradiances. The value of $n = 10$ is selected in the study to make the fitting results more reliable, i.e., a total amount of 40 normal samples are used for the parameter standardization. The parametric fitting result of PVM1 and PVM5 are evaluated by the average of 100 RMSEs, as shown in Table 8. Although the average RMSE value of V_m and V_{oc} are relatively large, the impacts on their actual values can be negligible.

The boxplot of four normalized feature variables is provided in Fig. 11 to describe the statistical distribution of experimental data samples of PVM1 compared with the simulated one in Fig. 5(b). It can be observed that some measured data deviate from the theoretical values due to various interferences in the real environment. However, the distribution of measured data is almost the same as that of the simulation data. This result verifies the effectiveness of the parameter normalization and provides a pssibility for the feasibility of replacing the measured data with simulated ones under the absence of data.

2) PERFORMANCE OF PV DIAGNOSTIC TECHNIQUES

In the experimental verification, 2,803 and 2,787 samples can be obtained by measuring I-V curves at each PVS per minute.

TABLE 8. Average RMSE of fitting results f or PVM1 and PVM5.

Fitting Evaluation	PVM1	PVM5
Average RMSE of V_m	3.6868	4.1358
Average RMSE of I_m	0.0448	0.0528
Average RMSE of V_{oc}	4.5283	1.1399
Average RMSE of R_s	0.6654	0.7281

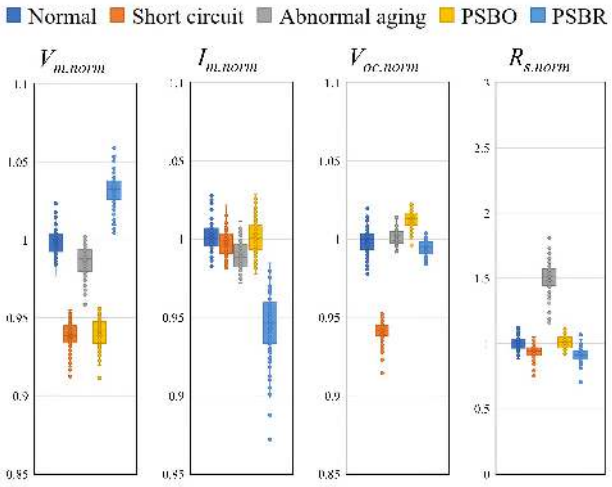

FIGURE 11. Boxplot of four normalized feature variables distribution from experimental data samples of PVM1.

In order to make full use of measured PV fault samples for fairly verifying the performance of the algorithm, this study conducts 100 times of random sample division for model training and testing. After determining the optimal number of iterations through 10-folds cross-validation, the average training and testing results are summarized in Table 9. Since the feature parameters are accurately converted to the ones at the STC, the training accuracy of the SAMME-CART algorithm is up to 100%. As for the testing results, the average testing accuracies for PVM1 and PVM5 are 98.80% and 97.87%, respectively; each type of fault classification accuracy is above 94.90 %, which demonstrates the overall performance of PV fault diagnoses in experimental verification.

In order to verify the generalization of the experimental PVS diagnostic model, the test data of these two PVS are exchanged. The fault diagnostic model of one PVS is tested with the fault test data of the other PVS. As shown in Table 10, the test results still have high recognition accuracy for testing different material solar modules. The results verify the generalization ability of the PVS diagnostic model. Note that, it is illustrated to a certain extent that fault samples of different PVS can be used interchangeably.

C. DISCUSSIONS

Compared with other studies based on I-V curves for fault diagnoses, the major significant difference in this study is that

TABLE 9. Detailed average classification accuracy of PVM1 and PVM5 experiment data samples.

Types of PVS fault	PVM1		PVM5	
	Training Accuracy (%)	Testing Accuracy (%)	Training Accuracy (%)	Testing Accuracy (%)
Normal	100.00	99.00	100.00	96.95
SCF	100.00	98.89	100.00	97.48
PSBO	100.00	99.34	100.00	96.91
PSBR	100.00	98.27	100.00	94.90
AAF	100.00	98.09	100.00	99.69
PSBO with PSBR	100.00	99.13	100.00	96.82
SCF with PSBO	100.00	99.03	100.00	99.31
SCF with PSBR	100.00	98.34	100.00	97.20
SCF with AAF	100.00	99.48	100.00	99.98
PSBO with AAF	100.00	98.19	100.00	99.13
PSBR with AAF	100.00	98.79	100.00	98.64
Average Accuracy	**100.00**	**98.80**	**100.00**	**97.87**

TABLE 10. Classification accuracy of experiment testing data from different modules.

Training Data	Testing Data	Accuracy (%)
PVM1	PVM5	95.16
PVM5	PVM1	91.37

characteristic parameters of I-V curves are converted to the ones at the STC via the PVS data at the normal state. The proposed method has the following advantages:

1) Higher fitting precision compared with traditional approximation equations.
2) Avoid the influence of different installation environment of PVS and the deviation of sensors placement.
3) Prevent misjudgment of the diagnostic model from the natural aging of a PVS via periodical parameters normalization.

More importantly, the high recognition feature can be obtained according to the data pre-processing, which greatly improve the accuracy and convergent speed of later classification algorithms. In addition, the parameter values of different PVS converted to the ones at the STC under fault conditions are nearly the same, which makes the classification algorithm and the training model to possess a higher generalization.

' Therefore, under the absence of fault cases data, fault diagnostic models can be established by using data from other practical sites or even simulation data. Compared with other machine-learning algorithms, the SAMME-CART ensemble learning combines multiple DTs to form the final strong classifier for reaching a higher accuracy. Furthermore, the over-fitting problem of the training model in traditional methods can be solved by pre-pruning the DTs and determining the optimal number of DTs through 10-fold cross-validation. In the real PVS with 13 modules connected in series, the small-scale shading of bird droppings or small

TABLE 11. Comparative Results Between Proposed Algorithm and Methods in [22], [26]–[27].

Case study		Proposed Algorithm	[27]	[26]	[22]
Year		2019	2018	2017	2016
PV system capacity		3.51kWp and 3.9kWp	0.3kWp	1.8kWp	0.66kWp
PV fault diagnosis technology	Fault feature	V_{oc}, V_{mp}, I_{mp} and R_s	12 features for optimal selection	I_{sc}, V_{oc}, I_{mp}, V_{mp}, n_n, R_s and RMSE	Several diagnostic parameters
	Feature normalization approach	Nonlinear Least Square Fitting	Theoretical Loss Rate	Approximate Per Unit Method	N/A
	Machine learning technique	SAMME-CART	MC-NFC	KELM	N/A
Under low irradiance conditions		✓	✓	✓	×
Insensitive to simulated model accuracy		✓	×	✓	✓
Consideration of hybrid faults		✓	×	×	✓
Generalization verification of diagnostic technology		✓	×	×	×
Adapting to PVS natural aging		✓	×	×	×

fragments can be even accurately identified by the proposed diagnostic method.

The comparative results between the proposed algorithm and the methods in [22], [26]–[27] are provided in Table 11. The intuitive look-up table of parameter thresholds is formed to detect PV faults in [22]. However, it is time consuming and difficult to adjust for different PVS. The traditional approximation formula is applied to standardize the diagnostic parameters in [26]. But this method cannot completely eliminate the influence of irradiations and temperatures on various modules. In the process of the feature normalization, the simulation model is used as the reference to calculate the theoretical ratio in [27]. However, the result of this method in [27] is limited by the accuracy of the simulation model. The performance of the methods in [22], [26]–[27] will be affected by the location and measured angle of sensors in a real PVS. Through reasonable parameter normalization, compared with the literatures [22], [26]–[27], the apparent advantages of the proposed method are the high generalization ability and the long-term effectiveness by using the low-cost normal data of a PVS. In addition, hybrid faults are studied and accurately classified thanks to the normalized features and the high performance of classification algorism.

Although the proposed method requires module-wise measurements which would be costly for large-scale systems, the same distribution of numerical simulations and experimental data provides a possibility for the feasibility of replacing experimental data by simulated ones for large-scale systems. Even though the proposed parameter normalization method could effectively standardize parameters from different irradiances and temperatures for a generalized diagnosis, its limitation is whether relevant sensors are equipped or not. Fortunately, operational & maintenance (O& M) companies usually install sunshine meters and thermometers in a PVS because irradiances and temperatures are important factors to judge whether the PV power generation is normal or not. Moreover, PV inverters without capturing the information of I-V curves will limit the feasibility of the proposed method. This issue can be easily solved by vacating a short time from the main program to execute the I-V curve capturing function in PV inverters. The proposed PV diagnostic technology combines with smart inverters integrated online I-V tracers would be the best scheme without additional hardware. If smart inverters can be popularized in the future, the application of the proposed method will become easier.

V. CONCLUSIONS

In this study, a new fault diagnostic technique for a photovoltaic system (PVS) based on measured I-V curves has been successfully designed. The proposed method consists of the extracted parameters normalization via the nonlinear particle-swarm-optimization trust-region-reflective (PSO-TRR) least-squares algorithm, and the pattern recognition via the classification algorithm via the stage-wise additive modeling using multi-class exponential (SAMME) loss function based on the classification and regression tree (CART). Four types of common faults including the short-circuit fault, the partial shading with the bypass-diode on (PSBO), the partial shading with the bypass-diode reversed (PSBR), and the abnormal aging as well as six types of their hybrid faults are considered for diagnosing. The effectiveness of the proposed method is verified by four specifications of PV simulation modules, and is also tested via two practical PVS with the power capacities of 3.51kWp and 3.9kWp. With the development of smart inverters, I-V curves of a PVS can be easily extracted by a smart inverter during the time gap of the voltage conversion and data storing in the cloud. The main contributions of this study are summarized as follows:

The electrical characteristics and I-V curves of PV faults under standard test conditions (STC) are analyzed, and the optimal characteristic parameters of PV fault diagnoses are extracted including the open-circuit voltage, the voltage and the current at the maximum power point tracking (MPPT), and the equivalent series resistance. The fault of partial shading through the analysis and verification is divided into two situations as PSBO and PSBR.

The proposed parameter normalization method based on the PSO-TRR nonlinear least-squares fitting converts characteristic parameters under different temperatures and irradiances into the ones at the STC, which eliminate the impact of varied environmental factors. Only low-cost data at the normal state are required to avoid the influence of PV installation environments as well as the deviation of the environmental measuring sensors. The parameter normalization equations can be updated periodically to adapt to the natural aging of a PVS.

The PV diagnostic model based on the SAMME-CART algorithm can achieve a higher accuracy in comparison with other machine-learning algorisms. Both numerical simulations and experimental results show superior classification and generalization performance than previous researches. The generalization ability is verified by various modules, and it concludes that the proposed fault diagnostic model still can maintain good accuracy by using the data from other PVS when the fault data is insufficient.

REFERENCES

[1] Y. Zhao, J.-F. de Palma, J. Mosesian, R. Lyons, and B. Lehman, "Line–line fault analysis and protection challenges in solar photovoltaic arrays," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3784–3795, Sep. 2013.

[2] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques," *IEEE J. Photovolt.*, vol. 5, no. 3, pp. 982–997, May 2015.

[3] D. S. Pillai and N. Rajasekar, "A compatibility analysis on NEC, IEC, and UL standards for protection against line–line and line–ground faults in PV arrays," *IEEE J. Photovolt.*, vol. 9, no. 3, pp. 864–871, May 2019.

[4] D. S. Pillai and N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in PV systems," *Renew. Sustain. Energy Rev.*, vol. 91, pp. 18–40, Aug. 2018.

[5] D. S. Pillai, F. Blaabjerg, and N. Rajasekar, "A comparative evaluation of advanced fault detection approaches for PV systems," *IEEE J. Photovolt.*, vol. 9, no. 2, pp. 513–527, Mar. 2019.

[6] M. Simon and E. L. Meyer, "Detection and analysis of hot-spot formation in solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 94, no. 2, pp. 106–113, Feb. 2010.

[7] Y. Hu, W. Cao, J. Ma, S. J. Finney, and D. Li, "Identifying PV module mismatch faults by a thermography-based temperature distribution analysis," *IEEE Trans. Device Mater. Rel.*, vol. 14, no. 4, pp. 951–960, Dec. 2014.

[8] D.-M. Tsai, S.-C. Wu, and W.-Y. Chiu, "Defect detection in solar modules using ICA basis images," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 122–131, Feb. 2013.

[9] S. Roy, M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "An irradiance-independent, robust ground-fault detection scheme for PV arrays based on spread spectrum time-domain reflectometry (SSTDR)," *IEEE Trans. Power Electron.*, vol. 33, no. 8, pp. 7046–7057, Aug. 2018.

[10] T. Takashima, J. Yamaguchi, K. Otani, T. Oozeki, K. Kato, and M. Ishida, "Experimental studies of fault location in PV module strings," *Sol. Energy Mater. Sol. Cells*, vol. 93, nos. 6–7, pp. 1079–1082, Jun. 2009.

[11] A. Chouder and S. Slivestre, "Automatic supervision and fault detection of PV systems based on power losses analysis," *Energy Convers. Manage.*, vol. 51, no. 10, pp. 1929–1937, 2010.

[12] R. Hariharan, M. Chakkarapani, G. S. Ilango, and C. Nagamani, "A method to detect photovoltaic array faults and partial shading in PV systems," *IEEE J. Photovolt.*, vol. 6, no. 5, pp. 1278–1285, Sep. 2016.

[13] C.-L. Kuo, J.-L. Chen, S.-J. Chen, C.-C. Kao, H.-T. Yau, and C.-H. Lin, "Photovoltaic energy conversion system fault detection using fractional-order color relation classifier in microdistribution systems," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1163–1172, May 2017.

[14] M. Dhimish, V. Holmes, B. Mehrdadi, M. Dales, and P. Mather, "Photovoltaic fault detection algorithm based on theoretical curves modelling and fuzzy classification system," *Energy*, vol. 140, pp. 276–290, Dec. 2017.

[15] F. Harrou, B. Taghezouit, and Y. Sun, "Improved kNN-based monitoring schemes for detecting faults in PV systems," *IEEE J. Photovolt.*, vol. 9, no. 3, pp. 811–821, May 2019.

[16] D. S. Pillai and N. Rajasekar, "An MPPT based sensorless line-line and line-ground fault detection technique for PV systems," *IEEE Trans. Power Electron.*, to be published. doi: 10.1109/TPEL.2018.2884292.

[17] Y. Hu, J. Zhang, W. Cao, J. Wu, G. Y. Tian, S. J. Finney, and J. L. Kirtley, "Online two-section PV array fault diagnosis with optimized voltage sensor locations," *IEEE Trans. Ind. Electron.*, vol. 62, no. 11, pp. 7237–7246, Nov. 2015.

[18] Z. Yi and A. H. Etemadi, "Line-to-line fault detection for photovoltaic arrays based on multiresolution signal decomposition and two-stage support vector machine," *IEEE Trans. Ind. Electron.*, vol. 64, no. 11, pp. 8546–8556, Nov. 2017.

[19] Z. Yi and A. H. Etemadi, "Fault detection for photovoltaic systems based on multi-resolution signal decomposition and fuzzy inference systems," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1274–1283, May 2017.

[20] Y. Zhao, R. Ball, J. Mosesian, J. F. D. Palma, and B. Lehman, "Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2848–2858, May 2015.

[21] H. Momeni, N. Sadoogi, M. Farrokhifar, and H. F. Gharibeh, "Fault diagnosis in photovoltaic arrays using GBSSL method and proposing a fault correction system," *IEEE Trans. Ind. Informat.*, to be published. doi: 10.1109/TII.2019.2908992.

[22] O. Hachana, G. M. Tina, and K. E. Hemsas, "PV array fault diagnostic technique for BIPV systems," *Energy Buildings*, vol. 126, pp. 263–274, Aug. 2016.

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

[24] S. Spataru, D. Sera, T. Kerekes, and R. Teodorescu, "Diagnostic method for photovoltaic systems based on light I–V measurements," *Sol. Energy*, vol. 119, pp. 29–44, Sep. 2015.

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

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

[27] A. Belaout, F. Krim, A. Mellit, B. Talbi, and A. Arabi, "Multiclass adaptive neuro-fuzzy classifier and feature selection techniques for photovoltaic array fault detection and classification," *Renew. Energy*, vol. 127, pp. 548–558, Nov. 2018.

[28] K. Ishaque and Z. Salam, "An improved modeling method to determine the model parameters of photovoltaic (PV) modules using differential evolution (DE)," *Sol. Energy*, vol. 85, pp. 2349–2359, Sep. 2011.

[29] J. J. Soon and K.-S. Low, "Photovoltaic model identification using particle swarm optimization with inverse barrier constraint," *IEEE Trans. Power Electron.*, vol. 27, no. 9, pp. 3975–3983, Sep. 2012.

[30] J. S. C. M. Raj and A. E. Jeyakumar, "A novel maximum power point tracking technique for photovoltaic module based on power plane analysis of I–V characteristics," *IEEE Trans. Ind. Electron.*, vol. 61, no. 9, pp. 4734–4745, Sep. 2014.

[31] J. D. Bastidas-Rodríguez, E. Franco, G. Petrone, C. A. Ramos-Paja, and G. Spagnuolo, "Model-based degradation analysis of photovoltaic modules through series resistance estimation," *IEEE Trans. Ind. Electron.*, vol. 62, no. 11, pp. 7256–7265, Nov. 2015.

[32] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, 1996.

[33] T. F. Coleman and Y. Li, "On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds," *Math. Programm.*, vol. 67, no. 2, pp. 189–224, Mar. 1994.

[34] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, "Approximate solution of the trust region problem by minimization over two-dimensional subspaces," *Math. Program.*, vol. 40, nos. 1–3, pp. 247–263, Jan. 1988.

[35] M. A. Branch, T. F. Coleman, and Y. Li, "A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems," *SIAM J. Sci. Comput.*, vol. 21, pp. 1–23, Aug. 1999.

[36] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, vol. 4, Nov./Dec. 1995, pp. 1942–1948.

[37] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micro Mach. Hum. Sci.*, Oct. 1995, pp. 39–43.
[38] Y. Shi and R. C. Eberhart, "A modified particle swarm optimizer," in *Proc. IEEE World Congr. Comput. Intell.*, May 1998, pp. 69–73.
[39] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proc. Congr. Evol. Comput. (CEC)*, Jul. 1999, pp. 1945–1950.
[40] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.
[41] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Mach. Learn.*, Jul. 1996, pp. 148–156.
[42] R. E. Schapire, "A brief introduction to boosting," in *Proc. 16th Int. Joint Conf. Artif. Intell.*, Jul. 1999, pp. 1401–1406.
[43] R. E. Schapire and Y. Freund, *Boosting: Foundations and Algorithms*. Cambridge, MA, USA: MIT Press, 2012.
[44] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class AdaBoost," *Statist. Interface*, vol. 2, no. 3, pp. 349–360, 2009.
[45] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York, NY, USA: Taylor & Francis, 1984.

JUN-MING HUANG received the B.S. degree in electrical engineering and automation from the Fujian University of Technology, China, in 2017. He is currently pursuing the joint M.S. degree with Fuzhou University in China and National Taiwan University of Science and Technology in Taiwan. His research interests include machine learning and photovoltaic fault diagnosis.

RONG-JONG WAI (M'99–SM'05) was born in Tainan, Taiwan, in 1974. He received the B.S. degree in electrical engineering and the Ph.D. degree in electronic engineering from Chung Yuan Christian University, Chung Li, Taiwan, in 1996 and 1999, respectively.

From 1998 to 2015, he was with Yuan Ze University, Chung Li, where he was the Dean of General Affairs, from 2008 to 2013, and the Chairman of the Department of Electrical Engineering, from 2014 to 2015. Since 2015, he has been with the National Taiwan University of Science and Technology, Taipei, Taiwan, where he is currently a Distinguished Professor, the Dean of General Affairs, and the Director of the Energy Technology and Mechatronics Laboratory. He is a chapter-author of the *Intelligent Adaptive Control: Industrial Applications in the Applied Computational Intelligence Set* (Boca Raton, FL: CRC Press, 1998) and the coauthor of the *Drive and Intelligent Control of Ultrasonic Motor* (Tai-chung, Taiwan, R.O.C.: Tsang-Hai, 1999), the *Electric Control* (Tai-chung, Taiwan, R.O.C.: Tsang-Hai, 2002), and the *Fuel Cell: New Generation Energy* (Tai-chung, Taiwan, R.O.C.: Tsang-Hai, 2004). He has authored over 170 conference papers, 185 international journal papers, and 57 inventive patents. His research interests include power electronics, motor servo drives, mechatronics, energy technology, and control theory applications. The outstanding achievement of his research is for contributions to real-time intelligent control in practical applications and high-efficiency power converters in energy technology.

Dr. Wai is a Fellow of the Institution of Engineering and Technology, U.K., and a Senior Member of the Institute of Electrical and Electronics Engineers, USA. He received the Excellent Research Award, in 2000, and the Wu Ta-You Medal and Young Researcher Award, in 2003 from the National Science Council, Taiwan. In addition, he was a recipient of the Outstanding Research Award from the Yuan Ze University, Taiwan, in 2003 and 2007, respectively; the Excellent Young Electrical Engineering Award and the Outstanding Electrical Engineering Professor Award from the Chinese Electrical Engineering Society, Taiwan, in 2004 and 2010, respectively; the Outstanding Professor Award from the Far Eastern Y. Z. Hsu-Science and Technology Memorial Foundation, Taiwan, in 2004 and 2008, respectively; the International Professional of the Year Award from the International Biographical Centre, U.K., in 2005; the Young Automatic Control Engineering Award from the Chinese Automatic Control Society, Taiwan, in 2005; the Yuan-Ze Chair Professor Award from the Far Eastern Y. Z. Hsu-Science and Technology Memorial Foundation, Taiwan, in 2007, 2010, and 2013, respectively; the Electric Category-Invent Silver Medal Award, in 2007, the Electronic Category-Invent Gold and Silver Medal Awards, in 2008, the Environmental Protection Category-Invent Gold Medal Award, in 2008, the Most Environmental Friendly Award, in 2008, the Power Category-Invent Bronze Medal Award, in 2012, and the Electronic Category-Invent Gold and Silver Medal Awards from the International Invention Show and Technomart, Taipei, Taiwan, in 2015; the University Industrial Economic Contribution Award from the Ministry of Economic Affairs, Taiwan, in 2010; the Ten Outstanding Young Award from the Ten Outstanding Young Person's Foundation, Taiwan, in 2012; the Taiwan Top 100 MVP Managers Award from *MANAGER Today Magazine*, Taiwan, in 2012; the Outstanding Engineering Professor Award from the Chinese Institute of Engineers, Taiwan, in 2013; the Green Technology Category-Scientific Paper Award from the Far Eastern Y. Z. Hsu-Science and Technology Memorial Foundation, Taiwan, in 2014, the Scopus Young Researcher Lead Award-Computer Science from Taiwan Elsevier, in 2014, the Outstanding Research Award from the National Taiwan University of Science and Technology, Taiwan, in 2016 and 2018, respectively, and the Most Cited Researchers Award, in 2016 (Field: electrical and electronics engineering).

WEI GAO was born in Pingtan, China, in 1983. He received the B.S. degree in electrical engineering and automation and the M.S. degrees in power system and automation from Fuzhou University, China, in 2005 and 2008, respectively. He is currently pursuing the Ph.D. degree with the National Taiwan University of Science and Technology, Taiwan. Since 2008, he has been a Lecturer with Fuzhou University. He has authored or coauthored about six journal papers (EI) and one book, and holds about three patents. His research area mainly focuses on generation technology of photovoltaic and faults diagnosis of power equipment.