

PAPER • OPEN ACCESS

A novel fault diagnosis method for photovoltaic array based on BP-Adaboost strong classifier

To cite this article: Y L Zheng *et al* 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **188** 012110

View the [article online](#) for updates and enhancements.

You may also like

- [Robust reconfiguration strategies for maximum output power from large-scale photovoltaic arrays under partial shading conditions](#)

Mona Sharma, Smita Pareek and Kulwant Singh

- [Analysing active power reserve strategies for photovoltaic systems under varying shading scenarios: a comparative study](#)

Pankaj Verma and Nitish Katal

- [Design methodology and implementation of stand-alone solar photovoltaic power system for daily energy consumption of 9.16 kWh](#)

Awoyinka Tunde Dare, David Timothy Wemimo, Somefun Tobiloba Emmanuel et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

A novel fault diagnosis method for photovoltaic array based on BP-Adaboost strong classifier

Y L Zheng^{1,2}, P J Lin^{1,2,4}, J L Yu^{1,2}, Y F Lai^{1,2}, Y H Lin³, Z C Chen^{1,2}, L J Wu^{1,2}, S Y Cheng^{1,2,4} and G D Chen^{1,2}

¹College of Physics and Information Engineering, and Institute of Micro-Nano Devices and Solar Cells, Fuzhou University, 350116 Fuzhou, China

²Jiangsu Collaborative Innovation Center of Photovoltaic Science and Engineering, 213164 Changzhou, China

³College of Computer and Information Sciences, Fujian Agriculture and Forest University, Fuzhou 350002, China

E-mail: linpeijie@fzu.edu.cn/sycheng@fzu.edu.cn

Abstract. Accurate fault diagnosis of photovoltaic (PV) array is important for effective operation of PV systems. The back propagation neural network (BPNN) based classifier model has wide application in fault diagnosis for PV array. Due to insufficient accuracy obtained by using single BPNN, this paper proposes a novel fault diagnosis scheme based on BP-Adaboost strong classifier. Firstly, several indicators constitute an effective feature vector which is applied to build several BPNN based weak classifier models. Secondly, Adaboost algorithm is adopted to build a strong classifier by combining those weak classifiers into the final output with certain weights. Four operation conditions including normal condition, short circuit fault, partial shade fault, open circuit fault can be accurately identified by the proposed method. Dataset from a 1.8 kW grid-connected PV system with 6×3 PV array are applied to experimentally test the performance of the developed method.

1. Introduction

Owing to environmental pollution and continuous growth in global energy depletion, clean renewable energy demand increases. Solar energy is a promising renewable energy due to its advantages of non-pollution and sustainability, whose main form of application is photovoltaic (PV) power generation [1]. Recently, the global PV installed capacity has greatly increased for the improvement of PV technology and the support of national policies. However, as the core components of photoenergy acquisition, PV arrays are exposed to harsh outdoor environment which may cause several faults. The faults may reduce the power generation efficiency, damage PV modules, and even lead to fire risk.

To address these issues, various fault diagnosis approaches for PV array have been proposed. The conventional method is comparing the difference between the measured and estimated variables. By analyzing the details of the expected and actual AC power, faults can be detected [2]. Drews *et al* develop a smart approach for fault diagnosis of PV array, the algorithm can detect the probable faults once the defined difference between the actual and simulated energy yield occurs [3]. Conventional method requires an accurate model, and the method may fail for the impacts of changeable environment. Therefore, methods based on machine learning are proposed to detect fault in PV arrays, since it doesn't require accurate mathematical models. Karmacharya *et al* propose a novel fault



diagnosis strategy, which applies the multi-resolution analysis (MRA) to extract the unique features of monitored signal as the inputs of a three-layer feed forward artificial neural network (ANN) classifier, which realizes the fault location of ungrounded PV power generation system [4]. The method proposed in [5] can train probabilistic neural network (PNN) faults classifiers with higher noise tolerance. We have been studying on PV fault diagnosis and achieve some results. In [6], a density peak-based clustering approach is proposed to identify faults automatically. Chen et al. introduce an efficient method based on kernel extreme learning machine (KELM) to achieve a high accuracy of faults diagnosis [7].

As a machine learning technique, ANN is widely used for PV array fault diagnosis [8]. Syafaruddin et al. present a fault diagnosis method based on three-layer ANN to locate the short circuit fault in PV modules [9]. Approach of BP neural network (BPNN) based on Levenberg-Marquardt (L-M) algorithm effectively detects open circuit, short circuit, partial shading and abnormal degradation faults in PV array [10]. However, single BPNN based classifier model has slow convergence and easily getting into local minimum [9]. In addition, the single BPNN based classifier model is a weak classifier due to the insufficient accuracy of classification [11]. Therefore, this paper proposes a fault diagnosis approach for PV arrays based on BP-Adaboost strong classifier to improve the performance of fault diagnosis and classification.

2. Strong classifier based on BP-Adaboost

The Adaboost is one of the most successful boosting algorithm [12]. The idea of the Adaboost algorithm is combining the outputs of several “weak” classifiers with certain weights to make the classification better [13]. Firstly, the data samples and the learning algorithm of weak classifier are given. Then m sets of data are randomly selected from the data samples as training data, the weight of each training data is set as $1/m$. Secondly, there are T same weak classifiers used for iterative operations. When each iteration finishes, weight distribution of each training data will be adjusted according to the classification result. After repeated iterative calculations, each weak classifier can be described as a corresponding classification function. For the result of classification, the better effect it has, the greater weight of the function. Finally, classification result of the strong classifier is weighted by all weak classifiers with a certain weight.

BP-Adaboost algorithm is a case of using the Adaboost to the boosting the performance of BPNN [13], BPNN is regard as the weak classifiers of strong classifier model. The flowchart, given in figure 1, describes the steps of forming a strong classifier

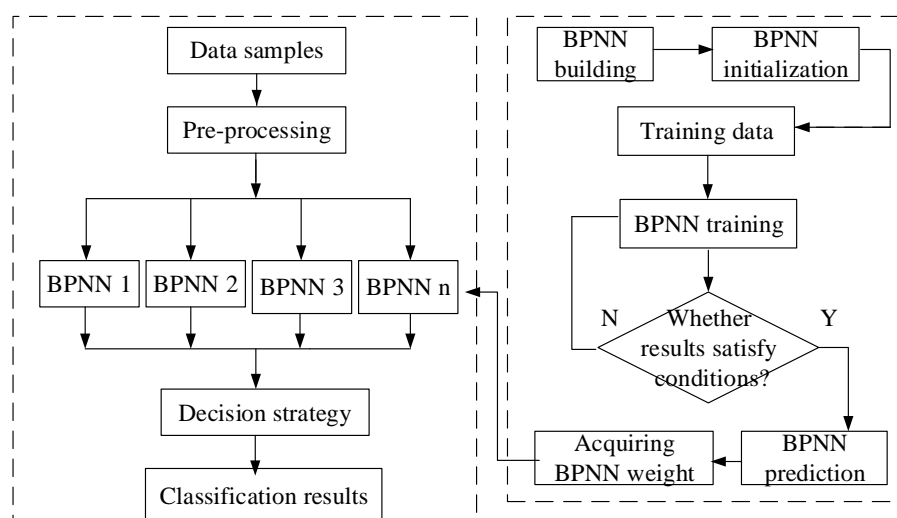


Figure 1. Flowchart of strong classifier based on BP-Adaboost.

Step 1: Data selection and BPNN initialization. Firstly, selecting m sets of data randomly from the samples as training data, which is $\{(x_i, y_i)\}_{i=1}^m$ ($y_i = [0, \dots, 0, 1, 0, \dots, 0]$), the distribution weight of the sample is $D_t(i) = 1/m$. Secondly, determining the node number of the input layer, hidden layer and output layer in the BPNN respectively. And the weight and threshold of BPNN are initialized.

Step 2: Data prediction with weak classifiers. Using $\{D_t(i) = 1/m\}_{i=1}^m$ as the priori probability function and training the t -th BPNN weak classifier with the training data by standard BP algorithm. The prediction error sum e_t of the weak classifier is calculated as follow [14].

$$e_t = \sum_i D_t(i) \quad i = 1, 2, \dots, m (g(t) \neq y) \quad (1)$$

where t represents the index of classifier, $g(t)$ is the classification prediction result, and y describes the expected classification result.

Step 3: Calculating the weight of prediction sequence. Using e_t obtained in step 2 to calculate the prediction sequence weight a_t [15].

$$a_t = \frac{1}{2} \ln\left(\frac{1-e_t}{e_t}\right) \quad (2)$$

Step 4: Updating weight of sample. Weight of sample in next training is adjusted according to the a_t obtained in step 3 [16]. The adjustment method is given as follows:

$$D_{t+1}(i) = \frac{D_t(i)}{B_t} * \exp[-a_t y_i g_t(x_i)] \quad i = 1, 2, \dots, m \quad (3)$$

where B_t represents a normalization factor which is used to ensure $\sum_i D_t(i) = 1$ when the weight ratio stays the same.

Step 5: Obtaining the function of strong classifier. After T times training, acquiring T weak Classification functions $f(g_t, a_t)$ which are combined to obtain the final strong classification function $h(x)$. The calculation equation is shown in equation (4) [14].

$$h(x) = \text{sign}\left[\sum_{t=1}^T a_t * f(g_t, a_t)\right] \quad (4)$$

3. The proposed fault classification scheme

Based on trial and error, the proposed approach is mainly composed of 10 BPNN based weak classifiers, the outputs of 10 BPNN are merged by certain weights to obtain the final output result. The parameters used in the proposed model are given as follows:

Input layer: Six indicators are selected as the effective feature vector of input layer, including irradiance (S_n), temperature of PV array (T_n), the voltage of PV array at maximum power point (V_{MPP}), the currents of PV string S at maximum power point (I_1 to I_5).

Hidden layer: the number of nodes in the hidden layer is 8, maximum learning times is 100, learning rate is 0.1, convergence error is 0.00004.

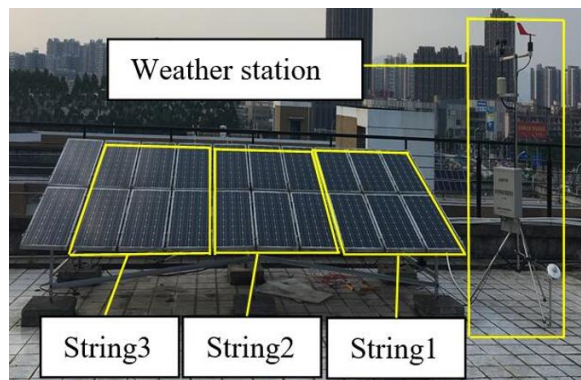
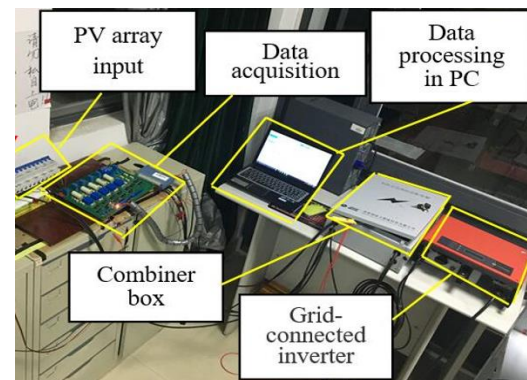
Output layer: Output layer describes the number of output classification categories. The operation conditions studied in this paper are as follows: normal condition (NORMAL), one string occurs open circuit fault (OPEN-1), two strings occur open circuit fault (OPEN-2), one module occurs short circuit fault in one PV string (SHORT-1), two module occurs short circuit fault in one PV string (SHORT-2), one module is shaded (SHADE-1), and two module is shaded (SHADE-2). Table 1 shows the correspondence between the output number of BPNN and the conditions of PV array.

Table 1. The correspondence between the output number of BPNN and the conditions of PV array.

Output status	Operation conditions
1	NORMAL
2	OPEN-1
3	OPEN-2
4	SHORT-1
5	SHORT-2
6	SHADE-1
7	SHADE-2

4. Experimental results and discuss

The experiment is carried out on a rooftop photovoltaic power station in laboratory. The system consists of PV array, weather station, combiner box, inverter, data acquisition module and data processing computer, as shown in figures 2 and 3. The PV array is composed of 3 strings connected in parallel ($S=3$), and each string has 6 modules connected in series. Under standard test conditions (STC), the open circuit voltage (V_{OC}) is 21.5V, the short circuit current (I_{SC}) is 6A and the output power is 100 W. The grid-connected inverter is GW2500-NS.

**Figure 2.** The physical map of PV array.**Figure 3.** Data collection system.

The data acquisition module is composed of sampling circuit and data acquisition card. The isolated Hall voltage sensor (LV-25P) and current sensors (HBC-LSP) are used to collect the voltage of the PV array and the currents of PV strings, respectively. The temperature sensor (PT100) is applied to acquire the temperature of PV array. Irradiance is obtained by placing an irradiance sensor (FZD-V1-2000) on the same inclined surface as the PV array, which can convert the irradiance information into voltage value for acquisition. In this experiment, the irradiance varies from 200 W/m² to 800 W/m². Then, setting the sampling frequency of the data acquisition card (USB-1608G) to 200 Hz. Collecting the samples data and storing it in the computer by the data acquisition card.

Once the PV system starts working, the maximum power point tracking (MPPT) algorithm of the inverter begins to find the optimal operation point automatically. Therefore, we acquire the dataset after PV system operating steadily which ensures the dataset is measured under maximum power point.

The six attributes, including S_n , T_n , V_{MPP} , I_1 , I_2 , and I_3 , are selected as the inputs for the classification model. A total 2584 data samples are acquired as the data set, including 391 samples under NORMAL, 359 samples under OPEN-1, 364 samples under OPEN-2, 365 samples under SHADE-1, 370 samples under SHADE-2, 372 samples under SHORT-1 and 363 samples under SHORT-2. 1800 (about 70%) samples are randomly chosen for training the model, and the rest (about 30%) samples for testing. Firstly, the function 'rand' is used to generate a random sequence consisting of 2584 random numbers from 0 to 1. Secondly, the random sequence is arranged from small to large and the indexes of each random number in the original sequence can be acquired by the function

‘sork’. The indexes compose a rand sequence consisting of integer numbers from 1 to 2584. Finally, the rand sequence is utilized as the index of samples data, by that, the samples data can be arranged randomly.

Before training, the proportional compression method is adopted to normalize the same type data set. Assuming a data set $x = [x_1, x_2, \dots, x_n]$, the procedure is shown in equation (5).

$$y = (y_{\max} - y_{\min}) \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} + y_{\min} \quad (5)$$

Where y represents the normalized data, x_{\max} is the maximum value in the data set, x_{\min} is the minimum value, y_{\max} is usually set to 1, and y_{\min} can be set to -1.

The proposed BP-Adaboost based fault diagnose approach is test by using MATLAB R2014b software. Figure 4 shows the training accuracy of the presented model. The blue “ Δ ” in the figure represents the experiment sample, the pink “*” represents the predicted results of training model. As can be seen from figure 4, there are 35 data points are wrongly classified, and the overall training accuracy of the model is 98.1%.

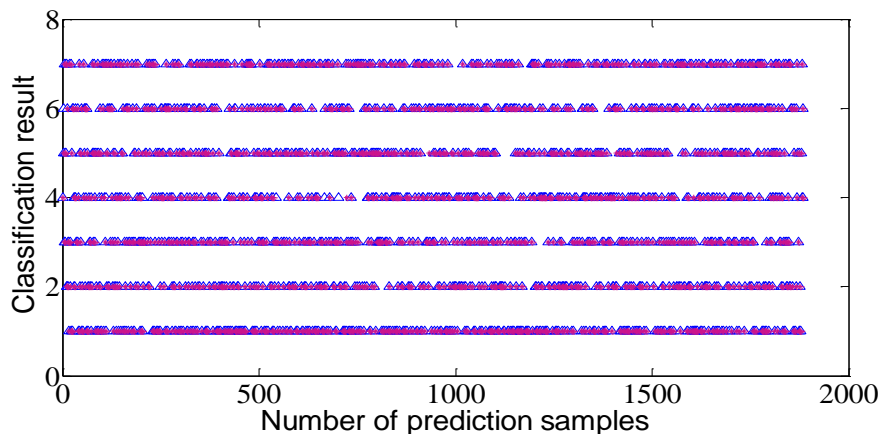


Figure 4. Experimental training results of the proposed model.

The prediction results of the proposed model are shown in figure 5. The overall prediction accuracy of the model is 97.7%.

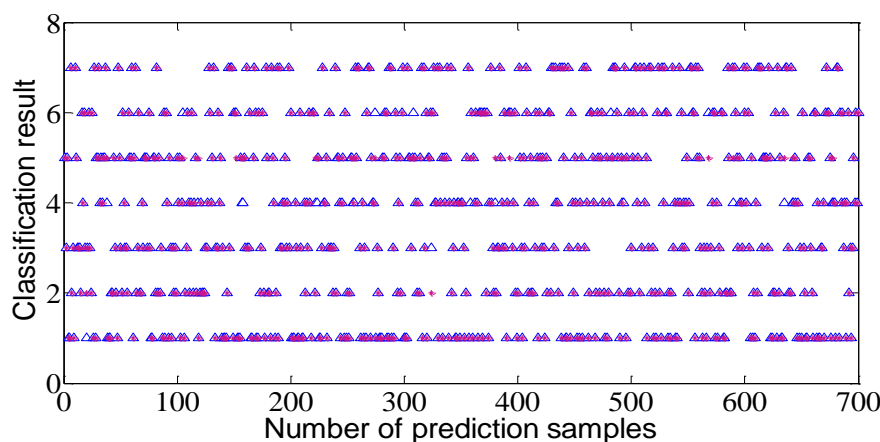


Figure 5. Experimental prediction results of the proposed model.

The detailed fault detection accuracy for the claimed seven conditions of PV array is summarized in table 2. In terms of training results, the classification accuracy of NORMAL, OPEN-1, OPEN-2, SHORT-2 and SHADE-2 is more than 99%, and the accuracy of SHORT-1 and SHADE-1 is about 94%. For the prediction results, the classification accuracy of NORMAL, OPEN-1, OPEN-2, SHORT-2 and SHADE-2 also reaches more than 97%. The accuracy of SHORT-1 and SHADE-1 is about 95%, which may be due to the fact that SHADE-1 and SHORT-1 have similar effect on the PV array.

Table 2. Experimental results of fault classification model.

Operating conditions	Training accuracy	Prediction accuracy
NORMAL	99.1%	98.0%
OPEN-1	100%	97.1%
OPEN-2	100%	100%
SHORT-1	94.4%	95.3%
SHORT-2	100%	100%
SHADE-1	93.8%	95.3%
SHADE-2	99.6%	97.8%

In addition, in order to compare the performance of the proposed model and that of single BPNN, the training and classification results of the 10 BPNN based classifiers are listed in table 3. The average training accuracy of the 10 classifiers is 96.8%, which is lower than the average training accuracy of the proposed strong classifier (98.1%). Moreover, the average prediction accuracy of the 10 classifiers is 96.6%, which is also worse than that of the strong classifier (97.7%). Therefore, the strong classifier based on BP-Adaboost is superior to single BPNN.

Table 3. Experimental training and prediction results of BPNNs.

Index of BPNN	Accuracy of training data	Accuracy of prediction data
1	94.5%	92.9%
2	95.8%	94.0%
3	97.5%	96.3%
4	97.7%	96.6%
5	97.1%	96.6%
6	96.5%	94.0%
7	98.4%	97.1%
8	96.0%	94.7%
9	97.4%	96.4%
10	97.2%	96.6%

To further verify the ability of the proposed method, PNN based model is also applied for comparison [17]. The accuracy of training and prediction for PNN are 94.75% and 95.57%, respectively. Therefore, the accuracy of the BP-Adaboost based model is obviously higher than that of PNN.

5. Conclusion

A fault diagnosis method for PV array based on BP-Adaboost strong classifier has been presented. Six

attributes including S_n , T_n , V_{MPP} , I_1 , I_2 , and I_3 are selected as the inputs of the classifier. Four types of operating cases including normal condition, short circuit fault, partial shade and open circuit fault can be accurately detected and classified. According to the experiment, the average classification accuracy of the 10 BPNN models is about 96.6%, the accuracy is 95.57% for PNN and the proposed Adaboost algorithm based model achieves best accuracy over 97%. Results demonstrate the promising performance of the proposed method for fault diagnosis in PV array.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61574038, 61601127, and 51508105), the Science Foundation of Fujian Science & Technology Department (Grant Nos. 2018J01774, 2018J01645, and 2016H6012) and the Science Foundation of Fujian Education Department (Grant No. JAT160073), and the Fujian Provincial Economic and Information Technology Commission of China (Grant Nos. 83016006 and 830020).

References

- [1] Madeti S R and Singh S N 2017 A comprehensive study on different types of faults and detection techniques for solar photovoltaic system *Sol Energy* **158** 161-85
- [2] Platon R, Martel J, Woodruff N and Chau T Y 2015 Online fault detection in PV systems *IEEE T Sustain Energ* **6** 1200-7
- [3] Drews A, *et al* 2007 Monitoring and remote failure detection of grid-connected PV systems based on satellite observations *Sol Energy* **81** 548-64
- [4] Karmacharya I M and Gokaraju R 2017 Fault location in ungrounded PV system using wavelets and ANN *IEEE T Power Deliver* **33** 549-59
- [5] Garoudja E, Chouder A, Kara K and Silvestre S 2017 An enhanced machine learning based approach for failures detection and diagnosis of PV systems *Energ Convers Manage* **151** 496-513
- [6] Lin P J, Lin Y H, Chen Z C, Wu L J, Chen L C and Cheng S Y 2017 A density peak-based clustering approach for fault diagnosis of PV arrays *Int J Photoenergy* 4903613
- [7] Chen Z C, Wu L J, Cheng S Y, Lin P J, Wu Y and Lin W C 2017 Intelligent fault diagnosis of PV arrays based on optimized kernel extreme learning machine and IV characteristics *Appl Energ* **204** 912-31
- [8] Pillai D S and Rajasekar N 2018 A comprehensive review on protection challenges and fault diagnosis in PV systems *Renew Sust Energ Rev* **91** 18-40
- [9] Syafaruddin, Karatepe E and Hiyama T 2011 Controlling of artificial neural network for fault diagnosis of photovoltaic array *International Conference on Intelligent System Application To Power Systems (ISAP) IEEE* (Hersonisos Crete Greece) pp 1-6
- [10] Wang Y Z, Chun-Hua W U, Zhou D Q and Li Z H 2013 A survey of fault diagnosis for PV array based on BP neural network *Power System Protection & Control* **41** 108-14
- [11] Wang J and Yu J Y 2011 Scientific creativity research based on generalizability theory and BP_Adaboost RT *Procedia Engineering* **15** 4178-82
- [12] Freund Y, Iyer R, Schapire R E and Singer Y 2003 An efficient boosting algorithm for combining preferences *J Mach Learn Res* **4** 933-69
- [13] Zhang Y D and Wu L N 2009 Stock market prediction of S & P 500 via combination of improved BCO approach and BP neural network *Expert Syst Appl* **36** 8849-54
- [14] Li N J, Cheng X, Zhang S F and Wu Z Y 2013 Recognizing human actions by BP-AdaBoost algorithm under a hierarchical recognition framewor *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) IEEE* (Vancouver BC Canada) pp 3407-11
- [15] Li H H, Chen Q S, Zhao J W and Wu M Z 2015 Nondestructive detection of total volatile basic nitrogen (TVB-N) content in pork meat by integrating hyperspectral imaging and colorimetric sensor combined with a nonlinear data fusion *LWT-Food Sci Technol* **3** 268-74

- [16] Zhang J, Yang Y P and Zhang J M 2016 A MEC-BP-Adaboost neural network-based color correction algorithm for color image acquisition equipments *Optik-International Journal for Light and Electron Optics* **127** 776-80
- [17] Akram M N and Lotfifard S 2015 Modeling and health monitoring of DC side of photovoltaic array *IEEE T Sustain Energ* **6** 1245-53