Fault diagnosis of PV array using adaptive network based fuzzy inference system

Yutao Gan, Zhicong Chen¹, Lijun Wu, Shuying Cheng and Peijie Lin

College of Physics and Information Engineering, Fuzhou University, 2 XueYuan Road, 350116 Fuzhou, China

¹Email: zhicong.chen@fzu.edu.cn

Abstract. A new online intelligent fault diagnosis method is proposed for PV arrays in this paper to improve the reliability and efficiency of PV systems. Firstly, a new seven-dimensional fault feature vector is extracted from the raw data of dynamic operating points of PV arrays including operating voltage, current, irradiance and temperature. Secondly, an optimized adaptive network based fuzzy inference system (ANFIS) is proposed as the fault diagnosis model. Lastly, the feasibility and superiority of the proposed ANFIS based fault diagnosis model are tested by both Simulink based simulation and real fault experiments on a laboratory PV system. Experimental results validate that the proposed ANFIS based method achieves a high performance and is superior to conventional back-propagation neural network (BPNN) based methods. The overall accuracy of the ANFIS based fault diagnosis model on the simulation and experimental dataset is 99.9% and over 97% respectively.

1. Introduction

To address the increasingly severe energy crisis, environment pollution and climate change, the installed capacity of PV systems is rapidly increasing in recent years. As the core of PV systems, PV arrays are susceptible to damage due to the complex installation environment, which may affect the power generation efficiency of PV array. Consequently, fault diagnosis of PV arrays is significant to improve the reliability and efficiency of PV systems [1]. At present, there are many kinds of fault diagnosis of PV arrays. From the perspective of data sources, they can be categorized into two types, the methods based on infrared thermal images and the methods based on output characteristics of PV arrays [1, 2]. The former can find the defects of the photovoltaic panel earlier when there is no major defect in the output characteristics of the PV array. The operating cost and data collection cost of this method are very high, which is not economic to wide-scale promotion. Data sources for diagnostic methods based on the output characteristics of PV arrays include I-V curves [3, 4], time series graph [5], and dynamic operating point voltage and current [1, 6]. The method based on I-V curve can collect more PV array information and detect PV faults more comprehensively [4], but it needs offline operation, which will interrupt the normal operation of PV systems. However, the fault diagnosis method based on dynamic operating points does not need to interrupt the normal operation of the PV systems. The time series graph signal has the advantage of high signal matching when the fault occurs but can only find the change of the PV array from normal working conditions to faulty working conditions. The dynamic operating points based method has a wider application scenario, because the dynamic operating point based fault diagnosis method can diagnose the normal state and the stable
fault state, and the time series graph signal based method can't do anything for the fault that has already become a steady state fault, because only the transients at the time of the fault can be detected.

In recent years, as the core of artificial intelligence, machine learning has been widely used in data processing due to the advantages of being able to process large amounts of data quickly. It is applied in the photovoltaic fault diagnosis method and is trained with labelled data, which has a very high accuracy rate of fault diagnosis. In machine learning, artificial neural networks (ANN) have strong self-learning ability and memory ability, and have strong mapping ability to the nonlinear output of PV arrays [7]. However, the shortcomings of neural networks are the inability to explain their own reasoning ability and reasoning basis, need a lot of data training, which will cause a big change in the results when the data changes slightly. Fuzzy logic inference systems (FIS) can apply fuzzy sets and fuzzy rules for reasoning. However, the FIS has no self-learning ability, and it is necessary to adopt the rules of expert experience. The formulation of the rules of expert experience greatly limits the application of the FIS, which greatly increases the difficulty of the work. The adaptive network based fuzzy inference system (ANFIS) combines the advantages of both the FIS and ANN, which does not require a large amount of training data and features fast training speed, fuzzy reasoning, and strong anti-interference ability which it is very suitable for fault diagnosis of nonlinear systems [8]. The main contribution of this study is to propose a new fault diagnosis method for PV array, which is based on the ANFIS and the dynamic operating point current together with the temperature and irradiance of the PV array. The proposed fault diagnosis method can continuously detect and diagnose the condition of the PV array without interrupting the working state of the PV arrays and the proposed algorithm model has better generalization performance and anti-interference performance.

2. Adaptive network based on fuzzy inference system for fault diagnosis of PV array

This section details the ANFIS based diagnosis model for detecting several common faults of PV arrays, including line-line fault, open-circuit fault, partial shadow fault and degradation fault. Firstly, a seven dimensional fault feature is extracted from the raw data as the input of the model. Secondly, the ANFIS model structure is detailed. Lastly, the ANFIS based model is trained and tested by the K-fold cross-validation method and the ergodic method.

2.1. Extraction of fault features

The proposed ANFIS based fault diagnosis is based on the raw monitored data of the operating point, including the operating voltage of PV array, operating current of each PV string, ambient temperature, and ambient irradiance. In order to establish high performance diagnosis model, the fault features should be carefully designed and extracted from the raw data. In this study, as defined in Eqs. (1) to (7), seven fault features are proposed, including the normalized PV array voltage ($V_n$), array current ($I_n$), array power ($P_n$), slope of the operating point ($S_n$), current dispersion rate ($C_n$), ambient temperature ($T_n$), and incident irradiance ($G_n$). In order to eliminate the influence of the environment condition, these seven fault features are all normalized by the corresponding simulation data.

\[
V_n = \frac{V_{mpp}}{V_{op}} \quad (1) \\
I_n = \frac{I_{mpp}}{I_{op}} \quad (2) \\
P_n = \frac{V_{mpp}I_{mpp}}{V_{op}I_{op}} \quad (3) \\
S_n = \frac{V_{op}I_{mpp}}{V_{mpp}I_{op}} \quad (4) \\
C_n = \frac{\mu}{\sigma} \quad (5) \\
G_n = \frac{G_a}{G_{stc}} \quad (6) \\
T_n = \frac{T_a}{T_{stc}} \quad (7) \\
V_{op} = N_s\{V_{stc}[1 + \beta(T_a - T_{stc})]G_a/G_{stc} + nU_0\ln(G_a/G_{stc})\} \quad (8) \\
I_{op} = N_pI_{stc}\{1 + \alpha(T_a - T_{stc})\}G_a / G_{stc} \quad (9)
\]
where, $V_{\text{mpp}}$ is the monitored operating array voltage and $V_{\text{op}}$ is the corresponding theoretical value defined in Eq. (8); $I_{\text{mpp}}$ is the monitored operating array current and $I_{\text{op}}$ is the corresponding theoretical value defined in Eq. (9); $\mu$ is the mean of the string current, and $\sigma$ is the standard deviation of string current; $G_i$ is the monitored incident irradiance, and $G_{\text{stc}}$ is the irradiance under standard test condition (STC), i.e., 1000 $W/m^2$; $T_a$ is the monitored ambient temperature, and $T_{\text{stc}}$ is the temperature under STC, i.e., 25$^\circ$C; $n$ of the ideal factor of the solar cell in the PV modules; $N_s$ is the number of solar cells in series in a PV string; $N_p$ is the number of solar cell strings in parallel in the PV array; $\alpha$ is the short-circuit current temperature coefficient, and $\beta$ is the open circuit voltage temperature coefficient; $U_t$ is the thermal voltage; $V_{\text{stc}}$ and $I_{\text{stc}}$ are the maximum power point (MPP) voltage and current in the STC condition.

2.2. Adaptive network based fuzzy inference system based fault detection and diagnosis model

Using the features selected in section 2.1 as the input of fault diagnosis model, the ANFIS based fault diagnosis model is established for photovoltaic fault diagnosis. The model has seven inputs and one output, which can be used to predict the class of the unlabeled data. The brief process for establishing the fault diagnosis model is illustrated in Figure 1.

Figure 1. Brief flowchart of the ANFIS based fault diagnostic model.

The ANFIS is a fuzzy Sugeno model put in the framework of adaptive systems to facilitate learning and adaptation [9], which consists of 5 layers in total as shown in Figure 2. The output of the nodes in the first layer is:

$$O_i^1 = \mu_{A_i}(X_i), i = 1, 2$$

$$O_i^1 = \mu_{B_i}(X_i), i = 1, 2$$

(10)

(11)

$\mu_{A_i}$ is fuzzy membership function, taking the Gaussian membership function as an example:

$$\mu_{A_i} = \exp[-(X - a_i)^2/b_i], i = 1, 2$$

(12)

where $a_i$ and $b_i$ are the parameters of the membership function, governing the gauss shaped functions accordingly, i represents the number of membership function of each inputs.

The output of the nodes in the second layer is:

$$O_i^2 = W_i = \mu_{A_i}(X_i) \ast \mu_{B_i}(X_i), i = 1, 2$$

(13)

The output of each node represents a regular excitation strength.

The output of node in the third layer is:

$$O_i^3 = M_i = W_i(W_1 + W_2), i = 1, 2$$

(14)
Each node represents the intensity of the stimulus normalization. The output of node in the fourth layer is:

\[ O_i^4 = f_i = M_i(p_iX_1 + q_iX_2 + r_i), i = 1,2 \]  

(15)

Each node indicates the output of each rule. \( p_i, q_i, \) and \( r_i \) are the design parameters that are determined during the training process.

The fifth layer calculates the sum of the fourth layer output nodes.

### 2.3. Training of the fault diagnosis model

Since the adaptive network based fuzzy inference system model for PV array fault diagnosis belongs to a supervised self-learning network, the main parameters are the number and the type of membership functions. The procedure to train the ANFIS model is shown as below and illustrated in Figure 3:

Step 1: The training set is further randomly divided into a training subset and a verification subset by K-fold cross-randomization;

Step 2: Traverse the number and the type of membership function of the ANFIS model in a specified range, so as to establish different ANFIS model;

Step 3: Verify the accuracy of the trained models with the verification subset, and select the model with the highest accuracy to determine the number and the type of membership functions;

Step 4: Use the training subset to train the ANFIS model with optimal parameters to obtain the final ANFIS fault diagnosis model.

### 3. Simulation study

**Figure 3.** Flowchart of parameter optimization for the ANFIS model.

**Figure 4.** PV array simulation with fault condition.
To verify the proposed fault diagnosis method, as illustrated in Figure 4, the normal status and six types of common faults of a laboratory PV array are simulated in Simulink to collect the data samples with the irradiance varying from 100-975 W/m\(^2\) (at the step of 25 W/m\(^2\)) and with the temperature varying from 25-70 °C (at the step of 2.5 °C). The simulated fault includes normal condition (N), line-line fault of string level with one module difference in the same string (LL1) (simulated by connecting a 0.001 ohms resistance between modules), line-line fault of array level with two modules difference (LL2) at two difference PV string (simulated by connecting a 0.001 ohms resistance between modules), open-circuit fault on one string (OC) (simulated by connecting a series resistor of 40000 ohms into the negative of the PV array), partial shading fault (PS) (simulated through setting the irradiance gains K=0.5), degradation fault of PV array level (D-A) (simulated by connecting 4 ohm resistor to the negative output of the PV array), degradation fault of PV string level (D-S) (simulated by connecting 4 ohm resistors between a string of PV and the negative of the PV array). Based on the collected raw data and the parameters of photovoltaic models, the fault features, including \(V_n, I_n, P_n, S_n, C_n, G_n, T_n\), are extracted to form the data samples to train the photovoltaic fault diagnosis model. The number of environment conditions for each fault is 684, and a data sample is collected for every environment conditions. Therefore, the final dataset contains 4788 data samples in total. The dataset is randomly divided into a training set and a test set. The training set accounts for 70% of the total data samples, and the test set accounts for 30% of the total data samples. The training dataset are further divided into training sets and verification sets by 5-fold crossover to search for optimal parameters of ANFIS. In order to avoid contingency, each parameter combination runs 100 times and the result is averaged, and then select the best combination of parameters as the optimal parameters for ANFIS. In this experiment, based on the data of Simulink, the results of finding the optimal parameters of ANFIS by K-fold are shown in Table 1.

### Table 1. The classification accuracy of traversal the type and the number of membership function in 100 times 5-fold crossover test for the simulation data sample.

| Number of MF | Type of MF | Training accuracy | Testing accuracy | Training accuracy | Testing accuracy | Training accuracy | Testing accuracy |
|--------------|------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|
|              | 3          | 4                 | 5                | 3                 | 4                | 5                 |                  |
| Trimf        | 99.32      | 99.16             | 99.75            | 99.58             | 99.81            | 99.76             |                  |
| Trapmf       | 99.26      | 99.24             | 99.24            | 99.18             | 99.52            | 99.40             |                  |
| Gbellmf      | 99.66      | 99.62             | 99.78            | 99.75             | 99.87            | 99.84             |                  |
| Gaussmf      | 99.63      | 99.62             | 99.78            | 99.72             | 99.88            | 99.83             |                  |
| Pimf         | 98.62      | 98.55             | 99.24            | 99.16             | 99.50            | 99.42             |                  |
| Dsigmf       | 99.31      | 99.26             | 99.76            | 99.70             | 99.87            | 99.83             |                  |
| Psigmf       | 99.31      | 99.26             | 99.76            | 99.70             | 99.87            | 99.83             |                  |

### Table 2. Comparison of classification accuracy of BPNN and ANFIS models in different types of faults for the simulation data sample.

| Item | BPNN | ANFIS |
|------|------|-------|
| OC   | 100.00 | 100.00 |
| LL1  | 100.00 | 100.00 |
| LL2  | 98.90  | 100.00 |
| PS   | 98.93  | 100.00 |
| D-S  | 92.92  | 99.04  |
| D-A  | 99.78  | 100.00 |
| N    | 100.00 | 100.00 |
According to the results of Table 1, we find that the result of parameter combination which is Gbellmf and each input 5 membership function are the best combination parameter of fault diagnosis using ANFIS. Therefore, we take this parameters combination as the optimal parameter of ANFIS model. In order to make the results more stable and reliable, the fault diagnosis model is trained and tested for 100 independent times. To verify the superior performance of the ANFIS based model, the common BPNN model is used for the comparison in this experiment. The BPNN model is trained for the same 100 independent times. The average results of the two methods are shown in Table 2.

4. Experimental study
In this section, a 2kw small grid-connected photovoltaic system installed on the roof of the North Building of the College of Physics and Information Engineering of Fuzhou University is used to experimentally verify the proposed photovoltaic fault diagnosis model. The same faults as detailed in previous section 3 is simulated on the laboratory PV array, and the data collection for each the fault condition is performed continuously for 3 hours with the sampling rate of 1 Hz. The collected raw data include the PV array I_{mpp}, V_{mpp}, PV array temperature T_a and environmental irradiance G_a. Therefore, 10800 raw data samples are collected for each fault condition, and thus there are 75600 raw data samples in total. And then, the fault features are extracted from these raw data to form the final data samples. The training and testing of the PV fault diagnosis model is the same as the simulation model as discussed and detailed in section 3. The obtained experimental results are listed in Table 3. Because the experimental data samples are noisy and the simulation data samples are ideal, the training accuracy and the test accuracy of the experimental data is slightly less than those of the simulation data. However, the experimental accuracy is still relatively high, which demonstrates that the PV fault diagnosis model is effective for real PV array. As shown in Table 4, the comparison results show that the overall accuracy of the ANFIS based fault diagnosis model is better than the BPNN based model.

| Number of MF | Training accuracy% | Testing accuracy% | Training accuracy% | Testing accuracy% | Training accuracy% | Testing accuracy% |
|--------------|--------------------|-------------------|--------------------|-------------------|--------------------|-------------------|
| Trimpf       | 90.02              | 90.06             | 91.25              | 91.29             | 92.81              | 92.76             |
| Trampf       | 90.15              | 90.14             | 91.24              | 91.18             | 92.52              | 92.40             |
| Gbellmf      | 92.25              | 92.26             | 93.46              | 93.47             | 96.84              | 96.95             |
| Gaussmf      | 92.27              | 92.17             | 94.48              | 94.42             | 97.88              | 97.68             |
| Pimf         | 91.06              | 91.05             | 92.04              | 92.06             | 95.46              | 95.42             |
| Dsigmf       | 91.09              | 91.02             | 92.17              | 92.16             | 94.38              | 94.34             |
| Psigmf       | 91.02              | 91.06             | 92.17              | 92.20             | 94.36              | 94.35             |

Table 3. The classification accuracy of traversal the type and the number of membership function in 100 times 5-fold crossover test for the experimental data samples.

| Item | BPNN Training accuracy% | BPNN Testing accuracy% | ANFIS Training accuracy% | ANFIS Testing accuracy% |
|------|-------------------------|------------------------|--------------------------|-------------------------|
| OC   | 97.63                   | 99.84                  | 98.54                    | 98.34                   |
| LL1  | 99.97                   | 99.98                  | 99.99                    | 99.99                   |
| LL2  | 80.51                   | 80.11                  | 94.02                    | 96.55                   |
| PS   | 89.42                   | 94.86                  | 98.53                    | 98.91                   |
| D-S  | 90.63                   | 86.01                  | 90.01                    | 89.91                   |
| D-A  | 99.44                   | 99.85                  | 99.89                    | 99.91                   |
| N    | 99.97                   | 99.96                  | 99.97                    | 99.98                   |

Table 4. Comparison of classification accuracy of BPNN and ANFIS in different types of faults for the experimental data samples.
5. Conclusions
This paper proposes a PV array fault diagnosis method based on the adaptive network based fuzzy inference system for online detection of PV array line-line faults, open circuit faults, degradation faults and partial shadow faults, which is based on the operating voltage, current, temperature and irradiance of PV arrays. Seven features are extracted from these raw data to eliminate the influence of environmental factors. In order to obtain the best diagnostic results, the averaging ANFIS parameters and K-fold cross validation are used to search the best parameters of the model. The PV array faults are simulated on a Simulink based model and a real laboratory PV system to obtain massive data samples, so as to verify the accuracy of PV array fault diagnosis model. The results show that the proposed ANFIS based PV fault diagnosis model has superior performance in terms of the accuracy, reliability and generalization performance in comparison to the BPNN based model. The overall accuracy of the proposed model based on the simulation data is 99.9%, while the overall accuracy for the model based on the experimental data is over 97%.

Acknowledgments
The authors would like to acknowledge the financial supports in part from the National Natural Science Foundation of China (Grant Nos. 61601127 and 61574038), the Fujian Provincial Department of Science and Technology of China (Grant Nos. 2019H0006 and 2018J01774), and the Foundation of Fujian Provincial Department of Industry and Information Technology of China (Grant No. 82318075).

References
[1] S R Madeti and S Singh 2017 Monitoring system for photovoltaic plants: A review Renewable and Sustainable Energy Reviews 67 1180-1207
[2] M K Alam, F H Khan, J Johnson and J Flicker 2013 PV faults: Overview, modeling, prevention and detection techniques, in Control and Modeling for Power Electronics (COMPEL), 2013 IEEE 14th Workshop on 1-7 IEEE
[3] Z Chen, L Wu, S Cheng, P Lin, Y Wu and W Lin 2017 Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and IV characteristics Applied Energy
[4] Z Chen, Y Chen, L Wu, S Cheng and P Lin 2019 Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions Energy Conversion and Management 198, 111793
[5] X Lu et al. 2019 Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph Energy Conversion and Management 196 950-965
[6] Z Chen et al. 2018 Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents Energy Conversion and Management 178 250-264
[7] M Dhimish, V Holmes, B Mehrdadi and M Dales 2018 Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection Renewable Energy 117
[8] K Polat and S Güneş 2007 An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease Digital Signal Processing 17(4) 702-710
[9] M Kuai, G Cheng, Y Pang and Y Li 2018 Research of Planetary Gear Fault Diagnosis Based on Permutation Entropy of CEEMDAN and ANFIS Sensors (Basel) 18(3) Mar 5