

PAPER • OPEN ACCESS

Fault Diagnosis Method of Photovoltaic Array Based on BP Neural Network

To cite this article: Junjie Wang *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **646** 012020

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

Fault Diagnosis Method of Photovoltaic Array Based on BP Neural Network

Junjie Wang¹, Shan Wang¹, Haixiong Liu¹, Jianbo Hong² and Dedong Gao^{1,2,*}

¹School of Mechanical Engineering, Qinghai University, Xining 810016, China

²Haixi Huahui Chemical and Machinery Co.,LTD

*Corresponding author: gaodd@zju.edu.cn

Abstract. Photovoltaic arrays are prone to various failures due to long-term work. In order to quickly and accurately diagnose the type of failure of the PV array and implement online monitoring of the PV array, this paper proposes the BP neural network for PV array fault diagnosis, and proposes a network search method when training BP neural network. And the K-cross-validation method is used to select the number of hidden layer nodes. The BP neural network fault diagnosis model designed and trained by this method is proved to have high precision.

1.Introduction

Solar energy is a kind of carbon-free renewable energy, which has been used more and more widely in the field of photovoltaics. However, since photovoltaic cells are installed in outdoor environments and they are exposed to sunlight and wind and rain throughout life of PV array, so the PV array is prone to various failures. If a method can be found to monitor the working status of the PV system in real time and quickly find various faults of the PV system then it can reduce the cost of maintenance of the PV power plant and reduce the burden on workers[1-3].

At present, methods for fault diagnosis of photovoltaic systems mainly include traditional diagnostic methods, model algorithms and intelligent algorithms[4].The traditional fault diagnosis methods mainly include two kinds: online diagnosis and offline diagnosis. Traditional fault diagnosis methods usually require other additional hardware equipment to complete the accurate diagnosis for PV system, and the cost is high, but the traditional method can accurately locate the fault of the PV system[5-8].Online diagnosis methods mainly include infrared thermography method and multi-sensor data fusion method. Infrared thermography method detects the fault by using the temperature difference on the surface of the photovoltaic panel. And it is necessary to set up the infrared imager to judge the photovoltaic panel by processing the image returned by the infrared imager. The existing fault can accurately be detected, but the cost is high. The infrared thermography method requires the imager to have high temperature resolution accuracy, and it is only suitable for the detection of hot spot fault of the photovoltaic panel[9-11].The multi-sensor data fusion method achieves the purpose of fault detection by installing multiple voltage and current sensors at each node of the photovoltaic system. By collecting the data returned by multiple sensors and integrating the data, the fault can be monitored in real time in the photovoltaic system. The multi-sensor data fusion method can accurately diagnose the location of the fault in the PV system, but the cost is too high due to the need to install the sensor at each node of the PV system. The offline fault diagnosis method mainly includes the



ground capacitance method and the time domain reflection method[12].The ground capacitance method locates the fault by measuring the capacitance of each node of the series battery to ground. This method can accurately locate the fault, but the photovoltaic system need to be offline when the fault is detected. The time domain reflection method need to input an electric pulse to the photovoltaic system, determine the position of the fault by the waveform and delay time of the returned signal, use the time domain reflection method can accurately determine the fault location, but because it is analysed by the signal waveform, so the requirements for the instrument are higher, the cost is higher, and online fault diagnosis cannot be achieved[13-15].

The intelligent algorithm and the model algorithm do not need to install other additional equipment, and only need to obtain the photovoltaic system output electrical parameters and real-time environmental parameters. Model algorithm fault diagnosis must be based on accurate modelling[16].Model algorithm method requires accurate establishment of the PV system model, the diagnostic accuracy is high, but the modelling is very difficult. Intelligent algorithm fault diagnosis methods mainly include cluster analysis method and BP neural network method. Clustering analysis is an unsupervised learning algorithm. In the fault diagnosis, the unsupervised learning of the sample data is performed first, then the data is divided into K categories, and the classification results are interpreted based on the corresponding prior knowledge. To discover the failure of the photovoltaic system, this method requires a certain prior knowledge, but BP neural network is a supervised learning machine[17, 18].There is no need for prior knowledge of PV array fault diagnosis using BP neural network. It only needs to collect the daily operation data of PV array, and BP neural network can solve any complex nonlinear problem[19-21].

The BP neural network is proposed to diagnosis the PV array faults in this paper. The BP neural network is a deep learning algorithm. It is widely used as pattern recognition algorithm. It has been used in photovoltaic system fault diagnosis. When performing PV system fault diagnosis, the data is usually not properly processed, so the accuracy of algorithm classification has not been improved reliably. Therefore, this paper proposes data protocol for training data. At the same time, when the BP neural network is established, the network search method and the K-cross-validation method are proposed to optimize the number of hidden layer nodes, which solves the problem that the hidden layer cannot be selected when using BP neural network for fault diagnosis[22-24].

2.PV array failure analysis and feature parameter selection

2.1. Failure analysis

Photovoltaic arrays work in outdoor environments and are exposed to strong ultraviolet and sand environments for a long time. Due to the special working environment of photovoltaic panels, various faults occur in photovoltaic arrays. Photovoltaic array failures mainly include hot spots, shadows, aging, short circuits and open circuits.

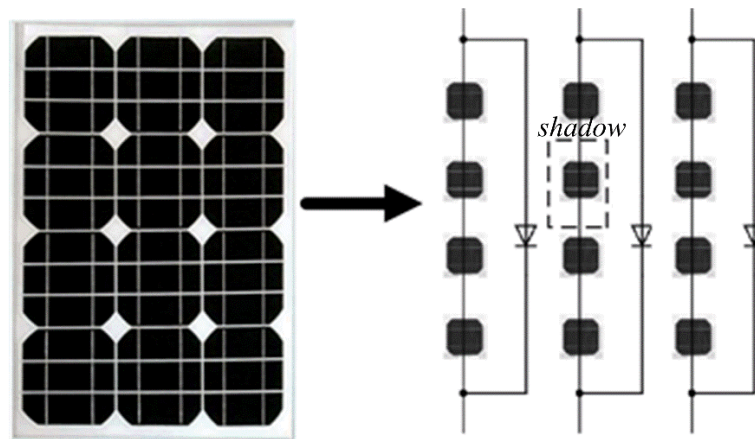
The hot spot fault is mainly caused by long-term shadow of a part of the photovoltaic panel. When a certain area of the panel is blocked, the battery cell in this area will be converted into a load and exist in the circuit. The battery unit will output a negative voltage. At the beginning, it consumes a certain amount of power and generates heat. The battery cells that have been in such a shade for a long time cause the battery cells to burn out. When the shading disappears, it cannot work normally again.

Aging is caused by the long-term operation of the panel and the erosion of wind and rain. The aging of the panel usually shows an increase in the series resistance of the panel. The output characteristic curve is usually similar to the shadow. The principle of shadow formation of the photovoltaic panel is shown in Figure 1. Since aging and hot spots are consistent on the output characteristic curve, aging, hot spots, and shadows are collectively classified as insufficient irradiation failure. Therefore, photovoltaic array failures include insufficient irradiation, open circuits, and short circuits.

The PV array failure analysis is shown in Table 1.

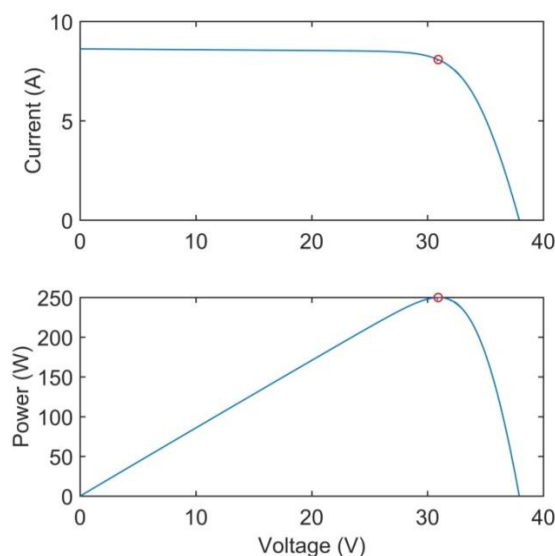
Table 1. Failure analysis.

Fault type	Cause of formation	Harm
Open circuit	Installation wiring error	Reduce power generation efficiency
Short circuit	Cable aging	Accelerate aging and cause fire
Insufficient irradiation	Aging, shadows and cracks, etc.	reduce power generation efficiency

**Figure 1.** Photovoltaic panel shadow failure diagram

2.2. Feature parameter selection

The photovoltaic array is composed of a plurality of photovoltaic panels. The output characteristic curve of the photovoltaic array directly reflects the working state of the photovoltaic array. The output characteristic curve of the photovoltaic array can be drawn by the short-circuit current, open circuit voltage, maximum power current and maximum power voltage of the photovoltaic cell, such as Figure 2 shows the output characteristics of the PV array. As can be seen from the figure, the PV array output characteristic curve is basically determined by the PV array short-circuit current, open circuit voltage, maximum power current and maximum power voltage. Therefore, this paper determines the most characteristic PV array fault characteristics using I_{sc} , U_{oc} , I_{mp} , and U_{mp} parameters.

**Figure 2.** PV array output characteristic curve

3. Modelling BP neural network

3.1. BP neural network

BP neural network is one of the intelligent algorithms widely used in the field of fault diagnosis, which can solve the problem of highly nonlinear fault pattern recognition. BP neural network has efficient self-learning, self-adaptive ability, and can complete contact memory. Since the input and output layer nodes of the BP neural network model can be determined according to the input parameters and the number of fault states, the BP neural network is suitable for solving complex fault diagnosis problems with multiple faults and multiple modes.

Since the PV array fault diagnosis problem is a highly nonlinear and complex fault diagnosis problem, there is a high degree of nonlinearity between the PV array output characteristic curve and various fault modes, and there is a high degree of coupling between various faults and output characteristic parameters. Sexuality, while BP neural network can complete the nonlinear mapping problem of input to output with arbitrary complexity, and has good generalization ability. Therefore, this paper proposes BP neural network for PV array fault diagnosis.

BP neural network is a multilayer feedforward error back propagation neural network system. BP neural network algorithm structure diagram shown in Figure 3. Its training learning process mainly includes two processes: the positive propagation process and the reverse propagation process. The positive propagation process is that the input data is directly transmitted to the output layer through the hidden layer. When the forward propagation is performed, the weights of the neurons in the neural network remain fixed. When there is an error between the output and the actual fault output, the network enters. In the anti-propagation process, the weight of the network during the anti-propagation process is adjusted according to the error so that the output is close to the desired output.

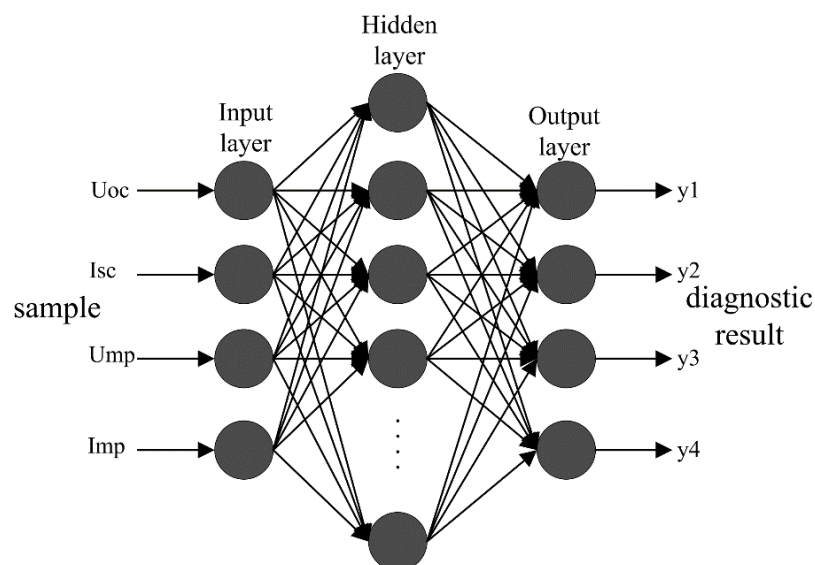


Figure 3. BP neural network structure

As can be seen from Figure 3, the BP neural network includes an input layer, an implicit layer, and an output layer. In the BP neural network model training, four problems need to be determined: the number of nodes of the input layer, the output layer and the hidden layer, and the hidden layer basis function. In actual use, the number of input layer nodes is determined according to the number of fault data characteristic parameters, and the number of fault data characteristic parameters is the number of input nodes. The number of output layer nodes is determined by the number of faults, and the number of faults is the number of output layer nodes.

There are three methods for determining the number of hidden layers, which are shown in Table 2 below. Since the method of determining the number of hidden layers is still not fully studied, there is

no most suitable method for determining the number of hidden layers. Therefore, this paper uses the network search method to determine the number of hidden layers. This method can find the optimal number of hidden layers. Before performing network search and optimization, we must determine the scope of optimization. Use the three methods in the table to calculate the number of hidden layer nodes, and take the maximum and minimum values as the scope of network. Searching for the range value, as can be seen from the above analysis, the number of input layer nodes is 4, and the number of output layer nodes is 4.

Table 2. Hidden layer node calculation function

Serial number	Implicit layer calculation formula	Implicit layer node value
1	$m = \sqrt{n+l} + \alpha$	m=3.8~12.8
2	$m = \log_2 n$	m=2
3	$m = \sqrt{nl}$	m=4

The hidden layer basis function uses a Gaussian function, which is a nonlinear function, which has the advantages of good convergence and fast calculation speed. The Gaussian function is shown in equation (1).

$$F(\sigma) = \exp\left(\frac{\|x - c\|^2}{-2\sigma^2}\right) \quad (1)$$

Where c is the initial training value and σ is the width of the Gaussian function.

By using a Gaussian function to transform the input into the hidden layer, it is known that the hidden layer value is as shown in equation (2).

$$s_i = \exp\left(\frac{\|x_i - c_i\|^2}{-2\sigma_i^2}\right) \quad (2)$$

Where s is the calculated value of the hidden layer.

Since there is no connection between the output and the input, and there is a linear relationship between the output and the hidden layer, the output calculation formula is as shown in equation (3).

$$y_i = \sum_{i=1}^4 \omega_i s_i \quad (3)$$

3.2. Hidden layer node number optimization

The number of hidden layer nodes is optimized by network search and K-cross-validation. The minimum range of network search is 2, the maximum value is 13, and the search step is 1, the number of nodes is increased each time. The number of nodes obtained by the search is verified using the K-cross validation method.

The K-cross-validation process is to mix the collected fault sample data into the same sample training set, and then divide the sample set into K-classes, and use one set of data each time the number of searched nodes is verified. As a test set, the K-1 group is left as a training set.

The fault diagnosis accuracy of BP neural network under the number of hidden layer nodes is shown in Figure 4. The figure shows that the model has the best effect when the number of hidden layer nodes is $m=5$, and the accuracy rate is 97 %.

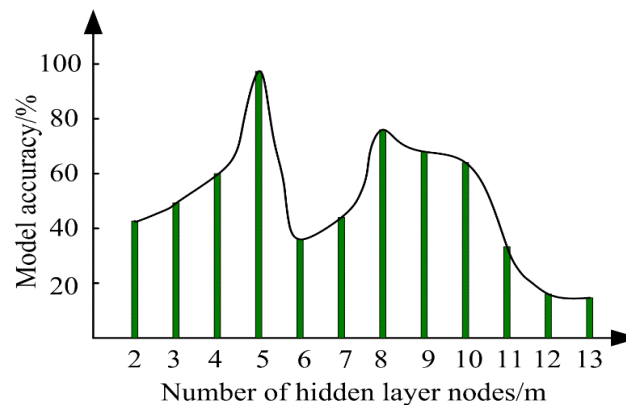


Figure 4. Implicit layer node optimization

4. Experiment analysis

4.1. Data collection

Collecting data is the first step in big data analysis. The use of BP neural network based PV array fault diagnosis requires a large amount of high quality training data. Therefore, a 3×4 photovoltaic array was built, as shown in Figure 5. The 3×4 photovoltaic array includes a total of 12 battery boards, and the type of the battery board is STS-156P-255W. Each battery board is composed of 60 battery cells. The basic parameters of the battery board are: $I_{sc}=8.95A$, $U_{oc}=37.8V$, $I_{mp}=8.32A$, $U_{mp}=30.65V$, $P_m=255W$.



Figure 5. 3×4 photovoltaic array

There are three kinds of fault data that need to be acquired in this experiment: insufficient irradiation, open circuit and short circuit. A total of four data are collected for data including the normal state of the PV array. In order to ensure the reliability of the experiment and collect high-quality data, this experiment must ensure that the irradiance is $1000W/m^2$ at the time of collecting the normal working data of the PV array. This experiment uses the PV array IV curve tester produced by Kewell Company for data acquisition. The Cowell IV curve tester model is IVT-30-1000, and the maximum current that can be tested is 30A. The maximum voltage that can be tested. It is 1000V.

In order to collect the PV array state data under the condition of insufficient irradiation, it is necessary to ensure that the irradiance is less than $1000 W/m^2$ at the time of collecting data, and at the same time, in order to ensure the uniformity between the experimental collection data and the working state of the official photovoltaic array, the experiment is being carried out. The PV array operating status data under different temperature and humidity conditions were collected.

In the open circuit data acquisition, since the state data obtained when the circuit is open at any different position is similar, the data acquisition of the PV array operating state is performed in three open modes, as shown in Figure 6 below.

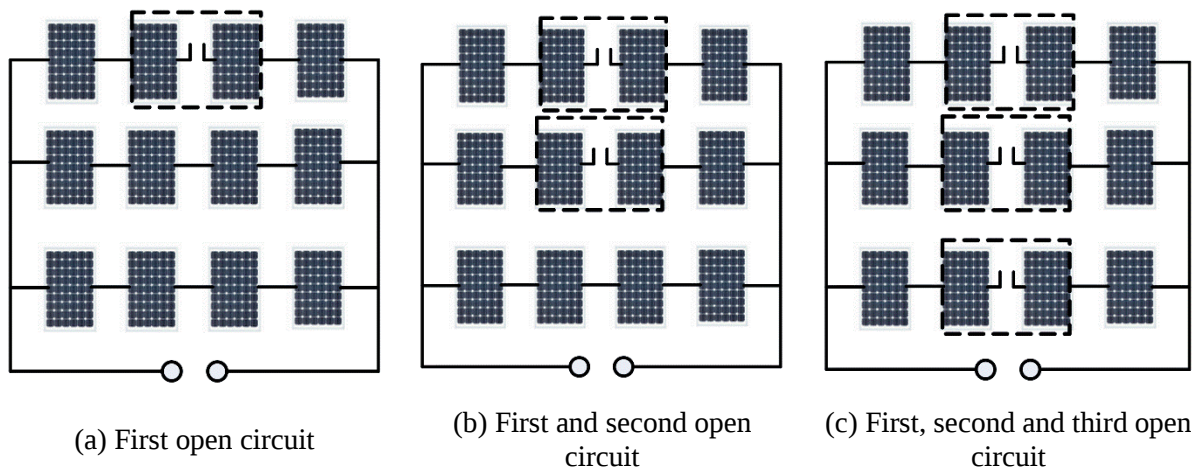


Figure 6. Open circuit experiment

In order to ensure that the battery board is not burned out during the short-circuit data acquisition, in the process of simulating the short circuit, at least one string of batteries will always be guaranteed to be free of short-circuit conditions. At the same time, it is necessary to ensure that a whole series of battery boards cannot be completely ensured. The true and random nature must ensure that the illumination is not less than 1000W/m², and the temperature and humidity are in various different situations. The process of short-circuit test is shown in Figure 7.

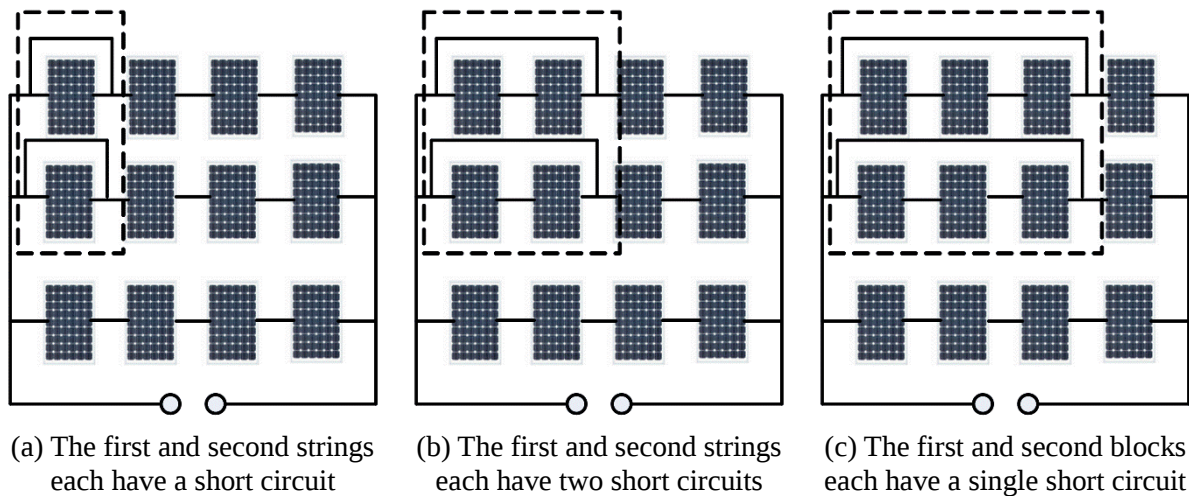


Figure 7. Short circuit experiment

4.2. Data normalization

Due to the limitations of the BP neural network model, the input data must be a unit value between [0, 1]. Therefore, the fault sample data obtained by the above experimental methods cannot be directly input into the algorithm for training, and must be normalized by data. The main calculation method of data normalization is shown in equation (4).

$$p = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

Where P is the normalized value of the to-be-stated value X.

4.3. Training and learning

In this paper, based on SPSS data analysis software, the neural network model is trained, and the preprocessed data is input into the BP neural network model. From the above, the number of hidden nodes in the neural network is 5, and the implicitness of BP neural network is determined. The number of layers is 1, and the number of input and output nodes is 4.

A total of 400 data were collected in this experiment, including 100 normal, insufficient, open and short circuit data. The data set was divided into two groups before the model training, 70% of which was used as the training set, 30%. Proportional data as a test set. The BP layer neural network activation function is a hyperbolic tangent function, and the output layer activation function is a softmax function. The maximum number of training times is set to 1000 times, the training step size is set to 0.05, and the training error is not more than 0.001. Figure 8 shows the error curve of the training data during the training process. It can be seen from the figure that the training error of the BP neural network will decrease with the increase of the training times. When the 420 trainings, the model error reaches 10^{-3} . The magnitude, at which point the desired target error is reached.

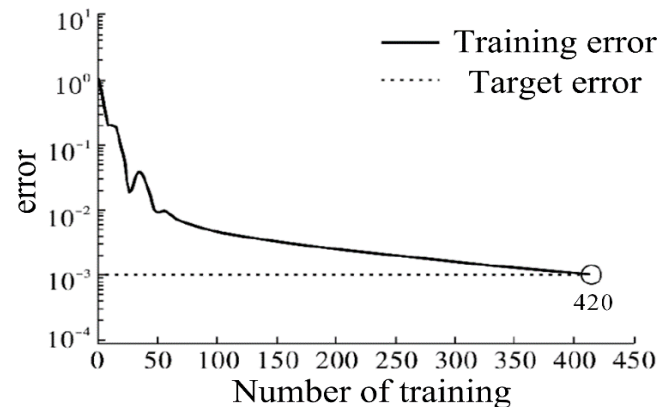


Figure 8. Training error

4.4. Test verification

In order to test the accuracy of the BP neural network fault diagnosis model after training, the remaining 120 data for the non-participating model training will be tested as a test set. The test results are shown in Table 3. It can be seen from the table that the normal data test accuracy is 80%, the insufficient irradiation data test accuracy is 91%, the open circuit data test accuracy is 96%, and the short circuit data test accuracy is 87%, where in the open circuit data test accuracy is the highest, and the model has the resolution capability. It is best, and the short-circuit data is more difficult to distinguish, and it is easy to blur the data with insufficient irradiation, but the test effect is still good.

Table 3. Test Results

label	0	1	2	3	Precision
0	27	0	3	0	80%
1	1	28	0	1	91%
2	0	0	29	1	96%
3	0	3	1	26	87%

5. Conclusions

In this paper, a fault diagnosis method for PV array based on BP neural network is proposed. After analysis, the open circuit voltage, short circuit current, maximum power voltage and maximum power current of the PV array are determined as input parameters of BP neural network. These four

parameters can be very good to reflect the working state of the PV array. The network search method and the K-cross-validation method are proposed to optimize the number of BP hidden nodes in the network. It is found that the model accuracy can reach 97% when the number of nodes in the hidden layer is 5. The network search method and the K-cross-validation method are combined, and the generalization of the BP neural network can be improved by using K-cross-validation method. However, due to the complexity of the working environment of the PV array, the complex and variability of the PV array output characteristic curve information seriously affects the accuracy of the PV array intelligent diagnosis algorithm. Therefore, the intelligent fault diagnosis algorithm based on BP neural network needs to be improved.

Acknowledgments

This research work is sponsored by the National Natural Science Foundation of China(No. 51566015), Qinghai Science and Technology Projects(No. 2019-ZJ-7002 and 2019-HZ-822).

References

- [1] Huang Zhiqiang and Guo Li. 2009 Research and implementation of microcomputer online fault detection of solar array *Computer Science & Education, 2009. ICCSE '09. 4th International Conference on*
- [2] Huang Z and Li G. 2009 Research and implementaion of microcomputer online fault detection of solar array *2009 4th International Conference on Computer Science & Education(ICCSE'09)* 1052-55
- [3] Cheng Z, Zhong D and Li B. 2011 Research on fault detection of PV array based on data fusion and fuzzy mathematics *2011 Asia-Pacific Power and Energy Engineering Conference (APPEEC). IEEE* 1-4
- [4] Cheng Z, Zhong D and Li B. 2011 Research on fault detection of PV array based on data fusion and fuzzy mathematics *2011 Asia-Pacific Power and Energy Engineering Conference(APPEEC)* 2011 1-4
- [5] Xiaoli Xu, Huan Wang and Yunbo Zuo. 2011 Method for Diagnosing Photovoltaic Array Fault in Solar Photovoltaic System *Power and Energy Engineering Conference (APPEEC), 2011 Asia-Pacific*
- [6] Guinjoan-Gispert F Velasco-Quesada G, Pique-Lopez R,et al. 2009 Electrical PV array reconfiguration strategy for energy extraction improvement in grid-connected PV systems *IEEE Transactions on Industrial Electronics* 2009,**56** 4319-31
- [7] Takumi Takashima, Junji Yamaguchi, Kenji Otani, Takashi Oozeki, Kazuhiko Kato and Masayoshi Ishida. 2008 Experimental studies of fault location in PV module strings *Solar Energy Materials and Solar Cells* **93**
- [8] Chowdhury S R and Saha H. 2010 Maximum power point tracking of partially shaded solar phtotvoltaic arrays *Solar Energy Materials&Solar Cells* **94** 1441-47
- [9] Agarwal V Patel H. 2008 MATLAB-based modeling to study the effects of partial shading on PV array characteristics *IEEE Transactions on Industrial Electronics* **23** 302-310
- [10] Ducange P, Fazzolari M and Lazzerini B. 2011 An intelligent system for detecting faults in photovoltaic fields *IEEE11th International Conference on Intelligent Systems Design and Applications. Cordoba, Argentina: IEEE* 1341-46
- [11] Togashi S Noguchi T. 2002 Short-current pulse-based maximum-power-point tracking method for multiple phtotvoltaic-and-converter module system *IEEE Transactions on Industrial Electronics* **49** 217-23
- [12] Noguchi, T., Togashi, S., Nakamoto and R. 2000 Short-current pulse based adaptive maximum-power-point tracking for photovoltaic power generation system *Industrial Electronics, 2000. ISIE 2000. Proceedings of the 2000 IEEE International Symposium on*
- [13] Lehman B Nguyen D. 2008 An adaptive solar photovoltaic array using model-based reconfiguration algorithm *IEEE Transactions on Industrial Electronics* **55** 2644-54

- [14] Inada T Miyatake M, Hiratsuka I, et al. 2004 Control characteristics of a fibonacci-search-based maximum power point tracker when a photovoltaic array is partially shaded *The 4th International Power Electronics and Motion Control Conference 2004(IPEM2004)* 2004 816-21
- [15] Yanli Liu, Bingfeng Li and Dan Zhong. 2010 Research on domestic PV module structure based on fault detection *Intelligent Control and Automation (WCICA), 2010 8th World Congress on*
- [16] Stellbogen D. 1993 Use of PV circuit simulation for fault detection in PV array fields *Conference Record of the Twenty Third IEEE on Photovoltaic Specialists Conference* 1993 1302-07
- [17] Lin C, Jie G and Wu Z. 2011 Design of networked monitoring system of PV grid-connected power plant *2011 International Conference on Electronic and Mechanical Engineering and Information Technology. IEEE*
- [18] Nian B, Fu Z and Wang L. 2010 Automatic detection of defects in solar modules: image processing in detecting *6th International Conference on Wireless Communications Networking and Mobile Computing. Chengdu, China: IEEE* 1-4
- [19] King D L, Kratochvil J A and Quintana M A. 2000 Applications for infrared imaging equipment in photovoltaic cell, module, and system testing *Conference Record of the Twenty-Eighth Photovoltaic Specialistis Conference* **456** 1487-90
- [20] Patel H and Agarwal V. 2008 MATLAB—based modeling to study the effects of partial shading on PV array characteristics *IEEE Transactions on Energy Conversion* **23** 302-10
- [21] Chao K H, Ho S H and Wang M H. 2008 Modeling and fault diagnosis of a photovoltaic system *Electric Power Systems Research* **78** 97-105
- [22] Boztepe M Karatepe E, Colak M. 2007 Development of a suitable model for characterizing photovoltaic arrays with shaded solar cells *Solar Energy* **81** 977-992
- [23] Nalin K, Gautam N D and Kaushika. 2002 Reliability evaluation of solar photovoltaic arrays *Solar Energy* **72**
- [24] Shigeo H, Toshiali Y and Tatsuo T. 2010 A New Stand-alone Hybrid Power System with wind Generator and photovoltaic Modules for Radio base station photovoltaic Energy Conversion *The IEEE International Conference* **2** 155-58