



10th International Conference on Applied Energy (ICAE2018), 22-25 August 2018, Hong Kong, China

An intelligent fault diagnosis method for PV arrays based on an improved rotation forest algorithm

Fuchang Han^a, Zhicong Chen^{a,*}, Lijun Wu^a, Chao Long^b, Jinling Yu^a, Peijie Lin^a, Shuying Cheng^a

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

^bInstitute of Energy, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

Abstract

With the exponential growth of global photovoltaic (PV) power capacity, it is essential to monitor, detect and diagnose the faults in PV arrays for optimal operation. This paper presents an improved rotation forest (RoF) algorithm classifiers ensemble hybridized with extreme learning machine (ELM) for fault diagnosis of PV arrays, which mainly consists of feature selection and classification. In the feature selection step, all the attributes are ranked by the ReliefF algorithm and the top-ranked attributes are chosen to create the new training data subset. In the classification step, the base classifier decision tree of the RoF is replaced by the extreme learning machine to form a new hybrid RoF-ELM ensemble classifier. In the RoF-ELM algorithm, the feature space is first split into several subspaces and the best number of feature subsets is found through the traversal search method. Then, the bootstrap algorithm is employed to carry out bootstrap resampling for each feature subspace, and the principal component analysis (PCA) is then used to transform the resampled samples. Finally, the ELM base classifier is exploited to build each classification model and the final decision is determined by the simple voting approach. By combining the RoF ensemble method with the ELM classifier, the proposed RoF-ELM algorithm not only overcomes the overfitting problem of the basic RoF algorithm, but also improves the generalization ability of the basic ELM. In order to experimentally verify the proposed approach, different types and levels of faults have been created in a laboratory small scale grid-connected PV power system to obtain the fault data samples. Experimental results demonstrate that the RoF-ELM can achieve higher diagnosis accuracy and reliability compared to the basic RoF and ELM algorithms.

© 2019 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of ICAE2018 – The 10th International Conference on Applied Energy.

* Corresponding author. Tel.: +86-158-0606-5152; fax: +86-0591-2286-6342.

E-mail address: zhicong.chen@fzu.edu.cn

Keywords: PV arrays; Fault diagnosis; ReliefF algorithm; Rotation Forest; Extreme learning machine

1. Introduction

As an alternative energy source, the solar renewable clean energy has received much attention in recent years [1]. According to the latest announcement of the International Renewable Energy Agency (IRENA), the global installed capacity of PV plants in the world has reached 390 GW by the end of 2017. However, PV power plants are subject to numerous kinds of faults mainly due to the outdoor harsh operating environment, which may cause an obvious energy loss as well as potential safety risk [2]. Therefore, real time monitoring based fault diagnosis is necessary to improve their productivity, reliability and safety. Many different methods have been developed to detect and classify the faults in PV systems. Among these methods, model-based approaches and intelligence-based algorithms have gained more and more attention recently [3].

Model-based approaches generally identify the faults based on the performance comparison between the predicted values from PV-array equivalent circuit models and the measured ones obtained from real time monitoring data. For example, Platon et al. presented a fault diagnosis algorithm based on the power loss between the model prediction and the measured results. However, multiple models for different irradiance ranges are developed for better performance [4]. Lin et al. proposed a reconfiguration technique based on power loss to identify the faults. However, it requires a large number of switches in a large PV system [5]. According to the previous literature, these model-based methods may not be effective to detect faults and their weakness has been found in paper [1]. Intelligence-based algorithms can realize the fault diagnosis using computational intelligence and machine learning techniques, which depends on the availability of labeled fault data samples. Hazra et al. employed three meta-heuristic optimization techniques to develop a novel PV fault diagnosis method [6]. Yi et al. proposed a fault detection approach based on multi-resolution signal decomposition (MSD) and fuzzy inference systems (FIS) [7]. In recent years, artificial neural networks (ANN), decision tree (DT), support vector machine (SVM), kernel based extreme learning machine (KELM) are commonly used classification algorithms for fault diagnosis of PV array [8]. Especially, the relatively new Rotation Forest (RoF) ensemble approach based on DT, has received much attention due to improved classification results, which was proved to have better accuracy than other ensemble methods (such as Bagging, AdaBoost, Random Forest) on some classification tasks [9].

In this paper, we propose an improved RoF algorithm based on ELM classifiers ensemble (RoF-ELM), which combines the advantages of RoF algorithm and ELM neural network. Firstly, In order to eliminate irrelevant and redundant attributes, all the attributes are ranked by using ReliefF algorithm and the top-ranked attributes are chosen to formulate new training dataset. And then, the proposed RoF-ELM is used to detect and classify the common faults. In order to verify the proposed method, several fault experiments were carried out on a laboratory small scale grid-connected PV power system to obtain experimental fault samples.

2. Fault analysis and features selection

2.1. Typical faults in PV systems

The PV array is subject to numerous kinds of faults, and some typical faults are considered in this paper to carry out fault experiments. The line-line (LL) faults refer to an unexpected short-circuit connection between two points of different potentials in PV array, which often occurs in many large-scale PV systems because PV-string cables are usually grouped together and passed into the same raceway [10]. There are two main fault scenarios: intra-string line-line (I-LL) fault and cross-string line-line (C-LL) fault. I-LL fault indicates a short circuit that occurs within the same PV string, while C-LL fault represents short-circuit connection on two different strings. Partial shadow (Shadow) faults are mainly caused since some of the PV cells/modules are blocked from the sunlight, which is dynamic and temporary in most cases. But if the shadow does not be cleared for a long time, those blocked PV cells/modules will become the load in the circuit and consume the energy generated by other PV modules, and even form hot spots. Open-circuit (OPEN) faults refer to the accidental disconnection in PV cells/modules/strings due to harsh environment condition or unexpected hardware damage, which reduces the output power of PV array

obviously because the open-circuit faults are essentially a reduction in the number of parallel PV strings. The degradation faults mean the gradual deterioration of PV component characteristics, which usually causes changes in parasitic resistances of PV modules, including the series resistance (R_s) and the shunt resistance (R_{sh}). In an aging module, R_s usually becomes larger while R_{sh} is relatively smaller. Other faults in PV systems, such as the ground faults, are not considered in this paper due to the simplicity of being cleared by existing protection devices.

2.2. ReliefF algorithm for features selection

The classification tasks in machine learning are binary, multi-class, multi-labeled and hierarchical. The presence of irrelevant or redundant features in high dimensional datasets possibly lowers the performance of machine learning algorithms [11]. Hence, the suitable feature selection becomes important for classification tasks. There are several measures for estimating attributes' quality, such as distance measure, information gain, Gini index, Relief algorithms (include Relief, ReliefF and RReliefF), minimum description length (MDL) and so on. Among these methods, Relief algorithms do not assume the conditional independence of the attributes, and hence are more appropriate in problems which possibly involve much feature interaction.

Table 1. Original and new attributes of PV faults.

Attributes	Unit	Description
I_1, I_2, I_3	A	Each PV-string current at MPP condition
V_{PV-mpp}	V	The voltage of PV array at MPP condition
V_{oc-Ref}	V	The open-circuit voltage of reference module
I_{sc-Ref}	A	The short-circuit current of reference module
I_{PV-mpp}	A	The current of PV array at MPP condition
P_{PV-mpp}	W	The power of PV array at MPP condition
I_{norm}	0 ~ 1	$I_{PV-mpp}/(3 * I_{sc-Ref})$
V_{norm}	0 ~ 1	$V_{PV-mpp}/(6 * V_{oc-Ref})$
FillFactor	0 ~ 1	$(V_{PV-mpp} * I_{PV-mpp}) / (18 * V_{oc-Ref} * I_{sc-Ref})$
C	Current discrete ratio	Coefficient of variance of all PV strings
S_1	Incremental derivative ratio	$(I_{PV-mpp} - 3 * I_{sc-Ref}) / (V_{PV-mpp} - 6 * V_{oc-Ref})$
S_2	Incremental derivative ratio	$(0 - I_{PV-mpp}) / (6 * V_{oc-Ref} - V_{PV-mpp})$
S_3	Incremental derivative ratio	$(I_{PV-mpp} - 3 * I_{sc-Ref}) / V_{PV-mpp}$

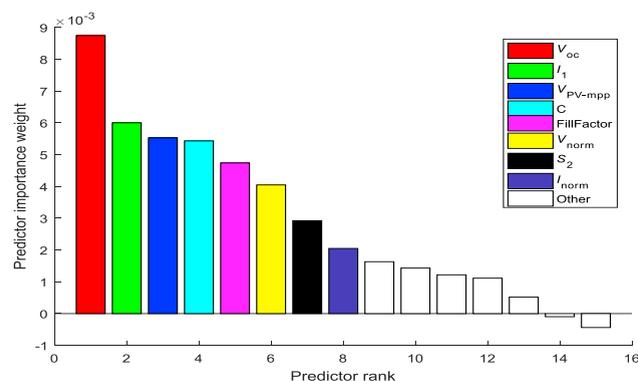


Fig. 1. All attributes ranked by the ReliefF algorithm

The core idea of Relief algorithm is to calculate the importance of every attribute according relief criterion, then select and rank the best features to constitute new training subset. In this study, the ReliefF algorithm is used to

select the good fault attributes. The original and some new attributes are summarized in Table 1. Note that the PV array consists of three parallel PV strings and each PV string is configured as six PV modules in series, which operates in maximum power point (MPP) condition, as detailed in 4.1. These attributes are ranked by using ReliefF algorithm, and then some top-ranked attributes are selected to develop the diagnosis model, as shown in Fig. 1. Eight attributes with high weight values ranking are chosen to build a new training subset from original training: V_{oc} , I_1 , V_{pv-mpp} , C , $FillFactor$, V_{norm} , S_2 , I_{norm} . Note that remaining attributes are ruled out, since they have high inter-correlation but low correlation with the class-label attributes.

3. The improved rotation forest for multiclass classification

3.1. Rotation forest

In recent years, the ensemble classifier techniques are rapidly growing due to their potential to greatly increase prediction accuracy of classification results. The base learning algorithms broadly used in building ensemble classifiers are neural networks and DT [12]. Although DT has very fast training speed, its training process is very unstable. To improve generalization property of DT classifiers, some ensemble classifiers (such as Bagging, Random Forest, Rotation Forest) have been proposed. These approaches try to build improved classifiers by the combination of weak and diversified classifiers. Note that RoF algorithm performs much better than other ensemble methods [9]. The main idea of RoF is to increase accuracy and diversity of base classifiers. A brief description of the algorithm is as follows: (1) the feature set is randomly split into K subsets and the diversity is promoted by using principal component analysis (PCA) for each of the subset. Hence, new attributes of base classifier are acquired with K axis rotations. (2) the unknown instances are classified and the final classification result is assigned to the winner class with the highest confidence value. In view of the advantages of the RoF algorithm in many other applications, the RoF algorithm is further explored for fault diagnosis of PV arrays in this paper.

3.2. The proposed RoF-ELM based fault diagnosis method

Each classifier in basic RoF is independently built using the DT method, and each DT is trained on the training dataset in a rotated feature space. However, the DT in RoF will lead to the overfitting and local optimum problem, which usually affects ensemble performance of RoF [13]. Therefore, it is important to select an optimal basic classifier to overcome above disadvantages. It is worth mentioning that recently proposed extreme learning machine (ELM) algorithm can resolve over fitting problem due to adaptation. The unique feature of ELM is to randomly assign the input weights and biases for hidden layer nodes, and the input weights and biases are never adjusted during the whole learning process. In addition, ELM has faster computation speed and higher generalization performance compared with traditional learning algorithms, which can handle multi-class classification problem through adjusting the number of output nodes [8]. Nevertheless, ELM also has some disadvantages. For instance, ELM lacks distributed implementation methods and the randomly assigned input weights and biases will cause unstable output results. Researchers found that the ensemble strategy is a direct way to solve the aforementioned shortcoming [14].

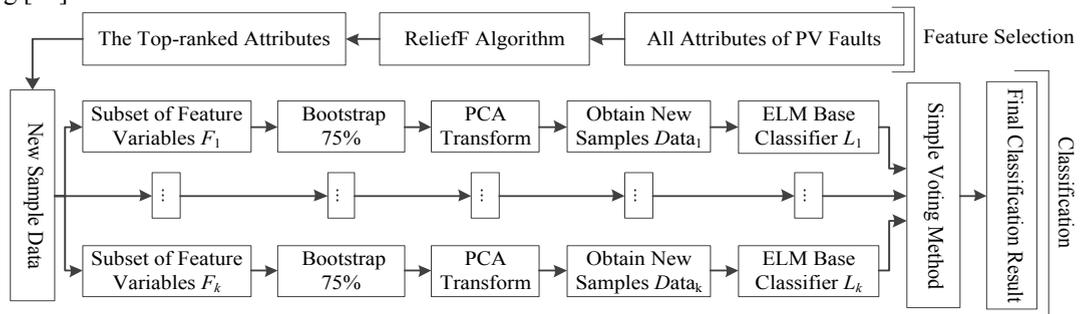


Fig. 2. Flow chart of the proposed RoF-ELM based fault diagnosis method.

In this paper, a new hybrid machine learning algorithm RoF-ELM is proposed for fault diagnosis of PV array, which combines the advantages of the basic RoF algorithm and ELM neural network. The proposed RoF-ELM algorithm not only overcomes the overfitting problem of RoF algorithm, but also improves the generalization capability of ELM neural network simultaneously. The proposed RoF-ELM based fault diagnosis for PV arrays mainly consists of two steps: feature selection and classification, as illustrated in Fig. 2. In the feature selection step, all attributes in table 1 are ranked using ReliefF algorithm and the top-ranked attributes are chosen to constitute new training subset. In the classification step, the best number of feature subsets (K) is found through the traversal search for a better performance, and then the RoF-ELM splits the feature space into K subspaces. After that, the bootstrap algorithm is employed to carry out data resampling (75% in this work) for each feature subspace. Then, the PCA is used to transform resampled data to further improve the quality of the data samples, after which the ELM base classifier is used to make the classification and prediction for each subset. Lastly, the final decision of the classification is determined by the simple voting method. After building the fault diagnosis model, the feature vectors of unknown (unused) data samples will be sending to the RoF-ELM ensemble classifier to test the model and the classification accuracy is calculated by the classification results of the each ELM base classifier.

4. Experimental setup and results

4.1. Setup of typical faults

A 2kW grid-connected PV system is used to verify the RoF-ELM algorithm. As shown in Fig. 3(a), the PV array consists of three parallel PV strings, i.e., String 1, String 2 and String 3, and each PV string is configured as six PV modules in series. Two PV modules on the left are used as the reference modules in this work, which provides real-time open-circuit voltage reference (V_{oc-Ref}) and short-circuit current reference (I_{sc-Ref}). The PV array operates in MPP condition under the inverter with MPPT algorithm. The PV-array voltage and each PV-string current are collected and stored continuously through a high-speed data acquisition card (DAQ), as shown in Fig. 3(b).

To increase the contamination range of above-mentioned typical faults, each fault type involves different mismatch levels. The setup schematic diagram of various faults is shown in Fig. 4(a). The I-LL faults are created by making short-circuit connection in one PV string, including the fault scenario with 1 module shorted (I-LL1) and the fault scenario with 2 modules shorted (I-LL2). The C-LL faults are created by making short-circuit connection in two different PV strings, including the fault scenario with 1 module shorted (C-LL1) and the fault scenario with 2 modules shorted (C-LL2). The degradation faults are emulated by inserting a series resistor of 4 ohms into PV circuit to simulate the larger parasitic series resistance (R_s), including degradation scenario at PV-string level (D-S) and degradation scenario at PV-array level (D-A). The Shade faults are simulated by covering some PV modules with shield panels (the translucent acrylic board in this work), including the fault scenario with 1 module blocked (Shade1) and the fault scenario with 2 modules blocked (Shade2). The OPEN faults are simulated by disconnecting the electrical connection wirings among the PV modules in one PV string. Therefore, 10 kinds of operating conditions including different types and levels of faults (I-LL1, I-LL2, C-LL1, C-LL2, D-S, D-A, Shade1, Shade2, OPEN) and the ones of normal case (NORMAL), are created to verify the proposed RoF-ELM method.

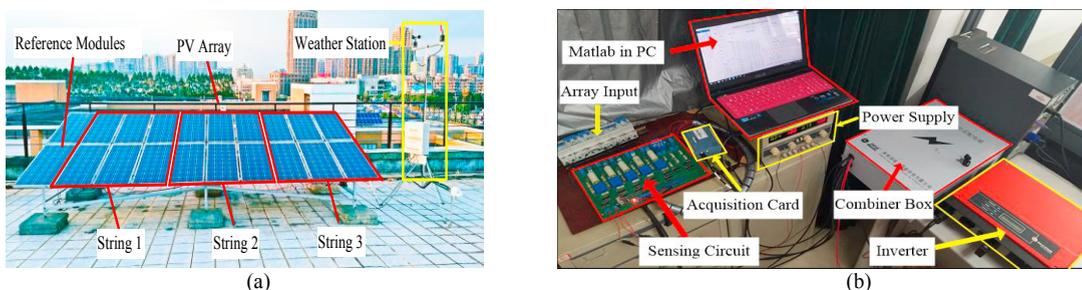


Fig. 3. (a) Illustration of the laboratory PV array; (b) Illustration of the real time monitoring system

4.2. Experimental results

The experiments are carried out on sunny day and each acquisition lasts for 180 minutes under changing irradiance. In this paper, a finite impulse response (FIR) low-pass filter is used to smooth the measurement for reducing noise interference and a down-sample procedure is used to achieve equal interval extraction for a low computational burden. Therefore 540 experimental data samples for each fault are acquired, and there are totally 5400 experimental data samples including the ones of normal case. Note that two parameters are significant for the success of RoF algorithm: the number of feature subsets (K) and the number base classifier in the ensemble (L). Besides, more base classifiers tend to produce the more accurate classification results, L = 6 is considered as the parameter value in this study. To find the best value of K parameter, the K parameter varies from 2 to 8 at the step of 1, and the corresponding average result of classification accuracy in 20 independent runs is shown in Fig. 4(b). We observe that the K=8 obtains the best classification results and the K parameter value is K=8.

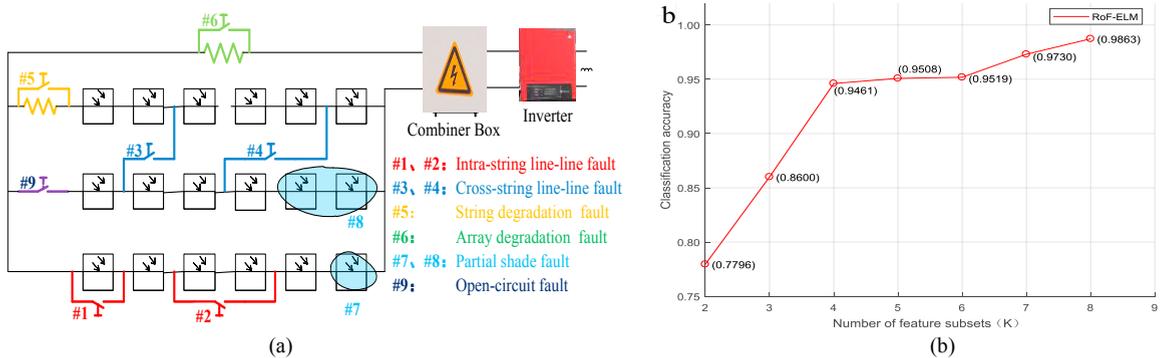


Fig. 4. (a) The setup schematic diagram of typical faults; (b) The classification accuracy of the RoF-ELM with different K values

To train and test the RoF-ELM algorithm, randomly selected 75% of the total sample data are used for training, whereas the remaining 25% samples are used for testing. To make the result more stable and reliable, the RoF-ELM algorithm is trained and tested for 20 independent times in this work, and the average results of 20 times for each operating condition are shown in Fig. 5(a). Experimental results show that the total classification accuracy of proposed RoF-ELM is 98.63%. In addition, both the basic ELM and basic RoF are similarly trained and tested for 20 independent times to further verify proposed RoF-ELM algorithm. The average results of 20 times for each operating condition are shown in Fig. 5(b). Experimental results demonstrate that the total classification accuracy of basic ELM is 98.07%, and total classification accuracy of basic RoF is 78.87%. It is obvious that RoF-ELM is more stable than both ELM neural network and RoF algorithm. Therefore, proposed RoF-ELM achieves a high diagnostic accuracy and reliability for PV array.

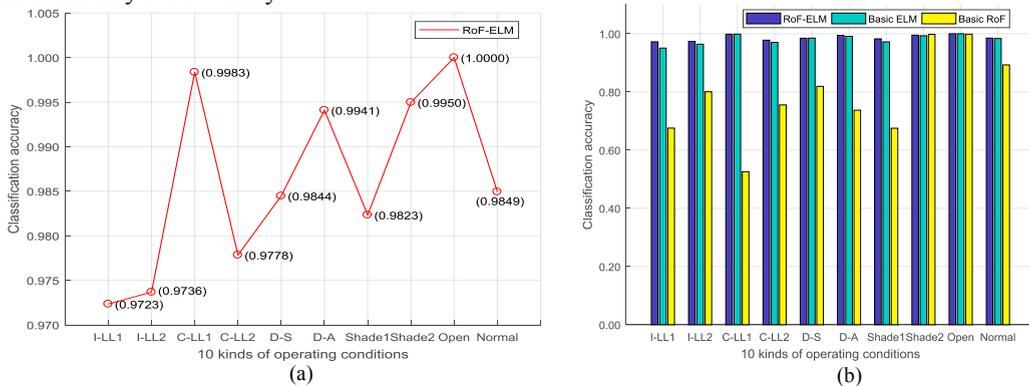


Fig. 5. (a) The classification accuracy of RoF-ELM for each operating condition; (b) Comparison among different algorithms

5. Conclusions

In this paper, a new hybrid algorithm RoF-ELM is proposed for fault diagnosis of PV arrays, which combines the advantages of the basic RoF algorithm and ELM classifier. The proposed fault diagnosis method mainly consists of two steps: feature selection and classification. In the feature selection step, the ReliefF algorithm is employed to select top-ranked attributes for creating the training dataset from the raw monitored data. In the classification step, the DT of the RoF algorithm is replaced with the ELM to overcome overfitting problem. . In order to verify the proposed method, different types and levels of faults experiments were carried out on a laboratory small scale grid-connected PV power system to obtain experimental fault data. Experimental results demonstrate that the total classification accuracy of optimal RoF-ELM is up to 98.63%, while the ones of the basic ELM and basic RoF are 98.07% and 78.87% respectively. Therefore, the proposed RoF-ELM based fault diagnosis method features superior performance and is promising for real applications.

Acknowledgements

The authors would like to acknowledge the financial supports by the National Natural Science Foundation of China (Grant Nos. 61601127, 51508105, and 61574038), the Fujian Provincial Department of Science and Technology of China (Grant No. 2016H6012, 2016H0016, and 2018J0106), the Fujian Provincial Economic and Information Technology Commission of China (Grant No. 830020 and 83016006), and the Science Foundation of Fujian Education Department of China (Grant No. JAT160073).

References

- [1] Kumar B P, Ilango G S, Reddy M J B, et al. Online fault detection and diagnosis in photovoltaic systems using wavelet packets[J]. *IEEE Journal of Photovoltaics*, 2018, 8(1): 257-265.
- [2] Chen Z, Wu L, Lin P, et al. Parameters identification of photovoltaic models using hybrid adaptive Nelder-Mead simplex algorithm based on eagle strategy[J]. *Applied Energy*, 2016, 182: 47-57.
- [3] Zhao Q, Shao S, Lu L, et al. A new PV array fault diagnosis method using fuzzy c-mean clustering and fuzzy membership algorithm[J]. *Energies*, 2018, 11(1): 238.
- [4] Platon R, Martel J, Woodruff N, et al. Online fault detection in PV systems[J]. *IEEE Transactions on Sustainable Energy*, 2015, 6(4): 1200-1207.
- [5] Lin X, Wang Y, Pedram M, et al. Designing fault-tolerant photovoltaic systems[J]. *IEEE Design & Test*, 2014, 31(3): 76-84.
- [6] Hazra A, Das S, Basu M. An efficient fault diagnosis method for PV systems following string current[J]. *Journal of Cleaner Production*, 2017, 154: 220-232.
- [7] Yi Z, Etemadi A H. Fault detection for photovoltaic systems based on multi-resolution signal decomposition and fuzzy inference systems[J]. *IEEE Transactions on Smart Grid*, 2017, 8(3): 1274-1283.
- [8] Chen Z, Wu L, Cheng S, et al. Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and IV characteristics[J]. *Applied Energy*, 2017, 204: 912-931.
- [9] Zhang C X, Zhang J S. RotBoost: A technique for combining Rotation Forest and AdaBoost[J]. *Pattern recognition letters*, 2008, 29(10): 1524-1536.
- [10] Dhoke A, Sharma R, Saha T K. PV module degradation analysis and impact on settings of overcurrent protection devices[J]. *Solar Energy*, 2018, 160: 360-367.
- [11] Chaudhary A, Kolhe S, Kamal R. An improved random forest classifier for multi-class classification[J]. *Information Processing in Agriculture*, 2016, 3(4): 215-222.
- [12] Zhang C X, Zhang J S, Wang G W. An empirical study of using Rotation Forest to improve regressors[J]. *Applied Mathematics and Computation*, 2008, 195(2): 618-629.
- [13] Chen T. An improved rotation forest algorithm based on heterogeneous classifiers ensemble for classifying gene expression profile[J]. *Advances in Modelling and Analysis B*, 2017, 60(1): 1-24.
- [14] Zhou Z, Song Y, Zhu Z, et al. Scene categorization based on compact SPM and ensemble of extreme learning machines[J]. *Optik-International Journal for Light and Electron Optics*, 2017, 140: 964-974.