A Robust Data-Driven Approach for Fault Detection in Photovoltaic Arrays

Heybet Kiliç*, Bilal Gumuş†, Behnam Khaki‡, Musa Yilmaz§ and Peter Palensky*
* TU Delft, † Dicle University, ‡ New York Power Authority, § Batman University

***Abstract*—In this paper, a robust data-driven method for fault detection in photovoltaic (PV) arrays is proposed. Our method is based on the random vector functional-link networks (RVFLN) which has the advantage of randomly assigning hidden layer parameters with no tuning. To eliminate the effect of measurement noise and overfitting in the training process which reduce the fault detection accuracy, the sparse-regularization method is utilized which uses l_2-norm with loss weighting factor to compute the output weights. To attain a strong robustness against the outlier samples, the non-parametric kernel density estimation is employed to assign a loss weighting factor. Through rigorous simulation studies, we validate the performance of our proposed method in detecting the short and open circuit faults based on only the output current and voltage measurements of PV arrays. In addition to a stronger robustness comparing with the least square-support vector machine, we also show that our proposed method provides 80% and 100% average detection accuracy for short circuit and open circuit, respectively.**

***Index Terms*—Canonical correlation analysis, fault detection, photovoltaic array, random vector-link network, sparse-regularization.**

I. Introduction

Despite advanced technologies in photovoltaic (PV) power generation deployment, PV systems are exposed to the faults that negatively affect their efficiency and profitability. Undetected faults in PV systems may lead to loss of power in the system as well as fire-hazards and safety issues. Conventional protection and fault detection methods used in PV systems include overcurrent protection devices (OCPDs), ground fault detection interrupters (GFDI) and fuses [1], [2]. However as demonstrated in the literature, the faults occurring under low mismatch and high impedance conditions do not change the output of PV arrays comparing to their operation under normal conditions, thus they are not detectable by the conventional methods and without suitable fault detection algorithms [3]. In addition, the blocking diodes protecting the system from back-feeding current and active maximum power point tracking (MPPT) module may mask the faults; in the case of blocking diodes, the induced back-feeding current is insufficient to melt the protective fuses; when MPPT is active, although the back-feeding current exceeds the fuse threshold, it is suddenly lowered down by MPPT, meaning that it is insufficient to melt the fuse [4].

In the literature, numerous PV fault detection and diagnosis methods have been proposed which are classified under two groups: model-based and data-driven. In the model-based approaches, the real-time measurement data is compared with the parameters of the analytical PV model to detect the faults [5]–[9]. Regarding the model-based methods, an effective kernel generalized likelihood ratio test method is used in [5] for the detection of the faults in low mismatch and partial shading conditions, and the authors in [6] propose a method based on I-V characteristics of PV system for the detection of short-circuit (SC) and open-circuit (OC) faults; a statistical approach is proposed in [7] using a one-diode model for the detection of SC, OC and partial shading faults; in [8], DC-side faults are detected by a method comparing estimated PV array parameters and actual measurements, in which the method recognizes the faulty condition using the eagle strategy-based hybrid adaptive Nelder-Mead simplex algorithm; to detect the ground, OC and bridge faults, the authors in [9] propose to use the fractional-order color relation classifier. Although the model-based methods require less equipment and are applicable to a variety of PV systems, they have low fault detection accuracy due to the difference between the actual system and the simulated model, similar operation under low mismatch and high impedance conditions, non-linear operation characteristics of PV systems and ever-changing environmental conditions.

In the data-driven approaches, however, the collected data from the PV system operation is used to extract the faults' features and detect them in real-time using measurement data [10]–[15]. Various methods based on signal processing, computational intelligence and machine learning techniques have been proposed by the researchers; unsupervised-based dilation and erosion clustering algorithms are proposed in [10] for the detection of SC and OC faults; using wavelet-based artificial neural network (ANN), an algorithm is developed in [11] for the fault detection on both DC and AC sides of the PV system; in [12], SC and partial shading faults are detected by the wavelet packet transforms using only MPPT information; the authors in [13] propose the support vector machine (SVM) based on the multi-resolution signal decomposition (MSD) to detect the SC faults, while an MSD-based fuzzy reference system is proposed in [14] for SC and ground faults detection; an algorithm called graph-based semi-supervised learning is proposed in [15] for the detection of SC and OC faults. Although those methods have shown an acceptable accuracy, they might not be practical due to the requirement of large training data, time-consuming training steps, parameter setting, sensitivity to the measurement noise, possible overfitting, and low generalization performance.

To address the issues discussed above, we propose a new method based on the random vector functional-link networks (RVFLN) for SC and OC faults detection in the PV system. Our proposed method works using the output current and voltage measurements of PV array. The features used to distinguish faulty conditions from normal operation are derived from the recorded current and voltage measurements. To decrease the computational complexity of the detection method and increase the efficiency of algorithm, the dimension of the extracted features are reduced by the canonical correlation analysis (CCA). Then, a sparse-RVFLN system is built, and finally the trained sparse-RVFLN system is tested using unseen actual data for model evaluation and performance. The contributions of this work are as follows:

- To the best of our knowledge, CCA for the first time is introduced for PV fault detection which helps to increase detection accuracy in case of hard-to-detect faults, e.g. high impedance and low mismatch SC faults.
- Our method processes large scale real-time data, and it shows a faster learning procedure than the reported conventional methods in the literature.
- RVFLN weights are calculated using sparse-regularization, which not only increases the fault detection accuracy, but also strengthens its robustness against the outlier samples resulting from overfitting, measurement noise, and sensor and measurement errors which are missing in most of the conventional

methods.

II. Proposed PV Fault Detection Method

In this section, we first show how feature vectors are constructed based on the output voltage and current measurements of PV array. Then, CCA, which provides the distinctive properties to algorithm for high impedance and low mismatch SC faults, is used to create a new feature space. Finally, the sparse-RVFLN classifier using $l_2 - norm$ and non-parametric kernel density estimation(NKDE) to reduce the impact of the noise are introduced.

A. Data Pre-processing

1) Feature Construction: To detect the faults occurring in PV array and distinguish them from normal system operation, the feature vectors must be constructed from the measured output current and voltage. To cover different faults, we construct 10 feature vectors which are listed in Table I.

TABLE I
FEATURE VECTORS USED FOR FAULT DETECTION.

Features	Notation	Formulation
$\mathbf{F}_1$	I_{norm}	$I_{MPP}/I_{SC-Array}$
$\mathbf{F}_2$	V_{norm}	$V_{MPP}/V_{OC-Array}$
$\mathbf{F}_3$	FF	$I_{MPP}V_{MPP}/I_{SC}V_{OC}$
$\mathbf{F}_4$	A_{curve}	$\int_0^{V_{OC}} IdI$
$\mathbf{F}_5$	X_s	dI/dV
$\mathbf{F}_6$	X_{s-irr}	$I_{norm}/V_{norm} * irr$
$\mathbf{F}_7$	ΔX_s at V_{OC}	$dI/dv\|_{V_{OC}}$
$\mathbf{F}_8$	ΔX_s at mV_{OC}	$dI/dv\|_{mV_{OC}}$
$\mathbf{F}_9$	ΔX_s at I_{SC}	$dI/dv\|_{I_{SC}}$
$\mathbf{F}_{10}$	ΔX_s at mI_{SC}	$dI/dv\|_{mI_{SC}}$

Among the feature vectors, the first two ($\mathbf{F}_1$ and $\mathbf{F}_2$) are obtained by normalizing the measured current and voltage to ensure the applicability of the proposed algorithm to the PV systems with various size and avoid re-tuning the parameters. I_{MPP} and V_{MPP} are, respectively, the maximum current and voltage, and $I_{SC-Array}$ and $V_{OC-Array}$ denote the SC current and OC voltage, respectively. The third feature ($\mathbf{F}_3$) is the fill factor indicating the maximum normalized power to be supplied by the PV array. Although both irradiation variation and faults change the fill factor, the faults in the system result in more abrupt changes comparing to the irradiation. Thus, using the fill factor is effective in detecting the SC faults. As the shape of I-V curve is significantly affected by partial shading, the fourth feature ($\mathbf{F}_4$), which refers to the area under the curve, is used to distinguish the faults from partial shading occurring under normal operating conditions. The fifth feature ($\mathbf{F}_5$) describes the derivative of conductance, and the sixth feature ($\mathbf{F}_6$) aims to reduce the effect of irradiation on conductance as any variation in irradiation changes $\mathbf{F}_5$. The seventh feature ($\mathbf{F}_7$), i.e. the rate of change of conductance at V_{OC}, and the eighth feature ($\mathbf{F}_8$) is mV_{OC}, which refers to the rate of change of conductance at the middle point between OC and maximum power points. These features are constructed to detect SC faults occurring under low mismatch and high impedance which are close to the normal operating conditions. The last two features ($\mathbf{F}_9$ and $\mathbf{F}_{10}$), which provide advantages in the detection of OC faults close to the normal operating conditions, refer to the rate of change of conductance at SC point and middle point between, SC and maximum power points, respectively.

B. Feature Extraction by CCA

If all feature vectors are used in the fault detection algorithm, computational complexity will critically increase and affect the efficiency and accuracy of the algorithm. Therefore, CCA is introduced to extract a number of features from high-dimensional correlated features.

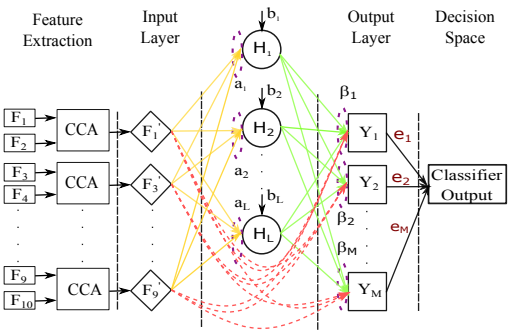

Fig. 1. CCA based RVFLN's structure.

CCA is a statistical technique used to find the low-dimensional linear projections of two maximally correlated high-dimensional random vectors [16], [17]. Let us assume that $\mathbf{F}_i = [F_{i1}, F_{i2}, \cdots, F_{iN}]^T \in R^n$ is the feature vector, and $\mathbf{Y}$ is the class label. If there is a main feature $\mathbf{F}_1$ and an auxiliary feature of $\mathbf{F}_2$ of dimension N and $c : R^n \rightarrow \{1, \cdots, M\}$ M classes classifier, providing the $(\mathbf{F}_1, \mathbf{F}_2, \mathbf{Y}) \sim F_{\mathbf{F}_1\mathbf{F}_2,Y}$, be an $R^n \times R^n \times \{1, \cdots, M\}$ pair, then CCA is applicable on $(\mathbf{F}_1, \mathbf{F}_2)$ pair to derive the projection $\mathbf{A}$ based on main $\mathbf{F}_1$ and auxiliary $\mathbf{F}_2$ features. The identification of CCA projection $\mathbf{A} = A(\mathbf{F}_1, \mathbf{F}_2)$ can be addressed as the problem of finding two sets of canonical vectors $\{a_i\}$ and $\{a_j\}$ to maximize the correlation between $a_i^{'}\mathbf{F}_1$ and $a_j^{'}\mathbf{F}_2$ for each $i, j = 1, 2, ..., r$. Dimension reduction for $\mathbf{F}_1$ is obtained from R^n to R^r before a classification operation. Then, a pair of conical vector set is defined as follows.

$$\{a_i, a_j\} = \text{argmin} \frac{a_i^{'}\Sigma_{\mathcal{F}_1\mathbf{F}_2}a_j^{'}}{\sqrt{a_i^{'}\Sigma_{\mathbf{F}_1}a_i}\sqrt{a_j^{'}\Sigma_{\mathbf{F}_2}a_j}} \quad (1)$$

A matrix is defined as the CCA projection matrix for $\mathbf{F}_1$, and $\mathbf{F}_1^p = A^{'}\mathbf{F}_1 \in R^r$ is the projected feature vector. Where $A^{'} = [a_1^{'}, \ldots, a_r^{'}]$.

Features $(\mathbf{F}_1, \mathbf{F}_3, \mathbf{F}_5, \mathbf{F}_7, \mathbf{F}_9)$ and $(\mathbf{F}_2, \mathbf{F}_4, \mathbf{F}_6, \mathbf{F}_8, \mathbf{F}_{10})$ are determined as main and auxiliary features, respectively.

C. RVFLN's Model Establishment

RVFLN which its structure is presented in Fig. 1 is developed as a single hidden layer feedforward networks [18]. RVFLN has been successfully used in various engineering practices [19]–[22] for the purposes of regression and classification since their development. The advantage of RVFLN is the random assignment of weights between the input layer and the hidden layer and the lack of need for tuning. Output weights are calculated with least squares (LS) algorithm. Practically, unlike the gradient-based learning techniques which require the learning parameters to be predefined and training may take several minutes to several hours to ensure the convergence, RVFLN can process data in real-time in most of the engineering applications.

If $\mathbf{F}_i^p = [F_{i1}^p, F_{i2}^p, ..., F_{iN}^p]^T \in R^n$ and $\mathbf{Y}_i = [Y_{i1}, Y_{i2}, ..., Y_{iM}]^T \in R^m$ are identified as the input and output vectors, respectively, RVFLN with hidden layer L can be defined as:

$$y_i\left(F_i^p\right) = \sum_{j=1}^{L} \beta_j h\left(< a_j, F_i^p > + b_j\right), i = 1, 2, ..., M, \quad (2)$$

where $y_i(F_i^p)$ is the output of RVFLN, $a_j = [a_{j1}, a_{j2}, \cdots, a_{jL}]^T \in R^l$ is the hidden weight vector which connects the input layer and

the hidden layer and is generated randomly in a specific probabilistic space, $\beta_j = [\beta_{j1}, \beta_{j2}, \cdots, \beta_{jM}]^T \in R^m$ is the output weight and must be calculated, $b_j = [b_{j1}, b_{j2}, \cdots, b_{jL}]^T \in R^l$ indicates the bias of hidden node, $< a_j, F_i^p >$ refers to the inner product of vectors a_j and F_i^p and $h(.)$ denotes the activation function which satisfies both conditions below:

$$\int_R h^2(x)\,dx \quad or \quad \int_R \left[h^{'}(x)\right]^2 dx < \infty. \tag{3}$$

The learning purpose of RVFLN is to minimize the error between the model outputs and actual outputs by calculating the β_j, a_j and b_j values which satisfy:

$$\sum_{j=1}^{L} \beta_j h(< a_j, F_i^p > + b_j) = y_i, j = 1, 2, ..., M, \tag{4}$$

which can be compactly defined as:

$$\mathbf{H}\beta = \mathbf{Y}, \tag{5}$$

where $\mathbf{H}$ is the hidden output matrix, β is the output weight matrix that needs to be evaluated and $\mathbf{Y}$ is the target output matrix. Those are specifically defined as follows:

$$\begin{aligned} &\mathbf{H}(a_1, \ldots, a_L, F_1^p, ..., F_M^p, b_1, \ldots, b_L) = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_M \end{bmatrix} \\ &= \begin{bmatrix} h(< a_1, F_1^p > + b_1) & \ldots & h(< a_L, F_1^p > + b_L) \\ \vdots & \ddots & \vdots \\ h(< a_1, F_M^p > + b_1) & \ldots & h(< a_L, F_M^p > + b_L) \end{bmatrix} \\ &\beta = \begin{bmatrix} \beta_1^T \ldots \beta_L^T \end{bmatrix}^T \quad \mathbf{Y} = \begin{bmatrix} y_1^T \ldots y_M^T \end{bmatrix}^T. \end{aligned} \tag{6}$$

After randomly assigning the parameters $(a_1, ..., a_L, b_1, ..., b_L)$ to train the network and find the optimal output weights, functional link networks are trained, and the optimal output weight network $\widehat{\beta}$ is obtained with the help of LS by:

$$\widehat{\beta} = \operatorname{argmin} \|\mathbf{H}\beta - \mathbf{Y}\|^2, \tag{7}$$

where $\widehat{\beta} = \mathbf{H}^{\dagger} Y$, and $\mathbf{H}^{\dagger} = \left(\mathbf{H^T H}\right)^{-1} \mathbf{H^T}$ is the Moore-Penrose generalized inverse of $\mathbf{H}$.

1) Sparse RVFLN: Sensor error and measurement noise occur in practical applications. We propose to calculate the RVFLN's output weights by sparse-regularization method in order to eliminate the noise and avoid overfitting that may negatively affect the fault detection accuracy. The spare-regularization RVFLN includes regularization term, i.e. $l_2 - norm$, in the objective function [23]–[25] and can be represented by:

$$\widehat{\beta} = \frac{1}{2}\|\beta\|_2^2 + \frac{D}{2}\sum_{i=1}^{N} \epsilon_i^2 \quad s.t \quad y_i - \mathbf{H}\beta = \epsilon_i, \tag{8}$$

where N is the number of the discrete training samples. Parameter D ensures regularization between $\|\beta\|_2^2$ and the training error and can be computed via Morozov's discrepancy principle [26]. The output weight norm and training error represent the structural and empirical losses, respectiviley, in (8). In order to improve robustness of the proposed method, NKDE and sparse-RVFLN with loss weighting factor are used. The loss weighting factor can attenuate the effect of noise for both small and long length samples which are also known as the low and high reliability samples, respectively. The long length samples presents the normal data while the low length samples include noise. Therefore, we modify (8) as the following:

$$\begin{aligned} \widehat{\beta} = \frac{1}{2}\|\beta\|_2^2 + \frac{D}{2}\sum_{i=1}^{N} q_i \|\epsilon_i\|_2^2 \\ s.t \quad y_i - \mathbf{H}\beta = \epsilon_i, \end{aligned} \tag{9}$$

where q_i is the loss weighting factor for the ith sample, and $\widehat{\beta}$ is defines as:

$$\widehat{\beta} = \begin{cases} \mathbf{H}^T(\frac{1}{\mathbf{D}} + \mathbf{QHH}^T)^{-1}\mathbf{QY}, & N < L \\ (\frac{1}{D} + \mathbf{H}^T\mathbf{QH})\mathbf{H}^T\mathbf{QY}, & N > L \end{cases}, \tag{10}$$

where q_i is the diagonal element of $\mathbf{Q}$ matrix and $i = 1, ..., N$. We set $\mathbf{Q}$ as the identity matrix and use NKDE to obtain the probability density function for the residuals. A Residual function for sparse-RVFLN can be define as:

$$r_j = \sum_{j=1}^{L} \beta_j h(< a_j, F_i^p > + b_j) - y_i, i = 1, 2, ..., N. \tag{11}$$

According to (11),the probability density function is represented as:

$$f(x) = \frac{1}{hN}\sum_{j=1}^{n} \phi\left(\frac{x - r_j}{h}\right), \tag{12}$$

where $h = 1.06\sigma N^{-1/5}$ indicates the width of the estimated window, σ represents the standard deviation and ϕ is Gaussian kernel function which is shown by:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}x^2. \tag{13}$$

Using (12), the probability density function $f(r_j)$ can be calculated for each residual, and $\mathbf{Q}$ can be set according to $f(r_j)$. Algorithm 1 shows the implementation of sparse-RVFLN for real-time fault detection in PV arrays.

Algorithm 1: CCA based sparse-RVFLN algorithm.

```
Construct feature vector F_i = [F_i1, F_i2..., F_iN]^T.
Define main and auxiliary features, and apply CCA to obtain
  projection feature vector F_i^p = [F_i1, F_i2..., F_iN]^T.
Calculate D parameter by Morozov's discrepancy principle.
Inputs: Projection feature vectors, number of hidden layers
  and D .
Generate randomly hidden node parameters (w_i, b_i).
Calculate hidden layer output H.
Set Q as the identity matrix.
for j=1 to L do
    for i=1 to N do
        Calculate residuals r_j using (11).
        Construct Q by (12) and (13).
    end
end
Calculate the output weight β̂ for each output node according
  to (10).
Output: Output weight β̂.
```

III. CASE STUDIES

In this section, we evaluate the performance of the proposed sparse-RVFLN method for PV fault detection through comprehensive numerical simulation and experiments.

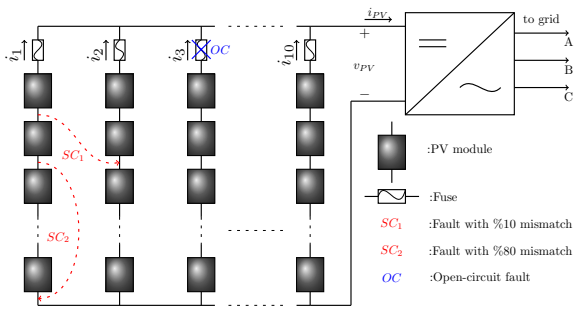

Fig. 2. Simulated PV array with potential SC and OC faults.

A. Numerical Setup

We implement the PV system in PSCAD/EMTDC and the proposed fault detection method in MATLAB for the numerical simulation. The ungrounded PV array system shown in Fig. 2 consists of 10 PV arrays connected in parallel each of which includes 10 series modules with bypass diodes. To show the performance of our fault detection method, we run the simulation studies for various operating and environment scenarios including different mismatch percentage levels, weather temperatures, irradiation values, and fault instances and impedances. In the following, more details are provided for the simulated scenarios.

1) SC Faults: They are line-to-line (LL) faults occurring under 2250 different conditions. The conditions include a combinations of different fault scenarios such as the irradiation levels $(200, 400, 600, 800, 1000)$ $[\mathrm{W/m^2}]$, different mismatch percentages $(10, 20, 30, 40, 50, 60)$ %, several temperatures $(10, 20, 30, 40, 50)$ $[^\circ\mathrm{C}]$, different fault impedances $(0, 5, 15, 25)$ $[\Omega]$, and positive, negative and negative-to-positive zero-crossing.

2) OC Faults: These faults are simulated for 150 different scenarios under the conditions similar to the SC faults. The OC faults are classified into two groups called Group1 and Group2. The former includes The OC faults close to the normal operating conditions, where only one string is faulty at a moment. On the other hand, Group2 includes the OC faults not close to the normal operating conditions, where multiple strings are faulty.

3) Normal Operating Conditions: To distinguish SC and OC faults from normal operating conditions, the PV system is simulated for 600 different normal operating scenarios under different temperatures, mismatch percentages, fault impedance, and irradiation conditions.The training data set includes 280, 20 and 75 scenarios for SC faults, OC faults and normal operation conditions, respectively.

B. Simulation Results and Discussion

Simulation results of the fault detection algorithm demonstrate the high accuracy of the proposed algorithm. Results are presented in Table II for SC faults.

TABLE II
RESULTS FOR THE SIMULATED SC FAULTS.

Mismatch Percentage (%)	Fault Impedance [Ω] 0	5	15	25	Average (%) Accuracy
60	100	100	96.7	91.5	97.1
50	100	98.9	92	90.3	95.3
40	97	95.4	76	74.2	85.7
30	92.4	88.2	70.5	60.3	77.8
20	88.8	73.7	52.2	47.3	65.5
10	74.3	53.5	46.4	40.3	53.6

The proposed method has the fault detection accuracy over 95% in cases with 50% or higher mismatch and all fault impedances for SC faults. As the fault impedance increases, and the mismatch percentage decreases, it is quite difficult to detect faults because they are close to normal operating conditions. For instance, inn cases of SC faults with less than 30% mismatch and more than 15Ω fault impedance, the algorithm has a fault detection accuracy of above 55%. On the other hand, the algorithm has the fault detection accuracy of 100% for OC faults which are close to SC faults and normal operating conditions (e.g. OC fault in one string), and OC faults occurring in multiple strings. The results of fault detection accuracy assessment in OC fault scenarios are given in Table III.

TABLE III
RESULTS FOR THE SIMULATED OC FAULTS.

Group #	Temperature [°C]	Irradiance [W/m²] 200	400	600	800	1000	Average Accuracy(%)
1	10	100	100	100	100	100	100
	50	100	100	100	100	100	100
2	10	100	100	100	100	100	100
	50	100	100	100	100	100	100

Fig. 3 and 4 show the SC and OC fault detection accuracy of the proposed method in comparison with LS-SVM and LS-RVFLN methods. According to the results, sparse-RVFLN has a better detection accuracy for both SC and OC faults owing to calculating the output weights of RVFLN by sparse estimation method.

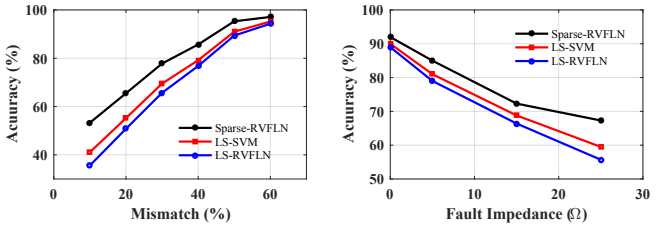

Fig. 3. Comparison of SC fault detection accuracy by Sparse-RVFLN, LS-SVM, and LS-RVFLN.

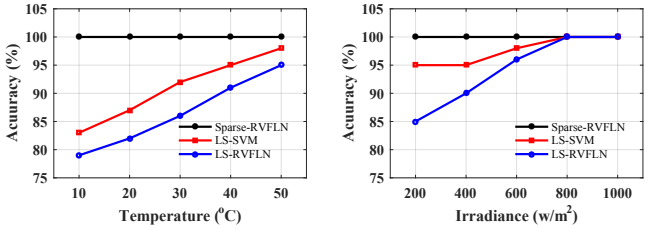

Fig. 4. Comparison of OC fault detection accuracy by Sparse-RVFLN, LS-SVM, and LS-RVFLN.

1) SC-OC Faults with Noisy Measurements: To test the robustness of the proposed method against noisy measurements, the samples are selected randomly at the ratios of 5%, 15%, 25%, 35% and 45% from the training dataset, and the noisy dataset is generated. detection accuracy results including noisy measurement are given in Table IV for SC and OC faults. The results show that fault detection accuracy of the other methods decreases due to the noisy measurements, while sparse-RVFLN provides an acceptable detection accuracy for both SC and OC faults.

2) Comparison of Computational Efficiency: Assigning input weights randomly and calculating the output weights with training, which is required by the conventional algorithms enable RVFLN to process and learn data fast which is suitable for practical application. Comparison of the proposed method and other conventional methods is presented in Table V.

TABLE IV
SC AND OC FAULTS DETECTION ACCURACY WITH NOISY MEASUREMENTS.

Fault Type	Method	Average Accuracy (%)	
		Noisy	Noiseless
SC	RVFLN	60.4	70.2
	Sparse-RVFLN	77.3	79.1
	LS-SVM	65.7	73.6
OC	RVFLN	83	89.1
	Sparse-RVFLN	98	100
	LS-SVM	86.2	91.4

TABLE V
COMPUTATIONAL EFFICIENCY OF RVFLN, SPARSE-RVFLN AND LS-SVM.

Method	Training Time (sec)	Testing Time (sec)
RVFLN	0.0125	0.0035
Sparse-RVFLN	0.1257	0.0045
LS-SVM	0.7264	0.9345

IV. CONCLUSION AND FUTURE WORK

In this paper, a new sparse-RVFLN method is proposed for the detection of SC and OC faults in PV arrays. New features are constructed and transformed by CCA to lower the dimension space and improve fault detection accuracy for hard-to-detect fault conditions such as low mismatch, high impedance, and active MPPT. In addition, sparse-regularization is used to calculate RVFLN's output weights to minimzie the effect of measurement noise and overfitting on the detection accuracy. The performance of the proposed method is verified by both simulation and experimental setup. In SC fault scenarios with low mismatch and high impedance fault, which are close to normal operating conditions, the proposed method provides a detection accuracy of minimum 77.5% for 30% low mismatch and all fault impedances, while it results in a detection accuracy of 100% in all OC fault scenarios. The proposed spare-RVFLN method is also compared with the conventional RVFLN and LS-SVM. The measurements with noise, in particular, reduce the fault detection accuracy of LS-SVM and RVFLN while it does not degrade the sparse-RVFLN performance. Furthermore, sparse-RVFLN has a higher computational efficiency than the conventional RVFLN and LS-SVM.

Notwithstanding the fact that the DC-side faults in PV arrays are investigated in this paper, numerous other faults occur in PV systems including ground SC, partial shading, arc and hot-spot faults. Therefore, we will focus on the development of the algorithms enabling the detection of those faults in our future research.

REFERENCES

[1] D. S. Pillai, J. P. Ram, N. Rajasekar, A. Mahmud, Y. Yang, and F. Blaabjerg, "Extended analysis on line-line and line-ground faults in pv arrays and a compatibility study on latest nec protection standards," *Energy Conversion and Management*, vol. 196, pp. 988–1001, 2019.

[2] D. S. Pillai and N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in pv systems," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 18–40, 2018.

[3] Y. Zhao, L. Yang, B. Lehman, J.-F. de Palma, J. Mosesian, and R. Lyons, "Decision tree-based fault detection and classification in solar photovoltaic arrays," in *2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*. IEEE, 2012, pp. 93–99.

[4] Y. Zhao, J.-F. De Palma, J. Mosesian, R. Lyons, and B. Lehman, "Line–line fault analysis and protection challenges in solar photovoltaic arrays," *IEEE transactions on Industrial Electronics*, vol. 60, no. 9, pp. 3784–3795, 2012.

[5] R. Fezai, M. Mansouri, M. Trabelsi, M. Hajji, H. Nounou, and M. Nounou, "Online reduced kernel glrt technique for improved fault detection in photovoltaic systems," *Energy*, vol. 179, pp. 1133–1154, 2019.

[6] S. R. Madeti and S. Singh, "Modeling of pv system based on experimental data for fault detection using knn method," *Solar Energy*, vol. 173, pp. 139–151, 2018.

[7] F. Harrou, Y. Sun, B. Taghezouit, A. Saidi, and M.-E. Hamlati, "Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches," *Renewable energy*, vol. 116, pp. 22–37, 2018.

[8] C.-L. Kuo, J.-L. Chen, S.-J. Chen, C.-C. Kao, H.-T. Yau, and C.-H. Lin, "Photovoltaic energy conversion system fault detection using fractional-order color relation classifier in microdistribution systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1163–1172, 2015.

[9] Z. Chen, L. Wu, S. Cheng, P. Lin, Y. Wu, and W. Lin, "Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and iv characteristics," *Applied energy*, vol. 204, pp. 912–931, 2017.

[10] S. Liu, L. Dong, X. Liao, Y. Hao, X. Cao, and X. Wang, "A dilation and erosion-based clustering approach for fault diagnosis of photovoltaic arrays," *IEEE Sensors Journal*, vol. 19, no. 11, pp. 4123–4137, 2019.

[11] I. M. Karmacharya and R. Gokaraju, "Fault location in ungrounded photovoltaic system using wavelets and ann," *IEEE Transactions on Power Delivery*, vol. 33, no. 2, pp. 549–559, 2017.

[12] B. P. Kumar, G. S. Ilango, M. J. B. Reddy, and N. Chilakapati, "Online fault detection and diagnosis in photovoltaic systems using wavelet packets," *IEEE Journal of Photovoltaics*, vol. 8, no. 1, pp. 257–265, 2017.

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

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

[15] Y. Zhao, R. Ball, J. Mosesian, J.-F. de Palma, and B. Lehman, "Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays," *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2848–2858, 2014.

[16] C. Shen, M. Sun, M. Tang, and C. E. Priebe, "Generalized canonical correlation analysis for classification," *Journal of Multivariate Analysis*, vol. 130, pp. 310–322, 2014.

[17] W. Zuobin, M. Kezhi, and G.-W. Ng, "Feature regrouping for cca-based feature fusion and extraction through normalized cut," in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 2275–2282.

[18] Y. . Pao and Y. Takefuji, "Functional-link net computing: theory, system architecture, and functionalities," *Computer*, vol. 25, no. 5, pp. 76–79, May 1992.

[19] M. Sahani and P. K. Dash, "Fpga-based online power quality disturbances monitoring using reduced-sample hht and class-specific weighted rvfln," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4614–4623, 2019.

[20] P. Zhou, Y. Lv, H. Wang, and T. Chai, "Data-driven robust rvflns modeling of a blast furnace iron-making process using cauchy distribution weighted m-estimation," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 9, pp. 7141–7151, 2017.

[21] R. Katuwal, P. N. Suganthan, and L. Zhang, "An ensemble of decision trees with random vector functional link networks for multi-class classification," *Applied Soft Computing*, vol. 70, pp. 1146–1153, 2018.

[22] W. Dai, Q. Chen, F. Chu, X. Ma, and T. Chai, "Robust regularized random vector functional link network and its industrial application," *IEEE Access*, vol. 5, pp. 16 162–16 172, 2017.

[23] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[24] A. E. Hoerl. and R. W. Kennard, "Ridge regression-1980: Advances, algorithms, and applications," *American Journal of Mathematical and Management Sciences*, vol. 1, no. 1, pp. 5–83, 1981.

[25] H. T. Huynh and Y. Won, "Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks," *Pattern Recognition Letters*, vol. 32, no. 14, pp. 1930–1935, 2011.

[26] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Siam, 2005, vol. 4.