Fault Classification in Photovoltaic Arrays Using Graph Signal Processing

Jie Fan[†], Sunil Rao[†], Gowtham Muniraju[†], Cihan Tepedelenlioglu and Andreas Spanias
SenSIP Center, School of ECEE, Arizona State University

***Abstract*—In this paper, we address the problem of fault classification in PhotoVoltaic (PV) arrays using a semi-supervised graph signal processing approach. Traditional fault detection and classification methods require large amounts of labeled data for training. In utility scale solar arrays, obtaining labeled data for different fault classes is resource intensive. We propose a graph based classification technique that relies on a limited amount of labeled data. We compare our results with the well known supervised machine learning classifiers such as the K-nearest neighbour classifier, random forest classifier, support vector machines, and artificial neural networks. We also show that the graph-based classifiers require lower training computational cost compared to the standard supervised machine learning algorithms. The proposed method also achieves good classification performance with unseen data. We validate our method on a real-time dataset and show significant improvements over existing approaches.**

***Index Terms*—Graph Signal Processing, Machine Learning, Photovoltaic Array, Solar Array Fault Classification.**

I. Introduction

In the last decade, new solar photovoltaic (PV) cell technologies have emerged [1], [2] for grid connected systems. However, automatic fault detection and accurate diagnosis of PV array systems is still an open problem [3]–[5]. PV arrays are generally installed in remote locations, and are often subjected to harsh weather conditions. The occurrence of PV faults is unpredictable and requires constant remote monitoring of several parameters. Even when over-current protection devices (OCPD), ground fault detection interrupters (GFDI) and smart monitoring devices (SMDs) with data transmission capabilities are integrated within the PV array system, recent studies [3], [6] have shown that these devices offer diagnosis for a limited set of commonly occurring faults. In this paper, we propose a novel graph-based method to identify PV faults with limited data.

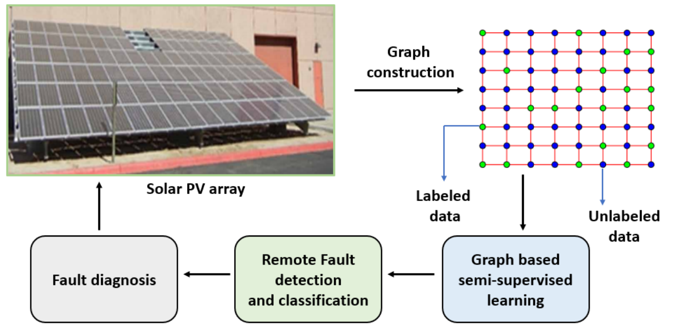

Figure 1: An illustrative block diagram of a graph-based semi-supervised fault classification and diagnosis.

In the following, we first briefly review machine learning methods for PV fault detection. An outlier detection method, solely based on the measured current was used to identify the faulty panel in [7]. This technique does not account for environmental factors such as humidity, shading, and soiling. Authors in [8], train an artificial neural network (ANN) to monitor the health of PV systems and to manage maintenance schedules based on the degradation of PV modules. A review of neural network (ANN) methods for PV arrays is presented in [9]. In addition, several studies [10]–[14] have proposed statistical outlier detection and neural network classifiers for fault detection. We propose here a graph signal processing based semi-supervised learning technique, which achieves good performance in fault classification with relatively limited data.

More specifically, we adopt a semi-supervised graph-based classifier to identify commonly occurring PV faults. We view the solar PV array as a connected graph, similar to Figure 1, and associate graph signals to represent the measurements of the PV modules. First, a classifier is

[†] These authors contributed equally.

The authors from Arizona State University are funded in part by the NSF CPS award 1646542 and the SenSIP Center, School of ECEE, Arizona State University.

978-1-7281-6389-5/20/$31.00 ©2020 IEEE

trained on the data available from the labeled nodes. The constructed graph is then used to propagate information from labeled samples to the unlabeled samples. Since graph-based methods are semi-supervised, the method requires lower computational cost than conventional supervised machine learning (ML) classifiers and ANNs.

Graph-based methods have been proposed for clustering [15], distributed estimation [16]–[18], localization [19] and outlier detection [20] problems in PV cyber-physical systems. Signal Processing and Machine Learning approaches have been proposed in [21]–[25] Recently, graph-based semi-supervised methods [4], [26] were proposed for fault detection in PV arrays. In contrast to those methods, our approach is based on graph signal processing (GSP) [27], [28], wherein, computing the inverse of matrices is avoided. Since matrix inverse scales as $\mathcal{O}(N^3)$, our method is computationally efficient, especially when the dataset dimensions are large, which is often the case for PV arrays.

The rest of the paper is organized as follows. In Section II, we describe the different faults occurring in PV arrays and the features extracted from the Sandia model to build the classifier. We provide a brief introduction to GSP in Section III. We explain the details of our graph-based semi-supervised classifier in Section IV. In Section V, we compare the performance of different classifiers to examine our ideas. We conclude the whole paper in Section VI.

II. Fault Diagnosis

In this section, we discuss the problem of fault classification in PV arrays. We identify 5 commonly occurring conditions in PV arrays, namely: standard test conditions (STC), shaded modules, degraded modules, soiled modules, and short circuit conditions. Our goal is to correctly classify the PV data into these classes via graph signal processing. In order to achieve this, we use the Sandia model [29] to develop an input feature matrix for the solar PV modules. The Sandia model involves parameters such as open circuit voltage V_{OC}, short circuit current I_{SC}, maximum voltage V_{MP} and maximum current I_{MP}, as shown in Figure 2. Additionally, we also obtain measurements for irradiance levels per hour per day and the corresponding temperature readings.

We use the PVWatts dataset [30] for our fault classification experiments. The PVWatts dataset is obtained for a period of one year, from January to December of 2006. The dataset includes Standard Test Conditions (STC) and four types of faults, namely: shading, degraded

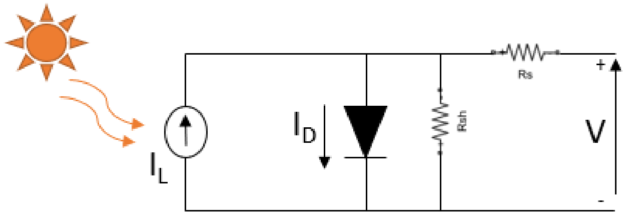

Figure 2: A circuit of the one diode model of a solar cell.

modules, soiling, and short circuit. When shading occurs, the measured power and irradiance are lower than STC, which is usually due to overcast conditions, cloud cover or building obstruction. Degraded modules are caused by wear and tear of aged PV modules. As a result, degraded modules cannot produce the standard rated power. Since PV modules are exposed to the outdoor environment, modules are soiled by dust, snow, bird droppings, etc, resulting in significant degradation of power output. Short circuits are a result of accidental shorting of PV modules due to faulty wires, equipment, etc. Short circuits not only result in power loss but are also a potential fire hazard. Therefore, the reliability of the PV systems can be significantly enhanced by the automatic diagnosis of these faults.

III. Background of Graph Signal Processing

For completeness, we review the key concepts of graph signal processing, including graph filter and graph shift operators.

A. Graph and Graph Signal

A graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ has N nodes $\mathcal{V} = \{1, 2, \cdots, N\}$, and described by an $N \times N$ matrix $\mathbf{A}$ which uses edge weights to characterize the relationships among all nodes. The graph signal is defined as $\mathbf{s} = [s_1, s_2, \cdots, s_N]^T$ and, based on the relationship among the nodes, GSP operators can be designed to conduct (propagate) the graph signal $\mathbf{s}$, throughout the graph.

B. Graph Shift and Graph Filter

GSP translates the traditional digital signal processing (DSP) concepts to the graph domain. Similar to the time shift operation in DSP filters, the graph shift operator is the base of the concept to design a graph filter. Consider a graph shift matrix $\mathbf{A}$, then the graph shift operation is given by

$$\widetilde{\mathbf{s}} = \mathbf{A}\mathbf{s} \tag{1}$$

There are numerous choices for the shift matrix $\mathbf{A}$, such as adjacency matrix, Laplacian matrix, normalized versions and other variations on these matrices. In DSP, the task of designing a conventional FIR filter involves finding the optimal filter taps for different time shift components. Similarly, in graph domain, an L^{th} order shift-invariant graph filter is defined as

$$\mathbf{H} = h(\mathbf{A}) = h_0\mathbf{I} + h_1\mathbf{A}^1 + \cdots h_L\mathbf{A}^L, \tag{2}$$

where h_i are scalar coefficients of the graph filter $\mathbf{H}$. Then we can conduct the graph filter operator $\mathbf{H}$ on the graph signal $\mathbf{s}$ as,

$$\mathbf{s}^{\text{fil}} = \mathbf{H}\mathbf{s}. \tag{3}$$

where, $\mathbf{s}^{\text{fil}}$ denotes the filtered graph signal. In this paper, the fault classification is achieved through a graph filtering process.

IV. Semi-supervised Graph-based Classification

In this section, we design a graph filter as a classifier to identify the specific types of faults in large scale utility arrays. We use an $N \times D$ matrix $\mathbf{X}$ to represent the initial dataset that has N samples and D features. Similarity among the nodes on the graph is represented by the graph shift matrix. We estimate similarity based on the Euclidean distance $\rho(\cdot)$ between the nodes, given by,

$$A_{i,j} = \rho(\mathbf{x}_i, \mathbf{x}_j), \tag{4}$$

where, $\mathbf{x}_i$ and $\mathbf{x}_j$ are i^{th} and j^{th} rows of $\mathbf{X}$. In [27], [28], the graph shift matrix is generated by,

$$A_{i,j} = \frac{\texttt{exp}(-\rho(\mathbf{x}_i, \mathbf{x}_j)/\sigma)}{\sum_{i=1}^{N}\texttt{exp}(-\rho(\mathbf{x}_i, \mathbf{x}_j)/\sigma)}, \tag{5}$$

where σ is a scaling coefficient. Note that the graph shift matrix obtained by equation (5) is the Hermitian transpose of the transition matrix of the graph [31].

The problem of fault classification translates to the node classification problem on the graph, wherein, each node belongs to a particular class. Consider $\mathbf{S}$ to be an $N \times K$ matrix that collects the labels of N samples, where each sample belongs to one of the K categories. For nodes with labels, $\mathbf{S}$ is one-hot encoded, i.e, if the i^{th} node belongs to j^{th} category, then $S_{i,j} = 1$ while the remaining elements of that row are 0. If a node is unlabeled, then all the elements in the corresponding row will be 0.

Given feature matrix X, graph shift matrix $\mathbf{A}$, and the node target class matrix $\mathbf{S}$, our goal is to find the graph filter $\mathbf{H}$, which classifies the nodes into true classes. The filter taps h_l of the filter $\mathbf{H}$ is computed by solving the convex objective function, given by

$$\mathcal{L} = \underset{\mathbf{h}}{\text{argmin}} \;\; ||\mathbf{R}\sum_{l=0}^{L} h_l\mathbf{A}^l\mathbf{S} - \mathbf{S}||_F, \tag{6}$$
$$\text{subject to} \;\; \mathbf{h} \in \Theta_h, \sum h_l = 1$$

where $||\cdot||_F$ represents Frobenius norm and $\mathbf{R}$ is an $N \times N$ diagonal matrix, wherein $R_{i,i} = 1$ if i^{th} sample is labeled, otherwise $R_{i,i} = 0$. The rectangular domain of filter coefficients Θ_h can be empirically decided. Since the objective function $\mathcal{L}$ given in equation (6) is a linear least square problem, it can be solved through its closed-form solution with the complexity $\mathcal{O}(N^3)$ or by an interior-point solver. After the filter $\mathbf{H}$ is well-trained, the classification result can be obtained by

$$\mathbf{s}^{\text{class}} = Q\left(\mathbf{s}^{\text{fil}}\right) = Q(\mathbf{H}\mathbf{s}), \tag{7}$$

where $Q(\cdot)$ is non-linear operator that transforms the largest value in each row to 1 and remaining elements to 0, and $\mathbf{s}^{\text{class}}$ denotes the class to which the node belongs. As the graph shift matrix contains information from both labeled and unlabeled data, the graph filter is a semi-supervised classifier. Through the objective function in equation 6, we train our graph filter by updating its eigenvalues. The eigenvectors of H are constant and decide the limit of the graph filter performance. Therefore, the performance of the GSP approach relies on the quality of H. The graph shift matrix is required to represent accurate similarity information among all nodes.

V. Simulation Results

In our experiments, we perform fault classification on PVWatts dataset. We obtain about 4400 measurements per class corresponding to the entire array. Therefore, this dataset has 22000 data samples, which corresponds to $N = 22000$ nodes in our graph. Each data sample corresponds to one of the five classes mentioned in Section II. We adopt a feature matrix $\mathbf{X}$ with 9 features for every node. The 9 features namely, V_{OC}, I_{SC}, V_{MP}, I_{MP}, fill factor, temperature, irradiance, gamma ratio, and maximum power, are derived from the Sandia model. These features are commonly used in the fault detection experiments [20], [32]. Our goal is to correctly classify each node to one of the 5 test conditions. We consider $\alpha\%$ of the samples to have labels and predict the labels for the rest of the nodes in the graph.

Table I: Comparison of various Classifiers with different labelling ratio for fault classification in PV arrays

Classification Error					
α	GSP	KNN	RFC	SVM	ANN
0.2	15.25 ± 4.77	16.14 ± 0.19	17.21 ± 0.31	19.55 ± 0.32	**14.13** ± **3.37**
0.3	**12.6** ± **3.13**	15.45 ± 0.11	15.98 ± 0.38	19.23 ± 0.19	15.72 ± 3.03
0.4	**10.32** ± **2.06**	14.98 ± 0.16	15.16 ± 0.18	19.30 ± 0.14	14.26 ± 2.72
0.5	**10.15** ± **1.97**	14.84 ± 0.26	15.08 ± 0.16	19.35 ± 0.04	12.48 ± 2.98
0.6	**9.97** ± **1.68**	14.39 ± 0.24	13.89 ± 0.54	18.95 ± 0.56	12.37 ± 2.37
0.7	**9.39** ± **1.67**	14.28 ± 0.43	13.46 ± 0.49	19.17 ± 0.36	12.86 ± 2.18

Firstly, we use $\mathbf{X}$ to generate the graph shift matrix $\mathbf{A}$ through equation (5). Next, we use the interior-point solver to solve the objective function given in (6), in order to compute the graph filter coefficients. Note that, the graph filter obtained is the fault classifier, which is then used to predict labels for the unlabeled data. Since we have the ground truth labels for all the nodes, we compute the overall error rate and use it as the metric to qualitatively evaluate the classifier's performance.

In addition to the proposed approach based on graph signal processing, we also apply conventional supervised machine learning classifiers, including random forest classifier (RFC), K-nearest neighbour classifier (KNN) and support vector machines (SVM), and the standard ANNs [33] to classify the PVWatts dataset. We trained an RFC classifier with 300 estimators with a depth of 50. The SVM classifier was trained with a radial basis kernel and the KNN classifier with 30 nearest neighbours. We considered standard ANNs with 4 hidden layers each with 100 neurons. We used *Relu* activation function for the hidden layer and a softmax layer for the output layer. ANN was trained using Adam optimizer with a learning rate of 0.01. These hyper-parameters were selected using brute force grid search and were found to have the best results in each case. Additionally, we examine the test accuracy's and error rates of all classifiers under different labelling ratios α from 0.2 to 0.7. The results are reported in Table I. We find that, in all cases of α, GSP method significantly outperforms the other methods. GSP had the best error rate performance among all classifiers with 9.32% error followed by ANNS with 9.61%. ANNs performed better than the conventional ML classifiers in all cases. Although the performance ANNs can be improved by adding more data and making the network deeper, it leads to expensive data collection and extra computational resources. KNN and RFC classifiers reach a minimum error rate of 14% and 13% respectively, falling short by about 4.5% with respect to GSP. SVM had the highest error rate among all the classifiers. The superior performance of GSP method can be attributed to the structural graph data along with the measurement data to construct the classifier.

VI. Conclusions

In this paper, we present a graph signal processing based fault classification method for the solar array systems. The proposed method constructs the classifier using the measured data as well as the structural connectivity of PV array topology. In addition, our method requires a significantly lower percentage of labeled data for classification and achieves good performance. To illustrate this point, we have shown a comparison of our graph-based method with the supervised machine learning methods such as KNN, RFC, SVM, and the ANNs. Experimental results show that the graph-based method requires the lowest training cost. In contrast to the conventional graph-based classifiers, our graph filter approach can be trained without calculating the inverse of the matrix, which significantly reduces the algorithm's complexity.

References

[1] S. Bacha, D. Picault, B. Burger, I. Etxeberria-Otadui, and J. Martins, "Photovoltaics in microgrids: An overview of grid integration and energy management aspects," *IEEE Industrial Electronics Magazine*, vol. 9, no. 1, pp. 33–46, 2015.

[2] V. Pires, J. Martins, D. Foito, and C. Hão, "A grid connected photovoltaic system with a multilevel inverter and a le-blanc transformer," *International Journal of Renewable Energy Research (IJRER)*, vol. 2, no. 1, pp. 84–91, 2012.

[3] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A comprehensive review of catastrophic faults in pv arrays: types, detection, and mitigation techniques," *IEEE Journal of Photovoltaics*, vol. 5, no. 3, pp. 982–997, 2015.

[4] 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.

[5] G. Muniraju, S. Rao, S. Katoch, A. Spanias, C. Tepedelenlioglu, P. Turaga, M. K. Banavar, and D. Srinivasan, "A cyber-physical photovoltaic array monitoring and control system," *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, vol. 5, no. 3, pp. 33–56, 2017.

[6] J. Flicker and J. Johnson, "Analysis of fuses for blind spot ground fault detection in photovoltaic power systems," *Sandia National Laboratories Report*, 2013.

[7] Y. Zhao, B. Lehman, R. Ball, J. Mosesian, and J.-F. de Palma, "Outlier detection rules for fault detection in solar photovoltaic arrays," in *2013 Twenty-Eighth Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*. IEEE, pp. 2913–2920, 2013.

[8] D. Riley and J. Johnson, "Photovoltaic prognostics and heath management using learning algorithms," in *2012 38th IEEE Photovoltaic Specialists Conference*. IEEE, pp. 1535–1539, 2012.

[9] S. Daliento, A. Chouder, P. Guerriero, A. M. Pavan, A. Mellit, R. Moeini, and P. Tricoli, "Monitoring, diagnosis, and power forecasting for photovoltaic fields: a review," *International Journal of Photoenergy*, vol. 2017, 2017.

[10] M. H. Ali, A. Rabhi, A. El Hajjaji, and G. M. Tina, "Real time fault detection in photovoltaic systems," *Energy Procedia*, vol. 111, pp. 914–923, 2017.

[11] L. L. Jiang and D. L. Maskell, "Automatic fault detection and diagnosis for photovoltaic systems using combined artificial neural network and analytical based methods," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–8, 2015.

[12] L. Bonsignore, M. Davarifar, A. Rabhi, G. M. Tina, and A. Elhajjaji, "Neuro-fuzzy fault detection method for photovoltaic systems," *Energy Procedia*, vol. 62, pp. 431–441, 2014.

[13] L. S. Bruckman, N. R. Wheeler, J. Ma, E. Wang, C. K. Wang, I. Chou, J. Sun, and R. H. French, "Statistical and domain analytics applied to pv module lifetime and degradation science," *IEEE Access*, vol. 1, pp. 384–403, 2013.

[14] S. Rao, A. Spanias, and C. Tepedelenlioglu, "Solar array fault detection using neural networks," in *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*. IEEE, pp. 196–200, Taiwan, May 2019.

[15] G. Muniraju, S. Zhang, C. Tepedelenlioglu, M. K. Banavar, A. Spanias, C. Vargas-Rosales, and R. Villalpando-Hernandez, "Location based distributed spectral clustering for wireless sensor networks," in *2017 Sensor Signal Processing for Defence Conference (SSPD)*. IEEE, pp. 1–5, London, Dec 2017.

[16] G. Muniraju, C. Tepedelenlioglu, A. Spanias, S. Zhang, and M. K. Banavar, "Max consensus in the presence of additive noise," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, pp. 1408–1412, Nov 2018.

[17] G. Muniraju, C. Tepedelenlioglu, and A. Spanias, "Distributed spectral radius estimation in wireless sensor networks," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, pp. 1–5, Nov 2018.

[18] G. Muniraju, C. Tepedelenlioglu, and A. Spanias, "Analysis and design of robust max consensus for wireless sensor networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 4, pp. 779–791, Dec 2019.

[19] X. Zhang, C. Tepedelenlioğlu, M. K. Banavar, A. Spanias, and G. Muniraju, "Location estimation and detection in wireless sensor networks in the presence of fading," *Physical Communication, Elsevier*, vol. 32, pp. 62–74, 2019.

[20] S. Katoch, G. Muniraju, S. Rao, A. Spanias, P. Turaga, C. Tepedelenlioglu, M. Banavar, and D. Srinivasan, "Shading prediction, fault detection, and consensus estimation for solar array control," in *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*. IEEE, pp. 217–222, Saint Petersburg, May 2018.

[21] U. S. Shanthamallu, A. Spanias, C. Tepedelenlioglu, and M. Stanley, "A brief survey of machine learning methods and their sensor and iot applications," in *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, Larnaca, July 2017.

[22] E. Pedersen, S. Rao, S. Katoch, K. Jaskie, A. Spanias, C. Tepedelenlioglu, and E. Kyriakides, "Pv array fault detection using radial basis networks," in *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE, Patras, July 2019.

[23] K. Jaskie and A. Spanias, "Positive and unlabeled learning algorithms and applications: A survey," in *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE, Patras, July 2019.

[24] H. Braun, S. T. Buddha, V. Krishnan, A. Spanias, C. Tepedelenlioglu, M. Banavar, S. Takada, T. Takehara, and T. Yeider, *Signal processing for solar array monitoring, fault detection, and optimization*. Morgan & Claypool Publishers, vol. 7, no. 1, pp. 1–95, September 2012.

[25] H. Braun, S. Buddha, V. Krishnan, C. Tepedelenlioglu, A. Spanias, M. Banavar, and D. Srinivasan, "Topology reconfiguration for optimization of photovoltaic array output," *SEGAN*, vol. 6, pp. 58–69, June 2016.

[26] H. Momeni, N. Sadoogi, M. Farrokhifar, and H. F. Gharibeh, "Fault diagnosis in photovoltaic arrays using gbssl method and proposing a fault correction system," *IEEE Transactions on Industrial Informatics*, 2019.

[27] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[28] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs: Graph filters," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp. 6163–6166, 2013.

[29] D. L. King, J. A. Kratochvil, and W. E. Boyson, *Photovoltaic array performance model*. United States. Department of Energy, 2004.

[30] A. P. Dobos, "PVWatts version 1 technical reference," National Renewable Energy Lab, Golden, CO (United States), Tech. Rep., 2013.

[31] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovačević, "Adaptive graph filtering: Multiresolution classification on graphs," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, pp. 427–430, 2013.

[32] Y. Zhao, F. Balboni, T. Arnaud, J. Mosesian, R. Ball, and B. Lehman, "Fault experiments in a commercial-scale pv laboratory and fault detection using local outlier factor," in *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*. IEEE, pp. 3398–3403, 2014.

[33] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.