PV Array Fault Detection Based on Deep Neural Network

Wei Gao
Department of Electrical and Computer Engineering
University of Denver
Denver, CO, USA 80210
wei.gao@du.edu

2021 IEEE Green Technologies Conference (GreenTech) | 978-1-7281-9139-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/GREENTECH48523.2021.00018

***Abstract*—This paper develops an artificial intelligence based method to detect different types of faults associated with photovoltaic (PV) arrays. This method is integrated with the powerful deep neural network including multilayer perceptron (MLP) and one-dimension convolutional neural network (1-D CNN). To test and validate the proposed method, a PV system which can simulate typical line-line, line-ground, open-circuit faults is first modeled via Matlab/Simulink and large amounts of normal and fault data are simulated. Then, extensive simulation data are fed into MLP and 1-D CNN to learn the characteristics of different types of faults, and thus detect and distinguish those faults. Finally, the results have shown the high accuracy and effectiveness of the neural network based PV array fault detection method.**

***Index Terms*—fault detection, photovoltaic, neural network**

I. Introduction

Renewable energy-based power system is in rapid development these years due to its clean nature and ability to address issues of limited conventional fossil fuels. Moreover, Microgrid including renewable energy is being paid more and more attention. In many Microgrids, photovoltaic (PV) is one of the major components, which has to guarantee its reliable operation mainly focusing on two aspects: converter control performance such as grid-forming characteristics and fault right through (FRT) ability. To mimic the advantages of the conventional power systems, Microgrid needs to provide sufficient inertia to improve the dynamic performance [1]. To better identify fault and clear related issues, Microgrid needs to provide a fast and reliable method of fault detection is desired.

Fault detection has always been a formidable challenge in power systems. With a lot of renewable energy resources utilized in the distribution system, it becomes more and more difficult to recognize and locate the faults. For example, faults can be caused by many different reasons in PV arrays such as line-line, line-ground, open-circuit, partial shading, arc faults, etc, thus leading to the whole PV system generation reduction suddenly by more than half, or even worse.

Generally, PV faults can be classified as electrical faults: line-line, line-ground, open-circuit, arc faults and non-electrical faults: shading and some aging related faults. Different faults have different characteristics and require different detection methods [2]. The scope of this paper is mainly focused on the most common electrical faults: line-line, line-ground and open-circuit faults.

Since the abnormality will be first reflected on the voltage and current waveform, a lot of methods utilize the transient waveform characteristics based analysis methods to detect the faults, such as time-domain analysis, frequency domain analysis, time-scale domain analysis, time-frequency domain analysis [3]. However, the inconvenience of those methods are obvious and more related signal processing knowledge is required.

Some statistics based methods have been developed for a period of time. A descriptive and inferential statistics based approach is proposed in [4] where the output energy of the inverters is utilized to supervise and monitor the operation of PV plants. Implementation of this method is complex and the generalization ability is relatively weak. An online fault detection method of solar array based on the deviation of the maximum current value from MPPT and the fault current value is proposed in [5], in which accurate devices such as Hall sensors are required to obtain the corresponding electrical information and thus this method turns out to be expensive in cost.

Machine Leaning based methods have also been used in fault detection. Classical machine learning algorithms such as support vector machine (SVM) is utilized in [6]. However, this type of machine learning algorithm is not as good as latest neural networks due to the rapid development of computer hardware resources nowadays. In many research areas, neural network has turned out to be very powerful machine learning technique, which will be continuously innovated with more state-of-the-art structure.

In this paper, a reliable PV system model is built in Matlab/Simulink and large amounts of simulations have proved the accuracy of this PV system model. Then, multilayer perceptron (MLP) neural network structure is first applied for the PV array fault detection task and shows the efficiency and easy to use characteristic in detecting different types of PV array faults. Besides, by employing the one-dimension convolutional neural network (1-D CNN) with additional convolutional layers in front of MLP, which can extract the useful and important information more efficiently, the accuracy of the neural network based fault detection method is further improved.

The contribution of this paper is: (1) a novel deep neural

This work is supported by U.S. NSF under grant 1711951.

2166-5478/21/$31.00 ©2021 IEEE
DOI 10.1109/GreenTech48523.2021.00018

network-based method to detect different types of PV array faults; (2) a comprehensive analysis of accuracy and effectiveness of PV array faults detection using different deep neural network structures.

The rest of the paper is organized as follows. Section II introduces modeling of the typical PV system and PV system faults. In Section III, the fundamental background knowledge of MLP and CNN is discussed. Section IV gives a case study to show the prominent performance of this neural network based PV array fault detection method. Finally, the conclusions of this paper are given in Section V.

II. Fundamentals of PV System

A. Typical PV System

PV cell is the basic component to convert the sunlight energy to electricity. To take full advantage of solar power, PV modules are formed through series and parallel connection of numerous PV cells to achieve higher voltage and current. To obtain more power, PV modules are further connected to form PV arrays. However, the output power from PV arrays is influenced by many factors such as solar irradiation, temperature and system structures. Thus, maximum power point tracking (MPPT) technique is used to track the maximum power point of the PV arrays under complex and varying weather conditions. As DC sources, PV arrays can either directly transfer energy to DC load through DC-DC converter or convert DC power to AC power through DC-AC converter, which needs to ensure the operating point of PV arrays at the maximum power point through controlling the IGBTs within the converters. Fig. 1 shows a typical PV system for supplying energy to a DC load with a MPPT controller. PV system models are developed in Matlab/Simulink, with well-documented models for PV cell and array, converter, MPPT controller, and load [7]. In real life, PV array in Fig. 1 can be expanded to large scale and thus fault detection within those PV arrays is very challenging, and the fault details will be explained in the next subsection.

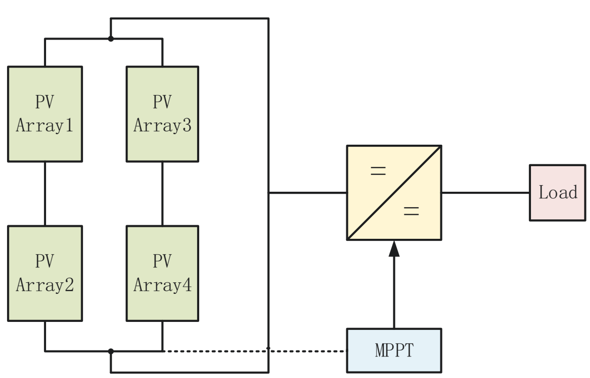


Fig. 1. Structure of typical DC-DC PV system

B. Typical Faults in PV System

As described, typical PV faults can happen within PV arrays in large PV power plant. In this section, three typical electrical faults in PV arrays are analyzed as shown in Fig. 2:

1) Line-line fault: Line-line fault usually happens between two neighboring arrays as a short circuit as shown in Fig. 2(a), leading to the increased current instantaneously in the string with faulted arrays. The conventional overcurrent relay uses this feature to detect the fault. However, the output power will drop as the voltage in the feeder will decrease.

2) Line-ground fault: Line-ground fault usually happens between the arrays and the ground because sometimes the conductor attached to PV arrays may be in contact with the ground unexpectedly. More frequently, it is caused due to the damage of the insulation materials in PV power plant. Fig. 2(b) shows a typical line-ground fault. The consequence is the lower voltage in this feeder, resulting in lower output power. The conventional overcurrent method can also be applied to this scenario. With the development of fault detection technique, some other methods appear without needing magnitude of the fault current such as the spread spectrum time domain reflectometry (SSTDR) [8].

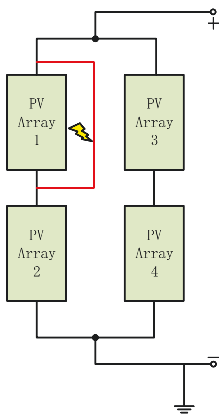
3) Open-circuit fault: Electrical devices are aging with time, especially for the breaker which links terminals of two PV arrays in PV system. It usually behaves as unstable connection between neighboring arrays. The typical open-circuit case can be seen in Fig. 2(c). Obviously, open-circuit fault will cause the whole system lose the entire power of the string. The output voltage of the system keeps at the same as normal condition; the output current reduces and thus the output power will decrease. Generally, conventional detection methods for this type of fault mainly focus on the time/frequency analysis of the transient waveform [9], which requires more advanced signal processing computing.

III. Fundamentals of Neural Network

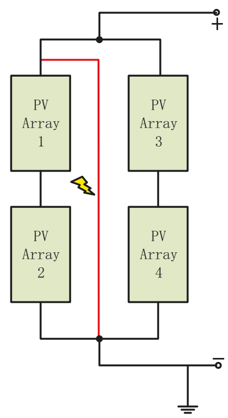
In recent years, neural network becomes more and more popular in many research fields due to the rapid evolution of computing technology and has shown its huge power in solving some formidable problems which are complex or even impossible using conventional methods. Two common neural network structures known as multilayer perceptron (MLP) and convolutional neural network (CNN) will be discussed in the next subsections and applied to the PV array fault detection problem.

A. Multilayer Perceptron (MLP)

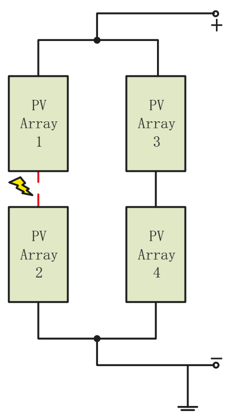
Multilayer perceptron is the basic deep neural network structure, which can be shown as in Fig. 3. MLP is powerful because it has many hidden layers consisting of perceptrons besides the basic input layer and output layer. To investigate the mechanism of MLP, it is necessary to start from signal neuron called perceptron. Each neuron works as a function,



(a) Line-line



(b) Line-ground



(c) Open-circuit

Fig. 2. Typical faults in PV system

taking inputs and generating outputs. A typical linear function operator of a neuron is:

$$f(x_i, ..., x_n) = \sum_{i=1}^{n} w_i * x_i + b \tag{1}$$

where x_i is input to a neuron at the hidden layers and output layer; n is the total input dimension of the layer; w_i is the weight of current neuron linked with all the input neurons; b is the bias. Weights are the cores of neural network because each neuron can generate different outputs based on those weights. It is obvious that the function is linear, so no matter how many layers the neural network has, it will behave as a single linear combination, leading to failure to complex tasks involving nonlinear systems. To make it more powerful to handle logical classification regression problems, additional nonlinear function also called activation function is used, which gives nonlinear fitting ability to the neuron. Up to now, many useful activation functions have been explored, and the most common ones are sigmoid, tanh and ReLU:

$$Sigmoid(x) = \frac{1}{1+e^{-x}} \tag{2}$$

$$Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

$$ReLU(x) = max(0, x) \tag{4}$$

In the example MLP shown in Fig. 3, there are 2 neurons in input layer, 4 neurons in hidden layer 1, 5 neurons in hidden layer 2 and 2 neurons in output layer. Each neuron has the linear function and nonlinear activation function as described, and all the neurons will generate all the way and transfer the corresponding output value to the next layer up to the output layer. This process is also called forward propagation. Assuming that the activation function for the hidden layers is ReLU, and for the output layer is sigmoid, the general output expression is:

$$\begin{aligned} \hat{y}_{i \atop i=1,...n_k} &= Sigmoid(\sum_{i=1}^{n_{k-1}} w_i * ...ReLU(\sum_{i=1}^{n_2} w_i * (\\ & ReLU(\sum_{i=1}^{n_1} w_i * x_i + b_1)) + b_2)... + b_{k-1}) \end{aligned} \tag{5}$$

where k is the number of layers; n_{k-1} is the number of neurons of the corresponding layer.

To obtain proper weights that can fit a nonlinear function better, the output of the forward propagation $\hat{y}_i$ is used to compare with the real output y_i of training data. First, typical mean square loss function between them is defined as:

$$L(y_i, \hat{y}_i) = \frac{1}{m} \sum_{j=1}^{m} (y_{ij} - \hat{y}_{ij})^2 \tag{6}$$

where m is the number of sample data. Based on the loss, different optimization methods can be applied into the neural network. A classical standard gradient descent (SGD) method is:

$$\bigtriangledown L = \partial \frac{L}{\partial x_i} \bigtriangledown x_i \tag{7}$$

This process is also called back propagation. With both forward propagation and back propagation, the neural network iterates again and again using large amounts of sample data until convergence in the training process.

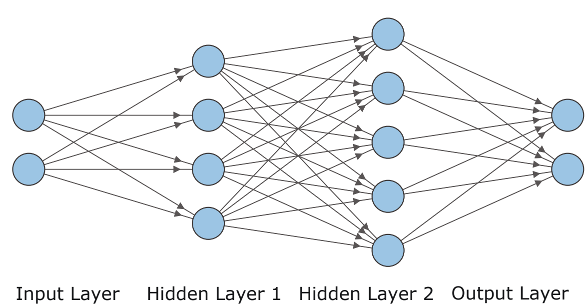


Fig. 3. Structure of Example Multilayer Perceptron (MLP)

B. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is used for image preprocessing when it was proposed firstly. With the basic MLP knowledge, CNN is added with a few convolutional layers in front of MLP (Dense) as shown in Fig. 4. In fact, the convolutional layers can be viewed as feature extractors. The input image will be sent to different channels and each channel will have many filters to expand the feature dimension. After each convolutional layer, there will be a pooling layer for reducing the expanded dimensions back to normal range and only keeping the most important information. Until last convolutional layer, all the channels' output will be converted to the normal fully connected layer for use of the next MLP. With the development of CNN, now the 1-D CNN is proposed and proven useful for extracting more important information in power system's 1-D waveform (time-scale waveform) instead of 2-D image.

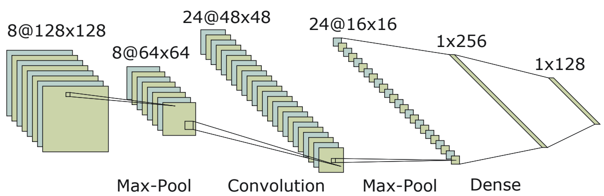


Fig. 4. Structure of Convolutional Neural Network (CNN)

IV. Case Study

A. Simulation Environment & Data Acquisition

a) Simulation Configuration:

In this paper, a DC PV system model with 4 arrays are simulated in Matlab/Simulink as shown in Fig. 1. For generalization, each PV array is in a different configuration as shown in Table I. The 4 PV arrays are connected together to transfer the maximum power to a DC load using MPPT controller. The simulation step for the power electronics is 1e-6 s and the total simulation time is 0.1 s. To make sure the sample data of the PV system model is reliable, one of simulation results under standard condition (Temperature: 25 °C, Solar irradiation: 1000 W/m^2) is shown in Fig. 5. The result shows that system can track the MPP signal from MPPT controller and output the maximum power quickly, and thus verifies the efficiency and accuracy of the MPPT control based PV simulation model.

b) Data Acquisition:

To obtain sufficient data for the training process of neural network, large amounts of simulations are executed via an automation tool. As the independent variables, solar irradiation of each PV array and the DC load resistance are set to be different values as shown in Table II. In this way, each type of fault has around 1875 sets of sample data, and the normal operation condition case has the same number of sets for sample data.

In this paper, only the output power waveform of the complete PV system is needed. However, if all the simulation data are exhibited as shown in Fig. 6, it is very challenging to classify those different types of faults. Before passing those sample data into neural network, some data preprocessing is necessary including disposing some bad/missing data, data normalization and data One-Hot Encoding. Disposing the bad/missing data is executed in the automation simulation tool to ensure integrity of the simulation data; the method of data normalization in this paper is converting the real output power to per-unit value, and thus making the value range of the inputs between 0 and 1, which is more appropriate for neural network training. Since the fault detection is a multi-classification task (needs to classify different types of faults, not only fault or normal case) for neural network, output of the neural network is a vector, in which each element represents one type of faults

TABLE I
Configuration of PV Arrays

Array #	Series[a]	Parallel[b]	Maximum power (W)
1	2	2	852.6
2	4	2	1705.2
3	2	3	1278.9
4	4	3	2557.8

[a]Series: Series-connected modules per string
[b]Parallel: Parallel-connected strings per array
Temperature: 25 °C
Solar irradiation: 1000 W/m^2

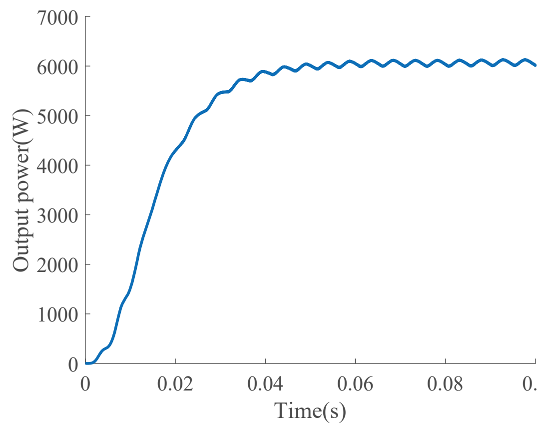


Fig. 5. Output power waveform of the PV system

TABLE II
Configuration of Dataset Acquisition

	Arr1	Arr2	Arr3	Arr4	Load	
Irra1	800	800	800	800	10	Res1
Irra2	900	900	900	900	20	Res2
Irra3	1000	1000	1000	1000	30	Res3
Irra4	1100	1100	1100	1100		
Irra5	1200	1200	1200	1200		

Arr: PV array number
Irra: Solar irradiation value (W/m^2)
Res: Resistance value (Ω)

or normal case. So One-Hot Encoding is appropriate to solve this problem. For example, in this paper, there are 4 types of outputs:

- Normal case - Labeling 0
- Line-line fault - Labeling 1
- Line-ground fault - Labeling 2
- Open-circuit fault - Labeling 3

By using One-Hot-Encoding technique, labeling of the data are set as a vector as follows:

- Normal case - Labeling [1 0 0 0]
- Line-line fault - Labeling [0 1 0 0]
- Line-ground fault - Labeling [0 0 1 0]
- Open-circuit fault - Labeling [0 0 0 1]

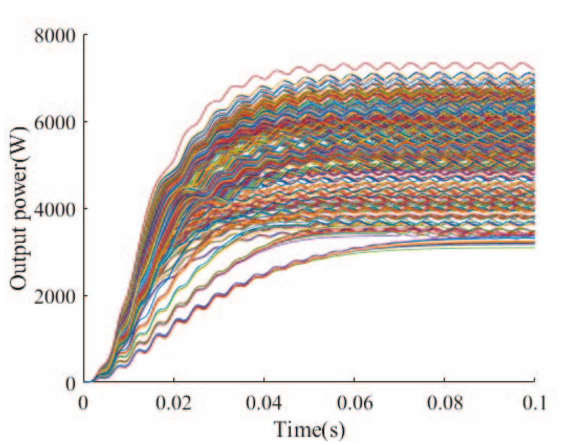


Fig. 6. Output power of the PV system under all the simulation cases

Since the simulation time step is 1e-6 s, there are 10e5 data points for entire 0.1 s simulation duration, meaning 10e5 dimensions for the input layer. However, on the one hand, not all the data information is necessary for training, leading to possible unexpected bad results due to so large data dimension; on the other hand, so large input dimensions are waste of computer resources, leading to slower training process. Therefore, the technique of taking one data point per 100 data points is applied to this dataset, which significantly reduces the input dimension to 1000.

Besides, the total dataset is split into three parts: training dataset, validation dataset and testing dataset with the corresponding ratio of 6:2:2. The purpose is that validation dataset can monitor the training process for preventing the neural network from overfitting. The testing dataset can verify the generalization ability of the trained neural network to those fault cases which have never been seen by the neural network.

B. *Results & Discussions*

In this paper, two different neural network structures (MLP and 1-D CNN) are utilized to validate the effectiveness of this neural network based method. The detailed settings of these two neural networks are listed in Table III.

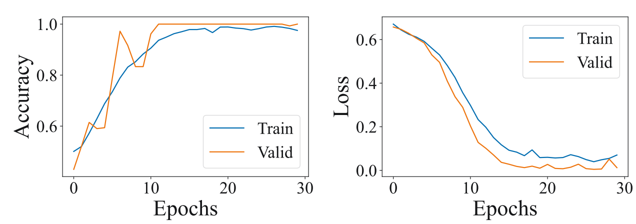
Correspondingly, results of the accuracy and loss during training process for MLP and 1-D CNN are shown in Fig. 7 and Fig. 8 respectively. The blue curves represent the accuracy and loss of training dataset, and the red curves represent the validation dataset for making sure that the training process of neural network is in the right direction. It can be seen that the training curves and validation curves are close to each other without a big gap, indicating that the neural networks are well trained. Meanwhile, both neural network structures can handle the fault detection task with high accuracy close to 100% with low loss near to 0. Furthermore, with additional convolutional layers as excellent feature extractor, the 1-D CNN has higher accuracy, lower loss and faster convergence speed than MLP as expected.

TABLE III
CONFIGURATION OF NEURAL NETWORK STRUCTURE

	MLP	1-D CNN
# of Neurons in Convolutional 1	/	80
# of Neurons in Convolutional 2	/	160
# of Neurons in Hidden 1	500	500
# of Neurons in Hidden 2	300	300
# of Neurons in Hidden 3	200	200
# of Neurons in Output	4	4
Activation in Convo/Hidden	ReLU	ReLU
Activation in Output	Softmax	Softmax
Dropout Rate	0.4	0.4
Loss	Category	Category
Optimizer	Adam	Adam
Learning Rate	0.0001	0.0001
Batch Size	32	32
Epochs	40	30

Category: Categorical Cross-entropy

To explore the generalization ability of the well-trained neural networks, the testing dataset which has not been seen by the neural networks are fed into them, and some evaluation criteria which are authoritative for evaluation of neural networks such as accuracy, precision, recall and F1 score are listed in Table IV. Therefore, neural network based PV array fault detection method is demonstrated to be highly effective.

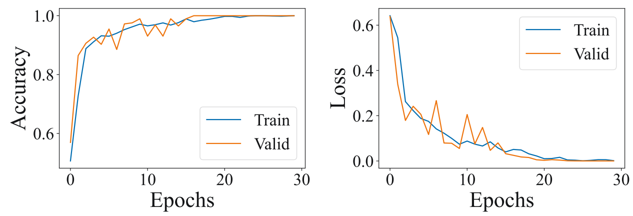


(a) Traning accuracy (b) Training loss

Fig. 7. Training accuracy and loss of MLP for PV array fault detection

V. CONCLUSIONS

This paper presents a neural network based PV array fault detection approach, in which the PV system with 4 arrays are supplying maximum power to a DC load under varying solar irradiation and DC load resistances. Typical electrical PV faults such as Line-line fault, Line-ground fault, and Open-circuit fault are simulated. Then, two different neural networks



(a) Traning accuracy (b) Training loss

Fig. 8. Training accuracy and loss of 1-D CNN for PV array fault detection

TABLE IV
EVALUATION CRITERIA OF MLP AND 1-D CNN UNDER TESTING DATASET

	MLP	1-D CNN
Accuracy	97.16%	99.12%
Precision	98.35%	99.65%
Recall	96.27%	98.73%
F1 score	97.38%	99.36%

(MLP and 1-D CNN) are applied into this PV system model to detect and distinguish different types of faults. Extensive testing sample cases are fed into the well-trained MLP and 1-D CNN to demonstrate the accuracy and effectiveness of this deep neural network based PV array fault detection method.

ACKNOWLEDGMENT

This work is supported by U.S. NSF under grant 1711951.

REFERENCES

[1] W. Gao. Microgrid control strategy based on battery energy storage system-virtual synchronous generator (bess-vsg). In *2020 IEEE Kansas Power and Energy Conference (KPEC)*, pages 1–6, 2020.
[2] 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*, 5(3):982–997, 2015.
[3] Mark Adamiak, Craig Wester, Manish Thakur, and Chuck Jensen. High impedance fault detection on distribution feeders. *GE Industrial solutions*, 2006.
[4] S. Vergura, G. Acciani, V. Amoruso, G. E. Patrono, and F. Vacca. Descriptive and inferential statistics for supervising and monitoring the operation of pv plants. *IEEE Transactions on Industrial Electronics*, 56(11):4456–4464, 2009.
[5] Huang Zhiqiang and Guo Li. Research and implementation of micro-computer online fault detection of solar array. In *2009 4th International Conference on Computer Science Education*, pages 1052–1055, 2009.
[6] Hong-Chan Chang, Shang-Chih Lin, Cheng-Chien Kuo, and Hao-Ping Yu. Cloud monitoring for solar plants with support vector machine based fault detection system. *Mathematical Problems in Engineering*, 2014, 2014.
[7] Wei Gao. Control strategy for a small-scale mircrogrid based on battery energy storage system-virtual synchronous generator (bess-vsg). *Thesis, University of Denver*, 2019.
[8] M. K. Alam, F. H. Khan, J. Johnson, and J. Flicker. Pv arc-fault detection using spread spectrum time domain reflectometry (sstdr). In *2014 IEEE Energy Conversion Congress and Exposition (ECCE)*, pages 3294–3300, 2014.
[9] X. Yao, L. Herrera, S. Ji, K. Zou, and J. Wang. Characteristic study and time-domain discrete- wavelet-transform based hybrid detection of series dc arc faults. *IEEE Transactions on Power Electronics*, 29(6):3103–3115, 2014.