




Received 30 November 2023, accepted 16 December 2023, date of publication 25 December 2023, date of current version 5 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3347199

RESEARCH ARTICLE

A Scalable Random Forest-Based Scheme to Detect and Locate Partial Shading in Photovoltaic Systems

ZAIN MUSTAFA¹, MAHER A. AZZOUZ¹, (Senior Member, IEEE), AHMED S. A. AWAD¹, (Senior Member, IEEE), AHMED AZAB⁴, AND MOSTAFA F. SHAABAN¹, (Senior Member, IEEE)

¹Department of Electrical and Computer Engineering, Faculty of Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada

²Department of Electrical Engineering, College of Engineering, Qatar University, Doha, Qatar

³Department of Electrical and Computer Engineering, College of Engineering, Sultan Qaboos University, Muscat 123, Oman

⁴Production and Operations Management Research Laboratory, Faculty of Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada

⁵Department of Electrical Engineering, Faculty of Engineering, American University of Sharjah, Sharjah, United Arab Emirates

Corresponding author: Ahmed S. A. Awad (a.awad1@squ.edu.om)


This work was supported in part by the Open Access Program from the American University of Sharjah.

ABSTRACT Photovoltaic (PV) systems are prone to partial shading (PS) due to the environmental factors that they function in such as vegetation, nearby structures, and clouds. All types of PS scenarios can lead to power loss and hot spots in the PV system due to module mismatch and heating of shaded cells. To mitigate the power loss that occurs due to PS, it is imperative to detect PS and its characteristics, such as the number of shaded modules and the associated shading factor (SF), in a reliable manner. This paper proposes a three-step framework to detect and locate PS, the number of shaded modules, and the SF in the PV system using a random forest (RF)-based approach. The proposed approach utilizes independent string current and voltage measurements to distinguish different PS scenarios. This approach allows for a scalable data acquisition through an uncoupled modeling scheme. PS, the number of shaded modules and the SF are deduced with accuracies of 99.5%, 92.3%, 90.2%, respectively. Further, the proposed approach is validated through two testing tiers, and its ability to detect multiple PS scenarios in a PV system has been highlighted. The results observed through different PS scenarios confirm the high reliability and demonstrate the effectiveness and scalability of the proposed RF-based approach.

INDEX TERMS Photovoltaic faults, partial shading, random forest, maximum power point tracking.

I. INTRODUCTION

The impact of increasing energy demands due to the rapid technological advancements has been globally realized, forcing the industry to focus on practices that could help mitigate the issues that result due to high energy demands [1]. Energy generation through renewable sources such as solar and wind are gaining heightened attention to combat the environmental repercussions of conventional fossil energy generation. These sources have seen a very high adoption

The associate editor coordinating the review of this manuscript and approving it for publication was Giambattista Gruosso¹.

rate in multiple global markets due to concerns surrounding conventional energy generation, especially solar energy. However, solar photovoltaic (PV) systems are developed in open and uncertain environments, which makes them prone to different abnormalities and vulnerabilities. Vulnerabilities can cause a decrease in power output and, if not detected, can cause the system to perform at a subpar level constantly. Partial shading (PS) is one of the vulnerabilities that every PV system is prone to, whether implemented in an open space or urban area. Hotspots generated when PS occurs can cause PV modules to be irreversibly damaged, impacting the power output of the system. Conventionally, PS is mitigated

by bypassing shaded modules through bypass diodes; this decreases the negative impact of PS on the PV system's power output [2].

Even with traditional mitigation approaches to decrease power loss due to PS, it is imperative to detect PS through a dependable monitoring approach. Conventionally, various electrical configurations and variations in voltage measurements have been proposed to address power loss due to PS, but these approaches can still be prone to errors since they depend on unique sensor placements [1], [2], [3], [4], [5]. In recent studies, researchers have employed various artificial intelligence (AI) techniques to detect PS as the first step in decreasing its impact on the system's power generation. These machine learning (ML) techniques aim to detect PS by analyzing data attained from running multiple PS simulations on a PV array model. The importance of early diagnosis of anomalies found in PV arrays has been consistently reinforced by different researchers. Since PS occurs in all PV systems, its early diagnosis can help initiate any measures taken to resolve any degree of effect it has on the system's power output [2].

Studies on different ML and deep learning (DL) algorithms have been carried out to detect PS in a PV array [8]. Literature that suggests these approaches depends on simulating a complete PV system through irradiance and temperature inputs to acquire data used to train these algorithms. Machine learning techniques such as Random Forest (RF) [9], [10], [11], [12], k-Nearest Neighbor [13], and support vector classification [4] have been considered to detect PS. There has been an effort to simply diagnose a PS as an anomaly on its own among other faults that occur throughout the PV array. The bulk of the ML approaches found only differentiate between PS and other faults, i.e., string-to-string and string-to-ground, in PV systems. Even though these algorithms can detect PS, they do not predict the shading factor (SF), or the number of modules shaded in each scenario. References [5], [6], [7], [8], [9], [10], and [11] have only considered PV arrays of small size to help accelerate the simulation and data acquisition phase; However, this assumption does not account for how the algorithm would react when dealing with a larger system.

The ease of access to incredible computing power and advanced libraries have allowed researchers to avail and improve these algorithms, in context of PS detection, at an incredible rate. These advances and increased computing resources have led the community to employ various DL techniques to detect PS. Recent literature has focused on detecting PS through general artificial neural networks [14], [15], convolutional neural networks (CNN) [16], residual neural networks (RNN) [17], and probabilistic neural networks [18], [19]. Like the conventional ML approaches, DL approaches are simply aiming at detecting PS and do not emphasize the detection of the associated SF and the number of shaded modules. Further, the other approaches have mainly focused on interpreting PS to differentiate it from faults found within the system. These approaches do

not determine the characteristics associated with the different PS scenarios. Among the recent efforts, DL algorithms are utilized more often with the expectation that they would outperform conventional ML algorithms. However, that does not hold true for all problems; the performance of ML algorithms varies depending on the application, dimensionality, and size of the dataset [20]. Considering these factors, RF tends to perform better on average than neural networks, specifically when the size of the dataset is a considerable bottleneck [21], [22].

Early detection of PS is of high importance to monitor and supervise a PV array, and to eliminate the need for predictive maximum power point tracking (MPPT) algorithms. PS caused due to cloud cover can cause up to 77% reduction in power output [23], leading to a major setback in the overall performance of the system during its lifetime. When AI techniques are deployed, a lot of initial preparation goes into building and training a model to work for a PV system. Previous studies found in the literature are carried out by simulating complete systems and, as such, their performance relies on the system that the algorithms are trained for. Decreasing the time needed for simulations and classifier training is a concern of great importance if AI techniques are to be adopted by large-scale PV arrays [24]. Since the current approaches are used per system, they are unable to detect if the shading ratio differs over different regions of the PV array, deriving a single generalized label to encompass the effect of PS over the PV array. Moreover, most of the previous approaches do not solve the problem of when partial shading occurs at different irradiance levels across the PV array and detecting the number of modules shaded [25]. The findings from recent major studies have been summarized and compared against the proposed approach in Table 1.

This paper proposes an approach that is carried out on a single string of the PV array to analyze PS, SF, and the number of shaded modules. Modeling and training one string allow this approach to be scalable, and hence, be applicable to any PV array that utilizes the same number of modules per string. To allow for this scalability, the MPPT is decoupled from the detection of PS. Thus, the data acquisition from one string is representative of the response of other strings. The proposed approach suggests using an RF algorithm on a comprehensive dataset acquired through automated simulation of a PV array modeled using PSCAD/EMTDC. The proposed approach can efficiently classify PS scenarios, SF, and the number of modules shaded with high accuracy of 99.5%, 92.3%, and 90.2%, respectively. For the purpose of this study, the SF is considered as the decrease in irradiance due to PS.

The contributions of this paper are as follows:

- A three-step approach is proposed to detect PS, the number of shaded modules, and the SF by analyzing string voltage and current measurements by adding a voltmeter and ammeter at the beginning of each string. Each category is further divided into classes, ensuring a detailed understanding of the PS scenario.

TABLE 1. Synthesis matrix comparing recent major studies with the proposed approach.

	Detect partial shading	Detect shading factor	Detect Shaded Modules	Locate shaded strings
Proposed approach	✓	✓	✓	✓
[4]	✓	X	X	X
[5]	✓	X	X	X
[6]	✓	X	X	X
[7]	✓	✓	X	X
[9]	✓	X	X	X
[10]	✓	X	X	X
[11]	✓	X	X	X
[12]	✓	X	X	X
[15]	✓	X	X	X

- A system architecture is proposed, that uses PSCAD /EMTDC to simulate the PV system, for data acquisition in an isolated manner by simulating a single string of the system instead of the complete 10 × 10 testbed. PS on each string is independently analyzed to capture the different levels of shading occurring throughout the system.

The rest of the paper is organized into six sections. The system description and effects of PS on the PV system are provided in Section II. The proposed algorithm is elaborated on in Section III. Data acquisition and the proposed methodology are explained in Section IV. Results obtained through different case studies are detailed in Section V. Sections VI and VII include a discussion on the results observed through testing and the conclusion, respectively.

II. PARTIAL SHADING IN PV ARRAYS

In this section, the proposed model as well as the impacts of partial shading on the PV systems are discussed.

A. SYSTEM DESCRIPTION

The model used for testing and validation is created using PSCAD/EMTDC and consists of a PV array with a central inverter topology. As shown in Fig. 1, the PV modules of the PV array are connected in parallel, leading to a central inverter. Each string is attached with a voltmeter and ammeter to capture voltage and current measurements. The PV array also includes blocking diodes on each string to prevent current circulation through the strings. The central inverter consists of two back-to-back converters: a DC-DC boost converter to track the system’s maximum power point (MPP) using an observe and perturb method and a DC-AC converter to regulate the DC-link voltage.

Fig. 1 illustrates a 10 × 10 PV array testbed that consists of ten parallel strings and each string contains ten series modules. Each module contains 60 cells connected in series and a bypass diode with a 0.4V drop. I_{Mi} is the measured current of the string, V_{Mi} is the measured voltage of the first module at each string, and V_{PV} and I_{PV} denote the total PV voltage and current, respectively. The focus of this paper is to develop a scalable detection algorithm to identify PS, the number of shaded modules, and the SF using the voltmeter and ammeter placed at the beginning of each string, as shown in Fig. 1.

The parameter specifications of the PV module under standard test conditions (i.e., 1000W/m² irradiance, 25°C temperature, and 1.5 air mass) are: open-circuit voltage $V_{OC} = 37.2V$, maximum power voltage and current are $V_{mp} = 29.8V$ and $I_{mp} = 7.8A$, respectively, and short-circuit current $I_{sc} = 8.48A$. The full specifications of the 235-W BP3 PV modules used for data collection are in [26].

B. EFFECTS OF PARTIAL SHADING

PS can be caused by various objects such as trees, surrounding structures, clouds, and debris, including snow, sand, and dirt [27]. When PS occurs, the PV modules affected receive a lower level of irradiation compared to the modules under normal circumstances. Since the modules per string are attached in a series formation, the magnitude of current through each module must be consistent. However, when a module within the series is shaded, it cannot produce the same magnitude of current as the other modules and, in turn, starts acting as a load [28]. Due to the difference in current, these shaded cells tend to cause hot spots in the PV system. Bypass diodes are implemented parallel to each module to mitigate hotspots caused due to PS and prevent reverse bias losses on modules affected by PS. A PV module without a bypass diode will experience a significant power loss under PS if reverse bias losses are not mitigated, a bypass diode prevents this from happening by providing a path for the current to flow around the shaded cells.

In PV modules connected in series formation per string, the voltage between the two modules defers while the current stays the same; however, in PS scenarios, the current of the string is bottlenecked by the shaded module. This affects the $I - V$ and $P - V$ characteristics of the series-connected modules [29]. The two characteristics of value, i.e., SF and the number of modules shaded, have an impact on the $I - V$ and $P - V$ curves of the PV array. Fig. 2 displays this impact when different numbers of modules are shaded by an SF of 50%.

PS affects both characteristics even when the number of modules increases even with a consistent level of shading. Variability in the SF has a similar result. Local maxima occur when the SF increases and the number of shaded modules is consistent. Fig. 3 presents the $I - V$ and $P - V$ characteristics when six modules in the PV strings are affected by different SFs. Considering that $I - V$ and $P - V$ characteristics are

impacted by both, the number of modules shaded and the SF, deriving both will help understand the optimal power output of the system.

Figs. 2 and 3 show the impact of PS on both the I-V and P-V curves; however, to successfully extract the correct features from the system for this study, it is important to understand its impact on the system voltage and current. Fig. 4 compares the impact on the system current when a PV module is shaded on a single string versus when the shading occurs on two strings. A higher drop in system current is seen when more than one string has a partially shaded module. Considering the impact that PS has on the system current, it is important to isolate the system and string parameters to reliably deduce PS scenario in an isolated manner. PS leads to a noticeable drop in power generation, Table 2 shows the percentage drop in power for a given shading factor.

TABLE 2. Percentage drop in power under stc for a given shading factor and shaded modules.

# of Modules Shaded	Drop in Power (%)	
	20% Shaded	80% Shaded
1 & 2	6.55	16.42
3 & 4	7.14	22.81
5 & 6	7.70	23.04
7 & 8	8.32	23.61
9 & 10	9.25	24.55

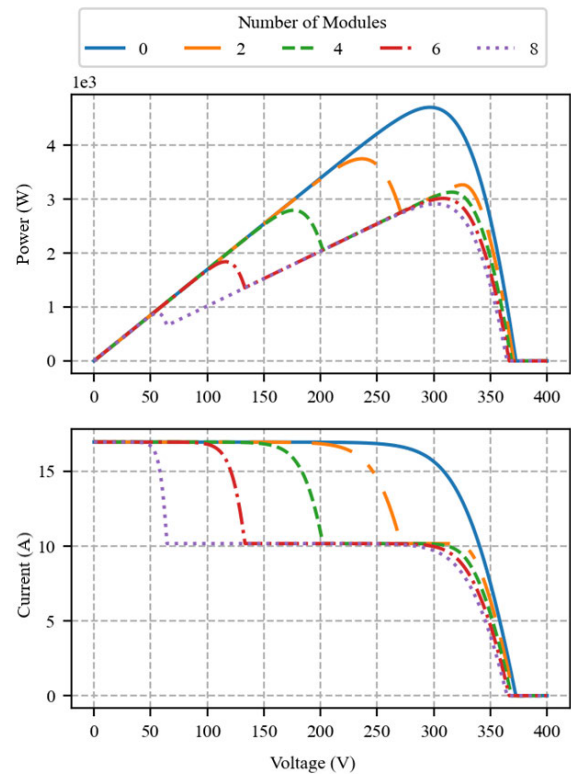


FIGURE 2. P-V and I-V curves when different numbers of modules are affected by PS.

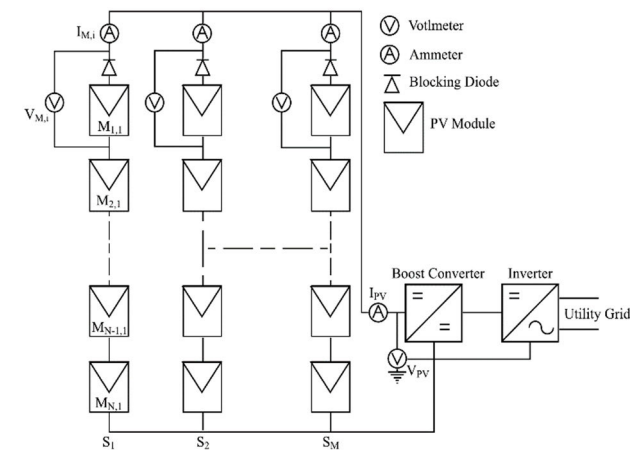


FIGURE 1. 10 × 10 PV array with a boost converter.

III. RANDOM FOREST ALGORITHM

RF algorithm is a learning ensemble technique created by generating multiple decision trees (DTs). A model based on a tree structure, shown in Fig. 5, represents a DT, where each node represents a feature, each branch a decision, and each leaf an output value. RF develops the DTs using resampled data from the original dataset and utilizes a random feature selection to decrease the correlation in the generated DTs. This resampling is done by bootstrapping randomly selected samples from the original dataset to produce a new dataset

of the same size to build each DT, and random subsets of features are considered when splitting at each node. The output generated by RF is derived by majority voting from the generated DTs. RF applications have a few drawbacks. For instance the trained algorithms may be time-consuming when making predictions on datasets that contain a large number of features or when training with datasets that have missing features [30]. However, the proposed approach does not face either of the drawbacks because when classifying each label, data from a small window is used for predictions and each sample extracted has a small number of features for training and testing.

Neural networks (NN), such as convolutional neural networks (CNNs), recursive neural networks (RNNs) and long short-term memory networks (LSTM), have been gaining a lot of momentum in recent years due to their ability to combine feature extraction and classification into a single process. However, there is no single algorithm that works optimally in all scenarios. RF algorithm can be used to train a reliable classifier using a relatively small dataset compared to other ML/DL techniques. They also perform better due to their ability to train quickly when working towards optimizing hyper parameters [20]. When considering an algorithm for a certain problem, it is important to look for factors such as performance, robustness, comprehensibility, and cost and time required for training the algorithm. Studies conducted in [20], [21], [22], and [31] compare the performance of

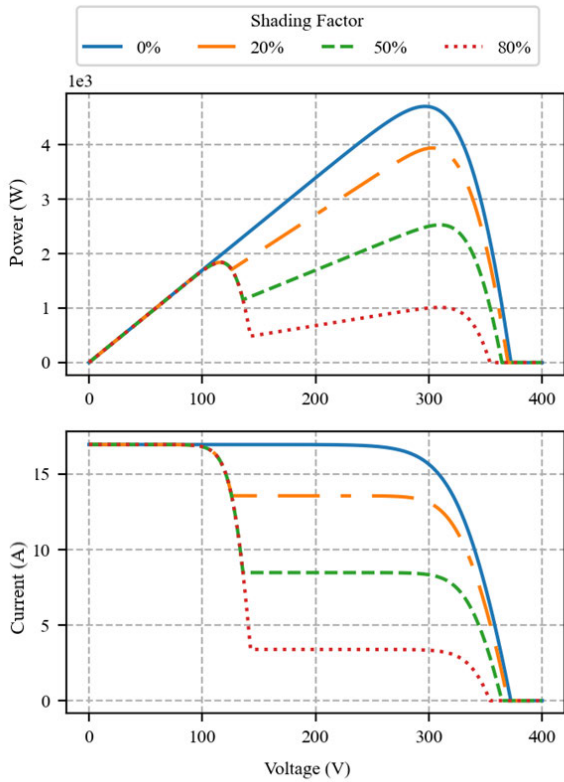


FIGURE 3. P-V and I-V curves of different SF affecting 6 modules per string.

RF and NN algorithms, concluding that in the majority of the cases, RF tends to have an advantage over NN. In this paper, RF took precedence due to its ability to outperform NN algorithms due to the size of the dataset and the available features.

In the random forest algorithm, forest F is an ensemble of decision trees where:

$$F(\theta) = \{h_m(\theta_m)\}, \quad m = 1, 2, \dots, M \quad (1)$$

where M is the total number of decision trees, θ represents the parameters in F , and h_m is the decision tree model where:

$$h_m(x) = \sum_{j=1}^J b_{jm}I \quad (2)$$

where J is the number of leaf nodes, b_{jm} is the predicted value for node jm . The proposed algorithm has been configured to consider a minimum number of 41 leaf nodes per decision tree. Each tree considers a subset of randomly selected samples to help reduce the correlation between the trees of the forest [30].

IV. PROPOSED METHODOLOGY

The proposed approach revolves around uncoupling a string from the system for data acquisition and training the RF algorithm. The voltage across the PV array is kept constant during the RF training and the algorithm application

to decouple the MPPT from the detection of partial shading. This decoupling makes the data acquisition from one string representative of the response of other strings. Notably, the number of modules simulated on this string needs to be consistent with the number of modules expected to be per string on the PV array testbed. Fig. 6 shows the model of the decoupled string used for the simulation of the PS scenario. This uncoupling enables the simulation of the PV array of a variable number of strings since the PS scenario is deduced via the data acquired by a single PV module string in the system. Since the number of strings that needs to be simulated does not increase with the system size, this approach is highly scalable and time-effective. The proposed approach is validated through a 10×10 PV system, however, it is applicable to a system of any size. A change in system size would require simulations to be carried out using the string of the system being used. For example, for a 5×5 system, the simulations would be carried out on the first string that contains five PV modules instead of the ten being used for the test bed. The model presented in Fig. 6 is used to generate a dataset of 1,425,948 unique samples, which the RF classifier is trained on to derive if the system is under PS, the number of modules affected by PS, and the SF. The acquired data goes through the flow summarized in Fig. 7 for classifier training and testing, detailed in Section IV-A.

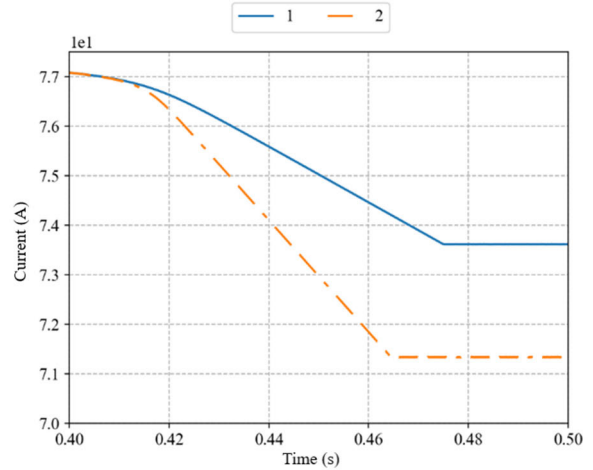


FIGURE 4. Impact on system current when the one module is shaded on different strings of the system.

A. DATA ACQUISITION

The modeled PV array is simulated by PSCAD/EMTDC is used to assess the proposed approach under PS scenarios, deriving the number of shaded modules and the SF from the data. The string current and voltage, irradiance, and temperature are captured as features from the simulated system for each time step. The data collected has been considered with a breadth of variance, considering PS scenarios for pre-shaded irradiance and temperature

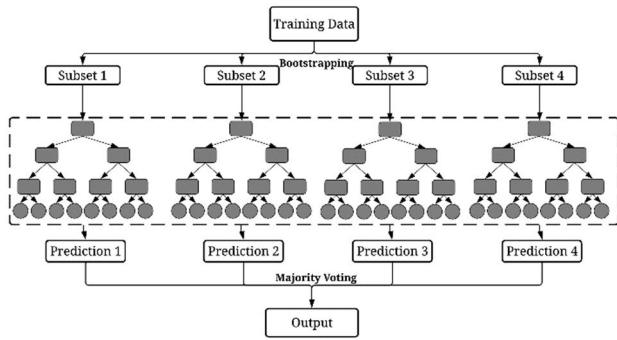


FIGURE 5. Random forest algorithm.

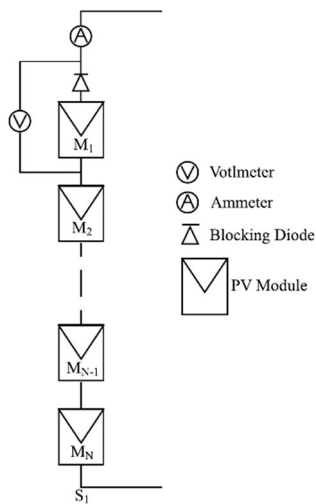


FIGURE 6. Model of the string extracted from the complete model for data acquisition.

values between $250 - 1000 \text{ W/m}^2$ with a time step size of 250 W/m^2 and $0 - 50^\circ \text{C}$ with a step size of 25°C , respectively. A smaller step size is selected for irradiance changes because partial shading is mainly dependent on irradiance changes, while the larger step size selected for the temperature is to cover a wide range of temperatures seen in different ambient conditions. The step size and the range of these values ensure that the simulated system covers a range of conditions that a real-time system may experience. This data is only extracted from the first string of the system during each scenario, enabling the diagnosis of PS and determining the number of modules shaded, and the SF on any string in the testbed. During each scenario, the string is split into two halves and a module is selected from each. The modules selected from the top half of the string stay consistent throughout the active scenario while the modules from the bottom half are incremented for each subsequent iteration. This method of data acquisition addresses all combinations of shaded modules occurring anywhere on the string. For example, to simulate a scenario under standard testing conditions (i.e., 1000 W/m^2 irradiance, 25°C temperature), a module is selected from the top half of the string for

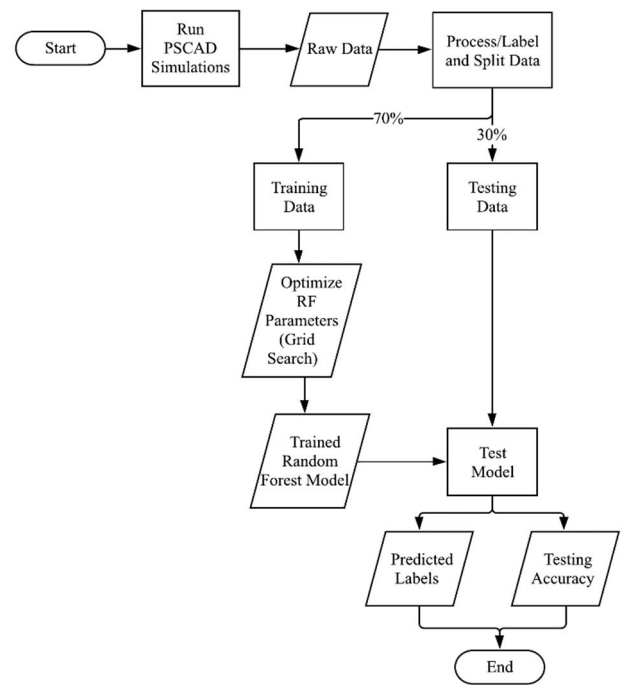


FIGURE 7. Flowchart of training and testing RF classifier to detect PS in PV arrays.

the first iteration, signifying that one module on the string is shaded. The second iteration then adds the last module from the bottom half and each subsequent iteration will add a module from the bottom half of the string until module 6 is reached. After module 6 the modules on the bottom half of the string are reset back to 0 and the modules on the top half are incremented by one. This process is repeated for each combination of irradiance, temperature, and shading factor values.

The data is acquired through detailed PSCAD/EMTDC simulations, which are well accepted in the industry and can accurately mimic the accuracy of real-world scenarios [3], [13], [25]. In addition, the proposed random forest algorithm is very robust when training on noisy data sets [20], [21], [31]. The experimental verification of the proposed approach was out of scope of this research work, due to limited resources, and it can be carried out in future work.

To encompass all the scenarios that could occur due to the variance in how PS occurs, consideration has been given to the level of SF and the number of shaded modules in the simulation. For each of the pre-shaded irradiances, simulation is carried out with 20% and 80% SF. To ensure that the RF model can accurately infer the number of shaded modules, the 20% and 80% shading configurations are applied to encompass shading occurring from one module to the string being completely shaded. After the simulations have been carried out, each sample is labeled under three categories: PS, modules shaded, and SF. These categories are then divided into classes. The PS category is divided into two classes: Normal and PS, where PS denotes the shaded sample.

Shaded modules are categorized into sets of two modules for each string, e.g. for a string of 10 modules, it is categorized into five classes. The shaded modules category contains the following classes: (1 & 2), (3 & 4), (5 & 6), (7 & 8), and (9 & 10). Each class represented in the shaded modules category signifies the range of modules that are shaded. The SF represents the drop in irradiance for each sample and is divided into two classes: 80% and 20%. These categories and classes are summarized in Table 3. The number of strings in the system does not impact the label classification for this approach; however, as the number of modules per string increases, the label used for the number of shaded modules would vary.

However, the raw data collected through the simulations is highly unbalanced. Since the simulations repeatedly run from the beginning, there are duplicated samples of the normal class generated for the initial time steps for each iteration; these duplicates tend to skew the dataset and have to be removed. Once the duplicates have been removed through data cleanup, the rest of the classes are balanced using the synthetic monitoring over-sampling technique (SMOTE) [32]. This technique synthesizes new samples of the minority class to build equal proportions of each class in all three labels. A balanced dataset is imperative to accurately measure and compare the efficiency of the proposed approach.

B. PARTIAL SHADING DETECTION ON 10 × 10 TESTBED

The proposed approach detects PS, the number of shaded modules, and SF in two distinct steps. In the first step, the RF classifier simply detects whether the system is functioning normally or whether there is a shaded module present. If PS is detected, the second step uses separate classifiers to detect the number of modules shaded, and the SF. Fig. 8 illustrates a flowchart outlining the detection process. The detection process is initiated once PS is detected through the process outline in Section IV-A, retrieving a subset of voltage and current measurements from the PV array and diagnosing the number of shaded modules and the shading factor. This process iterates over all the strings in the system and detects if any of them are partially shaded.

TABLE 3. Label classification for the dataset.

Category	Classes
Partial Shading	Normal, PS
Modules Shaded	(1 & 2), (3 & 4), (5 & 6), (7 & 8), (9 & 10)
Shading Factor	80%, 20%

V. PERFORMANCE EVALUATION

The performance evaluation of the RF classifier has been carried out through analyzing the system under the

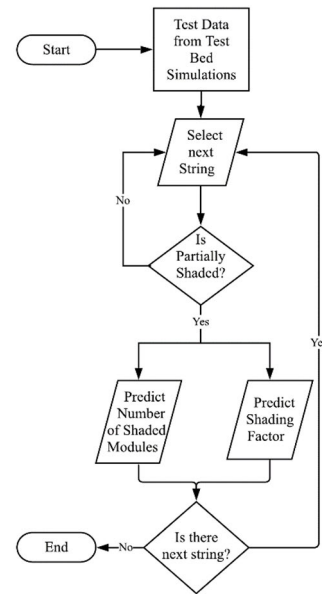


FIGURE 8. Process flow of PS detection in a 10 × 10 testbed simulation.

PS conditions, deriving the number of shaded modules and the SF, by acquiring the string current, system voltage, irradiance, and temperature from the PV array. The PV system, displayed in Fig. 1, is modeled and simulated using PSCAD/EMTDC, running on an AMD Ryzen 9 5900X processor at 3.70 GHz with 32 GB of RAM. An automation tool is employed to run the PS simulations using the PSCAD automation library and Python 3.x. The proposed algorithms are implemented using Scikit-learn libraries [33].

A. EVALUATION METRICS

The performance of the trained RF classifier is verified using five independent runs. These runs are carried out by splitting the dataset into a 70-30 split, using 70% of the simulated data to train the classifier and 30% to test the trained classifier. In addition to the test data acquired from simulating the uncoupled string to the PV array testbed, further performance evaluation of the classifier is carried out on an independent simulation of the PV array testbed.

This latter verification gauges the performance of the trained classifier on simulations conducted on multiple strings of the PV array testbed instead of just the initial string that the training data is acquired from.

The performance of the proposed RF algorithm is determined using accuracy, which has been the predominant metric to compare various studies in the literature [10], [16], [19]. While accuracy is considered the primary metric for this study, underlined metrics such as precision and specificity are also considered. These metrics are evaluated using the numbers of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) deduced through our 30% held-out data. For the proposed approach, accuracy (A)

is the number of samples that are classified correctly, precision (P) is the number of classified examples that are relevant, and specificity (S) is the number of negative samples that were correctly classified as negatives.

$$A = \frac{TP + TN}{TS} \tag{3}$$

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$S = \frac{TN}{TN + FP} \tag{5}$$

B. TESTING AND EVALUATION

The testing for the trained RF classifier is carried out on data in two separate tiers: 1) the 30% data held out from the data acquired through the PSCAD simulation; 2) a variation of independent PS scenarios carried out on the 10 × 10 PV array testbed. These variations are as follows:

- 1) *PS1*: 4 Modules are shaded on the 2nd string by 80%.
- 2) *PS2*: 2 Modules are shaded on the 4th string by 80%.
- 3) *PS3*: 10 Modules are shaded on the 6th string by 20%
- 4) *PS4*: a combination of PS1, PS2, and PS3 to detect multiple PS scenarios on a single PV array.

1) TIER 1–30% HELD-OUT DATA FROM TRAINING DATA

The performance of the classifiers is deduced through the 30% held-out data from the initial training dataset, and the average accuracy is captured over five separate runs for each label. The algorithms are tasked with deducing three different labels for any given sample: PS fault, the number of modules shaded, and the SF. Fig. 9 summarizes the comparison between RF, CNN, and LSTM algorithms for the simulated dataset. Both CNN and LSTM algorithms tend to perform well when detecting PS, but they are not as good as RF when detecting the number of modules shaded or SF. The accuracy of the RF algorithm is observed to be 99.5%, 92.3%, and 90.2% for PS detection, number of modules shaded, and SF, respectively. Whereas the performance of CNN and LSTM algorithm did not improve over 85.8% and 88.3%, respectively, when detecting the number of modules shaded. Considering the lagging accuracy of the CNN and LSTM algorithms, Tier 2 evaluations are carried out using the RF algorithm.

A confusion matrix for each output category is shown in Figs. 10 – 12. Each matrix represents the accuracy for each label and can be used to deduce TP, TN, FP, and FN samples to derive the precision and specificity values. The RF algorithm is compared against LSTM and CNN algorithms. It is important to note that the results of each matrix are independent of the other deductions. The classifier can reliably detect PS and deduce the number of shaded modules per string and SF. Fig. 11 further clarifies the labels that the model is unable to classify properly and have impacted the metric negatively. The bulk of these occurred when four or six modules are shaded in the simulation. The bulk of the false predictions made by the proposed RF algorithm are

seen when the samples between the ranges 3 to 4 and 5 to 6 modules are confused, where 5.4% and 6.3% of the samples from 3 to 4 class and 5 to 6 Mod, respectively, are classified incorrectly as the other label.

However, the outer labels of the matrix are deduced very reliably, displaying minimal errors. PS and SF are both inferred at a high accuracy as well and due to the lower number of classes for each category, the classifier can cleanly differentiate between each class in those categories.

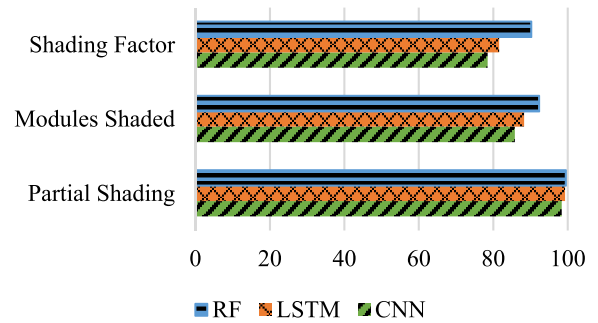


FIGURE 9. Accuracy of CNN, LSTM, and RF algorithms tested with 30% held-out data.

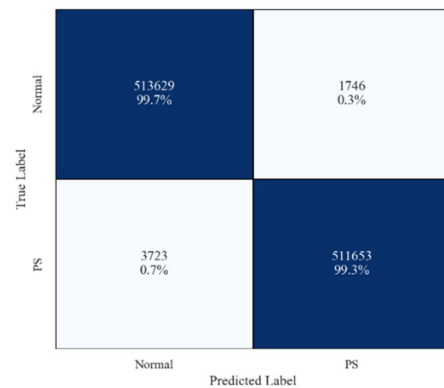


FIGURE 10. PS fault classification results for the proposed RF algorithm.

2) TIER 2–INDEPENDENT PS SCENARIOS

For Tier 2, simulation is conducted with different variations, as detailed in the introduction of this sub-section. The position of the shaded modules is randomly picked on each PV string for simulation. A single sample from the model present in Fig. 13 contains voltage and current information for ten strings. In the first three scenarios, each PS scenario is simulated independently for analysis, but in the last scenario, all three PS scenarios occur simultaneously over the system. This data is then pre-processed using the process defined in Fig. 8, enabling the RF algorithm to analyze each string separately throughout the system. For this tier, the original irradiance is assumed to 1000 W/m² and the SF is derived difference between the original and post-shading irradiance.

Simulating the PS scenarios for this tier results in 20 samples of each, since each PS fault occurs over a span

True Label	Normal	161718 99.1%	934 0.6%	505 0.3%	9 0.0%	0 0.0%	0 0.0%
	1 to 2	1329 0.8%	148565 91.0%	8868 5.4%	2624 1.6%	453 0.3%	1415 0.9%
	3 to 4	815 0.5%	10764 6.6%	140702 86.2%	8658 5.3%	1318 0.8%	997 0.6%
	5 to 6	337 0.2%	4606 2.8%	10270 6.3%	142393 87.2%	3998 2.4%	1651 1.0%
	7 to 8	21 0.0%	618 0.4%	987 0.6%	2375 1.5%	154540 94.7%	4713 2.9%
	9 to 10	6 0.0%	1135 0.7%	481 0.3%	569 0.3%	2462 1.5%	158602 97.2%
	Predicted Label	Normal	1 to 2	3 to 4	5 to 6	7 to 8	9 to 10

FIGURE 11. Number of shaded modules per string detection results through the RF algorithm.

True Label	Normal	58458 96.2%	1718 2.8%	567 0.9%
	80%	3755 1.8%	184747 88.0%	21398 10.2%
	20%	2250 1.1%	19583 9.3%	188067 89.6%
	Predicted Label	Normal	80%	20%

FIGURE 12. SF classification results through the RF algorithm.

of 20 ms every 1 ms is considered an independent sample. The trained RF algorithm is iterated ten times over a randomly picked sample, from the set of 20, to deduce a reliable accuracy for each label. Table 4 displays the accuracy observed for each PS scenario. For Scenario 4 in the table, the observed accuracy for each label depends on whether all three PS scenarios, that are running simultaneously, are accurately deduced. Testing conducted in Tier 2 shows that the proposed RF algorithm can reliably infer PS scenarios, the number of shaded modules, and the associated SF.

VI. DISCUSSION

The proposed data acquisition, simulation techniques, and the utility of the RF algorithm highlight the reliability of the proposed approach. It is observed in this study that the RF algorithm works very efficiently for the application at hand and proves its ability to deduce PS, the number of shaded modules, and the SF with tests conducted by two independent tiers. In comparison to other approaches found

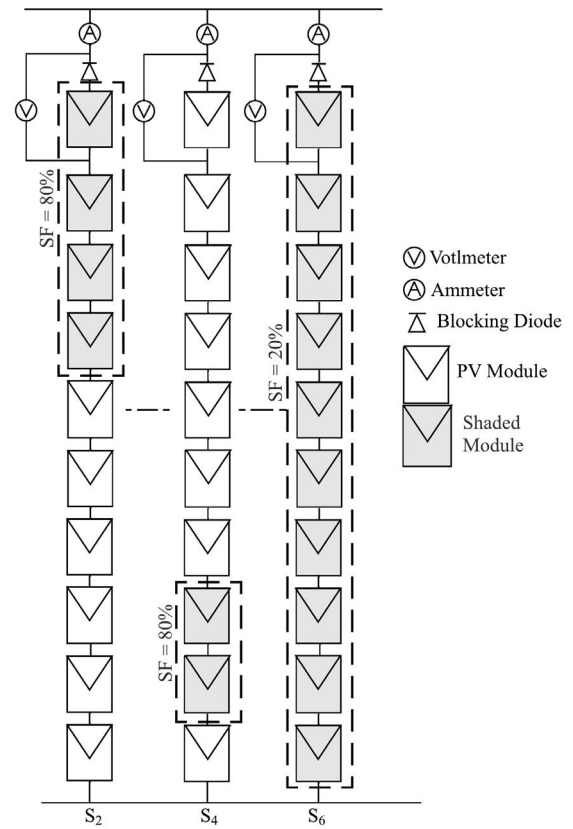


FIGURE 13. PV model to analyzing PS on multiple strings of a 10 x 10 PV system.

TABLE 4. Average classification accuracy of the independent PS scenarios.

Scenario	Predicted Accuracy (%)		
	PS	# of Modules	Shading Factor
1	100	90	90
2	100	90	100
3	100	100	90
4	100	80	90

in the literature, the proposed technique can train a classifier using the features solely available through the PV strings instead of relying on system features, enabling a higher level of independence and scalability. Acquired data considered a varying number of shaded modules and the SF that the irradiance differed by. The advantages of the RF algorithm highlighted in Section III are validated by comparing it with convolutional and recursive-based ANN's such as 1D-CNN and Bi-LSTM. Preliminary investigation demonstrates that, for the problem at hand, the RF algorithm takes significantly less time to train and can predict all three labels with greater reliability.

Compared to other approaches found in the literature, the proposed system improves the data acquisition, simulation, and training processes by uncoupling the PV strings from

the system. The uncoupling decreases the simulation time required to acquire the initial data by 60%. It also enables the proposed technique to be highly scalable since the string parameters from each parallel string are independent of the variation in system parameters due to different scenarios. Hyperparameters of the RF algorithm are optimized through a combination of grid search techniques [34] and the use of the Hyperopt library [35]. The trained RF algorithm is observed to achieve very reliable accuracy of 99.5%, 92.3%, and 90.2% for PS detection, number of modules shaded, and SF, respectively.

The proposed approach is further verified by conducting independent simulations on the testbed. These independent scenarios have verified that the proposed approach is able to detect PS in different strings of the PV system and enables the algorithm to detect PS over multiple strings simultaneously. Since each string is considered independently when deducing PS, the trained algorithm can also conclude the location of the shaded strings in a PV system. This approach complements different shading patterns highlighted in [36] via the results confirmed through Tier 2 since the PV strings from each pattern could be broken down to consider their effects independent of the system parameters.

It is observed through this study that the trained RF classifier can differentiate between the different classes in each category with great accuracy, as shown in Figs. 10–12. The performance of the trained RF algorithm is further validated by conducting two tiers of tests: Tier 1 uses the 30% held-out data from the training dataset, and Tier 2 uses data collected through independent simulations. The proposed RF algorithm is able to deduce single and multiple PS scenarios with high accuracy. After identifying the location of the shaded modules, the current and voltage measurements can be used to estimate the power-voltage characteristics of the shaded string. Since the power-voltage characteristics will vary depending on the number of modules shaded in each string, the proposed approach can be used for tracking the global MPP, as in [37].

Once shaded modules and strings are identified, the following corrective actions to improve the performance of the solar PV system may be take:

- Clean the shaded modules if the modules are shaded by dirt, leaves, or other debris.
- Trim the source of shading if the modules are shaded by trees, bushes, or other vegetation.
- Using an enhanced MPPT method to track the global maximum power point and avoid local maxima [37].
- Redesign the wiring configuration of the system, if applicable, changing the way the modules are connected to each other can help reduce the impact of partial shading [38].

VII. CONCLUSION

In this paper, an RF-based approach is proposed to detect PS, the number of shaded modules, and the associated SF.

The proposed approach decreases simulation time for the data acquisition and determines the location of the shaded string by isolating string parameters from the system parameters. An extensive dataset is generated by acquiring irradiance, temperature, string voltage and current from a PV model simulated by PSCAD/EMTDC. The model takes into consideration different irradiance, temperature, and SF to develop a reliable classifier. The trained classifier is validated through a multi-tiered testing process, considering independent simulations and detecting multiple PS scenarios occurring in the PV system. The proposed approach can achieve high accuracy of 99.5%, 92.3%, and 90.2% for PS detection, the number of modules shaded, and SF, respectively. The proposed approach is applicable for real systems because:

- By accurately detecting PS scenarios and quantifying their effects, the proposed approach provides a practical tool for system owners and operators to optimize energy production.
- The independent analysis of each partial shading scenario during validation, the results are highly representative of what can be expected in practical applications.
- The RF algorithm's high accuracy in detecting PS scenarios, number of shaded modules, and the SF, and the algorithm's performance [20], [21], [22] and resiliency to noisy data [20], [21], [31] demonstrates its reliability in practical scenarios.

The studies highlighted in this paper confirm that the location of the shaded string and the characteristics of PS scenarios can be deduced by acquiring current and voltage features from a PV string instead of the system. The results of this study could be extended to optimize MPPT for different shading scenarios detected in the PV system and detect peak power for each scenario.

ACKNOWLEDGMENT

This paper represents the opinions of the author(s) and does not mean to represent the position or opinions of the American University of Sharjah.

REFERENCES

- [1] N. V. Emodi, T. Chaiechi, and A. B. M. R. A. Beg, "The impact of climate variability and change on the energy system: A systematic scoping review," *Sci. Total Environ.*, vol. 676, pp. 545–563, Aug. 2019, doi: [10.1016/j.scitotenv.2019.04.294](https://doi.org/10.1016/j.scitotenv.2019.04.294).
- [2] J. Ma, X. Pan, K. L. Man, X. Li, H. Wen, and T. On Ting, "Detection and assessment of partial shading scenarios on photovoltaic strings," *IEEE Trans. Ind. Appl.*, vol. 54, no. 6, pp. 6279–6289, Nov. 2018, doi: [10.1109/TIA.2018.2848643](https://doi.org/10.1109/TIA.2018.2848643).
- [3] P. R. Satpathy, A. Sarangi, S. Jena, B. Jena, and R. Sharma, "Topology alteration for output power maximization in PV arrays under partial shading," in *Proc. Technol. Smart-City Energy Secur. Power (ICSESP)*, Mar. 2018, pp. 1–6, doi: [10.1109/ICSESP.2018.8376699](https://doi.org/10.1109/ICSESP.2018.8376699).
- [4] N. G. T. Ramirez and E. Q. B. Macabebe, "Identification of solar PV array partial shading patterns using machine learning," in *Proc. IEEE PES Asia-Pacific Power Energy Eng. Conf. (APPEEC)*, Dec. 2019, pp. 1–5, doi: [10.1109/APPEEC45492.2019.8994360](https://doi.org/10.1109/APPEEC45492.2019.8994360).
- [5] R. A. Guejia-Burbano and G. Petrone, "Partial shading detection on PV panels through on-line impedance spectroscopy," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur. (EEEIC/I&CPS Eur.)*, Sep. 2021, pp. 1–6, doi: [10.1109/EEEIC/ICPSEurope51590.2021.9584621](https://doi.org/10.1109/EEEIC/ICPSEurope51590.2021.9584621).

- [6] C. Li, Y. Yang, F. Fan, L. Xia, P. Peng, Y. Wang, K. Zhang, and H. Wei, "A novel methodology for partial shading diagnosis using the electrical parameters of photovoltaic strings," *IEEE J. Photovolt.*, vol. 12, no. 4, pp. 1027–1035, Jul. 2022, doi: [10.1109/JPHOTOV.2022.3173723](https://doi.org/10.1109/JPHOTOV.2022.3173723).
- [7] J.-H. Teng, H.-C. Wu, Z.-H. Wu, and W.-H. Huang, "Efficient partial shading detection for photovoltaic generation systems," *IEEE Trans. Sustain. Energy*, to be published, doi: [10.1109/TSTE.2023.3271298](https://doi.org/10.1109/TSTE.2023.3271298).
- [8] Z. Mustafa, "Fault and partial shading identification for photovoltaic systems using machine learning and deep learning approaches," M.S. thesis, Univ. Windsor (Canada), Windsor, ON, Canada, 2022.
- [9] A. Ziane, "Detecting partial shading in grid-connected PV station using random forest classifier," in *Artificial Intelligence and Renewables Towards an Energy Transition*, M. Hatti, Ed. Cham, Switzerland: Springer, 2021, pp. 88–95.
- [10] Z. Chen, F. Han, L. Wu, J. Yu, S. Cheng, P. Lin, and H. Chen, "Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents," *Energy Convers. Manage.*, vol. 178, pp. 250–264, Dec. 2018.
- [11] M. M. Badr, M. S. Hamad, A. S. Abdel-Khalik, R. A. Hamdy, S. Ahmed, and E. Hamdan, "Fault identification of photovoltaic array based on machine learning classifiers," *IEEE Access*, vol. 9, pp. 159113–159132, 2021, doi: [10.1109/ACCESS.2021.3130889](https://doi.org/10.1109/ACCESS.2021.3130889).
- [12] N.-C. Yang and H. Ismail, "Robust intelligent learning algorithm using random forest and modified-independent component analysis for PV fault detection: In case of imbalanced data," *IEEE Access*, vol. 10, pp. 41119–41130, 2022, doi: [10.1109/ACCESS.2022.3166477](https://doi.org/10.1109/ACCESS.2022.3166477).
- [13] F. Harrou, B. Taghezouit, and Y. Sun, "Improved kNN-based monitoring schemes for detecting faults in PV systems," *IEEE J. Photovolt.*, vol. 9, no. 3, pp. 811–821, May 2019, doi: [10.1109/JPHOTOV.2019.2896652](https://doi.org/10.1109/JPHOTOV.2019.2896652).
- [14] F. Salem and M. A. Awadallah, "Detection and assessment of partial shading in photovoltaic arrays," *J. Electr. Syst. Inf. Technol.*, vol. 3, no. 1, pp. 23–32, May 2016, doi: [10.1016/j.jesit.2015.10.003](https://doi.org/10.1016/j.jesit.2015.10.003).
- [15] V. K. Kolakaluri, A. Nayak, M. N. Aalam, and V. Sarkar, "Oscillation guided artificial neural network design for the partial shading detection on a photovoltaic array," in *Proc. IEEE Int. Conf. Power Electron., Drives Energy Syst. (PEDES)*, Dec. 2022, pp. 1–6, doi: [10.1109/PEDES56012.2022.10080358](https://doi.org/10.1109/PEDES56012.2022.10080358).
- [16] F. Aziz, A. Ul-Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A novel convolutional neural network-based approach for fault classification in photovoltaic arrays," *IEEE Access*, vol. 8, pp. 41889–41904, 2020, doi: [10.1109/ACCESS.2020.2977116](https://doi.org/10.1109/ACCESS.2020.2977116).
- [17] Z. Chen, Y. Chen, L. Wu, S. Cheng, and P. Lin, "Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions," *Energy Convers. Manage.*, vol. 198, Oct. 2019, Art. no. 111793.
- [18] A. Ul-Haq, H. F. Sindi, S. Gul, and M. Jalal, "Modeling and fault categorization in thin-film and crystalline PV arrays through multilayer neural network algorithm," *IEEE Access*, vol. 8, pp. 102235–102255, 2020, doi: [10.1109/ACCESS.2020.2996969](https://doi.org/10.1109/ACCESS.2020.2996969).
- [19] X. x. Wang, L. Dong, S. y. Liu, Y. Hao, and B. Wang, "A fault classification method of photovoltaic array based on probabilistic neural network," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2019, pp. 5260–5265, doi: [10.1109/CCDC.2019.8832338](https://doi.org/10.1109/CCDC.2019.8832338).
- [20] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Trees vs neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption," *Energy Buildings*, vol. 147, pp. 77–89, Jul. 2017, doi: [10.1016/j.enbuild.2017.04.038](https://doi.org/10.1016/j.enbuild.2017.04.038).
- [21] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *J. Mach. Learn. Res.*, vol. 15, no. 90, pp. 3133–3181, 2014.
- [22] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?" in *Proc. 36th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track*, 2022. [Online]. Available: https://openreview.net/forum?id=Fp7__phQszn
- [23] F. Khosrojerdi, S. Gagnon, and R. Valverde, "Identifying influential factors affecting the shading of a solar panel," in *Proc. IEEE Electr. Power Energy Conf. (EPEC)*, Oct. 2021, pp. 255–260, doi: [10.1109/EPEC52095.2021.9621688](https://doi.org/10.1109/EPEC52095.2021.9621688).
- [24] M. Mansouri, M. Trabelsi, H. Nounou, and M. Nounou, "Deep learning-based fault diagnosis of photovoltaic systems: A comprehensive review and enhancement prospects," *IEEE Access*, vol. 9, pp. 126286–126306, 2021, doi: [10.1109/ACCESS.2021.3110947](https://doi.org/10.1109/ACCESS.2021.3110947).
- [25] N. T. N. Trinh, D. T. Hung, N. H. T. Dat, and P. Q. Dung, "Application of artificial intelligence in detecting and classifying faults of solar panels," in *Proc. IEEE 9th Int. Conf. Commun. Electron. (ICCE)*, Jul. 2022, pp. 513–518, doi: [10.1109/ICCE55644.2022.9852089](https://doi.org/10.1109/ICCE55644.2022.9852089).
- [26] (2011). *BP 3 Series: 215, 220, 225, 230 and 235W Photovoltaic Modules Datasheet*. BP Solar. [Online]. Available: <https://shop.solardirect.com/pdf/solar-electric/modules/bp-215.pdf>
- [27] A. Abubakar, C. F. M. Almeida, and M. Gemignani, "A review of solar photovoltaic system maintenance characteristics," in *Proc. 14th IEEE Int. Conf. Ind. Appl. (INDUSCON)*, Aug. 2021, pp. 1400–1407, doi: [10.1109/INDUSCON51756.2021.9529669](https://doi.org/10.1109/INDUSCON51756.2021.9529669).
- [28] S. Laamami, M. Benhamed, and L. Sbita, "Analysis of shading effects on a photovoltaic array," in *Proc. Int. Conf. Green Energy Convers. Syst. (GECS)*, Mar. 2017, pp. 1–5, doi: [10.1109/GECS.2017.8066212](https://doi.org/10.1109/GECS.2017.8066212).
- [29] H. Patel and V. Agarwal, "MATLAB-based modeling to study the effects of partial shading on PV array characteristics," *IEEE Trans. Energy Convers.*, vol. 23, no. 1, pp. 302–310, Mar. 2008, doi: [10.1109/TEC.2007.914308](https://doi.org/10.1109/TEC.2007.914308).
- [30] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [31] S. Nawar and A. Mouazen, "Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line vis-NIR spectroscopy measurements of soil total nitrogen and total carbon," *Sensors*, vol. 17, no. 10, p. 2428, Oct. 2017, doi: [10.3390/s17102428](https://doi.org/10.3390/s17102428).
- [32] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953).
- [33] F. Pedregosa, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [34] S. M. LaValle, M. S. Branicky, and S. R. Lindemann, "On the relationship between classical grid search and probabilistic roadmaps," *Int. J. Robot. Res.*, vol. 23, nos. 7–8, pp. 673–692, Aug. 2004.
- [35] J. Bergstra, D. Yamins, and D. D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," *J. Mach. Learn. Res.*, vol. 28, pp. 115–123, Jun. 2013.
- [36] B. Yang, R. Shao, M. Zhang, H. Ye, B. Liu, T. Bao, J. Wang, H. Shu, Y. Ren, and H. Ye, "Socio-inspired democratic political algorithm for optimal PV array reconfiguration to mitigate partial shading," *Sustain. Energy Technol. Assessments*, vol. 48, Dec. 2021, Art. no. 101627, doi: [10.1016/j.seta.2021.101627](https://doi.org/10.1016/j.seta.2021.101627).
- [37] Y. Mahmoud and E. F. El-Saadany, "A novel MPPT technique based on an image of PV modules," *IEEE Trans. Energy Convers.*, vol. 32, no. 1, pp. 213–221, Mar. 2017, doi: [10.1109/TEC.2016.2629514](https://doi.org/10.1109/TEC.2016.2629514).
- [38] M. Z. S. El-Dein, M. Kazerani, and M. M. A. Salama, "Optimal photovoltaic array reconfiguration to reduce partial shading losses," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 145–153, Jan. 2013, doi: [10.1109/TSTE.2012.2208128](https://doi.org/10.1109/TSTE.2012.2208128).



ZAIN MUSTAFA received the B.Sc. degree in computer science and the M.Sc. degree in electrical engineering from the University of Windsor, Windsor, ON, Canada, in 2019 and 2022, respectively. He joined a leading automobile manufacturer, in 2020, as a Software Engineer, where he leads an engineering team that manages their portfolio of services to supplement the electric vehicle fleet and third-party electric charging networks. His research interests include machine learning, renewable energy sources, power system protection, and electric vehicles.



MAHER A. AZZOUZ (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees (Hons.) in electrical power engineering from Cairo University, Giza, Egypt, in 2008 and 2011, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2016. He was a Postdoctoral Fellow with the Power and Energy System Group, University of Waterloo. He is currently on leave with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He is with the Department of Electrical Engineering, Qatar University, Qatar. He is a Registered Professional Engineer in the Province of Ontario. His research interests include control of power electronic converters, power system protection, distribution system operation and planning, and renewable energy sources. He has been recognized as one of the best reviewers of IEEE TRANSACTIONS ON SMART GRID.



AHMED S. A. AWAD (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, Egypt, in 2007 and 2010, respectively, and the Ph.D. degree from the Electrical and Computer Engineering Department, University of Waterloo, Waterloo, ON, Canada, in 2014.

He is currently an Assistant Professor with the Electrical and Computer Engineering Department, Sultan Qaboos University, Muscat, Oman, and an Adjunct Associate Professor with the Electrical and Computer Engineering Department, University of Windsor, ON, Canada. He is a Registered Professional Engineer in the Province of Ontario, Canada. His research interests include the integration of sustainable energy technologies into power systems, electricity market equilibrium, and operation and control of power systems. He is an Editor of IEEE TRANSACTIONS ON SUSTAINABLE ENERGY.



AHMED AZAB received the P.Eng. degree in industrial and manufacturing systems engineering from the University of Windsor. He is currently the Director of the Production and Operations Management Research Laboratory and a Professor of industrial engineering with the University of Windsor. He is an Adjunct Professor with Nile University, Egypt. He is a Registered Professional Engineer in the Province of Ontario. He was a recipient and nominee for international and national research awards. His national and provincial research granting agencies and direct industrial research funds. He is an Associate Editor of *Expert Systems With Applications*.



MOSTAFA F. SHAABAN (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, Egypt, in 2004 and 2008, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2014.

Currently, he is an Associate Professor with the Department of Electrical Engineering, American University of Sharjah, Sharjah, United Arab Emirates, and an Adjunct Professor with the University of Waterloo. He has several publications in international journals and conferences. His research interests include smart grids, renewable DG, distribution system planning, electric vehicles, storage systems, and bulk power system reliability. He serves as an Associate Editor for *IET Smart Grid* and a reviewer for several refereed journals.

...