



AFRL-RQ-WP-TM-2024-0001

SERIES ARC FAULT IDENTIFICATION IN DC DISTRIBUTION BASED ON RANDOM FOREST PREDICTED PROBABILITY

**Vu Le, Chad Miller, Bang-Hung Tsao and Xiu Yao
IEEE**

**APRIL 2023
Final Report**

**DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited;
AFRL/PA-2022-4635; Cleared 29-09-2022.**

**AIR FORCE RESEARCH LABORATORY
AEROSPACE SYSTEMS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7542
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This paper was cleared for public release by AFRL Public Affairs, AFRL/PA and is available to the general public, including foreign nationals.

Copies may be obtained from the Defense Technical Information Center (DTIC)
(<https://discover.dtic.mil>).

AFRL-RQ-WP-JA-2023-0007 has been reviewed and is approved for publication in accordance with assigned distribution statement.

This paper is published in the interest of scientific and technical information exchange and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 04012023	2. REPORT TYPE Final	3. DATES COVERED	
		START DATE 02032015	END DATE 020320230
4. TITLE AND SUBTITLE SERIES ARC FAULT IDENTIFICATION IN DC DISTRIBUTION BASED ON RANDOM FOREST PREDICTED PROBABILITY			
5a. CONTRACT NUMBER In House	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER	
5d. PROJECT NUMBER	5e. TASK NUMBER	5f. WORK UNIT NUMBER Q1G2	
6. AUTHOR(S) Vu Le, Chad Miller, Bang-Hung Tsao and Xiu Yao			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UDRI, 300 College Park, Dayton, OH 45469 University at Buffalo, 230 Davis Hall, Buffalo, NY 14260			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Aerospace Systems Directorate Air Force Research Laboratory Air Force Materiel Command Wright-Patterson Air Force Base, OH 45433-7542		10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RQQE	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RQ-WP-TM-2024-0001
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited; AFRL/PA-2022-4635; Cleared 29-09-2022. Other requests for this document shall be referred to AFRL/RQQE, Wright-Patterson Air Force Base, OH 45433-7542.			
13. SUPPLEMENTARY NOTES PA Clearance Number: AFRL/PA-2022-4635; Clearance Date: 9-29-2023 This paper was published in the IEEE Journal of Emerging and Selected Topics in Power Electronics. 10.1109/JESTPE.2022.3228421			
14. ABSTRACT Series dc arc fault generates arcing noise making it challenging to identify the arc fault location and thus endangering the system's safety. This study proposes a series arc fault identification method in a dc zonal electrical distribution using Random Forest based local detectors to monitor constant power loads, and output predicted nominal and arc fault probabilities. With full communication capability among all local detectors, the master detector makes the final decision using Random Forest algorithms. The proposed fault identification method is experimentally verified with a zonal electrical distribution testbed that comprises three constant power loads.			
15. SUBJECT TERMS Circuit faults, Prediction algorithms, Fault detection, Fault location, DC arc fault, Random Forest, master detector, local detector			
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	SAR
			18. NUMBER OF PAGES 15
19a. NAME OF RESPONSIBLE PERSON Chad Miller			19b. PHONE NUMBER (Include area code) (937) 255-4122

Technical Memo for Intelligent Power Q1G2

Summary of the Memorandum

This technical memo serves for the basis of establishing final technical output of the In-House Work Unit Number (Q1G2). : Intelligent Power. Final documentation was published (AFRL-RQ-WP-JA-2023-0007) and cleared for public release (AFRL-2022-4635). Other supporting publications include: AFRL-RQ-WP-TR-2019-0016, AFRL-RQ-WP-TR-2018-0099, AFRL-RQ-WP-TP-2017-0032, and AFRL-RQ-WP-TM-2017-0073.

Background

This work unit was established in March 2015 to cover technical efforts related to in-house Intelligent Power tasks. Intelligent Power was divided into two subsections: Intelligent Power Generation and Intelligent Power Distribution. Intelligent Power Generation focuses on mechanical power generating technologies and approaches to dealing with faults directly at the source. Intelligent Power Distribution focuses on power distribution technologies and dealing with faults through-out the electrical power system. Other key technologies investigated deal with increasing the DC distribution voltage to enable high power loads, intelligent and optimal controls to enable higher power quality and more resilient systems, and Modeling and Simulation as an engineering tool for improved design and more detailed technology assessments.

Participation

Intelligent Power was an investigation into early 6.2 technologies that could improve Electrical Power System performance. Investigations looked into intelligent design, controls, and autonomy to improve electrical resilience and performance and involved many partnerships. These included partnerships with universities such as: the University of Dayton, Perdue, University at Buffalo, and the Ohio State University. Intelligent power also worked with Idaho National Research lab and NASA GRC.

High Impedance Arc Fault and AI/ML

A large portion of Intelligent Power research focused on detection and protection from high impedance arc faults. Artificial Intelligent and Machine Learning (AI/ML) was a significantly leverage tool in which new knowledge was created and solution approaches were proven in laboratory-based experiments. AI/ML was adapted to run in a real-time or faster than real-time environment in which difficult to detect arc faults were successfully detected and isolated using Random Forests. The tech publication: Series Arc Fault Identification in DC Distribution Based on Random Forest Predicted Probability is the final document reporting the knowledge discovered under the intelligent power work unit.

Series Arc Fault Identification in DC Distribution Based on Random Forest Predicted Probability

Vu Le, *Student Member, IEEE*, Chad Miller, *Member, IEEE*, Bang-Hung Tsao, *Member, IEEE* and Xiu Yao, *Member, IEEE*,

Abstract—Series dc arc fault generates arcing noise cross-talk to adjacent electrical loads, making it challenging to identify the arc fault location, which creates a fire hazard that endangers the system's safety. This study proposes a series arc fault identification method in a dc zonal electrical distribution using Random Forest based local detectors to monitor constant power loads, and output predicted nominal and arc fault probabilities. The predicted probabilities are then sent to a centralized master detector to obtain the final decision. With full communication capability among all local detectors, the master detector makes the final decision using Random Forest algorithms. If there is any disconnection in the communication links, the local detector can switch to operate independently by using the predicted arc fault probability to output the flag signal autonomously, while the master detector continues to operate with other local detectors. The proposed fault identification method was experimentally verified with a zonal electrical distribution that comprises three constant power loads.

Index Terms—DC arc fault, constant power load, Random Forest, master detector, local detector.

I. INTRODUCTION

THE emergence of advanced power electronics and resilient controllers enabled a wide range of zonal electrical distribution (ZED) applications, such as oil platform [1], integrated ship [2], and more electric aircraft (MEA) [3]. ZED operation is based on different zones interfaced with power electronics converters. Each zone may contain multiple distribution buses cascading from the sources to multiple series-connected, and parallel-connected loads [4]. While protection devices such as circuit breakers, fuses, and solid-state power controllers are commonly deployed for fault detection and system protection, series arc faults pose unique safety challenges. Series arc fault current is too low to trigger these conventional protection devices. Also, the current limiting effect of the power electronics interfaces in ZED can potentially obstruct the fault localization [1]. Therefore, series dc arc fault require its own detection algorithm and hardware separate from other fault types.

Series dc arc fault characteristics have been extensively studied based on input current and voltage in both time-domain and frequency domain [5]–[7]. An arc fault detection algorithm was implemented within the dc optimizer [8]. The algorithm consists of two steps to detect arc faults, including current and voltage. Once an arc fault is detected, this method controls the duty ratio to extinguish the arc fault. The algorithm achieved much slower detection latency, up to 200 ms. The electromagnetic radiation emits from the series arc fault has also been extensively studied as an arc fault detection method

[9], [10]. However, these methods required the antenna to be nearby the fault. In addition, it is difficult to accurately locate the arc fault based on the emitted radiations

In recent years, series arc fault detection has been developed using intelligent tools, such as machine learning (ML) and neural network, to provide flexibility and robustness in detection. A novel method using an image recognition algorithm is presented in [11]. This method converts Wavelet Transform (WT) coefficient matrix to a colormap index, then feeds it into a neural network to monitor and detect arc faults. The verification accuracy achieved up to 95.8 %, providing evidence that artificial intelligence algorithms effectively detect arc faults. However, WT and image processing consume a tremendous amount of computational power, which leads to longer detection latency time, and reduces the effectiveness of time-sensitive applications such as arc fault detection. Moreover, model based series dc arc fault detection methods have also been explored [12], [13]. However, model based methods heavily depend on system parameters and are not robust against parameter changes. System parameters can also be difficult to obtain.

A neural network incorporated with Fast Fourier Transform (FFT) was used to detect series arc fault in [14]. A Hidden Markov utilized Wavelet Transform as a feature to detect series dc arc fault in [15]. Ensemble machine learning (EML) algorithms were implemented on low-cost hardware as an arc fault detector and verified with different individual constant power loads (CPL) [16], [17]. However, these techniques are focused primarily on series arc fault detection in a single load, and arcing noise propagation was not considered. The effect of the arcing noise propagation was demonstrated in two parallel-connected strings of PV system [18], creating difficulty in identifying the location of the arc fault. The difficulty of series arc fault identification is further exacerbated in a more extensive electrical system, especially in ZED, where arcing noise also propagates from one zone to another.

Fault identification in a dc ZED can be generally carried out by using parameter estimation methods [19]–[21]. These approaches have been successfully tested for fault identification, such as short-circuit and parallel arc fault location. For instance, a combination of di/dt and line inductance was utilized and experimentally verified for a dc ZED [19], [20]. di/dt was utilized to detect any fault in the system, while the total inductance value was used to determine the location of the fault. This approach is promising when there is a large sudden current change when a fault has occurred. However, the series arc fault is a low fault current, which might not reach the

threshold of the preset di/dt . An active impedance estimation was utilized to identify fault location in a marine ZED [21]. It can be challenging to obtain the accurate impedance of an extensive system during operation. For further complication, series arc fault in dc distribution that is highly penetrated with power electronics can generate significantly different arc behaviors based on different converters [17] and controllers [22], potentially affecting the circuit parameters differently.

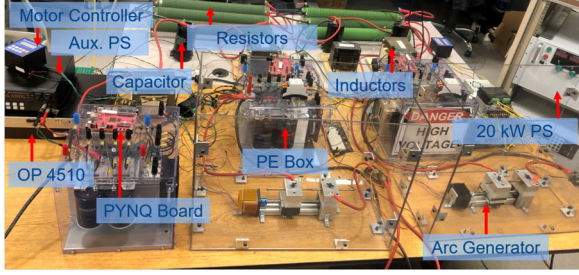
High-frequency series arcing noise propagates to different lines, and loads can be prevented using auctioneering diodes [23]. However, the low-frequency faulted current fluctuations caused by arc fault must be compensated by the adjacent loads or a source, making these currents resemble arc fault and cause mistriggered. Additional hardware, such as capacitors connected to the loads and the source, is required to identify series arc fault location [24]. This method uses Wavelet Transform to analyze the initial arc transient polarity of the capacitor currents. Recently, a line inductance estimation and a Kalman filter were developed to identify series arc fault and the lines affected by arcing noise [25]. [26] uses the correlation of current and voltage rate of change to identify arc fault location. The arc fault conditions were then simulated under low irradiances and successfully detected. These approaches suggesting current and voltage synchronization between different loads are required to identify series arc fault locations. At the same time, existing literature also suggested that a master detector (MD) is required to acquire an accurate arc fault identification in large systems [18], [27]. As a result, the communication infrastructure between an MD and LDs is crucial for accurate arc fault identification. The importance of communication can be demonstrated in [28], HVDC/AC transmission networks state estimation, where remote terminal units send information to the master units using either UDP or TCP/IP for a SCADA system. However, TCP/IP communication could fail due to handshake protocol never being reached, congestion [29], or data package being lost. Therefore, a redundant algorithm should be further explored to identify arc fault location when there is a disconnection between the MD and the LDs.

This paper proposes an advanced ML called Random Forest (RF) as a series dc arc fault identification algorithm in a large dc distribution system. RF algorithm have been widely adopted for many complex pattern recognition applications, such as partial discharge detection, voltage stability, and different types of faults detection [30], [31], [32]. A combination of RF and neural network was implemented for arc fault detection [33]. The RF algorithm was used to select the most important features to train the neural network. This technique was able to achieve high accuracy with ac loads. RF was one of the proposed ensemble algorithms used as an adaptive arc fault detection method in a previous study [34]. Unlike a single-model ML algorithm, RF uses multiple independent Decision Tree (DT) outputs to form a final decision. This action averages out the extreme cases to reduce bias and variance in the training process. Most importantly, the RF algorithm inherits one of the DT's attributes called predicted probability, which provides the flexibility to interpret the input data. Specifically, nominal probability and arc fault probability are used in the proposed method to identify arc fault location.

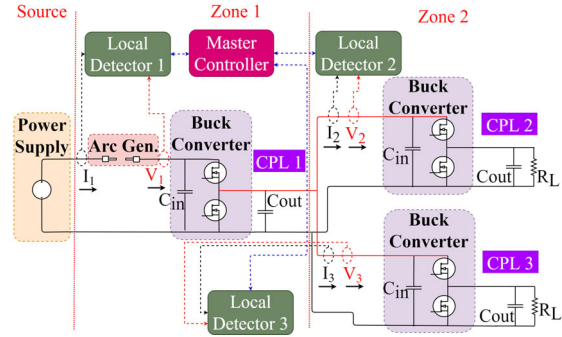
RF is an algorithm that can be used for classification ability or to generate effective features. A fault detection method in wind turbines was implemented using RF and XGboost algorithms [35]. This method uses RF's feature importance attribute to rank the features. The three highest rank features are selected to feed into the XGboost algorithm to determine the final decision. Similarly, a set of features were selected by using the RF algorithm as inputs for different neural networks for arc fault detection [33]. In contrast, a convolution neural network was used to extract features, and the RF algorithm was used for fault diagnosis [36].

Much existing literature on series arc fault detection utilized ML algorithm was selected with the default setting and overlooked the decision-making process, one of which is predicted probability. By default, 0.5 is the discrimination threshold value. For example, if an arc fault predicted probability is greater than 0.5, then the final decision is arc fault. Otherwise, the final decision is nominal. However, different applications have different electrical parameters, which influence the discrimination threshold, especially with an electrical system that is highly penetrated with power electronics, where switching noise, current ripples, and arcing propagation are presented. In addition, the ML algorithm predicted probability must be calibrated to achieve high confidence in the probabilities. The following study was able to identify arc fault location in parallel-connected CPLs by using the RF-based local detector (LD) predicted probabilities [37]. If the arc fault probability from any LD is larger than the default discrimination threshold, 0.5, it is decided to be the arc fault location. If more than one arc fault probability is larger than 0.5, then the largest arc fault probability determines the arc fault location. However, 0.5 is the borderline between nominal and fault, where it often fluctuates and generates mistriggered. Therefore, maximizing the margin between nominal and fault by selecting an optimal threshold is more desirable.

As the work in [34] laid a solid foundation on algorithm and hyperparameter selection and benchmarking, the complexity of multiple loads and zones have not been addressed, which is the focus of this paper. In this paper, techniques have been developed to address the arc noise propagation issue to ensure a fast and accurate fault localization in a zonal dc distribution systems. This work aims to advance the previous studies [34], [37] on EML as an arc fault identification algorithm by demonstrating the advantage of using predicted probability as RF's outputs instead of the binary 0 and 1 class identification that other studies used. Also, calibration techniques for ML algorithms were explored using two functions: Isotonics and Sigmoid. All ML algorithms were trained with different numbers of datasets and tested to determine better confident models. The best-calibrated model was then selected for further analysis using the accuracy score and F1 score, which provide the model's overall performance. However, the trade-off between different types of errors must be further analyzed to avoid arc fault misclassification. In addition, an RF-based was developed to identify the arc fault location. The arc fault identification algorithm was verified in a ZED with different scenarios regarding the communication statuses between LDs and the MD. This work demonstrates the possibility of using



(a) ZED experimental testbed with three CPLs.



(b) Circuit diagram of the experimental setup.

Figure 1. ZED hardware setup for data collection and experimental results.

low-cost hardware for fast and accurate arc fault identification in a zonal system that consists of three CPLs. The paper is organized in the following way: Section II presents the experimental testbed with different arc fault locations. Section III provides the analysis of predicted probability calibration and error evaluation. The best-performance ML algorithm is then implemented on LDs and MD with different communication statuses for arc fault identification simulation and experimental verification. Lastly, conclusion are presented in Section V.

II. SERIES ARC FAULT IN DC ZED TESTBED

Three CPLs were designed and built to emulate a ZED with three different zones: Source Zone contains a dc power supply, Zone 1 contains CPL 1 with 270 V input voltage, and Zone 2 contains CPL 2 and CPL 3 which are in parallel with 200 V input voltage, as shown in Fig. 1. Series arc fault was initiated only at the input of each CPL individually. The 270 V input voltage was selected to emulate a typical dc distribution system for MEA [38], [39]. Although this study focused only on MEA applications, the setup and system topology can represent many other dc applications, including microgrids [40] and electric vehicle charging stations [41]. The experimental ZED testbed was realized with a 20 kW Magna power supply, Opal 4510, two arc generators, and power electronics (PE) boxes that can be configured as buck converters, as shown in Fig. 1a. The buck converters and the resistors were controlled by the Opal 4510 to function as CPLs. The system diagram is shown in Fig. 1b.

The CPL parameters are listed in Table I. Current ratings for different CPLs were chosen so each zone can have sufficient power to sustain an arc fault while the power supply output current is within the 33 A overcurrent protection limit. The lower limit for each CPL current represents the current level when an arc fault is present. Voltage and current parameters were measured at 500 kS/s with a mixed-signal oscilloscope. The recording length was 10s. Experiments were conducted for arc located at each zone to study the following four aspects: arcing current, propagated arcing noise, arcing voltage, and input voltage drop.

There are two types of arc fault: parallel and series. Parallel arc fault typically produces a large current change, which can

Label	$V_{in}(V)$	$V_{out}(V)$	$I_{in}(A)$	$I_{out}(A)$
CPL 1	270	200	19 – 24	25 – 30
CPL 2	200	150	12.5 – 15	16 – 20
CPL 3	200	150	12.5 – 15	16 – 20

Table I. Experimental setup parameters.

be readily used to detect the fault, as has been extensively verified in [42], [43]. In addition, the high fault current amplitude could easily trip the circuit breakers or blow out the fuses. However, series arc fault is considered more of a fire hazard due to low fault current, making it harder for the circuit breakers to respond promptly. The prominent presence of power electronics in modern distribution systems adds a further challenge in series arc fault detection [27], [34]. It also reduces the performance of the distribution systems, where the arc fault consumes additional unnecessary power if not detected and isolated in time. While both types of arc faults are important to study, the detection mechanisms are very different, and this paper focuses only on series arc faults.

A. Arc Fault At Zone 1

Series arc faults are commonly initiated because of chemical, electrical, and mechanical deterioration of wire or interconnections that are in series with the electrical loads. In this study, an arc generator was placed in series at the input of the CPLs. Arc faults were created by separating two copper rods (electrodes) that were initially in contact with a stepper motor. Fig. 2 shows an example of arc fault at CPL 1. In this scenario, the ZED was initially operating normally, followed by an arc fault introduced at the input of CPL 1. With the presence of an arc fault, high frequency arcing noise can be observed in CPL 1 current I_1 . Due to the input and output capacitors in CPL 1, the high-frequency arcing noise does not propagate downstream zones. Series arc fault can be modeled as resistance [44] and its relation to the input voltage is expressed in (1). With close-loop controlled CPL loads, the output power is maintained constant regardless of voltage and current levels. If an arc fault is introduced in series with a CPL, the voltage drop across the arc will decrease the CPL's input voltage. For example, since the voltage across the electrodes increased from 0 V to approximately 20 V after arc was generated, V_{in} decreased from 270 V to 250 V. I_{in} will then increase to maintain the constant power. This

can be seen in Fig. 2 where CPL 1 average current amplitude increased by approximately 2 A after arc occurs, while CPL 1 output voltage V_{1out} remained nominal due to the closed-loop controller. This is the reason why Zone 2 was not affected by an arc fault in Zone 1: although V_{1in} dropped, V_{2in} , V_{3in} , and V_{source} remained the same. This can be considered one of ZED's advantages, preventing arc fault from propagating downward and containing arcing noise within a zone [45]. Since there is only one line in Zone 1, this arc fault scenario can be quickly detected using the existing EML-based arc fault detector [16].

$$V_{in} = V_{source} - I_{in}(R_{conductor} + R_{arc}) \quad (1)$$

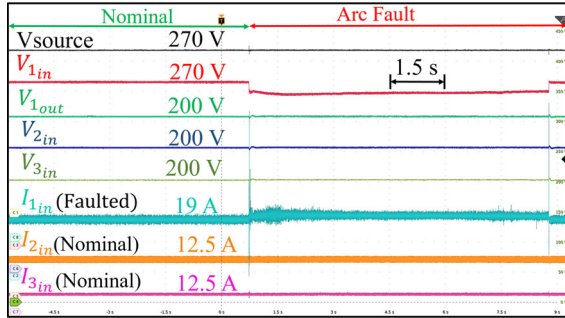


Figure 2. Arc fault initiated at CPL 1 in ZED.

Previous studies have demonstrated that an arc can be considered as a nonlinear resistor, with the arc voltage decided by both the arc gap length and arc current. Arc resistance is, therefore, also a nonlinear function of both arc current and gap length (with minor fluctuation affected by the ambient environment such as temperature and pressure) [27]. When the input voltage changes, it can potentially change the current, which then changes the arc voltage. Furthermore, the relationship between arc and input voltage is complementary when an arc fault is presented, making the input voltage vary based on the arc voltage level. In other words, the arc voltage drop can also be affected by the input voltage. Therefore, a fixed voltage drop threshold method in [46] can be less effective with a dynamic load such as CPL.

B. Arc Fault At Zone 2

Arc fault was also initiated at the input of CPL 2 in Zone 2. As it can be seen in Fig. 3, similar fault current patterns are observed for the input current of all three CPLs indicating that arcing noise has propagated from faulted CPL (CPL 2) to adjacent paralleled load (CPL 3), and upstream zone (CPL 1). Due to the input capacitor, the faulted input current suddenly dropped to 0, then quickly recovered, generating a current spike up to 47 A in I_{2in} . The adjacent unfaulted CPL 3, I_{3in} , generated an opposite current spike of -5 A, as shown in Fig. 4, indicating that 5 A was compensated for the faulted current spike. The current compensation occurred during the entire sustain arc, where arcing noise from CPL 2 and CPL 3 have the same noise patterns with opposite amplitude. This response potentially makes it more challenging to identify arc fault locations.

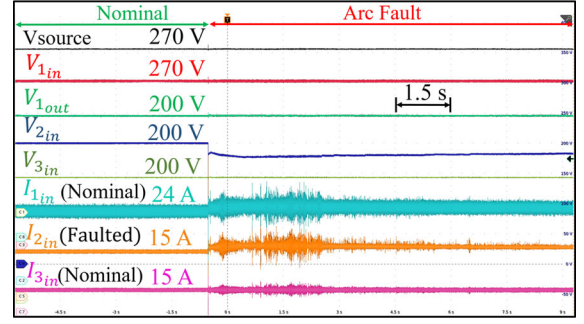


Figure 3. Arc fault initiated at CPL 2 in ZED.

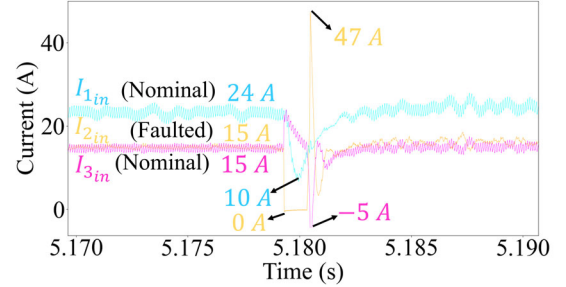


Figure 4. CPLs currents when arc fault initiated at CPL 2.

High frequency arcing noise can also be observed in all three CPLs input currents spectrogram, as shown in Fig. 5. It can be seen that the CPLs started with the nominal operation, where only a 10 kHz switching frequency is present. After 4.3 s, all three currents display similar fault patterns between 0 and 10 kHz. Especially in Zone 2, CPL 2 and CPL 3 experienced similar arcing noise intensity as if both CPLs were faulted. As a result, frequency-based detection method such as Wavelet Transform [15] and FFT [14], can be mistriggered on the unfaulted CPLs. Although the CPL 1 current noise seems to be slightly lower, due to the filtering effect of its input and output capacitors, this small difference can still cause a mistrigger of the CPL 1 protection device and cut off power to all downstream zones.

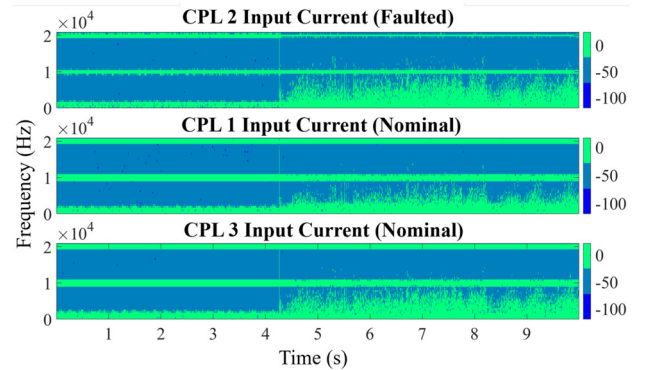


Figure 5. The spectrogram of all input currents .

III. ARC FAULT IDENTIFICATION METHODS

Series arc fault behavior is highly inconsistent in a power electronics system containing different CPLs types, current

levels, and controller responses [47]. This increases the arc fault detection complexity. Furthermore, the propagation of the arcing noise can trigger the detectors on adjacent unfaulted CPLs. Therefore, a more robust arc fault detection algorithm is needed. To increase the fault localization accuracy, effective features derived from both voltage and current measurements are utilized. Six time-domain features were used in total. The following five features were derived from the current measurements: average, median, variance, RMS, and the difference between the maximum and minimum values. In addition, the difference between maximum and minimum voltage values was also used. These features were then normalized together using the max-min method. This technique transforms the feature array, X , to an array of standard deviation:

$$X_{std} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (2)$$

then uses maximum and minimum values of the feature array to scale the standard deviation:

$$X_{scaled} = X_{std} * (X_{max} - X_{min}) + X_{min} \quad (3)$$

A high-level detector was also adopted to communicate with the fault detector at each CPL. This allows the high-level detector to gather arc fault information at each CPL to facilitate the arc fault localization.

The interest in ML-based arc fault detection has emerged in recent years [15], [34]. For example, using two arc fault detectors that output classes 0 or 1 to identify arc fault location was implemented [34] in two parallel-connected CPLs. Although arc fault location was successfully identified, the method utilized the default discrimination threshold of 0.5, which might not be an optimal threshold value in ZED, where the prior zone of the faulted CPL has similar fault patterns and potentially triggered the unfaulted CPLs.

In this study, RF's predicted probabilities ranging from 0 to 1 are utilized as the confidence level of arc fault detection at each CPL. The predicted probabilities are then sent to an MD to analyze and determine the arc fault location using another RF algorithm. In this approach, arc fault detectors continue to operate while the MD analyzes the arc fault location.

A. Class Predicted Probability

The Decision Trees (DT) algorithm is a building block that forms RF using bootstrap aggregation, where multiple DTs can be created from a single group of training datasets. From this group of training datasets, sub-datasets are randomly selected. These sub-datasets are then passed through multiple functions, such as entropy, information gain, and Gini index, to create different branches that form one DT algorithm. In particular, entropy can be used to measure the randomness of the sub-datasets:

$$Entropy(S) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (4)$$

It is easier to make a definitive conclusion if the entropy is low. It ranges from 0 to 1, and 1 is the highest entropy. For

instance, if 3900 out of 4000 nominal datasets are considered nominal (0), and 100 out of 4000 nominal datasets are considered a fault (1), it has an entropy of 0.1687, which can be considered minimal randomness. The entropy values then can be used for information gain at each node:

$$Information\ Gain = Entropy(T) - Entropy(T, X) \quad (5)$$

where (T, X) is a sub-branch of (T) . The goal is to find features that return the highest information gain and the smallest entropy to establish a node, which is determined by each split's quality between nominal and arc fault data. One popular method to measure each split's quality is Gini Impurity, as shown in (6). It calculates the randomly chosen data in a node that is incorrectly classified.

$$L_G(n) = 1 - \sum_{i=1}^J (p_i)^2 \quad (6)$$

where J is the number of classes, and P_i is the probability of randomly selecting a class i . In this study, J is 2: nominal and fault. Similar to information gain, the best split is chosen by maximizing the Gini gain, which is calculated by subtracting the weighted impurities of the branches from the original impurity. Once a node reaches 100 % purity, which contains only a single class, it stops generating sub-branches. The last nodes, also known as leaf nodes, consist of the number of samples of all possible classes, where predicted probabilities can be calculated to determine a class prediction.

However, the process of creating purity nodes can have negative impacts on the final prediction. One of them is over-fitting, where the DT has learned too well from the training data and will not be able to predict a new set of data accurately. To combat over-fitting, a large number of uncorrelated DTs should be used to have a better generalized classification model, which forms an RF algorithm. The RF algorithm process is summarized as follows:

- 1) The training data are randomized when training each DT algorithm. The training data are drawn with replacement, which means that some datasets can be drawn by multiple DTs in a single training session.
- 2) A random subset of the features are used for each training session to prevent same leaf nodes DT at each training session.
- 3) Each DT calculates predicted probabilities by determining how many samples of that class are at the leaf node, then dividing it by the total number of samples from all possible classes.
- 4) The final predicted class probabilities of an input sample are computed by averaging the predicted class probabilities of all DTs.
- 5) The final class prediction is determined by comparing the averaged predicted probabilities with the preset discrimination threshold.

B. Data Acquisition

The measurement from both nominal operating conditions and arc fault conditions were collected. The original recorded

10s long data was then divided into smaller datasets with 600 sample points to train and test an RF algorithm for each CPL. This RF algorithm is the most basic arc fault protection unit for each load branch. One of the metrics to measure the performance of the RF algorithm with the different number of training datasets is the accuracy score. The accuracy score measures the percentage of correct classifications. Generally, more training datasets produce a better ML model, producing higher accuracy scores. Fig. 6 illustrates the accuracy score of different discrimination threshold values of RF algorithms that were trained with different number of datasets. It can be seen that all models' accuracy scores increase as the number of training datasets increases. However, the accuracy scores seem to have stagnated at 10000 for training datasets with 97 %. As a result, 10000 datasets were used for all later benchmarks in this study.

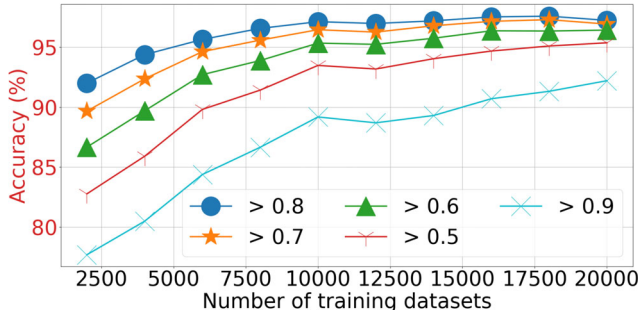


Figure 6. Accuracy vs different number of training datasets.

Table II illustrates the number of training and testing datasets for the RF algorithm. The training and testing processes were simulated 20 times. A randomization function was utilized to choose the training and testing datasets randomly without replacement; an individual dataset can only be selected once within one trial. 80 % of the data were used to train, and the remaining 20 % were to test. This data percentage split was proven optimal in [48]. At each trial, 4000 of nominal (before arc occurs) and 4000 arc fault datasets from faulted CPL and 2000 nominal datasets from unfaulted CPLs were chosen accordingly for a total of 10000 training datasets. Since arc fault was not initiated in the unfaulted CPL branch, all data collected from the unfaulted CPL branch is nominal data, and thus there was zero arc fault data from the unfaulted CPL, as shown in Table II. Although more nominal data can be easily obtained and available to train the algorithm, it does not increase the detection accuracy. In contrast, more arc fault datasets in training data can improve the classification ability, but obtaining arc fault data is more costly than nominal data. The number of fault training datasets matters more than the imbalance ratio between nominal and fault datasets.

	Training Data		Testing Data	
	Nominal	Arc	Nominal	Arc
Faulted CPL	4000	4000	1000	1000
Unfaulted CPL	2000	0	500	0
Total	6000	4000	1500	1000

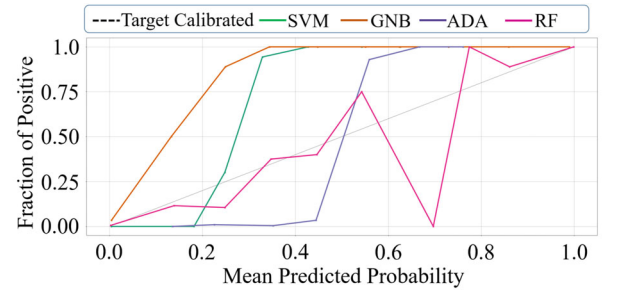
Table II. Experimental training data and testing data.

C. Predicted Probability Reliability Analysis

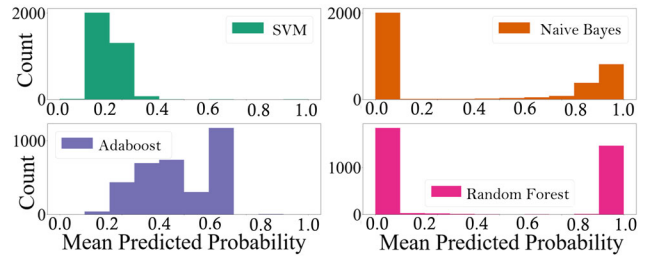
The predicted probabilities give ML algorithms the flexibility to interpret the probability of the input data belonging to each of the classes. However, the distribution of the calculated predicted probabilities does not always match the distribution of the target predicted probabilities for different algorithms [49]. For instance, when testing data containing 40 % of fault data and 60 % of nominal data, ideally, the trained algorithm should output a mean predicted probability of 0.4, as expressed below:

$$\text{Mean Predicted Probability} = \frac{(40 * 1 + 60 * 0)}{100} = 0.4 \quad (7)$$

However, the actual predicted probability typically deviates from that. This behavior can be illustrated using the reliability curve in Fig. 7a, where the y-axis is the percentage of fault datasets, and the x-axis is the mean predicted probability. The target predicted probabilities should show a diagonal reliability line from 0 to 1, with the average predicted probability (x-axis) matching the actual probability (y-axis).



(a) Reliability curve.



(b) Predicted probabilities distribution.

Figure 7. ML algorithms reliability curve.

Four selected ML algorithms were trained with nominal and fault datasets, and their reliability curves are presented in Fig. 7a. It can be seen that single-model ML algorithms, such as Support Vector Machine (SVM) and Naive Bayes (GNB), produce too small of predicted probability to classify fault data. For instance, SVM outputs 100 % of fault data when the mean predicted probability is 0.4 and beyond. Furthermore, it can be seen that the majority of data was less than 0.25 in the mean predicted probability distribution, as shown in the top left of Fig. 7b. This imbalance of probabilities generates inaccurately predicted probabilities that could potentially misinterpret nominal data containing propagated arcing noise. For instance, predictions with a true fault probability of 0.4 and beyond were outputted 1. GNB and Adaboost (ADA) produced

similar reliability curves with higher separation points at 0.4 and 0.65, respectively. It is clear that the RF algorithm produces a predicted probability line closest to the target ideal line. RF produced most probabilities toward 0 and 1, while most of the reliability curves are close to the target line, except for the region between 0.5 to 0.8. This discrepancy is caused by the nominal data containing propagated arcing noise, pushing the discrimination probability higher.

The advantages of RF can be further measured using the Brier loss score: the mean squared difference between the predicted probability and the actual outcome [50]. The value ranges from 0 to 1, and the lower Brier loss comes from better algorithms. Results in Table III shows that RF has the lowest Brier loss and the highest accuracy. Notice that the accuracy scores are not significantly different between RF and SVM, even though there is a significant difference in Brier loss. This is because the calibration process did not change the default discrimination threshold of all ML algorithms. However, making the predicted probabilities more accurate will be more useful when determining the final prediction under uncertainty. This study utilizes predicted probabilities to identify arc fault locations. Therefore, the predicted probability is an important performance indicator besides accuracy, in which RF outperforms the other three algorithms. In contrast, the higher recall and precision scores come from better algorithms. A precision score is the number of true positives divided by the number of true positives plus false positives. All algorithms achieved a relatively high precision score of 0.99. At the same time, the recall score is the number of true positives divided by the number of true positives plus false negatives. The SVM obtained the lowest recall score of 0.9, while RF obtained the highest score of 0.99.

Label	Brier Loss	Precision	Recall	Accuracy %
SVM	0.0504	0.9907	0.9018	94.85
GNB	0.0313	0.9992	0.9285	96.56
ADA	0.0271	0.9901	0.9860	98.75
RF	0.0088	0.9963	0.9897	98.98

Table III. Probability calibration parameters.

As mentioned earlier, even RF experiences discrepancies between predicted probability and target probability. To reduce these discrepancies, calibration techniques can be adopted [51]. Calibration is essentially a comparison between the actual output and the target (expected) output. This can be used to generate a distribution of predicted probability similar to that observed in training and testing data. Two commonly used methods: Sigmoid and Isotonic functions, are explored in this study for algorithm calibration.

1) *Sigmoid Function*: The Sigmoid function can be used to generate calibrated probabilities by using Platt's logistic model [52]:

$$P(y = 1|f_i) = \frac{1}{1 + \exp(Af_i + B)} \quad (8)$$

where y_i is the true label of the sample i and f_i is the output of the uncalibrated algorithm for sample i . A and B

are real numbers to be determined when fitting the function using maximum likelihood:

$$\arg \min_{A,B} (-\sum_i y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (9)$$

This method assumes the reliability curve can be corrected by applying a Sigmoid function, which is empirically verified in the case of SVM with different kernels [52]. However, it was suggested that the Sigmoid function works the best if the calibration error is symmetrical, where the algorithm outputs each class normally distributed with the same variance [53]. This becomes problematic when an ML algorithm produces highly imbalanced classification, and the outputs do not have equal variance. This issue can be seen with the sigmoid calibrated SVM across testing data of one, two, and three CPLs shown in Fig. 8, where the Sigmoid function did not change much for the SVM, GNB, and ADA. In contrast, RF reliability curves were corrected with the Sigmoid function, particularly from 0.5 to 0.8 of mean predicted probability across all three testing data sets. Sigmoid function works the best when the original curve resembles a Sigmoid curve, which was the case for the reliability curve of RF. SVM and GNB reliability curve resembled a transposed Sigmoid, which explains the limited calibration effectiveness for them. Moreover, Sigmoid method is in essence a regression of the predicted probabilities. Therefore, large deviations such as at 0.7 of the RF reliability curve get smoothed out.

2) *Isotonic Function*: Alternatively, the Isotonic function can be used to calibrate predicted probabilities that outputs a step-wise non-decreasing function, which can be assumed by:

$$\hat{m} = \sum (y_i - \hat{f}_i)^2 \quad (10)$$

where y_i is the true label of sample i and $\hat{f}_i$ is the output of the calibrated ML algorithm for sample i . This is a step-wise function that is expected to work better with linear separations like SVM. However, it is prone to overfitting with a small number of training datasets [54]. When training data is sufficient, it can achieve better performance [51]. The improvement can be seen from SVM and GNB, where their reliability curves are closer to the target calibrated curve across all three testing data sets, compared to Sigmoid function results, as shown in Fig. 8. In contrast, the Isotonic calibrated RF performs worse than the Sigmoid function, where the discrimination threshold goes from too large to too small between 0.5 and 0.8 of mean predicted probability.

Since uncalibrated RF has better performance than the other three algorithms based on predicted probability distribution, the calibration function that shows the largest improvement for the RF algorithm, which is the Sigmoid function, is then selected. In this study, the Sigmoid function has been used to calibrate all RF algorithms in LDs, before the output is sent to the MD.

D. Evaluation of Local Detector's Errors

Once the calibrated RF algorithm is obtained, the fault discrimination threshold plays a vital role in determining the

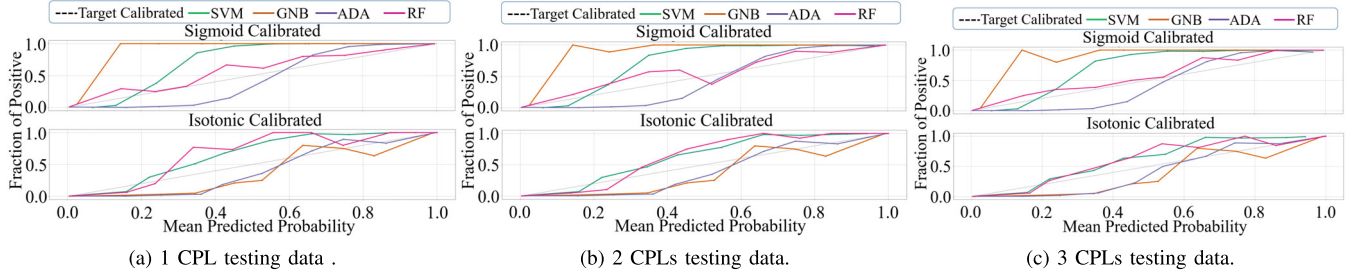


Figure 8. Probability calibration with different number of CPLs testing data.

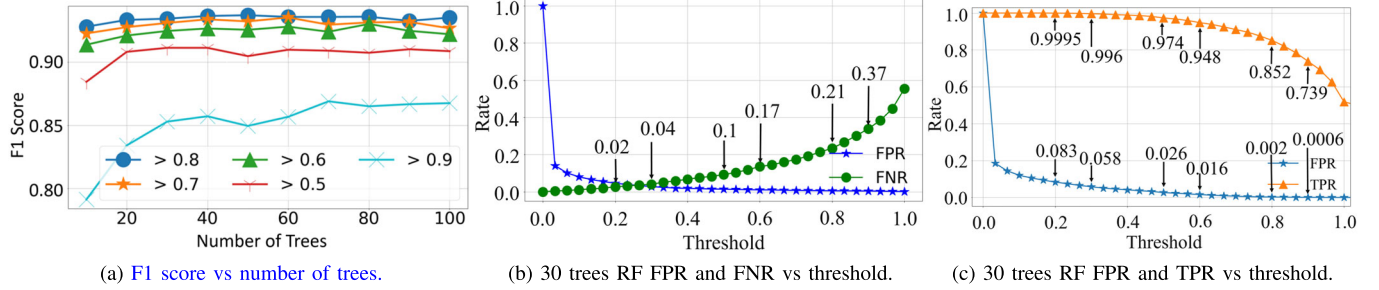


Figure 9. RF algorithm evaluation.

final prediction. When the predicted probability output by the RF algorithm in LD is higher than the set threshold, it is considered a fault has occurred in the corresponding branch. A low fault discrimination threshold can then cause mistriggers more often. A high threshold can potentially classify arc fault as nominal, which means the fault will not be detected in time. RF has a default discrimination threshold to determine class labels at 0.5. For instance, if a predicted probability of arc fault is greater than 0.5, then the final decision is arc fault for a single load, as can be seen in [33], [37]. However, the arcing noise propagates to unfaulted loads in a large electrical system and can produce an arc fault probability greater than 0.5. Therefore, this threshold must be fine-tuned to reduce the chance of mistriggers or missed faults.

Different discrimination threshold values were analyzed and benchmarked. One of the metrics to measure the level of errors is the F1 score. It is a measurement that combines the precision score and recall score of an algorithm by taking their harmonic mean, as shown in (11). It ranges from 0 to 1, and 1 is the best score. Fig. 9a illustrates an RF algorithm's F1 measurement with different numbers of trees and different threshold values. The threshold of 0.9 leads to the lowest F1 score. This is because many arc fault arc data were classified as nominal with such a high threshold value. The default threshold of 0.5 achieved the second-lowest F1 score. The threshold of 0.8 performed the best compared to others, with the F1 score of 0.94.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (11)$$

Four individual types of errors were further analyzed: true positive (TP), true negative (TN), false positive (FP), and false negative (FN):

- True positive rate (TPR): classify fault data as fault
- False positive rate (FPR): classify nominal data as fault
- True negative rate (TNR): classify nominal data as nominal
- False negative rate (FNR): classify fault data as nominal

To maintain the highest safety of the system, the FNR must be as low as possible to minimize the chance of fire hazards. The mistrigger rate (FPR) should also be at a manageable level. Therefore, a discrimination threshold should be selected to create a practical tradeoff between FPR and FNR for arc fault identification. Fig. 9b illustrates the FPR and FNR as the discrimination threshold increases from 0 to 1. At a high discrimination threshold value of 0.9, it has the highest FNR and the lowest FPR because many fault data were considered nominal. A low threshold of 0.2 has the lowest FNR of 0.02 and a high FPR. With the highest accuracy score at the threshold of 0.8, the FNR is at 0.21. We consider this FNR acceptable because if the threshold is lower to 0.6, accuracy is much lowered while the benefit of FNR is quite limited (0.04).

If the model has a high FPR, the algorithm classifies many nominal datasets as arc faults. This misclassification is mostly caused by arcing noise propagation to nominal CPLs, causing the features to resemble those of arc faults. Fig. 9c illustrates the trends of FPR and TPR with different threshold values. With a threshold value of 0.8, FPR is very low at 0.002. TPR is 0.852, which is high enough to ensure that most arc fault prediction is a true positive.

The area under the curve (AUC) represented the degree of separability between classes and was also calculated from the Receiver Operating Characteristic (ROC) curve for the 0.8 discrimination threshold algorithm. It ranges from 0 to 1, and 1 is the best score. All the models were trained with a different number of trees achieving 0.99, as shown in Fig. 10. The

above detailed error results further verify the effectiveness of a threshold value at 0.8. As a result, RF with a discrimination threshold of 0.8 is selected to perform arc fault identification at the LD level when the MD is unavailable.

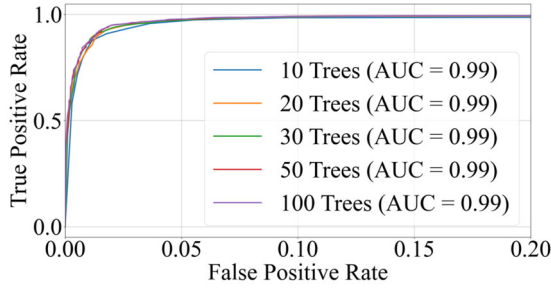


Figure 10. RF algorithms with Different numbers of trees ROC curve.

E. A RF-Based Master Detector

In order to accurately locate the faulty branch with arcing noise propagation, an MD that communicates with all LDs is developed in this study. An RF-based algorithm is developed using the predicted probability values generated and sent by LDs. Therefore, communication between the MD and the LDs is critical. It allows the MD to receive predicted probabilities from all the LDs in order to make the final decision. TCP/IP establishes the communication channel between LDs and MD. Each LD contains a unique IP address to identify the corresponding CPL. In this study, the possibility of communication failure is also considered. Three scenarios are presented below.

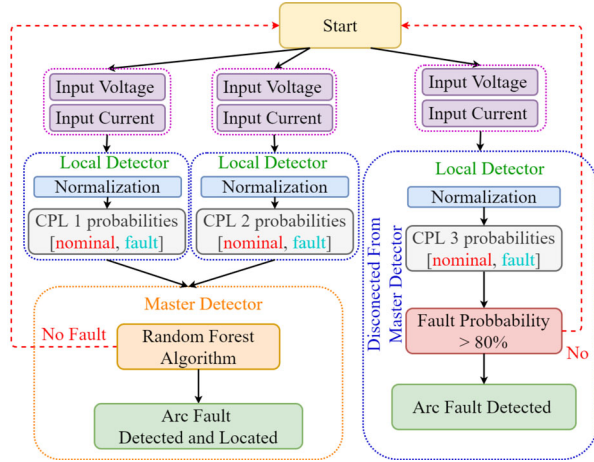


Figure 11. MD and LDs operation flow diagram.

1) **Complete Communication:** Each CPL is equipped with an RF-based LD directly communicating with the MD. The RF-based MD was trained with experimentally collected predicted probabilities for all possible classes with 30 trees. In this scenario, the communication from all three LDs (from CPL 1, CPL 2, and CPL 3) to MD is present. This allows the MD to receive predicted probabilities from all LDs and determine the final decision, as shown on the left side of Fig. 11. When an arc fault is detected, the MD outputs a flag signal corresponding to a CPL.

2) **No Communication:** The right side of the Fig. 11 shows the detection process when an LD operates independently because of communication failure to the MD. With a total loss of communication, all LDs will automatically switch to independent operation. In this way, if any LD generates an arc fault probability greater than 0.8, it outputs a flag signal.

Since no MD exists, two LDs can output high flag signals simultaneously caused by an arc fault with a strong arcing noise propagation. A sequential shutoff can be carried out using the flag signal to shut off the corresponding CPL. For instance, if the first CPL is shut off and arcing noise is isolated. Hence, arc fault probability from the second CPL is lowered, which will not be shut off. Unless arcing noise still exists, the second CPL can be shut off while the first CPL is turned back on.

3) **Partial Communication :** Finally, this arc fault identification method can operate when any LD is disconnected from the MD. In this case, there is one disconnected LD, while the other two LDs are still connected to the MD, as shown in Fig. 11. The disconnected LD can automatically switch to be independently operated, while MD continues to operate nominally without any change.

Similar to the previous scenario, sequential shutoff can be performed when multiple flag signals are presented simultaneously. The CPL corresponding to the MD's high flag signal can be shut off first. If the arcing noise is isolated, then the independent LD continues to operate nominally. Otherwise, the second CPL can be shut off while the first CPL is turned back on.

IV. EXPERIMENTAL VERIFICATION AND SIMULATION

A. Hardware verification in ZED testbed

Three PYNQ Z1 boards and an Intel-based Linux machine were deployed to implement the RF algorithms as the three LDs and MD, respectively. Each CPL is equipped with an LD that communicates with the MD using TCP/IP. The test setup is the same as introduced in Section II. The current and voltage signals were measured and collected with Verivolt differential voltage sensors and Allegro 50 A current sensors. Each RF-based LD board calculated the features and predicts the fault probabilities which were then sent to the MD. Once the MD obtained the fault probabilities from all LDs, an RF algorithm determines a final decision and outputs a flag signal.

1) **Sustained Series Arc Fault Detection:** In Fig. 12a, an arc fault was introduced at t_0 . The LD at CPL 1 flag signal turned to high (1) at t_1 . CPL 2 and CPL 3 flag signals remained low (0) the entire time. The detection latency is 20.8 ms. In Fig. 12b, arc fault was introduced at CPL 2 instead. Arcing noise propagated to other CPLs, and produced similar noise patterns in I_{1in} and I_{3in} . However, the developed MD is able to differentiate the real fault branch from the unfaulted branch suffering arcing noise propagation. It can be seen that only CPL 2 flag signal turned high (1), with the others remained low (0). The detection latency is 47.9 ms. In addition, as soon as the arc fault was extinguished, the CPL 2 flag signal turned back to low (0).

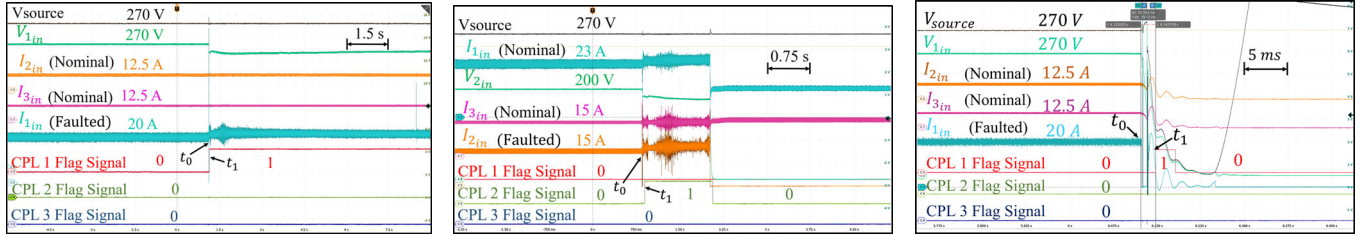


Figure 12. Arc fault identification in 270 V ZED testbed.

2) *Intermittent Arc Fault Detection*: Besides the sustained arc faults, intermittent arc faults can also occur when there is a temporary contact [55], [56]. An intermittent arc fault creates high current spikes, as shown in Fig. 12c. An intermittent arc fault was introduced at CPL 1 at t_0 . The faulted current, I_{1in} , reached approximately 80 A, and each spike lasted approximately 2 – 3 ms. Unlike a sustained arc fault in Fig. 12a, the initial arc fault spikes were not too high that the power supply was able to compensate for the fluctuations, and the unfaulted branches were unaffected. Whereas the high intermittent arc fault spikes were extremely high, causing other branches to compensate, creating fluctuations in the unfaulted branches initially. LD at CPL 1 detected the fault and output a high flag signal at t_1 with 25.5 ms latency time.

3) *Series Arc Fault Isolation*: Preliminary arc fault isolation was conducted using the MD's flag signal to control the CPL's operations. The flag signal was sent to Opal 4510 to turn off the PWM generator when the flag signal was high; otherwise, the CPL remained nominally operated. Fig. 13 illustrates arc fault identification and isolation in a two parallel-connected CPLs. Similar to previous tests, both CPLs were operated nominally, then series arc fault was introduced at CPL 1. The detector output a high flag signal responding to the arc fault, which then turns off the PWM generator. As a result, CPL 1 was shut off, while CPL 2 continued nominally. The isolation latency time was 24 ms.

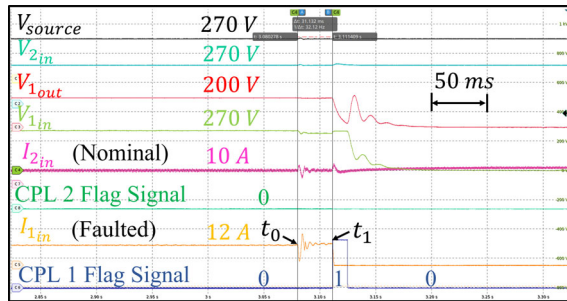


Figure 13. Series arc fault isolation in two parallel-connected CPLs.

4) *Discussion on Detection Latency time*: The detection latency can be broken down into three parts: pre-processing, predicting, and communication between the LDs and MD. The hardware's internal clock can be used to obtain a closer look and deeper understanding of the latency time at each step. The data window time is 3 ms, during which time data points are measured and stored. Then, it takes about 3 ms to calculate the features for LDs. This fast computation time is

the one advantage of using very simple yet effective features. The 30 trees RF-based LD algorithm has a predicting time of approximately 10 ms. It takes approximately 2 ms to 3 ms for the MD to receive fault probabilities from all the LDs. It takes another 1.8 ms to 2 ms for the 30 tree RF-based MD to output the final decision. The total time measured from the hardware ranges from 19.8 ms to 26 ms for a zonal distribution setup with the MD. However, if the first data window of an arc fault is not detected, the detection latency time can be prolonged until it is successfully detected.

B. Simulation in ZED testbed

The developed arc fault identification method was verified through simulation using experimental ZED data. The results are shown in Fig. 14. CPL 1 (red) in Zone 1 was operating at 23 A and 270 V. CPL 2 (green) and CPL 3 (blue) in Zone 2 were operating at 15 A and 200 V.

First, an arc fault was introduced in CPL 1 while CPL 2 remained normal the entire time. Communication from CPL 3 LD to MD was turned off at t_0 . When arc fault was introduced, CPL 1 input current suddenly dropped and recovered with additional arcing noise. The input voltage of CPL 1 also dropped due to the arc resistance. The arcing noise did not affect CPL 2 and CPL 3, as discussed in Section II. It can be seen that in this case, the MD successfully recognized the arc fault location to be at CPL 1, and therefore, the CPL 1 was turned off to clear the fault. Although CPL 3 LD is not communicating with MD, this independently operating LD output remained low because of the lack of arcing noise propagation from Zone 1 to Zone 2.

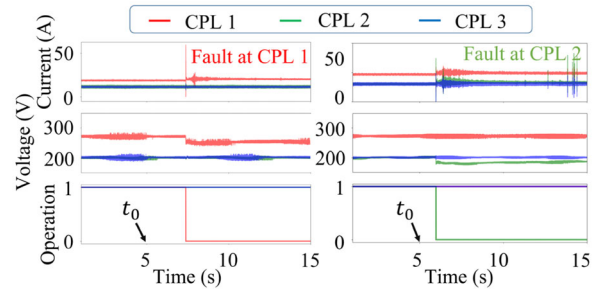


Figure 14. Arc fault identification RF-based MD simulation.

Next, an arc fault was introduced at CPL 2. Same as the last scenario, LD at CPL 3 was disconnected during the arcing at t_0 , as shown in the right column of Fig. 14. However, the arc fault identification method operated accurately, turning off

the faulted CPL 2 while the CPL 1 and CPL 3 were still on (1).

To further investigate the impact of communication availability, three scenarios: complete communication, partial communication, and no communication, were evaluated individually using metric scores: precision, recall, and accuracy, as shown in Table IV. Similar to performance metric benchmarks in Section III, testing datasets were selected randomly from the collected experimental data, where arc fault was introduced in either CPL 1 or CPL 2. In the partial communication scenario, the LDs from CPL 1 and CPL 2 were connected to the MD, while LD from CPL 3 operated independently. The partial communication accuracy value was obtained by averaging the accuracy of MD and the independent LD for CPL 3. In the no communication scenario, the accuracy was the average among all three independent LDs. It can be seen that the arc fault identification method performed more accurately when LDs were connected to the MD. The complete communication achieved the highest accuracy of 98.865 %.

Communication Status	Precision	Recall	Accuracy (%)
Complete	0.9839	0.9839	98.865
No	0.9837	0.9837	98.372
Partial	0.9838	0.9838	98.619

Table IV. MD methods metric scores.

V. CONCLUSION

A ZED testbed was designed and built to study the series arc fault behavior and the arcing noise propagation when introduced into the system. A large amount of arc fault was collected to train and test the ML algorithm as an arc fault detection algorithm. Consequently, the RF algorithm was selected as a local detector to monitor CPL and output predicted fault probability. The RF algorithm was thoroughly evaluated using accuracy and error evaluations. The accuracy was analyzed based on different numbers of trees and discrimination threshold values. While error evaluation was investigated using TPR, FPR, TNR, and FNR. The tradeoff between these four metrics was analyzed, and an optimized fault discrimination threshold was selected to reduce arc fault mistrigger.

Another RF algorithm was developed to accumulate the predicted probabilities from LDs to identify an arc fault location as a master detector (MD). All the LDs communicate with the MD using TCP/IP. The MD operates nominally and outputs flag signals if two or more LDs are connected. Otherwise, LDs are independently operated and output their flag signal. Three different arc fault identification scenarios were simulated regarding the communication status between LDs and MD.

The LDs and MD were then implemented on low-cost hardware. The LD utilizes the PYNQ Z1's FPGA functionality to receive data at a high sampling rate of 120 kS/s. All the computationally intensive functions were executed on the MD machine, containing a much faster CPU, Intel 10900K. The proposed LDs and MD approach has achieved outstanding

accuracy, reliability, and low detection latency time. The arc fault detection latency on two parallel-connected CPLs and ZED testbed range from 20.8 ms to 47.9 ms, which is applicable to UL 1699B standard. Integrating flag signal with power electronics control was carried out to isolate series arc fault with an isolation latency of 24 ms. Ultimately, the flag signal can be utilized to control circuit breakers or contactors for proper arc fault isolation.

ACKNOWLEDGMENT

This research is supported by the Air Force Research Laboratory (AFRL) Intelligent Power System (IPS) program, contract number: FA 8650-19-D2905

REFERENCES

- [1] G. Sulligoi, D. Bosich, A. Vicenzutti, and Y. Khersonsky, "Design of zonal electrical distribution systems for ships and oil platforms: Control systems and protections," *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5656–5669, 2020.
- [2] A. Maqsood and K. A. Corzine, "Integration of z-source breakers into zonal dc ship power system microgrids," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 5, no. 1, pp. 269–277, 2016.
- [3] K. J. Karimi, "Future aircraft power systems-integration challenges," *The Boeing Company*, 2007.
- [4] B. Rahrovi and M. Ehsani, "A review of the more electric aircraft power electronics," in *2019 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2019, pp. 1–6.
- [5] L. Zhu, J. Li, Y. Liu, and S. Ji, "Initial features of the unintended atmospheric pressure dc arcs and their application on the fault detection," *IEEE Transactions on Plasma Science*, vol. 45, no. 4, pp. 742–748, 2017.
- [6] S. Chen, X. Li, Z. Xie, and Y. Meng, "Time-frequency distribution characteristic and model simulation of photovoltaic series arc fault with power electronic equipment," *IEEE Journal of Photovoltaics*, vol. 9, no. 4, pp. 1128–1137, 2019.
- [7] M. Jalil, H. Samet, and T. Ghanbari, "Time-variant schwarz based model for dc series arc fault modeling in photovoltaic systems," *IEEE Journal of Photovoltaics*, 2022.
- [8] H.-P. Park, M. Kim, and S. Chae, "Smart dc optimizer for dc series arc fault detection and extinguishing," *IEEE Transactions on Power Electronics*, 2022.
- [9] S. Zhao, Y. Wang, F. Niu, C. Zhu, Y. Xu, and K. Li, "A series dc arc fault detection method based on steady pattern of high-frequency electromagnetic radiation," *IEEE Transactions on plasma science*, vol. 47, no. 9, pp. 4370–4377, 2019.
- [10] S. Wei, Q. Yang, Y. Qi, T. Zhou, W. Liao, and J. Sun, "A location method for arc fault based on electromagnetic radiation signals and tdoa," in *2022 5th International Conference on Energy, Electrical and Power Engineering (CEEPE)*. IEEE, 2022, pp. 427–431.
- [11] S. Zhang, N. Qu, T. Zheng, and C. Hu, "Series arc fault detection based on wavelet compression reconstruction data enhancement and deep residual network," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022.
- [12] X. Yao, V. Le, and I. Lee, "Unknown input observer-based series dc arc fault detection in dc microgrids," *IEEE Transactions on Power Electronics*, vol. 37, no. 4, pp. 4708–4718, 2022.
- [13] K. Gajula, L. K. Marepalli, X. Yao, and L. Herrera, "Recursive least squares and adaptive kalman filter-based state and parameter estimation for series arc fault detection on dc microgrids," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 10, no. 4, pp. 4715–4724, 2022.
- [14] J. A. Momoh and R. Button, "Design and analysis of aerospace dc arcing faults using fast fourier transformation and artificial neural network," in *2003 IEEE Power Engineering Society General Meeting (IEEE Cat. No. 03CH37491)*, vol. 2. IEEE, 2003, pp. 788–793.
- [15] R. D. Telford, S. Galloway, B. Stephen, and I. Elders, "Diagnosis of series dc arc faults—a machine learning approach," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1598–1609, 2016.
- [16] V. Le and X. Yao, "Ensemble machine learning based adaptive arc fault detection for dc distribution systems," in *2019 IEEE Applied Power Electronics Conference and Exposition (APEC)*. IEEE, 2019, pp. 1984–1989.

- [17] V. Le, X. Yao, C. Miller, and T.-B. Hung, "Arc fault detection in dc distribution using semi-supervised ensemble machine learning," in *2019 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, pp. 2939–2945.
- [18] J. Johnson, C. Oberhauser, M. Montoya, A. Fresquez, S. Gonzalez, and A. Patel, "Crosstalk nuisance trip testing of photovoltaic dc arc-fault detectors," in *2012 38th IEEE Photovoltaic Specialists Conference*. IEEE, 2012, pp. 001 383–001 387.
- [19] X. Feng, L. Qi, and J. Pan, "A novel fault location method and algorithm for dc distribution protection," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 1834–1840, 2017.
- [20] X. Feng, Q. Xiong, D. Wardell, A. L. Gattozzi, S. M. Strank, and R. E. Hebner, "Extra-fast dc distribution system protection for future energy systems," *IEEE Transactions on industry applications*, vol. 55, no. 4, pp. 3421–3430, 2019.
- [21] E. Christopher, M. Sumner, D. W. Thomas, X. Wang, and F. de Wildt, "Fault location in a zonal dc marine power system using active impedance estimation," *IEEE Transactions on Industry Applications*, vol. 49, no. 2, pp. 860–865, 2013.
- [22] L. Herrera, W. Zhang, and J. Wang, "Stability analysis and controller design of dc microgrids with constant power loads," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 881–888, 2015.
- [23] A. Maqsood, D. Oslebo, K. Corzine, L. Parsa, and Y. Ma, "Stft cluster analysis for dc pulsed load monitoring and fault detection on naval shipboard power systems," *IEEE Transactions on Transportation Electrification*, 2020.
- [24] Q. Xiong, X. Feng, A. L. Gattozzi, X. Liu, L. Zheng, L. Zhu, S. Ji, and R. E. Hebner, "Series arc fault detection and localization in dc distribution system," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 1, pp. 122–134, 2019.
- [25] K. Gajula and L. Herrera, "Detection and localization of series arc faults in dc microgrids using kalman filter," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2020.
- [26] M. Kavi, Y. Mishra, and M. Vilathgamuwa, "Dc arc fault detection for grid-connected large-scale photovoltaic systems," *IEEE Journal of Photovoltaics*, vol. 10, no. 5, pp. 1489–1502, 2020.
- [27] 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*, vol. 29, no. 6, pp. 3103–3115, 2013.
- [28] M. Ayadi, E. Maggioli, H. Leite, and H. Martins, "Communication requirements for a hybrid vsc based hvdc/ac transmission networks state estimation," *Energies*, vol. 14, no. 4, p. 1087, 2021.
- [29] P. K. Shinde and N. R. Chopde, "Performance analysis of tcp congestion control algorithms," *International Journal of Science, Engineering and Computer Technology*, vol. 4, no. 4, p. 100, 2014.
- [30] N. A. Muhamad, I. V. Musa, Z. A. Malek, and A. S. Mahdi, "Clas-sification of partial discharge fault sources on sf6 insulated switchgear based on twelve by-product gases random forest pattern recognition," *IEEE Access*, vol. 8, pp. 212 659–212 674, 2020.
- [31] H.-Y. Su, Y.-J. Lin, and C.-C. Chu, "Applications of decision tree and random forest methods for real-time voltage stability assessment using wide area measurements," in *Wide Area Power Systems Stability, Protection, and Security*. Springer, 2021, pp. 373–391.
- [32] R. Fezai, K. Dhibi, M. Mansouri, M. Trabelsi, H. Mansour, K. Bouzrara, H. Nounou, and M. Nounou, "Effective random forest based fault detection and diagnosis for wind energy conversion systems," *IEEE Sensors Journal*, 2020.
- [33] J. Jiang, W. Li, Z. Wen, Y. Bie, H. Schwarz, and C. Zhang, "Series arc fault detection based on random forest and deep neural network," *IEEE Sensors Journal*, 2021.
- [34] V. Le, X. Yao, C. Miller, and B.-H. Tsao, "Series dc arc fault detection based on ensemble machine learning," *IEEE Transactions on Power Electronics*, vol. 35, no. 8, pp. 7826–7839, 2020.
- [35] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and xgboost," *Ieee Access*, vol. 6, pp. 21 020–21 031, 2018.
- [36] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, p. 1088, 2019.
- [37] V. Le, X. Yao, C. Miller, and T.-B. Hung, "Series arc fault detection and localization in dc distribution based on master controller," in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, 2020, pp. 2499–2504.
- [38] K. Emadi and M. Ehsani, "Aircraft power systems: technology, state of the art, and future trends," *IEEE Aerospace and Electronic Systems Magazine*, vol. 15, no. 1, pp. 28–32, 2000.
- [39] W. Cao, B. C. Mecrow, G. J. Atkinson, J. W. Bennett, and D. J. Atkinson, "Overview of electric motor technologies used for more electric aircraft (mea)," *IEEE transactions on industrial electronics*, vol. 59, no. 9, pp. 3523–3531, 2011.
- [40] Q. Sun, Y. Li, D. Ma, Y. Zhang, and D. Qin, "Model predictive direct power control of three-port solid-state transformer for hybrid ac/dc zonal microgrid applications," *IEEE Transactions on Power Delivery*, 2021.
- [41] H. Tu, H. Feng, S. Srdic, and S. Lukic, "Extreme fast charging of electric vehicles: a technology overview," *IEEE Transactions on Transportation Electrification*, vol. 5, no. 4, pp. 861–878, 2019.
- [42] C. He, L. Mu, and Y. Wang, "The detection of parallel arc fault in photovoltaic systems based on a mixed criterion," *IEEE Journal of Photovoltaics*, vol. 7, no. 6, pp. 1717–1724, 2017.
- [43] M. Naidu, T. Schoepf, and S. Gopalakrishnan, "Arc fault detection scheme for 42-v automotive dc networks using current shunt," *IEEE Transactions on Power Electronics*, vol. 21, no. 3, pp. 633–639, 2006.
- [44] R. F. Ammerman, T. Gammon, P. K. Sen, and J. P. Nelson, "Dc-arc models and incident-energy calculations," *IEEE Transactions on Industry Applications*, vol. 46, no. 5, pp. 1810–1819, 2010.
- [45] K. Yuri, D. Norbert, D. Roger, E. Terry, P. Joseph, A. John, S. B. Curtin, B. Paul, B. Dushan, B. Joseph *et al.*, "Ieee standard for power electronics open system interfaces in zonal electrical distribution systems rated above 100 kw," 2012.
- [46] A. Shekhar, L. Ramírez-Elizondo, S. Bandyopadhyay, L. Mackay, and P. Bauera, "Detection of series arcs using load side voltage drop for protection of low voltage dc systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6288–6297, 2017.
- [47] X. Yao, "Study on dc arc faults in ring-bus dc microgrids with constant power loads," in *2016 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2016, pp. 1–5.
- [48] S. Hartshorn, "Machine learning with random forests and decision trees: A visual guide for beginners," *Kindle edition*, 2016.
- [49] B. Zadrozny and C. Elkan, "Transforming classifier scores into accurate multiclass probability estimates," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002, pp. 694–699.
- [50] G. W. Brier *et al.*, "Verification of forecasts expressed in terms of probability," *Monthly weather review*, vol. 78, no. 1, pp. 1–3, 1950.
- [51] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 625–632.
- [52] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
- [53] M. Kull, T. M. Silva Filho, and P. Flach, "Beyond sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration," *Electronic Journal of Statistics*, vol. 11, no. 2, pp. 5052–5080, 2017.
- [54] A. K. Menon, X. J. Jiang, S. Vembu, C. Elkan, and L. Ohno-Machado, "Predicting accurate probabilities with a ranking loss," in *Proceedings of the... International Conference on Machine Learning. International Conference on Machine Learning*, vol. 2012. NIH Public Access, 2012, p. 703.
- [55] A. Yaramasu, Y. Cao, G. Liu, and B. Wu, "Aircraft electric system intermittent arc fault detection and location," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 40–51, 2015.
- [56] J.-R. Riba, M. Moreno-Eguilaz, and J. A. Ortega, "Arc fault protections for aeronautic applications: A review identifying the effects, detection methods, current progress, limitations, future challenges, and research needs," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.