Hierarchical Data-Driven Protection for Microgrid with 100% Renewable Penetration

Preprint

Ahmed S. Zamzam and Jing Wang

National Renewable Energy Laboratory

Presented at the 2023 IEEE Energy Conversion Congress and Exposition (ECCE)
Nashville, Tennessee
October 29–November 2, 2023

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-84518
November 2023

Hierarchical Data-Driven Protection for Microgrid with 100% Renewable Penetration

Preprint

Ahmed S. Zamzam and Jing Wang

National Renewable Energy Laboratory

Suggested Citation

Zamzam, Ahmed S., and Jing Wang. 2023. *Hierarchical Data-Driven Protection for Microgrid with 100% Renewable Penetration: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5D00-84518. https://www.nrel.gov/docs/fy24osti/84518.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-84518
November 2023

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at NREL. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

Hierarchical Data-Driven Protection for Microgrids with 100% Renewables

Ahmed S. Zamzam
Power Systems Engineering Center
National Renewable Energy Laboratory
Golden, Colorado
Email: ahmed.zamzam@nrel.gov

Jing Wang
Power Systems Engineering Center
National Renewable Energy Laboratory
Golden, Colorado
Email: jing.wang@nrel.gov

***Abstract*—The accurate detection and isolation of faults is critical for the reliable operation of microgrids (MGs). Traditional protection approaches are even more challenged for 100% renewable MGs because inverter-based resources (IBRs) are the only sources for fault current which are usually low and unpredictable/non-uniform. This calls for new protection scheme that can identify IBR fault responses and detect faults in MGs. Data-driven based protection can learn the pattern of IBR fault responses and make the correct decision to identify faults. Therefore, this paper presents a data-driven approach for fault localization in island MGs. The approach builds a training dataset of comprehensive fault scenarios that can be used to learn fault characteristics from processed measurements. The localization task is modeled as a binary classification problem at each relay, which simplifies the learning process. Then, a hierarchical decision mechanism is used to identify the fault location. The proposed approach is assessed using an exemplary MG with several grid-forming (GFM) and grid-following (GFL) inverters, where accurate estimation of fault location is achieved. The data-driven based protection approach developed in this paper provides a generic framework and useful guidance for power system protection engineers to achieve reliable protection for MGs with 100% renewables.**

Keywords*—*Microgrids, renewable integration, protection, fault localization, data-driven methods.

I. Introduction

Protection of microgrids (MGs) is the most critical challenge to be resolved to ensure their reliable and safe operation for power system resilience. As more MGs become 100% renewables, this challenge is magnified [1]. Because inverter-based resources (IBRs) are the only source of fault current in such MGs, and their fault currents are usually low and unpredictable/non-uniform [2]. Especially the fault current levels of IBRs tend to be close to normal operation when the MG is in islanded mode. Such low current levels hinder the ability of traditional protection schemes to detect and isolate faults. In addition, the various control algorithms of inverters, which are dictated by the inverter control algorithms and by the MG operational conditions, make the fault response of inverters hard to characterize and detect. This calls for new protection scheme that can identify IBR fault responses and detect faults in MGs [3].

Due to the varying fault characteristics, classical protection schemes such as over-current or under-voltage protection may fail to perform accurate identification and isolation of faults. Additionally, adapting the settings of the protective relays often fails due to the fault responses variations depending on operational conditions that cannot be completely observed by the operator. For example, the fault response can change drastically with varying renewable energy penetration, which requires relays to obtain state information of renewable energy sources if it was to change the relay settings [4]. Recently, machine learning approaches have been proposed for fault classification and localization tasks [5], [6], [7], [8], [9], [10]. Existing work developed approaches that either assume centralized computations and learning [6], [7], [8], [10], which requires massive communication between relays, or approaches that focus on fault localization and classification in very small MGs, which can be protected by a single relay [5], [9]. Unfortunately, the majority of machine learning approaches proposed in the literature for identification and localization of faults suffer also from being trained under fixed loading and renewable penetration levels, which limits the approaches ability to perform accurately when these conditions change. These limitations hinder the applicability of the machine learning methods in MGs with varying sizes.

In this paper, we introduce a novel approach for fault localization in islanded MGs with 100% renewable penetration. The proposed method leverages a local learning technique, followed by a hierarchical localization approach to accurately identify fault locations. Our primary objective is to design a fault localization approach that can be effectively scaled for MGs with varying sizes and can adapt to diverse operational conditions and fault scenarios. To achieve this, we begin by developing a robust dataset generation process for simulating faults in MGs with significant inverter-based resource (IBR) penetration, ensuring the approach's applicability under varying conditions. The generated datasets are then used to train classifiers capable of distinguishing between upstream and downstream faults at any relay, irrespective of the fault characteristics. Furthermore, we introduce a hierarchical localization scheme that utilizes the results from local relay classifiers to pinpoint the fault location. This not only enables relays to localize faults without excessive data exchange but also ensures scalability for larger MG systems. The proposed method establishes a standardized approach that can be readily deployed in other MGs, offering a robust and efficient solution for fault localization in renewable-dominated microgrids.

Our contributions in this research have significant implications for the protection and operation of microgrids, particularly those with high renewable energy integration. Firstly, we address the pressing challenge of fault localization in islanded microgrids, which becomes even more critical as MGs transition to 100% renewables. By utilizing local learning and hierarchical approaches, our method overcomes the limitations of traditional protection schemes and global machine learning approaches that often rely on data measured from different locations. Secondly, we propose a scalable fault localization approach that requires only minimal communications between neighboring relays, offering a versatile solution that can be adapted to different microgrid configurations. Thirdly, our proposed dataset generation process ensures that the classifiers are trained on a wide range of fault scenarios, enhancing the robustness and accuracy of fault localization in diverse operational conditions. Fourthly, the localized communication requirement among neighboring relays significantly reduces the overhead and complexity of data transmission, leading to more cost-effective implementation of fault localization in large microgrid systems. Overall, our work provides a promising and practical solution to the most critical challenge of microgrid protection, contributing to the reliable and safe operation of future power systems with high renewable energy penetration.

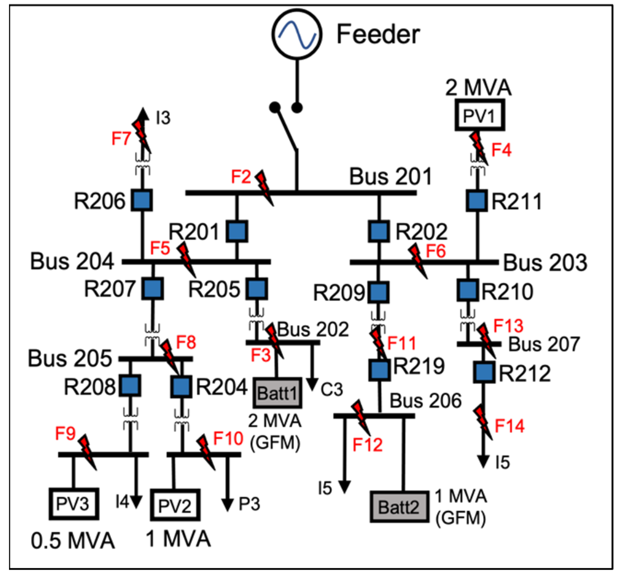


Fig. 1: Single-line-diagram of the example microgrid under study.

II. MICROGRID MODEL

An example MG is selected to develop data-driven based protection method for 100% IBR MGs. This MG under study is based on Feeder 2 of the Banshee MG which is a benchmark MG system to evaluate MG controller, protection, and cybersecurity. A schematic representation of the example MG is depicted in Fig. 1; more details on this MG can be found in [11].

This MG includes three grid-following (GFL) photovoltaic (PV) inverters and two grid-forming (GFM) battery inverters.Both GFL and GFM inverters are modelled as an average switching model with fixed DC voltage. This is to better represent the inverter dynamics than the controlled voltage source model for fault study and protection design. The GFL inverters can operate in different control modes, including external PQ control, volt-volt ampere reactive (VAR) control and fixed power factor control under maximal power point tracking (MPPT). The solar irradiance profiles are obtained from historical data, and it is assumed that all three PV systems have the same solar irradiance at any point in time. Eight representative days out of a year are selected to represent the yearly solar profiles. The GFM battery inverters work in both grid-connected and islanded operation with voltage control. The loads include both balanced and unbalanced loads. All the circuit breakers include a delay of 5 cycles to emulate the mechanical delay in real-world circuit breakers. In this section, we will discuss the modeling of GFM inverters, GFL inverters and loads in details.

A. GFM inverter modeling

Fig. 2 shows the control diagram of the GFM inverter. The GFM battery inverters are modeled in PQ control in grid-connected mode and VF control in islanded mode. And the GFM inverter uses droop control for both grid-connected (active and reactive power tracking) and islanded control (voltage and frequency control). So, there is no need to switch between grid-connected current control and islanded voltage control. This control is developed according to the latest GFM control technology of an inverter manufacture. Virtual impedance control is also added to improve the stability of the GFM inverter especially during contingency and fault events. Because it can adjust the voltage reference to have the GFM control reach the appropriate target voltage, thus not saturating the inverter and causing instability [12]. For voltage control, the traditional double-loop control structure is used, with outer-loop voltage control and inner-loop current control. This is typical control structure for droop based GFM control.

Note that the droop coefficients m_p and n_p are both 1e-5, and all variables are not in per unit. S_1 is the circuit breaker of the microgrid point of common coupling (PCC), and the integrator is only enabled if the microgrid is grid-connected. For the islanded microgrid, P_{ref} and Q_{ref} are set to be zero, so that the GFM inverters operate in droop control to share power.

B. GFL inverter modeling

In this study, the three PV GFL inverters can operate in any of the three operation modes. Fig. 3 shows the control diagram of the GFL PV inverters. The PV inverters use a phase-lock-loop (PLL) to synchronize to the grid voltage and inject active and reactive power into the grid based on the selected operation mode. Based on the selected operation mode, the corresponding power reference (P^* and Q^*) will be generated, and then the corresponding current reference in synchronous frame (I^*_{od} and I^*_{oq}). Note that dynamic current limiters with d-axis priority are applied in current references. The solar irradiance is used to add as the available power constraint to the PV inverters. Note that the PV inverters are modelled for IEEE 1547-2018 compliant, which means the fault current may vary depending on the terminal voltage/the significance of the fault situation. The detailed modeling of control and PWM of GFL PV inverters are necessary to capture the actual fault responses.

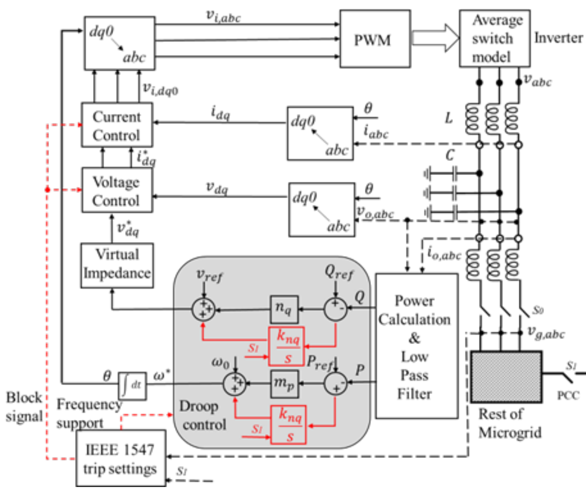


Fig. 2: Control diagram of the GFM battery inverters.

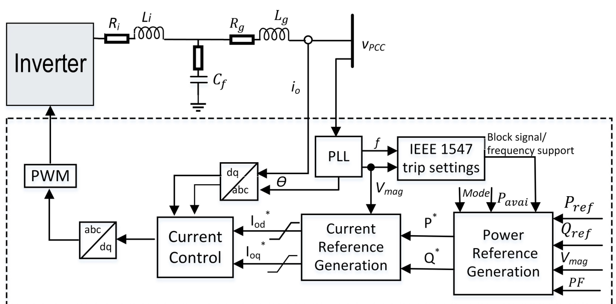


Fig. 3: Control diagram of the GFL PV inverters.

C. Load modeling

In general, there are three types of load, including the constant impedance, constant current and constant power loads. This benchmark microgrid includes both constant impedance and constant power loads. The constant power load is usually modelled as a current source to track the load profiles, however, the load may contribute fault current as it is grid voltage dependent. Therefore, all the loads are modelled as constant impedance loads, and their load values are modified externally through a script. Eight representative days out of a year are selected to represent the yearly load profiles. Note that the PV solar irradiance and load profiles are selected from the same days.

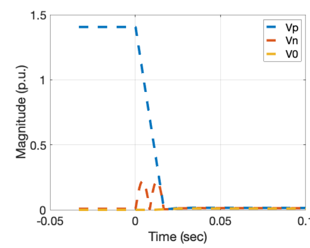
To make the load modelling more realistic, unbalanced loads are added in some of the big loads (e.g., P2). For those loads with unbalanced load, the total balanced and unbalanced load match the load capacity. The unbalanced load is assigned randomly between 0 and 0.3 p.u. of the total load, and the rest is for the balanced load. For the unbalanced load, the active and reactive power for each phase is assigned randomly with a factor smaller than 1, but total three phase is equal to 1.

III. Fault Localization Task

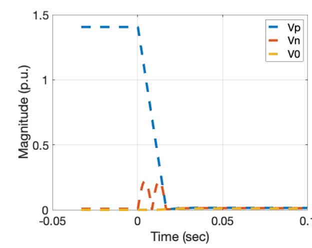
In the considered microgrid, the fault location can be in one of 13 possible locations. Naturally, different circuit breakers are responsible for isolation of faults depending on its location. Table I lists the relays that are responsible for clearing (isolating) each fault.

TABLE I: Circuit breakers responsible for clearing each fault.

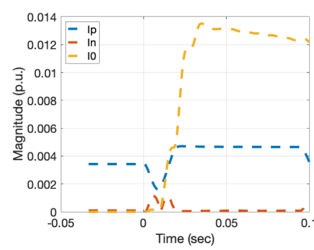
Fault	Circuit Breaker(s)	Fault	Circuit Breaker(s)
Fault 2	CB201, CB202	Fault 9	CB208
Fault 3	CB205	Fault 10	CB204
Fault 4	CB211	Fault 11	CB209, CB219
Fault 5	CB201, CB205, CB206, CB207	Fault 12	CB219
Fault 6	CB202, CB209, CB210, CB211	Fault 13	CB210, CB212
Fault 7	CB206	Fault 14	CB212
Fault 8	CB204, CB207, CB208		



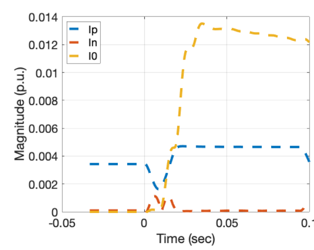
(a) Fault 2 voltage.



(b) Fault 5 voltage.



(c) Fault 2 current.



(d) Fault 5 current.

Fig. 4: The fault response recorded by CB207 in response to the same LLG at different locations (fault inception at time 0).

One major challenge faced in the protection of microgrid networks, especially when operated in isolation from the main grid, is the unidentifiability of the fault responses. That is, similar fault responses can be observed by a specific relay after different faults. This often leads to data-driven approach being unable to identify the appropriate relaying action in response to faults. The short distance between buses in microgrids also exaggerates this issue so that multiple relays in the same area can see similar fault responses. For example, the observed responses for fault 5 and fault 2 at circuit breaker (CB) 207 are almost identical; however, CB207 is supposed to isolate fault 5, but not fault 2 which is isolated by CB201. Fig. 4 shows the recorded fault responses of fault 5 and fault 2 for the same LLG fault with a very small fault resistance.

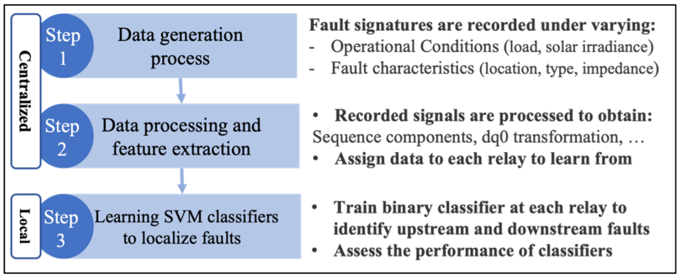


Fig. 5: Decentralized learning process.

IV. Proposed Learning Approach

Correctly identifying faults and performing the correct fault isolation procedure are crucial for the viability and resilience of MGs. The proposed approach tackles a multi-class classification problem where each class represents a possible fault location. Such problems are often notoriously hard to tackle especially when the number of possible locations increases. As mentioned, relays often observe identical fault responses for faults at different location, which makes tasking individual relays with identifying exact fault location an impossible task. Therefore, instead of using a completely centralized or decentralized data-driven fault localization method, we propose a decentralized-learning hierarchical-localization approach. The approach simplifies the process of localization into two steps: i) a decentralized learning phase where a classifier at each relay learns to classify faults into upstream and downstream faults, i.e., a binary classification problem, and ii) a hierarchical localization phase where the results of classifiers at the relays are combined to determine the location of the fault.

The main advantages of the proposed approach are as follows: i) the measurements at each relay are only processed locally instead of sharing all relays data with a centralized computing unit, ii) the learning task at each relay is a simple binary classification that is easy to learn with reduced number of data samples, and iii) combining the classification results amounts to only sharing 1 bit between neighboring relays, and thus, the localization can be done in a timely manner to isolate faults. We explain the details of the learning process and then explain the hierarchical localization approach next.

A. Decentralized Learning Process

In order to obtain classifiers that can classify faults into upstream and downstream faults, we utilize a training process that does not require sharing measurements between relays enabling it to be done in a distributed fashion.

1) Data generation process: To learn an effective representation of fault responses, data samples that are diverse and representative need to be collected. In general, many factors affect the fault response; these can be categorized into operational conditions and fault characteristics. The goal is to generate samples that cover very diverse scenarios of fault characteristics and operational conditions. The operational conditions include the load demand, the renewable generation availability, and the mode of operation of the grid-following inverters. The fault characteristics include the fault location, the fault type, and the fault impedance.

To have representative load profiles and PV generation, we use a data set of eight representative days from the measured data in [13]. During each simulation, one time instant is chosen at random from these days, where the load values as well as the PV irradiance are used to set the parameters in the simulations; therefore, the generated scenarios represent different operational conditions of the microgrid. In addition, for each scenario, the operational mode of each inverter is set to be external PQ control, volt-volt ampere reactive (VAR) control, or fixed power factor control at random. As a result of this diversity in the scenario generation process, in some scenarios the energy storage units are providing power to satisfy the loads, whereas in other scenarios the PV generation is exceeding the load demand and hence the energy storage units are absorbing power. Here, we do not vary the state of charge of the energy storage units because this timescale is beyond the protection relays' operational timescale.

In each scenario, after randomly setting the operational conditions, we select one of the 13 fault locations according to a discrete uniform distribution. In addition, we set the type of the fault randomly, i.e., the faulty phases (a, b, c, g), from a total of 11 possible fault types following a discrete uniform distribution. Then, the fault resistance value is set using:

$$z_{\text{fault}} = 0.01 + \beta_{\frac{1}{4},\frac{1}{4}} * (100 - 0.01),$$

where $\beta_{\frac{1}{4},\frac{1}{4}}$ is a realization from a Beta distribution with parameters $(\frac{1}{4}, \frac{1}{4})$. This distribution is chosen to model both cases of low-impedance and high-impedance faults. Therefore, the generated scenarios include samples of fault scenarios that are diverse and representative of faults possible to be encountered in operations.

2) Data processing and feature extraction: In each simulation, the voltage and current instantaneous values are measured at 6 KHz at each relay. These raw data include the voltage and current of all phases for the duration of the simulation, which is set to be 8.2 seconds. For each scenario, the fault is randomly placed at different points of time during the simulation period. After collecting the data measured by each relay, these measurements are post-processed to obtain variables that can aid relays in identifying faults. In our experiments, the following features are extracted for fault pattern recognition, including voltage and current sequence components (positive, negative and zero sequence), and voltage and current dq0 components. Those features are input to the classifiers.

Each fault is classified by each relay whether it is downstream or upstream from the relay location. For example, the relay at CB210 should aim to classify any instance of faults F13 and F14 as downstream faults, and classify all other instances as upstream faults. Notice that this is a simpler task than identifying a specific fault such as F13 because of the similar response observed for F14. To prepare the data for training, we attach the labels information according to which faults are upstream or downstream from each relay. In our simulation, we obtained 1000 samples representing different fault scenarios. The data of each relay is composed of the voltage and current measurements for a total of 8 cycles including two pre-fault cycles and 6 post-fault cycles. Given the 6 kHz sampling rate, this represents a total of 800 samples for each measured signal.

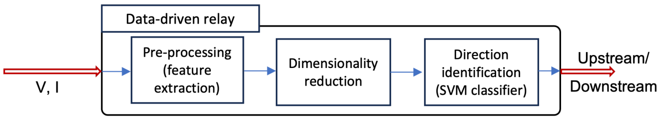


Fig. 6: Data-driven relay components.

3) Learning SVM classifiers to localize faults: We utilize support vector machine (SVM) classifiers that are trained on 700 fault samples that represent all fault types, locations, and impedances under diverse operational conditions. The remaining 300 are utilized for evaluating the performance of the proposed approach. The modelling and training of the classifiers is done using the `sklearn` package in Python. We utilize a polynomial kernel with degree 2. The regularization parameter is set separately for each relay using a k-fold validation procedure.

Due to the high-dimensional nature of the data, we use principal component analysis as a dimensionality reduction technique before passing the data to the classifier. To ensure a fair assessment of the performance, we use the training samples only to perform the principal component analysis, whereas the dimensionality reducing matrix is also used in testing. The number of features is reduced in all data samples to 300; thus, all support vector machine classifiers have the same size of 300 inputs. Fig. 6 shows the structure of the data-driven approach designed inside each relay to estimate if the fault is upstream or downstream relative to the relay location.

B. Hierarchical Fault Localization

Given how the local classifiers are designed, the location of the faults cannot be completely determined by any single classifier decision in many scenarios; therefore, it is required to combine the classification results of each relay to determine the fault location. In general, the 12 classifiers will provide an estimate on whether each fault is upstream or downstream from each relay. By combining the results of these classifiers, the location of the fault can be detected. For example, fault 11 should be classified as an upstream fault in all classifiers except those at relays 202 and 209, which should have this fault classified as a downstream fault. However, determining the fault location based on all 12 classifiers requires centralized operation, which can be challenging to perform in microgrids.

Therefore, we propose a decentralized operation method that can be used to determine the location of faults with communications only between neighboring relays. The proposed approach partitions the microgrid into multiple decision zones. Within each decision zone, a group of relays combine their classification results identify the fault location or the zone that needs to be checked. The decision zones are designed based on relays distance from the point of common coupling, are depicted in Fig. 7a, where decision zone #1 includes relays R201 and R202, and decision zone #2 includes relays R205, R206, and R207, and decision zone #3 includes R204 and R208, decision zone #4 includes R209, R210, and R211, decision zone #5 includes R219 only, and finally decision zone #6 includes R212 only.

TABLE II: Decision rules for Decision Zone #1.

CB201	CB202	Decision
0	0	Fault 2
0	1	DZ #4
1	0	DZ #2
1	1	Inadmissible

For each zone, upon obtaining the classification results, a decision is taken regarding the fault location. If the result identifies the fault the location, the algorithm should trigger the necessary tripping actions. If the result of the classification indicates that the fault is downstream in another decision zone, then this decision zone is triggered to perform its localization procedure. The process will continue until a fault is localized. Table II lists the decisions taken based on the relay classifiers in decision zone 1. If both relays 201 and 202 detect that the fault is upstream, then the fault is declared to be fault 2; however, if only relay 201 classifies the fault as downstream, then the decision zone 2 relays are informed that the fault localization task is passed down to them. Similarly, if only relay 202 classifies the fault as downstream, then the relays in decision zone 4 are informed. Notice that because the classifiers can have classification errors, there could be a scenario where both relays classify the fault as downstream. A maximum likelihood estimate can be used in this case, where we use a metric to quantify the confidence of each classifier based on their empirical accuracy. Note that if one utilize a soft classification approach, the result of the classifier can be used as an indication of the confidence of the model in the classification result.

Fig. 7b shows an example of the hierarchical localization method, where the relays in zone #1 (i.e., R201 and R202) see fault 8 as downstream from R201 and upstream from R202. So, the decisions in zone #2 are checked where only R207 detects the fault as downstream. Finally, in zone #4, both relays (i.e., R208 and R204) detect the fault as upstream, and hence, the location of the fault is determined to be fault 8. Note that for different locations, other zones in the hierarchy can be checked to determine the fault location.

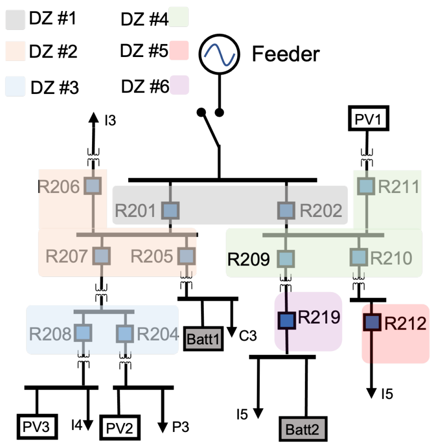
V. Numerical Results

In this section, we demonstrate the efficacy of the proposed fault localization approach. We used support vector machine classifiers at all relays. The implementation was done using the `sklearn` [14] using the polynomial kernel with the value of the regularizer chosen between 10 and 1000 using a k-fold cross-validation procedure. We first assess the performance of each classifier separately. Then, we present the accuracy of each decision zone.

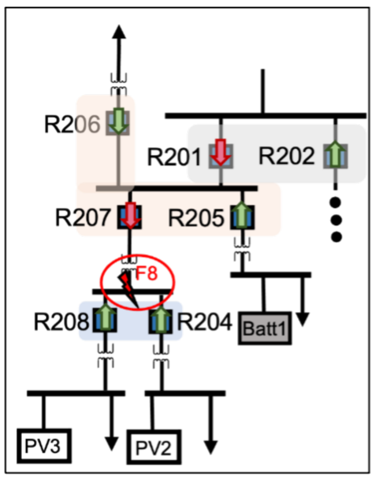
A. Individual Relay Performance

As discussed, the classifier at each relay is a binary classifier that aims to differentiate between upstream and downstream faults. For each classifier, we assess the performance using the accuracy, precision, and recall on a dataset of 300 fault testing scenarios recorded under varying operational conditions and fault characteristics. These measures are essential to fully assess the performance of each relay.

Table III presents the performance of each classifier. Note that in the data generation process, the probability of each



(a)



(b)

Fig. 7: The decision zones defined in the MG and an example of the hierarchical localization for Fault 8.

fault is $\frac{1}{13}$; thus, some classifiers have unbalanced data in the training, i.e., the ratio of positive samples is much less than negative samples. This emphasizes the importance of metrics such as recall and precision.

B. Hierarchical Decision Approach Performance

In the proposed approach, the identification of the fault location uses a hierarchical approach based on the decision zones defined in in Section IV-B. The hierarchical approach starts from a decision zone, then other decision zones are queried based on the results of the classifiers until the location of the fault is identified. We assess the performance of each decision zone to measures the accuracy of the overall approach. The localization of faults was carried out in 22.16 ± 9.9 milliseconds, which is less than two electric cycles.

The number of possible outcomes for each decision zone depends on the number of relays within the decision zone. For example, decision zones 2 and 4 encompass three relays, and

TABLE III: The performance of each classifier on the testing samples.

Classifier	Precision	Recall	Accuracy
CB201	81.00%	72.39%	82.06%
CB202	77.38%	80.25%	80.45%
CB204	72.22%	76.47%	94.97%
CB205	97.96%	100%	99.81%
CB206	100%	100%	100%
CB207	88.89%	66.67%	88.83%
CB208	59.09%	72.22%	92.18%
CB209	27.08%	54.17%	74.30%
CB210	81.81%	93.10%	95.53%
CB211	70.00%	70.00%	96.64%
CB212	100%	100%	100%
CB219	100%	60.53%	87.70%

TABLE IV: The accuracy of decisions in each decision zone.

Decision Zone	# of Relays	Accuracy
DZ #1	2	76.1%
DZ #2	3	82.6%
DZ #3	2	92.0%
DZ #4	3	73.7%
DZ #5	1	95.0%
DZ #6	1	99.3%

hence the number of possible outcomes is eight; therefore, an accurate decision for these two decision zones requires all three relays within the zone to correctly identify whether the fault is upstream or downstream from each relay. Table IV presents the results of the accuracy of each decision zone. For instance, in the hierarchical approach, if the fault is located at F4, in DZ#1 R201 and R202 will detect that the fault is upstream and downstream, respectively. Then, DZ#2 which includes R209, R210, and R211 will detected the fault upstream, upstream, and downstream, respectively. Thus, the fault can be localized as F4.

The results of the fault localization accuracy are comparable to those reported in the literature in [8], [5], [7], [6]; however, the proposed approach tackles two additional challenges. First, all the scenarios considered here are from islanded operation with 100% renewable penetration, which leads to a reduced magnitude of the fault currents and significant nonlinearity in the fault responses. Second, the scenarios used contain many high-impedance fault scenarios, that are challenging to detect and classify because they might not cause a noticeable change in the voltage and current levels.

VI. Discussions

The proposed approach uses an ML method to differentiate between upstream and downstream faults at each location to localize the faults in the MG. Designing a generic framework for data-driven-based protection in other MG systems requires several considerations. First, it is critical to develop a reliable and accurate model of the MG. This model should capture the dynamic behavior of the MG components, including inverters, energy storage systems, and loads. By simulating various fault scenarios on this model, it becomes possible to generate datasets for training the fault classifiers. The dataset generation process should also reflect the varying operations of the MG under consideration, such as variation in the load and renewables profiles as well as the fault characteristics.

Due to the radial (tree) structure of the considered MG, the definition of upstream and downstream directions at every relay can be simple. That is, the if the path between the fault and the point of common coupling (PCC) includes a specific relay, then the fault has to be considered downstream for the specific relay. Otherwise, if the path between the fault and the PCC does not include a relay, then the fault is considered upstream for the relay. Clearly, this approach is not applicable in networks with loops because the path between the fault and the PCC can be non-unique. This will require further developments to identify fault directions with respective to the relays in order to maintain the classification process as a simple binary classification. One option can be to use directions based on the current flows in pre-fault conditions, but this will require modifications of the labelling process in the data generation.

In the proposed approach, the communication requirements are limited to neighboring relays. This means that relays need to exchange information only with their adjacent relays to perform fault localization. This localized communication approach significantly reduces the amount of data transmission and ensures scalability for larger MG systems. As part of the generic framework, it is important to define the communication protocol and establish the necessary infrastructure to facilitate seamless information exchange between communicating relays. One advantage of the proposed approach is that the shared information between relays is only the classification results, i.e., 1-bit. This reduces the burden of establishing high-bandwidth communication link between relays which is required if measurements need to be shared, for example.

By considering the aforementioned points, the proposed data-driven approach can be adapted as a generic framework for designing data-driven-based protection in other MG systems. The framework provides a systematic and scalable approach for fault localization, leveraging local learning techniques and hierarchical decision-making processes. With the appropriate MG model, standardized differentiation of upstream and downstream faults, and addressing the communication requirements, the framework can be successfully deployed in various MG configurations, ensuring reliable and safe operation while accommodating different operational conditions and fault scenarios.

VII. Conclusions and Future Work

The protection of MGs with 100% penetration from IBRs remains a challenge with the adoption of MGs in future power networks. This paper discusses a machine learning-based approach to identify fault locations within MGs. A comprehensive dataset of fault scenarios was generated by varying the operational conditions as well as the fault characteristics. Next, a machine learning approach was developed that simplifies the task of each relay into a binary classification. Then, a hierarchical approach was presented to localize the faults using the decisions of the simple classifiers at the relays. While tackling challenging scenarios, the approach showed performance levels that match performance of approaches developed in the literature for much simpler distribution network protection tasks.

Future works include a localized decision making process that enable relays to make localization decision without requiring all relays in the network to classify fault locations. In addition, incorporating the proposed approach with a relay framework that also include a detection mechanism is required to establish data-driven relay architectures.

Acknowledgment

This work was authored by the National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at NREL. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

[1] U. Muenz, N. Xue, X. Wu, A. Mesanovic, L. Cockcroft, L. Dangelmaier, C. Heyde, Z. Dong, S. Chakraborty, and A. Ashok, "Operational security in 100% inverter-based power systems: Experiences from hawai 'i," *IEEE Electrification Magazine*, vol. 10, no. 3, pp. 10–19, 2022.

[2] B. Mahamedi and J. E. Fletcher, "Trends in the protection of inverter-based microgrids," *IET Generation, Transmission & Distribution*, vol. 13, no. 20, p. 4511, 2019.

[3] S. Venkata, M. J. Reno, W. Bower, S. Manson, J. Reilly, and G. W. Sey Jr, "Microgrid protection: Advancing the state of the art," *Sandia National Laboratories, SAND2019-3167, Albuquerque*, 2019.

[4] V. Telukunta, J. Pradhan, A. Agrawal, M. Singh, and S. G. Srivani, "Protection challenges under bulk penetration of renewable energy resources in power systems: A review," *CSEE journal of power and energy systems*, vol. 3, no. 4, pp. 365–379, 2017.

[5] J. Liang, T. Jing, H. Niu, and J. Wang, "Two-terminal fault location method of distribution network based on adaptive convolution neural network," *IEEE Access*, vol. 8, pp. 54 035–54 043, 2020.

[6] Y. D. Mamuya *et al.*, "Application of machine learning for fault classification and location in a radial distribution grid," *Applied Sciences*, vol. 10, no. 14, p. 4965, 2020.

[7] S. Paul, S. Grijalva, M. J. Aparicio, and M. J. Reno, "Knowledge-based fault diagnosis for a distribution system with high pv penetration," in *2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2022, pp. 1–5.

[8] M. J. Aparicio, M. J. Reno, P. Barba, and A. Bidram, "Multi-resolution analysis algorithm for fast fault classification and location in distribution systems," in *2021 IEEE 9th International Conference on Smart Energy Grid Engineering (SEGE)*. IEEE, 2021, pp. 134–140.

[9] M. Shafiullah, M. A. Abido, and T. Abdel-Fattah, "Distribution grids fault location employing ST based optimized machine learning approach," *Energies*, vol. 11, no. 9, 2018.

[10] A. Forouzesh, M. S. Golsorkhi, M. Savaghebi, and M. Baharizadeh, "Support vector machine based fault location identification in microgrids using interharmonic injection," *Energies*, vol. 14, no. 8, p. 2317, 2021.

[11] J. Wang, "Study of inverter control strategies on the stability of microgrids toward 100% renewable penetration," in *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2022, pp. 1–6.

[12] W. Jing, "Improved control strategy of grid-forming inverters for fault ride-through in a microgrid system," in *2022 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, 2022, pp. 1–7.

[13] S. Veda, “Eco-idea: Enhanced control and optimization of integrated distributed energy applications,” 1 2022. [Online]. Available: https://www.osti.gov/biblio/1843203

[14] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.