SAND2022-3650C

Knowledge-Based Fault Diagnosis for a Distribution System with High PV Penetration

Shuva Paul, Santiago Grijalva
Georgia Institute of Technology
Atlanta, Georgia, USA
Email: {spaul94, sgrijalva6}@gatech.edu

Miguel Jimenez Aparicio, Matthew J. Reno
Sandia National Laboratories
Albuquerque, New Mexico, USA
Email: {mjimene, mjreno}@sandia.gov

***Abstract*—Identifying the location of faults in a fast and accurate manner is critical for effective protection and restoration of distribution networks. This paper describes an efficient method for detecting, localizing, and classifying faults using advanced signal processing and machine learning tools. The method uses an Isolation Forest technique to detect the fault. Then Continuous Wavelet Transform (CWT) is used to analyze the traveling waves produced by the faults. The CWT coefficients of the current signals at the time of arrival of the traveling wave present unique characteristics for different fault types and locations. These CWT coefficients are fed into a Convolutional Neural Network (CNN) to train and classify fault events. The results show that for multiple fault scenarios and solar PV conditions, the method is able to determine the fault type and location with high accuracy.**

***Index Terms*—Fault detection, fault classification, fault location, deep learning, solar PV.**

I. Introduction

Fast and accurate fault location is critical for distribution system protection and recovery. Accurate fault diagnosis can help maintenance personnel ensure fast restoration, reduce economic damage, and enhance the reliability of the power system. Distribution systems with high penetration of solar PV exhibit a broader range of behavior given the variability of renewable sources. Thus fault detection and estimation methods must be robust enough to handle multiple conditions. Faults in distribution systems with high penetration of PV usually have less fault impedance, which makes detection more challenging. Moreover, higher PV penetration introduces a wide range of power quality issues due to nonlinear power electronics-based devices and loads [1]. In some cases, distributed resources in distribution systems have led to misoperations of conventional relays while detecting, localizing, and classifying faults [2], [3].

In this paper, an efficient method is presented for detecting, localizing and classifying faults in distribution systems with high penetration of solar PV, based on machine learning (ML) methods. The main contributions of this work are:

- Development of a robust database of fault signal traces for a distribution system with high solar PV penetration.
- Development of a machine learning (ML)-based fast fault detection method.
- Development of a deep learning (DL)-based framework for accurate localization and classification of faults.

II. Background

In this section, the main methods for distribution system fault location are briefly discussed.

A. Monitoring Electrical Measurements

Distance protection is one of the most widespread protection scheme in the transmission level. This technique utilizes the impedance calculated from currents and voltages measured so to estimate the location. It is not applicable to distribution systems since it is difficult to isolate the impedance that correspond to individual lines.

B. Traveling Waves and their Characteristics

This method utilizes the time differences between consecutive arrivals of the traveling waves caused by faults in the distribution network. Continuous Wavelet Transform (CWT) or Discrete Wavelet Transform (DWT) are usually required as part of the process to support fault location. [4], [5]. However, travelling waves for the distribution systems have a large number of reflections. Hence, it is difficult to identify individual reflections and their origin in distribution systems.

C. Machine Learning Techniques

Various machine learning techniques including Support Vector Machines (SVM), Artificial Neural Networks (ANN), Extreme Learning Machines (ELM), fuzzy logic, and stacked autoencoder, can be used for fault location and fault type classification [6]–[9]. The state-of-the-art studies use Convolutional Neural Networks (CNN) to extract features from the current signals and then use the extracted features to train the model for fault location [1], [10]. However, these techniques require detailed investigations to observe the impact of high penetration of Solar PVs into the distribution systems.

Although the aforementioned types of methods report good performance for fault detection and classification in traditional distribution networks, the system scenarios that include high penetration of solar PV require improved protection strategies.

III. Proposed Method

A. Problem Description

This research aims to develop a fault diagnosis framework for distribution systems with high PV penetration capable of detecting, locating, and classifying different types of faults. Five measurement devices are deployed in the system to record

the measurements needed to train the ML and DL models and to perform the detection, location, and classification tasks in their area of operation. These devices are not synchronized.

Three types of faults are considered: Single-Line-to-Ground (SLG) faults, Line-to-Line (LL), and Three-Phase (3P) faults. The faults are simulated to occur at several nodes in the distribution system. The sampling frequency used for measurement devices is 10 MHz. After the measurements are collected, the ML model employs the Isolation Forests (IF), a powerful technique to detect anomalies in the collected signals, which can indicate the existence of a fault. The signal is cropped ± 0.5 ms the fault is detected to occur. The 1 ms gives enough time to collect necessary information regarding the fault dynamics to support accurate fault diagnosis. Next, CWT matrices are calculated from this 1 ms signal since they contain all the information regarding the inherent frequencies of the traveling waves. Finally, the DL framework deploys the CNN to provide both the location of the faults and their types. Figure 1 presents the overall workflow, which is applied separately for fault location and classification.

B. Fault Simulation

Three types of faults, namely SLG, LL, and $3P$ faults are simulated in the IEEE 34-bus distribution test feeder. The $3P$ faults are the most severe types of faults in this system. These faults are simulated in PSCAD for different combinations of parameters as shown in Table I in order to generate the fault signals needed for training.

Table I: Parameter details for fault simulation

Parameter	Value
Type of faults	SLG, LL, $3P$
Resistance	0.01Ω, 0.1Ω, 1Ω, 1.5Ω, 2Ω, 5Ω, 10Ω
Incidence angle	1 ms, 2 ms
Irradiation	600 W/m^2, 1000 W/m^2
Temperature	$28°C$, $50°C$

The simulation of faults in PSCAD involves two steps: transient to steady-state and fault transient. In the first stage, the simulation is conducted for 2 seconds without the fault to ensure that the system arrives at a steady state. A snapshot is taken and saved after 2 seconds of simulation. In the second stage, 2 ms per fault case are simulated. The snapshot is considered as the starting point, and the fault occurs at 1 ms and 2 ms. Hence, the first half of the measurement data records the regular operation, and the last half records the incipient fault current transients. The recorded data from the simulation infers a group of signal measurements. The three-phase current measurements of the five measurement devices are extracted from this group of signal measurements for further processing.

C. Continuous Wavelet Transform

The Wavelet Transform (WT) is a powerful mathematical tool used in Digital Signal Processing (DSP). The convolution of the product between a signal $f(t)$ and the daughter wavelet is known as the Continuous Wavelet Transform (CWT) of that signal. The CWT can analyze the frequency components

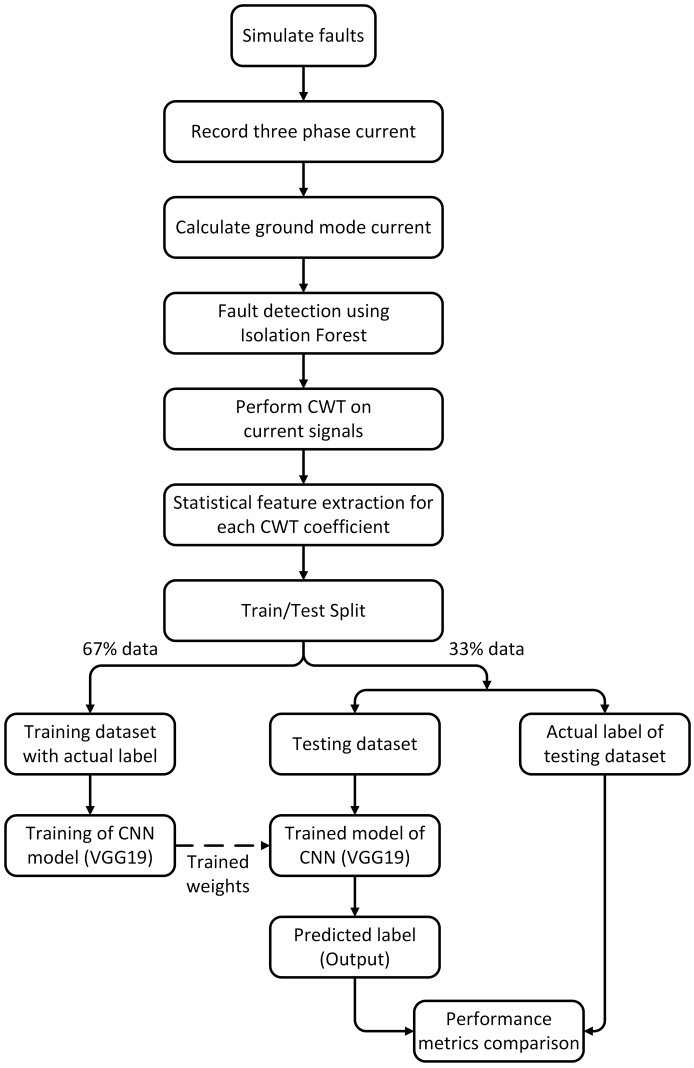

Figure 1: Overall workflow for the fault detection and location/classification.

of a signal for a specific time. The CWT provides high-frequency resolution for low scales, which allows an accurate representation of the TWs. The CWT of a signal $x(t)$ is defined as follows:

$$CWT_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-b}{a}\right) \mathrm{dt} \qquad (1)$$

where $\psi(t)$ is the mother wavelet which is scaled by a coefficient a, and translated using a translation coefficient b. The CWT of this signal results in a rectangular matrix where the number of rows represents the number of scales (each of the scales is inversely related to a frequency), and the number of columns represents the number of samples (frequency spectrum of the wave at each time instants) [11].

D. Isolation Forest (IF) for Fault Detection

Because the fault signatures of the measurements behave differently than the measurements recorded during the regular operation, the fault detection is inherently an anomaly

detection process. Typically, statistical, clustering, and nearest neighbor-based techniques are used for anomaly detection. An isolation-based approach, namely, Isolation Forest (IF) for fault detection is adopted, which was proposed in 2008 [12].

Isolation Forest isolates the observations by splitting the dataset. It develops a forest of random distinct itrees (isolation trees), where each of the trees has to decide based on the observation whether it is an anomaly or not. IF works in two stages. In the first stage, the IF model is trained and it constructs the forest of random itrees. In the second stage (scoring phase), the IF assigns an anomaly score to all the observations in the dataset. The anomaly score is computed using the following formula:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{2}$$

where,

$$E(h(x)) = \frac{\sum_{i=1}^{t} h_i(x)}{t} \tag{3}$$

Here, x, $h(x)$, and $E(h(x))$ represent the observation, path lengths, and average path length of x over t itrees, respectively. $c(n)$ stands for the average path length of the unsuccessful search in the Binary Search Tree (BST).

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \tag{4}$$

where, $H(i) = \ln(i) + \gamma$. Here, γ is the Euler's constant. IF determines whether an observation x is an anomaly or not based on the following condition:

$$x = \begin{cases} \text{Anomaly}, & \text{if } s(x,n) \sim 1 \\ \text{Not anomaly}, & \text{if } s(x,n) < 0.5 \end{cases} \tag{5}$$

If the $s(x,n)$ is close to one, it belongs to the anomaly group. On the other hand, if $s(x,n)$ is less than 0.5, it belongs to the group of normal data points [13]. The ranges of the anomaly score can be scaled based on the design of the algorithm.

E. CNNs for Fault Diagnosis

CNNs are well known for their superior performance in image processing and classification. The CNNs are composed of successive convolutional layers, maximum pooling layers, average pooling layers, dense layers, etc. There are several architectures of CNNs, namely RESNET50, VGG16, VGG19, NASNET, etc. The VGG19 architecture of CNN is adopted, which is a deep CNN used for image classification. It has improved training time, and a higher number of FLOPs (floating point operations per second) compared to other architectures. In this research, the CWT matrices are treated as images. The VGG19 is a variant of the VGG model, composed of 19 layers including 16 convolutional layers, 3 fully connected layer, 5 MaxPool layers, and 1 SoftMax layer.

Convolutional operations are executed in the convolutional layers, which extract the information from the images while maintaining the spatial relationship of pixels. The convolutional operation is conducted between the input image and a filter. The resulting matrix from the convolution is called a "feature map", which consists of successive iterations of the convolution across the whole input image. The model learns the coefficients of these filters during the training process. As mentioned earlier, the measurements are recorded for 0.5 ms before and after the ground mode arrival time (total 1 ms), the CWT matrices of that 1 ms are saved for the training process of the VGG19. The VGG19 model is being trained with the measurements from each measurement device for fault location/classification. Given that there are 5 measurement devices located across the system, 5 models for fault location and 5 models for fault classification are trained.

IV. Numerical Results and Discussion

In this section, the case description and results from the fault detection method and fault location and classification based on CNN are discussed. All the measurement devices are trained under all the fault scenarios.

A. Case Description

The IEEE 34-bus case is adopted for fault simulations. The simulations are performed using PSCAD using the python *Automation Library*. Figure 2 shows the IEEE 34-bus case with the fault locations, the measurement devices, and 3 solar PV installations of $200kW$ each. Measurement devices s, 1, 2, 3, and 4 are located at nodes 800, 850, 828, 832, and 860, respectively.

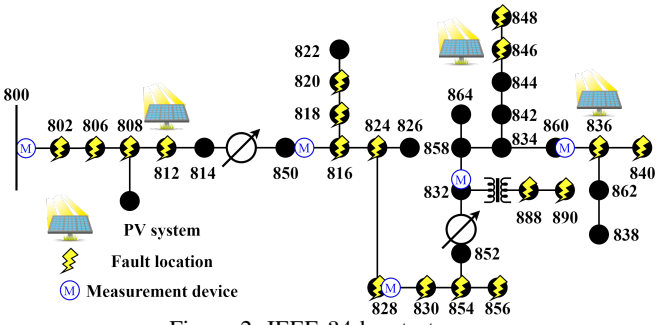

Figure 2: IEEE 34-bus test case.

In order to have a robust training model, a combinatorial dataset containing comprehensive information about the incipient fault current transients was generated. A total of $13,440$ fault cases (including measurements recorded by the 5 measurement devices) were simulated. For each case, the corresponding CWT matrices were obtained.

B. Simulation Results

In order to evaluate the classification performance of the CNN architecture (VGG19), we use precision, recall, F_1-score, and accuracy.

1) Fault Detection: The Isolation Forest model is trained for fault detection with normal data (no fault measurements). The three-phase measurements are summarized into the ground mode. Then, the ground mode current is used for detection. It is easier to represent the ground mode current compared to the three-phase fault. Figure 3 shows the ground mode current of node 806 under a three-phase fault. The measurement is recorded by measurement device 2.

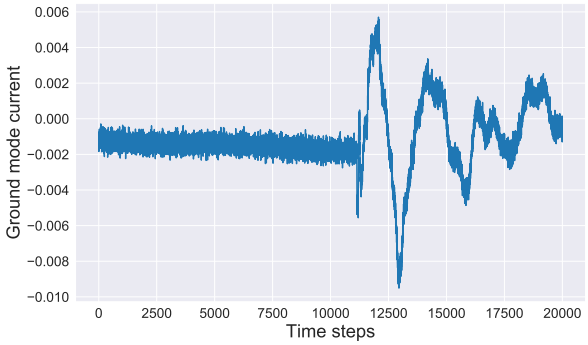

Figure 3: Ground mode current of node 806 with noise (timestep 0.1 μs).

Once training is completed, all the data (including normal and faulty observations) are fed to the trained model for testing. Figure 4 illustrates the anomaly score region after testing. The gray-colored region in Figure 4 represents the outlier region. Any value that lies in the gray-colored region is considered as fault measurement. Similarly, to detect events/faults for the other nodes, the current measurements need to be used to train the IF, calculate anomaly score, and detect the fault based on the anomaly score.

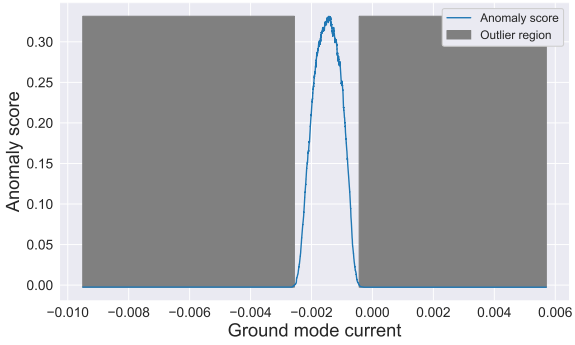

Figure 4: Anomaly region for the ground mode current of node 806

2) Fault Location: After the fault is detected, its location is determined. For this purpose, the CWT matrices are fed into the VGG19 model. The results of fault location classification for measurement device 4 are shown in Figure 5. The diagonal elements of the confusion matrix show the successful classifications, while the off-diagonal elements represent unsuccessful classification. For the majority of the nodes, the predictions were correctly classified. However, the model showed poor performance for nodes 888 and 890.

Table II: Classification report for fault location of measurement device 4.

	Precision	Recall	F_1-score
Accuracy			0.91
Average	0.91	0.91	0.91

Table II presents the overall classification accuracy, and averages of precision, recall, and F_1-score. The overall accuracy of the model is 91%. The performance metrics for the

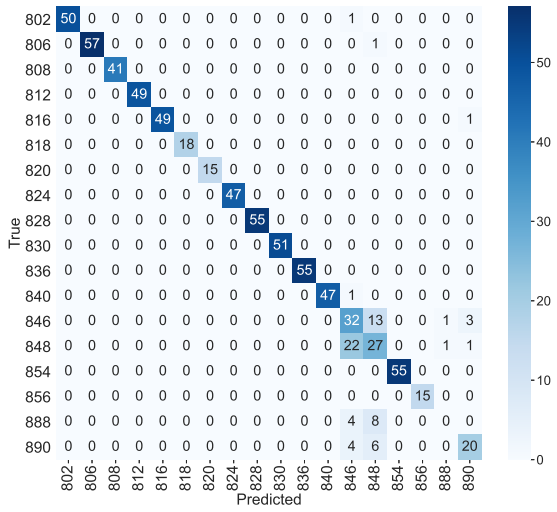

Figure 5: Confusion matrix for fault location at measurement device 4.

Table III: Fault location metrics for all the measurement devices using the VGG19 model

	s	1	2	3	4
Precision	76%	89%	90%	89%	91%
Recall	75%	88%	89%	89%	91%
F_1-score	75%	88%	89%	89%	91%
Accuracy	75%	88%	89%	89%	91%

other measurement devices are shown in Table III. The overall accuracy of measurement devices s, 1, 2, 3, and 4 was 75%, 88%, 89%, 89%, and 91%, respectively.

3) Fault Classification: Once the fault is located, the fault is classified using new VGG19 CNN models.

Table IV: Detailed classification report for type classification of measurement device s.

	Precision	Recall	F_1-score
SLG	0.97	0.93	0.95
3P	0.97	0.94	0.96
LL	0.94	0.99	0.96

Table IV presents the detailed classification report in terms of the performance matrices (i.e., precision, recall, and F_1-score) for individual fault types.

Table V: Classification report for type classification of measurement device s.

	Precision	Recall	F_1-score
Accuracy			0.96
Average	0.96	0.96	0.96

Table V shows the average precision, recall and F_1-score of the fault types. From Table V, it is found that the overall accuracy of the fault classification in measurement device s is 96%.

Table VI: Fault type classification

	s	1	2	3	4
Precision	96%	99%	99%	99%	99%
Recall	96%	99%	99%	99%	99%
F_1-score	96%	99%	99%	99%	99%
Accuracy	96%	99%	99%	99%	99%

Table VI presents the fault classification performances for the other measurement devices. All the measurement devices, except measurement device s, achieved an accuracy of 99% while classifying the faults.

V. Discussion

In this section, we compare the results from this paper with our previous work [14] and other state-of-the-art research on fault location and classification.

Table VII: Results comparison for fault location

Measurement device	s	1	2	3	4
Proposed method	75%	88%	90%	89%	91%
[14]	40.51%	78.44%	64.65%	85.34%	93.10%

Table VII and VIII compare the outcome of this paper with our previous work in [14].

Table VIII: Results comparison for type classification

Measurement device	s	1	2	3	4
Proposed method	96%	99%	99%	99%	99%
[14]	93.97%	94.83%	93.10%	93.10%	93.11%

In both cases, the accuracy of fault location and classification is improved considering a large number of fault scenarios (compared to [14]). For fault location, the accuracies have increased by 34.49%, 9.56%, 25.35%, and 3.66% for measurement device s, 1, 2, and 3, respectively. For fault classification, the performance accuracies have increased by 2.03%, 4.17%, 5.90%, 5.90%, and 5.89% for measurement device s, 1, 2, 3, and 4, respectively.

Apart from the previous work, the results in this paper are comparable to [1], [9], [10], [15] in terms of the complexity of the system in consideration (size of the system, number of PVs, number of faults scenarios, etc.) and performance of the classification models.

VI. Conclusions and Future Work

Fault detection, location, and type classification are essential in distribution system protection and service restoration. This paper applied fault detection, location, and classification techniques on a distribution network with high PV penetration. The knowledge-based technique was implemented using ML and DL algorithms. CWT was performed together with statistical measures to extract meaningful information from the measured current signal. Those features were used to train the DL algorithm (VGG19). The performances of the ML models are evaluated in terms of precision, recall, F_1-score, and accuracy, and found satisfactory compared to our previous work (86.6% and 98.4% average accuracy for fault location and classification, respectively) and current state-of-the-art works.

Acknowledgements

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA000352.

References

[1] P. Rai, N. D. Londhe, and R. Raj, "Fault classification in power system distribution network integrated with distributed generators using cnn," *Electric Power Systems Research*, vol. 192, p. 106914, 2021.

[2] H. A. Tokel, R. A. Halaseh, G. Alirezaei, and R. Mathar, "A new approach for machine learning-based fault detection and classification in power systems," in *2018 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, 2018.

[3] T. S. Abdelgayed, W. G. Morsi, and T. S. Sidhu, "A new approach for fault classification in microgrids using optimal wavelet functions matching pursuit," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4838–4846, 2018.

[4] F. V. Lopes, K. M. Dantas, K. M. Silva, and F. B. Costa, "Accurate two-terminal transmission line fault location using traveling waves," *IEEE Transactions on Power Delivery*, vol. 33, no. 2, pp. 873–880, 2018.

[5] X. Chen, X. Yin, and S. Deng, "A novel method for slg fault location in power distribution system using time lag of travelling wave components," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 12, no. 1, pp. 45–54, 2017.

[6] 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.

[7] 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, 2021.

[8] O. W. Chuan, N. F. Ab Aziz, Z. M. Yasin, N. A. Salim, and N. A. Wahab, "Fault classification in smart distribution network using support vector machine," *Indonesian Journal of Electrical Engineering and Computer Science*, 2020.

[9] G. Luo, Y. Tan, M. Li, M. Cheng, Y. Liu, and J. He, "Stacked auto-encoder-based fault location in distribution network," *IEEE Access*, vol. 8, pp. 28043–28053, 2020.

[10] 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. 54035–54043, 2020.

[11] H. Jia, "An improved traveling-wave-based fault location method with compensating the dispersion effect of traveling wave in wavelet domain," *Mathematical Problems in Engineering*, vol. 2017, 2017.

[12] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422, 2008.

[13] M. U. Togbe, M. Barry, A. Boly, Y. Chabchoub, R. Chiky, J. Montiel, and V.-T. Tran, "Anomaly detection for data streams based on isolation forest using scikit-multiflow," in *Computational Science and Its Applications – ICCSA 2020*, (Cham), pp. 15–30, Springer International Publishing, 2020.

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

[15] Y. D. Mamuya, Y.-D. Lee, J.-W. Shen, M. Shafiullah, and C.-C. Kuo, "Application of machine learning for fault classification and location in a radial distribution grid," *Applied Sciences*, vol. 10, no. 14, 2020.