A Rule-based Intelligent Method for Identifying Single-Line-to-Ground Fault Location in Distribution Systems

G. Y. Li, P. J. Hsu, and G. W. Chang

Dept. Electrical Engineering

National Chung Cheng University

Chaiyi, Taiwan

*wchang@ee.ccu.edu.tw

***Abstract*—Accurate fault location in distribution systems remains a considerable challenge due to factors such as multi-terminal line configurations, load unbalance, neutral grounding conditions, and uncertainties in load modeling. This paper proposes a simulation-based fault-location method that considers single-line-to-ground faults occurring at various positions along distribution feeders. When a fault occurs, advanced metering devices installed at substations are capable of capturing the transient voltage waveforms induced by the fault, which include fundamental-frequency components, high-frequency transients, and harmonics. The method extracts the high-frequency transient components using FFT, identifies fault types using wavelet transform, and further integrates rule-based processing with machine-learning classification algorithms to determine the faulted zone and estimate the fault location.**

***Index Terms*—Deep learning, ensemble learning, fault classification, fault location, spectral analysis.**

I. Introduction

In smart grid distribution systems, the rapidity and accuracy of fault location techniques are essential for effective protection schemes. Accurate fault location helps minimize downtime, avoid asset damage, and expedite system restoration. In the literature, various fault-location algorithms have been proposed. In [1] and [2], the authors proposed traveling-wave-based methods, where the relative distance between the peak points of high-frequency traveling waves is calculated and compared with known feeder distances to determine the faulted area and potential fault location. In [3], the authors introduced a method for fault location in multipoint distribution networks by estimating the wave arrival time at measurement points to determine the fault distance. In [4], wavelet transform was used to process transient waves propagating through transmission lines, and GPS-based synchronization was applied.

For voltage-sag-based methods, the authors in [5] used voltage and current measurements from substations for fault location. An iterative process based on power-flow calculations updates the voltages, currents, and load of each line segment. Reference [6] described a state-estimation-based method that assumes the short-circuit current equals the measured substation current; voltages and currents from the main feeders and distributed energy resources are used for the state estimation. However, the obtained estimation errors remain relatively large. In [7] and [8], a generalized impedance-based fault location technique for distribution networks was developed. Such methods require only simple calculations and feature low implementation costs, but they suffer from multiple estimation issues. The authors in [9] proposed intelligent fault location using intelligent electronic devices (IEDs) and remote terminal units (RTUs) to obtain real-time data and system status, although the cost of IED deployment in distribution systems remains a concern.

With strong pattern-recognition capabilities, machine-learning (ML) techniques have become promising alternatives to conventional fault-location methods [10]. The authors in [11] proposed a support vector machine (SVM)-based approach to partition the distribution system into several regions and identify the faulted section. In [12], SVM was used to identify the faulted feeder in intelligent distribution networks. A hybrid approach was developed in [13], where an artificial neural network (ANN) was trained using voltage signals decomposed at different scales as learning features. The authors of [14] proposed a hybrid ML approach using wavelet scattering networks for SLG fault location with single-ended measurements of a small system. Classification and regression problems are solved using SVMs, targeting faulty laterals and exact fault distances.

This study presented in this paper develops a hybrid high-frequency and ML-based single-line-to-ground (SLG) fault location method. The measured three-phase transient voltages (V_a, V_b, V_c) at the substation are transformed into V_α, V_β, and V_0 components using Clarke transformation. The Fast Fourier Transform (FFT) is applied to extract zero-mode frequency-domain features, and low-frequency components are filtered using a high-pass filter. Due to the distribution-dependent resonance characteristics along the feeder, each location has distinct local resonant frequencies, which are extracted as feature inputs for ML algorithms. The rule-based techniques and the ML algorithms are then applied to detect the faulted zone and the fault location.

The remainder of this paper is organized as follows: Section II describes the methodological framework of the study, which is based on rule-based zone partitioning with six ML algorithms and ensemble learning techniques. Case studies are then discussed in Section III. Finally, the conclusions are provided in Section IV.

This work was supported in part by the Ministry of Science and Technology of Taiwan under Grant MOST 108-2221-E-194-031-MY2.

979-8-3315-8138-1

II. METHODOLOGICAL FRAMEWORK OF THE STUDY

A. *Applying Intelligent Methods for Fault Location in Distribution Feeders*

When short-circuit faults occur in a distribution system, the magnitudes of current and voltage abruptly change at multiple locations along the feeder, resulting in transient phenomena. If a fault occurs at any location in a distribution feeder, it will produce the transient that propagates to the measurement point at the substation. The frequency components measured at the substation include the fundamental, harmonics, and high-frequency ones, where the high-frequency components can be used to identify the fault zone in the study. The waveform processing techniques are applied to classify fault types, including using Clarke transform to derive the zero-mode voltage for SLG fault detection and wavelet transform for current waveforms to identify the fault type. Fig. 1 illustrates the frequency spectrum when single-line-to-ground faults occur at different busses in a distribution system after filtering the low-frequency components, where the dominant resonance points are at 5423 Hz, 5355 Hz, and 4230 Hz, respectively.

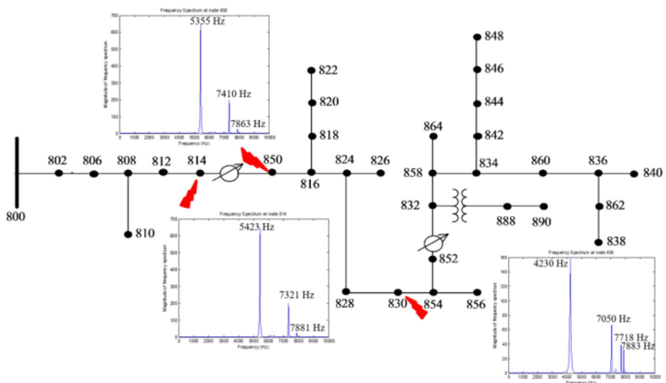

Fig. 1. Resonant high-frequency spectra of fault-induced zero-mode voltage measured at substation bus #800.

During the fault interval, the three-phase voltages and currents measured at the substation and along the feeder are preprocessed and transformed into the frequency domain using the Fast Fourier Transform (FFT). After feature extraction, fault zones are defined through a combination of expert rules and unsupervised learning techniques. The extracted spectral features serve as the input to the ML model, while the output corresponds to the predicted fault zone. Once the fault zone is identified, the fault distance is estimated from the origin of that zone. The flowchart of the proposed fault-location method is illustrated in Fig. 2.

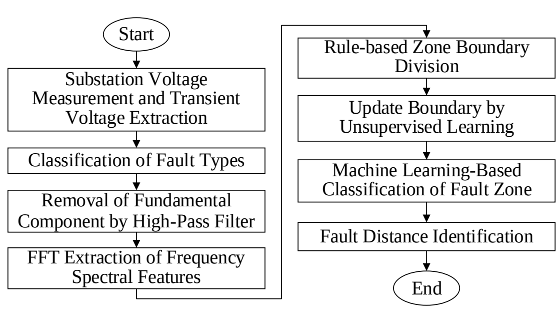

Fig. 2. Flowchart of the proposed fault-location method.

B. *Ground Fault Type and Fault Phase Detection*

In a three-phase four-wire distribution system, the voltages measured at the substation include V_a, V_b, and V_c. The Clarke transformation is applied to convert the measurements into the orthogonal components V_α, V_β, and V_0. By analyzing the spectral characteristics of the aerial mode (α, β) and the zero mode (0), the fault type can be determined, such as single-line-to-ground (SLG), double-line-to-ground (DLG), line-to-line (LL), and three-phase faults. A summary of these mode-to-fault relationships is provided in Table 1.

TABLE 1. SELECTION OF MODAL SIGNALS CORRESPONDING TO FAULT TYPES

Fault Type	Modal Signal
Single-Line-to-Ground	V_0
Double Line	V_α, V_β
Double-Line-to-Ground	V_0
Three-Phase	V_α, V_β

The measured fault-induced voltage at the substation is used for classifying the ground fault. The sum of the measured three-phase voltages, after performing the power-invariant Clarke transform to obtain the zero-mode voltage of (1) can be used to determine whether it is a ground fault.

$$v_0 = \frac{1}{\sqrt{3}}(v_a + v_b + v_c) \tag{1}$$

In the proposed approach, the fault type is determined by comparing healthy and faulted phase currents. A wavelet transform and statistical feature-based faulty-phase detection method, similar to [15], is then adopted.

C. *FIR High-Pass Filtering*

In the spectrum of voltage signals, the fundamental component typically exhibits the highest magnitude. Therefore, it must be removed to properly isolate the high-frequency components required for accurate fault-location analysis. In addition, harmonic frequencies may induce harmonic resonance when they coincide with the natural resonance frequencies of the power network. In most cases, transient resonance components appear above 3 kHz. For a 60-Hz system, frequencies below 3000 Hz should be attenuated to focus on high-frequency resonance behavior driven by feeder configuration and fault location. The input for this step is the fault-induced zero-mode voltage waveform.

D. *Spectrum Extraction Using FFT*

The FFT is a widely adopted tool in power system analysis for both steady-state and transient signal evaluation. It efficiently converts time-domain voltage measurements into frequency-domain components and enables accurate estimation of the magnitude and phase associated with each characteristic frequency. By performing frequency decomposition, the FFT transforms the measured waveform into a spectrum in which the magnitude corresponds to the squared amplitude contribution of each frequency component.

In the proposed method, fault-location information is not explicitly observable in the fault-induced zero-mode voltage waveform. Instead, the signal is transformed into the frequency domain to identify resonance characteristics that are

979-8-3315-8138-1

closely related to the fault location. A sampling frequency of 20 kHz is selected to ensure that the extracted spectrum contains sufficient resolution to capture resonance peaks distributed across high-frequency ranges. These resonance components, influenced by feeder configuration, grounding type, and fault distance, serve as key discriminative features for accurate ML-based fault localization.

E. Rule-Based Zone Boundary Division

In order to identify the faulted zone boundary when the SLG fault occurs, the distribution feeder is divided into several smaller fault zones. This zone segmentation allows the classifier to distinguish resonance characteristics within a reduced search space, thereby enhancing prediction performance. The boundary determination follows predefined engineering rules, such as line section length, impedance variations, and branch connection points. A simple distribution feeder configuration used as an example for zone boundary division is shown in Fig. 3 for illustration.

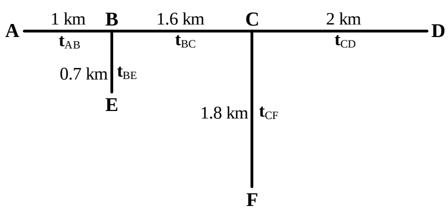

Fig. 3. Typical branch configuration of a distribution feeder.

The zone boundary segmentation is determined based on the voltage transient propagation time along the distribution line. For example, the propagation time of a specific line section can be expressed by (2).

$$\begin{aligned} &\text{From Bus A to bus B}: t_{AB} \\ &\text{Bus A to bus C}: t_{AB} + t_{BC} \\ &\text{Bus A to bus E}: t_{AB} + t_{BE} \\ &\text{Bus A to bus F}: t_{AB} + t_{BC} + t_{CF} \end{aligned} \tag{2}$$

In Fig. 4, it can be observed that the travel time t_{AB} is uniquely distinguishable. The propagation times of segments BE and BC overlap within the first 0.7 km; however, the remaining 0.9 km of the BC segment exhibits a distinct propagation characteristic. Similarly, for segments CD and CF, the propagating times t_{CF} and t_{CD} overlap over the first 1.8 km, whereas the remaining 0.2 km of segment CD provides a unique time signature. Therefore, the distribution feeder shown in Fig. 2 can be partitioned into three fault zones.

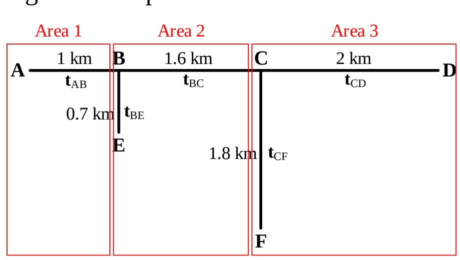

Fig. 4. Boundary zone (i.e. area) division of branches in a distribution feeder.

Although the propagation-time differences along distribution lines are extremely small and therefore insufficient for reliable zone discrimination in the time domain, the frequency-domain representation provides much stronger distinguishability. Even a slight variation in propagation distance can lead to a significant shift in resonance frequencies due to their inverse relationship. Consequently, fault-location features become more separable in the spectral domain.

F. Unsupervised Boundary Refinement Using K-means Clustering

Although the initial rule-based zone boundaries provide a reasonable segmentation of the feeder, they cannot fully separate zones with similar spectral features. Therefore, K-means clustering is employed to refine the boundary regions by identifying clusters of transient-frequency features that exhibit high similarity. The objective is to determine the most representative cluster centroids through iterative computation, minimizing the sum of squared distances between each data point and its assigned centroid. This process yields an approximately optimal clustering result and significantly improves zone discrimination accuracy. The K-means clustering procedure is illustrated in Fig. 5.

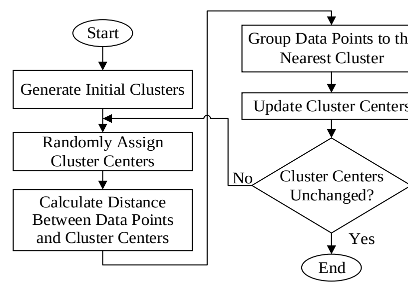

Fig. 5. Flowchart of the K-means clustering algorithm.

Since each fault zone consists of different series and parallel combinations of R, L, and C elements, faults occurring within the same zone tend to exhibit similar frequency-domain resonance signatures. By integrating rule-based segmentation with the K-means clustering algorithm, the IEEE 34-bus distribution network can be divided into nine zone clusters with different geometric regions, as shown in Fig. 6. The corresponding nodes included in each zone are summarized in Table 2.

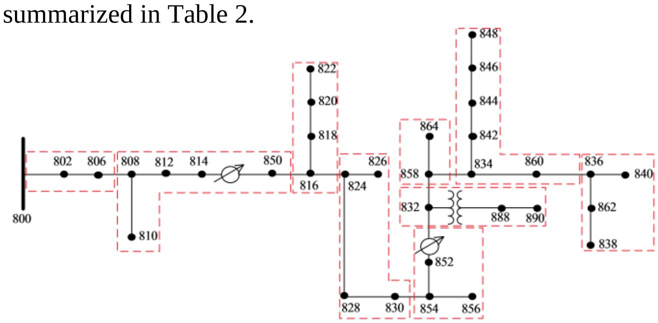

Fig. 6. Illustration of zone boundary division for IEEE 34-bus system.

TABLE 2. ZONE DIVISION OF THE IEEE 34-BUS DISTRIBUTION SYSTEM

Zone	Node
1	802, 806, 808
2	810, 808, 812, 814, 850, 816
3	824, 816, 818, 820, 822
4	826, 824, 828, 830, 854
5	856, 854, 852, 832
6	890, 888, 832, 858
7	864, 858, 834
8	836, 860, 834, 842, 844, 846, 848
9	840, 836, 862, 838

G. Machine Learning-Based Fault-Zone Classification

The input to the six machine-learning classifiers consists of normalized spectral feature vectors derived from the fault-induced zero-mode voltage. These input vectors are used to train the models including Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Decision Tree (CART), Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). Since an individual classifier may not achieve satisfactory performance across all fault cases, an ensemble learning strategy is adopted to obtain a more reliable and accurate fault-zone classification result.

Ensemble learning techniques can generally be categorized into bagging, boosting, and stacking. Bagging is considered the simplest approach in terms of training-data manipulation, while boosting is widely applied to improve weak learners through sequential refinement [16]. To prevent uncertain or ambiguous zone prediction, a stacking-based ensemble framework is implemented in this study. In the stacking architecture, the classification results from all base learners are fed into a final-stage meta-classifier, which is implemented using XGBoost to generate the optimal zone classification output. The stacking ensemble architecture is shown in Fig. 7.

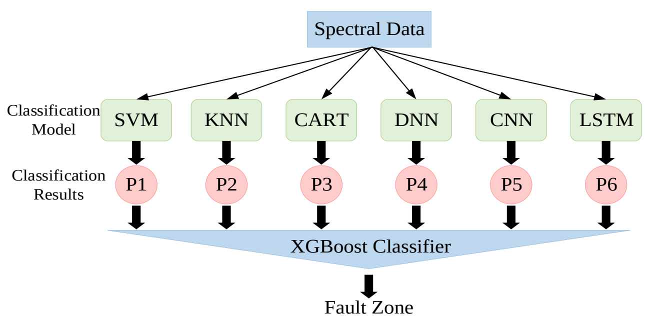

Fig. 7. Stacking ensemble architecture for fault-zone classification.

H. Fault-Distance Estimation

In a distribution system, the main feeder usually branches into multiple laterals with different line lengths. Using the IEEE 34-bus system in Fig. 5 as an example, three-phase voltage signals are measured only at the substation. Figs. 8 and 9 illustrate the frequency spectra of the zero-sequence voltage component V_0 generated by single-line-to-ground (SLG) faults occurring on the main feeder at distances of 1300 m and 11,000 m from the substation, respectively.

The frequency components in the spectrum are excited by localized transient high-frequency resonances. As shown in Figs. 8 and 9, when the fault location is farther away from the substation, the dominant resonance frequencies shift toward lower values. This frequency-distance dependency enables fault-distance estimation through regression analysis after the faulted zone is identified. In the proposed method, a stacking-based ensemble learning framework similar to Fig. 7 is employed to accurately determine the fault distance by mapping the extracted spectral features to their corresponding actual locations along the feeder.

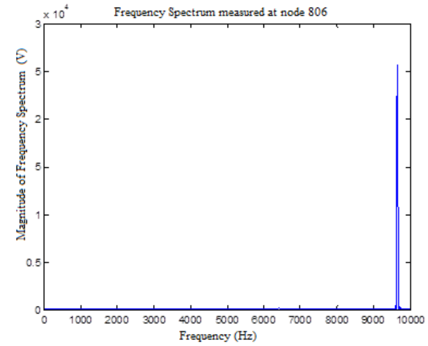

Fig. 8. Spectrum of the IEEE 34-bus system for a SLG fault located 1300m away from the substation (dominant frequency is near 9800 Hz).

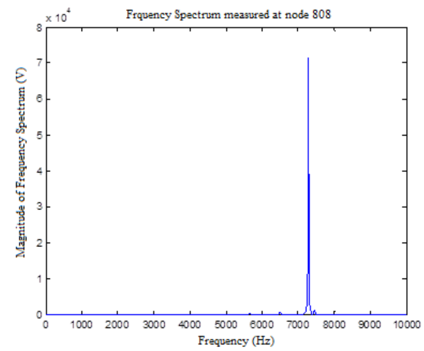

Fig. 9. Spectrum of the IEEE 34-bus system for a SLG fault located 11000m away from the substation (dominant frequency is near 7300 Hz).

III. Test Results

In this study, EMTP/ATP is used to model the IEEE 34-bus benchmark distribution feeder shown in Fig. 5. Transient waveforms under SLG fault conditions are simulated, and the recorded signals are processed in MATLAB to extract frequency-domain features. In the ATP simulation-produced fault events, the total dataset of fault samples is randomly separated into two parts: 80% of fault events are set as training data, and 20% of events are set as testing data. The total number of training samples is 624, and the number of test samples is 157. Subsequently, six different machine-learning models are implemented in Keras for fault-zone classification and fault-distance estimation to achieve fast fault localization [17]. In addition, the stacking-based ensemble learning strategy is incorporated to improve the location accuracy.

A. Classification of Fault Zones

The proposed method utilizes spectral features as input to six ML classifiers. The extracted spectra are normalized, and the dataset is randomly split into 80% for training and 20% for testing. Among the base learners, the CNN achieves the highest standalone performance with an accuracy of 96.1%, although its computational cost is notably higher. Traditional ML models also achieve above 90% accuracy with significantly shorter inference time. As shown in Fig. 10, the stacking ensemble model combines the strengths of both deep learning and traditional classifiers, improving the fault-zone classification accuracy to 98.7%, thereby enabling highly reliable fault-zone identification.

979-8-3315-8138-1

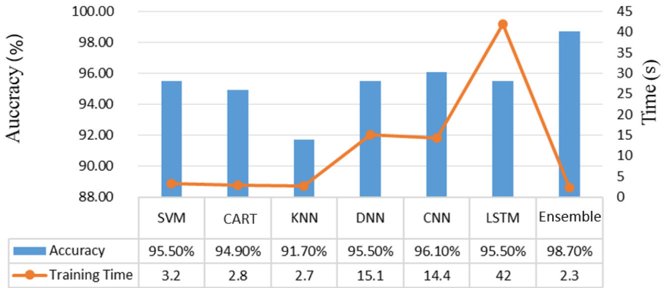

Fig. 10. Faulted zone classification accuracy for IEEE 34-bus system.

B. *Fault Distance Identification*

Once the fault zone is correctly identified, the fault distance is measured starting from the origin of the predicted zone. Therefore, spectral features generated at different fault locations (in meters) within the selected zone are used as inputs to the six ML models again to estimate the fault distance. The performance is evaluated using the absolute error and the relative error percentage, as given in (3).

$$\text{Error (\%)} = \frac{|\text{Actual Value-Estimated Value}|}{\text{Actual Value}} \times 100\% \tag{3}$$

For the IEEE 34-bus system, a single-line-to-ground fault is applied within Zone 3 (between Buses 816 and 822), and the stacking-based ensemble learning model is used to estimate the fault distance. As summarized in Table 3, all fault distances are measured as the total path length from Node 800 of Fig. 6 to the fault location. It can be observed that the proposed ensemble approach achieves an average absolute error of 6.303 m, corresponding to a relative error percentage of only 0.017%.

TABLE 3. FAULT-DISTANCE ESTIMATION ERROR FOR IEEE 34-BUS SYSTEM

Zone	Actual Fault Distance (m)	Estimated Fault Distance (m)	Absolute Error (m)	Relative Error (%)
3	339.570	347.429	7.859	0.025
	1083.585	1088.824	5.239	0.016
	2094.816	2085.909	8.907	0.026
	2765.477	2754.998	10.479	0.03
	4326.860	4319.001	7.859	0.022
	5841.087	5840.039	1.048	0.003
	6752.767	6753.029	0.262	0.001
	8618.166	8629.484	11.318	0.028
	10165.030	10164.402	0.628	0.002
	13233.607	13243.039	9.432	0.021
	Total Average Error		**6.303**	**0.017**
	Training Time (sec)		**2.427**	

IV. CONCLUSION

In this work, an intelligent fault-location method for distribution systems is proposed. When a SLG fault occurs, the transient zero-mode voltage waveform measured only at the substation is analyzed using mathematical transformations and FFT to extract resonance frequencies. Multiple resonance peaks form spectral feature vectors, which are then processed through zone-boundary division and ML classification. The stacking ensemble learning approach is further adopted to enhance the classification performance. Simulation results demonstrate that the proposed method achieves up to 98.7% fault-zone classification accuracy in the test system. After identifying the fault zone, the fault distance can be accurately estimated. These simulation results confirm that the proposed approach can be effectively applied to more complex distribution systems.

REFERENCES

[1] H. Hizam, P. A. Crossley, P. F. Gale, and G. Bryson, “Fault Section Identification and Location on A Distribution Feeder Using Travelling Waves,” in *Proc. 2002 IEEE Power Eng. Soc. Summer Meeting*, vol. 3, pp. 1107-1112, 2002.

[2] C. Trindade, L. Fernanda, W. Freitas, and J. C. Vieira, “Fault Location in Distribution Systems Based on Smart Feeder Meters,” *IEEE Trans. Power Del.*, vol. 29, no. 1, pp. 251-260, 2014.

[3] R. Kumar and D. Saxena, “A Traveling Wave Based Method for Fault Location in Multi-Lateral Distribution Network with DG,” in *Proc. 2018 IEEE ISGT Asia*, pp. 7-12, 2018.

[4] A. Tabatabaei, M. R. Mosavi, and A. Rahmati, “Fault Location Techniques in Power System Based on Traveling Wave Using Wavelet Analysis and GPS Timing,” *Electrical Review*, vol. 88, no. 6, pp. 347-350, 2012.

[5] R. Pereira and A. Fernandes, “Improved fault location on distribution feeders based on matching during-fault voltage sags,” *IEEE Trans. Power Del.*, vol. 24, no. 2, pp. 852-862, 2009.

[6] J. D. C. Guillen, “Fault Location Identification in Smart Distribution Networks with Distributed Generation”, Master’s Thesis, Florida State University, 2015.

[7] G. D. Ferreira, D. D. S. Gazzana, A. S. Bretas, A. H. Ferreira, A. L. Bettiol, and A. Carniato, “Impedance-Based Fault Location for Overhead and Underground Distribution Systems,” in *Proc. 2012 NAPS*, pp. 1-6, 2012.

[8] K. Sun, Q. Chen, Z. Gao, D. Liu, and G. Zhang, “Generalized Impedance-Based Fault Distance Calculation Method for Power Distribution Systems,” in *Proc. 2014 IEEE APPEEC*, pp. 1-4, 2014.

[9] C. Trindade, L. Fernanda, W. Freitas, and J. C. Vieira, “Fault Location in Distribution Systems Based on Smart Feeder Meters,” *IEEE Trans. Power Del.*, vol. 29, no. 1, pp. 251-260, 2014.

[10] Mallinath, S. Dutta, N. S. Jayalakshimi, and V. K. Jadoun, “Shifting of research trends in fault detection and estimation of location in power system,” *IEEE Access*, vol. 13, pp. 45678-45691, 2025.

[11] R. Perez, C. Vásquez, and A. Viloria, “An Intelligent Strategy for Faults Location in Distribution Networks with Distributed Generation,” *J. Intell. Fuzzy Syst.*, vol. 36, no. 2, pp. 1627-1637, 2019.

[12] H. Livani, C. Y. Evrenosoglu, and V. A. Centeno, “A Machine Learning-Based Faulty Line Identification for Smart Distribution Network,” in *Proc. 2013 NAPS*, pp. 1-5, 2013.

[13] P. Ray, D. P. Mishra, and D. D. Panda, “Hybrid Technique for Fault Location of A Distribution Line,” in *Proc. 2015 Annual INDICON*, pp. 1-6, 2015.

[14] C. G. Arsoniadis and V. C. Nikolaidis, “A Machine Learning Based Fault Location Method for Power Distribution Systems Using Wavelet Scattering Networks,” *Sustain. Energy Grids Netw.*, vol. 29, 2024.

[15] J. Zhang, Z. Y. He, S. Lin, Y. B. Zhang, Q. Q. Qian, “An ANFIS-based fault classification approach in power distribution system,” *Int. J. Electr. Power Energy Syst.*, vol. 49, pp. 243-252, Jul. 2013.

[16] X. Liu, Z. Liu, G. Wang, Z. Cai, and H. Zhang, “Ensemble Transfer Learning Algorithm,” *IEEE Access*, vol. 6, pp. 2389-2396, 2018.

[17] Keras. (2024). *Keras: The Python Deep Learning API*. [Online]. Available: https://keras.io/.

979-8-3315-8138-1