A Travelling Wave-Based Fault Locator for Radial Distribution Systems Using Decision Trees to Mitigate Multiple Estimations

L. S. Lessa*, C. V. C. Grilo*, A. L. Moraes*, D. V. Coury*, R. A. S. Fernandes†
* Department of Electrical and Computer Engineering, São Carlos School of Engineering
University of São Paulo, São Carlos, Brazil
† Department of Electrical Engineering, Federal University of São Carlos, São Carlos, Brazil

***Abstract*--Electrical systems have been facing transformations, such as distributed generation insertion, system expansion and regulatory standards in order to increase reliability and quality of the power supply. Thus, fault location methods must be updated to ensure accuracy in estimating the location of electrical faults. The delay in restoring the system causes damage to utilities and consumers. Considering this, the current work presents an approach capable of locating faults accurately in radial distribution systems. At first, the distance is estimated using the travelling wave theory with data from measurements from two terminals. Next, due to the radial characteristic of the system, the proposal aims to mitigate the problem of multiple estimation of faults. Thus, features are extracted from the voltage and current signals, which are used as inputs of decision trees to identify the fault region. The proposed approach was validated in a medium voltage distribution system, in which the results presented an average error of 0.79% (with a standard deviation of 0.4%) in estimating the fault distances and an average accuracy above 88.7% in identifying the region under fault. Thus, it was demonstrated that the proposed methodology is efficient to locate faults, mitigating the problem of multiple estimation.**

***Keywords*: Distribution systems, fault location, multiple estimation, decision trees, travelling waves.**

I. Introduction

DISTRIBUTION systems have been constantly modified according to technological advances. The standards that regulate the power sector are increasingly rigorous and they prioritize the quality of the service of the power supply. The electrical system reliability is improved when faults are detected and isolated quickly and accurately by protective devices. To decrease the interruption time of the power supply after a fault, several studies have been carried out to develop reliable and accurate fault locators [1], [2], [3].

Over the last decade, electrical systems have grown due to the increasing number of distributed generations, mainly represented by wind and photovoltaic power plants. These sources are connected completely or partially in large distribution systems via inverters. This fact increased the topology complexity of electrical systems and the diversity in system parameters. As a consequence, the difficulties and challenges for locating faults were increased [4].

Looking to the future of protection in electrical systems, detection, location and fault isolation leading to the restoration of the service are pillars for distribution systems, open space for devices that operate in a self-healing mode [5]. In the past, many fault location methods have been developed, some of them based on the apparent impedance calculation. Although their accuracy is affected by problems such as high impedances and fault inception angles, these methods are still in use [4].

On the other hand, travelling wave-based location techniques have proven to be efficient in the current context of distributed systems [6], which is represented by constant modification of the system load profile and diversification of the network topology [7]. Considering the advances in technology, modern travelling wave recorders are capable of operating with high sampling rates (between 1.25 MHz and 20 MHz), digital signal processing, synchronization and data exchange. Thus, these meters allow fault locators based on the travelling wave theory to overcome barriers, aiming to consolidate this kind of meter in the market [8].

However, it is worth mentioning that distribution systems have branches and shunt loads that poses an even greater challenge for fault locators due to the problem of multiple estimation [9]. Recently, many authors have sought new methods to reduce this problem, as proposed in [10] and [11]. The methodology proposed in both studies made use of the concept of low voltage zones, which is based on voltage measurements and fault location estimation using the system impedance matrix. The proposed approach used several meters in the system and an adaptive threshold, based on the calculation of the apparent impedance, associated with the fault distance to indicate regions of voltage sag, and thus reduce the problem of multiple locations of the fault.

Other studies proposed to reduce multiple estimation regions based on data mining [12], [13], which received the voltage and current signals recorded by smart meters. Based on the extracted signals, a feature vector was created that helped to define the relationship between the missing events and the regions.

As an alternative to the methods presented for multiple estimation reduction, the approach proposed by [9] made use of only one phase involved in the fault. This approach used positive sequence components and an adaptive threshold to reduce regions of multiple faults. Furthermore, the authors analyzed the behavior of the methodology as a function of the reduction in the number of meters.

As a solution to the difficulties encountered in the related literature, this article proposes an application of a fault location

This work was supported by the São Paulo Research Foundation (FAPESP) [Grant Number 2021/04872-9] and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) [Finance Code 001].

Paper submitted to the International Conference on Power Systems Transients (IPST2023) in Thessaloniki, Greece, June 12-15, 2023.

technique based on travelling waves together with decision trees to determine the distance and region of the system under fault, respectively. The location technique adopted made use of signals recorded through meters, allocated at the ends of the system. On the other hand, multiple estimation mitigation made use of data from a virtual meter positioned in the substation and connected meters at the ends of the system. These data were delivered to a set of specialized classifiers in each type of fault. It is also worth mentioning that each classifier is responsible for a region of the system.

In short, the contributions of this work are as follows:

- applicability of the technique of locating faults based on travelling waves for distribution systems;
- proposal of the necessary structure for fault location in distribution systems, from the stage of data acquisition, precise calculation of the fault location, as well as the mitigation of the multiple estimation problem;
- mitigation of the multiple estimation problem using machine learning algorithms.

The work is organized as follows. Section II presents the test system and details about the simulations performed. Section III presents the complete structure and steps for online and offline execution of the proposed methodology. The results with discussions are presented in Section IV. Finally, the conclusions are presented in Section V.

II. Modelled and Simulated Radial Distribution System

Fig.1 presents the first test system based on the CIGRE electrical distribution system (topology I), which is inspired by a real distribution system located in southern Germany. This system consists of a balanced and symmetrical medium voltage network, composed by two 20 kV feeders (T1 and T2) and 14 buses. For the execution of this study, this system was modeled in the PSCAD™/EMTP™ software and adapted to operate at a frequency of 60 Hz to meet the Brazilian context. Thus, all parameters of the system lines, represented for the frequency of 50 Hz, were recalculated to adjust the frequency of 60 Hz. Meters were allocated at the ends of the feeders, called *M1* to *M6*, to collect the system operation data.

In addition, the CIGRE system was segmented into 5 non-overlapping regions (*S1* to *S5*). These areas are important for the multiple estimation mitigation process.

III. Proposed Methodology

In this section, details of the proposed methodology for locating the fault and mitigation of the multiple estimation of faults are presented, which can be visualized in Fig. 2. It is worth mentioning that the stages for fault detection and classification were not implemented in this work, as some approaches in the literature present high precision for these tasks [14], [15].

A. Data acquisition

In this stage, in accordance with [3], the signals were sampled at a frequency of 12 MHz and each meter recorded voltages and currents of all phases of the system. Following this, the recorded signals were stored in a database accessed for the offline step execution. It is important to mention that due to the fact that the fault locator required measurements from two terminals, it was considered that the meters were synchronized by a GPS (Global Positioning System). Furthermore, in this measurement strategy, the meters must be allocated at the substation and at the end of each branch.

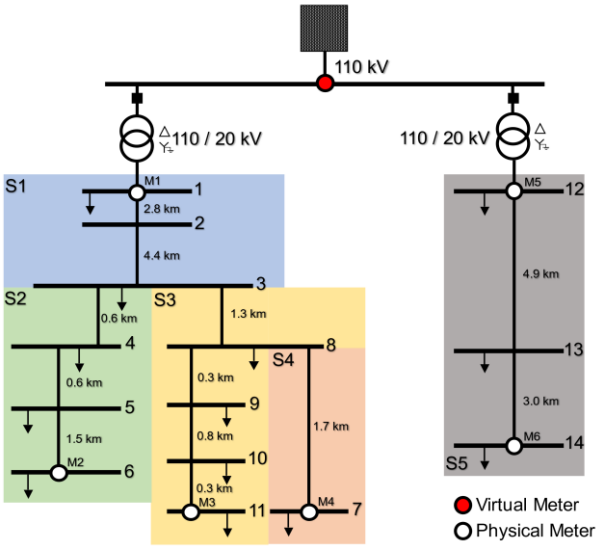

Fig. 1. One-line diagram showing the first test system and division of regions (topology I).

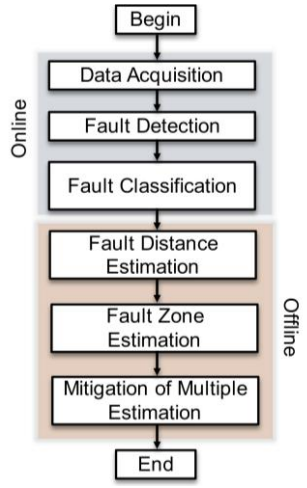

Fig. 2. Flowchart representative of the proposed methodology.

B. Fault detection and classification

The latest methodologies adopted for fault detection and classification were based on intelligent algorithms. These steps can be performed directly on the relay and implemented according to [14] and [15]. Considering that the fault detection rate reached in these papers is 99.9%, the models implemented in this paper did not use data representing non-fault situations.

C. Fault distance estimation

This subsection details the steps taken to estimate the distance from the fault using two terminals. Fig. 3 describes the

steps performed in this process.

1) Modal Transformation: in distribution systems, signals are best observed by decoupling phases into modal components, adopting the modal transformation technique. Modal transformation allows the three-phase system to be treated as a system with three single-phase circuits using the three decoupled modes - ground mode (mode 0) and two aerial modes (α and β modes) - each with its own characteristics [6].

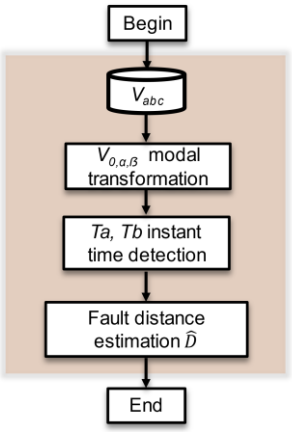

Fig. 3. Flowchart representing the steps performed to estimate the distance of the faults.

The modes were obtained by (1), which were a transformation matrix applied to three-phase systems:

$$\begin{bmatrix} V_0 \\ V_\alpha \\ V_\beta \end{bmatrix} = \frac{1}{3} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 0 & \sqrt{3} & -\sqrt{3} \end{bmatrix} \cdot \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix}, \quad (1)$$

where V_0, V_α and V_β are modal vectors of voltage and V_a, V_b and V_c the voltage phase vectors.

The propagation speed for each mode was calculated by (2), as demonstrated in [1]:

$$v = \frac{1}{\sqrt{LC}}, \quad (2)$$

in which L and C are obtained from the series impedance and shunt capacitance of the sequence components used to represent the electrical system.

2) Time instant detection: the Wavelet transform is widely accepted for a great variety of signals that are not periodic and that may contain sine signals and impulsive signals, a common feature in electrical power systems. Thus, the application of Wavelet to locate faults in electrical systems is adequate, because it can measure the time instants of the reflections of the travelling waves generated during a fault [1], [3].

In practice, the Wavelet transform entails grouping pairs of filters (low pass and high pass) at each scale step. These steps can be considered as successive approximations of the same function, where each approximation details information from a given frequency range. This successive filtering process is called Multi-Resolution Analysis.

To detect the arrival time of a travelling wave, each instant of the signal was analyzed with a threshold, established by means of the comparison with the maximum value of the detail coefficients, for each meter, from a signal in a steady-state situation. To detect the correct time, the value established for the threshold was increased by 10%. Depending on the network noise level, the time obtained was the maximum value of the signal in up to 24 samples after considering the threshold. In this technique, only the time instant of arrival of the first wave, reflected in each of the two terminals, was detected.

3) Fault distance estimation: after detecting the highest value using the threshold established to the terminal meter *A*, the same procedure was repeated with the signal recorded by the other meter (*B*). These instants were denoted as T_a and T_b, as presented in Fig. 4 for a fault case.

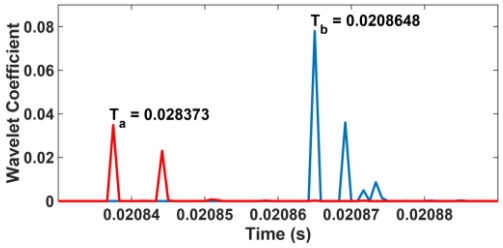

Fig. 4. Detection of the time instants of the reflected waves.

After detecting the time instants, the delay was calculated using (3):

$$T_{delay} = T_b - T_a, \quad (3)$$

Thus, the distance by the travelling wave was calculated and subtracted from the total length of the line [1], [2], based on the formulation of the distance given by (4):

$$\hat{D} = \frac{l - Vp \times T_{delay}}{2}, \quad (4)$$

in which $\hat{D}$ is the estimated distance from the fault (in meters); Vp is the wave propagation speed in the stretch for α mode (in m/s); l is the length of the section covered by the meter (in meters); T_{delay} is the time difference between the measuring terminals (in seconds).

D. Determining possible fault regions

From the distance of the fault, estimated by the locator, it was possible to determine in advance in which possible regions of the system the fault could have occurred. Considering the one-line diagram of Fig. 1, if the fault occurred at 7.5 km from the *M1* meter, the possible regions of the fault would be *S2* or *S3*.

As presented in the next subsection, each system region had a specific algorithm for analyzing the occurrence or not of a fault. Therefore, based on the distance estimated by the fault locator, a pre-selection of possible faulted regions was performed. Thus, only the expert algorithms of the sensitized regions were activated. Furthermore, if no fault occurred, the algorithms were kept on standby.

E. Mitigation of multiple estimations

This subsection presents the methodology used to deal with the problem of multiple estimation. As seen before, the presence of branches in distribution systems leads to a greater difficulty in dealing with the exact fault location. In order to address this problem, this article presents an approach based on decision trees that complements the fault locator response by identifying the faulted region, as illustrated in Fig. 5.

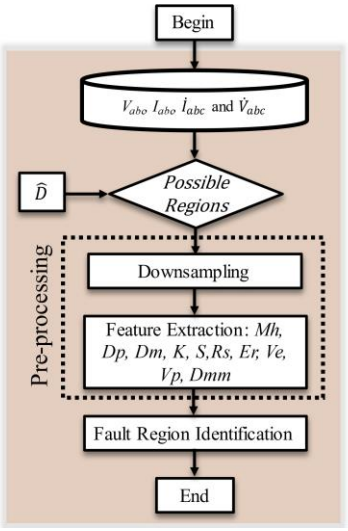

Fig. 5. Flowchart of multiple estimation mitigation routine.

1) Pre-processing: to perform this subroutine, the voltage and current signals presented by a virtual meter were considered, which emulates the behavior of a real meter positioned at the substation. By applying Kirchhoff's laws, the voltages and currents were obtained at a point of interest of the system. In addition to the virtual meter data, voltage and current data originating from real meters located at the ends of the system, downstream of the substation, were considered. Moreover, the application of a down-sampling step in the acquired signals was considered to adjust the sampling to 256 samples/cycle (15,360 Hz), and thus contribute to reducing the computational burden of this subroutine.

From the voltage and current signals obtained by the meters, the zero sequence components were obtained, represented by V_0 and I_0, respectively. Afterwards, the superimposed components were calculated involving the voltages and instantaneous currents of each of the phases (V_{a_sup}, V_{b_sup}, V_{c_sup}, I_{a_sup}, I_{b_sup} and I_{c_sup}) and zero sequence voltages and currents (V_{0_sup}, I_{0_sup}), only for the virtual meter. These calculations, represented by the difference between the post- and pre-fault signals, reduce the impact on the system against the loading variations [16].

In addition, to reduce the influence of high frequency transients on the calculation of superimposed components, the second signal cycle prior to the beginning of the fault was considered as a pre-fault and the third signal cycle after the fault as a post-fault.

Afterwards, features of the previously presented signals were extracted. Thus, the features proposed in [17] were extracted, as follows:

- **Based on statistics:** Harmonic Mean (Mh), Standard Deviation (SD), Mean Deviation (Dm) and Kurtosis (K);
- **Based on the amount of signal information:** Entropy (S), Shannon Entropy (Es) and Rényi Entropy (Er);
- **Based on signal amplitude:** Root mean square (V_{RMS}), Peak value (Vp), and Difference between maximum and minimum window (Dmm).

Finally, the ratio of the voltage and phasor current of each phase was performed to represent the apparent post-fault impedance ($\dot{Z}_{app}$); and the ratio between the superimposed phase voltage and current, to represent the apparent overlapping impedance of the system ($\dot{Z}_{app\_sup}$). Each of these features was separated into real (R) and imaginary (X) to represent the data delivered to the intelligent algorithm.

2) Identifying the fault region: in this study the machine learning algorithm known as Extra-Tree-Classifier (or known as Extremely Randomized Trees Classifier) was used [17], which is capable of dealing with classification problems. This algorithm created Decision Trees at random, combining the results of each tree to find the final decision. A Decision Tree has a structure similar to a flowchart, hierarchically structured and comprising a set of interconnected nodes. For each data sample presented for the tree, each internal node performs a conditional test of the type "*if <condition>, then ...; else ...*", to determine an output response [19].

As presented later, the reference distribution system of this study was divided into regions. The proposed algorithm had the task of identifying the region of the system in which the fault possibly occurred, thus mitigating multiple estimations. To do this, the problem was modeled as a binary case, in which class 1 represents the occurrence of the fault in each region and class 0 to non-occurrence. For each type of fault an ensemble of Extra-Trees was established, in which each of them was responsible for a region of the system.

The classifiers were implemented in Python language using the *ExtraTreesClassifier* function of the Scikit-Learn library. The algorithms were configured to operate with unbalanced database and with a number of estimators equal to 100.

For training and validation of the proposal, 2,484 fault cases were simulated in the PSCAD™/EMTDC™. Based on these cases, a 10-fold cross-validation [20] was performed to train and validate the models, while 30% of the data were used to test them. More details about the simulated cases are presented next.

F. Simulation of fault conditions for topology I

To validate the present work, the study was divided into two parts, namely: (i) estimation of the fault distance and (ii) mitigation of multiple estimations. To analyze the performance of the fault locator, LG, LL, LLG and LLL faults were considered, distributed into 10, 25, 40, 50, 75 and 90% of the length of each section of the system. The fault resistances ranged from 0, 10 and 25 Ω and the fault inception angles were 0, 45 and 90°.

On the other hand, in order to analyze the mitigation of multiple estimation of faults, the same types of faults were

considered, which were applied in the regions *S1* to *S5* in distances of 100 in 100 meters for each section, with fault resistances randomly varying between 0.001 and 30 Ω and with fault inception angles ranging randomly between 0° and 100°. These parameters were adapted from [21].

G. Performance evaluation metrics

In this study, only the estimated distance from the fault was adopted as a metric to evaluate the behaviour of the fault locator. Therefore, the relative error (5) was calculated by:

$$Error_{rel}\ [\%] = \frac{D - \widehat{D}}{l} \times 100 \tag{5}$$

where D is the actual distance from the fault. However, in order to obtain an overview, the average percentage error was still considered.

To evaluate the performance results of each classifier used to mitigate the multiple estimation, individually for each region, the use of confusion matrices was considered [22]. Through a confusion matrix, the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN) resulting from the classification process for the test data were reported and compared. Given this information, the accuracy (Acc) of the model was calculated, according to (6):

$$Acc\ [\%] = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{6}$$

Additionally, the percentage of multiple estimation reduction (R%), proposed by [12], was calculated as follows:

$$R\% = \frac{\sum D_r(n)}{\sum D_{sme}(n)} \times 100, \tag{7}$$

where n is the total number of simulated faults; $D_r(n)$ represents the distance between the beginning of the region under multiple estimation to the fault location, taking into account only the cases in which the faults were correctly located by the algorithm; and $D_{sme}(n)$ is similar to $D_r(n)$, however, considering all the faults correctly located or not.

IV. Results

The results obtained from the proposed methodology considering topology I are presented in this section. First, the results of the fault locator based on travelling waves are analyzed, and then the results of multiple estimation mitigation are presented.

A. Location of faults

In all the results presented below, the average relative errors for all cases of the fault were considered. Only average errors were considered to evaluate the performance, as the location techniques based on travelling waves are robust and have virtually no considerable influences for inception angles and the resistances of the fault.

Fig. 6 shows the performance of the locator in relation to the distance from the fault occurrence to the meter. Errors were higher for faults very close to the meter due to the high frequencies resulting from the fault. Considering the pairs of meters corresponding to the feeder derived from the T1 transformer, the largest errors did not exceed 1.5%. The largest error of 2.7% was obtained for the pair of feeder meters derived from the T2 transformer. The major errors occurred due to the difficulty of detecting the instant time of the first wavefront reflected at the location of the fault. Faults applied exactly at half the length of the line between the pair of meters obtained the smallest errors ranging between 0.04% and 0.36%, almost representing an exact location of the fault.

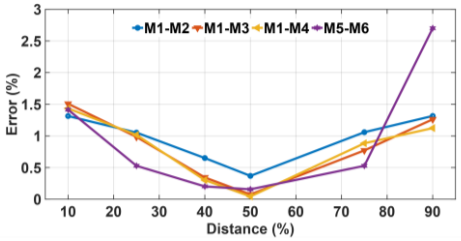

Fig. 6. Locator accuracy for different distances from the electrical system (topology I).

Another performance analysis of the proposed algorithm was carried out for 10 different types of faults, which were applied in all pre-established distances, prioritizing the relationship of the probability of occurrence and severity provided by the fault of the electrical network. The result of the algorithm's performance in relation to the type of fault is observed in Fig. 7.

The greatest influences occurred for LG and LLG faults. The biggest errors were 1.09% and 1.03%, for LG and LLG faults, located by the M5-M6 meter pair. The smallest errors were 0.52% and 0.53% in LL and LLL faults. Although the largest error obtained was 2.7%, the results validate the fault locator when used in distribution systems with radial topology, that is, for a scenario that presents greater complexity when detecting time instants of the travelling waves.

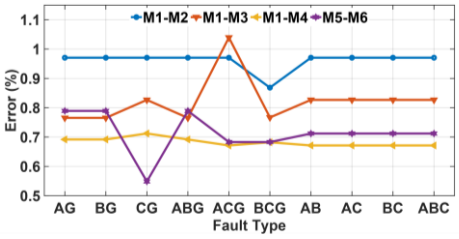

Fig. 7. Locator accuracy in relation to the type of fault (toplogy I).

B. Mitigation of multiple estimations

Unlike the estimation of the fault distance, the sub-routine for mitigating multiple fault estimation is a stochastic process that needs to go through a learning stage. This subroutine was performed to identify in which region of the system the fault occurred. Figures 8 to 11 show the individual performance of each classifier for a given region and type of fault. "F" indicates the presence of a fault in the region, while "NF" indicates the absence of a fault. The first element of the matrix (TP) denotes

that there was a fault in the region and the estimator correctly identified it as such. The second element (FP) denotes that there was no fault in the region, but the estimator incorrectly identified it as faulty. The third element (FN) denotes that there was a fault in the region, but the estimator failed to identify it. The fourth element (TN) denotes the absence of a fault in the region and the estimator correctly identified it as such.

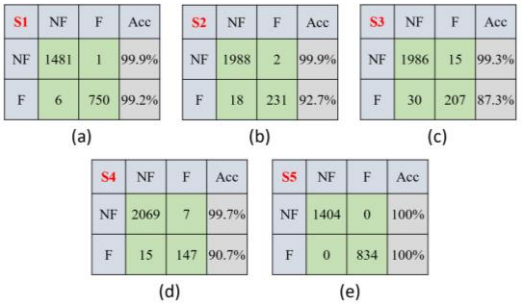

Fig. 8. Results for fault Line-Ground.

Observing the results for single-phase faults, represented in Fig. 8, it can be observed that the classifiers individually achieved an average accuracy of 95.1% to indicate whether a fault occurred (class 1) in a given region of the system. For regions *S2*, *S3* and *S4*, for example, the accuracies found for TP were 92.7%, 87.3% and 90.7%, respectively. Moreover, in relation to TN (no occurrence of fault in the region), the classifiers individually reached accuracies between 99.3% and 100% for single-phase faults, highlighting the results attributed to regions *S1* and *S5*, with performances equal to 100%.

Concerning biphasic faults (Figures 9 e 10), the performance was quite similar to the scenario of single-phase faults. In general, the accuracies varied within the range of 99.3% and 100% for the identification of TN and 80.6% and 100% for TP regarding non-grounded biphasic faults; and accuracies between 99.2% and 100% for TN and between 87.8% and 100% for TP regarding grounded biphasic faults. Therefore, it can be concluded that the absence of ground in biphasic faults implied in a subtle reduction in the performance when identifying the faulty region, especially concerning regions *S3* and *S4*.

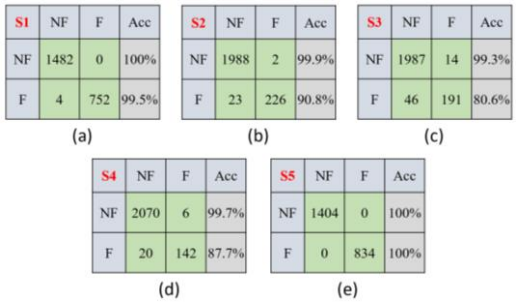

Fig. 9. Results for fault Line-Line

The individual results of the classifiers for three-phase faults are presented in Fig. 11. For TN identification, accuracies between 99.4% and 100% were achieved; while for TP accuracies, they were between 91.3% and 100%.

Although the aforementioned results indicate a high accuracy, they only show the individual performance of each classifier for a given faulty region. Above all, the main scope of this work is to indicate in which of the regions of the system the fault occurred. The performance obtained specifically for each type of fault is shown in Fig. 12.

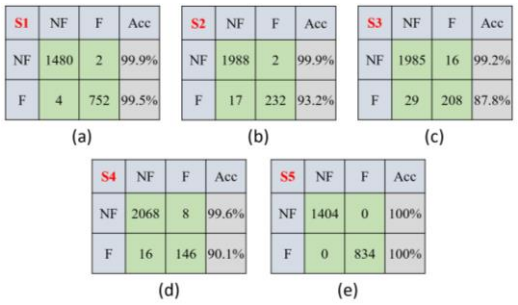

Fig. 10. Results for fault Line-Line-Ground.

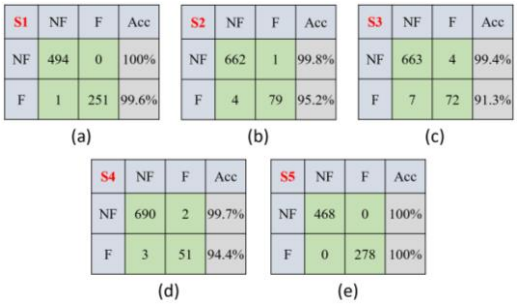

Fig. 11. Results for fault Line-Line-Line.

Based on the mitigation results (Fig. 12), an average accuracy of 88.7% was obtained in the correct identification of the fault region. Among the four types of faults tested, there was a higher identification accuracy for LLL faults (92.6%) and a lower accuracy for LG faults (85.6%).

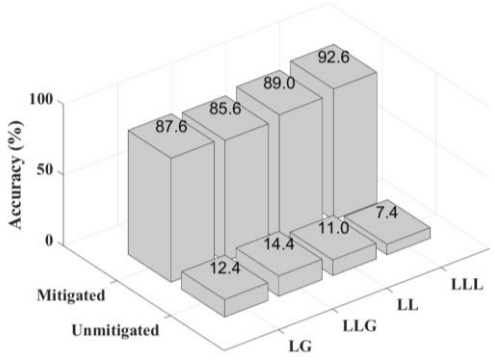

Fig. 12. Multiple estimation mitigation results for topology I.

It is important to note that the unmitigated faults, indicated in Fig. 12 represent the cases in which the algorithm either indicated more than one region as a fault location or did not indicate any region even if there was a fault in any of them. For LG type faults, for example, 11.4% of these were not identified in the regions subject to multiple estimation and 1.0% returned with more than one indicated region. In the case of LLG faults, 13.9% were not identified and 0.5% with more than one region

indicated. For LL faults, 9.8% were not identified and 1.2% obtained more than one indicated region. And for LLL faults, 6.5% of faults were not identified and 0.9% showed more than one identified region as output (Table I).

Furthermore, Table II presents the evaluation of the results using the R% metric. LLL faults continue to represent the cases in which there was the greatest reduction in the multiple estimation (93.5%), while the LLG had the lowest reduction (85.9%).

TABLE I
RESULTS IN TERMS OF NON-IDENTIFICATION OF REGION UNDER FAULT OR UNMITIGATED MULTIPLE ESTIMATION FOR TOPOLOGY I.

Fault / Type of problem	LG	LLG	LL	LLL
No region identified	11.4%	13.9%	9.8%	6.5%
Multiple regions identified	1.0%	0.5%	1.2%	0.9%

TABLE II
MITIGATION RESULT IN TERMS OF R% FOR TOPOLOGY I.

Fault type	LG	LLG	LL	LLL
R%	89.7	85.9	90.0	93.5

Finally, it is important to comment that, despite different approaches and test distribution systems, the results are in line with those obtained in [9], [12] and [11]. While the present study reached an average percentage reduction value (R%) of 89.8%, the work by [9] obtained an average reduction value of 91.9%. Regarding [12], the average percentage of reduction achieved was 94.2%. However, no cross-validation technique was applied to verify the robustness of the proposed approach. In [11], in turn, the R% calculated by [9] showed that the average percentage of reduction was less than 80%, which was much lower than the performance attained in this study.

V. TESTS FOR A DIFFERENT TOPOLOGY

In order to validate the reliability of the proposed framework, the methodology was implemented in a second topology for the distribution network shown in Fig. 13 (topology II).

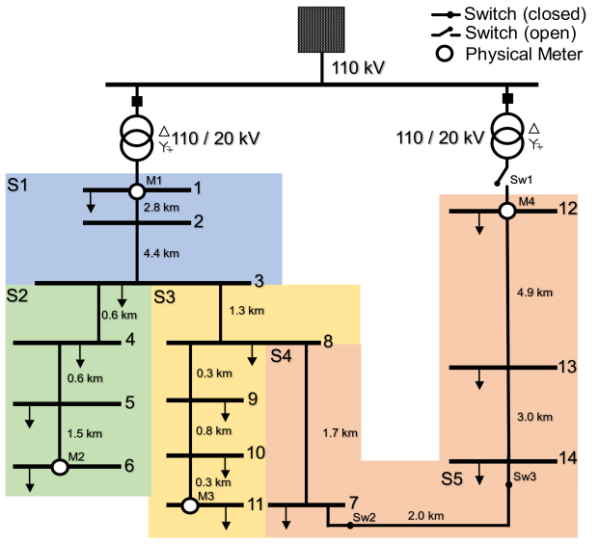

Fig. 13. The second test system and division of regions (topology II).

This system differs from the one in Fig. 1 due to the presence of an extended branch belonging to zone S4 and the meters were relocated at the ends of each branch.

Tests were carried out considering faults to the new topology, varying fault resistances and angles. The location methodology using two terminals achieved better performance in fault location for topology II. This can be observed in Fig. 14, where faults applied in a branch with 20.1 km between meters M1-M4 in the system of Fig. 2 (topology II) presented more accurate calculations of the fault distance if compared to the pair of M1-M4 in topology I (Fig.1).

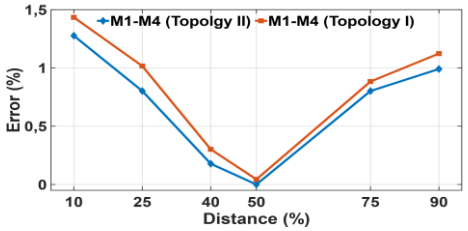

Fig. 14. Fault location errors for topologies I and II considering meters M1-M4.

The performance in topology II reached a fault location 11% better than in topology I, and for faults located near the middle of the line, the error was almost zero. It should be noted that for the methodology adopted, the greatest difficulty lies in locating faults near the meters due to the difference in propagation time between the meters. Furthermore, other combinations of meter pairs (different from M1-M4 mentioned) did not show significant variations from the values depicted in Fig. 6 for different fault locations.

Regarding the problem of multiple estimation, it is worth noting that changing the system topology did not have any adverse effect on the algorithm's performance, as shown in Fig. 15. The results of the conducted tests indicate that an accuracy rate of over 92.3% was achieved in correctly detecting the fault zone, which is higher than the value obtained for topology I. Furthermore, it can be observed that the best and worst detection performances were still found for LLL and LLG faults, respectively. Detailed results of the mitigation are shown in Table III.

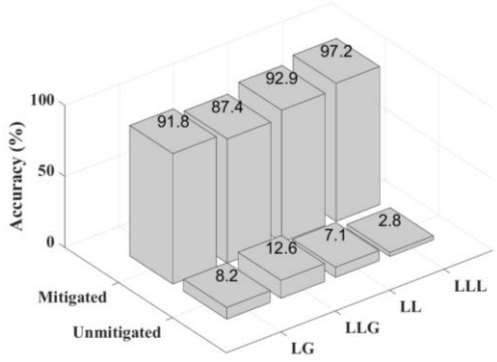

Fig. 15. Multiple estimation mitigation results for topology II.

The results based on the R% metric are presented in Table II. It is worth noting that the values obtained are similar to those found for topology I. Only a slight increase of approximately 3% was noted for LG, LL, and LLL faults, and a 1.1% increase was observed for LG faults.

TABLE III
RESULTS IN TERMS OF NON-IDENTIFICATION OF REGION UNDER FAULT OR UNMITIGATED MULTIPLE ESTIMATION FOR TOPOLOGY II.

Type of problem \ Fault	LG	LLG	LL	LLL
No region identified	8.0%	11.4%	6.8%	2.6%
Multiple regions identified	0.2%	1.2%	0.3%	0.2%

TABLE IV
MITIGATION RESULT IN TERMS OF R% FOR TOPOLOGY II.

Fault type	LG	LLG	LL	LLL
R%	92.2	87.0	92.9	96.9

VI. CONCLUSIONS

This article proposed an application of a fault location technique based on travelling waves for radial distribution systems together with decision trees to determine the faulty region of the system, respectively. To do this, a methodology was proposed that integrates the accuracy of a fault locator based on travelling waves with an ensemble of decision trees responsible for identifying the faulty region with high precision and, consequently, mitigate the problem of multiple estimation. Unlike the approaches found in the literature, a virtual meter was considered and allocated in the substation to obtain the voltage and current signals used as inputs (followed by a feature extraction stage) of the multiple estimation mitigation technique. Considering the results, it was observed that the proposed approach was able to locate the faults precisely, mitigating the problem of multiple estimations. It is important to mention that the method presented in this work can be adapted to different topologies. For each analyzed system, it is essential to divide the previous segmentation into non-overlapping regions. Then, for the multiple estimation method, the decision tree model for each region should be retrained, considering the information collected by the meters installed at the ends of each branch.

VII. ACKNOWLEDGMENTS

The authors would like to thank the São Carlos School of Engineering, University of São Paulo, Brazil for the facilities provided.

REFERENCES

[1] 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.

[2] E. C. M. Maritz, J. M. Maritz, and M. Salehi, "A travelling wave-based fault location strategy using the concepts of metric dimension and vertex covers in a graph," *IEEE Access*, vol. 9, pp. 155 815–155 825, 2021.

[3] L. Xie, L. Luo, Y. Li, Y. Zhang, and Y. Cao, "A traveling wave-based fault location method employing vmd-teo for distribution network," *IEEE Transactions on Power Delivery*, vol. 35, no. 4, pp. 1987–1998, 2020.

[4] IEEE, "Ieee guide for determining fault location on ac transmission and distribution lines," *IEEE Std C37.114-2014 (Revision of IEEE Std C37.114-2004)*, pp. 1–76, 2015.

[5] G. Liu, T. Jiang, T. B. Ollis, X. Li, F. Li, and K. Tomsovic, "Resilient distribution system leveraging distributed generation and microgrids: a review," *IET Energy Systems Integration*, vol. 2, no. 4, pp. 289–304, 2020.

[6] R. Kumar and D. Saxena, "A traveling wave based method for fault location in multi-lateral distribution network with dg," in 2018 *IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2018, pp. 7-12.

[7] K. Prabakar, A. Singh, M. Reynolds, M. Lunacek, L. Monzon, Y. N. Velaga, J. Maack, S. Tiwari, J. Roy, C. Tombari, I. Mendoza Carrillo, X. Yang, S. Iyengar, W. John, D. Anand, A. Gopstein, and D. Vaidhynathan, "Use of traveling wave signatures in medium-voltage distribution systems for fault detection and location," *Technical Report*, 2021.

[8] F. V. Lopes, R. L. A. Reis, K. M. Silva, A. Martins-Britto, E. P. A. Ribeiro, C. M. Moraes, and M. A. M. Rodrigues, "Past, present, and future trends of traveling wave-based fault location solutions," in 2021 *Workshop on Communication Networks and Power Systems (WCNPS)*, 2021, pp. 1-6.

[9] Y. V. Tresso, R. A. Fernandes, and D. V. Coury, "Reducing multiple estimation for fault location in medium voltage distribution networks," *Electric Power Systems Research*, vol. 199, 2021.

[10] F. C. L. Trindade, W. Freitas, and J. C. M. Vieira, "Fault location in distribution systems based on smart feeder meters," *IEEE Transactions on Power Delivery*, vol. 29, no. 1, pp. 251–260, 2014.

[11] F. C. Trindade and W. Freitas, "Low voltage zones to support fault location in distribution systems with smart meters," *IEEE Transactions on Smart Grid*, vol. 8, pp. 2765–2774, 2017.

[12] J. V. Sousa, E. A. Reche, D. V. Coury, and R. A. Fernandes, "Cloud computing in the smart grid context: an application to aid fault location in distribution systems concerning the multiple estimation problem," *IET Generation, Transmission & Distribution*, vol. 13, no. 18, pp. 4222–4232, Sep. 2019.

[13] E. A. Reche, J. V. d. Sousa, D. V. Coury, and R. A. S. Fernandes, "Data mining-based method to reduce multiple estimation for fault location in radial distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3612–3619, 2019.

[14] M. Shafiullah and M. A. Abido, "S-transform based ffnn approach for distribution grids fault detection and classification," *IEEE Access*, vol. 6, pp. 8080–8088, 2018.

[15] M. Mishra and P. K. Rout, "Detection and classification of micro-grid faults based on hht and machine learning techniques," *IET Generation*, Transmission & Distribution, vol. 12, no. 2, pp. 388–397, 2018.

[16] R. Aggarwal, D. Coury, A. Johns, and A. Kalam, "A practical approach to accurate fault location on extra high voltage teed feeders," *IEEE Transactions on Power Delivery*, vol. 8, no. 3, pp. 874–883, 1993.

[17] F. A. S. Borges, R. A. S. Fernandes, I. N. Silva, and C. B. S. Silva, "Feature extraction and power quality disturbances classification using smart meters signals," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 824–833, 2016.

[18] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine Learning, vol. 63, no. 1, pp. 3–42, 2006.

[19] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Amsterdam: 3a ed., Amsterdam: Morgan Kaufmann, 2011.

[20] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014.

[21] E. A. Reche, J. V. d. Souza, D. V. Coury, and R. A. S. Fernandes, "Data mining-based method to reduce multiple estimation for fault location in radial distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3612-3619, 2019.

[22] R. O. Duda and D. G. Stork, "Pattern classification," *2a ed., New York: Wiley*, vol. 10, no. 4, pp. 3612–3619, 2001.