Fault location in radial distribution systems based on decision trees and optimized allocation of power quality meters

André Luís da Silva Pessoa
Mário Oleskovicz
Department of Electrical and Computer Engineer
University of São Paulo, USP
São Carlos, Brazil
alsp@usp.br, olesk@sc.usp.br

***Abstract*—One of the reasons for not achieving satisfactory indices of Power Quality (PQ) is due to the discontinuity of power supply in Distribution Systems (DS), usually caused by the occurrence of short-circuits. In this context, to characterize these occurrences, a database was compiled by simulations in the IEEE 34-bus DS using the ATP (Alternative Transients Program) software. In these simulations, the type, location and fault impedance were used as parameters. The voltages and currents of all three phases of the power quality meters optimally allocated in the DS were considered. Based on these measurements, the J48 decision tree algorithm was used to identify in which area of the 34-bus DS the single-phase faults occurred. In order to use the J48 decision tree, the WEKA (Waikato Environment for Knowledge Analysis) software was used. Promising results demonstrated the effectiveness of the proposed algorithm to locate the single-phase short-circuit situations considered.**

***Index Terms*—decision trees, distribution systems, fault location, power quality meters.**

I. Introduction

Power Quality (PQ) is an area of Electrical Engineering in which its scope is difficult to determine as it covers a wide range of aspects from generation to transmission and electric power distribution, as well as the end consumers of the product ”electricity”. Thus, as PQ affects all end users of the electrical system, power utilities should take into account the relationship between the system users equipment with the available voltage levels. The energy end consumers, mainly the industrial ones, should also respect the other users connected to the Distribution Systems (DS) by controlling the possible sources of disturbances associated with PQ and mitigating the problems that can arise when connecting their linear or nonlinear loads to the network [1].

As presented in [1], PQ is an important aspect for power systems, significantly affecting the operation, safety, efficiency and rehabilitation of the entire system. Among the various definitions, the term PQ has been generally used to express the quality of voltage supply according to pre-established standards. Taking this into account, PQ can be defined as the measurement, analysis and improvement of the voltage in a bus in order to keep its form close to the sine wave, with fixed amplitude and frequency. Utilities and end users agree that more than 60% of PQ problems are caused by natural and unpredictable events, such as: short-circuits, lightning, ferroresonance and geomagnetic induced currents, among others.

Short-circuits are among the various causes of PQ disturbances in power systems. In this scenario, in order to reduce the damage to consumer loads, it is essential that these short-circuits are accurately and quickly located [2]. It is worth noting that, depending on the DS protection philosophy applied by many power distributors, the precise location of short-circuits is based purely on information provided by the system users and tests to delimit the area where the disturbance occurred [3].

The initial searches for fault location were aimed at the transmission lines. They were based on the apparent impedance calculations, which were defined based on the current and voltage phasors under analysis. These techniques were developed and improved after the introduction and widespread use of micro processed equipment [4]. Subsequently, techniques based on artificial intelligence were developed aiming at locating the faults as efficiently, or even more, than traditional methods.

Specifically focusing on DS, research carried out by [5] used artificial intelligence techniques for fault location. In this research, signals resulting from short-circuits were processed by using the Wavelet Transform (WT) and were then applied to Artificial Neural Networks (ANN) and *Fuzzy Logic* to locate faults. [6] presents an approach using ANN and Discrete Wavelet Transform (DWT) considering the fourth level of decomposition with Daubechies 4 as the mother wavelet. [7] used multi-agent systems to determine in which area of the system the faults occurred. [8] developed a methodology based on the decision tree algorithms C4.5, ID3, Gini and Chi, and also used a single measuring point in the substation of the tested system. Regarding an approach based on a previous allocation of power quality meters, [9] developed an analytical method to determine the location of fault occurrence.

Considering that faults must be quickly and accurately located, this research proposes using a J48 decision tree algorithm to locate the area and the fault occurrence zone, based on data extracted from three-phase voltage and current signals recorded by PQ meters optimally allocated in the system. Another point that this research intends to better answer is related to multiple estimations of fault location from a specific point, usually a measurement point on the electrical system of interest [10]. This multiple estimation can occur due to the various paths using the same length that can exist until the actual short-circuit position, considering the main feeder and the various lateral branches that there may be in a distribution system.

Besides this introduction, this paper is divided into three other sections. Section 2 presents the methodology adopted in this research. Section 3 presents the results and Section 4 draws the relevant conclusions of this research.

II. Methodology

In this research, the model of the IEEE 34-bus DS was considered, where the three measurement points are the optimal allocation of PQ meters. This optimum allocation of PQ meters results from previous work [11] and is not the focus of this article. After that, the system was divided into three areas. In sequence, eleven specific zones to locate possible short-circuits were defined based on these three areas.

For the moment, only single-phase short-circuits were simulated and considered in this research. Depending on the type of fault, the area and the zone where each fault occurred were determined by using the J48 decision tree [12]. Each of the main steps of the developed algorithm will be presented and described in this paper.

It should be noted that all data processing was performed on a PC core i7 @3.6 GHz and 16 GB RAM.

A. Modeled system and simulations

In order to use the J48 decision tree for fault location, a prior model of the IEEE 34-bus DS was used [13]. An interface between the ATP [14] and MATLAB [15] was considered in the simulations to handle and process the required information more efficiently.

For this research, an optimized PQ meter allocation [11], considering only three PQ meters, provided sensitivity to observe all the voltage sags with a remaining voltage of at least 0.9 p.u. on the 34-bus DS. Based on the position of these three PQ meters, the 34-bus DS was divided into three areas, as shown in Figure 1.

In each simulation of the short-circuits used, a sampling rate of 256 samples per cycle was considered. In order to estimate the fault distance from one of the three meters allocated, a post-fault cycle of three-phase voltages and current signals was considered. As for the fault positions, they were used in DS branches with less than 800m at 1%, 25%, 50%, 75% and 99 % of its length. For line lengths longer than 800m, the short-circuit situations were applied every 200 meters. It is worth mentioning that a threshold of 200 meters was adopted because of the future aim of exact estimation of the distance to fault, within an average error of approximately 200m.

In addition to the fault position variation, the values of fault impedances 0.0001 Ω, 10 Ω, 20 Ω, 30 Ω, 40 Ω and 50 Ω were also considered to generate a dataset to train the algorithm. To generate a test dataset, faults applied every 50m and the fault impedance values of 5 Ω, 15 Ω, 25 Ω, 35 Ω and 45 Ω were used.

For this research, it was considered that all the processing of the information from the three allocated PQ meters will be concentrated in the fault location system, as shown in Figure 2.

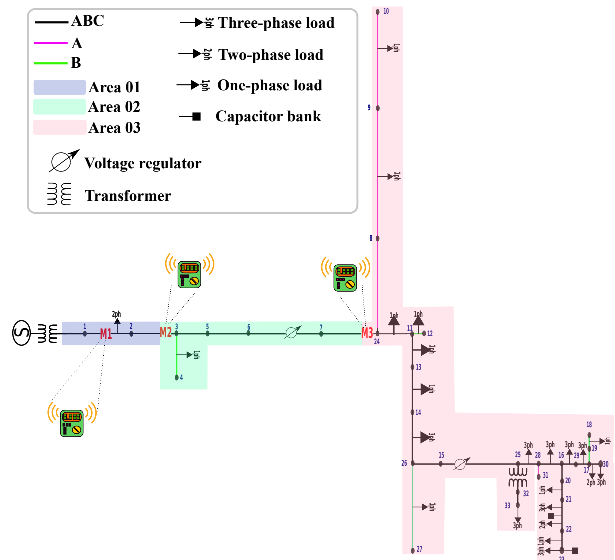

Fig. 1. IEEE 34-bus distribution system with PQ meter positions.

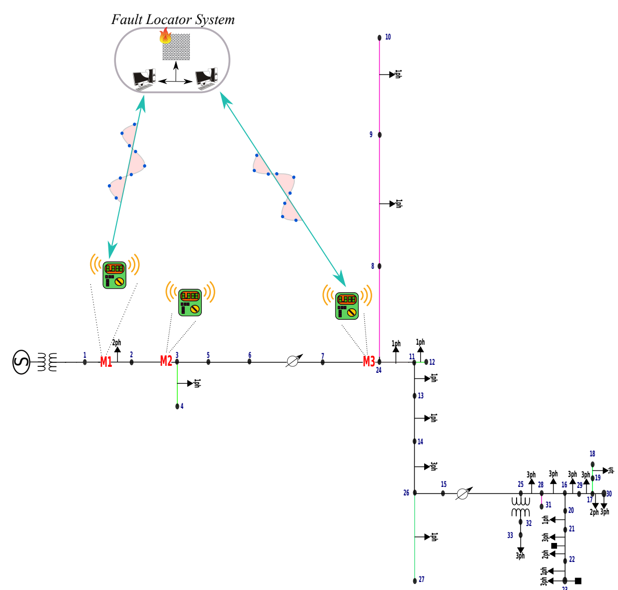

Fig. 2. Data acquisition from PQ meters installed in the system.

For each of the three allocated meters, a specific area was defined. Thus, the meter of this specific area will be considered for the precise location of short-circuits depending on the area where the fault occurred. Thus, after detecting the fault occurrence and its location regarding the short-circuit type, the zone where the fault occurred will be estimated based on data from the three-phase voltage and current signals of the meter allocated in the area to which the zone belongs.

Therefore, eleven zones were characterized using the following criteria:

i) In each of the three areas delimited by the three allocated meters, a zone delimiting the main feeder was assigned; and
ii) Each lateral branch was considered as a zone.

Figure 3 presents the 34-bus DS divided into eleven zones. Table I shows the areas related to each type of single-phase fault considered.

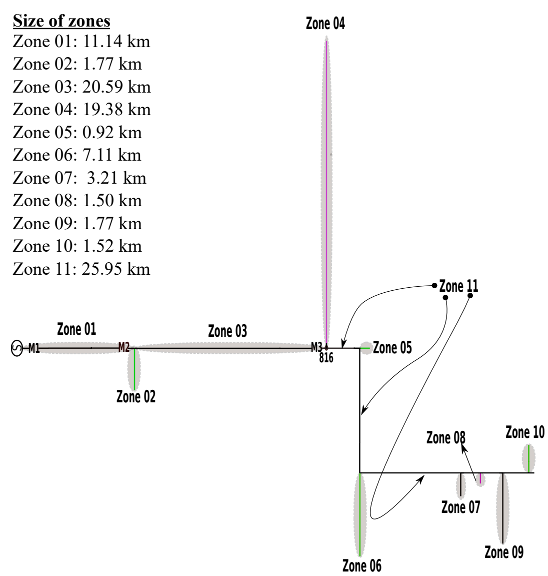

Fig. 3. 34-bus IEEE distribution system divided into 11 zones.

TABLE I
ZONES RELATED TO SINGLE-PHASE FAULTS WHICH OCCURRED IN AREA 03

Fault type	Zones related
A-ground	01 , 03, 04, 07, 08, 09 and 11
B-ground	01, 02, 03, 05, 06. 07, 09, 10 and 11
C-ground	01, 03, 07, 09 and 11

B. Selected parameters

In the context of this research, various parameters were calculated after a fault detection cycle. The calculated parameters were:

i) *RMS* (Root Mean Square) value;
ii) the amplitudes and phases of components of the fundamental frequency; and
iii) the energy of the first two coefficients of the 4th level of Wavelet Packet Transform (WPT) decomposition [16], using Daubecheis with support 4 (db4) as the mother wavelet.

As shown, to define the nomenclature of these parameters, the following criteria were adopted:

i) RMS_i for $i \in \{IA, IB, IC, VA, VB, VC\}$ indicates the RMS value of the three-phase voltage and current signals;
ii) Amp_i and Ph_i for $i \in \{IA, IB, IC, VA, VB, VC\}$ respectively indicate the values of the amplitudes and phases of the three-phase voltage and current signals; and
iii) E_{ji} for $j \in \{01, 02, 03, \cdots, 16\}$ and $i \in \{IA, IB, IC, VA, VB, VC\}$ indicate the energy of coefficients of the 4th level of WPT decomposition of three-phase voltage and current signals.

Therefore, for example:

i) RMS_{IA} is the RMS value of current signals recorded by the meters installed in the system related to phase A;
ii) Amp_{VC} and Ph_{VC} are respectively the amplitudes and phases of the fault voltage signals recorded by the meters installed in the system related to phase C; and
iii) E_{01IB} and E_{02IB} are the energy of the first two coefficients of the WPT fault current signals recorded by the meters installed in the system related to phase B.

C. Identifying the fault occurrence zone

To use the J48 decision tree, WEKA software [17] was used. For the research carried out, decision trees were built for areas 01, 02 and 03 to determine if the phase-to-ground faults occurred in one of the eleven zones previously described. To illustrate, Table II shows the rules to determine whether the single-phase fault phase *A*-to-ground in area 03 occurred in zone 09. Then, by the rules illustrated for zone 9, it turns out that the algorithm indicating the fault occurrence zone can be compiled from a set of IF-THEN conditions.

TABLE II
A RULE OBTAINED AFTER APPLYING THE J48 DECISION TREE ALGORITHM TO IDENTIFY WHETHER THE SINGLE-PHASE FAULTS INVOLVING PHASE A WITH GROUND CONNECTION APPLIED TO AREA 03 OCCURRED IN ZONE 9.

Condition	Rules	Zone
1	$Amp_{IB} > 61.49$ and $E_{02VC} \leq 0.050038$ and $E_{02IA} \leq 0.048028$ and $E_{02VB} \leq 0.009954$ and $E_{04IC} \leq 0.023392$ and $RMS_{IC} \leq 43.562$	**09**
2	$Amp_{IB} > 61.49$ and $E_{02VC} \leq 0.050038$ and $E_{02IA} \leq 0.048028$ and $E_{02VB} \leq 0.009954$ and $E_{04IC} > 0.023392$	**09**
3	$Amp_{IB} > 61.49$ and $E_{02VC} \leq 0.050038$ and $E_{02IA} \leq 0.048028$ and $E_{02VB} > 0.009954$ and $E_{16VB} > 0.004245$ and $E_{14IC} \leq 0.002428$ and $Fas_{VA} \leq 0.73513$	**09**
4	$Amp_{IB} > 61.49$ and $E_{02VC} \leq 0.050038$ and $E_{02IA} \leq 0.048028$ and $E_{02VB} > 0.009954$ and $E_{16VB} > 0.004245$ and $E_{14IC} > 0.002428$	**09**

III. RESULTS

In this section, the results obtained considering the three single-phase fault location in the IEEE 34-bus distribution system will be presented.

For areas 02 and 03, by the presence of lateral branches, the possibility of multiple estimations from the meters allocated in the respective areas can be observed. As there are no lateral branches in area 01, there is therefore no possibility of multiple estimations of the fault from the measurements provided by meter 1. Since the objective of this research is to identify the fault zone and to solve the problem of multiple fault estimations in a better way, the focus of this article is to locate B-to-ground faults that occurred in area 02 and faults A, B and C-to-ground that occurred in area 03.

Initially, emphasis will be placed on the location of phase A-to-ground faults that can occur in area 03. Considering a J48 decision tree algorithm and using training and test data from phase A-to-ground faults applied in area 03, the time to build the model of the J48 decision tree was 0.15 s and the time to evaluate the test set was 0.18 s. Another point that should be emphasized is that out of the 114 parameters presented to the J48 decision tree algorithm to provide all the rules for the location of phase A-to-ground faults, for area 03 only 14 of these parameters were used by the algorithm.

The initial choice of presenting results from area 3 is accounted for by the extent and complexity (several lateral branches) that this area presents. Therefore, the other situations of single-phase faults (B and C to-ground) will also be analyzed in area 3. To illustrate, the training dataset for area 03, considering single-phase faults (phases A, B and C to ground) has 1,956 cases, 1,638 cases and 1,314 cases, respectively. The test dataset for the same condition and phases has 5,155, 4,135, and 3,195 cases, respectively.

In area 2, only the faults involving phase B to ground will be evaluated as this is the only single-phase situation that can result in multiple estimations in area 2 due to zone 2 (Figure 3). The training dataset for area 02, considering only phase B to ground has 690 cases. The test dataset for the same condition has 2,688 cases.

To initially demonstrate the performance of the J48 decision trees to identify faults in the 34-bus DS, a phase A-to-ground fault was applied in area 9 of Figure 3. The oscillography was obtained by applying this fault, as shown in Figure 4.

From this short-circuit situation, the following values were obtained: 63.0041 for Amp_{IB}; 0.0446 for E_{02VC}; 0.0426 for E_{02IA}; 0.0070 for E_{02VB} and 0.0268 for E_{04IC}. From the values of these parameters and the rules shown in Table II, it can be concluded by condition 2 that this fault occurred in zone 09 of area 03.

It is worth mentioning that an accuracy rate of 99.42% was obtained for the 5,155 simulated cases of phase A-to-ground fault by applying the J48 decision tree. Thus, in only 30 applied short-circuits, it was not possible to correctly classify, from the generated rules, the zone in which they occurred. The root mean square error obtained was 0.0477.

Figures 5, 6 and 7 show the confusion matrices concerning the phase-to-ground faults that occurred in area 3. The confusion matrices shows where there is a correct location and identifies the errors made by the algorithm. The row of the confusion matrix shows the output zone (predicted zone) and the column is the target zone (true zone). For example, in Figure 5, 2,632 faults were correctly classified as occurred in zone 11, 8 faults that occurred in zone 4 were classified incorrectly in zone 11 and 4 faults that occurred in zone 9 were classified incorrectly in zone 11. Considering these confusion matrices, the success rate for phases A, B and C-to-ground, respectively, were 99.4%, 98.3% and 99%.

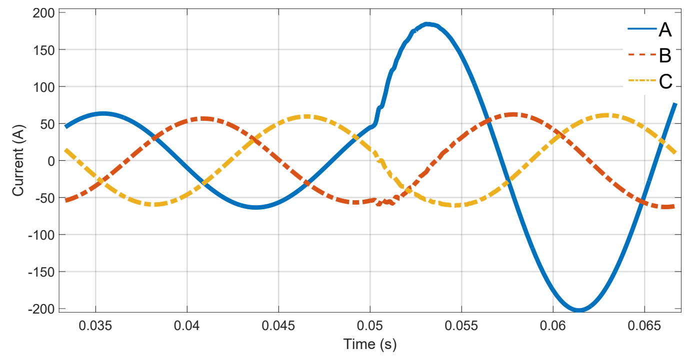

Fig. 4. Oscillography obtained from a phase A-to-ground fault applied in zone 09.

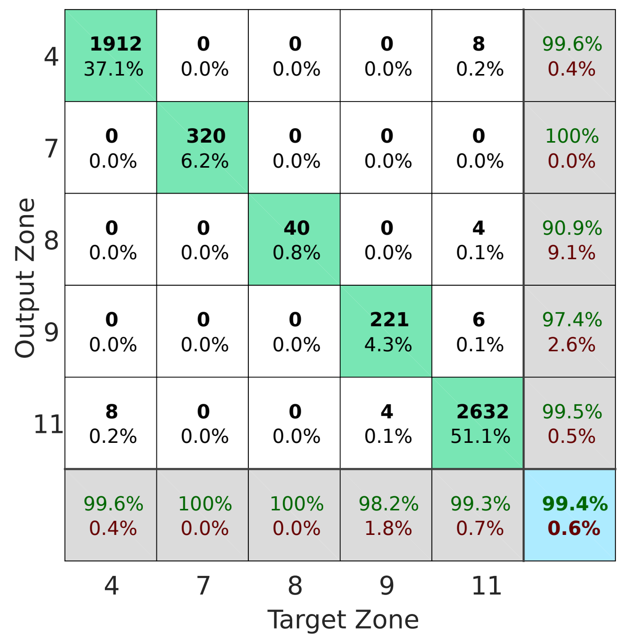

Fig. 5. Confusion matrix for localization of phase A-to-ground faults which occurred in area 3.

Figures 8, 9, 10 and 11 are ROC (Receiver Operating Characteristic) curves used to evaluate the performance of the J48 algorithm considering the fault location applied in areas 3 and 2. These graphs show the algorithm performance for each zone and identifies in which zone it was more difficult to locate the faults. Thus, this type of graph will only be presented for areas 2 and 3, which have more than one zone.

Based on the Area Under the Curve (AUC), analyzing the ROC curves, the efficiency of the fault locator could be determined.

Therefore, it is considered an outstanding location when

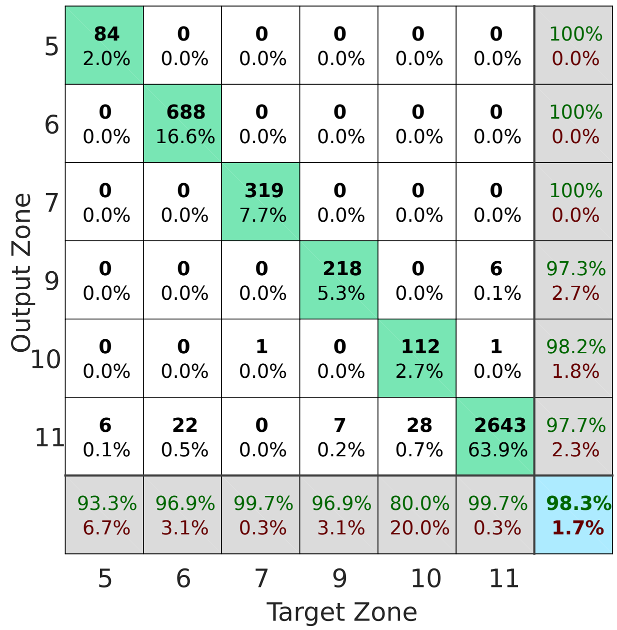

Fig. 6. Confusion matrix for localization of phase B-to-ground faults which occurred in area 3.

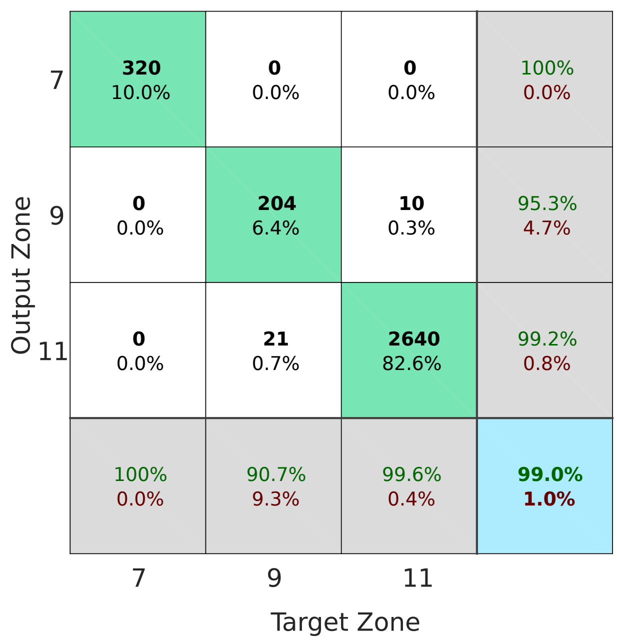

Fig. 7. Confusion matrix for localization of phase C-to-ground faults which occurred in area 3.

AUC > 0.9, an excellent location when 0.8 ≤ AUC ≤ 0.9, an acceptable location when 0.7 ≤ AUC ≤ 0.8, a fair location when 0.6 ≤ AUC ≤0.7, a poor location when 0.5 ≤ AUC ≤ 0.6 and a wrong location when AUC = 0.5 [18].

Figures 8, 9 and 10 show the performance of the J48 algorithm for the location of single-phase faults applied in area 3. Figure 11 refers to the phase B-to-ground faults applied in area 2.

In all these figures, the AUC values are above 0.9. In Figure 11, the performance of the algorithm to identify faults that occurred in the main branch (zone 3) was already similar to the fault location that occurred in the lateral branch (zone 2).

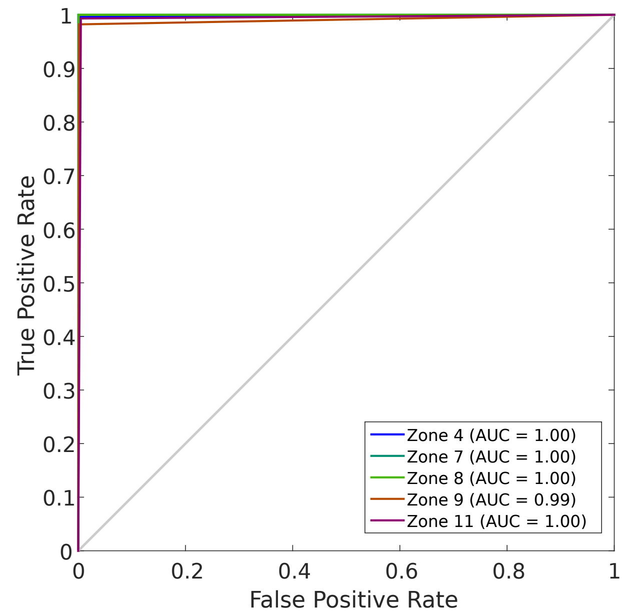

Fig. 8. ROC curve to locate phase A-to-ground faults which occurred in area 3.

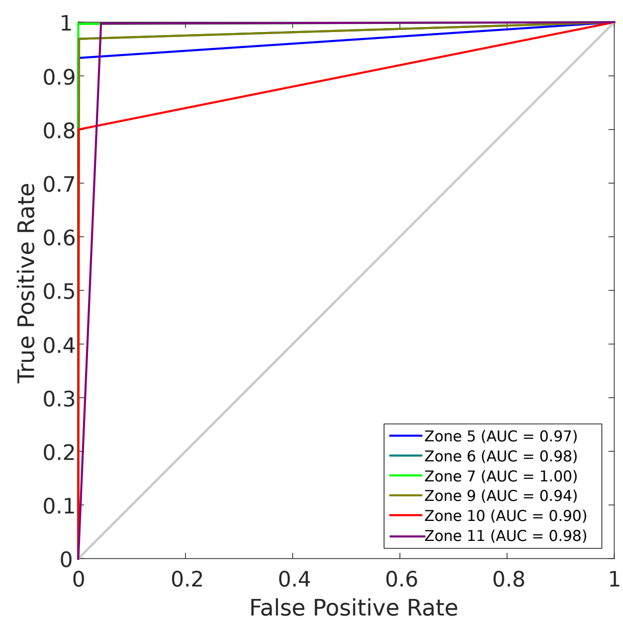

Fig. 9. ROC curve to locate phase B-to-ground faults which occurred in area 3.

IV. Conclusions

This research presented promising results using decision trees to estimate the location of single-phase faults on the IEEE 34-bus distribution system. In areas 2 and 3 defined in the distribution system, and for now, considering only single-phase faults, the J48 decision tree algorithm presented excellent

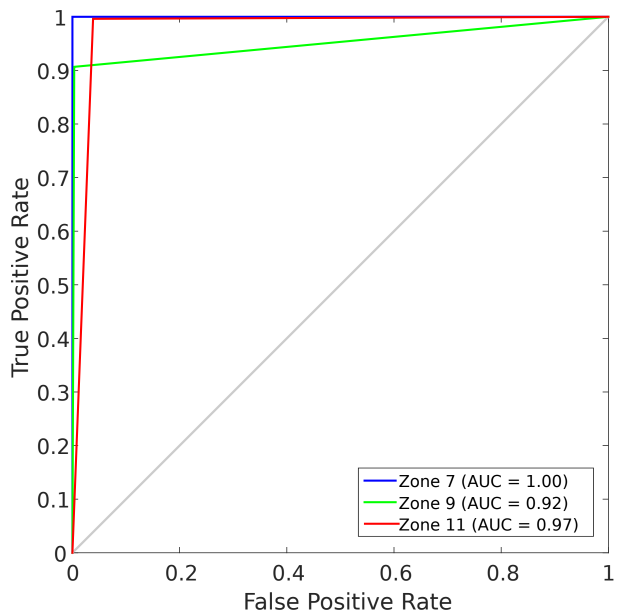

Fig. 10. ROC curve to locate phase C-to-ground faults which occurred in area 3.

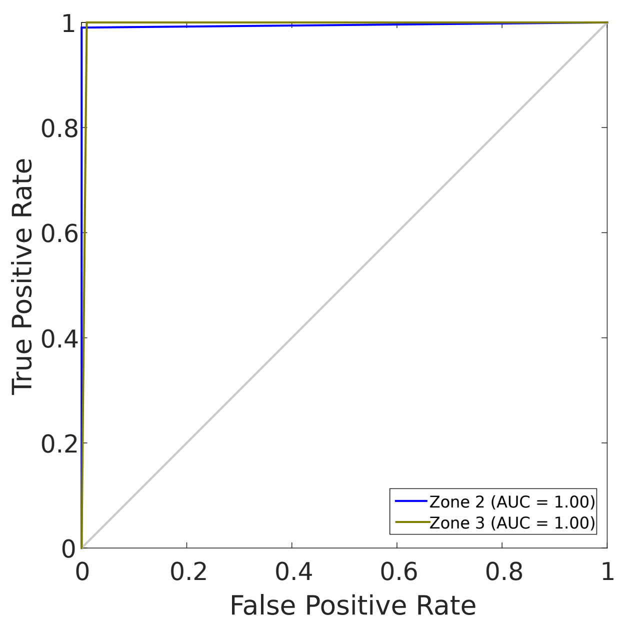

Fig. 11. ROC curve to locate phase B-to-ground faults which occurred in area 2.

results in identifying the occurrence zone on the IEEE 34-bus DS. The results were mainly investigated using confusion matrices, as well as ROC curves. The confusion matrix shows where there is a correct location and identifies the errors made by the algorithm. Based on the Area Under the Curve (AUC), analyzing ROC curves, the efficiency of the fault locator could be determined.

Acknowledgment

The authors would like to acknowledge the Department of Electrical and Computer Engineering, São Carlos School of Engineering, University of São Paulo (Brazil), for the research facilities provided to conduct this project. Our thanks also extend to the financial support received from CAPES and CNPq (Governmental Brazilian Institutions).

References

[1] E. Fuchs and M. A. Masoum, *"Power quality in power systems and electrical machines"*. Academic press, 2011.

[2] P. Farzan, M. Izadi, C. Gomes, M. Kadir, M. Hesamian, M. Radzi *et al.*, "Short circuit power based fault location algorithm in distribution networks," in *Power Engineering and Optimization Conference (PEOCO), 2014 IEEE 8th International*. IEEE, 2014, pp. 105–109.

[3] J. Zhu, D. L. Lubkeman, A. Girgis *et al.*, "Automated fault location and diagnosis on electric power distribution feeders," *Power Delivery, IEEE Transactions on*, vol. 12, no. 2, pp. 801–809, 1997.

[4] L. Sousa Martins, J. Martins, V. F. Pires, and C. Alegria, "The application of neural networks and clarke-concordia transformation in fault location on distribution power systems," in *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES*, vol. 3. IEEE, 2002, pp. 2091–2095.

[5] A. Rafinia and J. Moshtagh, "A new approach to fault location in three-phase underground distribution system using combination of wavelet analysis with ann and fls," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 261–274, 2014.

[6] K. Lout and R. K. Aggarwal, "Current transients based phase selection and fault location in active distribution networks with spurs using artificial intelligence," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

[7] J. Ghorbani, M. A. Choudhry, and A. Feliachi, "Fault location and isolation using multi agent systems in power distribution systems with distributed generation sources," in *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*. IEEE, 2013, pp. 1–6.

[8] A. Zapata-Tapasco, J. Mora-Flórez, and M. C. de Almeida, "Fault location in power distribution systems using a learning approach based on decision trees," in *Transmission & Distribution Conference and Exposition-Latin America (PES T&D-LA), 2014 IEEE PES*. IEEE, 2014, pp. 1–6.

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

[10] M. M. Saha, J. J. Izykowski, and E. Rosolowski, *Fault location on power networks*. Springer Science & Business Media, 2009.

[11] D. P. S. Gomes, "Methodology and simplification of the search space for power quality monitor allocation facing voltage variationsvariations," Master's thesis, University of São Paulo, 2016 (in portuguese).

[12] T. R. Patil and S. Sherekar, "Performance analysis of naive bayes and j48 classification algorithm for data classification," *International Journal of Computer Science and Applications*, vol. 6, no. 2, pp. 256–261, 2013.

[13] D. T. Feeders, "Ieee pes distribution system analysis subcommittee's, distribution test feeder working group," 2013.

[14] L. Prikler and H. Hoidalen, "Atpdraw manual," 2002.

[15] M. MatLab, "The language of technical computing," *The MathWorks, Inc. http://www. mathworks. com*, 2012.

[16] P. Ray, D. P. Mishra, and D. D. Panda, "Hybrid technique for fault location of a distribution line," in *2015 Annual IEEE India Conference (INDICON)*. IEEE, 2015, pp. 1–6.

[17] G. Holmes, A. Donkin, and I. H. Witten, "Weka: A machine learning workbench," in *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*. IEEE, 1994, pp. 357–361.

[18] R. Shatnawi, W. Li, J. Swain, and T. Newman, "Finding software metrics threshold values using roc curves," *Journal of software maintenance and evolution: Research and practice*, vol. 22, no. 1, pp. 1–16, 2010.