Optimized Method for Locating the Source of Voltage Sags

José Carlos Filho, Fabbio Anderson da Silva Borges, Ricardo de Andrade Lira Rabelo, *Member, IEEE,* Ivan Saraiva Silva, and Antônio Oséas de Carvalho Filho

Abstract—Short-Duration Voltage Variations (SDVVs) are the power quality disturbances (PQD) that mainly affect industrial systems, and are originated for various reasons, in particular short circuits over large areas, even those originating in remote points of the electrical system. The location problem aims to indicate the area or region or distance from the substation that is connected to the source causing the voltage sags, and is a fundamental task to ensure good power quality. One of the strategies used to determine the location of sources causing SDVVs and for an implementation of machine learning algorithms in modern distribution networks, called Smart Grids. Monitoring a Smart Grid plays a key role, however mostly it generates a large volume of data (Big Data) and as a result, multiple challenges arise due to the properties of this data such as volume, variety and velocity. This work presents an optimization through genetic algorithm to select meters which already exist in the Smart Grid, using a voltage sag location method in order to reduce the data obtained and analyzed throughout the localization process. Optimization was evaluated through a comparison with a non-optimized localization method, this comparison showed a difference between the hit rates of less than 1%.

Index Terms—Voltage sag, Clustering algorithm, Genetic Algorithm, Disturbance location, Smart Grids, Big Data.

I. INTRODUCTION

S Hort-Duration Voltage Variations (SDVVs) are the power quality disturbances (PQD) that mainly affect industrial systems, and are originated from the start-up of large motors, capacitor failures and, in particular, short circuits over large areas, even those originating at remote points of the electrical system [1].

Among the various phenomena related to PQD, voltage sags stand out due to their frequency of occurrence when compared to voltage rise and voltage interruption events. Voltage sinking is the momentary or temporary reduction of the voltage amplitude in relation to the reference voltage. Voltage sags are directly linked to the continued operation of the electrical system and their presence can damage a large amount of equipment if the protection systems do not work properly [2], [3].

Location problem aims to indicate the area or region or distance from the substation that is connected to the source causing voltage sinking and consists of a fundamental task to ensure good power quality, because once located it is possible to take measures to minimize the effects, such as installation, adjustment and coordination of the system protection devices [4].

This causative source can be a short circuit, energizing large loads, and transformer magnetization. Among these causal sources, short circuit, also called system fault, is the most common cause of this event. However, localization is not a trivial task, since voltage sags occur in a short time interval and propagate throughout the system [1], [5].

One of the strategies used to determine the location of the sources causing SDVVs is the implementation of machine learning algorithms. These methods are characterized by analyzing the characteristics of the distribution system during the occurrence of the disturbance and automatically determining the location of the event. These characteristics come from the real-time monitoring system available in modern distribution networks, called Smart Grids.

Monitoring plays a key role, but mostly generates a large volume of data (Big Data) and as a result multiple challenges arise due to the properties of this data such as volume, variety and velocity [6]. Therefore, while machine learning methods need to analyze a large volume of data, there is also a concern about the volume of data generated by the system. For the increase of data in the communication channels can cause latency problems and packet errors, and requires a robust communication architecture, whose data is collected by the smart meters installed in the network and sent to an operation control center.

The size of the message packet can vary greatly depending on the type and amount of data the energy provider needs. For simple functions, such as remote reading the consumption of energy, a small message packet size may be sufficient, while for sophisticated functions, such as deriving consumption profiles and identifying the type of electrical devices that are turned on at a given time, more data is needed [7]. According to [8] a typical smart meter message size is about 100 bytes and the traffic sent to the data concentrator can range from 1600 to 2400 bytes per reading interval.

Authors of paper [9], proposed a method that models the

fault as a special load temporarily connected to the faulted node and performs a forward and backward load flow to calculate the fault current and also the voltage drops at the nodes that have measurements. During the fault localization process, the proposed algorithm massifies the fault at each node one at a time, it performs a forward and backward load flow updating the fault current in each iteration and calculates an error index and for each bar.

The method proposed in [10], incorporates all measured values into calculations through a state estimation-based algorithm to improve accuracy and performance in the face of uncertainties and imperfections. The algorithm first evaluates the input data and eliminates gross measurement errors. It then assumes the fault at each of the suspected nodes and performs a set of state estimations to find the node closest to the fault location. By incorporating all available measurement resources into the calculations, the method significantly reduces the impact of random measurement errors and load estimation uncertainties and improves the accuracy and robustness of the results. However, compared to previous load flow-based methods, such as the method proposed in [9], the state estimation-based method requires more computational time.

The method proposed in [11] presents two types of fault location methods for radial and non-radial overhead distribution systems that are based on solving a set of equations derived from the elements of the bus impedance matrix, expressing the substation voltage and current as a function of fault location and fault resistance. The algorithm must be applied to each line section, providing a list of possible fault locations and requires some additional information, such as customer fault reports or fault indication signs, in order to find the most likely location.

As [12], based upon the voltage drop, the methods must estimate the fault resistance or current during or before performing the fault localization process. In addition, it is necessary to examine all the nodes in the system to find out which is defective.

In [13] the authors propose a fault localization algorithm for distribution networks with Distributed Generation (DG), which considers the fault localization process as an optimization problem where the fault localization and resistance are unknown variables. They developed a genetic algorithm-based technique to obtain an optimal solution to this problem. The proposed method ignores the fault estimation step and searches a limited number of nodes to find out what is faulting.

Considering the techniques found in the literature, it is pertinent to evaluate that there is no consent on an approach to be used for the localization of SDVVs disturbances, besides, in most of them, the method does not select the amount of meters installed in the network in an optimized way, a fundamental task, since the larger the amount of meters sending information the higher the cost of implementing the method. Thus, it is relevant to analyze a method that can be applied in a real-time monitoring system.

To localize the disturbance, it is ideal to have meters at all points, but this strategy leads to the generation of a large volume of data both for traffic and for analysis. Thus, the objective of this work is to analyze an optimization method that selects the meters capable of detecting the effects of a fault in the system and generate the necessary data for localization. The optimization algorithm will be evaluated starting from a localization algorithm that is proposed in the article [14]. The method uses the data to group the meters through a clustering algorithm generating a region, then classifies these regions as either with or without fault.

In the optimization process a Genetic Algorithm (GA) is applied to select the quantity and the optimal location of meters belonging to the Smart Grid, reducing the quantity of meters that will compose a region. With this, this work brings contributed to a reduction in the cost and volume of collected data without losing the quality of the method.

In the sequence this study is described using the following division. We talk about the basic Genetic Algorithm in section II. Then we present the methodology used in the optimization process in section III. Section IV shows how the short circuit simulations were implemented to obtain the database. In section V performance comparisons and result discussions are displayed. Finally in the section VI the work is concluded presenting some perspectives.

II. Genetic Algorithm

The Genetic Algorithm (GA) is one of the first population-based stochastic algorithms proposed in history [15]. The main operators of GA are selection, crossover and mutation (Figure 1).
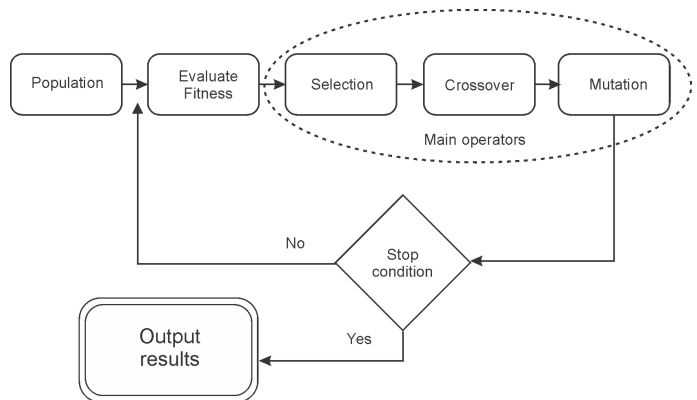

Fig. 1.   Basic structure of a genetic algorithm.

GA is a population-based algorithm. Each solution corresponds to a chromosome and each parameter represents a gene. GA evaluates the fitness of each individual in the population using a fitness function (objective) [16]. To improve the poor solutions, the best solutions are chosen randomly with a selection mechanism (e.g. roulette wheel). This operator is more likely to choose the best solutions because the probability is proportional to the fitness (objective value). What increases the avoidance of great locations is the probability of choosing bad solutions as well. This means that if good solutions get stuck in a local solution, they can be extracted with other solutions [17].

III. Methodology

The methodology used presents the task of selecting meters, which consists of obtaining the smallest number of meters

and the installation location in the electrical system, ensuring the observability of the system and good effectiveness of the proposed method of locating voltage sag events. This task is fundamental, because it is intended reducing the volume of data that will transit on the network and that will compose the data of the Big Data analysis. To this end, a GA will be used whose evaluation function is to set the balance between the accuracy of the localization process and a weight referring to the quantity of selected nodes.

The implemented algorithm considers the typical phases of GA: Initialization (Population), Objective-Function or Fitness, Selection, Crossover and Mutation, in this session the Objective-Function phase will be detailed and briefly the Initialization phase. For the other phases the default configuration of the library used will be used, where Selection uses the tournament selection strategy. Concerning the crossover phase, the SBX (simulated binary crossover operator) was used and for mutation phase, a polynomial mutation operator was used [18].

A. Initialization (Population)

During initialization the population creation is done randomly where each individual is represented by a vector, eachbar in the system is represented by a cell in this vector, withvalues between 0 and 1, where 0 indicates that the bar wasnot selected and 1 that the bar was selected as shown in the Figure 2, the size of the vector will depend on the size of thenetwork.

Individual (Vector of the selected bars)

Bars	802	806	808	812	814	816	824	828	830	832	834	836	840	842	844	846	848	850	852	854	858	860	862
Individual	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0

Fig. 2. Vector of the selected bars.

B. Objective-Function (Evaluate Fitness)

The objective function uses the localization strategy which is divided into two phases. The first phase is responsible for obtaining and analyzing the generated clusters, the clustering algorithm is responsible for making inferences about the propagation of the voltage sags event in the distribution network, dividing the unlabeled input data into groups according to the similarity between the disturbances evidenced in the network and independent of the system topology. The second phase corresponds to the application of the decision rule-based system, which is responsible for analyzing the characteristics of each of the generated clusters and identifying which of them provides the region of location of the voltage sinking disturbance (more detail of the phases is described in [14]), With the following changes:

- Processed data is only from the selected bars
- At the end of the second phase you get the accuracy subtracted by a weight (*p*):

$$p = bs\left(\frac{1}{tb}\right) \tag{1}$$

where *bs* is the number of selected bars and *tb* is the total number of bars, i.e. the size of the vector.

1) Weight of the Objective Function: The need to use the weight in the objective function is tostrike a balance between the number of bars and the location accuracy, eliminating situations where two or more individualshave the best accuracy found but with very different numbers of selected bars.

The whole process is repeated as shown in the Figure 3 until it reaches a certain number of generations (the stopping condition of the algorithm used), ending with the indication of how many and which bars are to be selected.
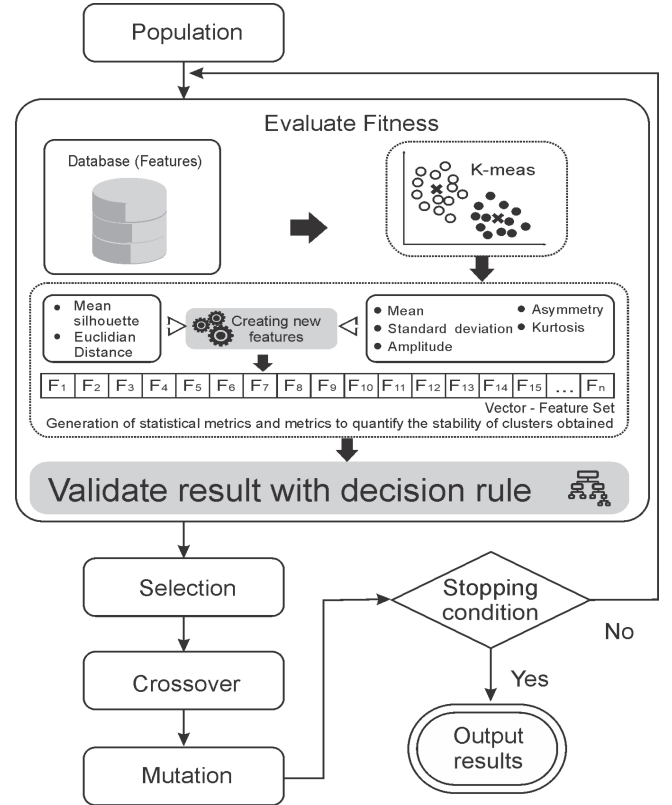

Fig. 3. Basic structure of a genetic algorithm with a function to evaluate proposed fitness.

IV. SIMULATIONS

In this section, the simulations of how the database, the method of locating the source of voltage sags, and finally the optimization process were performed.

A. Simulation of Voltage Sags

In order to evaluate the proposed algorithm, it was necessary to model and simulate the IEEE 34-Bus Test Feeder [19] through the ATP (Alternative Transients Program) software. Having modeled this system (Figure 4), simulations were performed by changing the system's behavior to generate voltage sags with different durations and magnitudes. Thus, it was possible to obtain a database to validate the proposed algorithm.

It is worth mentioning that all voltage sags were generated as a result of three-phase short-circuits with a fault duration of 4 cycles, different fault impedances (10, 20, 30, 40 e 50 Ω), with an incidence angle of 0.
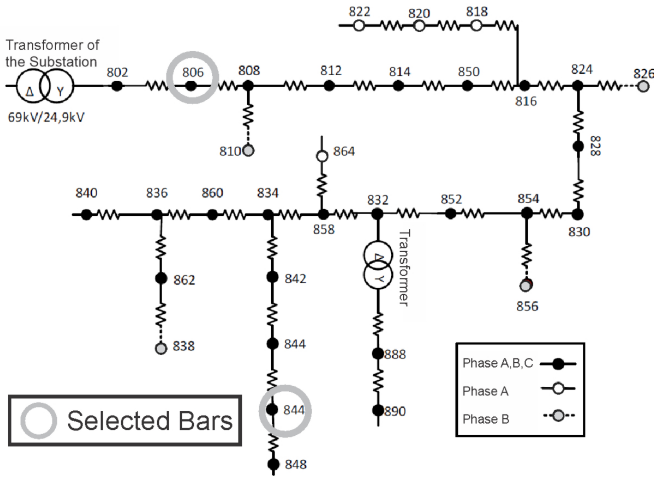
Fig. 4. Indication of the selected bars in the IEEE 34 Node Test System diagram, Adapted from [20]

The short circuit is simulated on each of the 23 three-phase bars of the system, considering a sampling rate of 256 points/cycle. This gives a total of 115 simulations. After each simulation the measured voltage values in each node of the system are stored. Thus, an information bank necessary to test the proposed algorithm is obtained. The process of building the database is carried out by means of communication between ATP® and MATLAB®.

B. Optimization by Genetic Algorithm

The optimization simulation was performed with each of the 115 test scenarios, even if the error bar is not one of the selected bars, Since any and every fault incident in the system is propagated through the network, the chosen bars expressed reflection values of this fault and this data from the chosen bars will be analyzed.

The implementations of the programming routines were developed in python language using the library *Pymoo* [18] for GA with the following settings: population size 30, random initialization, elimination of duplicate individuals and stopping condition 50 generations. For the evaluation function, the clustering algorithm used was the *k-means* with k equal to 3, the choice of this configuration was based on the performance of the method presented in [14], the library *sklearn* [21] was used to perform the k-means algorithm and obtain the accuracy value.

V. RESULTS OF THE PROPOSED METHODOLOGY

The following will discuss the results obtained using the entire proposed approach for locating voltage sag disturbances in the IEEE 34-Bus Test Feeder system. The efficiency of this methodology is influenced by the clustering algorithm used and the number of clusters. The fragmentation of the distribution network into several clusters leads to a greater segmentation of the search space and, consequently, can influence the effectiveness of the rules-based system. On the other hand, in a scenario where the distribution system is divided into a

small number of clusters, each cluster aggregates a significant amount of bars. In this situation the output of the decision-based algorithm will point to an extensive region of origin, composed of several nodes.

A. Result of Optimization

At this point , the results obtained will be discussed using the optimization proposal (GA) to select among all bars existing in the system a reduced amount of bars and identifying strategic locations of these bars in order to reduce the volume of data that will be transmitted and analyzed. It is important to highlight that the clustering algorithm executed in the objective function is the k-means with k equal to 3 due to its better performance displayed in [14].

The Table I shows the best results of each generation (best individual) with emphasis on the generations 12, 27 and 45 that represent the best configurations in relation to the quantity and positions of the bars chosen, as the choice criteria are tied in the three generations having a result of 0.818 for the objective function (evaluation), 99.17% of accuracy and 0.17391 of weight, GA produced the first result found in the case of the 12 generation.

Still observing the Table I we can verify that the position of the selected bars has influence in the result, as an example the 5th generation despite having 4 bars selected the accuracy value is 89.55% which is quite below the result produced by GA which was 99.17% that also has 4 bars selected. To better visualize the GA result the Figures 5, 6 and 7 are graphical representations adapted to the IEEE 34-bus test feeder model of satisfactory results.
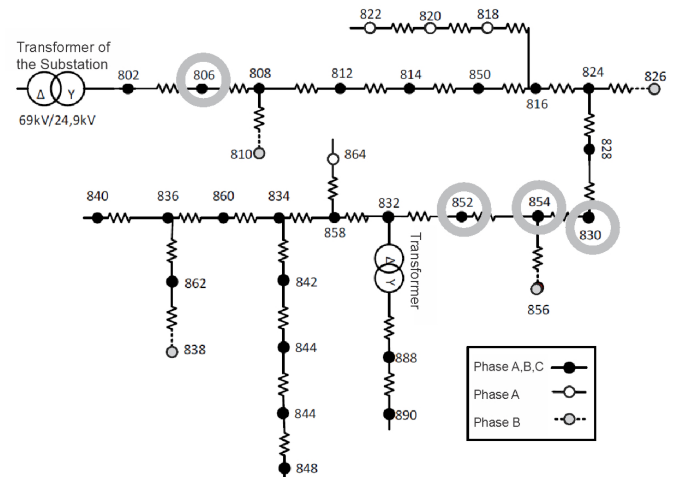

Fig. 5. Graphical representation of the bars selected in the optimization process 12th-generation

In the Table II it is possible to visualize the best performance of the proposed non-optimized method K-means with k=3 and PART presented in [14] with 99.42% of accuracy but with use of the data of the 23 three-phase bars of the network. The accuracy of the optimized methodology has a very small decrease of only 0.26%, totaling 99.16% of accuracy, but with the use of data from only 4 bars, which is advantageous since

TABLE I
BEST RESULTS PER GENERATION

Generation	Bars	Return OF	Accuracy (%)	Weight
1	8	0,600	94,77%	0,34783
2	8	0,635	98,26%	0,34783
3	6	0,686	94,70%	0,26087
4	6	0,677	93,79%	0,26087
5	4	0,722	89,55%	0,17391
6	4	0,765	93,94%	0,17391
7	4	0,757	93,11%	0,17391
8	4	0,783	95,68%	0,17391
9	4	0,809	98,26%	0,17391
10	4	0,809	98,33%	0,17391
11	4	0,781	95,45%	0,17391
12	4	0,818	99,17%	0,17391
13	4	0,800	97,42%	0,17391
14	4	0,809	98,26%	0,17391
15	4	0,809	98,26%	0,17391
16	4	0,800	97,35%	0,17391
17	4	0,800	97,35%	0,17391
18	4	0,800	97,42%	0,17391
19	4	0,792	96,59%	0,17391
20	4	0,799	97,27%	0,17391
21	4	0,809	98,26%	0,17391
22	4	0,800	97,42%	0,17391
23	4	0,800	97,42%	0,17391
24	4	0,800	97,42%	0,17391
25	4	0,793	96,67%	0,17391
26	4	0,799	97,27%	0,17391
27	4	0,818	99,17%	0,17391
28	4	0,775	94,92%	0,17391
29	4	0,791	96,52%	0,17391
30	4	0,801	97,50%	0,17391
31	4	0,800	97,42%	0,17391
32	4	0,800	97,42%	0,17391
33	4	0,801	97,50%	0,17391
34	4	0,817	99,09%	0,17391
35	4	0,800	97,35%	0,17391
36	4	0,800	97,35%	0,17391
37	4	0,792	96,59%	0,17391
38	4	0,790	96,44%	0,17391
39	4	0,783	95,68%	0,17391
40	4	0,809	98,33%	0,17391
41	4	0,790	96,44%	0,17391
42	4	0,800	97,42%	0,17391
43	4	0,817	99,09%	0,17391
44	4	0,791	96,52%	0,17391
45	4	0,818	99,17%	0,17391
46	4	0,817	99,09%	0,17391
47	4	0,790	96,44%	0,17391
48	4	0,809	98,33%	0,17391
49	4	0,801	97,50%	0,17391
50	4	0,801	97,50%	0,17391

there is no need to send and analyze data from 19 bars of the system.

TABLE II
ACCURACY OF THE NON-OPTIMIZED AND OPTIMIZED METHOD

Algorithm	Hit Rate (%)	Bus
K-means with k=3 and PART	99,42	23
Genetic Algorithm	99,16	4

As the voltage sags are short duration system failures, the speed of data interpretation is an important factor to be considered, the optimization process has a small loss of accuracy but a much smaller volume of data to be analyzed,
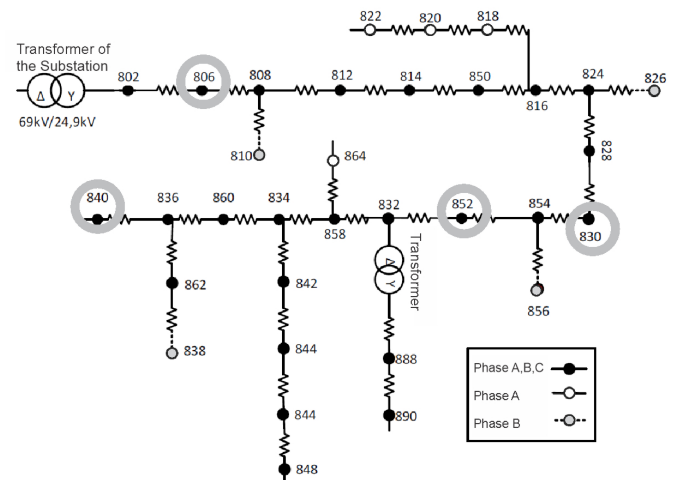

Fig. 6. Graphical representation of the bars selected in the optimization process 27th-generation
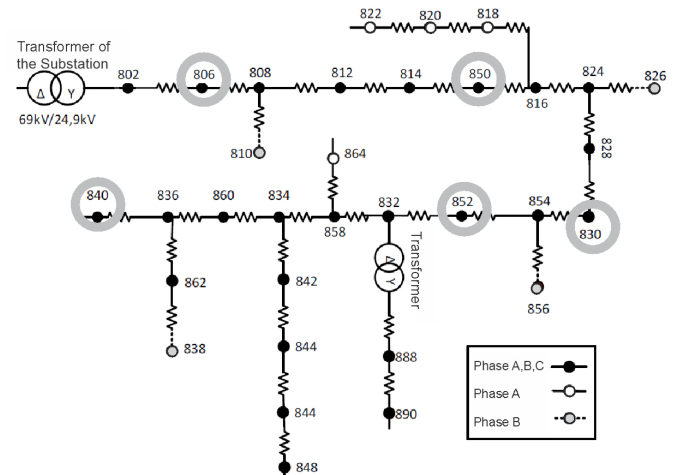

Fig. 7. Graphical representation of the bars selected in the optimization process 45 th-generation

which gives credence to a greater speed in the analysis of the information and consequently in the speed of finding the location of the fault.

VI. CONCLUSION

In this work, an optimization algorithm (GA) was applied to select the best measurement locations in order to decrease the amount of data for both network traffic and analysis and maximize the accuracy of the localization method. A comparison between the optimized and the non-optimized method was performed and the loss in accuracy was very small with a difference of only 0.26%, which is considered satisfactory taking into consideration that it was possible to reduce the volume of data that will be transmitted and analyzed by the methodology with the emission of data from only 4 bars, which was discarded from the localization process 19 bars of the system compared to the non-optimized method that uses 23 three-phase bars. For future studies, it is intended

to analyze the common method larger bar system and verify the scalability of the optimized methodology.

REFERENCES

[1] J. Miranda Filho, J. M. de Carvalho Filho, A. P. Paiva, P. V. G. de Souza, and S. Tomasin, "A pca-based approach for substation clustering for voltage sag studies in the brazilian new energy context," *Electric Power Systems Research*, vol. 136, pp. 31–42, 2016, DOI: 10.1016/j.epsr.2016.02.012.

[2] S. M. Hietpas and M. Naden, "Automatic voltage regulator using an ac voltage-voltage converter," *IEEE Transactions on Industry Applications*, vol. 36, no. 1, pp. 33–38, 2000, DOI: 10.1109/28.821793.

[3] A. Safdarian, M. Fotuhi-Firuzabad, and M. Lehtonen, "A general framework for voltage sag performance analysis of distribution networks," *Energies*, vol. 12, no. 14, p. 2824, 2019, DOI: 10.3390/en12142824.

[4] Z. Galijasevic and A. Abur, "Fault location using voltage measurements," *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 441–445, 2002, DOI: 10.1109/61.997915.

[5] G. Chang, J. Chao, S. Chu, and C. Chen, "A new procedure for tracking the source location," in *Power Engineering Society General Meeting, 2007. IEEE*. IEEE, 2007, pp. 1–4, DOI: 10.1109/PES.2007.385828.

[6] H. M. Safhi, B. Frikh, and B. Ouhbi, "Energy load forecasting in big data context," in *2020 5th International Conference on Renewable Energies for Developing Countries (REDEC)*. IEEE, 2020, pp. 1–6, DOI: 10.1109/REDEC49234.2020.9163901.

[7] N. Andreadou, E. Kotsakis, and M. Masera, "Smart meter traffic in a real lv distribution network," *Energies*, vol. 11, no. 5, p. 1156, 2018, DOI: 10.3390/en11051156.

[8] M. Kuzlu and M. Pipattanasomporn, "Assessment of communication technologies and network requirements for different smart grid applications," in *2013 IEEE PES innovative smart grid technologies conference (ISGT)*. IEEE, 2013, pp. 1–6, DOI: 10.1109/ISGT.2013.6497873.

[9] R. A. F. Pereira, L. G. W. da Silva, M. Kezunovic, and J. R. S. Mantovani, "Improved fault location on distribution feeders based on matching during-fault voltage sags," *IEEE Transactions on Power Delivery*, vol. 24, no. 2, pp. 852–862, 2009, DOI: 10.1109/TPWRD.2009.2014480.

[10] S. Jamali and A. Bahmanyar, "A new fault location method for distribution networks using sparse measurements," *International Journal of Electrical Power & Energy Systems*, vol. 81, pp. 459–468, 2016, DOI: 10.1016/j.ijepes.2016.02.046.

[11] S. Jamali, A. Bahmanyar, and H. Borhani-Bahabadi, "A fast and accurate fault location method for distribution networks with dg using genetic algorithms," in *2015 Smart Grid Conference (SGC)*. IEEE, 2015, pp. 110–114, DOI: 10.1109/SGC.2015.7857419.

[12] A. Bahmanyar, S. Jamali, A. Estebsari, and E. Bompard, "A comparison framework for distribution system outage and fault location methods," *Electric Power Systems Research*, vol. 145, pp. 19–34, 2017, DOI: 10.1016/j.epsr.2016.12.018.

[13] S. A. M. Javadian, M.-R. Haghifam, and N. Rezaei, "A fault location and protection scheme for distribution systems in presence of dg using mlp neural networks," in *2009 IEEE Power & Energy Society General Meeting*. IEEE, 2009, pp. 1–8, DOI: 10.1109/PES.2009.5275863.

[14] J. Lima Filho, F. A. da Silva Borges, R. d. A. L. Rabelo, I. S. Silva, R. P. T. Junior, and A. O. de Carvalho Filho, "Methods for voltage sag source location by cluster algorithm and decision rule labeling with a comparative approach of k-means and dbscan clustering algorithms," in *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, 2020, pp. 1–8, DOI: 10.23919/SpliTech49282.2020.9243823.

[15] S. Mirjalili, "Genetic algorithm," in *Evolutionary algorithms and neural networks*. Springer, 2019, pp. 43–55.

[16] R. Arulmozhiyal and K. Baskaran, "Speed control of induction motor using fuzzy pi and optimized using ga," *International Journal of Recent Trends in Engineering*, vol. 2, no. 5, p. 43, 2009.

[17] B. S. Chaudhari and M. Zennaro, *LPWAN Technologies for IoT and M2M Applications*. Academic Press, 2020.

[18] J. Blank and K. Deb, "pymoo: Multi-objective optimization in python," *IEEE Access*, vol. 8, pp. 89 497–89 509, 2020, DOI: 10.1109/ACCESS.2020.2990567.

[19] R. D. T. Feeders, "Ieee distribution planning working group report," *IEEE Trans. Power Syst*, vol. 6, no. 3, pp. 975–985, 1991.

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

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

José Carlos Filho received the B. Sc. degree in Information System from the Faculty of Business Activities of Teresian (FAETE) in 2013 and a Master's degree in computer science from the Federal University of Piauí, Brazil, in 2019. He is currently a Ph.D. student in computer science from the Federal University of Piauí and professor at the Estácio de Sá University, Teresina, Piauí, Brazil.

Fábbio A. S. Borges was born in Teresina, Brazil, in 1989. He received the graduation degree in computer science from the Unified Teaching Center of Teresina, Teresina, in 2010, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of São Paulo, São Carlos, Brazil, in 2013 and 2017, respectively. Currently, he is an Adjunct Professor with the State University of Piaui. His research interests include the field of Smart Grids, power quality, and intelligent systems.

Ricardo A. L. Rabelo received the B.Sc. degree in computer science from the Federal University of Piauí, Brazil, in 2005, and the Ph.D. degree in power systems from the São Carlos Engineering School, University of São Paulo, Brazil, in 2010. He is currently a Professor with the Federal University of Piauí (UFPI), Brazil. He has authored or coauthored over 150 papers in refereed international journals and conferences. His research interests include Smart Grid, Internet of Things, intelligent systems (computational intelligence) and optimization (linear and nonlinear programming).

Ivan Saraiva Silva received a Ph.D. in Computing from Pierre et Marie Curie (Paris VI) in 1995. He is currently Associate Professor III at the Federal University of Piauí (UFPI).

Antonio Oseas de C. Filho received a Ph.D. in Electrical Engineering at Federal University of Maranhão - Brazil in 2016. Currently, he is a professor at the Federal University of Piaui (UFPI). His research interests include medical image processing, machine learning and deep learning.