# Fault Section Identification in Distribution Networks with DFIG and PMSG Generators Using Current Transients

J. C. Peqqueña Suni, M. G. S. P. Paredes, M. V. de Paula, E. Ruppert Filho, and J. A. Martinez-Velasco

***Abstract*—This paper presents a methodology for fault section identification (FSI) in distribution networks with embedded wind power generation. The phase currents are measured only at the distribution substation (DS), using a waveform window of two cycles (one before and one after the fault detection). The proposed approach is divided into two stages: the first stage, Fault Identification (FI), aims to identify whether a short-circuit fault lies on a main feeder or one of the branches effectively addressing the challenge of multiple fault locations that may arise when several branches correspond to the estimated fault point; the second stage, Fault Location (FL), estimates the distance between the DS and the fault location. The algorithm employs discrete wavelet transform (DWT) in combination with artificial neural networks (ANNs). Energy and Relative Energy Entropy, both in per unit ($E_{PU}$ and $REE_{PU}$), are proposed and calculated from DWT decomposition, with regularization indexes applied to $E_{PU}$ and $REE_{PU}$. These indexes serve as input to multi-layer ANN models, which work as classifiers for FI and predictors for FL. Various fault scenarios with different fault inception angle, fault type, fault resistance and fault location are simulated using MATLAB® software and the IEEE 34-node benchmark feeder as test system. The results demonstrate that the proposed methodology performs effectively the FSI task, achieving an accuracy of up to 95% for FI and a maximum error of 5.2% for FL.**

**Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/9488**

***Index Terms*— Distribution Network, Fault Section Identification Machine Learning, Discrete Wavelet Transform, Wind Power.**

## I. Introduction

ELECTRIC power systems are subject to different events, such as severe weather conditions, contact with animals, falling tree branches or trees themselves, human accidents, or equipment malfunction. The short-circuits faults in distribution networks are considered the leading causes of supply interruption to system users and are responsible for power quality disturbances [1].

Traditionally, protection systems and fault location methods for distribution networks have assumed that the network is single-source and has a radial topology. However, these characteristics change due to the presence of bidirectional power flows caused by the insertion of distributed generation (DG); this changes the way these systems are operated and adds more complexity to the problem of identifying the faulted section [2], [3]. Some methodologies that account for the effect of integrating DG into the distribution network, as well as the challenges in the protection system [4], [5].

Fault diagnosis is based on machine learning techniques, such as Artificial Neural Networks (ANNs), Support Vector Machine (SVM) [6], bagged tree and Adaptive Boosting (AdaBoost) [7]. ANN-based architectures can be used to solve problems related to fault identification and fault location [8]-[13]. To accurately detect a fault and then locate the physical section in which the fault occurs, some techniques use measurements at the distribution substation (DS) and DG points; the physical location of the fault is identified by comparing the results with different ANN architectures from representative data sets [14]. Some techniques use Discrete Wavelet Transform (DWT), in combination with ANNs, as a tool for processing fault signals and feature extraction [10], [15], [16]. Most of those techniques analyze the effect of DG using a simplified synchronous machine model and assuming a direct interconnection with the distribution network, without considering an inverter-based interface [17], [18]. Other techniques consider the two types of technologies: inverter non-interfaced and inverter-interfaced [19].

The inherent characteristics of primary distribution networks still impose limitations on fault section identification (FSI) techniques. Actually FSI, being the fault located in the main feeder or in a branch, remains an unsolved challenge, particularly when measurements are confined to the DS [20], [21].

Economic factors for electrical utilities justify the use of measuring and supervision equipment installed only in the DS, which imposes restrictions on the remote acquisition of data (voltage, current) from the distribution network [22]. The challenges increase due to uncertainties in the input data, such as conductor impedances and the loading of each branch,

The associate editor coordinating the review of this manuscript and approving it for publication was Gabriel Pinto (*Corresponding author: Juan Carlos Peqqueña Suni*).

This research was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Juan Carlos Peqqueña Suni is with the Universidade Federal do Ceará, Sobral, Brazil (e-mail: jcarlosps@ufc.br).

M. G. S. P. Paredes, M. V. de Paula, and E. R. Filho are with the Universidade Estadual de Campinas, Brazil (e-mails: marinagp@unicamp.br, mvpaula@unicamp.br, and ruppert@unicamp.br).

J. A. Martinez-Velasco is retired, formerly with the Universitat Politecnica de Catalunya, Spain (e-mail: juan-antonio.martinez@upc.edu).

besides to inherent errors due to the fault impedance and the system load level at the time the fault occurs.

The main contributions of this paper can be summarized as follows: (i) the development of a non-iterative predictive method for FSI in distribution networks with wind power generation, considering inverter-based interfaces for DFIG and PMSG machines, (ii) the use of instantaneous phase currents measurements monitored only at the DS (2 cycles of 60 Hz), (iii) the application of time-frequency techniques aiming to capture transient information in the current signal during faults, and (iv) the use of regularization indexes on input samples for modular ANN-based algorithms.

This paper has been organized as follows, Section II provides some details about the test system used, the proposed FSI methodology is described in Section III, while simulation results are presented and discussed in Section IV, a brief discussion final is presented in section V, finally, Section VI summarizes the main conclusions.

## II. Test System

Fig. 1 shows the diagram of the test system used in this paper. It is based on the IEEE 34-node test system [23] with a wind power generator connected to node 844. Two types of models are considered for representing the wind generator: a doubly fed induction generator (DFIG) and a permanent-magnet synchronous generator (PMSG). Some details about these models are provided below.

### *A. DFIG-based Wind Turbine Generator Model*

The model of an ideal voltage source behind a sub-transient reactance is not an accurate representation of a wind power generator; a better representation must include the modeling of electronic power converters. When a short circuit occurs in the network, the behavior of these types of generators is not similar to that of a conventional synchronous generator. The model used for representing the DFIG model in this paper was developed and implemented in Simulink® by Richard Gagnon [24]. This model consists of a wound rotor induction generator with rotor- and grid-side converters, both using IGBT (Insulated Gate Bipolar Transistor) technology.

DFIG-based generators are sensitive to high induction currents caused by network faults. The most common system to avoid any risk is the crowbar system. When wind turbine generators must remain connected during faults, the crowbar resistance must be high enough to ensure the lowest reactive power consumption [25]. The above DFIG model does not have a crowbar system; consequently, the model is inaccurate when a short-circuit fault occurs in the network and implementing a crowbar system in the DFIG model, between the rotor windings and the rotor-side converter, becomes a requirement. Fig. 2 shows the model implemented in this paper: the normally open-circuit breaker (CB2) will close when the rotor currents reach a high value and the energy will be dissipated through the inserted bank of resistors; at the same instant, the normally closed-circuit breaker (CB1) will open, thus disconnecting the rotor-side converter from the rotor circuit for its protection [26].

The contribution of the DFIG model to short-circuit currents is greater when the crowbar system is not included compared to when it is included [27]. This also influences the contribution to the substation's short-circuit current, as the crowbar system resistor alters the equivalent impedance seen from the substation. Consequently, the methodology proposed in this paper enhances the accuracy of the DFIG model by including the crowbar system.

### *B. PMSG-based Wind Turbine Generator Model*

The PMSG-based generator model used in this paper was developed and implemented in Simulink® by Richard Gagnon and Jacques Brochu [24]. The wind power generator consists of

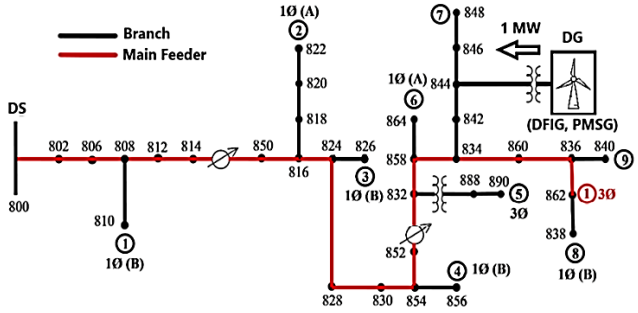



Fig. 1. IEEE 34-node test system [23] and DG integration point.

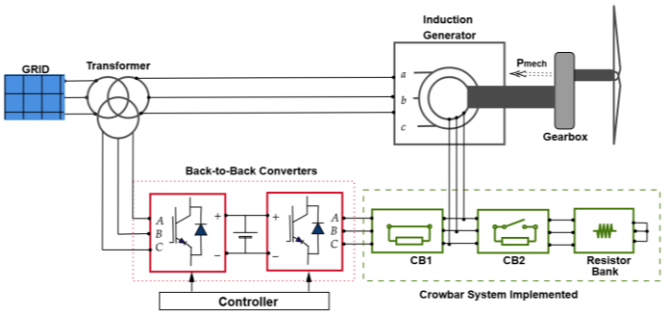



Fig. 2. Crowbar system implementation in the wind turbine DFIG-based [24],[25].

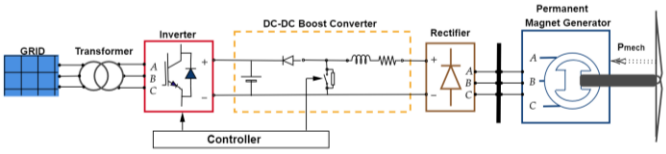



Fig. 3. Wind turbine PMSG-based representation [24].

a PMSG machine connected to a diode rectifier, a DC-DC PWM boost converter, and an IGBT DC-CA PWM inverter, as shown in Fig. 3. This technology can extract maximum energy from the wind at low wind speeds by optimizing turbine speed. In this paper, the wind speed is kept constant at 15 m/s and the DC-DC converter control system is used to maintain the generator speed at 1 p.u [24].

## III. Applied FSI Methodology

Fig. 4 summarizes the proposed FSI methodology, which is detailed below.

### *A. Data Generation*

Two data sets, obtained respectively with the DFIG and

PMSG models, were created from short-circuit fault simulations carried out with MATLAB® Simulink® software [24]. A simulation step of 10 µs for a sampling frequency of 100 kHz, 1667 samples per 60 Hz cycle, was used; this is sufficient to cover the high frequency transients produced by a short-circuit fault. In order to reduce the size of the input datasets, the sampling frequency was then decreased to 7.680 kHz with 128 samples per cycle.

Short-circuit faults are classified into four main categories: phase to ground faults (A-G, B-G and C-G), phase-phase faults (AB, BC, CA), phase-phase to ground faults (AB-G, BC-G and CA-G) and three-phase faults (ABC). The faults were distributed across 24 equally spaced locations (2.5 km) starting at 1 km along the 57.5 km three-phase main feeder (MF); 11 equally spaced locations (1.5 km) along the single-phase branch 2 (B2); 11 equally spaced locations (5%) were also considered for the remaining single-phase (B1, B3, B4, B6 and B8) and three-phase (B7 and B9) branches. B5 was not considered because it is a secondary voltage level branch.

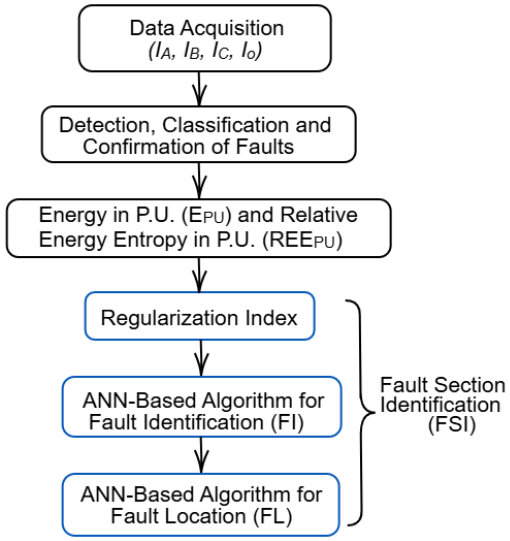



Fig. 4. FSI methodology scheme.

For each location, fault resistances of 0, 15, 25, 35 and 50 Ω were applied together with nine fault onset angles every 11.25° between 0° and 90°. In total, there were 10800 fault cases (10 fault types) simulated for the MF, 495 cases for single-phase branches (1 fault type) and 4950 cases for three-phase branches (10 fault types) for each system model.

Wind generators (1.5 MW for DFIG and 2 MW for PMSG) were connected to the test system at node 844, limiting in both cases the power injected by the wind turbine to 1 MW. For each simulated fault case, data of instantaneous phase and homopolar currents ($I_A$, $I_B$, $I_C$, $I_0$) were stored, using current values monitored and measured only at DS.

### *B. Detection, Classification and Confirmation of Faults*

The procedure for detecting, classifying and confirming short-circuit faults is based on the comparison between instantaneous current values (samples) measured in real time and values of the same current stored in the device's memory up to one 60 Hz cycle before (i.e. 128 samples). Since the imbalance is very evident between the three phases of the test system, the current thresholds were estimated by using the current value of the phase that has the highest load (in our case phase A). A value of 25% of the nominal current is selected as the threshold for each phase to determine the instant at which the fault begins. The classification of the fault type uses the faulty phase(s) current(s) as well as the homopolar current, the fault confirmation is done in the following samples [28], [29].

This method, called over-current [30], in combination with DWT and ANNs [31], [32] gains in simplicity and speed when compared to other machine learning methods that despite having great precision, demand a large amount of data for the training process and require high computational power [33], [34].

The three instantaneous phase currents ($I_A$, $I_B$, $I_C$) are continuously monitored and measured at the DS; their values are used to calculate the homopolar current ($I_0$) according to the following form:

$$\bar{I}_0 = \frac{\bar{I}_A+\bar{I}_B+\bar{I}_C}{3} \quad (1)$$

The calculated homopolar current will indicate whether the detected short-circuit fault is of the shunt/parallel type (with contact to earth) or the series type (without contact to earth). Fig. 5 shows an example of the proposed procedure for detection, classification and confirmation of a case that corresponds to a single-phase fault (A-G).

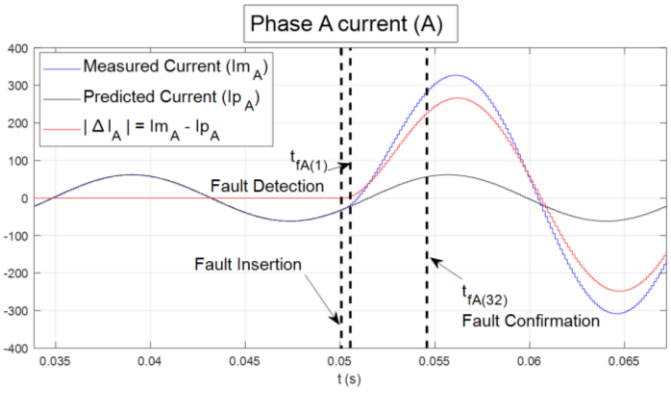



Fig. 5. Fault insertion, transient detection and fault confirmation on 31 following samples.

### *C. Feature Extraction*

The main goal when extracting relevant features from data is to reduce the amount of information from the original current waveforms. The feature extraction process consists of finding distinct parameters of the waveform, with significant information, which can represent the fundamental characteristics of the signal [10], [32].

For such purposes, this paper takes advantage of the DWT and the wavelet family Daubechies-4 (Db4) using the following multiresolution signal decomposition algorithm [35]:

$$DWT(m,n) = \frac{1}{\sqrt{2^m}}\sum_k f(k)\psi\left(\frac{n-k2^m}{2^m}\right) \quad (2)$$

where *f(k)* is the discrete signal represented as a function of its coefficients, $\psi$ is the mother wavelet, *m* and *n* are respectively

the time and the scale parameters, $k$ is both the discrete time and the number of coefficients in the DWT, $2^m$ is the variable for scaling, $k2^m$ is the variable for shifting and $1/\sqrt{2^m}$ is the energy normalization component to ensure the same scale as the mother wavelet [10], [35].

Due to the large amount of data processed, the level $j$ of decomposition of high frequencies generated by short-circuit faults was calculated using the following form [35]:

$$f_c = 2f_0 = \frac{f_{sam}}{2^{j+1}} \tag{3}$$

This saves time in processing by having less input data, discarding data at low frequencies, and always aiming at not losing precision results by increasing the speed of the analysis. For fundamental and sampling frequencies of $f_o$=60 Hz and $f_{sam}$=7680 Hz respectively we have:

$$j = \log_2\left(\frac{f_{sam}}{f_0}\right) - 2 = 5 \tag{4}$$

Table I shows the frequency bands up to the 5th used in this paper, where $a_j(n)$ and $d_j(n)$ are respectively the approximation and detail coefficients derived from the implementation of DWT using a multiresolution signal decomposition algorithm [35]. However, directly using these wavelet coefficients as input would result in a complex and time-consuming analysis. To overcome this, the entire spectrum of wavelet coefficients (2 cycles of 60 Hz with 256 samples, one cycle before and one cycle after fault detection) was used to calculate the energy in p.u. ($E_{PU}$) and the relative energy entropy in p.u. ($REE_{PU}$), as detailed below.

TABLE I
FREQUENCY BANDS OF WAVELET DECOMPOSITION LEVELS

| Level ($j$) | $a_j(n)$ Scale Filter (Hz) | $d_j(n)$ Wavelet Filter (Hz) |
|---|---|---|
| **1** | 0 - 1920 | 1920 - 3840 |
| **2** | 0 - 960 | 960 - 1920 |
| **3** | 0 - 480 | 480 - 960 |
| **4** | 0 - 240 | 240 - 480 |
| **5** | 0 - 120 | 120 - 240 |

*C.1. Energy in p.u. ($E_{PU}$)*

The energy of each wavelet detail coefficient ($d_j$) of the signal recorded at level $j$ and instant $k$ ($k$ = 1, 2, …, $N$), where $N$ is the number of coefficients in the $j$th level, is [10]:

$$E_{jk} = \left|d_j(k)\right|^2 \tag{5}$$

The total signal energy at level $j$:

$$E_j = \sum_{k=1}^{N} E_{jk} \tag{6}$$

The wavelet energy in p.u., ($E_{PU}$), is proposed as:

$$E_{PUjIp} = \frac{E_{jIp}}{E_{jIA}+E_{jIB}+E_{jIC}+E_{jI0}} \tag{7}$$

where the current $I_p \in \{I_A, I_B, I_C, I_0\}$.

*C.2. Relative Energy Entropy in p.u. ($REE_{PU}$)*

To obtain the $REE$ representing the energy distribution in each sample of the signal spectrum first calculate [10]:

$$P_{jk} = \frac{E_{jk}}{E_j} \tag{8}$$

The $REE$ for each level $j$, where the current $I_p \in \{I_A, I_B, I_C, I_0\}$, can be defined as:

$$REE_{jIp} = -\sum_{k=1}^{N} P_{jkIp}\, log\, P_{jkIp} \tag{9}$$

The $REE_{PU}$ is then calculated as follows:

$$REE_{PUjIp} = \frac{REE_{jIp}}{REE_{jIA}+REE_{jIB}+REE_{jIC}+REE_{jI0}} \tag{10}$$

Equations (7) and (10) form the basis for the feature extraction and selection process; the calculated $E_{PU}$ and $REE_{PU}$ are used as inputs to train and validate the ANNs.

*D. Proposed Regularization Indexes*

Normally, $E_{PU}$ of the faulty phases at all levels of wavelet decomposition should be higher than in the healthy phases, while $REE_{PU}$ in the faulty phases should be lower than that in the healthy phases. According to the results obtained, there are cases where this is not fulfilled at all levels, so the samples must be regularized to ensure a clear difference between the phases involved in the fault and the healthy ones.

After testing other mathematical operations, the ones that suited the behavior of our data samples were rooting and potentiation. Therefore, the criterion proposed was to estimate empirically $E_{PU}$ and $REE_{PU}$ regularization indexes and apply them to the five levels of wavelet decomposition with the ten types of faults. Table II shows the estimated regularization indexes, the rooting index ($m$) for $E_{PU}$ and the potentiation index ($n$) for $REE_{PU}$. The regularization indexes are only calculated for the $E_{PU}$ and $REE_{PU}$ of the faulty-phase currents.

Fig. 6 shows the application to an AB-G fault using the DFIG system for the level 1 of wavelet decomposition, corresponding

TABLE II
REGULARIZATION INDEX

| Test System | Rooting ($m$) $\sqrt[m]{E_{PUjI_p}}$ | Potentiation ($n$) $\left(REE_{PUjI_p}\right)^n$ |
|---|---|---|
| **DFIG** | 11 | 1.9 |
| **PMSG** | 10 | 2.5 |

## IV. SIMULATION RESULTS

*A. Fault Identification (FI) Stage*

The aim of the FI stage is to identify the branch/feeder in which the fault is located. The number of neurons of the ANN output layer depend on the type of fault and the test system topology: from the comparison of Tables III and IV, it is easy to deduce that the number of neurons of the output layer of each ANN is equal to the number of columns of Table III in which all the faulty phases are involved. Note that the number of input neurons (Input-Layer) is 40 for all ANNs.

The final ANN architecture (Input Layer - Hidden Layer - Output Layer) was based on the analysis of the performance of the confusion matrix and its accuracy. The accuracy was calculated according to the following form [36]:

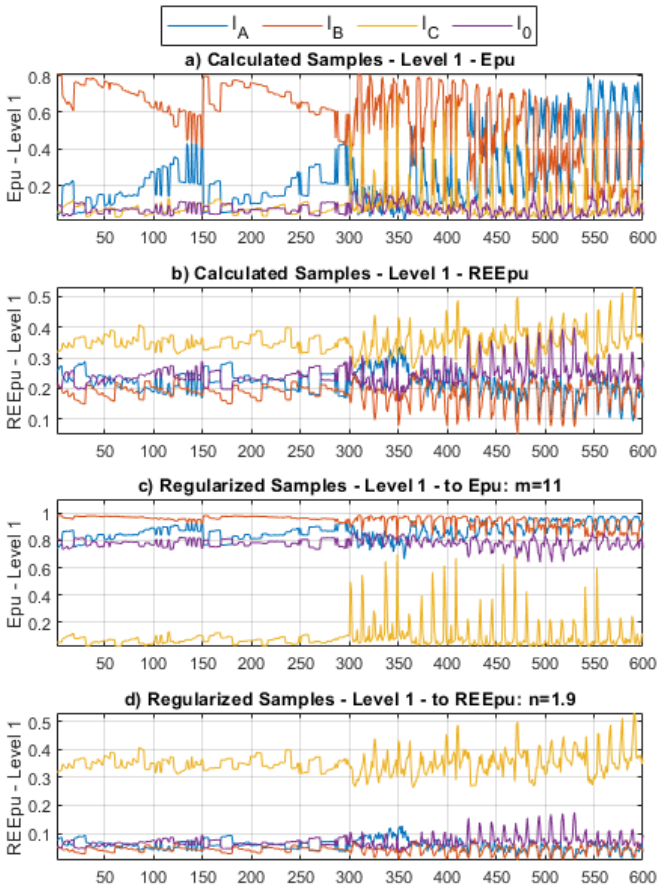



Fig. 6. Regularization of AB-G samples to level 1 - DFIG test system to 600 input samples (simulated cases for training) divided into 150 from branch 7, 150 from branch 9 and 300 from main feeder. Once samples were regularized (see part (c) and (d)), a clear difference between the phases involved in the short-circuit fault and those not involved can be observed. A similar procedure is applied to all types of faults and network configurations.

$$Acc\ (\%) = \frac{\sum_i (CM_{ii})}{total\ data} . 100 \qquad (11)$$

where $CM_{ii}$ are the diagonal elements of the confusion matrix (see Fig. 8).

According to Table IV, the lowest accuracy was obtained when analyzing B-G faults; it was respectively 73.5% and 64.6% for DFIG and PMSG test systems. The ANNs were less accurate with this fault type because there were 7 binary outputs which represented each branch and feeder where phase B should be considered. Remember that the measurements are made only at the DS, and the measured values account for measurements from single-phase, two-phase and three-phase faults without knowing the branch in which the fault did occur. A strong phase imbalance at the time of the fault can affect the input samples to ANNs, since they depend on the values of the instantaneous currents measured only at the DS; similar input samples can be generated from faults at different branches and feeder, negatively influencing the performance of the ANN.

Figs. 7 and 8 provide some information about the performance of this stage when used for analyzing B-G faults with the DFIG test system. Fig. 7 shows the ANN architecture with the best accuracy (73.5%) when analyzing B-G faults.

Fig. 8 shows the confusion matrix that corresponds to the training stage for B-G faults simulated with the DFIG test system. The matrix has 7 classes; each class represents a network branch (including the main feeder) where phase B exists (see Table III). According to Table V, 1200 B-G fault scenarios are simulated for training and 1890 for testing.

The matrix provides the detection accuracy (main diagonal) and the classification errors (off-diagonal elements) of each class/branch. The figure actually shows an 8x8 matrix, where row 8 and column 8 summarize the results from the other rows and columns. The first element of the diagonal, $a_{11}$, is 147 and represents the total number of correct hits for class 1 (Branch 1) out of a total of 150 real samples of this class; 3 hits were incorrectly classified as class 7. Element $a_{81}$ shows a 98% true positive rate of class/branch 1; the value just below represents the false positive rate (false alarm value) and is obviously 2%. Similar reasoning can be followed with the other classes. Element $a_{18}$ shows that the ANN accuracy for class/branch 1 is 96.1% (147 out of 153 predicted samples), while the value below is the complementary value in percentage terms; in this case 3.9%. Element $a_{88}$ is the ratio between the sum of all true positive hits of all classes and the total number of samples/cases, see equation (11); its value is 89.8%.

TABLE III
IEEE 34-NODE NETWORK TOPOLOGY

| Ø | MF | B1 | B2 | B3 | B4 | B6 | B7 | B8 | B9 |
|---|---|---|---|---|---|---|---|---|---|
| A | 1Ø | | 1Ø | | | 1Ø | 1Ø | | 1Ø |
| B | 1Ø | 1Ø | | 1Ø | 1Ø | | 1Ø | 1Ø | 1Ø |
| C | 1Ø | | | | | | 1Ø | | 1Ø |

TABLE IV
ANN-FI TEST PERFORMANCE

| Fault | DFIG | | PMSG | |
|---|---|---|---|---|
| | IL-HL-OL | Acc (%) | IL-HL-OL | Acc (%) |
| A-G | 40-17-5 | 79.1 | 40-19-5 | 68.1 |
| B-G | 40-23-7 | 73.5 | 40-15-7 | 64.6 |
| C-G | 40-15-3 | 85.1 | 40-12-3 | 80.0 |
| AB-G | 40-12-3 | 84.4 | 40-17-3 | 78.7 |
| BC-G | 40-12-3 | 85.9 | 40-12-3 | 79.4 |
| CA-G | 40-12-3 | 82.5 | 40-15-3 | 79.1 |
| AB | 40-05-3 | 94.8 | 40-09-3 | 81.0 |
| BC | 40-15-3 | 85.4 | 40-10-3 | 81.1 |
| CA | 40-12-3 | 85.8 | 40-08-3 | 83.1 |
| ABC | 40-08-3 | 83.4 | 40-05-3 | 82.2 |

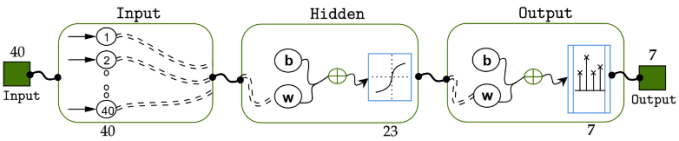



Fig. 7. ANN-FI architecture selected for B-G fault - DFIG test system.

*B. Fault Location (FL) Stage*

In this stage, ANN-based fault location algorithms have the distance between the DS and the fault location as output. The selected ANN architectures for FL were based on the best

performance of the ANNs using linear regression.

TABLE V
TOTAL SAMPLES USED FOR EACH FAULT TYPE

| Fault | N° cases - Training | N° cases - Testing |
|---|---|---|
| **A-G** | 900 | 1440 |
| **B-G** | 1200 | 1890 |
| **C-G** | 600 | 990 |
| **AB-G** | 600 | 990 |
| **BC-G** | 600 | 990 |
| **CA-G** | 600 | 990 |
| **AB** | 600 | 990 |
| **BC** | 600 | 990 |
| **CA** | 600 | 990 |
| **ABC** | 600 | 990 |

A linear regression is based on the following equation

$$y = ax + b \tag{12}$$

where $a$ is the slope, $b$ is the initial condition, $x$ is the target and $y$ is the output.

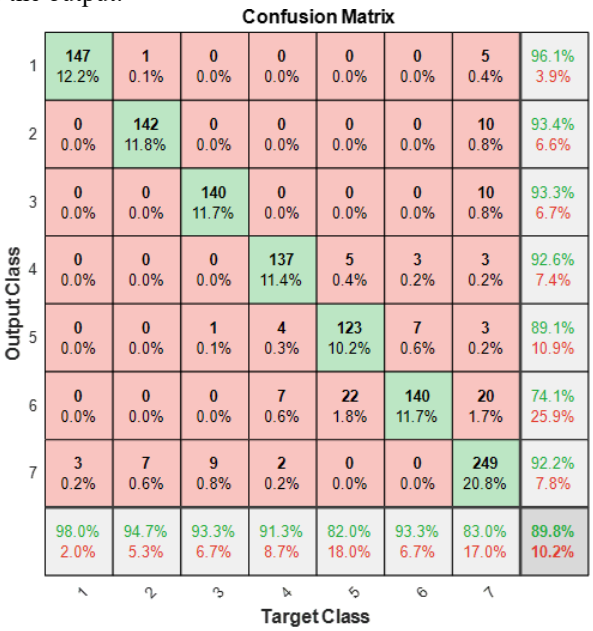



Fig. 8. ANN-FI confusion matrix for B-G fault – DFIG test system – Training stage [36].

The regression coefficient (correlation), $R_c$, can be used to describe the relationship between the ANN output values and the target values. A perfect performance of an ANN architecture would ideally achieve a value of $R_c$ equal to 1, with $a$ equal to 1 and $b$ equal to 0.

Fig. 9 depicts results corresponding to BC faults on the main feeder using the PMSG test system. The fit line is the linear regression of the test samples. The ideal dashed line, Y(Output) = T(Target), can be used for comparison [37].

Fig. 10 shows the $R_c$ values obtained with fault locations in the main feeder. The worst $R_c$ values were approximately 0.9 for the validation of the ANN architecture selected for analyzing two-phase ungrounded faults (AB, BC, CA) with PMSG. The selected ANN architectures have a single hidden layer that varies between 5 and 12 neurons. The criterium used for this stage was to select the smallest neural network configuration with good enough performance. The number of neurons of the input layer was kept fixed at 40 while the number of neurons of the output layer was always 1. The maximum number of epochs was 1000. The activation functions in the hidden and output layers were respectively *tansig* and *purelin*. The early stopping (*ES*) method was used to train all ANNs, ensuring that they are not "overtrained" (overfitting) and have a good ability to generalize. The test was performed with input samples from an independent database; that is, databases different from those used from training and validation.

Table VI summarizes performance results by providing the maximum estimated distance error in percentage, calculated using the following form:

$$\%Error = \frac{|Real\ Location - Estimated\ Location|}{Total\ Length} x100 \tag{13}$$

where *Total Length* is the distance between the terminal node of the main feeder/branch in which the fault occurs and the DS. Table VII provides information about the length of all sections (branch/feeder) of the test system; see Fig. 1.

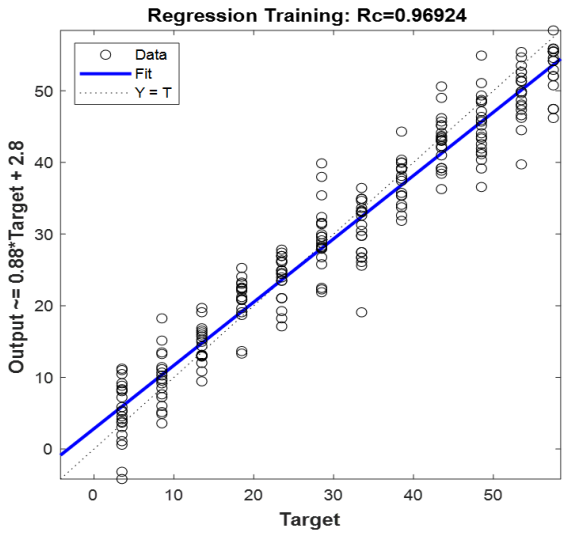



Fig. 9. Linear regression - ANN-FL architecture for BC faults on main feeder - PMSG test system.

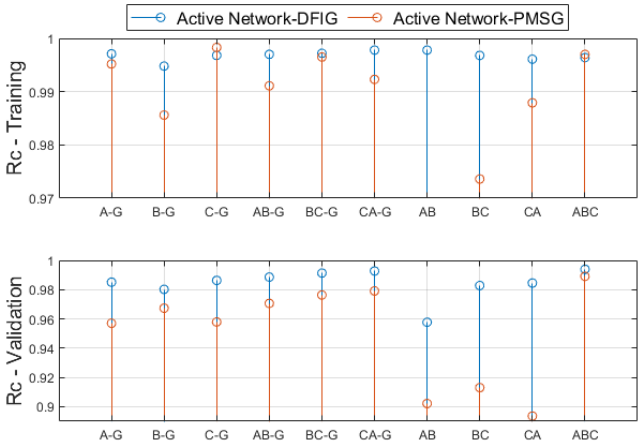



Fig. 10. ANN–FL $R_c$ – Three-phase main feeder – PMSG test system.

TABLE VI
ANN-FL TEST PERFORMANCE

| FL | Fault | DFIG Max. Error (%) | PMSG Max. Error (%) |
|---|---|---|---|
| **B1, B4** | B-G | 5.17 | 3.92 |
| **B7** | AB, BC, CA | 2.67 | 2.41 |
| **MF** | B-G, C-G, CA | 5.21 | 5.21 |

TABLE VII
TOTAL LENGTHS OF LINES (BRANCH/FEEDER)

| N° | Line section | Total length | |
|---|---|---|---|
| | | Line start (km) | Long length (km) |
| **1** | Main feeder 800-862 | 0 | 57.5000 |
| **1** | Branch 808-810 | 11.1374 | 1.7691 |
| **2** | Branch 816-822 | 31.7266 | 19.3853 |
| **3** | Branch 824-826 | 34.8386 | 1.7691 |
| **4** | Branch 854-856 | 41.4832 | 7.1110 |
| **5** | Branch 832-890 | 52.7120 | 3.2187 |
| **6** | Branch 858-864 | 54.2055 | 0.4938 |
| **7** | Branch 834-848 | 55.9825 | 1.7678 |
| **8** | Branch 862-838 | 57.5004 | 1.4813 |
| **9** | Branch 836-840 | 57.4151 | 0.2621 |

TABLE VIII
PERFORMANCE COMPARISON WITH SOME EXISTING METHODS

| Method | Max. Error (%) |
|---|---|
| Proposed method | 5.21 |
| Impedance-based [3] | 6.01 |
| ANN & WPT & Transient-based [6] | 4.68 |
| ANN & Impedance-based [8] | 4.57 |
| ANN & DWT & Transient-based [10] | 1.60 |
| Adaptive Impedance-based [19] | 1.10 |
| ANN & DWT & Transient-based [26] | 7.00 |
| SVM & WSN & Transient-based [38] | 1.50 |

ANN: Artificial Neural Networks, DWT: Discrete Wavelet Transform, SVM: Support Vector Machine, WSN: Wavelet Scattering Networks, WPT: Wavelet Packet Transform.

Note that the maximum percentage error is about 5.2%, and the less accurate scenarios correspond to faults close to the DS: less than 6 km. Very likely, this can be improved by increasing the number of fault locations on the main feeder.

Table VIII presents a comparative analysis of the proposed method alongside several existing methods from the literature. It is important to note that identifying the best approach for FL in distribution networks is challenging, as the most suitable method depends on several factors, including the level of network modernization. The ultimate goal is to provide the reader with a concise overview of the potential and direction of the proposed methodology.

## V. DISCUSSION

The results presented in this study (see also Table VIII) indicate that certain methods yield better results for FL, even when tested on a distribution network with different configurations. Notably, the potential for improving the performance of the ANN model as a classifier appears to be higher than its ability as a predictor. Enhancing its performance could be achieved by increasing the number of hidden layers in the ANN model.

One common limitation in any ANN-based FL approach is its dependence on the measurement location and the network topology. The ANN models must be retrained to accommodate the new measurement set to the new scenarios. To reduce this limitation, alternative methods that do not rely on training data could be explored during the FL stage. Before this, evaluating the robustness of the FI and FL stages would help identify how sensitive the methodology is to changes in the distribution network.

Another significant challenge is the lack of a comprehensive training database based on real fault data. This limitation, which is typical of ANN-based FL approaches, stems from the scarcity of real fault data in distribution systems.

Moreover, in larger and complex active distribution networks, the further the fault point is from the DS, the more challenging it becomes for our methodology to produce accurate results, particularly in the FL stage. This challenge is also common in methods that rely solely on measurements from the DS. A potential solution lies in incorporating data from strategically placed devices along the distribution system, such as fault detectors, smart meters, and synchrophasors (PMUs). Additionally, data from protection devices like overcurrent relays, recloser or interruption fuses could be integrated to improve fault location accuracy. This would be particularly beneficial when applying the methodology in microgrids or smart grids, where communication systems, data synchronization, and energy management play a crucial role.

## VI. CONCLUSION

A methodology for FSI of short-circuit faults in active distribution networks has been presented considering two types of wind power generator models (DFIG and PMSG) with inverter-based network interfaces. To validate the proposed FSI methodology, the IEEE 34-node test feeder was used in combination with MATLAB® Simulink® as simulation software.

The FSI methodology was divided into two stages, fault identification (FI) and fault location (FL), and based on coefficients obtained from DWT decomposition using Db4 mother wavelet. Energy per unit ($E_{PU}$) and relative energy entropy per unit ($REE_{PU}$) calculated from the DWT results were used as input to train and validate ANN models. The application of regularization indexes was very important to clearly highlight the difference between faulty and healthy phases.

The results exhibited an acceptable performance of ANN architectures with a maximum accuracy of 95% for IF stage and a maximum error of 5.2% for FL stage, taking into account the number of samples (128 samples per cycle) used for training and validation at each stage: one cycle before and one cycle after fault detection. The performance of ANNs was analyzed using confusion matrices and linear regression.

The robustness of the proposed methodology was demonstrated across various fault conditions, including different fault resistances, types, initiation angles, and locations. The methodology relies solely on instantaneous phase current signals measured at the distribution substation, using the existing current transformers. As a result, there would be no additional cost associated with its implementation.

A future extension of the proposed methodology could be based on the use of data from fault indicators strategically

installed in the analyzed distribution system or real-time data from μPMUs for application in microgrids.

An ideal FSI algorithm could be based on a single ANN (with probably more than one hidden layer) that would cover all the stages, fault types and locations. Future work using deep learning techniques [38] should consider such an option, and perhaps adding voltage measurements at DS [39].

## ACKNOWLEDGEMENTS

This paper uses models for representing wind power generators that were downloaded from the MATLAB®/MathWorks web site [24]. The authors are very grateful to the developers of those models for making them publicly available.

This paper is part of the Doctoral thesis written by J. C. Peqqueña Suni at the State University of Campinas (Brazil) [40].

## REFERENCES

[1] S. F. Alwash, V. K. Ramachandaramurthy and N. Mithulananthan, "Fault-Location Scheme for Power Distribution System with Distributed Generation," *IEEE Transactions on Power Delivery*, 30(3) 1187-1195, 2015, doi:10.1109/TPWRD.2014.2372045.

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

[3] J.J. Mora-Florez, R.A. Herrera-Orozco, A.F. Bedoya-Cadena, "Fault location considering load uncertainty and distributed generation in power distribution systems", *IET, Generation Transmission, Distribution* 9, 287–295, 2015, doi.org/10.1049/iet-gtd.2014.0325.

[4] P. Mandi tereza and R. Bansal, "Renewable distributed generation: the hidden challenges - a review from the protection perspective," *Renewable and Sustainable Energy Reviews*, 58, 1457–1465, 2016, doi:10.1016/j.rser.2015.12.276.

[5] M. Shih, A. Conde, Z. Leonowicz, and L. Martirano, "An adaptive overcurrent coordination scheme to improve relay sensitivity and overcome drawbacks due to distributed generation in smart grids," *IEEE Transactions on Industry Applications*, 53(6), 5217–5228, 2017, doi:10.1109/TIA.2017.2717880.

[6] D. S. Pessoa, A.L., Oleskovicz, M. & Martins, P.E.T. "Sensibility analysis of a fault location method based on ANN, WPT and Decision Tree in distribution Systems". *J. Control Autom. Electr. Syst.* 31, 990–1000, 2020, doi.org/10.1007/s40313-020-00597-6.

[7] H. Maruf, F. Müller, M. Hossan, and B. Chowdhury, "Locating faults in distribution systems in the presence of distributed generation using machine learning techniques," *9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, NC, USA, 1-6, 2018, doi:10.1109/PEDG.2018.8447728.

[8] R.H. Salim, K.R. Caino de Oliveira, A.D. Filomena, M. Resener, A.S. Bretas, "Hybrid fault diagnosis scheme implementation for power distribution systems automation", *IEEE Trans. Power Deliv.* 23, 1846–1856, 2008, doi: 10.1109/TPWRD.2008.917919.

[9] M. Usman, J. Ospina, and M. Faruque, "Fault classification and location identification in a smart distribution network using ANN," *IEEE Power & Energy Society General Meeting (PESGM)*, Portland, OR, USA, 1-6, 2018, doi:10.1109/PESGM.2018.8586471.

[10] A. Adewole, R. Tzoneva, and S. Behardien, "Distribution network fault section identification and fault location using wavelet entropy and neural networks," *Applied Soft Computing*, 46, 296-306, 2016, doi:10.1016/j.asoc.2016.05.013.

[11] M. Kim, J. An, Y. Oh, S. Lim, D. Kwak, and J. Song, "A method for fault section identification of distribution networks based on validation of fault indicators using artificial neural network," *Energies*, 16, 5397, 2023, doi:10.3390/en16145397.

[12] Y. Tuna and A. Ali, "Convolutional neural network-assisted fault detection and location using few PMUs," *Electric Power Systems Research*, 235, 110705, 2024, doi:10.1016/j.epsr.2024.110705.

[13] H. Rezapour, S. Jamali, and A. Bahmanyar, "Review on artificial intelligence-based fault location methods in power distribution networks," *Energies*, 16, 4636, 2023, doi:10.3390/en16124636.

[14] L. Acácio, P. Guaracy, T. Diniz, D. Araujo, and L. Araujo, "Evaluation of the impact of different neural network structures and data input on fault detection," *IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America)*, Quito, Ecuador, 1-5, 2017, doi:10.1109/ISGT-LA.2017.8126699.

[15] J. Yu, Y. Hou, A. Lam, and V. Li, "Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks," *IEEE Transactions on Smart Grid*, 10(2), 1694-1703, 2019, doi:10.1109/TSG.2017.2776310

[16] Y. Mamuya, Y. Lee, J. Shen, M. Shafiullah, and C. Kuo, "Application of machine learning for fault classification and location in a radial distribution grid," *Applied Sciences*, 10, 4965, 2020, doi:10.3390/app10144965.

[17] A. H. Orozco, J. M. Flórez, and S. P. Londoño, "An impedance relation index to predict the fault locator performance considering different load models," *Electric Power Systems Research*, 107, 199-205, 2014, doi.org/10.1016/j.epsr.2013.10.007.

[18] C. O. Henao, A. Bretas, A. H. Orozco, R. Ch. Leborgne and D. Schwanz, "Inverter-based DG impact on impedance-based fault location algorithms," *11th IEEE/IAS International Conference on Industry Applications*, Brazil, 1-6, 2014, doi:10.1109/INDUSCON.2014.70594252014.

[19] C. O. Henao, A. Bretas, J. M. Quintero, A. H. Orozco, J. P. Rivera and J. C. Velez. "Adaptive impedance-based fault location algorithm for active distribution networks." *Applied Sciences*, 2018, doi:10.3390/APP8091563.

[20] F. Mohammadi, G. Nazri, and M. Saif, "A fast fault detection and identification approach in power distribution systems," *International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET)*, Turkey, 1-4, 2019, doi: 10.1109/PGSRET.2019.8882676.

[21] G. G. Santos, T. S. Menezes, P. H. A. Barra, and J. C. M. Vieira, "An efficient fault diagnostic approach for active distribution networks considering adaptive detection thresholds," *International Journal of Electrical Power & Energy Systems*, 136, 107663, 2022, doi:10.1016/j.ijepes.2021.107663.

[22] O. Naidu, R. Gore, N. George, and S. Ashok, "A new approach for fault location on modern distribution systems with integrated DER," *Biennial International Conference on Power and Energy Systems: Towards Sustainable Energy (PESTSE),* India, 1-6, 2016, doi:10.1109/PESTSE.2016.7516499.

[23] IEEE 34 Node Test Feeder, "IEEE PES AMPS DSAS Test feeder working group," Available: https://site.ieee.org/pes-testfeeders/resources/

[24] G. Richard, "Wind farm – DFIG and PMSG detailed model," *MathWorks 2024b*. Available: https://www.mathworks.com/help/physmod/sps/ug/wind-farmdfig-detailed-model.html.

[25] M. Salles, K. Hameyer, J. Cardoso, A. Grilo, and C. Rahmann, "Crowbar system in doubly fed induction wind generators," *Energies*, 3, 738-753, 2010, doi:10.3390/en3040738.

[26] K. Lout and R. K. Aggarwal, "Performance analysis of a novel AI based approach to fault classification and location in an active distribution network with Type 3 and Type 4 wind turbine connections," *12th IET International Conference on Developments in Power System Pro*tection *(DPSP)*, Denmark, 2014. 1-6, doi:10.1049/cp.2014.0021.

[27] A. A. Piedy del Mar, "Analysis of the impact of the crowbar protection on short-circuit level and quality index". *Renewable Energy & Power Quality Journal (RE&PQJ)*, 1, 813-818, 2017, doi.org/10.24084/repqj15.473.

[28] Technical Guide, "Numerical distance protection," *Schneider Electric*. Available: https://www.rza.by/upload/iblock/da5/P44x_EN_T_I95_v.C7-D4-D5-D6.pdf

[29] R. Bouchet, O. Saad, and A. Xémard, "Module de protection générique sous EMTP-RV," *Hydro-Québec*, Canada, 2007. Available: https://www.emtp.com/products/modules/protection-toolbox.

[30] R. Das, "Determining the Locations of Faults in Distribution Systems", Doctoral thesis, Saskatchewan Univ., Canada, 1998. Available: http://hdl.handle.net/10388/etd-10212004-001150.

[31] S. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, 38(2), 617-643, 1992, doi:10.1109/18.119727.

[32] M. Dashtdar, R. Dashti, and H. Shaker, "Distribution network fault section identification and fault location using artificial neural network," *5th International Conference on Electrical and Electronic Engineering (ICEEE), T*urkey, 273-278, 2018, doi:10.1109/ICEEE2.2018.8391345.

[33] H. Okumus and F. Nuroglu, "A random forest-based approach for fault location detection in distribution systems," *Electrical Engineering*, 103, 257-264, 2021, doi:10.1007/s00202-020-01074-8.

[34] P. Stefanidou-Voziki, N. Sapountzoglou, B. Raison, and J. Dominguez-Garcia, "A review of fault location and classification methods in distribution grids," *Electric Power Systems Research*, 209, 108031, 2022, doi:10.1016/j.epsr.2022.108031.

[35] D. Borras, M. Castilla, and N. Moreno, "Wavelet and neural structure: a new tool for diagnostic of power system disturbances," *IEEE Transactions on Industry Applications*, 37(1), 184-190, 2001, doi:10.1109/28.903145.

[36] R.O. Duda, P.E. Hart, and D.G. Stork, "Pattern Classification", *John Wiley & Sons*, 2001. *Journal of Classification* 24, 305–307, 2007, doi:10.1007/s00357-007-0015-9.

[37] M. H. Beale, M. T. Hagan, & H. B. Demuth. "Neural network toolbox. User's Guide", *MathWorks*, 2, 77-81, 2010. Available: https://ge0mlib.com/papers/Books/04_Neural_Network_Toolbox_Reference.pdf.

[38] Ch. G. Arsoniadis and V. C. Nikolaidis, "A machine learning based fault location method for power distribution systems using wavelet scattering networks," *Sustainable Energy, Grids and Networks*, 40, 101551, 2024, doi:10.1016/j.segan.2024.101551.

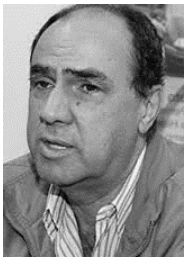
[39] V. Rizeakos, A. Bachoumis, N. Andriopoulos, M. Birbas, A. Birbas, "Deep learning-based application for fault location identification and type classification in active distribution grids", *Applied Energy*, 338, 120932, 2023, doi:10.1016/j.apenergy.2023.120932.

[40] J. C. Peqqueña Suni, "Sistema híbrido de detecção e diagnóstico de faltas de curto-circuito em redes de distribuição de energia elétrica primárias", Doctoral thesis, State University of Campinas, Brazil, 2024. Available: https://hdl.handle.net/20.500.12733/20540.

**Juan Carlos Peqqueña Suni** received the Ph.D. degree in 2024 in electrical engineering from State University of Campinas (UNICAMP), Campinas in Brazil. He is currently a professor of Electrical Engineering at the Federal University of Ceará (UFC), Sobral campus in Brazil, where he has been teaching and conducting research in electrical power systems since 2014. His areas of interest include synchronous machine parameters, analysis of distribution systems and renewable energies using artificial intelligence.

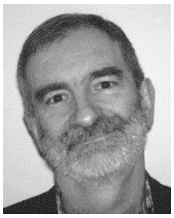
**Marina Gabriela Pérez Paredes** received a B.S. degree in electronic engineering from the Technological University of Peru, Lima, Peru, and M.S. and Ph.D. degrees in electrical engineering from the University of Campinas, Campinas, Brazil, in 2013 and 2018, respectively. From 2015 to 2016 I worked at the Hitachi Research Laboratory, Japan, in its green mobility field. She is a postdoctoral researcher at the Faculty of Mechanical Engineering at the University of Campinas (UNICAMP), Brazil. Her research focuses on interests include renewable energy, electric machine drives, and electric vehicles.

**Marcelo Vinícius de Paula** received the B.S. degree in electrical engineering from the Federal University of Goiás, Brazil, in 2016, the M.Sc. degree in electrical engineering from University of Campinas (UNICAMP), Campinas, Brazil, in 2018, and Ph.D. in mechatronics engineering from UNICAMP, Campinas, Brazil, 2022. Currently, he is a Professor at Faculty of Mechanical Engineering of UNICAMP. He involves in the areas of electric machines and drives, power electronics, and transportation electrification. His research interests include renewable energy, electric machine drives and electric vehicles.



**E. Ruppert Filho** received the B.Sc degree in Electrical Engineering in 1971, MSc degree in 1974 and PhD degree in 1982 at Computer and Electrical Engineering School of the Campinas State University in Brazil (UNICAMP). During his professional life he worked as project engineer and/or as consultant for several large companies such as Itaipu, Petrobras, General Electric, Alstom, Copel, CPFL, and Elektro in Brazil and abroad. He has been with the Campinas State University (UNICAMP) in Campinas, Brazil, since 1972 as Professor and Researcher. His research interests are power electronics, superconductor current limiters, electrical power systems, distributed generation, electric machines and motor drives. He

has published many technical papers in international journals and conferences and has advised several M.Sc. and PhD thesis along his career.



**Juan A. Martínez-Velasco** received the Ph.D. degree in 1982 from the Universitat Politecnica de Catalunya (UPC), Barcelona, Spain. He was with the Department of Electrical Engineering, UPC. He is retired since September 2017. His teaching and research interests included transmission and distribution, power system analysis, and Electro Magnetic Transients Program (EMTP) applications.