A Neural Network-Based Classifier for Identifying and Locating Neutral Wire Breaks in Low Voltage Distribution Networks

Rubén Carmona-Pardo*†, Álvaro Rodríguez del Nozal*, Esther Romero-Ramos*‡
* Department of Electrical Engineering, Universidad de Sevilla, Seville, Spain
† Ingelectus Innovative Electrical Solutions, Seville, Spain
‡ ENGREEN Laboratory of Engineering for Energy and Environmental Sustainability, Universidad de Sevilla, Seville, Spain.

***Abstract*—The breakage of the neutral conductor in low voltage distribution networks is a major concern for distribution companies. This breakage causes significant voltage deviations that can damage the connected equipment as well as jeopardizing people. The detection and localization of the breakage is a major challenge as it does not always manifest in the same way. This work presents a methodology based on artificial intelligence for the detection and localization of neutral conductor breaks in distribution networks. Two neural networks are trained in attempt to solve each of the challenges. For this purpose, measurements commonly taken by smartmeters such as power and nodal voltages are used. The methodology is evaluated in simulation exhibiting a good performance.**

***Index Terms*—Distribution system; LV networks; Neural networks; Neutral breakage**

I. Introduction

Loss of neutral conductor in low voltage distribution networks is a major concern of utilities [1], [2]. This series open circuit fault causes overvoltages that damage electronic equipment, electric motors, lighting loads, appliances, etc. Another consequence can be the creation of hazardous touch voltages on exposed conductive parts, putting people's safety at risk [1].

Neutral loss conditions can arise for different reasons: overloading, load unbalancing, loose termination of neutral conductor due to poor workmanship, poor maintenance, etc. The adverse effects caused by neutral conductor breakage depend on the pre-fault load and unbalance levels, as well as the neutral grounding system adopted by the utility, the ground resistance values, and the location of the fault [3]. A massive presence of distributed low-carbon technologies may even worsen the starting conditions previous to the loss of neutral [4].

In most cases, these fault situations are only detected when reports of supply interruption are received from customers. Then, network operators must determine whether or not these outages are due to a broken neutral fault and, if so, where in the network they have occurred. There are numerous patents for open-neutral detection devices to be installed at the feeder end or at the loads terminals, a good summary of which is given in [5]. This solution is quite expensive as these protection devices must be distributed throughout the LV network. The authors in [5] propose using a single detection device located at the head of the feeder. Its working is based on evaluating the 3rd harmonic current, which changes significantly when open neutral fault occurs. This proposal is validated experimentally in [6].

The work presented in [7] proposes to use Fourier transform analysis of the measured neutral-earth voltage (NEV) and system current. From this analysis it is possible to calculate a new index whose amplitude and sign allows to know if the neutral opening has occurred or not and its location. This methodology requires recording measurements at each node of the LV network, with the added difficult task of establishing a reference ground for NEV measurement.

A great variety of methodologies are proposed in [8], all of them based on using the data recorded by smart meters. Simple algorithm methods based on comparing a specific electric magnitude or index with a limit previously tuned are evaluated, as well as a combination of these basic methods. Some of these solutions seems to work quite efficiently, but most of them imply installing new software in already installed smart meters. Also, machine learning methodologies are considered, more specifically a decision tree-based algorithm and a neural network-based methodology, the first one performing better. One limitation of these artificial intelligence-based solutions is that need quite information of each smart meter, until 19 parameters. For confidentiality reasons, most of the proposed solutions are not detailed.

This work is focused on the resolution of two main problems, namely *broken neutral fault detection* (BNFD) and *broken neutral fault location* (BNFL). That is, given an LV network with a European layout, i.e., three-phase four-wire systems, this work aims to identify whether a broken neutral fault has occurred and, if so, to determine the line segment where the problem has occurred. In this way, the utility can preserve the proper functioning of the network, as well as economize on system repair due to the easy location of the

This work was supported by the Ministerio de Ciencia e Innovación (MCIN) Agencia Estatal de Investigación (AEI) 10.13039/501100011033 through the "European Regional Development Fund (ERDF) A way of making Europe" under Grant PID2021-124571OB-I00; and by the European Commission through the project SUNRISE under grant agreement 101079200.

fault. The proposed solution takes advantage of the massive information collected by smart meters, making use of them for the training of neural networks that can predict the problems tackled. Thus, the main contributions of the paper are:

- Motivate the adverse effects of a neutral wire breakage in a LV distribution network.
- The development of an AI-driven methodology for identifying neutral faults in LV distribution networks.
- The construction of a second methodology aimed at effectively locating neutral faults within LV distribution networks.
- The use of limited and accessible data recorded by smart meters and metering equipment located in secondary substations as input data for the proposed methodologies.

Section II analyzes from a theoretical point of view the consequences of a neutral break in a basic unbalance low voltage network. This analysis helps to focus the design of the solution methodology to be defined. Section III describes the problem from an artificial intelligence perspective, including the definition of the problem, the benchmark LV network considered and the dataset used. In Section IV the proposed method is introduced together with all the techniques used to enhance the approach. Section V evaluates the performance and robustness of the proposed solution under simulation and Section VI summarizes the main results of the conducted research and future work.

II. Motivation of the problem

European LV distribution networks are radially operated three-phase four-wire networks. In these networks, there are both single-phase and three-phase customers. Although the distribution company makes an effort to distribute the single-phase customers appropriately between the phases, the arbitrary consumption of the different customers generates imbalances in the network. These unbalances result in the sum of the currents of the three phases at the nodes not being zero, and therefore, there is a return current to the head-end through the neutral conductor that relieves the nodal voltages while maintaining operating margins.

Occasionally, the neutral conductor breaks, which is usually difficult to identify and locate, especially if we focus on urban networks where the cables are buried. In this situation, two possible cases can arise. First, if the consumption node has a ground connection, the currents caused by unbalances can use that path to return to the head-end. However, the values of the grounding resistances are often very high, which complicates current flow. Second, there is no path for the flow of this current, so Kirchhoff's current law must be satisfied at that node, i.e. $i_a + i_b + i_c = 0$. This fact generates serious problems in the voltages at that node. In both cases, the neutral-ground voltage of the node may increase considerably, compromising the isolation of the equipment connected to the network at that point.

To illustrate the problem, consider a system in which the low voltage side of a secondary substation (modeled as a three-phase fully balanced voltage source) feeds an unbalanced load by means of an aluminum power line of 500 m length. In this example a 63 kW of active power demand is considered being the 36 % demanded in phase a and the 32 % in phases b and c, respectively. In addition, a perfect grounding with zero resistance is considered at the secondary substation and a resistance of 40 Ω at the load node. Two power flow problems are solved over the example network defined above. For both cases, the voltages at the head-end (red) as well as the voltages at the load terminals are shown. On the one hand, the line voltages are shown in cyan solid line, while the phase-neutral voltages are shown in yellow dashed lines. Finally, the blue dashed lines represent the phase-ground voltages. First, when

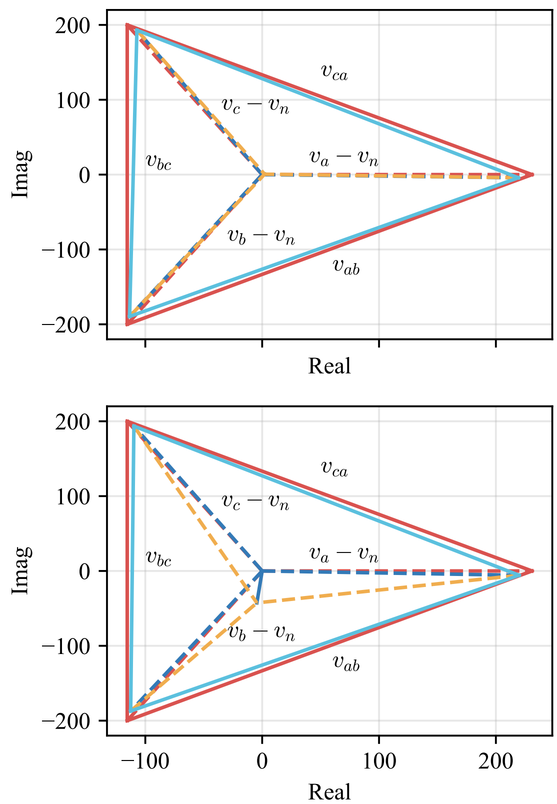

Fig. 1. Voltage values for the test case with neutral (up) and neutral breakage (down) in the power line. Voltages at the head-end are represented in red solid liness; the line voltages at the load are shown in cyan solid lines; the phase-neutral voltages are represented in yellow dashed lines; and the phase-ground voltages are depicted in blue dashed lines.

the neutral conductor is intact, the voltage at that point reaches a value of 2 V. This value is considered practically negligible and, as can be seen, the triangle of voltages at the head-end and at the load is practically identical and is only affected by the losses in the line (reducing its size) and the small unbalance in the load (rotating the triangle). Second, when the neutral breakage occurs, the voltage at that point increases to about 50 V. The phase-to-ground voltages remain practically constant, however, when the neutral is displaced, overvoltages can be observed in the phase-to-neutral voltages at phases a and c, as well as an undervoltage at phase b.

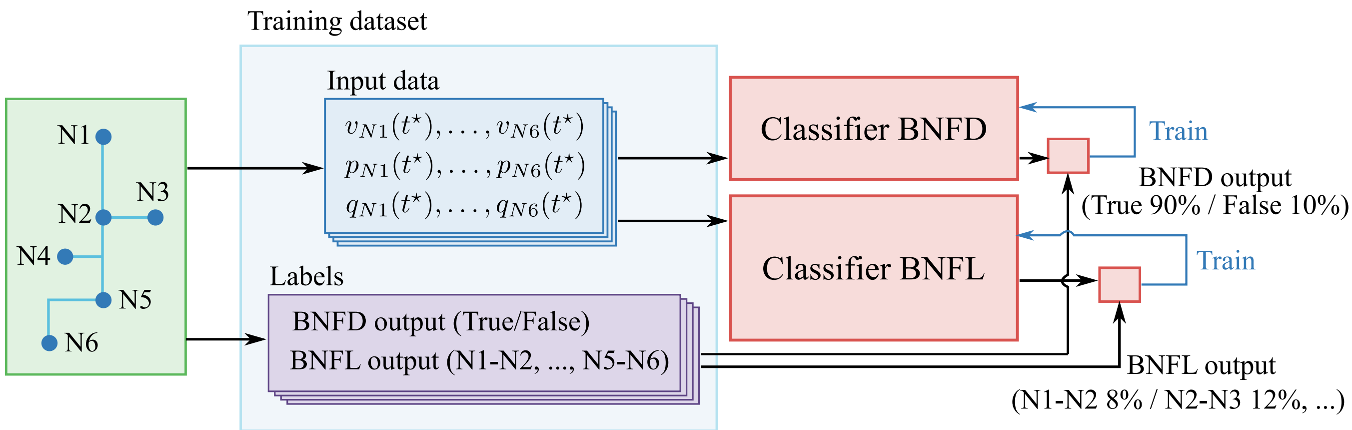

Fig. 2. Dataset generation and classifier training procedure.

Note that phase-to-neutral overvoltages on phases a and c can damage connected equipment as well as cause automatic disconnection of the customer to ensure safe operation of the network. These voltage problems could be even worse when distributed generation is taken into account, as they can lead to higher pre-fault unbalance scenarios. The neutral grounding system adopted by the utility, the ground resistance values, and the location of the fault are additional factors that influencing the level of the adverse effects caused by neutral conductor breakage [3].

When this problem is analyzed in larger networks, the neutral brakage effect may go unnoticed, producing milder effects on voltage levels depending on where the breakage occurs. However, if a neutral breakage occurs in more than one section of the network, it can have serious consequences, involving long periods of supply interruption to carry out repairs. Thus, having a suitable tool for the identification and location of neutral conductor breaks seems to be a very necessary tool.

III. Problem Statement

This section is structured into three subsections. Firstly, the two problems addressed are described from an artificial intelligence perspective, specifying the classification schemes used for each problem's resolution. Additionally, the input requirements for the solution engine are defined, along with the output format. Secondly, the network under study is presented, outlining its topology, grid connection points, and grounding configurations. Finally, the methodology employed for data acquisition from this network under different operating modes is outlined.

A. Problem description

As already mentioned, the objective of this work is to solve two main problems that can arise in LV distribution networks: *broken neutral fault detection* (BNFD) and *broken neutral fault location* (BNFL). It is important to note that both problems are related. The location of the neutral breakage will be addressed only if it has occurred.

To address both problems, Section IV proposes a deep learning-based method. However, before delving into the methodology, it is essential to reframe these issues from an electrical context to a data-driven perspective. Both problems fall under the classification paradigm, where the objective is to categorize input data into predefined classes or categories. The first problem, BNFD, is a binary classification task, determining whether the neutral is damaged or intact. The second problem, BNFL, presents a more intricate multiclass classification challenge, where data needs to be assigned to distinct categories indicating the specific line section in which the fault occurred. The output of the classifiers gives a probability to each of the possible outputs. The sum of all these probabilities is equal to 1. Thus, the classifier for the BNFD problem will assign one probability to the case of a neutral breakage and another probability to to the case in which the neutral wire is intact. On the other hand, in the BNFL problem, the output will be the probability of breakage in each section of the network.

The input data of both classifiers is a collection of the available information of the current state of the network. In this work, it is considered accessible the nodal voltages and the active and reactive power measurements of the smart meters installed in the network and the voltage magnitude recorded at the secondary substation. The dataset should be divided into the one used to train the classifier, i.e. training dataset, and the one used to evaluate the performance of the method, i.e. testing dataset. The dataset must be labeled since a supervised learning approach is used, that is, it must be specified the desired output of the classifier: the state of the neutral conductor and, in case of breakage, the line segment where the fault has occurred.

First, the training dataset will be used to train the classifier, i.e. to correctly adjust its internal variables in order to match the desired outputs (those labels included in the dataset). Second, once the training process is completed, the testing dataset is used to test the performance of the method by comparing the classifier outputs with the dataset labels. The analysis of the performance will be deeper explained together with the simulation results.

An illustrative schematic of the dataset generation and the training process is shown in Fig. 2. As can be seen, the dataset is constructed from the input data to the classifier (voltage and power measurements taken from the network) and the labels corresponding to these inputs (whether there is a break and its location). Note that the dataset is composed of this set of measurements for several time instants (with different loads and neutral wire state). After the dataset is generated, the input data is sent to the classifiers, which compare their estimates of the neutral state with the labels in the dataset. The discrepancies obtained are used to reassign values to their internal variables in the training process.

B. Network benchmark

The considered benchmark network is part of the 400 V European LV benchmark distribution network proposed by the CIGRE Task Force C06.04.02 [9]. In particular, the residential sub-network whose topology is depicted in Fig. 3. In this network, there are several zero-injection buses without neutral conductor grounding (such as R5 or R7). In order to facilitate the NBFL problem, it has been decided to avoid these nodes by adapting the line models to achieve a fully equivalent network.

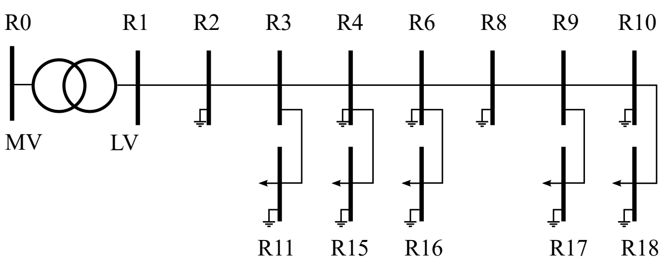

Fig. 3. Topology of residential European LV benchmark network

The benchmark system under consideration is a three-phase four-wire underground network designed to accommodate both single-phase and three-phase prosumers. This network spans a length of 570 meters and is powered by a 20/0.4 kV-500 kVA power transformer configured in a delta-star arrangement. To ensure proper grounding, various LV grounding configuration exist, as documented in [10]. One prevalent approach involves the use of a multi-grounded neutral scheme, wherein the neutral conductor is strategically grounded at multiple points throughout the network, a strategy well-documented in [11]. In our specific scenario, grounding resistors with values of $3\,\Omega$ and $40\,\Omega$ have been incorporated at the secondary transformer and certain intermediate nodes, respectively, as detailed in [9]. The nodes with a grounding connection as well as the customers grid connection nodes are shown in Fig. 3.

In the absence of real network data, the decision has been made to utilize this benchmark for the study. To ensure the problem closely resembles reality and obtain the necessary smart meter data, an accurate grid model has been constructed following the procedure explained in [12]. Subsection III-C details the process of generating consumption/generation scenarios that will conform the used datasets.

C. Generation of the dataset

Two datasets have been generated as mentioned in the previous section. On the one hand, the training dataset contains multiple network states labeled with the neutral conductor status. On the other hand, the testing dataset contains labeled information to evaluate the model performance. Each network state included in the dataset has been obtained by solving the power flow problem in the proposed network. For this purpose, the fully balanced base load scenario represented in Table I has been considered.

TABLE I
BASE LOAD SCENARIO

Node	S (kVA)	$cos\varphi$
R1	200	0.85
R11	15	0.85
R15	52	0.85
R16	55	0.85
R17	35	0.85
R18	47	0.85

Starting from the base load case, random scenarios have been generated by modifying the power demanded in each node and phase between 80 % and 120 % of the starting value. It is considered that the status of the network can be known once every hour (based on the information acquired by the smart meters), and therefore, one scenario can be included in the dataset every this period of time.

In order to evaluate the influence of the training dataset on the performance of the proposal, datasets constructed from one week of data, one month, six months, one year, one and a half years and two years are considered. It is important to highlight that larger datasets include those of smaller size. Thus, for example, the entire dataset with one week of data is included in the dataset built from one month of information.

Finally, different percentages of neutral breakage exist in each of the datasets. In particular, datasets with 10%, 9%, 8%, 7%, 6%, 5% and 4% of the scenarios presenting a breakage are considered. Note that these percentages are of small value since in a scenario closer to reality the breakage of the neutral conductor should not appear with regularity.

Thus, there are 6×7 datasets with all combinations of size (six possibilities depending on the time extension) as well as fault probability (seven possibilities).

IV. PROPOSED METHOD

This section presents the proposal implemented to solve the BNFD and BNFL problems. First, the structure of the classifier based on neural networks is presented, identifying the considered topology as well as its activation functions. Subsequently, the training process for the neural networks is defined, detailing the data augmentation and data scaling tools used.

A. Classifier architecture

The two classifiers used in this work are composed of a neural network with a sequential model and a dense connection, that is, the neurons that make up each layer of the neural network are connected with all the neurons of the next layer. Both classifiers solve a Supervised Learning problem considering limited labelled data. The neural network based classification models have been trained considering a binary crossentropy loss function with a learning rate of 0.001. Table II details the structure of both neural networks identifying the number of layers, neurons, training parameters and the activation function used. For a more detailed explanation about the neural networks hyperparameters reader is referred to [13].

TABLE II
NEURAL NETWORK STRUCTURE

Layer	Output shape	Training params	Activ. func.
1	128	7424	ReLu
2	256	33024	ReLu
3	128	32896	ReLu
4 (BNFD)	1	129	Sigmoid
4 (BNFL)	12	129	Sigmoid

Note that both neural networks, either for the BNFD or BNFL problem, have the same structure, only having a difference in the last layer due to the type of classifier used. While in the case of the BNFD problem, the classifier is binary with one output (probability of breakage), in the case of the BNFL problem, the classifier is multi-class and presents a probability for the breakage of each network segment. In the case of the network studied in this work, there are 12 segments as shown in Fig 3.

It is important to note that different neural network structures have been considered by modifying their hyperparameters: number of layers, number of neurons, activation functions, etc. Similarly, the networks have been trained with different loss functions and learning rate values. Among all the results obtained, the best performance has been achieved with the ones presented in this section.

B. Dataset enhancement

The training of neural networks depends on data quality. This section covers data preprocessing techniques for extending and modifying datasets to improve the classifier performance with the addressed problems. Starting with the input data described in the previous section (voltage and power measurements at different network nodes), two strategies will be employed: data augmentation to expand the training dataset and data scaling to enhance classifier performance.

Data augmentation consists of expanding the initial dataset by generating altered instances from the original data. Its main purpose is to allow effective training when the input dataset is small, however, it is considered also a prudent approach for mitigating overfitting or optimizing model performance. In addressing this problem, efforts have been made during the data generation process to create a training dataset that closely resembles data typically accessible from a real network. This has resulted in training datasets where the majority class represents the absence of neutral breakage. The scarcity of scenarios involving failures may consequently result in a reduced ability to accurately detect these failures. To address this issue, a set of new synthetic data for the minority classes (instances with neutral breakage) has been generated. For this purpose, all the instances of the minority class of the dataset have been taken in order to subsequently apply a random uniform noise to each of the measurements for each instance, obtaining new synthetic instances that extend the base case dataset.

Once the training dataset has been extended based on the data augmentation method, all the information included in the dataset must be scaled. The data scaling method consists in adjusting the input values of the dataset to be within a specific range. The main benefits of scaling the datasets information are [14]: (i) stability is improved during the training process since all the information in the dataset and its optimality gradients take values on the same scale; (ii) numerical problems such as overflows that can arise when working with data with different scales are avoided; (iii) data scaling allows better compatibility with activation functions; and (iv) when features are on comparable scales, the neural network can learn patterns and relationships more effectively. Thus, a *MinMaxScaler* has been applied to all the datasets gathered, transforming all the dataset features within the range $[0, 1]$.

V. METHOD PERFORMANCE

This section presents the results obtained after applying the proposed methodology to the network under study. For this purpose, multiple training of the classifiers have been carried out, taking into account the datasets presented in Section III-C and the neural network structures defined in Table II. This section first presents the metrics used to objectively evaluate the quality of the results obtained. After that, the results of the NBFD problem are analyzed and, finally, those obtained for the NBFL problem. A comparative analysis has not been performed because previous equivalent works based on the use of artificial neural networks [8] do not provide details on the proposed solutions that would allow their implementation. All simulations have been executed in a personal computer, with a 2.24 GHz AMD 16-Core Processor and 64 GB RAM, using *OpenDSS* to generate the dataset and *Python* with the artificial intelligence library *TensorFlow* to train and evaluate the developed neural networks.

A. Performance metrics and evaluation procedure

To evaluate the performance of the classifiers a second dataset was generated, namely testing dataset. The same procedure used in the case of the training dataset detailed in Section III-C has been followed. This dataset contains new scenarios that differ from those included in the training dataset. In this way, it is possible to evaluate with sufficient certainty the quality of the classifiers outputs. It is important to note that this dataset has also been preprocessed with data scaling, but unlike the training dataset, no data augmentation strategy has been applied to it, as it does not make sense to apply it in this case.

To evaluate the results in the case of the BNFD problem, there are four possible cases. If the classifier has been fed with input data from a scenario with neutral breakage, it can either identify the breakage (TP: True Positive) or fail to identify it (FN: False Negative). If, on the other hand, it has been fed with input data from a scenario without a neutral breakage, it can either succeed (TN: True Negative) or identify a non-existent breakage (FP: False Positive). Thus the accuracy and the precission of the classifier can be defined as:

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}}. \end{aligned}$$

Note how the accuracy shows the success of the classifier in either identifying a break or confirming the good condition of the network, with respect to the total dataset. On the contrary, the precision only evaluates the correct neutral break identification with respect to all the neutral break identifications triggered by the classifier.

On the other hand, in the case of the BNFL problem, only the precision term will be used to quantify the number of times the classifier correctly identifies the line section in which the fault occurred (CP: Correct Predictions), with respect to the total number of cases, i.e AP (All Predictions):

$$\text{Precision} = \frac{\text{CP}}{\text{AP}}.$$

Note that in this case the classifier only acts when it is sure that there is a neutral breakage in the network.

TABLE III
BNFD PROBLEM CLASSIFIER ACCURACY.

Rate	2.0 y	1.5 y	1.0 y	0.5 y	1.0 m	1.0 w
10%	94.2	93.0	92.0	90.5	84.8	89.4
9%	92.5	93.0	91.5	93.0	86.8	86.8
8%	94.2	93.2	92.8	87.0	87.2	88.2
7%	94.2	93.8	91.2	90.2	86.2	86.2
6%	93.0	92.0	92.8	86.2	87.0	87.2
5%	93.5	93.8	87.2	91.0	86.0	84.8
4%	92.5	93.0	91.5	87.0	85.5	87.2

TABLE IV
BNFD PROBLEM CLASSIFIER PRECISION.

Rate	2.0 y	1.5 y	1.0 y	0.5 y	1.0 m	1.0 w
10%	95.4	96.5	95.4	93.6	74.6	84.4
9%	94.2	91.9	89.0	92.5	78.6	78.6
8%	95.4	91.3	91.9	91.3	80.3	81.5
7%	96.0	94.2	88.4	93.1	79.2	78.0
6%	91.9	94.2	94.2	91.3	80.3	79.8
5%	96.5	94.8	83.8	92.5	76.9	73.4
4%	98.3	91.9	89.6	89.6	75.7	79.8

TABLE V
BNFD PROBLEM CLASSIFIER TRAINING SPENT TIME (MIN).

Rate	2.0 y	1.5 y	1.0 y	0.5 y	1.0 m	1.0 w
10%	1.406	2.429	3.117	3.484	3.569	3.607
9%	1.406	2.456	3.158	3.518	3.601	3.640
8%	1.422	2.470	3.162	3.536	3.617	3.655
7%	1.446	2.513	3.212	3.568	3.649	3.691
6%	1.390	2.430	3.114	3.467	3.548	3.586
5%	1.368	2.450	3.154	3.506	3.583	3.622
4%	1.380	2.408	3.121	3.489	3.570	3.607

B. BNFD problem results

The results obtained in the BNFD problem are presented in the Tables III and IV, which shows, respectively, the accuracy and precision for the classifiers generated from the different training datasets. As can be seen, in the case of the accuracy, the results remain practically constant being slightly better for larger training datasets and failure rates, always performing above 85%. Nevertheless, when evaluating precision results (a metric focused on measuring the effectiveness in predicting true positives, i.e. broken neutral fault), the results of Table IV are highly sensitive to the size of the training datasets and the fault ratio. Additionally to these results, Table V shows the time spent in training the model for each data availability context evaluated. As can be observed, the smaller the original data set, the longer the computation times, which is reasonable since it is necessary to generate more synthetic scenarios to reinforce learning. Nevertheless, the values here presented evince the scalability of the methodology in terms of computational requirements. An illustrative representation of these metrics are presented in Fig. 4.

In view of the results, several conclusions can be drawn. Firstly, it is logical that better accuracy values than precision values are obtained, since the dataset used for classifier training comprises more situations in which the neutral wire remains intact and is therefore better suited to identify these sitiuations rather than the other ones. Second, when analyzing the precision results, it is evident that the performance of the classifier decreases with the size of the dataset. However, it is worth noting that thanks to the use of data augmentation technique, it has been possible to achieve precisions above 70% even for datasets with only one week of information, and above 80% for datasets comprising half a year or longer.

C. BNFL problem results

In order to properly analyze the results, it is important to remember that in this case the classifier output assigns a probability of breakage to each section based on the input data. Using this information, the user can consider that the breakage has occurred in that section to which a higher probability has been assigned, or set a threshold to that probability below which the information provided by the classifier is not considered consistent. For example, if the classifier identifies that the break has occurred in a particular section of the network with a probability of the 90 %, it seems to be quite consistent, however, if the maximum probability assigned to a section is 10 % it seems that the classifier has failed to identify the location of the break with certainty.

In this problem, given its complexity, we have chosen to train the classifier only with datasets with a 10 % presence of neutral conductor breakage. Based on that information, the results obtained are the ones plotted in Fig. 5 (not setting any

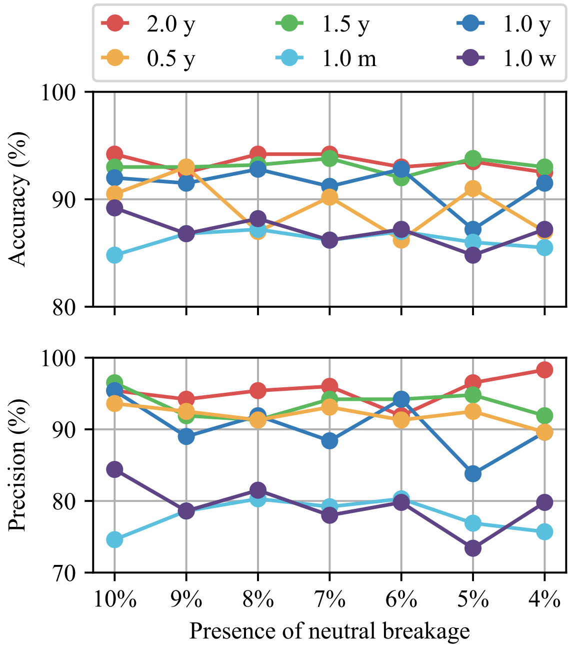

Fig. 4. BFND problem classifier accuracy and precision.

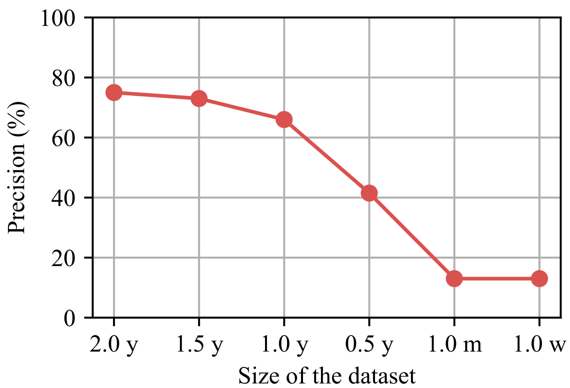

Fig. 5. BNFL precison.

threshold value and choosing that section of the network to which the classifier assigns a higher probability of breakage). Note that in this case, unlike the BNFD problem, the obtained precisions are much lower, even more so when small datasets are considered. For training data with an acquisition time of less than one year, the results obtained are rather weak, with detection rates of no more than 50 %, and therefore their use in this type of problem is discarded. However, when the classifier is trained on data from more than one year, the probabilities of correctly detecting the network fault increase to rates nearly 80%, which are quite effective values.

Next, we choose to analyze the behavior of the classifier that has achieved the best performance, i.e. the one trained with a dataset collected from two years and with an occurrence of defects in the 10% of the scenarios. Table VI shows the precision obtained for the classifier when different thresholds are set on the ouput returned by the classifier, i.e. considering inconsistent those responses of the classifier in which no section of the network is assigned a breakage probability higher than the one set by the threshold. It has been confirmed that, although there would be some cases in the testing dataset for which the proposed solution could not return a conclusive location, it is possible to improve the performance of the classification model by being more rigorous in determining that a broken neutral fault has been located.

TABLE VI
BNFL PRECISION WITH RESPECT TO THE CHOSEN THRESHOLD.

Threshold	None	0.5	0.6	0.8	0.9	0.95	0.99
Precision	75.0	77.7	78.2	79.0	81.5	84.6	85.6

Finally, Fig. 6 shows the confusion matrix of the evaluated classifier (the one trained with two years of data and a fault rate of the 10%). A confusion matrix is a compact representation of the performance of a classification model. In the specific case presented in this work, each row represents each of the line sections in which a neutral breakage has occurred (according to the testing dataset) and each column the prediction made by the classifier for each case. Each cell shows the percentage of times a break occurred in the corresponding row section and the classifier concluded that the breakage occurred in the corresponding column section. Thus, if the matrix were an identity matrix, the classifier would have been correct about the location of the breakage for all cases. Off-diagonal elements are defined as false positives, i.e. identification of breakages in erroneous sections. In view of the results, it can be concluded that the classifier is correct in most cases and, in case of misidentifications of the fault location, it is usually confused with line segments adjacent to the fault, perfectly adapting to the requirements imposed. This would respond to the behavior reflected in 5 where very high accuracy values would not be achieved because, on some occasions, faults are assigned to the adjacent segment.

VI. Conclusions

This work has presented a strategy for the detection and location of neutral breaks in low voltage electrical networks. Thanks to the use of data-based classification models and artificial intelligence techniques it is possible to identify and locate these defects, facilitating the repair work of the distribution company and avoiding major problems in the safety and equipment of consumers.

For this purpose, the work has presented two classifiers based on neural networks that solve two specific problems: the identification of neutral breakages and their localization. The results obtained in simulation have been evaluated for different training datasets with different sizes and defect rates presence in data. The construction of these datasets has been brought as close as possible to reality by considering that most of the instances are taken when there is not fault in the network.

	R1-R2	R2-R3	R3-R4	R4-R6	R6-R8	R8-R9	R9-R10	R3-R11	R4-R15	R6-R16	R9-R17	R10-R18
R1-R2	0.500	0.450	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.050	0.000
R2-R3	0.467	0.533	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R3-R4	0.000	0.053	0.895	0.000	0.000	0.000	0.000	0.000	0.053	0.000	0.000	0.000
R4-R6	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R6-R8	0.000	0.000	0.000	0.000	0.545	0.364	0.000	0.000	0.000	0.000	0.000	0.091
R8-R9	0.000	0.000	0.000	0.000	0.750	0.250	0.000	0.000	0.000	0.000	0.000	0.000
R9-R10	0.000	0.000	0.000	0.000	0.000	0.000	0.400	0.000	0.000	0.000	0.000	0.600
R3-R11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.933	0.067	0.000	0.000	0.000
R4-R15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
R6-R16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
R9-R17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
R10-R18	0.000	0.000	0.000	0.000	0.000	0.000	0.333	0.056	0.000	0.000	0.000	0.611

Fig. 6. BNFL confusion matrix.

The results obtained are quite promising, reaching an accuracy of more than 90 % in the identification of defects. Likewise, defect localization is achieved with an accuracy of over 80 %. Moreover, when the classifier establishes an erroneous location, the defect usually occurs in segments adjacent to the identified one. It should be noted that the detection and localization of the defect is performed with limited data from a specific time instant of the network status. Therefore, if the evaluation of the defects is considered with temporal data of days or weeks, the results would be even more conclusive and probably would improve the results here obtained. In addition, input data considered are electrical magnitudes accessible in current LV networks: active and reactive power and voltage magnitudes collected by smart meters and voltage magnitude in secondary substation. This gives the solution a strong practical feature.

Finally, the application of this method to a real network with data acquired directly from smart meters is considered as a future line of work. For this kind of real cases, model-based data augmentation, instead of purely data-based as is the case here, is expected to be promising in improving the dataset, after demonstrating throughout this work the effectiveness of applying an extension on the dataset.

References

[1] Mashangu H. Xivambu, “Impact of floating neutral in distribution systems”, 19th International Conference on Electricity Distribution (CIRED), Vienna, 21-24 May 2007, paper 0300.

[2] N. Kagan et al., “Methodology for support and analysis of indemnity requests due to electrical equipment damaged in Eletropaulo customers”, 10th International Conference on Harmonics and Quality of Power. Rio de Janeiro, Brazil, 2002, pp. 304-309 vol.1.

[3] C. D. Halevidis and E. I. Koufakis, “Power flow in PME distribution systems during an open neutral condition”, in IEEE Transactions on Power Systems, vol. 28, no. 2, pp. 1083-1092, May 2013.

[4] Frantzeskakis, Syllas et al,.“Loss of Neutral in Low Voltage Electrical Installation with connected DG units – consequences and Solutions”, 25th International Conference on Electricity Distribution (CIRED), Madrid, 3-6 June 2019, paper 833.

[5] J. Yong, C. Zhou, B. Yang and X. Wang, “A harmonic-based approach for open-neutral fault detection in low voltage systems”, 2012 IEEE 15th International Conference on Harmonics and Quality of Power, Hong Kong, China, 2012, pp. 53-58.

[6] L. Jiang and J. Yong, “Validation of the Harmonic-Based Method for Detecting Open-Neutral Fault in Low-Voltage Systems”, in IEEE Transactions on Industry Applications, vol. 54, no. 4, pp. 3145-3152, July-Aug. 2018

[7] Mohd Abdul Talib Mat Yusoh, Ahmad Farid Abidin, Zuhaila Mat Yasin, Sim Sy Yi, “Identification of the source location Neutral to Earth Voltage (NTEV) rise on the commercial building”, Ain Shams Engineering Journal, Volume 10, Issue 2, 2019, Pages 389-401,

[8] Christoforos Menos-Aikateriniadis, “Methods to identify broken neutral fault in LV distribution grids by using existing smart meters infrastructure”, Master of Science Thesis, KTH School of Industrial Engineering and Management, August 2019.

[9] CIGRE Task Force C6.04.02, “Benchmark systems for network integration of renewable and distributed energy resources”. Technical Brochure 575, 2014

[10] M. Mitolo, M. Tartaglia and S. Panetta, “Of International Terminology and Wiring Methods Used in the Matter of Bonding and Earthing of Low-Voltage Power Systems”, in IEEE Transactions on Industry Applications, vol. 46, no. 3, pp. 1089-1095, May-june 2010

[11] J. Csatár and A. Dán, “Neutral Voltage Comparison of Different Grounding Configurations and Calculation Methods in Multi-Grounded Low Voltage Network”, Periodica Polytechnica Electrical Engineering and Computer Science, vol 61, no. 1, pp. 77–81, 2017

[12] A. Rodríguez del Nozal, E. Romero-Ramos, and Á. L. Trigo-García, “Accurate Assessment of Decoupled OLTC Transformers to Optimize the Operation of Low-Voltage Networks”, Energies 12, no. 11: 2173, 2019.

[13] Lawrence, J. (1993). “Introduction to neural networks”. California Scientific Software.

[14] Ketkar, N. (2017). “Deep Learning with Python. A Hands-on Introduction”. Apress.