A generalizable and sensor-independent deep learning method for fault detection and location in low-voltage distribution grids

Sapountzoglou, Nikolaos; Lago, Jesus; De Schutter, Bart; Raison, Bertrand

DOI
10.1016/j.apenergy.2020.115299

Publication date
2020

Document Version
Final published version

Published in
Applied Energy

Applied Energy 276 (2020) 115299

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

A generalizable and sensor-independent deep learning method for fault detection and location in low-voltage distribution grids

Nikolaos Sapountzoglou[a,1,*], Jesus Lago[b,c,d,1], Bart De Schutter[b], Bertrand Raison[a]

[a] Univ. Grenoble Alpes, CNRS, Grenoble INP, G2Elab, 38000 Grenoble, France
[b] Delft Center for Systems and Control, Delft University of Technology, Delft, the Netherlands
[c] Algorithms, Modeling, and Optimization, Energyville, Genk, Belgium
[d] Energy Technology, Flemish Institute for Technological Research (VITO), Mol, Belgium

HIGHLIGHTS

- We propose a deep learning methodology to detect and localize faults in LV grids.
- The method is generalizable and not limited by the number of sensors.
- It is the first method to localize high-impedance faults in LV grids.
- An analysis of the hindering factors is presented.
- Deep neural networks are shown to outperform other methods from the literature.

ARTICLE INFO

Keywords:
Fault detection
Fault location
Low-voltage distribution grids
Smart grids
Neural networks
Deep learning

ABSTRACT

Power outages in electrical grids can have very negative economic and societal impacts rendering fault diagnosis paramount to their secure and reliable operation. In this paper, deep neural networks are proposed for fault detection and location in low-voltage smart distribution grids. Due to its key properties, the proposed method solves some of the drawbacks of the existing literature methods, namely a method that: 1) is not limited by the grid topology; 2) is branch-independent; 3) can localize faults even with limited data; 4) is the first to accurately detect and localize high-impedance faults in the low-voltage distribution grid. The generalizability of the method derives from the non-grid specific nature of the inputs that it requires, inputs that can be obtained from any grid. To evaluate the proposed method, a real low-voltage distribution grid in Portugal is considered and the robustness of the method is tested against several disturbances including large fault resistance values (up to 1000 Ω). Based on the case study, it is shown that the proposed methodology outperforms conventional fault diagnosis methods: it detects faults with 100% accuracy, identifies faulty branches with 83.5% accuracy, and estimates the exact fault location with an average error of less than 11.8%. Finally, it is also shown that: 1) even when reducing the available measurements to the bare minimum, the accuracy of the proposed method is only decreased by 4.5%; 2) while deep neural networks usually require large amounts of data, the proposed model is accurate even for small dataset sizes.

1. Introduction

Power outages can lead to serious consequences of both economic and societal nature, ranging from production loss to risk to health and safety [1,2]. Weather conditions [3], equipment failures [4], accidents and unpredictable events such as vandalism, hacking or equipment theft, are some of the causes of power outages. Faults in distribution grids account for 80% of the customer electricity interruptions, with single-phase-to-ground faults being the most frequent type of fault and three-phase faults the most severe one [5].

The system average interruption duration index (SAIDI), i.e. the average time a costumer has no electricity service, is used to measure the reliability of distribution grids. In Europe, in 2016, most countries presented a SAIDI inferior to 100 min per year per customer, with a clear improvement tendency over the last years [6]. This is mainly attributed to the fact that smarter and more efficient functionalities,

* Corresponding author.
E-mail address: nikolaos.sapountzoglou@g2elab.grenoble-inp.fr (N. Sapountzoglou).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.apenergy.2020.115299
Received 13 January 2020; Received in revised form 3 April 2020; Accepted 30 May 2020

introduced by the smart grid concept, are being integrated in distribution grids.

In a recent study [7], ENEDIS, the main French distribution system operator (DSO), in collaboration with ADEeF, the French DSO association, reported profits of €3.3 millions per year from the implementation of self-healing tools in a *medium voltage (MV)* smart distribution grid. At the same time, the benefits of just a situation awareness tool in a Finnish DSO substantially reduced the cost of power outages [8]. Based on these real examples, it is clear that fault diagnosis tools are not only needed for grid stability but they are also paramount to reduce costs and to increase profits.

According to the different definitions of a smart grid [9-12], one of its key features is the ability to self-heal with the aid of advanced metering and communication tools and intelligent monitoring, aiming at a more secure, cost-effective and reliable operation. The backbone of self-healing strategies are the fault detection and fault location processes. However, despite the serious effects of power outages described above, many utilities still rely on customer phone calls to detect or localize a fault [13]. As a result, researchers have long been trying to automatize these processes.

In this context, the available fault location methods can be divided in three main categories: a) the conventional methods, including impedance-based and traveling-wave methods, b) the knowledge-based methods that use artificial intelligence and c) hybrid methods. From the first category, impedance-based methods are the most widely used basically due to their simplicity. Traveling-wave methods, although initially applied to transmission systems, have been used for distribution systems as well. Reviews of the conventional methods are provided in [14-16]. In the second category, different aspects of artificial intelligence have been employed to tackle the fault location problem including: a) artificial neural networks [17-19], b) support vector machine neural networks [18,20] and c) fuzzy logic [21,22]. Finally, neural networks and fuzzy logic have been used in combination with conventional methods to create hybrid tools [23-25].

1.1. LV grid characteristics

The LV grid is the final link that connects customers with the distribution substation. Although initially designed to follow the “fit and forget" doctrine, DSOs face now several challenges because of the integration of renewable energy sources and the bidirectional flow of energy they entail. Particularly, due to the necessity of the installation of renewable energy sources to fight climate change, DSOs are being forced to shift their attention to the monitoring of the LV grid.

Compared to the MV distribution grid, the LV grid presents a more complex structure. In detail, the LV grid presents five big differences [26]:

- Radial structure with an increased number of branches.
- Multi-phase and unbalanced operation.
- Unbalanced distribution of loads and generation units both per phase and topologically.
- A big variety of conductors connecting the nodes of the grid with a wide range of resistance (R) and reactance (X) values.
- The resistive nature of distribution lines ($R/X > 1$).

As a result, despite the desire of the DSOs for automated fault detection and location techniques, the above characteristics in combination with the limited availability of sensors in LV grids, pose considerable obstacles in the development of such techniques.

1.2. Literature review

In spite of the reasons listed above, researchers have mostly focused on the development of fault detection and location methods for the MV grid, neglecting the more complex case of the LV grid. Particularly, only a handful of methods have been recorded for the LV distribution grid, with the oldest one dating back to just 2012.

In [27] a fault detection and location method based on a current phase-angle difference analysis is proposed; in terms of fault location, the method is very limited as only the faulty sector is identified without a precise estimation of the fault location. Another approach was tested in [28] where the operational status of smart meters is taken into account to determine the fault location area. Moreover, in [29] a method to detect and localize non-technical losses was developed by comparing current values from smart meters with the current value at the transformer level; an estimation of the non-technical losses, e.g. electricity theft, location is achieved with an accuracy of 85%. Furthermore, the authors of [30] propose the use of the current signal injection technique, a more conventional method, to monitor deviations of impedance characteristics in the nodes of the grid and to locate the faulty node/area. Another approach is described in [31], and it involves the use of the Park vector representation of the voltage sag produced after a short-circuit fault to identify the fault type but also to get an idea of the proximity of the fault location. A fault diagnosis method based on gradient boosting trees is proposed in [32] but it is limited to a branch identification and does not provide a distance estimation. In addition, an attempt to apply a graphic method in LV grids was made in [33] with the fault location process based on the extracted voltage profile across the faulty branch. Finally, a conceptual method of fault detection and location is proposed in [34] based on data from sparse sensors along the grid.

1.3. Motivation

To ensure the success of the energy transition, fault diagnosis methods are paramount to mitigate the reliability issues of renewable energy sources. In this context, methods that detect, identify and locate faults have been widely researched across several domains: photovoltaics [35,36], electric batteries and electric vehicles [37], thermal storage [38], wind farms [39], fuel cells [40,41], gas turbines [42], gas supply [43], thermal power plants [44], or even household appliances [45]. Moreover, on higher voltage levels, researchers have also tried to increase the reliability of the transmission grid [46] and to predict cascading effects and blackouts [47]. However, despite the success of fault diagnosis methods in all these applications, reliable and accurate methods for fault detection and location in LV grids are still missing.

In particular, one of the main issues of the existing fault detection methods is that they are grid-specific. As it is clear from the literature review, despite the generalization properties of the conventional methods for MV grids (impedance-based or travelling wave), none of them have been applied to LV grids. This is not just a coincidence: impedance-based methods present limited accuracy and identify multiple possible locations for a fault [13,16]; this problem renders them impractical for the case of LV grids where multiple branches exist. Similarly, traveling-wave methods can also be seriously affected by the presence of multiple branches, which hinder the distinction between waves [13,16]. Finally, both impedance-based and traveling-wave methods depend on line parameters that in LV distribution grids vary a lot; as a result, their accuracy would not be very good.

The existing methods, while they overcome the issues of conventional methods with the use of data-driven approaches, have several issues of their own. Particularly, while they are in general more accurate, they are very specific to the grid topologies under study as they cannot be easily generalized to new grids, and require a lot of data from smart meters. In addition, with the exception of two studies [29,33], the existing methods do not provide an estimation of the fault location, but instead they are limited to the identification of the faulty area/line.

Another problem with the existing fault detection and location methods in LV grids, is that they all consider a limited number of fault scenarios in their case studies. For example, in terms of fault locations, only one fault location is considered in [27,28,48], two in [30], five in

[29] and ten in [33]. As many parameters can influence the fault location process, such a small number of fault scenarios is not enough to validate the existing methods.

Finally, the last limitation in the current literature is that, even for the methods proposed for MV grids, most methods only focus on faults with low fault resistance values that rarely surpass the 100 Ω. Taking into account that faults with a higher fault resistance can occur, e.g. in the case of non-technical losses (fault resistances can be higher than 500 Ω) or when a downed conductor touches the earth (fault resistances reach up to 1500 Ω), the need for fault detection and location techniques that cover these cases, i.e. large fault resistances and LV distribution grids, is evident.

1.4. Contribution

In this paper, in order to fill the described scientific gap, a deep learning method is proposed that addresses all these issues. In detail, the contribution of the paper is fourfold and proposes:

1. A fault detection and fault location method that can perform three different tasks: detect the occurrence of a fault and identify a faulty feeder, identify a faulty branch, and localize the position of the fault.
2. A data-driven method that, while being very accurate, is not limited by the grid topology nor the number of sensors, but that can detect and localize faults independently from the grid structure or available information.
3. A method that can detect and localize faults even when the data are very limited. In particular, the average accuracy decreases only slightly in the case of extremely limited measurements.
4. The first method that can accurately detect and localize faults of high fault resistance values in the LV distribution grid.

As neural networks have proven to be more accurate than analytical methods [18,19,49,50] in solving the fault location problem in MV grids, four of these methods are used as a benchmark [18,48,50,51]. Moreover, as a direct comparison with methods designed for MV is not always fair, one conventional method for the LV case [33] is also considered.

The paper is organized as follows: first, Section 2 introduces the field of deep learning. Then, Section 3 defines the proposed method for fault detection and fault location. Next, Section 4 describes the considered case study to evaluate the proposed method. Section 5 presents the results and discusses the merits of the proposed method. Finally, Section 6 compares the accuracy of the proposed method with others in the literature.

2. Deep learning

As one of the aspects of this work is the use of *deep learning (DL)* and *deep neural networks (DNNs)*, in this section, a brief overview on this topic is provided.

2.1. Introduction to deep learning

In recent years, the research on neural networks has achieved several breakthroughs that have lead to what is now known as deep learning. In particular, due to these breakthroughs, the usage of neural networks whose depth is no longer limited to a single hidden layer is now possible. These deeper neural networks have systemically proven to be better at estimation problems in several applications due to their better generalization properties [52].

While this success of DL models initiated in computer science applications, e.g. image recognition [53], speech recognition [54], or machine translation [55], the benefits of DL have also spread in the last years to several energy-related applications [56-64]. Among these areas, time series forecasting is arguably the field that has benefited the most [56,59,61,63,64].

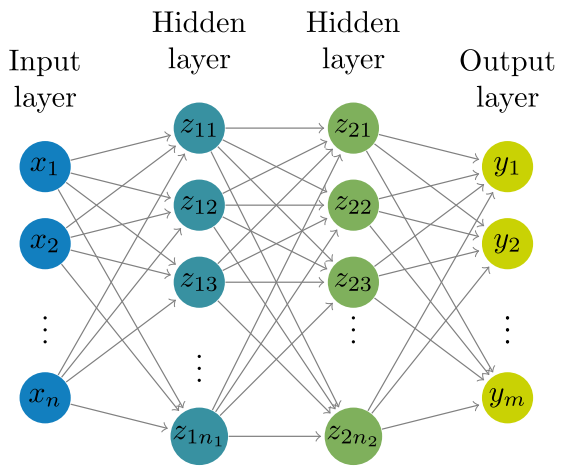


Fig. 1. Example of a DNN.

As mentioned in the introduction, despite the success of DL in all these energy-related areas, there has not yet been, to the best of our knowledge, an attempt to bring its ideas and models to the field of fault detection and fault location in LV distribution grids.

2.2. Deep neural networks

In general, a neural network is nothing else than a model $F(\mathbf{X}, \mathbf{W})$, with parameters $\mathbf{W}$, that uses some input features $\mathbf{X}$ in order to predict some variable of interest $\mathbf{Y}$ [62]. Thus, to use the neural network for a given task, i.e. to predict $\mathbf{Y}$, one only needs to gather a dataset $\mathscr{S}_{\mathscr{T}} = \{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^{N}$ and use this dataset to estimate the optimal parameters $\mathbf{W}^{\text{å}}$ that best fit the dataset. Let us define the input of a neural network by $\mathbf{X} = [x_1, \cdots, x_n]^T$ and the output by $\mathbf{Y} = [y_1, \cdots, y_m]^T$. Let us also define the number of neurons of the k^{th} hidden layer by n_k and by $\mathbf{z}_k = [z_{k1}, \cdots, z_{kn_k}]^T$ the state vector in the same layer. Using these definitions, a general DNN with two hidden layers can be represented as in Fig. 1.

In this model, the parameters $\mathbf{W}$ are the weights establishing the mapping connections between the different neurons of the network. In detail, the mapping equation of a general neuron i in the k^{th} layer is given by:

$$z_{ki} = f_{ki}(\mathbf{W}_{ki} \cdot \mathbf{z}_{k-1} + b_{ki}) \tag{1}$$

where f_{ki} represents the activation function of the neuron, $\mathbf{z}_{k-1}$ the values of the neurons of the previous layer, i.e. $k-1$, $\mathbf{W}_{ki}$ the matrix of weights establishing the connection between all the neurons of layer $k-1$ and neuron i in the k^{th} layer, and where b_{ki} is the so-called bias parameter of the neuron. Typical activation functions are the sigmoid function, the hyperbolic tangent function, or the rectified linear unit.

2.3. Training

The process of estimating the model weights is usually called training. Given the previously defined set $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^{N}$, the network training is done by solving a general optimization problem with the following structure:

$$\underset{W}{\text{minimize}} \sum_{k=1}^{N} gk(Y_k, F(X_k, W)) \tag{2}$$

where g_k is the problem-specific cost function. For grid fault diagnosis, this cost function varies depending on the specific task (more details on this will follow later in Section 3.3.4).

2.4. Hyperparameter and feature selection

In addition to the weights, the network has several parameters that need to be selected before the training process. Typical parameters include the number of neurons of the hidden layers, the number of hidden layers, the type of activation functions or the learning rate of the stochastic gradient descent method. To distinguish them from the main parameters, i.e. the network weights, they are referred to as the network hyperparameters.

Besides hyperparameters, DNNs need to perform another selection before the training process: the input features. Particularly, an adequate set of input features is key to obtain accurate models: a large set of irrelevant features will lead to inaccurate models that are hard to train; similarly, a small set that misses relevant input features will produce models with low accuracy.

The reason why both hyperparameter and features need to be selected before the training process is because these parameters cannot be optimized using a derivative-based method. In detail, while the best approach to optimize the weights of the neural network is to use some state-of-the-art gradient descent method, features and hyperparameters need to be optimized using black-box methods since the relation between features/hyperparameters and the output of the DNN cannot be defined (in general) using a differentiable function.

In practice, to optimize the hyperparameter and features, a search method is usually employed. This method iteratively selects different combinations of hyperparameters and features, trains the neural network with each of these different combinations, and finds the set of hyperparameters and features that lead to the most accurate model. To do so and to avoid data contamination, the dataset is usually divided into three subsets:

- Training dataset: the dataset used for training the DNN, i.e. estimating the weights.
- Validation dataset: the dataset used for optimizing hyperparameters and features.
- Test dataset: an out-of-sample dataset[2] that is used to evaluate the final method and compare against existing ones.

2.5. Further contributions of deep learning

An important thing to note is that deep learning is more than just deep neural networks. In particular, while the success of these models has been usually linked to the depth of the networks, the field has also benefited from a series of developments and contributions that have made possible the training of deep networks and the attainment of more accurate models. In this context, there are three key developments that have been crucial in the success of DL:

- Optimization algorithms: the traditional optimization algorithm for training neural networks, i.e. gradient descent method with back-propagation, had multiple limitations when training deep neural networks, e.g. computational cost or getting stuck at bad local minima. The appearance of several stochastic gradient descent methods, e.g. Adam [65], facilitated the training of deep networks, the attainment of better local minima, and the reduction of the computation cost of training.
- Regularization techniques: without appropriate regularazation, due to the large amount of parameters, DNNs can easily overfit the training data. To avoid that, new regularization techniques, e.g. dropout [66], had to be developed.
- Activation functions: training deep networks with standard activation functions, e.g. sigmoid and hyperbolic tangent, leads to networks that are not that accurate. The appearance of new activation functions, e.g. *the rectifier linear unit (ReLU)*, lead to more accurate neural structures.

As a result, when developing deep neural networks and other deep learning models, it is important to consider all these factors in order to obtain accurate networks that can be trained efficiently.

3. Fault detection and location method

The basis of the proposed algorithm is to use a DL model in order to solve the different fault diagnosis tasks. However, in order to obtain an accurate model with the desired properties, it is not enough to train a DL model with grid data. Particularly, the method has to be designed to be robust, to be grid-independent, and to be accurate even when the information is limited. To obtain such a method, careful consideration has to be paid to the model design.

3.1. Grid independence

One of the key properties of the proposed method is that it is independent of any grid topology, i.e. it can be estimated using data from multiple grids and branches and it is not restricted by the number of sensors nor the grid topology.

Standard data-driven models usually lack the above property as their inputs are based on the real measurements on the grid. As an example, if we consider a model that uses voltage measurements along the grid, it is clear that the number of inputs will vary for each branch, e.g. a branch with five measurements will provide five inputs while a branch with two measurements will provide two inputs. In this context, it is not easy to derive a generalizable model that can be applied to both branches. Instead, it is easier to derive a model for each branch.

While the described approach works reasonably well, it prevents data-driven methodologies to generalize to new grids. In particular, while they might perform well for a given grid topology, they might fail to locate faults when conditions change. In addition, due to their grid dependence, they cannot be used in other grids than the one where they have been estimated, and they always require new data when the models are deployed in a new grid. This obviously poses a problem as gathering new data is not always possible and, even when it is, it usually has associated costs.

In this paper, the proposed method avoids the issue of non-generalizability via two of its componenets: a) a simple pre-processing step and b) considering branch-independent inputs. In detail, any input to the model representing measurements along the grid/branch is pre-processed via an interpolation function to obtain a set of inputs that always has the same size. By doing so, the method becomes generalizable as the inputs of any grid/branch always have the same size. In addition, the feature set avoids branch specific features, e.g. branch length or branch resistance and reactance, so that the model can be applied to a different grid as its inputs do not dependent on the branch/grid topology. Finally, with these two components, not only does the method becomes generalizable, but the DNN becomes more accurate due to regularization. In detail, it has been empirically shown that by forcing DNNs to learn multiple related tasks, the performance and learning speed can be improved [62,67,68]. Hence, as the DNN is trained to solve different tasks, i.e. to identity faults from different branches/grids, its performance is expected to improve.

This interpolation function maps any set of measurements to a set of values representing the same measurements as if they were coming from equally spaced sensors inside each branch. In particular, if we consider the case of three interpolated measurements per branch, we would interpolate the real measurements to obtain one measurement at the beginning, another at the end of the branch, and a third one in the middle. While the distance between the sensors would be obviously

[2] A dataset that is never used during training and that is used to evaluate the accuracy of the method in unseen data. This is needed to ensure the method does not overfit.

equal within a branch, the interpolated distance would naturally change between branches. An example of this interpolation method is provided in Fig. 2, where branch 1 and 8 equipped with seven and five voltage sensors respectively, after the interpolation, end up both with five voltage measurements this time at fixed locations within each branch. More specifically, since in this case the desired number of measurements per branch was selected to be five, a choice that will be explained later, consequently the fixed interpolated sensor locations inside each branch were at the beginning, the middle and the end of each branch and at the intermediary locations of 25% and 75% of the branch length. As mentioned above, this interpolation method guarantees the uniformity of the input dataset and renders the method generalizable to other grids since it does not depend on the number of sensors per branch. In addition, it should be noted that a minimum of two measurements per branch are required for the interpolation step. Given the fact that one of those two can be the one at the substation level, at least one more along each branch would be required for the interpolation method to work. To further clarify this interpolation process the detailed algorithm is provided in Algorithm 3.

Algorithm 1 Creation of Voltage matrix – Interpolation

#i: branch number
#n: desired number of interpolated voltages
#V^m: *vector of voltage measurements for branch i*
procedure VOLTAGE_INTERPOLATION (V^{m}, i, n)

#Obtain locations of measured voltages in branch i
$X = [X_1, \cdots, X_k]$ = `sensorLocations`(i)

#Generate equally spaced locations X^{int}
#of the virtual/interpolated voltages
l = `readBranchLength`(i)
$X^{\mathrm{int}} = [0, \frac{l}{n-1}, \frac{2l}{n-1}, \cdots, l]$

#Generate vector of virtual/interpolated voltages
V^{int} = `interpolate`(V^{m}, X, X^{int})

Return: V^{int}

#n is the desired number of interpolated voltages
procedure GENERATE_VOLTAGE_MATRIX(n)
$V^{\mathrm{grid}} = []$
#B: total number of branches
for $i = 1$: B **do**

#Read measured voltages at branch i
#k: total number of nodes in branch i
$V^{\mathrm{m}} = [V_1^{\mathrm{m}}, \cdots, V_k^{\mathrm{m}}]$ = `readVoltages`(i)

#Generate virtual/interpolated voltages
$V^{\mathrm{int}} = [V_1^{\mathrm{int}}, \cdots, V_n^{\mathrm{int}}]$= `voltage_interpolation`($V^{\mathrm{m}}$, i, n)

#Append interpolated voltages to matrix
$V^{\mathrm{grid}} = \begin{bmatrix} V_1^{\mathrm{int}} \\ V^{\mathrm{grid}} \\ V_n^{\mathrm{int}} \end{bmatrix}$ ⋮

Return:V^{grid}

It is important to note that, unless the voltage decay is linear, the interpolation method adds an extra error to the measurements. This is however not important as: a) the error introduced by the interpolation method is smaller than the measurement error and b) the DNN is trained with the noisy data so the this noise is implicitly included in the model.

Besides the interpolation step the model considers branch-independent features. In particular, branch-specific features, e.g. branch length, type of conductors or branch resistance and reactance, are disregarded. By doing so, the model can be applied to a different grid as its inputs are not branch-specific.

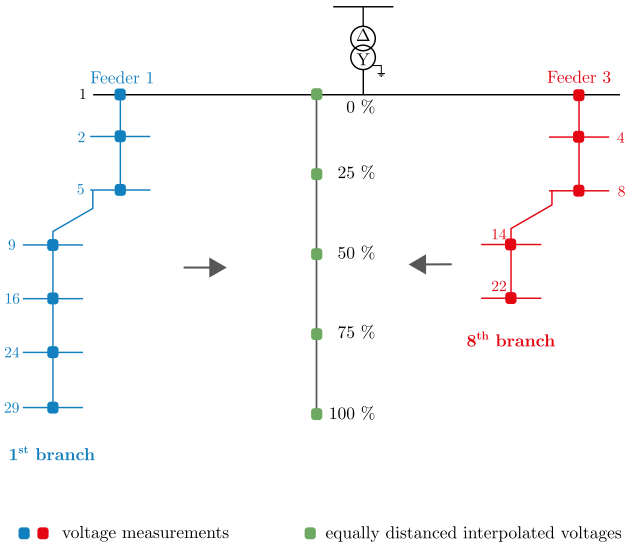


Fig. 2. Representation of the interpolation scheme to obtain equally spaced measurement.

3.2. Limited data

A second key property of the algorithm is that it can detect and locate faults even in a grid where the amount of data is limited, e.g. in a grid where measurements are only available at the beginning of the feeder and at each terminal point.

The main problem to remain accurate even when data is limited is the fact that data-driven methods tend to overfit under those conditions. Therefore, to accurately detect faults even with limited data, the method needs to extract the maximum amount of information from a given set of inputs, while avoiding the extraction of information that is too specific to that set of inputs.

To do so, a prediction model is proposed that has the potential of generalizing to different data, together with data from different sources. For the model that can generalize to different data, a DNN is considered. For data from different sources, data from different branches and data with different types of faults and noise levels are considered.

The motivation behind this is that it has been empirically shown that DNNs can learn features that can, to some extent, generalize across tasks [62,67,69]. In this case, by having a DNN that learns to locate faults under different situations, a model is obtained that can generalize and extract useful information even when data are limited. There are some possible hypotheses that could explain why this methodology improves the performance:

1. The simplest explanation is the amount of data: as more data are available, the DNN can learn more relevant features. Moreover, as the data are related, the DNN has more data to learn features that are common to all sources.
2. A second reason is regularization: by solving different tasks, the DNN is forced to learn features useful for all sources and to not overfit to the data of a single source.

3.3. Prediction model

While the diagnosis methodology involves different components, the key element of the proposed method is the prediction model itself. As motivated in the previous sections, a DNN is considered for multiple reasons, namely, its generalization capabilities, which allow to identify faults even with limited data, and the success of these type of models in

multiple applications.

However, to obtain a highly accurate model, employing a DNN is not enough. Particularly, the inputs, hyperparameters, and training of the DNN need to be carefully optimized.

3.3.1. Model inputs

One of the most important aspects to have an accurate prediction model is to select the optimal set of input features. In the proposed method, in order to always consider a set of optimal inputs, a feature selection method is considered. In particular, for any given diagnosis task, the method considers the information that the DSO might have available and then, during training, it performs a feature selection using the tree-structured Parzen estimator (TPE) [70]. The selection of this algorithm has been done because of its success in other energy-related applications [62-64]. To define this set of available input features, three aspects need to be taken into account: a) how the method detects and locates faults, b) the requirements of the method to be generalizable, and c) the fact that faults are time-dependent events.

For the first consideration, independently of the diagnosis task, the fact that the method detects faults by evaluating the status of each grid branch needs to be taken into account. Therefore, the set of input features has to contain data that models the status of each branch.

For the second consideration, as motivated before, the feature set has to avoid branch specific features, e.g. branch length or branch resistance and reactance, and employ interpolated measurements. In particular, for the model to be applied to a different grid, the input of the network cannot be branch-specific. Similarly, if the number of voltage measurements depend on the branch, any proposed method would be, not only grid-specific, but branch-specific. To address this, the proposed method interpolates the measurements along each branch to represent the same measurements as if they were coming from equally spaced sensors. Hence, for an application in a different grid it suffices to respect the format of the input features without modifying other elements of the method.

It should be underlined here that the scenario of identical branches is entirely theoretical as in practice they are heterogeneous in various ways: a) length, b) number and length of sectors, c) line characteristics of every sector and d) connected PV and loads per phase and node. On top of that, as mentioned earlier, the method does not take into consideration branch-specific parameters such as branch length. So even in the case of two branches within the same feeder with the same length, the algorithm will not be affected.

For the third consideration, the set of input features should be able to model, not only the status of each branch, but also the evolution of this status. As an example, to identity whether a fault has occurred, the method should not only consider the status of the grid under the fault, but also the status of the grid before the fault; by doing so, the method can more reliably identify faults as it has a representation of the grid under healthy conditions.

Based on these three considerations, to detect and locate a fault at time t, the set of all possible input features is defined as:

- N_i branch voltages at time t obtained by interpolation of the real branch measurements into five equally spaced measurements. Selecting N_i is a design choice. In the case study of Section 4, it was observed that $N_i = 5$ is a good enough parameterization.
- The current at time t at the beginning of the feeder where the branch is located.
- The generation and load in the distribution grid.
- The same four elements (voltage, current, generation and load) but 5 min before t.In particular, data are recorded on intervals of 50 ms. Then, to evaluate if a fault is happening at a given moment, the method considers the measurements of the voltage, current, and load/generation 5 min before. The goal of considering previous measurements is to have a recent snapshot of the grid during healthy operation. Thus, both pre-fault and post-fault values are part of the input dataset.

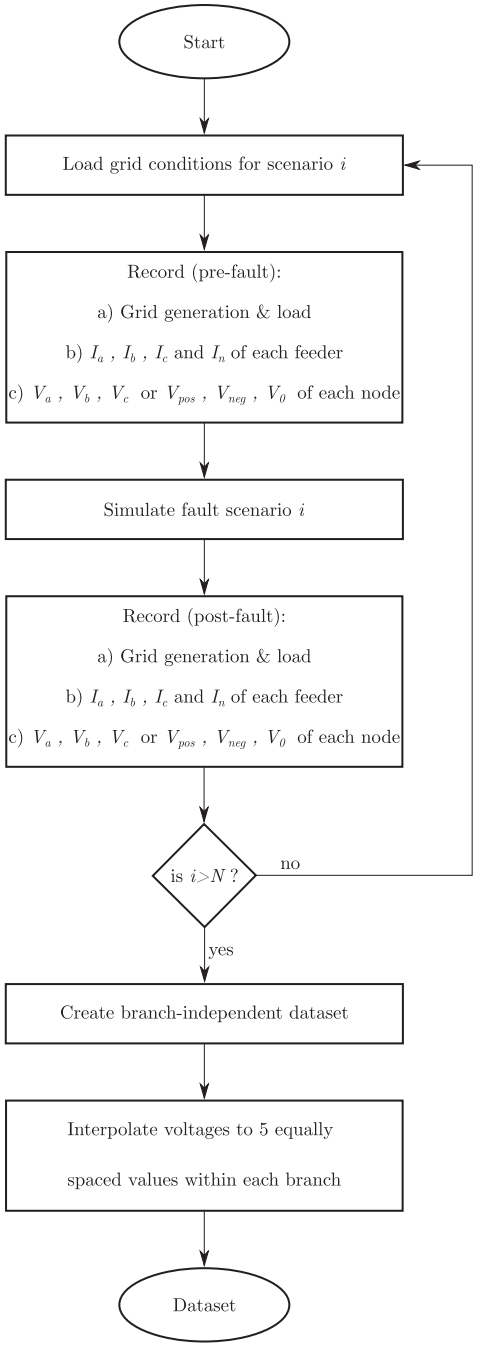


Fig. 3. Flowchart of the model input dataset creation process. N indicates the number of the total studied fault scenarios.

The complete process of creating the input feature dataset is described in Fig. 3.

Although, smart meters nowadays are capable of providing measurements every 1 to 10 s [71], due to restrictions imposed by the current telecommunication technology in the supervisory control and data acquisition system (SCADA), this measurement frequency decreases to only once every 15 min [72]. Even though, ideally, having a picture of the grid just the moment before the fault occurrence would increase the reliability of the method, the 5 min interval that was chosen above is justified as a realistic expectation in the near future.

As a final remark, it is important to note that the optimal set of features might vary from task to task, e.g. the optimal features for detecting faults might differ from the set for locating faults.

Table 1
Model hyperparameters and their possible ranges.

Hyperparameter	Range
Batch normalization	{0, 1}
Dropout	[0, 1]
Learning rate	$[10^{-4}, 10^{-1}]$
Activation function	{ReLU, softplus, tanh, sigmoid, selu, PReLU, LeakyReLU}
Number of layers	{1, 2, ⋯,7}
Neurons per layer	{25, 26, ⋯,400}
Normalization	{None, Uniform, Gaussian}

3.3.2. Hyperparameter optimization

Besides optimizing the inputs, to obtain an accurate model, it is paramount to optimize the DNN topology. For the proposed method, to obtain an optimal DNN structure, an optimal selection is performed using a large set of hyperparameters.

For the optimization, the hyperpamaters of the DNN are simultaneously optimized together with the input features, i.e. the tree-structured Parzen estimator algorithm [70] is employed to simultaneously obtain the optimal hyperparameters and the optimal features. The use of the tree-structured Parzen estimator to do this selection was motivated by the success of the method in other energy-related studies [63,64].

To maximize accuracy of the model, the method considers a large set of hyperparameters that models almost all possible DNN configurations. These hyperparameters are listed in Table 1 together with their possible ranges.

Most of the hyperparameters on the table are self-explanatory. The exception are the activation function and the data normalization. For the activation functions, we refer to [73] for a detailed definition. For the data normalization, *Uniform* uniformly normalizes the data to [0, 1] and *Gaussian* standardizes the data to follow a Normal distribution.

3.3.3. Training

In order to estimate the optimal network, i.e. optimal DNN weights $\mathbf{W}$, optimal input features and optimal hyperparameters, the same procedure is repeated:

1. Consider a dataset $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^N$ containing data pairs representing all possible input features $\mathbf{X}$ and the relevant output $\mathbf{Y}$ to be identified (see also Section 3.3.4).
2. Divide $\mathscr{S}_{\mathscr{T}} = \{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^N$ in three subsets: training (60%) + validation (20%) + test (20%).
3. Perform the feature and hyperparameter optimization using the training and validation dataset:
   - The training dataset is used to solve (2.3) and estimate the weights $\mathbf{W}$.
   - The validation dataset is used as an out-of-sample dataset to select the optimal features and hyperparameters.
4. Using the optimal network, evaluate its performance on the test dataset.

To solve (2), the Adam optimizer is used [65]. In addition, to avoid overfitting, i.e. estimate a DNN that fits perfectly to the training dataset but it cannot generalize to new data, the network is trained in combination with early stopping and out-of-sample data to evaluate its performance.

3.3.4. Network outputs and cost function

While the procedure to optimize the features, hyperparameters, and network weights is independent of the fault diagnosis task, the output of the DNN and its cost function varies with the problem. Several steps are necessary to pin point the exact location of the fault. First of all, the occurrence of the fault should be detected, then the faulty feeder and branch should be identified and finally, the exact distance of the fault within the faulty branch should be estimated. Following the above sequence three fault diagnosis tasks were designed for the DNN.

- **Fault detection and feeder identification:** detecting whether a feeder has a fault can be modeled with a classification network. In particular, for each branch-related input X, the output Y is defined to be 1 or 0 to respectively denote that the branch has a fault or that the branch has no fault and is in a healthy feeder (the data from those branches who are healthy but in a faulty feeder were excluded). Then, for training, the neural network can simply minimize the binary cross-entropy loss (standard loss for classifying between two classes). In real time, this network can simply be applied to all the branches of a feeder to identify whether there is a fault on the feeder.
- **Branch identification:** identifying the faulty branch in a faulty feeder is a very similar task to identifying a faulty feeder. However, instead of labeling 1 and 0 the branches in and out of faulty feeders, only branches within a faulty feeder are considered. For all these branches, the output is defined as 1 for the faulty branch and as 0 for the healthy branch in a faulty feeder (data from healthy branches in healthy feeders were excluded). In real time, once the feeder is identified with the fault detection and feeder identification network, this network is run to identify the faulty branch. As it is a binary classification problem, the network considers again the binary cross-entropy loss.
- **Fault distance estimation:** unlike the previous two tasks, estimating the distance at which the fault occurs is no longer a classification problem but a regression problem. To solve it, the subset of faulty branches is considered. Then, each output Y is defined as the distance at which the fault occurred. As the method has to be length-independent, the faulty distance is normalized between [0,1]. For the loss function, the standard mean squared error is considered. In real time, once the branch has been identified by the previous networks, this network indicates the fault distance.

3.4. Representation

To provide a better understanding of the method, Fig. 4 represents the different components of the proposed methodology and how they relate to each other.

3.5. Generalizability of the method

The generalizable property of the method does not imply that the method can be estimated in one grid and then use in the context of a completely different grid. In particular, while the method can be applied to different grid topologies or different grid characteristics, it is advisable to retrain the algorithm (if possible) for each grid considered. This ensures that the accuracy is maximized as the algorithm is tailored to the specific characteristics of the grid. While being less critical, it is also advisable to retrain the algorithm if a permanent change occurs in the grid, e.g. a line is upgraded.

In the same context, as the method is based on identifying faults per branch basis, no retraining is needed for the case of grid maintenance or servicing. Particularly, if a fault occurs when works are being carried out in another feeder or another branch, the method accuracy will not be affected as its accuracy only depends on the branch under fault. The latter remains unaffected by the servicing in another feeder/fault.

It is important to note that the retraining step is not mandatory: the method is still expected to detect faults when changes occur. Particularly, as it is trained with general inputs, e.g. normalized/interpolated voltages, some of the characteristics employed by the method to identify a fault, e.g. large voltage drop, would be independent of the grid characteristics. However, if not retrained, the

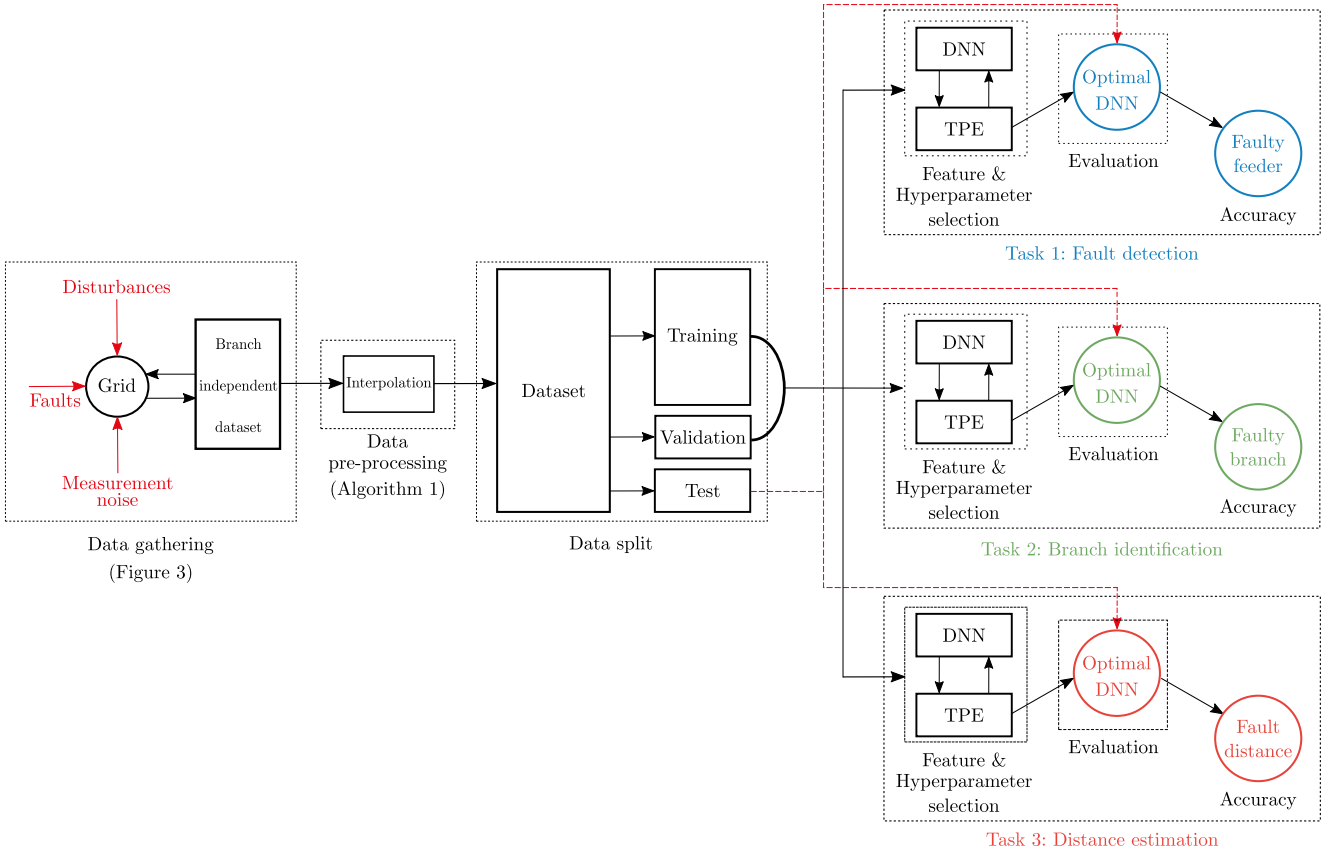


Fig. 4. Conceptual representation of the method.

accuracy of the method will likely decrease, with the drop in accuracy dependent on the severity of the changes in the grid.

4. Case study

As mentioned in the introduction, the proposed method is evaluated using a real LV distribution grid in Portugal. In this section, the case study and the experimental setup are properly defined.

4.1. Grid features

The considered three-phase-four-wire LV grid with a solidly grounded neutral has two distinct characteristics: a) *heterogeneity*: the grid consists of non-homogeneous distribution lines, i.e. conductors of different lengths, resistaces and reactances are used to connect the nodes and branches of the grid; b) *imbalance*: there is an asymmetrical distribution of the loads and *photovoltaic (PV)* systems both topologically and per phase. Moreover, a total of 48 consumers and 18 photovoltaic systems are connected to the grid via single phase connections. The LV grid schematic is presented in Fig. 5 and the described features are listed in Table 2.

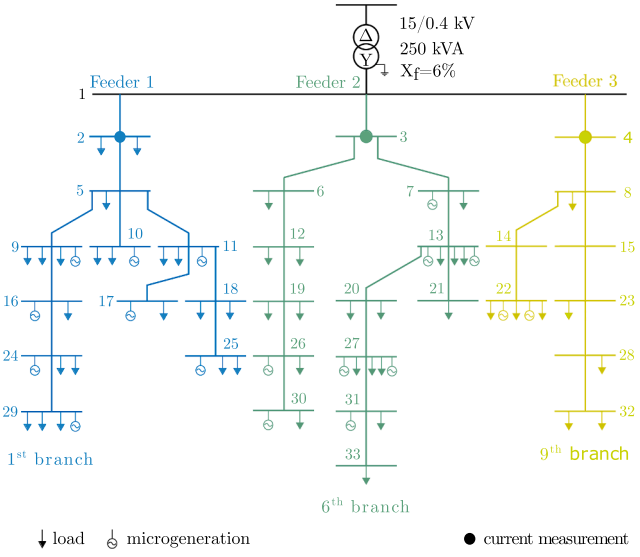


Fig. 5. Single line diagram of the LV distribution grid.

4.2. Grid measurements

As presented in Fig. 5, phase and neutral RMS current measurements are considered in the beginning of each feeder while voltage RMS measurements are considered to be available on every node of the grid. For the voltage, in addition to phase RMS measurements, a symmetrical component analysis is performed to compare which of the two is able to provide more accurate information for the fault location process.

It is important to note that, as explained in Section 3.1, any measurements along the grid, i.e. the voltages, are transformed to five equally distanced points by linear interpolation. Moreover, despite having sensors at every node in the grid, the method is also tested under the assumption of limited information.

4.3. Simulation environment

In order to perform the study, a realistic simulation framework of the real LV grid was employed. The simulation framework was provided by the company Efacec [29] as part of their software suite for modeling

Table 2
Grid characteristics: minimum and maximum values of branch length, conductor length and R/X ratio, current and voltage sensors, and total per phase ($Ph_{A,B,C}$) contracted and installed power.

Grid Characteristic		Value		Unit
Branch length		185–640		m
Conductor length		35–210		m
Conductor R/X		4.27–47.53		–
Current sensors		3		–
Voltage sensors		33		–
	Ph_A	Ph_B	Ph_C	
Installed load	75.90	96.60	89.70	kVA
Installed generation	17.90	15.96	23.27	kW

LV distribution grids. This simulator can be used for both normal and faulty operating conditions. The simulation environment uses the phasor simulation mode in order to reduce computation time as the voltage and current measurements are based on RMS values. In terms of data sampling, the environment has a configurable sampling frequency that for this study was chosen as 50 ms.

In addition, as smart meters are designed with a class 1 accuracy (± 1%) [29], the simulation environment adds Gaussian errors with a standard deviation of 1% to the current and voltage measurements.

4.4. Grid effects

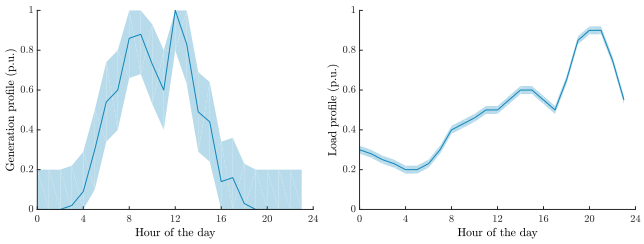
In order to perform the different experiments and to evaluate the method against different disturbances, the LV grid was simulated considering different effects:

1. *Fault location*: to evaluate the effect of the location of the fault, faults were simulated for nine locations within each one of the thirty two sectors, i.e. 288 fault locations were considered.
2. *Fault types*: in distribution grids, single-phase-to-ground faults are the less severe but at the same time the most frequent (they represent 70% of the fault cases). On the other hand, three-phase faults are the most severe but also less frequent (they represent only 5% of the fault occurrences [5]). Consequently, to assess the effect of the type of fault, the most frequent and most severe faults are studied. Particularly, for each of the 288 fault locations, 4 fault types were considered: the three single-phase-to-ground short-circuit faults (AG, BG and CG) and the three-phase short-circuit fault (ABC).
3. *Simultaneity factor*: since not all the loads of every consumer are going to be activated simultaneously, a variable simultaneity factor was considered [74]:

$$\mathrm{SF} = \frac{\max(D_{\mathrm{system}})}{\sum_{i=1}^{N} D_i^{\max}}, \tag{3}$$

where D is the load demand and N the number of loads. While SF stabilizes around 0.5 in residential areas of developed countries [74], the smaller the amount of consumers the more likely it is to suffer deviations. Since the considered grid only has 48 consumers, for each fault location and type, three values of SF were considered: 0.3, 0.5 and 0.8.

4. *Time of the day*: to account for the load and generation variability, the statistical distribution of the daily generation and load were considered (Fig. 6a and Fig. 6b). Then, for each fault location/type and SF, generation and load values for each hour of the day were sampled.
5. *Fault resistance*: to study the effect of the fault resistance, 8 random resistances were sampled for each possible combination of the other four effects. To sample fault resistances, a log-uniform distribution between 0.1 and 1000 Ω was considered.



(a) PV generation profile (b) Load profile

Fig. 6. Generation and load profiles for one day.

4.5. Data recording and dataset generation

To study the faults, the state of the grid was recorded 150 ms after the fault occurrence. This is a design choice to ensure that: a) that the faults are at the early stages of their steady-state and b) that no protective element acts as inverters are expected to isolate the generators at 200 ms after a fault occurs [75].

Moreover, based on the described values for the effects, the LV grid was recorded under 663,552 faulty scenarios:

$$\begin{aligned} &32 \ \text{sectors} \times 9 \ \text{faults per sector} \times \\ &4 \ \text{fault types} \times 3 \ \text{simultaneity factors} \times \\ &24 \ \text{load/gen. values} \times 8 \ \text{fault resistance values} \\ &= 663552 \ \text{scenarios}. \end{aligned} \tag{4}$$

Besides recording data under faults, the LV grid was also simulated to generate and record data representing its operating status under healthy conditions. Particularly, data representing its nominal working regime are also needed in order to train the DNN. To generate these healthy data, as the generation and load are the only possible variable effects, the grid was simulated during normal operating conditions for 65,000 generation and load profiles that were randomly sampled.

As a result, to conduct the experiments, the considered dataset contains 663,552 datapoints representing faulty conditions and 65,000 datapoints representing healthy operation. However, it is important to note that this is just the total dataset size as, depending on the experiment, the employed datasets to train the DNN are small subsamples of this large one (see Section 4.7 for further details).

4.6. Evaluation metrics

The proposed method is designed to solve three different tasks: fault detection and identification of a faulty feeder, identification of the faulty branch, and estimation of the location of the fault. Therefore, depending on the task, different metrics are needed in order to evaluate the accuracy and performance of the method.

4.6.1. Faulty feeder detection

For the first task, as it is a standard classification problem, two standard metrics are considered: the accuracy and the F1 score [76]. Let us define by tp the number of true positives, i.e. the number of times that a faulty feeder is correctly identified, by tn the number of true negatives, i.e. the number of times a healthy feeder is correctly identified, and by fp/fn the number of false positives/negatives, i.e. the number of times a faulty/healthy feeder is identified but the feeder is healthy/faulty. Then, the accuracy of the method is defined as:

$$\mathrm{Accuracy} = \frac{\mathrm{tp} + \mathrm{tn}}{\mathrm{tp} + \mathrm{tn} + \mathrm{fp} + \mathrm{fn}} \tag{5}$$

Similarly, the F1 score is defined as:

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{precision} \ \ \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}, \tag{6a}$$

$$\text{precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}, \tag{6b}$$

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}. \tag{6c}$$

4.6.2. Faulty branch identification

For the faulty branch identification, while it is also a classification problem, the resulting metrics are slightly different. In particular, the goal is to identify which of the grid branches is the one with a fault (the occurrence of the fault itself was already detected in the previous task). Therefore, the number of times a branch is correctly identified as faulty, i.e. tp, must necessarily be the same as the number of times that the other branches are identified as healthy, i.e. tn. Similarly, the number of fp must be the same as fn. Therefore, it holds that:

$$\text{Accuracy} = \text{F1} = \frac{\text{tp}}{\text{tp} + \text{fn}} = \text{recall} = \text{precision} \tag{7}$$

In other words, the accuracy is simply defined as the number of times the faulty branch is correctly identified divided by the total number of times a fault occurs.

4.6.3. Fault location

Since the location of the fault is provided as the distance of that location from the beginning of the branch, the accuracy of the method is evaluated in terms of standard fault distance estimation error [77]:

$$\text{distance error}(\%) = \frac{|d_{\text{estimation}} - d_{\text{actual}}|}{l_{\text{total}}} \cdot 100 \tag{8}$$

where $d_{\text{estimation}}$ and d_{actual} respectively represent the predicted and the real distances, and l_{total} is the total length of the branch.

4.7. Objectives of the study and experiments definition

In order to evaluate the robustness and accuracy of the method against all the different effects acting on the grid, seven different individual experiments are performed. In particular, the accuracy of the method and its variation are analyzed for different: a) depths of the DNN model, b) voltage measurements, c) resistance values, d) fault types, e) load demands, f) fault location, g) number of measurements, and h) dataset sizes.

4.7.1. Effect of DNN depth

One of the premises of using a DL model is the fact that deep models can generalize better and lead to more accurate predictions. Thus, as a first experiment, it is necessary to to validate that claim by analyzing the dependence of the accuracy with the DNN depth.

To analyze the accuracy of detecting faulty feeders, a dataset is built comprising the 65,000 datapoints representing healthy operation and 65,000 datapoints representing faults (the latter are randomly sampled from the larger dataset comprising 663,552 datapoints). For identifying the branch and locating the fault, the full dataset with 663,552 datapoints is employed. In all three cases, as defined in Section 3.3.3, the dataset is split into training, validation, and test datasets.

4.7.2. Effect of voltage measurement type

A common practice in the literature is to use positive, negative and zero voltage components instead of the phase voltages. In unbalanced operation, where negative and/or zero sequence components of the voltage and current might be present, the use of symmetrical components ensures that no information is lost. In this experiment, to validate this claim, the effect in accuracy of using the two types of voltage measurements (symmetrical components vs. phase measurements) is assessed.

The datasets used are the same as the ones defined in Section 4.7.1 for studying the effect of the DNN depth. As for the DNN depth, considering the results of the previous experiment, the current experiment is performed using a DNN with optimal depth, i.e. 3 hidden layers for branch identification and 4 hidden layers for fault location.

4.7.3. Effect of fault resistance value

As motivated in the introduction, while methods in the literature focus on low-impedance faults, high-impedance faults are harder to locate. More specifically, this difficulty derives from the fact that with an increase of the fault resistance, the amplitude of the currents flowing though the faulty branch is severely decreased. Those very small currents will in their turn decrease the voltage drop across the faulty branch, bringing the voltages to a level very close to that of normal operating conditions. To validate the importance of studying high-impedance faults and to show the relevance of the proposed method, the dependence of the accuracy with the fault resistance value is studied.

The datasets used are the same as for the other two experiments. However, to compare the dependence of the accuracy with regards to fault resistance, the datapoints are grouped based on their fault resistance value. In particular, as the fault resistances are sampled from a logarithmic uniform distribution, they are grouped into six datasets according to the following fault resistance intervals: [0.1, 1), [1, 10), [10, 50), [50, 100), [100, 500), [500, 1000).

For the DNN depth, 3 and 4 hidden layers are once again considered. For the type of voltage measurements, considering the results of the previous experiment, the positive, negative and zero voltage components are selected.

4.7.4. Effect of type of fault

The effects and characteristics of a fault do not only depend on the resistance value of the fault, but also on the type of the fault itself. Balanced and unbalanced faults can have different effects on the grid voltage, especially in the negative and zero sequence components where deviations from the normal operating conditions will be noticeable under unbalanced faults (single-phase-to-ground faults in this case) [78]. Therefore, to study the accuracy and robustness of the method, its performance across the four considered fault types is compared (see Section 4.4 for details).

The dataset is again the same as for the first two experiments. However, the datapoints are grouped based on the fault type. For the DNN depth and voltage measurements, again 3 and 4 hidden layers and positive, negative and zero voltage components are considered.

4.7.5. Effect of load demand

A fifth factor that impacts the effects and characteristics of a fault is the level of the grid load. As mentioned before (see Section 4.4 for details), the grid load depends on the statistical behavior of consumers, which is a function of the hour of the day and the simultaneity factor. The higher the load demand, the higher the regular voltage drop (under normal operation) will be across a branch [79]; this makes the voltage profile between normal operation during high load demand and faulty operation for high-impedance faults (low voltage drop during a fault) very similar, and in turn the fault diagnosis process more difficult.

It is important to note that this study focuses on the analysis of the load but not the generation because the load is the dominant factor. Particularly, being an inverter controlled source, the contribution of the PV units to the fault current is expected to be limited (the current is limited at 1.1–1.4 p.u. of normal operation during a fault) [78]. Consequently, the effect of an increased load demand is expected to dominate over that of increased penetration of PVs.

As the load depends on the simultaneity factor, the same dataset as before is used but with the datapoints grouped based on the three simultaneity factors, i.e. 0.3, 0.5 and 0.8. For the DNN depth and voltage measurements, again 3 and 4 hidden layers and positive, negative and zero voltage components are considered.

4.7.6. Effect of fault location

Another influencing factor that is identified in this study, is the location of the fault. Particularly, the fault depends on the branch itself and the distance within the branch. In detail, since the grid is heterogeneous (see Section 4.1), the longer a branch is the higher the number of different conductors connecting the nodes are; this in turn increases the heterogeneity of the grid and hinders the fault location process. At the same time, nodes further away from the beginning of the feeder will experience higher voltage drops than those located at the beginning [79]; thus, the location of the fault within the faulty branch itself is also influencing the accuracy of the method. For both cases, emphasis is given in the fault distance estimation task as it is more complex and vulnerable against these parameters.

In terms of the dataset, the full dataset is considered. For the DNN depth and voltage measurements, 4 hidden layers and positive, negative and zero voltage components are considered.

4.7.7. Effect of number of measurements

One of the main advantages of the method is that it can perform accurately even when the grid information is limited. To assess this claim and to show that the method can indeed perform well under the assumption of limited information, the accuracy of the method is tested when the only measurements available are at the beginning and at the end of each branch. Then, the performance is compared with the case of having voltage data at all intermediate nodes. As before, for the DNN depth and voltage measurements 3 and 4 hidden layers, and positive, negative and zero voltage components are considered.

4.7.8. Effect of dataset size

A potential argument against the proposed method is the fact that it might require large amounts of data. Particularly, as it is based on a DNN, it could be argued that the model is only accurate if it is estimated using a large dataset.

To refute this claim and to show that the method is accurate even for small dataset sizes, the method accuracy is also studied as a function of the amount of available data. In detail, the same large dataset is considered as for the previous experiment; then, a test dataset of 30,000 scenarios is randomly sampled using an uniform distribution. From the remaining datapoints not included in the test dataset, another uniform random sampling is performed to build subsets of different sizes. Finally, the performance of the method is evaluated across these datasets of different sizes: the DNN is trained with these differently-sized datasets and all models are evaluated using the defined test dataset. As before, for the DNN depth and voltage measurements 3 and 4 hidden layers and positive, negative and zero voltage components are considered.

4.8. Case study limitations

Five potential limitations of the case study are being addressed here. These limitations concern: the type of studied faults, the line charging current phenomenon, the earthing system of the grid, the use of flexible alternating current transmission system devices and the source of the data.

First of all, with regards to the studied faults, since the DNN is trained for short-circuit single-phase-to-ground and three-phase faults the presence of another type of fault (double-phase-to-ground or phase-to-phase) may affect the fault location method but not the fault detection one, as will be explained in the next section. In the case of arc faults a re-training of the DNN is required.

Secondly, in the simulation model, line charging currents, a phenomenon that depends on two parameters: a) the line length (the phenomenon starts having an impact after several kms) and b) the line capacitance, are completely negligible in low-voltage distribution grids. More specifically, in this simulation model, the two reasons that render this phenomenon negligible are: a) the fact that the maximum length of a branch is 640 m and b) as in all low-voltage distribution grids, in this case too, the distribution lines present a mainly resistive nature ($R/X > 1$) thus limiting the effect of the line capacitance on the charging currents.

Additionally, it is important to note that, without loss of generality and for the sake of simplicity, the case study is based on the most popular earthing strategy for LV distribution grids: the solidly grounded neutral, in this case, through a 40 Ω resistance. However, this does not mean that the proposed method is limited to this type of earthing system. As the method is generalizable, a new earthing type would only require a model re-estimation so that the model becomes tailored to the new earthing type.

Furthermore, in the considered grid model, the installation of flexible alternating current transmission system devices is not taken into account as it can prove a very expensive solution for low-voltage grids. The presence of such devices serves at improving the power quality and will mitigate the amplitude of the produced voltage sag during a fault. However, since the method is able to detect with 100% accuracy even very high impedance faults, in terms of fault detection, its accuracy will not be affected. Similarly for the distance estimation of low-impedance faults where the effect of such devices will be more visible, as long as this effect is considered in the modeling of the system and the training phase of the DNN it is not expected to affect significantly the accuracy of the algorithm.

Finally, as stated before, the last limitation of this case study is that the method was evaluated on data acquired from simulations based on a real distribution grid of Portugal and not on data from field tests.

5. Results and discussion

To analyze and discuss the results, the eight different experiments are presented individually. However, as the dependence of the accuracy with regards to the voltage type and fault type did not produce very interesting results, the analysis of these two experiments is presented in 8 and 9. In addition, for the sake of simplicity, the individual results of each experiments are not listed in full detail but instead summarized and explained briefly. Nonetheless, for the sake of completion, the complete results of all experiments are included in 10.

It is important to note that for each experiment the accuracy of the method is analyzed individually for two of the tasks that it addresses: branch identification and fault location. For the other task, i.e. fault detection, the results are summarized in a single section as the accuracy of the method for detecting faults does not vary across the experiments.

5.1. Fault detection

As mentioned before, the results for fault detection are not individually presented for each experiment since the accuracy of the method for detecting faults does not vary across experiments.

In detail, independently of the number of layers, the method achieved an 100% accuracy in detecting the fault occurrence and identifying the feeder under fault. This result is also observed across all fault resistance values, simultaneity factors, fault locations, number of measurements, and dataset sizes. When compared with the accuracy for the other two tasks, this result might seem surprising. However, it is not so: when a fault occurs in a feeder, all the branches of the feeder get disturbed. The voltage deviations across the branches of the faulty feeder and the current increase in the beginning of the faulty feeder provide the DNN with sufficient information to distinguish faulty from normal operation and to identify the faulty feeder from the healthy ones.

5.2. Effect of DNN depth

The first experiment to be discussed involves the method accuracy as a function of the DNN depth. Fig. 7 summarizes the results of this

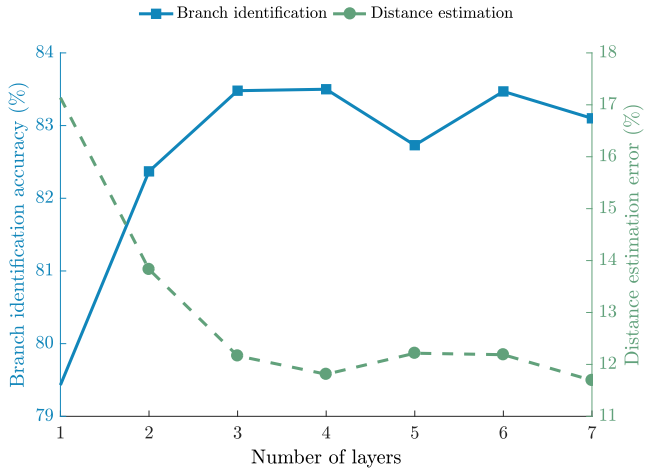


Fig. 7. Branch identification accuracy and fault location error for different number of layers.

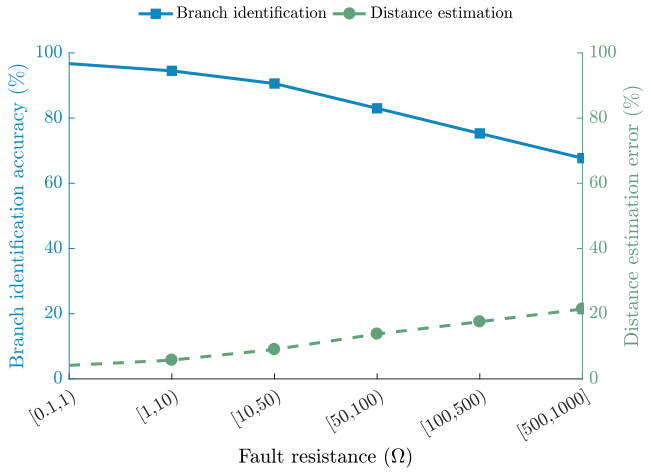


Fig. 8. Branch identification accuracy and fault location error for different fault resistance values using 3 and 4 hidden layers respectively.

experiment for both branch identification and fault location. For both tasks, it can be observed how the accuracy improves as the number of layers increases and then it plateaus when it reaches a certain depth.[3] This plateau, i.e. the optimal depth, is reached at 3 and 4 hidden layers for branch identification and fault location, respectively.

These results validate the premise that using a DL model is very important as it can generalize better and lead to more accurate predictions. In particular, it can be clearly observed how shallow networks, i.e. with 1–2 hidden layers, are not accurate enough and how deep models are needed in order to fully capture the fault dynamics in the LV grid.

Another interesting result (not displayed here but listed in 10) is that the depth becomes more important for higher values of fault resistance. Particularly, deeper neural networks are much more accurate at locating high-impedance faults than shallow networks: e.g., while the difference in branch identification accuracy between 1 and 3 hidden layers is only 0.17% for fault resistances between 0.1 and 1 Ω, this difference reaches a maximum of 7.33% for fault resistances between 10 and 50 Ω. However, for extra high fault resistance values, i.e. fault resistances between 500 and 1000 Ω, though still deeper networks are significantly better (the difference between 1 and 3 layers is 3.77%), the accuracy increase they offer is less significant compared to high fault resistance values (i.e. 10–500 Ω). This can in principle be expected as higher fault resistances represent cases where it is more difficult to draw conclusions of whether a branch is under fault and to estimate the location of the fault. In the extreme case of extra high fault resistance values (between 500 and 1000 Ω) even deeper networks have a harder time locating faults thus making the improvement less significant compared to high fault resistance values (i.e. 10–500 Ω).

5.3. Effect of fault resistance value

As a second experiment, the accuracy dependence w.r.t. to fault resistance is analyzed. Fig. 8 displays the results of this experiment for both branch identification and fault location. As can be observed, the larger the fault resistance, the lower the accuracy when identifying faulty branches and locating faults.

The first thing to note is that the proposed method is highly accurate. Particularly, even for high fault resistance values, the accuracy is reasonably high: the method identifies faulty branches with an accuracy between 95 and 70%, and estimates the fault location with an error between 5 and 20%.

[3] Note that the curves shapes are inverted as one metric represents the accuracy and the other the error.

These results validate the importance of developing methods for high-impedance faults. Particularly, while large resistance values are obviously harder to predict, the literature is limited to methods for low fault resistance values.

5.4. Effect of load demand

For the third experiment, the effect of the load demand is analyzed. Fig. 9 shows the results of this experiment for the branch identification task. As expected, a decrease of the branch identification accuracy with an increase of the simultaneity factor, i.e. an increase in the grid load, can be observed. This decrease is noticed across all the fault resistance values with the difference between an SF = 0.3 and SF = 0.8 ranging from 3.8 to 13.2%.

Similarly, the results for the distance estimation task are presented in Fig. 10. In this case, to show the dominance of the load over the generation, the method estimation error is not only plotted against the different values of simultaneity factor, but also against the hours of the day and the considered generation and load profiles. Regarding the hour of the day, three regions of load and generation combinations can be identified: low demand and generation - region (I); medium demand and high generation - region (II); high demand and low generation - region (III).

The first observation that can be made is that, as with the branch identification task, the accuracy of the method decreases with an increase of the simultaneity factor. In addition, the dominance of the load over the PV generation on the accuracy of the method is evident as all error estimation curves follow the trend of the load profile curve. In other words, the distance estimation error increases with an increase of the load demand but not with an increase of the generation.

These result are expected as with an increase of the activated loads in the grid, i.e. higher simultaneity factor, voltage drops across the branches during healthy operation will be more significant, thus bringing faulty and normal operation voltage values closer together. As a consequence, it becomes harder to identify a faulty from a healthy branch and locate a fault.

5.5. Effect of fault location

For the fourth experiment, the effect of the location of the fault in terms of the faulty branch and the distance within the faulty branch itself are presented. For the sake of simplicity, the focus is put on the fault distance estimation task as it is more complex and vulnerable against these parameters.

Fig. 11 shows the results in terms of the estimation error as a

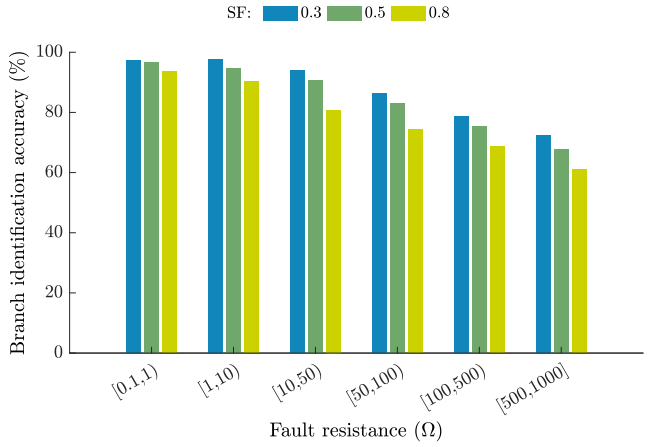


Fig. 9. Branch identification accuracy for different simultaneity factor values using 3 hidden layers.

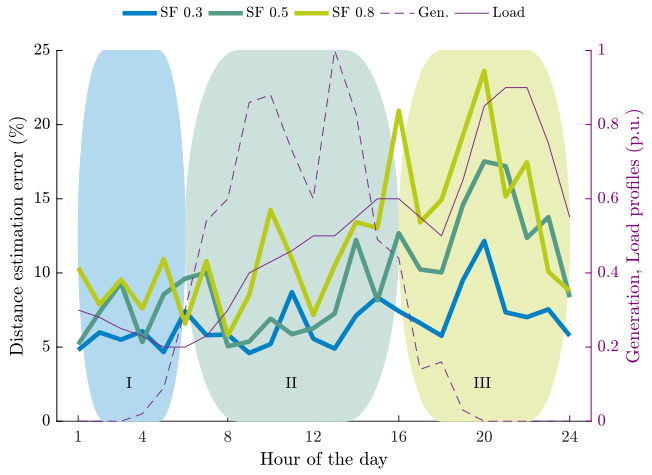


Fig. 10. Distance estimation error for different simultaneity factor values throughout the day using 4 hidden layers. The main generation (dashed purple line) and load profiles (solid purple line) are provided in the secondary y axis. Three distinct regions (I, II and III) of different performance are marked. The dominant effect of the load demand over the PV generation is shown as the SF curves follow the trend of the load curve and not so much that of generation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

function of the branch and the fault resistance. Branches are sorted in an ascending order with lengths ranging from 185 to 640 m. To give a different notion of the accuracy of the method, the distance estimation results are provided in meters instead of percentages. From these results, two important observations can be made: a) as expected, the fault distance estimation error increases with increasing branch length; b) this effect gets amplified with the increase of the fault resistance. This can be further observed from the maximum and minimum errors, which respectively occur at the longest branch with the largest fault resistance values and shortest branch with smaller fault resistance values.

Fig. 12 presents the estimation error as a function of the fault location within the branch and the grid load. As before, two observations can be made: a) as expected, the error increases as the fault gets located towards the end of the branch; b) this effect gets amplified with the increase of the activated loads in the grid which will increase the voltage drop across the branch. This can be further observed from the maximum and minimum errors: the minimum error is 5.68% and occurs for faults in the beginning of the branch with SF = 0.3; similarly, the maximum error is 17.26% and occurs for faults at the end of the branch with SF = 0.8.

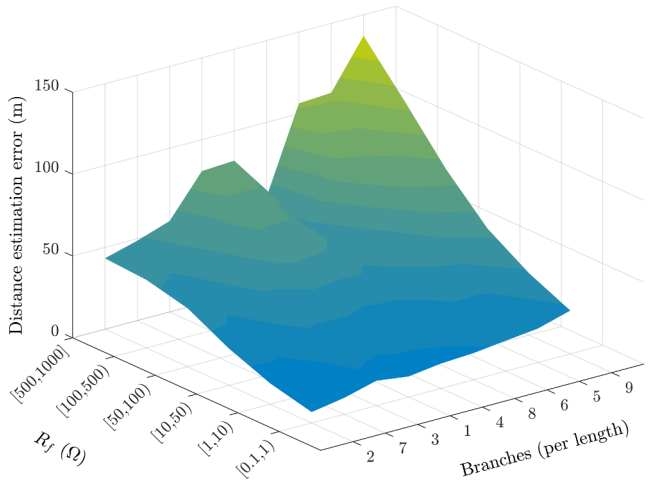


Fig. 11. Distance estimation error for different types of branches and fault resistance values using 4 hidden layers. Branches are sorted in an ascending order in terms of length.

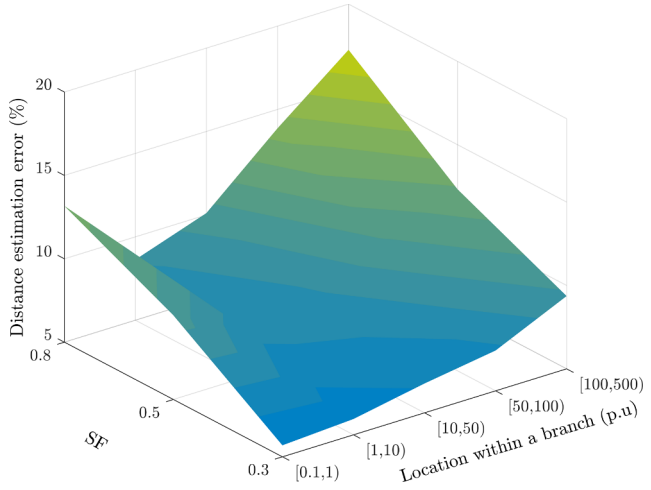


Fig. 12. Distance estimation accuracy for different locations of the fault within a branch.

The complexity of the LV grid compared to MV distribution grids or even transmission grids is evident, as mentioned in the motivation. The increased heterogeneity and imbalance that the LV grids present, hinder the fault location methods. However, even under the worst conditions the developed DL method yields a very good accuracy in locating the fault.

5.6. Effect of number of measurements

One of the key properties of the proposed method is that it can accurately locate faults even when the sensorial data along the grid are limited. To test this hypothesis, the accuracy of the method is evaluated assuming that sensors are only available at the beginning of each branch (one common voltage measurement for all branches at the substation level) and at the end of each branch (one voltage measurement at the terminal node of each branch). Figs. 13 and 14 respectively display the results of this experiment for branch identification and fault location.

As can be observed, even when the number of sensors is limited, the average decrease in branch identification accuracy is just 4.6% and the average increase in fault location error is just 4.4%. Considering that the accuracy variations when varying fault resistances are larger, it can be argued that these changes in accuracy are very minor. Therefore,

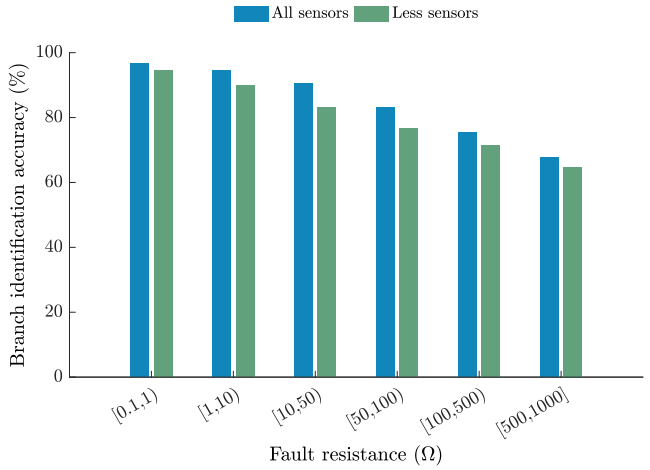


Fig. 13. Branch identification accuracy for less available measurements using 3 hidden layers.

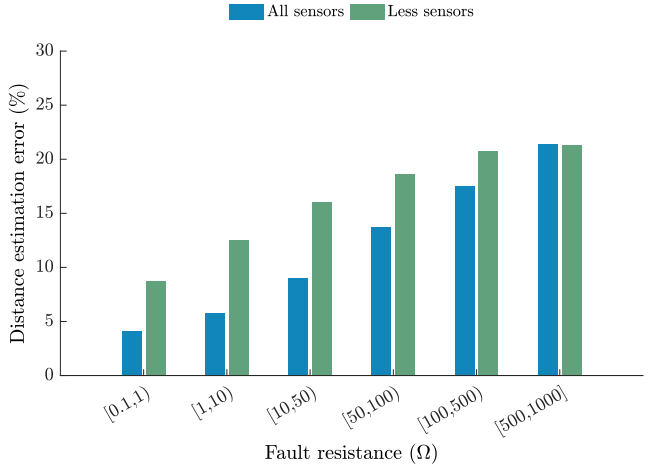


Fig. 14. Distance estimation error for less available measurements using 4 hidden layers.

these results show that the proposed method does indeed accurately locate faults even when the number of sensors is limited.

5.7. Effect of dataset size

As a last experiment, the accuray of the method with regards to the dataset size is analyzed. Fig. 15 displays the results of this experiment for both branch identification and fault location.

As it could be expected, the lower the dataset size, the lower the accuracy. Yet, the difference is not as large as one might expect: even when reducing the size of the training dataset by a factor of 100 (from 300,000 to 3000 datapoints), the average branch accuracy is still 73% and the fault location error is 13%.

Considering the accuracy of other methods from the literature (see Section 6), and considering that the smallest training dataset test is 10 times smaller than the test dataset (30000 datapoints), it is clear that the method performs reasonably well even when the amount of data is small.

These results are very important as they refute the most important argument against the proposed method, namely, that the method might be innacurate for small dataset sizes as it is based on a DNN.

5.8. Further case studies

It could be argued that using more that one case study is necessary

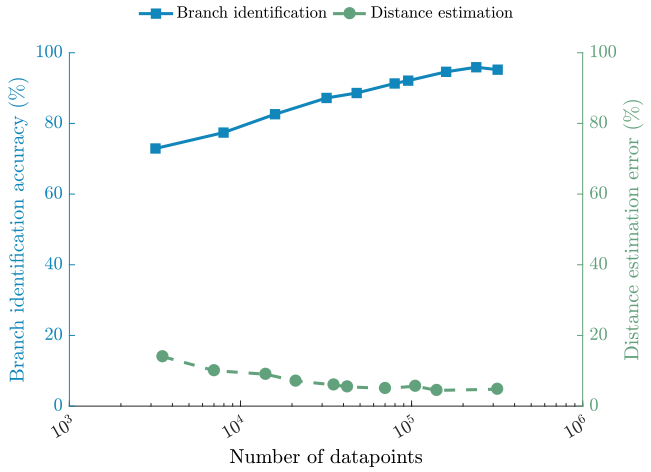


Fig. 15. Branch identification accuracy and distance estimation error for different dataset sizes using 3 and 4 hidden layers respectively.

to claim that the method is generalizable. However, the generalizable property of the method is defined due to the type of inputs that it requires, i.e. inputs that are non-grid specific and that can be obtained from any grid. In this context, while the accuracy of the method might depend on the case study, the generalizability of the method does not.

This is not to say that further case studies are not useful. Particularly, it would be useful to compare the accuracy of the method in different applications to study how the accuracy varies for different grids. However, such comparison is out of the scope of this paper as it requires new grid simulators, data that we do not have, and an analysis of the influence in the accuracy of each grid characteristic that would go beyond the current scope.

6. Comparison with literature methods

As a final comparison, the accuracy of the proposed method is analyzed and compared against other methods from the literature. It is important to note that, since the method is able to detect a fault occurrence with 100% accuracy, no comparative analysis is performed for this task. Instead, the comparative analysis is divided into two parts: one for faulty branch identification and the other for fault location estimation. The objective of this study is not only to provide a comparative analysis of the method performance, but to also identify the differences and limitations in the case studies from literature.

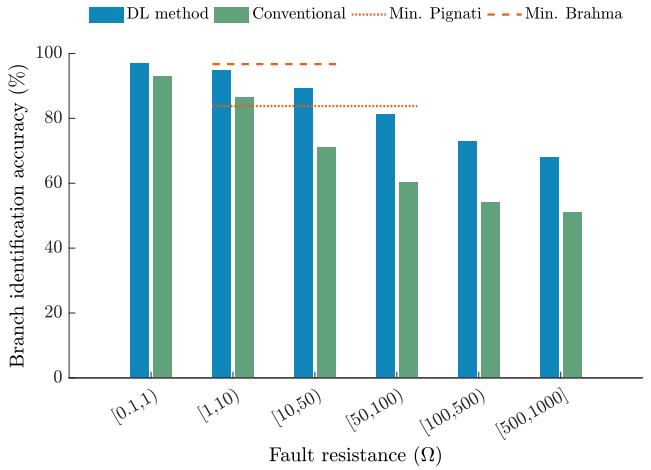


Fig. 16. Average branch identification accuracy for different fault resistance values, 3 hidden layers and an SF = 0.5. The results are compared with the LV conventional method and with two MV methods.

Table 3
Comparison of case studies for branch identification methods.

Parameters	Brahma [51]	Pignati [80]	This case study
Grid	12.4 kV (MV), U.S.A.	10 kV (MV), The Netherlands	400 V (LV), Portugal
Fault types	1ph-G, 2ph-G, ph-ph, 3ph	1ph-G, 2ph-G, 3ph	1ph-G, 3ph
Fault resistance (Ω)	1–5 (ph-ph, 3ph), 1–50 (1ph-G, 2ph-G)	1, 100, 1000[a]	0.1–1000 (63772 different values)
Fault location within the sector	0.05, 0.5, 0.95	0.25, 0.5	9 locations within each sector
Measurements, inputs	synchronized I, V at the point of common coupling of DG	I, V from phase measurement units on every node	RMS I at the beginning of the feeder, RMS V at every node
Noise in measurements	–	0.016% for voltage, 1.2% for current	1%

[a] 1000 Ω only for one case.

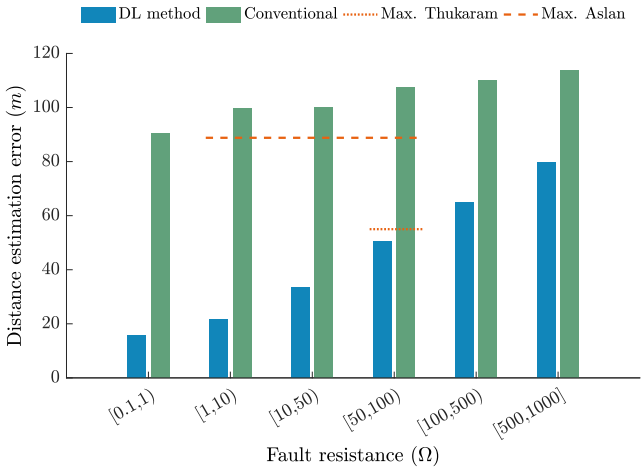


Fig. 17. Average distance estimation error for different fault resistance values, 4 hidden layers and an SF = 0.5. The results are compared with the LV conventional method and with two MV methods.

It is important to note that, since the case studies are different, the accuracy differences between the methods cannot be strictly measured and this comparison is purely qualitative. In addition, most of the methods from the literature are proposed for the MV grid case, which obviously differs significantly from the LV case.

6.1. Branch identification

Two methods designed for MV grids [51,80] and one [33] for LV grids are employed to compare the branch identification accuracy of the proposed method.

The first method [51] proposes a general faulty branch identification approach based on synchronized voltage and current measurements at the point of common coupling of the distributed generators. The second method [80] is based on phase measurement unit real-time state estimation and is able to identify the faulted lines of a distribution grid. Finally, the third method [33] compares the minimum voltage within each branch of the faulty feeder and considers as faulty, the branch with the highest voltage drop.

Fig. 16 presents the comparative analysis of the four branch identification methods mentioned above: the proposed DL method, the conventional method for LV grids, and the methods for the MV grid case. The first obvious observation from Fig. 16 is that the proposed method outperforms the conventional method by an average of 14.56%, effect that is even more visible for higher fault resistance values. At the same time, the DL method presents on average better accuracy than the minimum threshold of [51] and performs slightly worse than [80]. Yet, even with these similar accuracies, it can be argued that the proposed method is better than the MV methods as: a) the MV methods from the literature are only studied for a limited number of fault resistance values; b) the comparison with methods from the MV grid case is not exactly fair as the MV case is arguably less complex.

To provide a more thorough comparison, the differences between the case studies are presented in Table 3. From Table 3, the superiority of the case study of this paper in terms of considered influencing parameters (e.g. fault resistance, fault locations and measurement noise) and the number of the total considered fault scenarios is evident.

6.2. Distance estimation

As with the branch identification task, two methods designed for MV distribution grids [18,48] and one conventional one for LV grids [33] are employed to compare the fault location error of the proposed method.

In [18], neural networks in combination with a support vector machine method are used to create two fault location schemes. In [48], a single-end impedance-based method based on time-domain formulation is proposed. Finally, [33] proposes a graphic method based on the extracted voltage profile.

The results of this comparative analysis are gathered in Fig. 17. To have a fair comparison, the error metrics are displayed in terms of *meters* and not percentual errors as the lines in MV grids are orders of magnitude longer than those of LV grids. Particularly, long lines would increase the denominator of (9) and reduce the percentage error, which in turn would lead to deceiving results. The first conclusion that can be drawn from Fig. 16 is that the proposed DL method outperforms the conventional method by an average of 59.29 m. Additionally, not only does the proposed method perform better than a conventional LV method, but it also outperforms both MV methods.

To provide a better comparison, the differences between the case

Table 4
Comparison of case studies for fault location methods.

Parameters	Thukaram [18]	Aslan [50]	This case study
Grid	11 kV (MV), 52-bus distribution system	34.5 kV (MV), simplified distribution feeder	400 V (LV), Portugal
Fault types	1ph-G, 2ph-G, ph-ph, 3ph	1ph-G, 2ph-G, ph-ph, 3ph	1ph-G, 3ph
Fault resistance (Ω)	50, 60, 70, 80, 90, 100	2, 5, 10, 15, 20, 30, 40, 50, 60, 80, 100	0.1–1000 (63772 different values)
Fault location	0.2, 0.5, 0.8 (within each branch)	every 2.5 km (total length of 40 km)	288 locations (9 within each sector)
Measurements, inputs	I, V at substation and statuses of circuit breakers, relays	I, V at one end of the line	rms I at the beginning of the feeder, rms V at every node
Noise in measurements	–	–	1%

studies are presented in Table 4 where the superiority of the case study of this paper in terms of considered influencing parameters (e.g. fault resistance, fault locations and measurement noise) and the number of the total considered fault scenarios is demonstrated once again.

7. Conclusions

In this paper, a method based on *deep neural networks* (*DNNs*) has been proposed as a solution to detect and locate faults in *low-voltage* (*LV*) smart distribution grids. The method improves upon the existing literature by addressing four limitations of existing methods: a) not being limited by the grid topology but instead being generalizable to other grid structures; b) being the first method to accurately detect and localize high-impedance faults in the LV distribution grid; c) being able to accurately detect and localize faults even when data and the number of sensors are limited; d) being able to perform not one but three different identification tasks: detect a fault occurence and simultaneously identify a faulty feeder, identify a faulty branch, and locate the position of the fault.

To test the accuracy of the method and to study its main properties, a LV distribution grid in Portugal has been considered. Moreover, to evaluate the robustness of the method, its accuracy has been measured against different effects: a) depth of the DNN model, b) voltage measurements, c) fault resistance values, d) fault types, e) load demand, f) fault location, g) number of measurements, and h) dataset sizes.

Using the experimental results, it is shown that the proposed method excels in detecting and locating faults as it can detect a faulty feeder with a 100% accuracy, identify a faulty branch with a 84% average accuracy, and estimate the actual fault location within a branch with a average error of 12%. In addition, the method accuracy is compared against other literature methods and it is shown that the proposed method can obtain a similar or better performance than state-of-the-art approaches.

Based on the several simulation experiments the main properties of the method are demonstrated: a) the method is accurate even when the number of sensors is limited as even in the extreme case of having only two available measurements per branch the method accuracy is decreased by only 4–5%; b) the method accurately estimates the location of high-impedance faults; c) the method can be used even when the available datasets are small. As a final result, the importance of deep models is demonstrated by showing how the depth of the DNN plays a significant role in obtaining accurate results.

As future research, the case study will be expanded to the case of double-phase-to-ground faults and phase-to-phase faults. Moreover, the method will be tested against different grid topologies.

CRediT authorship contribution statement

Nikolaos Sapountzoglou: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Jesus Lago:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Bart De Schutter:** Supervision. **Bertrand Raison:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 675318 (INCITE). Thanks are also due to Mr. Konstantinos Kotsalos and the company Efacec in Maia, Portugal for providing us with the available data of a real semi-rural LV distribution grid of Portugal.

Appendix A. Effect of voltage measurement type

In this section, the accuracy dependence with regards to the type of voltage measurements is analyzed. Fig. A.18 displays the results of this experiment for branch identification and fault location. As can be observed, using positive, negative and zero voltage components instead of the more standard phase voltages leads to a more accurate identification of faulty branches and location of faults.

It is important to note that, while the difference in accuracy between the two types of measurements is minor, these results do validate the proposed hypothesis. Namely, that the use of symmetrical components ensures that no information is lost during unbalanced operation, and that as a result more accurate results can be obtained.

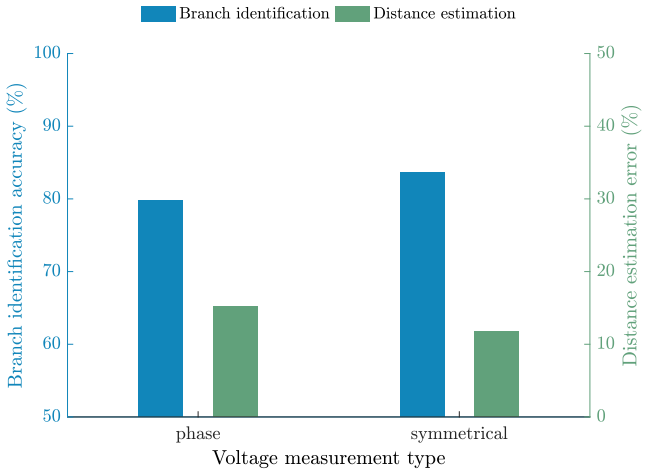


Fig. A18. Average branch identification accuracy and fault location error for different voltage measurement types using 3 and 4 hidden layers respectively.

Appendix B. Effect of type of fault

In this section, the accuracy dependence with regards to the type of fault is analyzed. Fig. B.19 displays the results of this experiment for both branch identification and fault location. As can be observed, for single-phase-to-ground faults the accuracy is independent of the underlying faulty phase. However, three-phase faults are easier to locate as their effect on the grid is more severe.

While this result might seem surprising (one might expect the accuracy to depend on the type of fault), it simply validates the robustness of the proposed method. Particularly, the proposed method is able to accurately identity all faults independently of the underlying cause for the fault. This also proves that the method is immune to the unbalanced per phase distribution of loads and PVs as they do not seem to affect the results of the single-phase-to-ground faults.

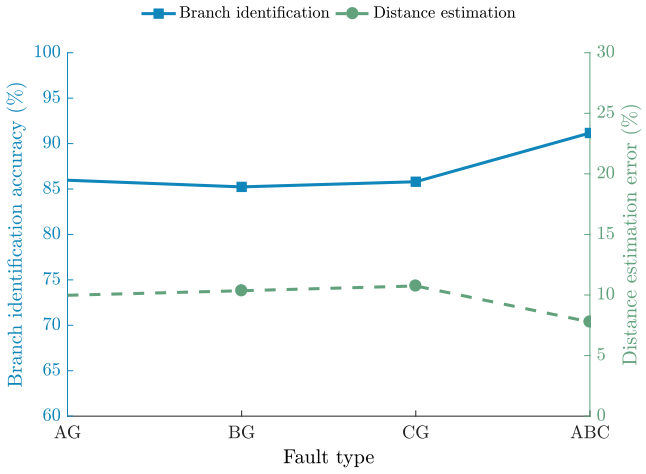


Fig. B19. Average branch identification accuracy and fault location error for different fault types using 3 and 4 hidden layers respectively.

Appendix C. All experimental results

In this Appendix, we provide a comprehensive list of the obtained results. In detail:

- Tables C5 and C6 respectively represent the dependency of the branch identification accuracy and the distance estimation error w.r.t. the DNN depth.
- Tables C7 and C8 respectively represent the dependency of the branch identification accuracy and the distance estimation error w.r.t. the voltage measurement type.
- Tables C9 and C10 respectively represent the dependency of the branch identification accuracy and the distance estimation error w.r.t. the fault resistance and the grid load.
- Tables C11 and C12 respectively represent the dependency of the branch identification accuracy and the distance estimation error w.r.t. the fault type and the grid load.
- Tables C13 and C14 respectively represent the dependency of the branch identification accuracy and the distance estimation error w.r.t. limiting the number of sensors in the grid to the minimum.
- Tables C15 and C16 respectively represent the dependency of the branch identification accuracy and the distance estimation error w.r.t. the size of the training dataset.
- Finally, Tables C18 and C17 list the dependency of the distance estimation error w.r.t. to the fault location. Particularly, Table C17 depicts the dependency w.r.t. the normalized fault location within a branch and Table C18 the dependency w.r.t. the faulty branch.

Table C5
Effect of number of layers on the branch identification accuracy. The metric is computed as the average across all simultaneity factors with the use of symmetrical components as measurement tool.

Fault	Branch identification accuracy (%) Number of layers							Difference
resistance	1	2	3	4	5	6	7	L 3 - L 1
[0.1, 1]	95.73	95.07	95.90	95.83	96.17	95.87	96.03	0.17
[1, 10)	92.23	93.17	94.13	94.37	94.50	93.77	94.30	1.90
[10, 50)	83.30	87.43	88.47	88.10	87.47	87.83	87.10	5.17
[10, 50)	73.87	80.63	81.20	81.30	80.07	81.37	81.10	7.33
[100, 500)	68.07	72.60	74.20	74.13	73.00	74.00	73.37	6.13
[500, 1000]	63.20	65.33	66.97	67.57	65.10	67.97	66.53	3.77
Average	79.40	82.37	83.48	83.55	82.72	83.47	83.07	4.08

Table C6
Effect of number of layers on the fault distance estimation error. The metric is computed as the average across all simultaneity factors with the use of symmetrical components as measurement tool.

Fault	Distance estimation error (%) Number of layers							Difference
resistance	1	2	3	4	5	6	7	L 1 - L 4
[0.1, 1]	7.51	6.13	4.45	4.38	4.23	3.95	3.68	3.13
[1, 10)	10.26	7.99	6.11	5.73	5.54	5.56	5.22	4.53
[10, 50)	15.51	11.74	9.85	9.21	9.38	9.92	9.38	6.30
[10, 50)	20.16	15.56	13.72	13.49	14.23	14.25	13.95	6.67
[100, 500)	23.43	19.07	17.67	17.16	18.12	18.17	17.41	6.27
[500, 1000]	25.95	22.50	21.20	20.90	21.79	21.27	20.51	5.05
Average	17.14	13.83	12.17	11.81	12.22	12.19	11.69	5.33

Table C7
Effect of the voltage type on the branch identification accuracy. The metric is computed as the average accuracy for 3 hidden layers and an SF = 0.5.

Fault	Identification accuracy (%)	
Resistance	Phase	Symmetrical
[0.1, 1]	91.6	96.7
[1, 10)	89.9	94.5
[10, 50)	83.8	90.6
[50, 100)	77.1	83.0
[100, 500)	70.7	75.3
[500, 1000]	65.6	67.7
Average	79.78	83.90

Table C8
Effect of the voltage type on the distance estimation error. The metric is computed as the average error for 4 hidden layers and an SF = 0.5.

Fault	Estimation Error (%)	
Resistance	Phase	Symmetrical
[0.1, 1]	8.01	4.15
[1, 10)	9.86	5.76
[10, 50)	13.09	9.04
[50, 100)	18.02	13.77
[100, 500)	20.38	17.53
[500, 1000]	21.58	21.45
Average	79.78	83.90

Table C9
Effect of the fault resistance and the load demand on the branch identification accuracy. The metric is computed as the average accuracy for 3 hidden layers and symmetrical components as measurement tool.

Fault	Identification accuracy (%) SF			
resistance	0.3	0.5	0.8	Avg
[0.1, 1]	97.4	96.7	93.6	95.9
[1, 10)	97.5	94.5	90.4	94.1
[10, 50)	94.0	90.6	80.8	88.5
[50, 100)	86.3	83.0	74.3	81.2
[100, 500)	78.7	75.3	68.6	74.2
[500, 1000]	72.3	67.7	60.9	67.0

Table C10
Effect of the fault resistance and the load demand on the distance estimation error. The metric is computed as the average error for 4 hidden layers and symmetrical components as measurement tool.

Fault	Estimation error (%) SF			
resistance	0.3	0.5	0.8	Avg
[0.1, 1]	2.41	4.15	6.58	4.38
[1, 10)	2.55	5.76	8.87	5.73
[10, 50)	5.19	9.04	13.39	9.21
[50, 100)	10.16	13.77	16.55	13.49
[100, 500)	14.23	17.53	19.71	17.16
[500, 1000]	19.03	21.45	22.23	20.90

Table C11
Effect of the type of fault and load demand on the branch identification accuracy. The metric is computed as the average accuracy for 3 hidden layers and symmetrical components as measurement tool.

Fault	Identification accuracy (%) SF			
type	0.3	0.5	0.8	Avg
AG	89.8	86.8	81.3	86.0
BG	89.9	85.8	80.0	85.2
CG	89.5	87.1	80.8	85.8
ABC	94.1	92.3	87.1	91.2

Table C12
Effect of the type of fault and load demand on the distance estimation error. The metric is computed as the average error for 4 hidden layers and symmetrical components as measurement tool.

Fault	Estimation error (%) SF			
type	0.3	0.5	0.8	Avg
AG	9.75	7.17	13.02	9.98
BG	10.36	7.25	13.45	10.35
CG	11.28	7.61	13.36	10.75
ABC	7.98	5.14	10.21	7.78

Table C13
Effect of limiting the number of sensors on the branch identification accuracy. The metric is computed as the average accuracy for 3 hidden layers, an SF = 0.5 and symmetrical components as measurement tool.

Fault	Identification accuracy (%)	
Resistance	All sensors	Limited sensors
[0.1, 1]	96.7	94.4
[1, 10)	94.5	89.9
[10, 50)	90.6	83.2
[50, 100)	83.0	76.7
[100, 500)	75.3	71.3
[500, 1000]	67.7	64.8
Average	84.6	80.1

Table C14
Effect of limiting the number of sensors on the distance estimation error. The metric is computed as the average error for 4 hidden layers, an SF = 0.5 and symmetrical components as measurement tool.

Fault	Estimation error (%)	
Resistance	All sensors	Limited sensors
[0.1, 1]	4.15	8.72
[1, 10)	5.76	12.52
[10, 50)	9.04	16.08
[50, 100)	13.77	18.60
[100, 500)	17.53	20.73
[500, 1000]	21.45	21.28
Average	11.95	16.32

Table C15
Effect of the size of the training dataset on the branch identification accuracy.The metric is computed as the average accuracy for 3 hidden layers, an SF = 0.5 and symmetrical components as measurement tool.

Number of training datapoints	Branch identification accuracy (%)
3184	72.9
7961	77.4
15,922	82.6
31,845	87.2
47,767	88.6
79,612	91.3
95,535	92.1
159,225	94.6
238,838	95.9
318,451	95.2

Table C16
Effect of the size of the training dataset on the distance estimation error.The metric is computed as the average error for 4 hidden layers, an SF = 0.5 and symmetrical components as measurement tool.

Number of training datapoints	Distance estimation error (%)
3518	13.99
7037	10.04
14,074	9.02
21,112	7.09
35,186	6.01
42,224	5.46
70,373	5.01
105,560	5.63
140,747	4.45
318,451	4.79

Table C17
Effect of the normalized fault location within a branch on the distance estimation error. This metric is computed independently for each fault resistance interval as the average error across all simultaneity factors, for 4 hidden layers and with symmetrical components as measurement tool.

Fault	Estimation error (%) SF			
type	0.3	0.5	0.8	Avg
[0, 0.2)	5.69	10.04	13.17	9.63
[0.2, 0.4)	5.95	7.96	8.67	7.53
[0.4, 0.6)	6.76	8.28	10.11	8.38
[0.6, 0.8)	7.42	10.23	13.87	10.51
[0.8, 1]	9.37	12.33	17.28	12.99

Table C18
Effect of the fault location in terms of the faulty branch on the distance estimation error. This metric is computed independently for each fault resistance interval as the average error across all simultaneity factors, for 4 hidden layers and with symmetrical components as measurement tool.

Faulty	Branch	Distance estimation error (m) Fault resistance						
branch	length (m)	[0.1, 1]	[1, 10)	[10, 50)	[50, 100)	[100, 500)	[500, 1000]	Avg
2	185	6.62	12.62	22.54	35.65	41.61	43.31	27.06
7	285	9.99	13.27	21.44	30.06	38.43	48.80	27.00
3	290	15.23	21.04	32.23	43.05	51.87	55.60	36.50
1	330	12.55	19.15	32.22	55.29	65.13	80.77	44.19
4	340	15.19	19.21	29.50	52.99	65.05	81.92	43.98
8	390	16.25	20.99	30.28	43.56	47.23	54.43	35.46
6	455	18.54	25.34	38.12	44.61	79.89	106.35	52.14
5	470	20.79	25.14	42.06	70.67	85.97	107.54	58.70
9	640	26.48	37.85	53.63	78.29	108.37	137.07	73.62

References

[1] Linares P, Rey L. The costs of electricity interruptions in Spain. Are we sending the right signals? Energy Policy 2013;61:751–60. https://doi.org/10.1016/j.enpol.2013.05.083.

[2] Meier A, Ueno T, Pritoni M. Using data from connected thermostats to track large power outages in the United States. Appl Energy 2019;256:113940https://doi.org/10.1016/j.apenergy.2019.113940.

[3] Jufri FH, Widiputra V, Jung J. State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies. Appl Energy 2019;239:1049–65. https://doi.org/10.1016/j.apenergy.2019.02.017.

[4] Tsianikas S, Zhou J, Birnie DP, Coit DW. Economic trends and comparisons for optimizing grid-outage resilient photovoltaic and battery systems. Appl Energy 2019;256:113892https://doi.org/10.1016/j.apenergy.2019.113892.

[5] Gönen T. Electric power distribution engineering. Boca Raton: Taylor & Francis; 2014.

[6] CEER. Report on Regulatory Frameworks for European Energy Networks, Tech. Rep. C18-IRB-38-03. Brussels, Belgium: Council of European Energy Regulators; 2019.

[7] Economic assessment of smart grids solutions: Analysis carried out by the distribution network operators, Executive Summary 2017, ENEDIS, ADEeF; 2017.

[8] Kupila T, Ihonen T, Keränen T, Anttila L. Efficient coordination in major power disruption. CIRED - Open Access Proc J 2017;2017(1):1064–7. https://doi.org/10.1049/oap-cired.2017.0628.

[9] Shabanzadeh M, Parsa Moghaddam M. What is the smart grid? Definitions, perspectives, and ultimate goals. Power System Conference 2013. https://doi.org/10.13140/2.1.2826.7525.

[10] Zia MF, Elbouchikhi E, Benbouzid M. Microgrids energy management systems: A critical review on methods, solutions, and prospects. Appl Energy 2018;222:1033–55. https://doi.org/10.1016/j.apenergy.2018.04.103.

[11] Hussain A, Bui V-H, Kim H-M. Microgrids as a resilience resource and strategies used by microgrids for enhancing resilience. Appl Energy 2019;240:56–72. https://doi.org/10.1016/j.apenergy.2019.02.055.

[12] Fontenot H, Dong B. Modeling and control of building-integrated microgrids for optimal energy management – a review. Appl Energy 2019;254:113689https://doi.org/10.1016/j.apenergy.2019.113689.

[13] Bahmanyar A, Jamali S, Estebsari A, Bompard E. A comparison framework for distribution system outage and fault location methods. Electr Power Syst Res 2017;145:19–34. https://doi.org/10.1016/j.epsr.2016.12.018.

[14] IEEE guide for determining fault location on AC transmission and distribution lines, IEEE Std C37.114-2014 (Revision of IEEE Std C37.114-2004) (2015) 1–76 doi:10.1109/IEEESTD.2015.7024095.

[15] Mora-Flòrez J, Meléndez J, Carrillo-Caicedo G. Comparison of impedance based fault location methods for power distribution systems. Electr Power Syst Res 2008;78(4):657–66. https://doi.org/10.1016/j.epsr.2007.05.010.

[16] Zidan A, Khairalla M, Abdrabou AM, Khalifa T, Shaban K, Abdrabou A, et al. Fault detection, isolation, and service restoration in distribution systems: State-of-the-art and future trends. IEEE Trans Smart Grid 2017;8(5):2170–85. https://doi.org/10.1109/TSG.2016.2517620.

[17] Souza JCS, Rodrigues MAP, Schilling MT, Filho MBDC. Fault location in electrical power systems using intelligent systems techniques. IEEE Trans Power Delivery 2001;16(1):59–67. https://doi.org/10.1109/61.905590.

[18] Thukaram D, Khincha HP, Vijaynarasimha HP. Artificial neural network and support vector machine approach for locating faults in radial distribution systems. IEEE Trans Power Delivery 2005;20(2):710–21. https://doi.org/10.1109/TPWRD.2005.844307.

[19] Dehghani F, Nezami H. A new fault location technique on radial distribution systems using artificial neural network. In: 22nd International conference and exhibition on electricity distribution (CIRED 2013); 2013. p. 1–4. doi:10.1049/cp.2013.0697 .

[20] Janik P, Lobos T. Automated classification of power-quality disturbances using SVM and RBF networks. IEEE Trans Power Delivery 2006;21(3):1663–9. https://doi.org/10.1109/TPWRD.2006.874114.

[21] Das B. Fuzzy logic-based fault-type identification in unbalanced radial power distribution system. IEEE Trans Power Delivery 2006;21(1):278–85. https://doi.org/10.1109/TPWRD.2005.852273.

[22] Majidi M, Etezadi-Amoli M, Fadali MS. A novel method for single and simultaneous fault location in distribution networks. IEEE Trans Power Syst 2015;30(6):3368–76. https://doi.org/10.1109/TPWRS.2014.2375816.

[23] Salim RH, de Oliveira KRC, Filomena AD, Resener M, Bretas AS. Hybrid fault diagnosis scheme implementation for power distribution systems automation. IEEE Trans Power Delivery 2008;23(4):1846–56. https://doi.org/10.1109/TPWRD.2008.917919.

[24] Galijasevic Z, Abur A. Fault location using voltage measurements. IEEE Trans Power Delivery 2002;17(2):441–5. https://doi.org/10.1109/61.997915.

[25] Momoh JA, Dias LG, Laird DN. An implementation of a hybrid intelligent tool for distribution system fault diagnosis. IEEE Trans Power Delivery 1997;12(2):1035–40. https://doi.org/10.1109/61.584434.

[26] Teng Jen-Hao. A direct approach for distribution system load flow solutions. IEEE Trans Power Delivery 2003;18(3):882–7. https://doi.org/10.1109/TPWRD.2003.813818.

[27] Niu G, Zhou L, Pei W, Qi Z. A novel fault location and recognition method for low voltage active distribution network. 2015 5th International conference on electric utility deregulation and restructuring and power technologies (DRPT) 2015. p. 876–81. https://doi.org/10.1109/DRPT.2015.7432418.

[28] Sun K, Chen Q, Gao Z. An automatic faulted line section location method for electric power distribution systems based on multisource information. IEEE Trans Power Delivery 2016;31(4):1542–51. https://doi.org/10.1109/TPWRD.2015.2473681.

[29] Marques L, Silva N, Miranda I, Rodriges E, Leite H. Detection and localisation of non-technical losses in low voltage distribution networks. In: Mediterranean conference on power generation, transmission, distribution and energy conversion (MedPower 2016); 2016. p. 1–8. doi:10.1049/cp.2016.1079.

[30] Pasdar AM, Sozer Y, Husain I. Detecting and locating faulty nodes in smart grids based on high frequency signal injection. IEEE Trans Smart Grid 2013;4(2):1067–75. https://doi.org/10.1109/TSG.2012.2221148.

[31] Orcajo GA, Cano JM, Melero MG, Cabanas MF, Rojas CH, Pedrayes JF, et al. Diagnosis of electrical distribution network short circuits based on voltage Park's vector. IEEE Trans Power Delivery 2012;27(4):1964–72. https://doi.org/10.1109/TPWRD.2012.2210448.

[32] Sapountzoglou N, Lago J, Raison B. Fault diagnosis in low voltage smart distribution grids using gradient boosting trees. Electr Power Syst Res 2020;182:106254https://doi.org/10.1016/j.epsr.2020.106254.

[33] Sapountzoglou N, Raison B, Silva N. Fault detection and localization in LV smart grids. In: 2019 IEEE Milan PowerTech; 2019. p. 1–6. doi:10.1109/PTC.2019.8810799.

[34] Silva N, Basadre F, Rodrigues P, Nunes MS, Grilo A, Casaca A, et al. Fault detection and location in Low Voltage grids based on distributed monitoring. In: 2016 IEEE international energy conference (ENERGYCON); 2016. pp. 1–6. doi:10.1109/ENERGYCON.2016.7514000.

[35] Chen Z, Wu L, Cheng S, Lin P, Wu Y, Lin W. Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics. Appl Energy 2017;204:912–31. https://doi.org/10.1016/j.apenergy.2017.05.034.

[36] Jia K, Gu C, Li L, Xuan Z, Bi T, Thomas D. Sparse voltage amplitude measurement based fault location in large-scale photovoltaic power plants. Appl Energy 2018;211:568–81. https://doi.org/10.1016/j.apenergy.2017.11.075.

[37] Wang Z, Hong J, Liu P, Zhang L. Voltage fault diagnosis and prognosis of battery systems based on entropy and Z-score for electric vehicles. Appl Energy 2017;196:289–302. https://doi.org/10.1016/j.apenergy.2016.12.143.

[38] Wang Z, Wang Z, He S, Gu X, Yan ZF. Fault detection and diagnosis of chillers using Bayesian network merged distance rejection and multi-source non-sensor information. Appl Energy 2017;188:200–14. https://doi.org/10.1016/j.apenergy.2016.11.130.

[39] Li J, Yang Q, Mu H, Blond SL, He H. A new fault detection and fault location method for multi-terminal high voltage direct current of offshore wind farm. Appl Energy

2018;220:13–20. https://doi.org/10.1016/j.apenergy.2018.03.044.
[40] long Wu X, Xu Y-W, Xue T, qi Zhao D, Jiang J, Deng Z, et al. Health state prediction and analysis of SOFC system based on the data-driven entire stage experiment. Appl Energy 2019;248:126–40. doi:10.1016/j.apenergy.2019.04.053.
[41] Pahon E, Steiner NY, Jemei S, Hissel D, Moçoteguy P. A signal-based method for fast PEMFC diagnosis. Appl Energy 2016;165:748–58. https://doi.org/10.1016/j.apenergy.2015.12.084.
[42] Tahan M, Tsoutsanis E, Muhammad M, Karim ZA. Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review. Appl Energy 2017;198:122–44. https://doi.org/10.1016/j.apenergy.2017.04.048.
[43] Fu X, Li G, Zhang X, Qiao Z. Failure probability estimation of the gas supply using a data-driven model in an integrated energy system. Appl Energy 2018;232:704–14. https://doi.org/10.1016/j.apenergy.2018.09.097.
[44] Yuyama A, Kajitani Y, Shoji G. Simulation of operational reliability of thermal power plants during a power crisis: Are we underestimating power shortage risk? Appl Energy 2018;231:901–13. https://doi.org/10.1016/j.apenergy.2018.09.089.
[45] Rashid H, Singh P, Stankovic V, Stankovic L. Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour? Appl Energy 2019;238:796–805. https://doi.org/10.1016/j.apenergy.2019.01.061.
[46] Veldhuis AJ, Leach M, Yang A. The impact of increased decentralised generation on the reliability of an existing electricity network. Appl Energy 2018;215:479–502. https://doi.org/10.1016/j.apenergy.2018.02.009.
[47] Kamali S, Amraee T. Blackout prediction in interconnected electric energy systems considering generation re-dispatch and energy curtailment. Appl Energy 2017;187:50–61. https://doi.org/10.1016/j.apenergy.2016.11.040.
[48] Alamuti MM, Nouri H, Ciric RM, Terzija V. Intermittent Fault Location in Distribution Feeders. IEEE Trans Power Delivery 2012;27(1):96–103. https://doi.org/10.1109/TPWRD.2011.2172695.
[49] Javadian SAM, Nasrabadi AM, Haghifam M, Rezvantalab J. Determining fault's type and accurate location in distribution systems with DG using MLP neural networks. In: 2009 International conference on clean electrical power; 2009. p. 284–9. doi:10.1109/ICCEP.2009.5212044.
[50] Aslan Y. An alternative approach to fault location on power distribution feeders with embedded remote-end power generation using artificial neural networks. Electr Eng 2012;94(3):125–34. https://doi.org/10.1007/s00202-011-0218-2.
[51] Brahma SM. Fault location in power distribution system with penetration of distributed generation. IEEE Trans Power Delivery 2011;26(3):1545–53. https://doi.org/10.1109/TPWRD.2011.2106146.
[52] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.
[53] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Proceedings of the 25th international conference on neural information processing systems, NIPS'12 USA: Curran Associates Inc.; 2012. p. 1097–105. https://doi.org/10.1145/3065386.
[54] Hinton G, Deng L, Yu D, Dahl GE, Mohamed A, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Process Mag 2012;29(6):82–97. https://doi.org/10.1109/MSP.2012.2205597.
[55] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. 3rd International conference on learning representations, ICLR. 2015.
[56] Liu H, Chen C. Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction. Appl Energy 2019;254:113686https://doi.org/10.1016/j.apenergy.2019.113686.
[57] Filipe J, Bessa RJ, Reis M, Alves R, Póvoa P. Data-driven predictive energy optimization in a wastewater pumping station. Appl Energy 2019;252:113423https://doi.org/10.1016/j.apenergy.2019.113423.
[58] Vázquez-Canteli JR, Nagy Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. Appl Energy 2019;235:1072–89. https://doi.org/10.1016/j.apenergy.2018.11.002.
[59] Hong Y-Y, Rioflorido CLPP. A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. Appl Energy 2019;250:530–9. https://doi.org/10.1016/j.apenergy.2019.05.044.
[60] Zhang J, Yan J, Infield D, Liu Y, Sang Lien F. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Appl Energy 2019;241:229–44. doi:10.1016/j.apenergy.2019.03.044.
[61] Brusaferri A, Matteucci M, Portolani P, Vitali A. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. Appl Energy 2019;250:1158–75. https://doi.org/10.1016/j.apenergy.2019.05.068.
[62] Lago J, De Ridder F, Vrancx P, De Schutter B. Forecasting day-ahead electricity prices in Europe: The importance of considering market integration. Appl Energy 2018;211:890–903. https://doi.org/10.1016/j.apenergy.2017.11.098.
[63] Lago J, De Ridder F, De Schutter B. Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms. Appl Energy 2018;221:386–405. https://doi.org/10.1016/j.apenergy.2018.02.069.
[64] Lago J, De Brabandere K, De Ridder F, De Schutter B. Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data. Sol Energy 2018;173:566–77. https://doi.org/10.1016/j.solener.2018.07.050.
[65] Kingma DP, Ba J. Adam: A method for stochastic optimization, arXiv eprint; 2014. arXiv:1412.6980.
[66] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929–58.
[67] Jaderberg M, Mnih V, Czarnecki WM, Schaul T, Leibo JZ, Silver D, et al. Kavukcuoglu, Reinforcement learning with unsupervised auxiliary tasks. 5th international conference on learning representations. 2017.
[68] Li X, Zhao L, Wei L, Yang M-H, Wu F, Zhuang Y, et al. DeepSaliency: Multi-Task Deep Neural Network model for salient object detection. IEEE Trans Image Process 2016;25(8):3919–30. https://doi.org/10.1109/TIP.2016.2579306.
[69] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, editors. Advances in Neural Information Processing Systems 27. Curran Associates, Inc.; 2014. p. 3320–8.
[70] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. In: Advances in neural information processing systems; 2011. p. 2546–54.
[71] Sanduleac M, Lipari G, Monti A, Voulkidis A, Zanetto G, Corsi A, et al. Next generation real-time smart meters for ICT based assessment of grid data inconsistencies. Energies 2017;10(7):857. https://doi.org/10.3390/en10070857.
[72] Parvez I, Sarwat AI, Wei L, Sundararajan A. Securing metering infrastructure of smart grid: a machine learning and localization based key management approach. Energies 2016;9(9):691. https://doi.org/10.3390/en9090691.
[73] Chollet F. Keras; 2015. https://github.com/fchollet/keras.
[74] Bayliss C, Hardy B. Transmission and distribution electrical engineering. 4th ed. Oxford: Newnes; 2012.
[75] Yang J, Wang Y. Review on protection issues of low-voltage distribution network with multiple power-electronic-converter-interfaced distribution energy resources. In: International conference on renewable power generation (RPG 2015); 2015. pp. 1–6. doi:10.1049/cp.2015.0327.
[76] Hu R, Granderson J, Auslander D, Agogino A. Design of machine learning models with domain experts for automated sensor selection for energy fault detection. Appl Energy 2019;235:117–28. https://doi.org/10.1016/j.apenergy.2018.10.107.
[77] Saha MM, Izykowski J, Rosołowski E. Fault location on power networks power systems. London: Springer; 2010.
[78] El Khatib M, Hernandez Alvidrez J, Ellis A. Fault analysis and detection in microgrids with high PV penetration. Tech. Rep. SAND2017-5472. Sandia National Lab; 2017.
[79] Kim I. The effect of load current on a three-phase fault. In: 2016 IEEE power energy society innovative smart grid technologies conference (ISGT); 2016. p. 1–4. doi:10.1109/ISGT.2016.7781169.
[80] Pignati M, Zanni L, Romano P, Cherkaoui R, Paolone M. Fault detection and faulted line identification in active distribution networks using synchrophasors-based real-time state estimation. IEEE Trans Power Delivery 2017;32(1):381–92. https://doi.org/10.1109/TPWRD.2016.2545923.