Voltage dip diagnosis in electrical distribution systems using extreme learning machines: an empirical evaluation

Emiliano Reynares[a,b,*], Emmanuel Sangoi[c,d], Jorge Ruben Vega[c,d], María Laura Caliusco[a,b], María Rosa Galli[e]

[a] *Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información (CIDISI), Universidad Tecnológica Nacional Facultad Regional Santa Fe (UTN FRSF), Lavaisse 610, S3004EWB Santa Fe, Argentina*
[b] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290, C1425FQB CABA, Argentina*
[c] *Centro de Investigación y Desarrollo en Ingeniería Eléctrica y Sistemas Energéticos (CIESE), Universidad Tecnológica Nacional Facultad Regional Santa Fe (UTN FRSF), Lavaisse 610, S3004EWB Santa Fe, Argentina*
[d] *Instituto de Desarrollo Tecnológico para la Industria Química (INTEC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ruta Nacional 168, Km. 0, Paraje "El Pozo, 3000 Santa Fe, Argentina*
[e] *Instituto de Desarrollo y Diseño (INGAR), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) - Universidad Tecnológica Nacional Facultad Regional Santa Fe (UTN FRSF), Avellaneda 3657, S3002GJC Santa Fe, Argentina*

Abstract

An inefficient operation of the utility grid can be inferred from low power quality indexes. Under such conditions, reductions in the useful life of equipments and loads, grid instabilities, service interruptions and breakdowns can be expected, which would produce meaningful economic losses for the end-users as well as for the power supply company. In this context, it is useful to have a diagnosis tool able to identify the underlying cause of a given disturbance. This article proposes a one-step extreme learning machine-based method for the diagnosis of voltage dips detected on an electrical distribution grid. The proposed method was evaluated on the basis of synthetic data generated by the simulation of a real power grid. The performance of 10 hyperparameter combinations on

*Corresponding author
Email address: `ereynares@frsf.utn.edu.ar` (E. Reynares), `esangoi@frsf.utn.edu.ar` (E. Sangoi), `jrvega@frsf.utn.edu.ar` (J.R. Vega), `mcaliusc@frsf.utn.edu.ar` (M.L. Caliusco), `mrgalli@santafe-conicet.gov.ar` (M.R. Galli) (María Rosa Galli)

Preprint submitted to IFAC Journal of Systems and Control *September 18, 2019*

the diagnosis of 20 different classes of voltage dips was assessed through two different strategies, i.e.: 1) a one-for-all approach, with a common classifier for all disturbance classes; and 2) a one-against-all approach, with a classifier per disturbance class. It was proven that strategy 2) has a better generalization ability and lower training and testing times. The obtained results suggest the potential of the proposed approach for power quality disturbance diagnosis, and open the challenge of formulating further hypotheses to be assessed.

Keywords: power distribution line, power quality, voltage dip diagnosis, machine learning, artificial neural network, extreme learning machine

1. Introduction

The interest in power quality has increased among electric utilities and consumers in the last decades, as a result of the technological advancement and consumer awareness. Ten years ago, the duration and frequency of the discontinuity of electricity supply were the main power quality issues. Today, the monitoring and control of variations in the frequency and voltage amplitude are equally important.

The diffusion of non-linear electric equipment contributes to the degradation of power quality, requiring more attention from the network operators (Decanini et al., 2011). Power quality problems are the consequences of the increasing use of solid state switching devices, nonlinear and power electronically switched loads, unbalanced power systems, lighting controls, computer and data processing equipment as well as industrial plant rectifiers and inverters. These electronic loads cause quasi-static harmonic dynamic voltage distortions, inrush, pulse type current phenomenon with excessive harmonics and high distortion (Manimala et al., 2012). At the same time, such devices are essential for the daily functioning of our society, being widely used in production, control, and communication processes as well as commercial, domestic and leisure devices.

Low power quality causes inefficient operation of electrical networks and infrastructures as well as the affected loads, resulting in breakdowns and reductions in the useful life of the equipment, malfunctions, instabilities, and interruptions; resulting in breakdowns and product life reductions can result in significant economic losses for the end-user and for the power company (Zhang et al., 2012; Sánchez et al., 2013).

A study carried out in Unites States of America (USA) showed that industrial and digital business firms lost $45.7 billion per year due to power quality issues (Seymour & Horsley, 2008; Meher & Pradhan, 2010). A power quality survey carried out in European Union (EU) analyzed the economic consequences of unreliable electrical power for industrial and service sectors representing over 70% of the EU-25 economic output. The main purpose was to estimate costs of wastage generated by inadequate power quality for those sectors for which electrical power is critical. Costs were classified according to the main power quality disturbances, i.e, (i) voltage dips and swells, (ii) short interruptions, (iii) long interruptions, (iv) harmonics, (v) surges and transients, and (vi) flicker, unbalance, earthing and electromagnetic compatibility problems. The survey presents two main conclusions. First, poor power quality generates a total loss of 150 billion euros annually in the EU-25. And second, voltage dips and short interruptions are the disturbances with the greatest negative economic impact, accounting for 57% of the total loss (Manson & Targosz, 2008).

This situation has raised the need to monitor and improve the power transmission and distribution networks. The ultimate goal of monitoring is to avoid electrical equipment damage and to identify the sources of power quality disturbances. But while the sources and causes of such disturbances must be identified and controlled before taking an appropriate mitigating action, the typical practices to diagnose them strongly rely on visual inspection of waveform disturbances and operator assessment. The analysis of such a volume of data is a difficult task - not to say almost impossible - even for experimented engineers (Khokhar et al., 2015b, 2017). Furthermore, it is rather difficult to determine the cause of a voltage dip, which usually gives rise to controversies between consumers and the electrical company. Within this perspective, it becomes evident the need for diagnostic tools able to identify possible causes of voltage disturbances and to provide assistance in decision-making processes so as to define corrective actions (Decanini et al., 2011). This requires to develop techniques for automatic disturbance identification, to be applied directly or by means of feature extraction and pattern recognition.

Several surveys reporting the application of signal processing and artificial intelligence techniques for the (semi) automatic detection, classification and diagnosis of power quality disturbances can be found in the literature (Strack et al., 2018; Montoya et al., 2016; Khokhar et al., 2015b; Mahela et al., 2015; Saini & Kapoor, 2012; Granados-Lieberman et al., 2011; Saxena et al., 2010). These proposals can be generally divided into two groups.

The first group involves the approaches aimed to classify the type of disturbance - e.g., voltage sags/swells, interruption, etc. A Support Vector Machine (SVM) approach combined with Discrete Wavelet Transform (DWT) to classify disturbances is explored in Ekici (2009). In Meher & Pradhan (2010), Wavelet Transform (WT) is used to extract the features of disturbances, which are then classified by means of a Fuzzy Product Aggregation Reasoning Rule (FPARR). Salem et al. (2010) presents a rule-based system for intelligent classification of disturbances using the Stockwell-Transform (S-Transform) features. Wang & Tseng (2011) classifies disturbances by means of DWT and Extension Genetic Algorithm (EGA). Zhang et al. (2011) applies Discrete Fourier Transform (DFT) and a Rule-Based Decision Tree (RBDT). Decanini et al. (2011) shows the use of DWT, MultiResolution Analysis (MRA), Entropy Norm (EN), and a Fuzzy Adaptive Resonance Theory Neural Network to classify the disturbances. Pires et al. (2011) presents an approach based on Principal Component Analysis (PCA) and Neuro-Fuzzy based automatic classification of disturbances. A classification approach based on the S-Transform and a Probabilistic Neural Network (PNN) is proposed in Huang et al. (2012b). Manimala et al. (2012) classify the disturbances using Wavelet Packet Transform (WPT) and SVM. Sánchez et al. (2013) describes a Genetic Algorithm (GA) that optimizes the S-Transform for analysis and classification of the disturbances. Ozgonenel et al. (2013) assesses the performance of a disturbance classification approach based on Ensemble Empirical Mode Decomposition (EEMD) and SVM for feature extraction and classification. Shareef et al. (2013) presents an image processing-based technique to detect disturbances. Biswal & Dash (2013) introduces variants of S-Transform algorithm and Decision Tree (DT) to select optimal set of features for extraction of the decision rules, which are then used for identification of disturbances. In Hajian & Foroud (2014), Hyperbolic S-Transform (HST), DWT and Orthogonal Forward Selection (OFS) have been used for extraction and selection of features, comparing the performance of SVM, PNN, and K-Nearest Neighbor (K-NN) classifiers. Biswal et al. (2014) proposes an Empirical-Mode Decomposition (EMD) and Hilbert transform (HT)-based method for disturbance classification. Liu et al. (2015) uses a Rank-Wavelet SVM method combined with a EEMD process for the classification task. Khokhar et al. (2015a) presents an approach using DWT and Modular Probabilistic Neural Network (MPNN). Kanirajan & Kumar (2015) presents the application WT combined with Radial Basis Function Neural Network and Particle Swarm Optimization (RBFNN-PSO) algorithms. A disturbance feature selection and recognition method based on Optimal Multi-resolution Fast S-Transform (OMFST) and Classification and Regression Tree (CART) algorithm is proposed in Huang et al. (2016). Reddy & Sodhi (2016) uses a rule-based S-Transform for feature extraction and Adaptive Boost (AdaBoost) as classifier. Strack et al. (2016) assesses a logical classifier, a fuzzy classifier, and a MultiLayer Perceptron (MLP) for voltage disturbances classification. A method using WPT and Extreme Learning Machines (ELM) has been proposed in Naik et al. (2016). Disturbances are classified in Sahani & Dash (2018); Sahani et al. (2016c,b,a) by using WT, Hilbert-Transform (HT), and some variations of ELM. Nayak et al. (2016) presents an ELM and a modified Teaching-Learning-Based Optimization (TLBO) approach for the Local Linear Radial Basis Function Neural Network (LLRBFNN) model to perform the classification. In Zhang et al. (2016), Modified S-Transform (MST) and ELM algorithms are applied to disturbance classification. Khokhar et al. (2017) proposes DWT for feature extraction and PNN-based Artificial Bee Colony (PNN-ABC) as classifier. Ma et al. (2017) introduces a Deep Learning-based classifier, known as Stacked AutoEncoder (SAE). Jaen-Cuellar et al. (2017) focuses on the use of the micro-genetic algorithms (MGA), applying a one-step approach that directly performs feature extraction and classification. Mahela & Shaik (2017) presents a method based on S-Transform and Fuzzy C-means clustering initialized by DT. A Variational Mode Decomposition (VMD) aided Fischer linear Discriminant Analysis (FDA)-based feature selection with reduced kernel ELM has been used as classifier in Chakravorti & Dash (2018). De & Debnath (2018) presents a fuzzy-based method using cross-correlation technique for detection and classification of disturbances. An approach based on Tunable-Q Wavelet Transform (TQWT) and a dual Multiclass Support Vector Machines (MSVM) has been proposed for disturbance detection in Thirumala et al. (2018). Chen et al. (2018) proposes an approach based on dual strong tracking filters (STFs) and rulebased ELM for disturbance detection and classification. Jamali et al. (2018) presents the use of the Sequential Forward Selection (SFS), GA and maximum Relevance Minimum Redundancy (mRMR) algorithms for the selection of disturbance features, which are then input to the most popular models - e.g., Linear Discriminant Analysis (LDA), Naive Bayes, SVM, PNN, MLP, K-NN, etc - for the classification. An ELM classifier coupled with DWT-Entropy and histogram methods for disturbance classification is proposed in Ucar et al. (2018). A deep learning-based compressed sensing and sparse autoencoder approach is used in Liu et al. (2019) for classification of single and complex disturbances.

The goal of the second group of techniques is to diagnose the disturbances by identifying their underlying causes. One of the main attempts in this direction can be found in Bollen et al. (2007). It analyzes a deterministic and a statistical classification approach. The work distinguishes five classes of disturbances on the synthetic dataset: voltage dips with drop in one phase, voltage dips with identical drop in two phases, voltage dips due to three phase faults, voltage dips with a different drop in two phases, and voltage disturbances due to transformer energizing. Erişti & Demir (2010) presents a Wavelet Multi-Resolution Analysis (WMRA) and SVM based approach for the diagnosis of simulated phase-to-ground faults, phase-to-phase faults, three-phase fault, load switching, capacitor switching and transformer energizing events. In Erişti & Demir (2012), WT is used at the feature extraction stage, while the disturbances are classified by means of SVM and diagnosed by using a DT algorithm. Real power system event data is used to evaluate the performance of the strategy, which considers the following disturbances: faults, self-extinguishing faults, line energizing, interruption, and transformer energizing. A feature selection technique for the diagnosis of a dataset of real disturbances is proposed in Erişti et al. (2013). A data set feature vector is obtained by applying multi-resolution analysis (MRA) and k-Means based Apriori algorithm feature selection is used. The obtained optimal feature vector is the input a Least Squares Support Vector Machine (LS-SVM) classifier. A S-Transform and ELM based diagnosis approach is proposed in Erişti et al. (2014) and its performance is assessed on noisy conditions over real and synthetic datasets.

The previous paragraphs show the large number of existing initiatives regarding the detection, classification and diagnosis of power quality disturbances. However, most of them classify disturbances according to their types - e.g.: dip, sag, interruption, etc. - instead of doing it according to their underlying causes. Although such approach is useful for the development of methods, the availability of tools for classifying disturbances according to the event that generated them has a greater practical value. At the same time, the identification of the underlying causes of disturbances is a more difficult and challenging task. It requires not only signal analysis but also the information of power network configuration or settings, comprising signal analysis knowledge, power system knowledge, and information on power network configurations and settings (Bollen et al., 2007).

The aforementioned issues are addressed in this article by performing a set of experiments aimed at assessing the feasibility of an artificial intelligence technique for the diagnosis of voltage dips. The proposal is rooted on the *one-step application* of the ELM algorithm for single-hidden layer feed-forward neural networks. The *one-step application* refers to the lack of any segmentation and feature extraction pre-processing stage: the three-phase voltage waveforms are directly shown to the classifier to identify the underlying cause of the observed disturbance. The classification task is performed over synthetic data generated by the simulation of a real power distribution grid, assessing the performance of 10 hyperparameter combinations of the technique. The experiments consider 20 different types of voltage dips. They are divided into three main groups: fault-induced dips, self-extinguishing faults, and motor starting events. Eleven types of dips were simulated for fault-induced dips, three types for self-extinguishing faults, and six types for motor starting events. 1,000 dips were simulated for each type of dip under different conditions of fault resistance, fault duration, and time evolution of the current.

The rest of the work is organized as follows. Section 2.1 and Section 2.2 introduces the power quality concepts and the Extreme Learning Machine technique, respectively. The experiment dataset is depicted in Section 3 and the experiment setup is detailed in Section 4. The results and the performance metrics applied in their evaluation are shown in Section 5. Insights about the work and the related literature are presented in Section 6. Conclusions and outlook on future research are shown in Section 7.

2. Background

2.1. Power quality

Different and sometimes conflictive definitions of Power Quality (PQ) can be found. IEEE states that "PQ refers to a wide variety of electromagnetic phenomena that characterize the voltage and current at a given time and at a given location on the power system" (IEEE, 2009). IEC defines PQ as "[the set of] characteristics of the electricity at a given point on an electrical system, evaluated against a set of reference technical parameters" (IEC, 2003).

This work adheres to the definition most widely considered in the bibliography: PQ comprises voltage and current quality, which are concerned with deviations of the actual voltage/current from the ideal voltage/current. Ideal voltage/current is a sinusoidal voltage/current waveform with constant amplitude and constant frequency of a given nominal value. These statements allow the simple and straightforward definition of Power Quality Disturbance (PQD) as any deviation of voltage or current from the ideal one. A disturbance is a voltage disturbance or a current disturbance, but it is often not possible to distinguish them given that any change in current is accompanied by a change in voltage and viceversa (Bollen & Gu, 2006).

Voltage disturbances are generally categorized according to their duration and magnitude, considering short and long interruptions, short-duration overvoltages and undervoltages, and long-duration overvoltages and undervoltages. A voltage disturbance is caused by a sudden current increase at a specific point of the electrical grid and along a specified period. The main cause of the overcurrents are faults, originated from short-circuits raised by insulation breakdown, switching operations or atmospheric discharges. Switching of large shunt capacitor banks and starting of large loads (e.g., motors), can also produce large changes in current similar in effect to those of short-circuit states. Supply systems are equipped with protective devices to disconnect the short circuit from the source of energy, which allows the almost immediate recovery of the voltage at every point except the disconnected ones. Some types of faults are self-clearing: the short circuit disappears and the voltage recovers before the line disconnection can take place (Bagdadioglu & Senyucel, 2010; Hanzelka, 2008; Bollen & Gu, 2006). In consequence, the depth and duration of a disturbance perceived by a given user is strongly related to the cause of the overcurrent and the point of the grid where it occurs.

This work focuses on the disturbances known as "voltage dips" or "voltage sags". A voltage dip/sag is a sudden reduction of the voltage at a particular point of an electricity supply system below a specified dip threshold within a short time period, followed by its recovery after a brief interval. The threshold value is the voltage r.m.s value defined in order to determine the start and end of a voltage dip. IEEE defines voltage sag as a decrease of 10 - 90% in the rms voltage at the power frequency for duration of half cycle to 1 minute (IEEE, 2009). IEC defines a voltage dip as a sudden reduction in the supply voltage to a value between 90% and 1% of the nominal voltage followed by a recovery between 10 milliseconds and 1 minute. In polyphase systems it is assumed that a three-phase dip starts when the voltage in the first disturbed phase falls below the threshold and ends when voltages in all phases are equal to or above the threshold. The duration of a voltage dip depends on the specified threshold value. The rest of this work will use the term voltage dip to refer to both concepts, although some details in terms of amplitude and duration differ slightly(IEC, 2004).

2.2. Extreme Learning Machine (ELM)

ELM is a relatively new learning algorithm for Single-hidden Layer Feed-forward Neural networks (SLFNs), which provides good generalization performance at extremely fast learning speed -in most cases thousands of times faster than conventional learning techniques (Huang et al., 2006b,a). Due to its remarkable efficiency, simplicity, and impressive generalization performance, ELM has been applied in a variety of domains, such as biomedical engineering, computer vision, system identification, robotics, and power systems control and operation (Zhang et al., 2018; Das et al., 2018; Lin et al., 2017; Bequé & Lessmann, 2017; Rubiolo et al., 2018; Huang et al., 2015; Song & Zhang, 2013; Avci, 2013; Avci & Coteli, 2012).

ELM randomly chooses hidden nodes and analytically determines the output weights of SLFNs. Thus, the only free parameters that need to be learned are the weights between the hidden layer and the output layer of the neural model.

For L arbitrary distinct samples $(\mathbf{x}_l, \mathbf{y}_l)$, where $\mathbf{x}_l = [x_{l1}, x_{l2}, ..., x_{lN}]^T \in \mathbf{R}^N$ is the l-th input vector and $\mathbf{y}_l = [y_{l1}, y_{l2}, ..., y_{lM}]^T \in \mathbf{R}^M$ is the l-th target vector, a SLFN with P hidden nodes and activation function $f(\cdot)$ is mathematically modeled as:

$$\sum_{i=1}^{P} \mathcal{B}_i f(\mathbf{w}_i^T \cdot \mathbf{x}_l + b_i) = \mathbf{o}_l \tag{1}$$

where $\mathcal{B}_i$ is the weight vector connecting the i-th hidden node and the output nodes, $\mathbf{w}_i = [w_{i1}, w_{i2}, ..., w_{iN}]^T$ is the weight vector connecting the i-th hidden node and the input nodes, b_i is the threshold of the i-th hidden node, and $\mathbf{o}_l$ is the output value of the l-th node.

The SLFN can approximate these L samples with zero error as $\sum_l \|\mathbf{o}_l$-$\mathbf{y}_l\| = 0$, for which there must exist $\hat{\mathcal{B}}_i$, $\hat{\mathbf{w}}_i$, and $\hat{b}_i$ so that:

$$\sum_{i=1}^{P} \hat{\mathcal{B}}_i f(\hat{\mathbf{w}}_i^T \cdot \mathbf{x}_l + \hat{b}_i) = \mathbf{y}_l \tag{2}$$

The above L equations can be compactly written as:

$$HB = Y \tag{3}$$

where

$$H = \begin{bmatrix} f(\hat{\mathbf{w}}_1^T \cdot \mathbf{x}_1 + \hat{b}_1) & \dots & f(\hat{\mathbf{w}}_P^T \cdot \mathbf{x}_1 + \hat{b}_P) \\ \vdots & \ddots & \vdots \\ f(\hat{\mathbf{w}}_1^T \cdot \mathbf{x}_L + \hat{b}_1) & \dots & f(\hat{\mathbf{w}}_P^T \cdot \mathbf{x}_L + \hat{b}_P) \end{bmatrix} \tag{4}$$

$$B = [\hat{\mathcal{B}}_1^T, ..., \hat{\mathcal{B}}_P^T]^T, \tag{5}$$

$$Y = [\mathbf{y}_1^T, ..., \mathbf{y}_L^T]^T \tag{6}$$

H is the hidden layer matrix of the neural network, and its i-th column is the i-th hidden node output with respect to inputs $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_L$. The output connection weights can be obtained by solving $min_B \|HB$-$Y\|$. Then:

$$\hat{B} = H^{\dagger} Y \tag{7}$$

where $H^{\dagger}$ is the Moore-Penrose generalized matrix inverse of the hidden layer matrix H.

The ELM algorithm can be summarized as follows:

a) Inputs:

- L-size training set $(\mathbf{x}_l, \mathbf{y}_l)_{l=1,2,...,L}$, where $\mathbf{x}_l \in \mathbf{R}^N$ and $\mathbf{y}_l \in \mathbf{R}^M$,

- the number P of hidden nodes,
- and the activation function $f(\cdot)$.

b) Steps:

- Select arbitrary values for the input weights $\mathbf{w}_i \in \mathbf{R}^N$ and thresholds $b_i \in \mathbf{R}$ for the P hidden neurons.
- Calculate the hidden layer matrix H of the SLFN.
- Obtain the optimal $\hat{B}$ by solving the $\hat{B} = H^{\dagger} Y$ equation

c) Output: For any input vector $\mathbf{x}_l \in \mathbf{R}^N$, the output value $\mathbf{o}_l$ is obtained through the formula $\sum_{i=1}^{P} \hat{\mathcal{B}}_i f(\hat{\mathbf{w}}_i^T \cdot \mathbf{x}_l + \hat{b}_i)$.

As proved in Huang et al. (2006b), if the $f(\cdot)$ activation function of the hidden layer is infinitely differentiable and the number P of hidden nodes is less than or equal to the number L of samples, then the input weights $\mathbf{w}_i \in \mathbf{R}^N$ and thresholds $b_i \in \mathbf{R}$ can be randomly selected and the SLFN approximates the training data with ε error, i.e.: $\|HB\text{-}Y\| \leq \varepsilon$.

After the input weights and the hidden layer biases are chosen randomly, the SLFNs is a linear system and the output weights are analytically determined through generalized inverse operation of the hidden layer output matrices. In this way, ELM obtains the global optimal solution by fitting $\hat{B}$, avoiding to fall into local optimums.

Different from traditional iterative learning algorithms, ELM not only reach the smallest training error but also the smallest norm of weights. According to the Barlett (1998) study on the generalization performance of feedforward neural networks reaching smaller training error, the smaller the norm of weights is, the better generalization performance the networks tend to have. Thus, ELM obtains not only the minimum square training error but also the best generalization scores on new samples (Huang et al., 2006a,b).

3. Experimental dataset

Diagnosis of voltage disturbances is performed by classifying voltage dips according to their underlying causes. The classification task is performed over synthetic data generated by the simulation of a real power distribution grid, assessing the performance of several hyperparameter combinations of the SLFN/ELM based technique. The grid is placed on the outskirts of the city of Santa Fe (Argentina), and supplies electricity to a populous coastal neighborhood (named *Alto Verde*) and a scientific research and industrial installations campus ($PTLC$)[1]. The distribution grid has an aerial distributor of 13.2 kV in a horizontal coplanar arrangement that starts in the transformer station and ends at PTLC, with intermediate derivation that supplies electricity to the Alto Verde neighborhood. The lines are made of aluminum conductor steel reinforced (ACSR) of $50/8mm^2$ with the following parameters:

- $R_0 = 0.7406\ ohm/km$, $X_0 = 1.624\ ohm/km$, for homopolar impedance, and
- $R_1 = 0.5962\ ohm/km$, $X_1 = 0.3555\ ohm/km$, for direct sequence impedance.

The electricity source is a power substation of 132 kV, with a single/three-phase short-circuit power of 3,500/3,100 MVA. A three-winding transformer (30/30/30 MVA) is used to reduce the high-voltage to 13.2 kV in the feeder of interest, which also has a neutral grounding reactor with $R = 11.80\ ohm$ and $X = 2.00\ ohm$. Figure 1 depicts the grid.

The distribution grid was modeled with ATP/EMTP and ATPDraw [2] software for the simulation of 20,000 voltage dips. Each simulation lasts 500 milliseconds, with a sampling time of 1 millisecond and a discrete simulation interval of 10 microseconds. The simulation consisted in producing overcurrents at a specific location of the grid - the fault point - and measuring the instantaneous

[1] Parque Tecnológico Litoral Centro (PTLC). More information on `http://www.ptlc.org.ar/`

[2] More information on `http://www.emtp.org/`

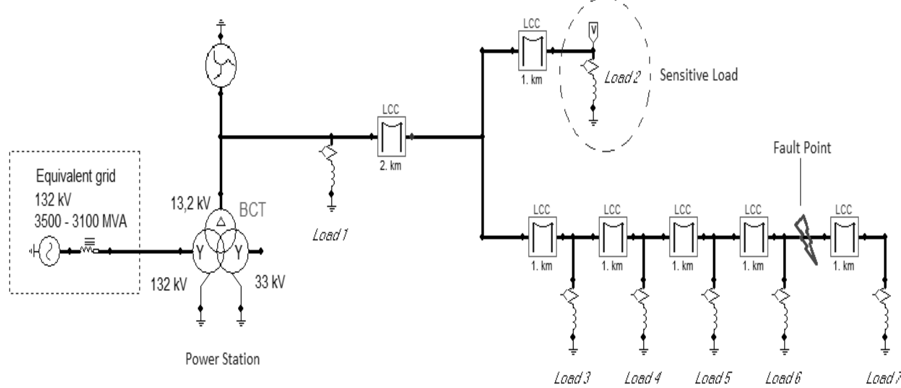


Figure 1: Electric grid model used in the experiments.

voltages on another location - the sensitive load -. An intermediate point of the derivation line to the neighborhood was stated as the fault point, while PTLC represents the sensitive load. Both location points can be identified in Figure 1.

This article considers 20 different types of voltage dips. They are divided into three main groups: fault-induced dips, self-extinguishing faults, and motor starting events. Eleven types of dips were simulated for fault-induced dips, three types for self-extinguishing faults, and six types for motor starting events. 1,000 dips were simulated for each type of dip under different conditions of fault resistance, fault duration, and time evolution of the current, i.e.: duration time and fault impedance values were randomly selected from zero-mean Gaussian distributions. Table 1 presents the details of the voltage dip types and the nomenclature used in the context of the classification problem. Finally, it remains to be noted that the simulations considers peak load demands and the loads were modeled as constant impedances, according to the parameters depicted in Table 2.

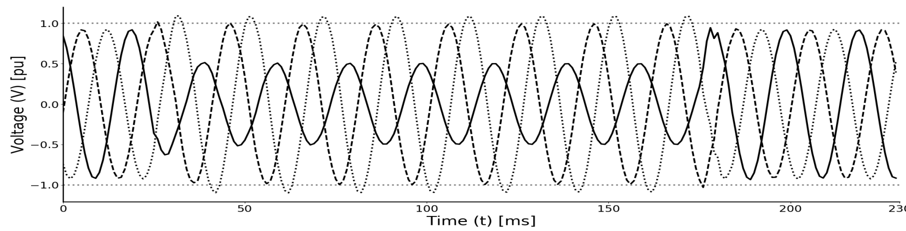
Figure 2(a)-Figure 8(b) present the three-phase voltage waveforms of a simulation of each type of voltage dip analyzed. The vertical axis shows the phase-to-ground voltage per unit. Although each simulation last 500 milliseconds, the horizontal axis (time axis) length was shortened to highlight the window where the disturbance and the fault-clearing happens.

Table 1: Types of voltage dips simulated, and their nomenclature for the classification task.

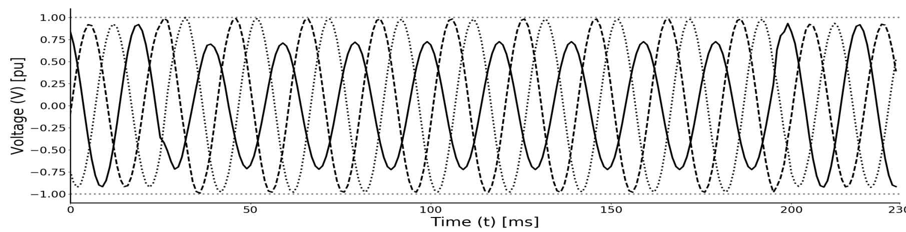
Class	Voltage dip type description
C1	Single line-to-ground fault. Three-phase fault-clearing. Linear fault resistance.
C2	Single line-to-ground fault. Single-phase fault-clearing. Serial nonlinear fault resistance.
C3	Single line-to-ground fault. Single-phase fault-clearing. Parallel nonlinear fault resistance.
C4	Double line-to-ground faults. Three-phase fault-clearing. Always the same two phases for each simulation.
C5	Double line-to-ground fault. Tree-phase fault-clearing. Two different phases for each simulation.
C6	Three-phase fault. Three-phase fault-clearing.
C7	Simultaneous single line-to-ground fault on all phases. Independent single fault-clearing. Different fault resistances and different clearing times for each phase.
C8	Simultaneous single line-to-ground fault on all phases. Independent single fault-clearing. The same fault resistances but different clearing times for each phase.
C9	Self-extinguishing single line-to-ground fault. Nonlinear fault resistance.
C10	Self-extinguishing line-to-line fault. Linear fault resistance.
C11	Self-extinguishing three-phase fault. Linear fault resistance.
C12	Simultaneous line-to-ground fault on all phases. Electric arc with fault ionization.
C13	Simultaneous line-to-ground fault on all phases. Electric arc with fault deionization.
C14	Simultaneous line-to-ground fault on all phases. Electric arc with random resistance variation.
C15	Balanced three-phase starting current ($\sim$ *180A*).
C16	Unbalanced three-phase starting current ($\sim$ *300A*).
C17	Unbalanced three-phase starting current ($\sim$ *300A*). Different current variation for each phase.
C18	Balanced three-phase starting current ($\sim$ *300A*).
C19	Single-phase starting current ($\sim$ *120A*).
C20	Single-phase starting current on two phases simultaneously ($\sim$ *120A*).

Table 2: Electric parameters for modeling demands as constant impedance loads

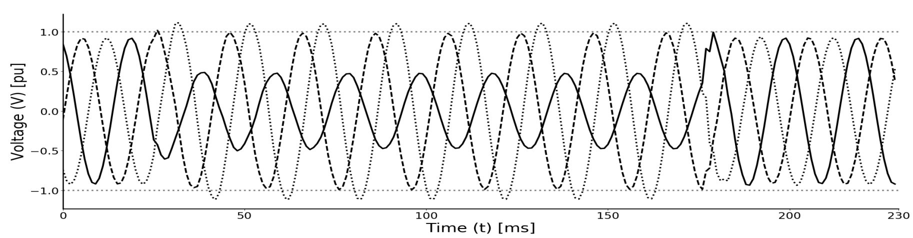
Load	1	2	3	4	5	6	7
R [ohm]	8.92	134.46	313.74	235.31	470.61	313.74	941.22
X [ohm]	2.93	44.19	103.12	77.34	154.68	103.12	309.36



(a) Single line-to-ground fault. Three-phase fault-clearing. Linear fault resistance.

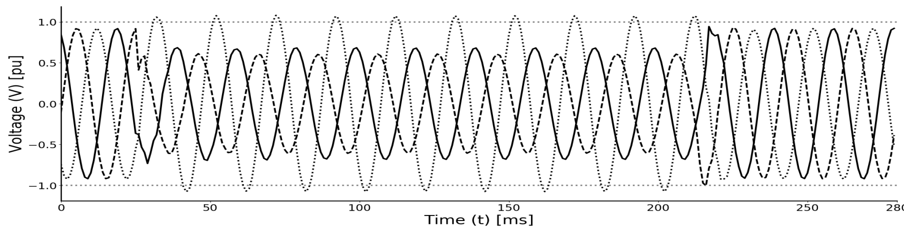


(b) Single line-to-ground fault. Single-phase fault-clearing. Serial nonlinear fault resistance.

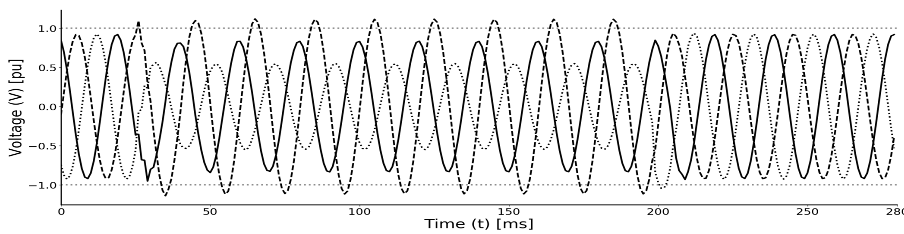


(c) Single line-to-ground fault. Single-phase fault-clearing. Parallel nonlinear fault resistance.

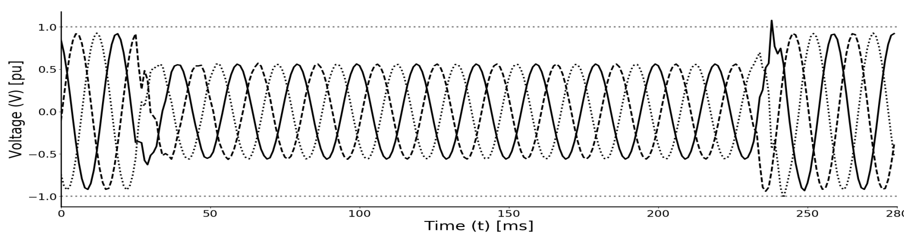
Figure 2: The three-phase voltage waveforms of: (a) a C1 dip type example, (b) a C2 dip type example, and (c) a C3 dip type example.



(a) Double line-to-ground faults. Three-phase fault-clearing. Always the same two phases for each simulation.



(b) Double line-to-ground faults with three-phase fault-clearing. Two different phases for each simulation.

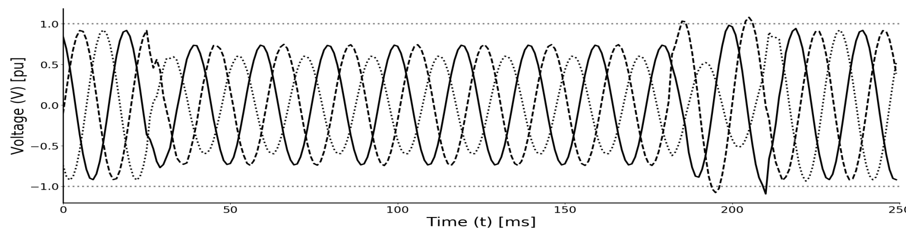


(c) Three-phase faults. Three-phase fault-clearing

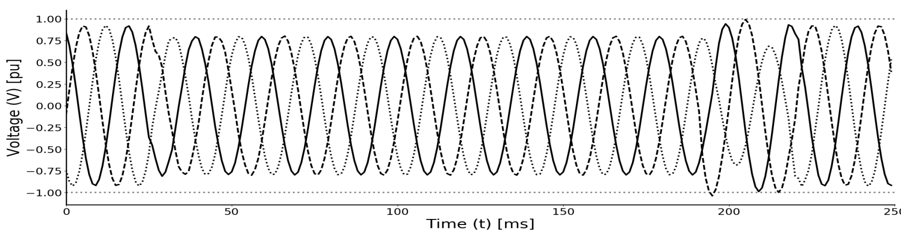
Figure 3: The three-phase voltage waveforms of: (a) a C4 dip type example, (b) a C5 dip type example, and (c) a C6 dip type example.

4. Experimental setup

The SLFN/ELM based approach is evaluated through two strategies aimed at testing different combinations of the following hyperparameters: (i) the number of hidden neurons of the SLFN, and (ii) the activation function. Both strategies apply the single classification method to state one correct class for



(a) Simultaneous single line-to-ground fault on all phases. Independent single fault-clearing. Different fault resistances and different clearing times for each phase.

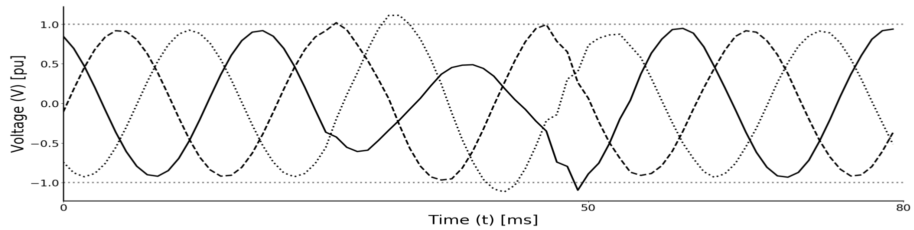


(b) Simultaneous single line-to-ground fault on all phases. Independent single fault-clearing. The same fault resistances but different clearing times for each phase.

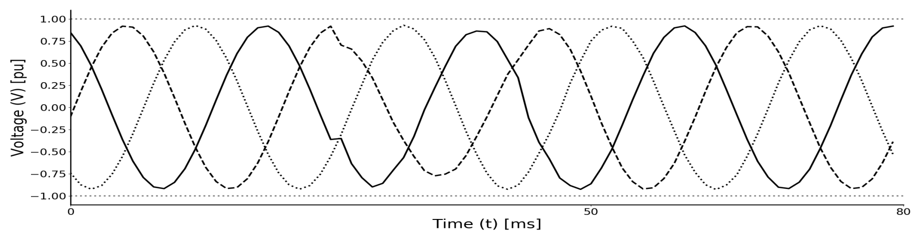
Figure 4: The three-phase voltage waveforms of: (a) a C7 dip type example, and (b) a C8 dip type example.

each sample, (ii) and the *Leave-One-Out Cross-Validation* (LOOCV) procedure to select the model structure with the optimal number of hidden neurons. The LOOCV procedure is computed by the *Predicted Residual Error Sum of Squares* (PRESS) statistic (Raschka, 2018; Akusok et al., 2015).

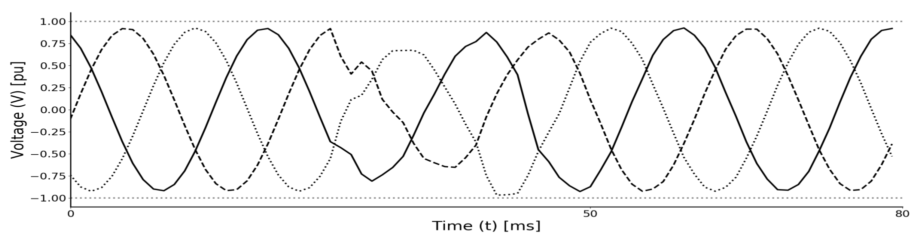
Strategy 1 applies a one-for-all approach. A single classifier is trained to classify the 20,000 simulated voltage dips according to the 20 classes, testing 9 different combinations of the hyperparameters: 1,500 (*1.5k*); 3,000 (*3k*); and 5,000 (*5k*) neurons in the hidden layer, applying sigmoid activation function (*sigm*), hyperbolic tangent activation function (*tanh*), or RBF activation function with L^{∞} norm (rb^{∞}). This strategy applies the one-hot encoding approach by defining an output matrix of Boolean values, showing the class - i.e., the underlying cause - of a given voltage dip. Therefore, an output matrix of 400,000



(a) Self-extinguishing single line-to-ground fault. Nonlinear fault resistance.



(b) Self-extinguishing line-to-line fault. Linear fault resistance.

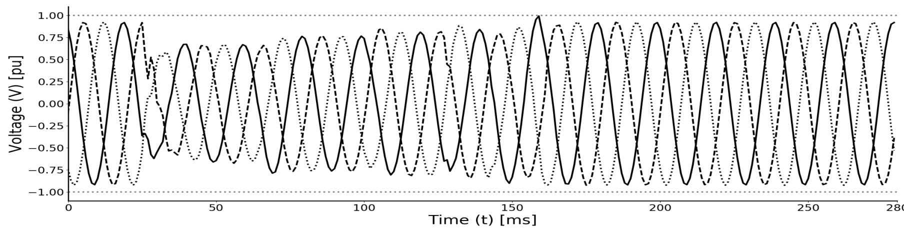


(c) Self-extinguishing three-phase fault with linear fault resistance.

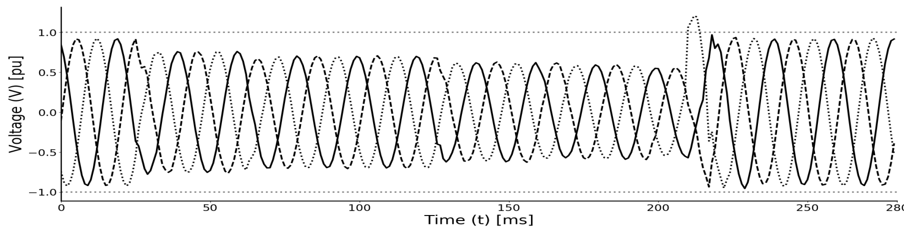
Figure 5: The three-phase voltage waveforms of: (a) a C9 dip type example, (b) a C10 dip type example, and (c) a C11 dip type example.

cells is obtained: 20,000 rows for the voltage dips and 20 columns for the classes.

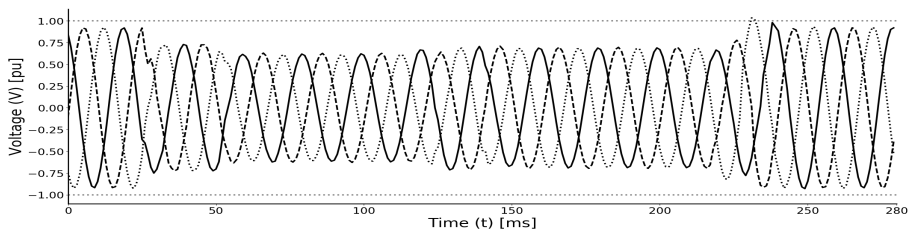
Strategy 2 applies a one-against-all approach. A single classifier is trained per class, the samples of the class are positive samples and the remaining ones are negatives. In consequence, 20 SLFN/ELMs are defined to classify the 20,000 simulated voltage dips according to the 20 adopted classes. Each SLFN/ELM is trained to detect a specific class as the cause of a given voltage dip. This



(a) Simultaneous line-to-ground fault on all phases. Electric arc with fault ionization.



(b) Simultaneous line-to-ground fault on all phases. Electric arc with fault deionization.



(c) Simultaneous line-to-ground fault on all phases. Electric arc with random resistance variation.

Figure 6: The three-phase voltage waveforms of: (a) a C12 dip type example, (b) a C13 dip type example, and (c) a C14 dip type example.

experiment comprises just one setting of the hyperparameters: 750 neurons applying *tanh*. This strategy also applies the one-hot encoding approach, but each of the 20 SLFN/ELMs must learn the relationship between the voltage dips and a single specific class. Then, each classifier defines an output matrix of Boolean values showing whether a given voltage dip is caused by the class for which the model was trained. Therefore, it is obtained 20 output matrices of

40,000 cells each one: 20,000 rows for the voltage dips and 2 columns showing whether *they were/they were not* produced by the specific class.

In this way, the diagnosis of voltage dips is performed by presenting each simulation -comprising 500 milliseconds of the three-phase voltage waveforms- to the different settings of the SFLM/ELM based technique. The voltage waveforms and the class label representing their underlying causes are shown to the classifiers, so that the model can learn the relationship between them. For all settings of both experiments, inputs comprise 500 samples of instantaneous voltages at each phase. As can be noted in Section 2, although ELM originally belongs to the set of regression methods it can be easily adapted for classification problems (Huang et al., 2012a). For the single classification method applied in this work, the index of the output node with the highest output indicates the predicted label of the input.

Strategy 1 and Strategy 2 are implemented using the *Python*[3] *HP-ELM* library. HP-ELM is a high performance neural network toolbox for solving large problems, which supports datasets of any size and GPU acceleration with modest memory consumption and fast non-iterative optimization Akusok et al. (2015) [4]

5. Results

The performance of the approach is assessed by means of three widely used classification task metrics: precision, recall, and F1 score.

Precision (P) is defined as $tp/(tp+fp)$. Recall (R) is defined as $tp/(tp+fn)$. In both expressions, tp is the number of true positives, fp is the number of false positives, and fn is the number of false negatives. Intuitively, P is the ability of the classifier to not label as positive a sample that is negative, while R measures its ability to find the positive samples.

[3]More information on `https://www.python.org/`

[4]Python implementation of the experiments can be found in https://gitlab.com/emireys/elm4vdc-experiments.git

F1 score (F1) is defined as $2 * (P * R)/(P + R)$, and represents a weighted average of precision and recall metrics where the relative contribution of both metrics to F1 score are equal. As it can be inferred from the previous mathematical expressions, the scores for all these metrics range from 1 (best) to 0 (worse).

Since F1 score can be used as a joint representation of precision and recall metrics, the experiment results will be mainly assessed by analyzing Table 3. It shows the average values of F1 scores per class over 100 runs of the 10 different combinations of hyperparameters considered in the experiments. The best values for each class along all proposed hyperparameter combinations are highlighted in grey color. The last two rows of the table show the average and standard deviation values for each hyperparameter combination and for all classes under study. Evenly, the best average and standard deviation values are highlighted in grey color.

By analyzing the rows, it is possible to notice those classes that are easily identifiable by means of any of the proposed hyperparameter combinations. Specifically, C_1, C_2, C_3, C_4, C_5, C_9, C_{10}, C_{11}, C_{15}, C_{16}, C_{19}, and C_{20} classes have F1 scores close to 1, independently of the hyperparameter combination. For C_6 class, only the hyperparameter combination comprising 1,500 neurons in the hidden layer with rb^{∞} activation function has a F1 score lower than 0.91, while the remaining combinations show F1 scores close to 1. The same characteristics are recognizable in F1 scores for C_{12} class, where only the hyperparameter combinations comprising 1,500 and 5,000 neurons in the hidden layer with rb^{∞} activation function has a F1 score lower than 0.92.

Instead, the different hyperparameter combinations have quite disperse F1 scores - i.e., between 0.60 and 0.87 - for C_7 class. A similar dispersion can be seen for C_8 and C_{17} classes, where F1 scores fall in the [0.63; 0.87] and [0.67; 0.82] ranges, respectively.

For C_{13} class, it is observed that only those hyperparameter combinations with the rb^{∞} activation function have F1 scores that do not exceed 0.81, while the other combinations have values at least 0.91.

Table 3: Average values of the F1 score per class over 100 runs of each hyperparameters combination.

	F1 score									
	1.5k sigm	3k sigm	5k sigm	1.5k tanh	3k tanh	5k tanh	1.5k rb$^{\infty}$	3k rb$^{\infty}$	5k rb$^{\infty}$	750 tanh
C_{1}	0.98	0.98	0.98	0.98	0.98	0.99	0.97	0.98	0.98	0.98
C_{2}	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.97	0.97	0.99
C_{3}	0.98	0.98	0.98	0.98	0.98	0.99	0.97	0.98	0.98	0.98
C_{4}	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.97	0.97	1.00
C_{5}	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.97	0.97	1.00
C_{6}	0.97	0.98	0.98	0.97	0.98	0.98	0.89	0.91	0.91	0.97
C_{7}	0.82	0.80	0.74	0.82	0.82	0.78	0.62	0.64	0.60	0.87
C_{8}	0.83	0.82	0.75	0.83	0.83	0.79	0.66	0.67	0.63	0.87
C_{9}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C_{10}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C_{11}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C_{12}	0.97	0.97	0.94	0.98	0.98	0.97	0.90	0.92	0.90	0.99
C_{13}	0.98	0.97	0.91	0.98	0.98	0.96	0.81	0.80	0.74	0.99
C_{14}	0.92	0.92	0.85	0.93	0.93	0.90	0.76	0.79	0.75	0.89
C_{15}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C_{16}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C_{17}	0.67	0.70	0.72	0.71	0.74	0.75	0.78	0.82	0.82	0.81
C_{18}	0.80	0.81	0.81	0.82	0.83	0.83	0.84	0.86	0.85	0.90
C_{19}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C_{20}	1.00	0.99	0.98	1.00	1.00	0.99	0.94	0.93	0.91	1.00
avg.	0.95	0.95	0.93	0.95	0.95	0.95	0.90	0.91	0.90	0.96
std.	0.09	0.09	0.10	0.08	0.08	0.08	0.12	0.11	0.12	0.06

The C_{14} class shows F1 scores of at least 0.90 for the hyperparameters combinations comprising 1,500 and 3000 neurons with *sigm* activation function, and the *tanh* activation function with any number of neurons. Remaining combinations have F1 scores that do not exceed 0.89.

All hyperparameter combinations for C_{18} class have F1 scores in the [0.80, 0.90] range.

The analysis of the columns of Table 3 helps to identify the hyperparameter combination comprising 750 neurons in the hidden layer with *tanh* activation function as the best option for the classification task under study. This conclusion is based on 3 criteria:

1. It gets the best F1 score more times than any other combination. Specif-

ically, in 14 of 20 classes.

2. It gets the best average F1 score along the 20 classes, as shown in the penultimate row of the table.

3. It presents the lowest dispersion of F1 scores, as shown in the last row of the table.

The analysis of training and test times strengthen the choosing of this option. Fig. 9(a) and Fig. 9(b) compare the average training and test times over 100 runs of the 10 hyperparameters combinations. It is evident that at training and testing stages the option comprising 750 neurons in the hidden layer with *tanh* activation function is much faster than the remaining alternatives.

This hyperparameter combination applies the one-against-all approach. I.e., a single classifier is trained per class, where the samples of the class are positive samples and the remaining ones are negative. In consequence, Fig. 9(a) and Fig. 9(b) show the average training/test time over 100 runs of the 20 trained models comprising 750 neurons and *tanh* activation function. For the remaining options -which apply the one-for-all approach-, the average training and test times over 100 runs of the single classifier are shown.

The raw values of the training/testing times are presented as additional material to this article, showing that the aforementioned hyperparameter combination is up to two orders of magnitude faster than the remaining ones. Simulation model of the electrical grid, dataset with 20,000 simulated voltage dips, and performance metrics reports are available on [dataset] Reynares (2019).

6. Discussion

Although there is a large number of initiatives regarding the detection, classification and diagnosis of PQDs, most of them classify disturbances according to their types instead of doing it according to their underlying causes. The diagnosis approach is a difficult and challenging task, but it also comprises greater practical value. Erişti & Demir (2010, 2012); Erişti et al. (2013) and Erişti et al. (2014) depict a set of proposals in this direction.

Erişti & Demir (2010) classifies 1,200 synthetic PQDs simulated on the WECC 9-bus system model according to 6 causes: phase-to-ground fault, phase-to-phase fault, three-phase fault, load switching, capacitor switching, and transformer energizing. Erişti & Demir (2012) and Erişti et al. (2013) classify a dataset comprising 538 disturbances monitored in a real power system into 5 causes: fault, self-extinguishing fault, line energizing, non-fault interruption, and transformer energizing. In Erişti et al. (2014),the approach proposes to classify a mixture of 5,458 real and synthetic disturbances, considering the same 5 causes of the aforementioned works.

All these proposals also share the same general structure. They comprise (i) the pre-processing stage, (ii) the feature extraction stage, and (iii) the classification stage. At the pre-processing stage, normalization and segmentation task are performed. Segmentation decreases the size of data under analysis and enables the obtention of its distinctive features. Normalization converts voltage waveform values to a relative scale - i.e., per unit. The feature extraction stage allows to recognize and prioritize the distinctive features of the dataset. Just after having completed all these activities, it is possible to apply some classification procedure of the disturbances. This structure increases the complexity of the proposed solutions, and entails greater computational and time resources, both at the design and execution stage.

The structure proposed in this article follows the opposite approach. No normalization, segmentation, nor feature extraction stage. The diagnosis task is performed by presenting the three-phase voltage waveforms directly to the ELM based classifier, without intermediate steps. This allows to reduce the processing loads and simplifies the overall technique. The strategy not only allows to diagnose disturbances according to a reduced set of generic causes, it focuses on the diagnosis of voltage dips assessing 20 slightly different failure scenarios under different operative conditions of a real power network.

Despite the lack of the normalization, segmentation, and feature extraction stages, the proposed technique achieves high levels of performance. Specifically, the technique performs almost perfect classifications in at least half of the causes considered, and achieves an average performance of 96% on the full dataset. The experiment setup allows to rule out the possibility of achieving these results just by chance. Two different strategies and ten hyperparameter combinations are considered, running a hundred repetitions of each possible option by randomly varying the elements that constitute the training and test sets. The classification task performance is assessed by means of three widely used classification task metrics. The training and test times are also considered as criteria to select the best option for the classification task under study.

Unlike many published works, and in order to facilitate the eventual reproduction of experiments, this article provides the ATP/EMTP power network model and all the details about the failure scenarios under different operative conditions, the dataset with the simulated voltage dips, the source code of the experiments, and their performance metrics. The software frameworks, libraries, and tools being used for performing the experiments are open sourced and freely accessible. Publishing of these resources enables the experiments of this work to be fully reproducible. The simulation model of the power network, the dataset with 20,000 simulated voltage dips, and performance metrics reports are available on [dataset] Reynares (2019)[5].

7. Conclusions

This work shows the potential of a SLFN/ELM based approach for the diagnosis of voltage dips on power distribution lines by identifying the causes of the disturbances. Previous works have shown the potential of ELMs as classifiers, and this article empirically confirm such insight by assessing several performance dimensions of the algorithm on an extensive set of experiments. The experiments also allow to recognize the strategy that presents the best

[5]The dataset and Python implementation of the experiments can be found on `https://gitlab.com/emireys/elm4vdc-experiments`

predictive performance, and at the same time, it is the most attractive from a pragmatic point of view. I.e., the one-against-all strategy is a scalable approach, and the 750-*tanh* hyperparameter combination turns out a simple model with high predictive scores at short training and test times.

Scalability is an extremely important issue since in a real environment new causes for disturbances could be identified and the diagnostic technique should be able to incorporate them into their internal models. The one-against-all approach enables to extend such models by building SLFN/ELMs specifically trained to diagnose the disturbances according to the newly identified cause. The re-training of the previous models is required in this case, but as it was previously shown, this is quite fast. In contrast, the gradual incorporation of new knowledge is not possible when the one-for-all strategy is applied. It requires the building and training of a full new model from scratch to consider new causes for disturbances, and this is quite costly in terms of time and effort.

It is important to mention that the hyperparameters settings assessed in this works also respond to the intention to contrast the performance of the one-for-all (Strategy 1) and one-against-all (Strategy 2) approaches. According to that intention, Strategy 2 only evaluates the half of the minimum number of hidden neurons considered by Strategy 1 and applies only one of the available activation funcions. However, despite this drastic reduction in the complexity of the model, Strategy 2 is presented as the best option for the task under study in terms of generalization performance, training times and testing times. These results made evident the impact of each approach on the performance metrics, architectural features -in terms of scalability-, and extensibility -through gradual incorporation of new knowledge- of the overall solution.

It is known that a laboratory experiment suffers from a certain lack of realism, but is also recognizable that a field study that involves the monitoring and diagnosis of voltage events is quite expensive both from a technical and legal point of view. Even taking into account these limitations, the results encourages researchers on the field to formulate further hypotheses to be tested in subsequent experiments.

A first scenario to be analyzed involves the assessment of the proposal under different noisy conditions. An even more interesting proposal is to assess the classification performance of the models on a power network different from which they were trained. Such experiments would allow to test the impact of a power network topology and operative condition over the diagnosis model, getting insights about one of the main issues regarding the application of machine learning techniques on power quality diagnosis: the portability of the technique and the models obtained through different power systems, network topologies and operating conditions.

Acknowledgements

This work was supported by Agencia Santafesina de Ciencia, Tecnología e Innovación (ASaCTeI) under IO 2010-058-16, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and Microrredes Eléctricas Hibridas con Alta Penetración de Energías Renovables (MEIHAPER) thematic network (CYTED 717RT0533).

References

Akusok, A., Bjork, K. M., Miche, Y., & Lendasse, A. (2015). High-Performance Extreme Learning Machines: A Complete Toolbox for Big Data Applications. *IEEE Access*, *3*, 1011–1025. doi:10.1109/ACCESS.2015.2450498.

Avci, E. (2013). A new method for expert target recognition system: Genetic wavelet extreme learning machine (gawelm). *Expert Systems with Applications*, *40*, 3984–3993. doi:10.1016/j.eswa.2013.01.011.

Avci, E., & Coteli, R. (2012). A new automatic target recognition system based on wavelet extreme learning machine. *Expert Systems with Applications*, *39*, 12340–12348. doi:10.1016/j.eswa.2012.04.012.

Bagdadioglu, N., & Senyucel, O. (2010). *Service Quality Regulation in Electricity Distribution*. doi:10.3846/bm.2010.102.

Barlett, P. L. (1998). The sample complexity of pattern classification with neural networks: the size of the weight is more important than the size of the network. *IEEE Transactions on Neural Networks*, *5*, 989–993.

Bequé, A., & Lessmann, S. (2017). Extreme learning machines for credit scoring: An empirical evaluation. *Expert Systems with Applications*, *86*, 42–53. doi:10.1016/j.eswa.2017.05.050.

Biswal, B., Biswal, M., Mishra, S., & Jalaja, R. (2014). Automatic classification of power quality events using balanced neural tree. *IEEE Transactions on Industrial Electronics*, *61*, 521–530. doi:10.1109/TIE.2013.2248335.

Biswal, M., & Dash, P. (2013). Detection and characterization of multiple power quality disturbances with a fast s-transform and decision tree based classifier. *Digital Signal Processing: A Review Journal*, *23*, 1071–1083. doi:10.1016/j.dsp.2013.02.012.

Bollen, M., & Gu, I. (2006). *Signal Processing of Power Quality Disturbances*. IEEE Press Series on Power Engineering, Vol. 30. John Wiley Sons.

Bollen, M. H. J., Gu, I. Y. H., Axelberg, P. G. V., & Styvaktakis, E. (2007). Classification of underlying causes of power quality disturbances: Deterministic versus statistical methods. *Eurasip Journal on Advances in Signal Processing*, *2007*. doi:10.1155/2007/79747.

Chakravorti, T., & Dash, P. (2018). Multiclass power quality events classification using variational mode decomposition with fast reduced kernel extreme learning machine-based feature selection. *IET Science, Measurement and Technology*, *12*, 106–117. doi:10.1049/iet-smt.2017.0123.

Chen, X., Li, K., & Xiao, J. (2018). Classification of power quality disturbances using dual strong tracking filters and rule-based extreme learning machine. *International Transactions on Electrical Energy Systems*, *28*. doi:10.1002/etep.2560.

Das, P., Bisoi, R., & Dash, P. (2018). Data decomposition based fast reduced kernel extreme learning machine for currency exchange rate forecasting and trend analysis. *Expert Systems with Applications*, *96*, 427–449. doi:10.1016/j.eswa.2017.10.053.

De, S., & Debnath, S. (2018). Real-time cross-correlation-based technique for detection and classification of power quality disturbances. *IET Generation, Transmission Distribution*, *12*, 688–695. doi:10.1049/iet-gtd.2017.0507.

Decanini, J., Tonelli-Neto, M., Malange, F., & Minussi, C. (2011). Detection and classification of voltage disturbances using a fuzzy-artmap-wavelet network. *Electric Power Systems Research*, *81*, 2057–2065. doi:10.1016/j.epsr.2011.07.018.

Ekici, S. (2009). Classification of power system disturbances using support vector machines. *Expert Systems with Applications*, *36*, 9859–9868. doi:10.1016/j.eswa.2009.02.002.

Erişti, H., & Demir, Y. (2010). A new algorithm for automatic classification of power quality events based on wavelet transform and svm. *Expert Systems with Applications*, *37*, 4094–4102. doi:https://doi.org/10.1016/j.eswa.2009.11.015.

Erişti, H., & Demir, Y. (2012). Automatic classification of power quality events and disturbances using wavelet transform and support vector machines. *IET generation, transmission & distribution*, *6*, 968–976.

Erişti, H., Yıldırım, Ö., Erişti, B., & Demir, Y. (2013). Optimal feature selection for classification of the power quality events using wavelet transform and least squares support vector machines. *International Journal of Electrical Power & Energy Systems*, *49*, 95–103. doi:https://doi.org/10.1016/j.ijepes.2012.12.018.

Erişti, H., Yıldırım, Ö., Erişti, B., & Demir, Y. (2014). Automatic recognition system of underlying causes of power quality disturbances based on S-

Transform and Extreme Learning Machine. *International Journal of Electrical Power and Energy Systems*, *61*, 553–562. doi:10.1016/j.ijepes.2014.04.010.

Granados-Lieberman, D., Romero-Troncoso, R., Osornio-Rios, R., Garcia-Perez, A., & Cabal-Yepez, E. (2011). Techniques and methodologies for power quality analysis and disturbances classification in power systems: A review. *IET Generation, Transmission and Distribution*, *5*, 519–529. doi:10.1049/iet-gtd.2010.0466.

Hajian, M., & Foroud, A. (2014). A new hybrid pattern recognition scheme for automatic discrimination of power quality disturbances. *Measurement: Journal of the International Measurement Confederation*, *51*, 265–280. doi:10.1016/j.measurement.2014.02.017.

Hanzelka, Z. (2008). Voltage dips and short supply interruptions. In A. Baggini (Ed.), *Handbook of Power Quality* (pp. 79–134). doi:10.1002/9780470754245.ch4.

Huang, G., Huang, G. B., Song, S., & You, K. (2015). Trends in extreme learning machines: A review. *Neural Networks*, *61*, 32–48. URL: http://www.sciencedirect.com/science/article/pii/S0893608014002214. doi:https://doi.org/10.1016/j.neunet.2014.10.001.

Huang, G., Zhou, H., Ding, X., & Zhang, R. (2012a). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *42*, 513–529. doi:10.1109/TSMCB.2011.2168604.

Huang, G.-B., Chen, L., & Siew, C. K. (2006a). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, *17*, 879–892.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006b). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*, 489–501. URL: http:

//www.sciencedirect.com/science/article/pii/S0925231206000385. doi:https://doi.org/10.1016/j.neucom.2005.12.126.

Huang, N., Peng, H., Cai, G., & Chen, J. (2016). Power quality disturbances feature selection and recognition using optimal multi-resolution fast s-transform and cart algorithm. *Energies*, *9*.

Huang, N., Xu, D., Liu, X., & Lin, L. (2012b). Power quality disturbances classification based on s-transform and probabilistic neural network. *Neurocomputing*, *98*, 12–23.

IEC (2003). *Electromagnetic compatibility (EMC). Part 4-30: Testing and measurement techniques - Power quality measurement methods. International Standard IEC 61000-4-30:2003.*. IEC.

IEC (2004). *Electromagnetic compatibility (EMC). Testing and measurement techniques - Voltage dips, short interruptions and voltage variation immunity tests. International Standard IEC 61000-4-11:2004.*. IEC.

IEEE (2009). *IEEE Recommended Practice for Monitoring Electric Power Quality. IEEE Std 1159-2009.*. IEEE.

Jaen-Cuellar, A. Y., Morales-Velazquez, L., de Jesus Romero-Troncoso, R., Moriñigo-Sotelo, D., & Osornio-Rios, R. A. (2017). Micro-genetic algorithms for detecting and classifying electric power disturbances. *Neural Computing and Applications*, *28*, 379–392.

Jamali, S., Farsa, A. R., & Ghaffarzadeh, N. (2018). Identification of optimal features for fast and accurate classification of power quality disturbances. *Measurement*, *116*, 565–574.

Kanirajan, P., & Kumar, V. S. (2015). Power quality disturbance detection and classification using wavelet and rbfnn. *Applied Soft Computing*, *35*, 470–481.

Khokhar, S., Mohd Zin, A., Memon, A., & Mokhtar, A. (2017). A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network. *Measurement: Journal of the International Measurement Confederation*, *95*, 246–259. doi:10.1016/j.measurement.2016.10.013.

Khokhar, S., Mohd Zin, A., Mokhtar, A., Bhayo, M., & Naderipour, A. (2015a). Automatic classification of single and hybrid power quality disturbances using wavelet transform and modular probabilistic neural network. In *2015 IEEE Conference on Energy Conversion, CENCON 2015* (pp. 457–462). IEEE. doi:10.1109/CENCON.2015.7409588.

Khokhar, S., Mohd Zin, A., Mokhtar, A., & Pesaran, M. (2015b). A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances. *Renewable and Sustainable Energy Reviews*, *51*, 1650–1663. doi:10.1016/j.rser.2015.07.068.

Lin, L., Wang, F., Xie, X., & Zhong, S. (2017). Random forests-based extreme learning machine ensemble for multi-regime time series prediction. *Expert Systems with Applications*, *83*, 164–176. doi:10.1016/j.eswa.2017.04.013.

Liu, H., Hussain, F., Yue, S., Yildirim, O., & Yawar, S. J. (2019). Classification of multiple power quality events via compressed deep learning. *International Transactions on Electrical Energy Systems*, . doi:10.1002/2050-7038.12010.

Liu, Z., Cui, Y., & Li, W. (2015). A classification method for complex power quality disturbances using eemd and rank wavelet svm. *IEEE Transactions on Smart Grid*, *6*, 1678–1685. doi:10.1109/TSG.2015.2397431.

Ma, J., Zhang, J., Xiao, L., Chen, K., & Wu, J. (2017). Classification of power quality disturbances via deep learning. *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, *34*, 408–415. doi:10.1080/02564602.2016.1196620.

Mahela, O., Shaik, A., & Gupta, N. (2015). A critical review of detection and classification of power quality events. *Renewable and Sustainable Energy Reviews*, *41*, 495–505. doi:10.1016/j.rser.2014.08.070.

Mahela, O. P., & Shaik, A. G. (2017). Recognition of power quality disturbances using s-transform based ruled decision tree and fuzzy c-means clustering classifiers. *Applied Soft Computing*, *59*, 243–257.

Manimala, K., Selvi, K., & Ahila, R. (2012). Optimization techniques for improving power quality data mining using wavelet packet based support vector machine. *Neurocomputing*, *77*, 36–47. doi:10.1016/j.neucom.2011.08.010.

Manson, J., & Targosz, R. (2008). *European Power Quality Survey Report*. Technical Report Leonardo Energy - Europpean Copper Institute. URL: http://www.leonardo-energy.org/resources/846/european-power-quality-survey-report-5835c642a4a68.

Meher, S., & Pradhan, A. (2010). Fuzzy classifiers for power quality events analysis. *Electric Power Systems Research*, *80*, 71–76. doi:10.1016/j.epsr.2009.08.014.

Montoya, F., García-Cruz, A., Montoya, M., & Manzano-Agugliaro, F. (2016). Power quality techniques research worldwide: A review. *Renewable and Sustainable Energy Reviews*, *54*, 846–856. doi:10.1016/j.rser.2015.10.091.

Naik, C., Hafiz, F., Swain, A., & Kar, A. (2016). Classification of power quality events using wavelet packet transform and extreme learning machine. In *2016 IEEE 2nd Annual Southern Power Electronics Conference, SPEC 2016*. IEEE. doi:10.1109/SPEC.2016.7846169.

Nayak, P., Mishra, S., Dash, P., & Bisoi, R. (2016). Comparison of modified teachinglearningbased optimization and extreme learning machine for classification of multiple power signal disturbances. *Neural Computing and Applications*, *27*, 2107–2122. doi:10.1007/s00521-015-2010-0.

Ozgonenel, O., Yalcin, T., Guney, I., & Kurt, U. (2013). A new classification for power quality events in distribution systems. *Electric Power Systems Research*, *95*, 192–199. doi:10.1016/j.epsr.2012.09.007.

Pires, V. F., Amaral, T. G., & Martins, J. F. (2011). Power quality disturbances classification using the 3-d space representation and pca based neuro-fuzzy approach. *Expert Systems with Applications*, *38*, 11911–11917.

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*, .

Reddy, M., & Sodhi, R. (2016). A rule-based s-transform and adaboost based approach for power quality assessment. *Electric Power Systems Research*, *134*, 66–79. doi:10.1016/j.epsr.2016.01.003.

[dataset] Reynares, E., Emiliano; Sangoi (2019). Electrical grid model, 20,000 simulated voltage dips according to 20 causes, and performance metrics of assessed models. mendeley data, v1. doi:http://dx.doi.org/10.17632/mbmvmd47bs.1.

Rubiolo, M., Milone, D. H., & Stegmayer, G. (2018). Extreme learning machines for reverse engineering of gene regulatory networks from expression time series. *Bioinformatics*, *34*, 1253–1260. URL: http://dx.doi.org/10.1093/bioinformatics/btx730. doi:10.1093/bioinformatics/btx730.

Sahani, M., & Dash, P. (2018). Automatic power quality events recognition based on hilbert huang transform and weighted bidirectional extreme learning machine. *IEEE Transactions on Industrial Informatics*, *14*, 3849–3858. doi:10.1109/TII.2018.2803042.

Sahani, M., Mishra, S., Ipsita, A., & Upadhyay, B. (2016a). Detection and classification of power quality event using hybrid wavelet-hilbert transform and extreme learning machine. In *Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016*. IEEE. doi:10.1109/ICCPCT.2016.7530185.

Sahani, M., Mishra, S., Ipsita, A., & Upadhyay, B. (2016b). Detection and classification of power quality event using wavelet transform and weighted extreme learning machine. In *Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016*. IEEE. doi:10.1109/ICCPCT.2016.7530177.

Sahani, M., Upadhyay, B., Beura, R., & Mishra, S. (2016c). Detection and classification of power quality event using wavelet transform and extreme learning machine. In *Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016*. IEEE. doi:10.1109/ICCPCT.2016.7530176.

Saini, M., & Kapoor, R. (2012). Classification of power quality events - a review. *International Journal of Electrical Power and Energy Systems*, *43*, 11–19. doi:10.1016/j.ijepes.2012.04.045.

Salem, M. E., Mohamed, A., & Samad, S. A. (2010). Rule based system for power quality disturbance classification incorporating s-transform features. *Expert Systems with Applications*, *37*, 3229–3235. URL: http://www.sciencedirect.com/science/article/pii/S0957417409008380. doi:https://doi.org/10.1016/j.eswa.2009.09.057.

Sánchez, P., Montoya, F., Manzano-Agugliaro, F., & Gil, C. (2013). Genetic algorithm for s-transform optimisation in the analysis and classification of electrical signal perturbations. *Expert Systems with Applications*, *40*, 6766–6777. doi:10.1016/j.eswa.2013.06.055.

Saxena, D., Verma, K., & Singh, S. (2010). Power quality event classification: an overview and key issues. *International Journal of Engineering, Science and Technology*, *2*, 186–199. URL: http://www.ajol.info/index.php/ijest/article/view/59190. doi:10.4314/ijest.v2i3.59190.

Seymour, J., & Horsley, T. (2008). The seven types of power problems. *EngineerIT*, (pp. 60–65).

Shareef, H., Mohamed, A., & Ibrahim, A. (2013). An image processing based method for power quality event identification. *International Journal of*

Electrical Power and Energy Systems, *46*, 184–197. doi:10.1016/j.ijepes.2012.10.049.

Song, Y., & Zhang, J. (2013). Automatic recognition of epileptic eeg patterns via extreme learning machine and multiresolution feature extraction. *Expert Systems with Applications*, *40*, 5477–5489. doi:10.1016/j.eswa.2013.04.025.

Strack, J., Carugati, I., Orallo, C., Donato, P., Maestri, S., & Carrica, D. (2018). Analysis of drawbacks and constraints of classification algorithms for three-phase voltage dips. *International Review of Electrical Engineering*, *13*, 23–36. doi:10.15866/iree.v13i1.13391.

Strack, J., Carugati, I., Orallo, C., Donato, P., Maestri, S., & Surez, J. (2016). Evaluation of three-phase voltage dips classification algorithms based on computational intelligence. In *2015 16th Workshop on Information Processing and Control, RPIC 2015*. IEEE. doi:10.1109/RPIC.2015.7497065.

Thirumala, K., Prasad, M. S., Jain, T., & Umarikar, A. C. (2018). Tunable-q wavelet transform and dual multiclass svm for online automatic detection of power quality disturbances. *IEEE Transactions on Smart Grid*, *9*, 3018–3028. doi:10.1109/TSG.2016.2624313.

Ucar, F., Alcin, O. F., Dandil, B., & Ata, F. (2018). Power quality event detection using a fast extreme learning machine. *Energies*, *11*. doi:10.3390/en11010145.

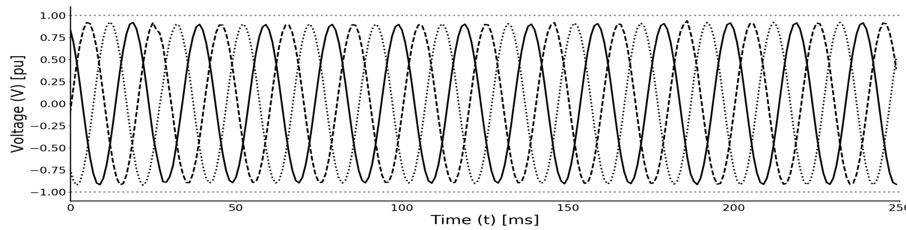
Wang, M.-H., & Tseng, Y.-F. (2011). A novel analytic method of power quality using extension genetic algorithm and wavelet transform. *Expert Systems with Applications*, *38*, 12491–12496. doi:10.1016/j.eswa.2011.04.032.

Zhang, L., Jiang, Y., Liu, N., & Zhang, B.-H. (2012). The study on power quality evaluation based on data analysis. *International Journal of Advancements in Computing Technology*, *4*, 586–592. doi:10.4156/ijact.vol4.issue19.70.

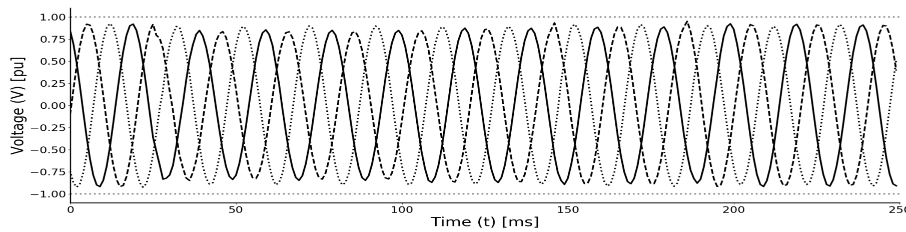
Zhang, M., Li, K., & Hu, Y. (2011). A real-time classification method of power quality disturbances. *Electric Power Systems Research*, *81*, 660–666. doi:10.1016/j.epsr.2010.10.032.

Zhang, S., Li, P., Zhang, L., Li, H., Jiang, W., & Hu, Y. (2016). Modified s transform and elm algorithms and their applications in power quality analysis. *neurocomputing*, *185*, 231–241.

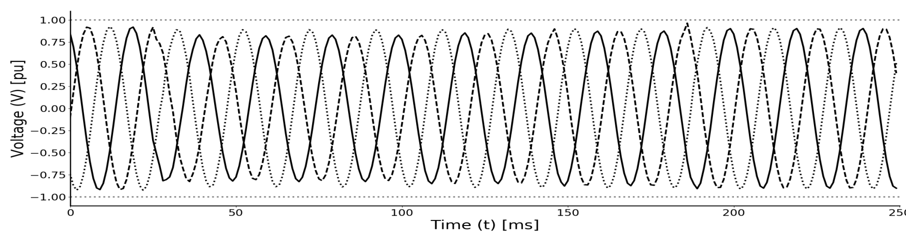
Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., & Cichocki, A. (2018). Multi-kernel extreme learning machine for eeg classification in brain-computer interfaces. *Expert Systems with Applications*, *96*, 302–310. doi:10.1016/j.eswa.2017.12.015.



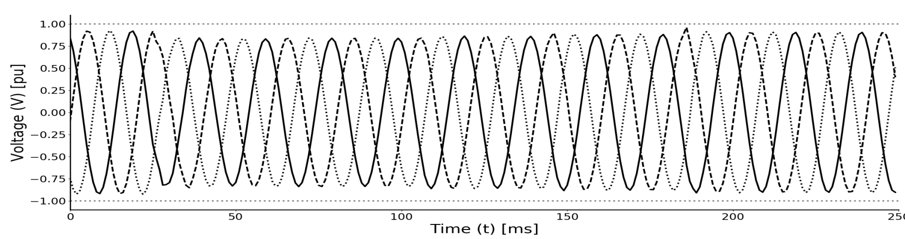
(a) Balanced three-phase starting current (∼ *180A*).



(b) Unbalanced three-phase starting current (∼ *300A*).

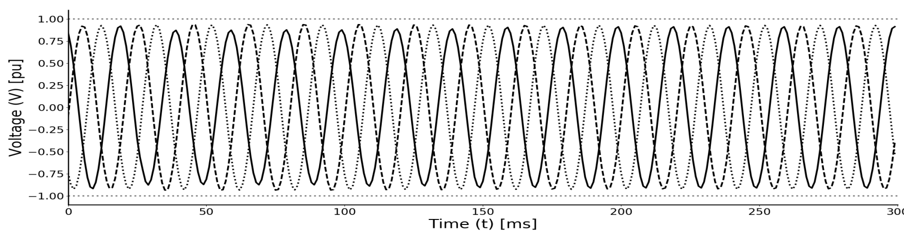


(c) Unbalanced three-phase starting current (∼ *300A*). Different current variation for each phase.

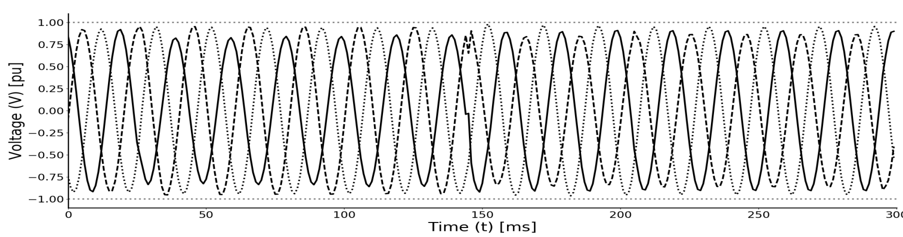


(d) Balanced three-phase starting current (∼ *300A*).

Figure 7: The three-phase voltage waveforms of: (a) a C15 dip type example, (b) a C16 dip type example, (c) a C17 dip type example, and (d) a C18 dip type example.

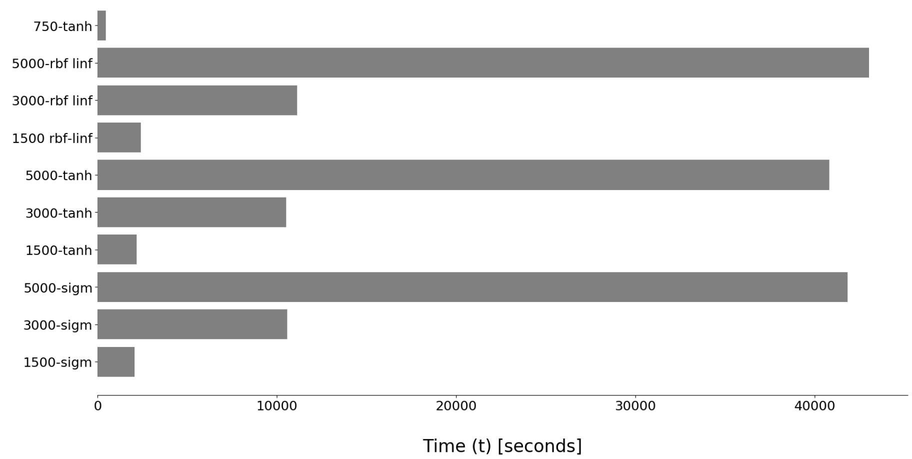


(a) Single-phase starting current ($\sim$ *120A*).

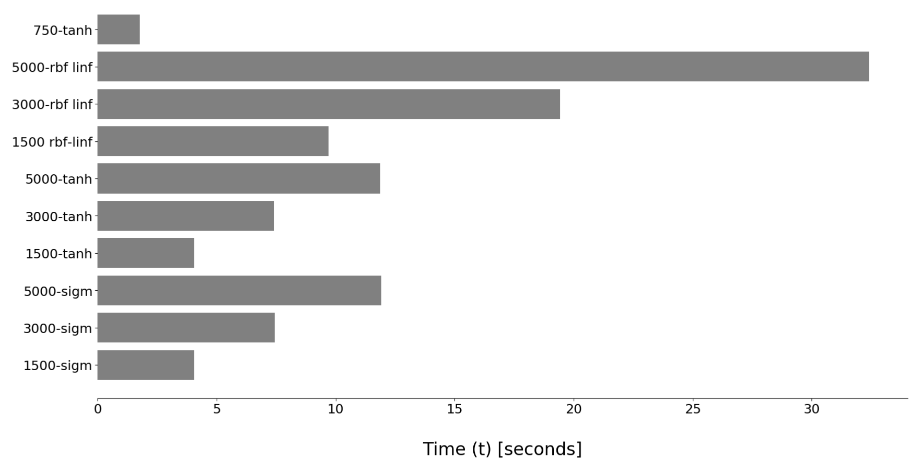


(b) Single-phase starting current on two phases simultaneously ($\sim$ *120A*).

Figure 8: The three-phase voltage waveforms of: (a) a C19 dip type example, and (b) a C20 dip type example.



(a) Average values of training time for each hyper-parameter combination



(b) Average values of testing time for each hyper-parameter combination

Figure 9: Comparison of average values of training and testing time for each hyper-parameter combination.