

Predicting the Grounding Topology Based on Grounding Impedance and the Pattern Recognition Framework: A case study on one to four ground rods in straight line

Francisco Alexandre A. Souza, and Tobias R. Fernandes Neto and Francisco Rodrigo P. Magalhães and Felipe Bandeira Silva and Ricardo Silva Thé Pontes

Abstract—This paper presents a system which predicts the grounding topologies of grounding systems (GS) with one, two, three or four ground rods in straight line, when no prior information about the grounding topology, soil or rod properties is given. This is achieved through the exploration of the information contained into the impulse response of the GS together with the pattern recognition framework. The proposed system is composed by four main elements, an excitation system, a high-speed data acquisition system, a feature extraction and a pattern recognition model. Experimental setup comprises the evaluation of the accuracy of the proposed system in discriminating among four grounding configurations, composed by the topologies of one, two, three and four ground rods in straight line. This paper also proposes a new data driven approach to improve the classification performance, in pattern classification models, when dealing with undersampling problem in the acquired signal used to extract the features.

Index Terms—grounding electrodes, lightning, transient response, pattern recognition

I. INTRODUCTION

Grounding systems (GS) plays a vital role in electrical network. The correct design of GS are of vital importance for safety and protection of personnel, equipments and facilities, as well to the correct operation of the electricity supply network, etc. The selection of correct GS topology (i.e. number and disposition of rods) is dependent on several factors, which includes the soil resistivity, the available area to install the GS, the season-weather characteristics of the region, the demands of the project, etc [1]. Once the GS topology is defined, its installation should be properly implemented. However, in some cases, this does not happen, mainly because of errors during the project execution, such as installing an incorrect number of rods (usually less than specified a priori) and/or by improperly connecting the rods (e.g. bad clamping or broken connections).

There are two possible ways to minimize this issue. The first one is based on the visual inspection of the installed GS, under the drawback that the visual inspection does not allow

Francisco Alexandre A. Souza, and Tobias R. Fernandes Neto, and Felipe Bandeira Silva, and Ricardo Silva Thé Pontes; Department of Electrical Engineering, Laboratory of Energy Efficiency in Motor Driven Systems - LAMOTRIZ, Federal University of Ceará, Fortaleza - CE, Brazil, and Francisco Rodrigo P. Magalhães; University of Fortaleza (UNIFOR), Center of Technological Sciences; Fortaleza - CE, Brazil

the analysis of the buried part. Moreover, there is also the possibility that the GS have already been covered or cemented, thereby making impossible the visual inspection of the GS. The second one is based on the value of GS resistance. Having the GS resistance and the information on the soil resistivity, defined as ρ , it is possible to infer the GS topology. However, this information on the soil properties is not always available, mainly in GS systems of transmission poles where the GS topology is kept fixed. For example, in Brazilian utilities a single ground rod 2.4 m long, with 0.0150 m diameter, is frequently applied throughout medium-voltage lines at the service entrances of low-voltage consumers. For pole-mounted distribution transformers protected by surge arresters, the typical grounding configurations applied are three 2.4 m long parallel rods in straight line, with 0.0150 m diameter, spaced in intervals of 3 m [2]. Then, under this kind of situation it would be beneficial to have a method/system to check whether the configuration of GS is arranged properly or not, by indicating the exact or at least the approximated GS under the soil, and with no prior information on the soil characteristics. Thus, to tackle this problem, a new system for inspecting GS is proposed in this paper. The proposed system is composed by four other subsystems: a excitation, acquisition, feature extraction and pattern recognition subsystems. The core of the proposed system is based on the modeling of GS topology based on the GS impedance. Moreover, the proposed system does not requires any prior information on the soil and/or rod properties, and the proposed system requires only two auxiliary electrodes to be composed.

Several works attempted to study the behavior of the GS impedance when submitted to a lightning strike [3]–[9]. The objective in all of these works were very clearly, model/predict the performance of a GS topology having as input the characteristics of the soil under consideration, the characteristic of the GS topology and the lightning strike current curve. In all these works, the objective was to find a appropriate way to model the harmonic grounding impedance $Z(j\omega)$, or the grounding impedance in time $z(t) = \mathcal{F}^{-1}\{Z(j\omega)\}$, where \mathcal{F}^{-1} is the discrete inverse Fourier transform, by setting the following function $f(\cdot)$:

$$Z(j\omega) = f(\omega, \Delta, \Theta) \quad (1)$$

where $j = \sqrt{-1}$, $\omega = 2\pi f$, ω is the angular frequency and f

is the frequency (in hertz), Δ represents the soil properties, which usually are given by ρ , ϵ and μ , the resistivity (in ohms · meter), permittivity (in farads per meter) and permeability (in henry per meter) of the soil, respectively. The parameter Θ accounts for the GS topology characteristics, which usually are given by the following parameters, l the grounding rod length (in meters), a the rod radius (in meters), d is the depth of burial of horizontal rods (in meters), D is the distance between vertical grounding electrodes and the number of ground rods, represented by s . Thus, $\Delta = \{\rho, \epsilon, \mu\}$ and $\Theta = \{l, a, d, D, s\}$. Another two important aspects which are also taken into consideration while modeling Eq. (1) are the effects of soil ionization [10] and the dependence of soil parameters ρ and ϵ , over frequency, as detailed described in [9], [11] (see [12] for further understanding on implications of frequency dependence of soil parameters in real case situations, e.g. short circuits in transmission line caused by lightning strikes). The most common approaches to model $f(\cdot)$ are based on the on circuit theory, transmission line theory, full wave models based on field equations and the Numerical Electromagnetic Code [3]–[9], [13]–[15] or hybrid time-frequency approaches to deal with the soil ionization and dependence of soil parameters over frequency [10], [11].

From the knowledge about $Z(j\omega)$ several quantities related to the performance of a GS in high frequencies are computed, usually the impulse impedance Z and the impulse coefficient A . Moreover, it is already know that the transient response (high frequency signal) has several characteristics on the rod properties and soil properties [16], [17].

Instead of modeling the GS impedance, given by Eq. (1), as a function of soil and GS topology parameters, the objective here is to model the GS topology, described by the parameter Θ , in function of the GS impedance (without lack of generality the grounding impedance will refereed in this work as the harmonic impedance $Z(j\omega)$ or time impedance $z(t)$). Assume that $\Theta^1, \dots, \Theta^C$ are the representation of different GS topologies, then to compose the proposed system it is necessary to find the following map function $h(\cdot)$

$$\Theta^* = h(\mathbf{x}) \quad (2)$$

which maps the characteristics/features \mathbf{x} extracted from the impedance $Z(j\omega)$ signal to the GS topology domain Θ , where $\Theta^* \in \{\Theta^1, \dots, \Theta^C\}$. As can be seen in Eq. (2), the only information provided to predict the GS topology is the GS impedance, no prior information on soil properties, described by the parameter Δ , is utilized.

Modeling $h(\cdot)$ based on first-principle models can be difficult and so far it has not been investigated yet. In this work, to overcome the complexity and the time demanding of first-principle models, a black-box/data-driven model, based on pattern recognition framework is going to be employed to model the function $h(\cdot)$. The pattern recognition framework requires a set of N examples $\Phi = \{(\mathbf{x}_i, \Theta_i) | i = 1, \dots, N\}$ to learn the function $h(\cdot)$, where \mathbf{x}_i are the features extracted from the GS impedance $Z_i(j\omega)$ of a GS topology with characteristics given by $\Theta_i \in \{\Theta^1, \dots, \Theta^C\}$, in a soil with properties described by Δ_i . There are several applications of pattern recognition framework in the power systems field, e.g.

power system disturbance classification [18], [19], detection of power islands [20], [21] and prediction of seasonal variation of ground resistance [22]. In all the applications it is clear the lack of physical modeling between the output of interest and the characteristics (features) which describe the output. Then, given a set of data, the pattern recognition framework is capable to find the mapping function between the input and output to perform the required task.

However, despite the advantages of black-box modeling over the physical modeling, the accuracy of Eq. (2) is highly dependent on the amount of examples N and the number C of different GS topologies to discriminate. In the case of this work, the collection of data set Φ requires the mounting of the different grounding topologies (the ones to be predicted $\{\Theta^1, \dots, \Theta^C\}$) in different soils, which is costly due the necessity to mount and unmount the different GS's, where $N = n_s \cdot C$ with n_s being equal to the number of soils/places where the GS topologies $\{\Theta^1, \dots, \Theta^C\}$ were mounted and C the total number of GS topologies to be predicted, then the number of collected points tends to be small. Thus, to increase the accuracy of modeling (2), a special case in the experimental part was considered. Instead of predicting the exact GS topology, the special case predicts the approximated GS topology. This is done by grouping GS topologies which are similar to each other (this similarity is defined by the user, for example, in the experimental part the GS topologies with one and two ground rods were considered as a group and the GS topologies with three and four ground rods were considered as a other group). This simplification reduces the number C of GS topologies to be predicted, then reducing the complexity of the learning and increasing the accuracy in modeling $h(\cdot)$. The objective of this approach can be defined as:

$$\Lambda^* = g(\mathbf{x}), \quad (3)$$

where $\Lambda^* \in \{\Lambda^1, \dots, \Lambda^J\}$ is the predicted approximated GS topology, where $\Lambda^j \subset \{\Theta^1, \dots, \Theta^C\} \setminus \sum_{i=1}^j \cup \Lambda^{i-1}$ with $\Lambda^0 = \emptyset$ and $j = 1, \dots, J$, with $J < C$. Instead of predicting the exact topology, the proposed system will also predict the approximated GS topology. In a practical perspective it is enough for the authors to discriminate among topologies with one or two ground rod/s and three or four ground rods.

The proposed system also should be able to extract the GS impedance and then, based on the GS impedance, predict the exact or approximated GS topology which that impedance belongs by using Eqs. (2) and (3). The proposed system is subdivided in four subsystems: excitation, acquisition, feature extraction and pattern recognition subsystems. The excitation system is responsible to inject a impulse voltage signal, similar to a lightning stroke, into the installed GS. The acquisition system is responsible to acquire the response signals (voltage and current, trough two auxiliary electrodes) at a high sample rate rate. The transient part of voltage and current signals are acquired with undersampling (the transient part contains the most information on the GS properties Θ [8]), and to reduce the impact of undersampling problem in modeling (2) and (3) a data-driven approach is proposed in this work. The feature extraction module is responsible to extract the

features \mathbf{x} from the acquired signal of the GS impedance, determined through the voltage and current responses. The extracted features are used as input of a pattern recognition model in the classification system, that outputs the information regarding the topology of the GS. To model the functions $h(\cdot)$ and $g(\cdot)$ in Eqs. (2) and (3) the following pattern recognition models were evaluated in the experimental part, the Naive Bayes, Adaptive Boosting (Adaboost) [23], Decision Trees (DT) [24], and Random Forest Classifiers (RF) [25] and a method based on l_2 norm.

In order to evaluate the proposed system, a set of 26 controlled experiments, in distinct soils were conducted and the configurations with one, two, three and four ground rods in horizontal line were mounted and the impedance $Z(j\omega)$ for each of these configurations were extracted and stored (i.e. $n_s = 26$ and $C = 4$). All the ground rods are cooper made and have $l = 2.4$ m length with $a = 0.0150$ m diameter, and for the configurations with two, three and four rods, the rods were spaced in intervals of $D = 3$ m. The data collection was done in different seasons and in different locations.

This paper is organized as follows. Section II describes the proposed methodology. Section III presents the experimental results. Finally, Section V gives the conclusion remarks.

II. PROPOSED SYSTEM

The objective of this paper is to build a system to predict the exact and approximated GS topology by using Eqs. (2) and (3). For that purpose, the proposed system was designed based on four other subsystems: excitation, acquisition, feature extraction and pattern recognition subsystems. The excitation subsystem is responsible to apply a impulse voltage signal similar to a lightning stroke in the installed GS (this enable the acquisition of high frequency components of impedance). The acquisition subsystem is responsible to acquire the response signals (voltage and current). The feature extraction subsystem is responsible to extract the features from the impedance signal $Z(j\omega)$. The extracted features are used as input of a pattern recognition subsystem, that outputs the prediction regarding the exact and approximated GS topology. All of these subsystems are managed by a dedicated computer. Each of these subsystems are detailed described as follows.

A. Excitation Subsystem

The objective of excitation system is to enable the acquisition of the impedance $Z(j\omega)$ of the installed GS.

The impedance $Z(j\omega)$ is extracted by applying a voltage signal similar to a lightning strike to the GS. According to [26, Chapter 6] the lightning stroke waveform can be approximated by a double exponential as follows:

$$v_{in}(t) = V_0 (e^{-\alpha t} - e^{-\beta t}), \quad (4)$$

where V_0 is the peak value and α and β are constants that define the front time t_f and tail time t_t , respectively. According to [26], the following approximations can be used:

$$t_f \approx \frac{1}{\beta}, \quad (5)$$

$$t_t \approx \frac{1}{\alpha}. \quad (6)$$

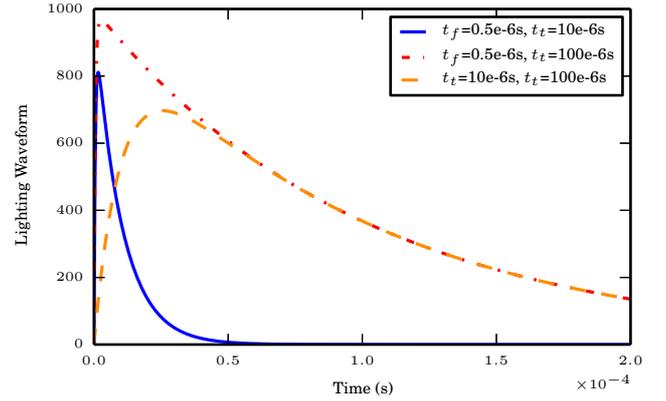


Fig. 1. Lighting stroke waveform for different value of β and α .

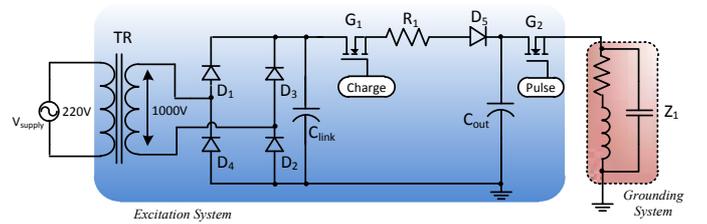


Fig. 2. Architecture of the proposed excitation system.

The common values of the front time of a lightning strikes is between $0.5 \mu\text{s}$ to $10 \mu\text{s}$, while its peak value decays to 50% after $30 \mu\text{s}$ to $200 \mu\text{s}$. Fig. 1 shows the lightning strike waveform for different values of α and β and $V_0 = 1000$ V, the values of α and β range in the common values of a lightning stroke.

The excitation system circuit, proposed in this paper, is depicted in Fig. 2. This circuit can approximate the typical waveform of lightning stroke, given by (4). The system of Fig. 2 can be described by two main steps. The first one is the capacitor charging C_{out} step and the second one is the application of the capacitor voltage into the GS. In the first step, the MOSFET G_1 is closed and G_2 is kept open. The transformer TR step-up the voltage from 220 V to 1 kV, approximately. Then, the voltage is rectified, by the full bridge rectifier composed by the diodes D_1 , D_2 , D_3 , D_4 and capacitor C_{link} , so that the output capacitor C_{out} is charged, the role of resistor R_1 is to limit the current to the capacitor C_{out} . In the second step, the MOSFET G_1 is opened and MOSFET G_2 is closed, then the output capacitor C_{out} (charged before in the first step), will generate an impulse voltage over the GS impedance Z_1 through the MOSFET G_2 that is closed.

As can be seen in Fig. 3, the output of the proposed excitation system has a similar behavior of a double exponential, given by (4) and illustrated by Fig. 1. As the output voltage of the excitation system, Fig. 2, is uncontrolled, the parameters t_f and t_t of the double exponential waveform are dependent of the soil and rod properties. The peak of voltage applied to the GS in all experiments is in the order of 1 kV.

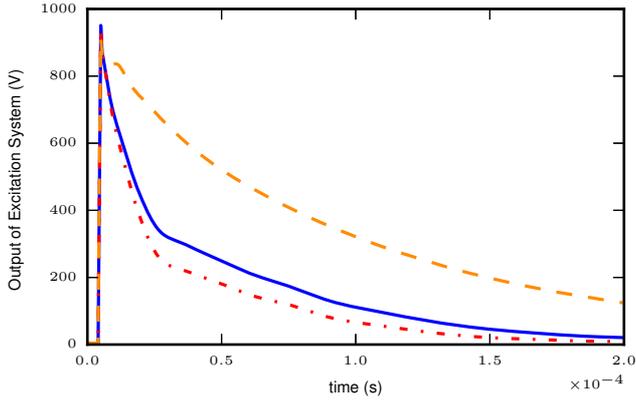


Fig. 3. Experimental results of the voltage waveform applied in different soils

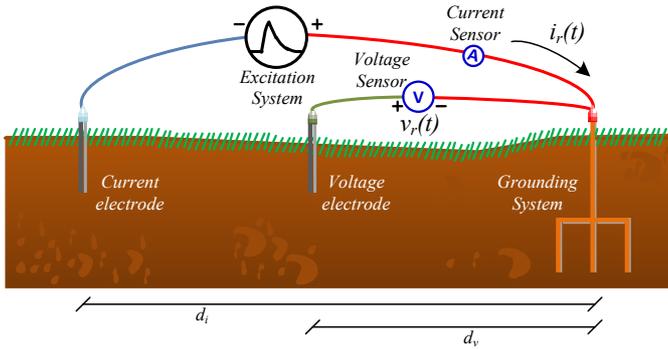


Fig. 4. Architecture of the experimental setup.

B. Acquisition Subsystem

The data acquisition system is responsible to acquire the data necessary to compute the impedance $Z(j\omega)$.

The scheme of the experimental setup, to acquire the impedance $Z(j\omega)$, is shown in Fig. 4. As can be seen, two signals are acquired as responses, the voltage signal $v_r(t)$ and the current signal $i_r(t)$. After the voltage $v_r(t)$ and current $i_r(t)$ responses are acquired, by the data acquisition system, the time domain impedance of the GS, given by

$$z(t) = \frac{v_r(t)}{i_r(t)}, \quad (7)$$

is determined. These signals are collected by two auxiliary electrodes, in line, placed from distances d_v and d_i from the GS under test. This is based on the 3-point method fall of potential test for measuring the resistivity of the soil, but differently of this methodology the value of d_v is kept fixed in all experiments. In this paper, the distances d_v and d_i were defined as $d_v = 12.5$ m and $d_i = 20$ m. The current and voltage electrodes were disposed in straight line in all experiments. These distances were defined empirically. However, others can use different values for these distances, under the constraint that it should keep it fixed for all experiments.

The transient response, which is assumed to contain the majority of information necessary to predict the GS topology, is in the order of micro seconds (μs). To allow the acquisition of the transient signal, this work acquires the data at a sample

rate of 2MSa/s. and sample interval $\Delta t = 0.5 \mu s$ under this acquisition rate, by using the data acquisition system U2531A from Agilent. The measurement period was set to $T = 25$ ms. To capture the voltage and current signals two Hall Effect transducers from LEM were used, the sample rate of voltage and current sensors are equal to 500 kHz and 1 MHz, respectively. However, as it is going to be discussed as follows, there are some issues regarding the frequency of operation of the voltage and current sensors. The maximum acquisition-rate of the voltage is 500 kHz and the current 1 MHz. The experimental results suggests that the transient part of both the voltage and current signals are acquired with undersampling, this can be visualized in Figs. 5(a) and 5(b), respectively. This figure shows four acquired voltage and current signals of a GS topology composed by one ground rod. Each of these curves of $v_r(t)$ and $i_r(t)$ were acquired sequentially and in the same soil during the experiments. The first line figure shows that the first five samples of both, voltage and current have a heavy variation in their values. Figs. 5(c) and 5(d) shows the mean and standard deviation values of $v_r(t)$ and $i_r(t)$, respectively, of thirty acquired signals of the GS topology with one ground rod. As can be seen, the first samples have a large variance, indicating the undersampling problem and increasing the uncertainty on the acquired transient signal. To overcome this issue during the modeling of Eqs. (2) and (3), a data-driven approach was adopted. For each acquired voltage $v_r(t)$ and $i_r(t)$ signals, in a sequence of n_r acquired data, represented by $\{(v_{rj}(t), i_{rj}(t))\}_{j=1}^{n_r}$, applied in Eq. (2), there will be $\{\Theta_1, \dots, \Theta_{n_r}\}$ predicted GS topologies, where $\{\Theta_1^*, \dots, \Theta_{n_r}^*\} \subseteq \{\Theta^1, \dots, \Theta^C\}$. Then, the predicted GS topology will be given the majority vote rule, as follows:

$$\Theta^* = \text{mode}(\{\Theta_1, \dots, \Theta_{n_r}\}) \quad (8)$$

In other words, the predicted GS topology is the one that appears most frequently in the predicted GS topologies from n_r runs (the same was applied in Eq. (3)). This approach has show to reduce the uncertainty associated with the undersampling characteristics of the acquired current and voltage signals. The experimental results will discuss this issue, exhibiting that this strategy increases the accuracy-rate due to the reduce of the uncertainty associated with the transient undersampling problem.

C. Feature Extraction Subsystem

The objective of the feature extraction is to define the features \mathbf{x} extracted from the impedance signal $Z(j\omega)$ to be used in the models $h(\cdot)$ and $g(\cdot)$ in Eqs. (2) and (3). In this work, several features are going be extracted from the acquired data and evaluated later regarding its accuracy in modeling the GS topologies.

As discussed before, the pattern recognition models are black-box models, thus they lack on the explicit formulation of the features \mathbf{x} to model the exact and the approximated GS topology Θ and Λ . However, from the grounding theory it is known that many information, regarding the rod and soil properties, are contained in the impedance signal $z(t)$. Thus two characteristic/features are going to be extracted from the DFT

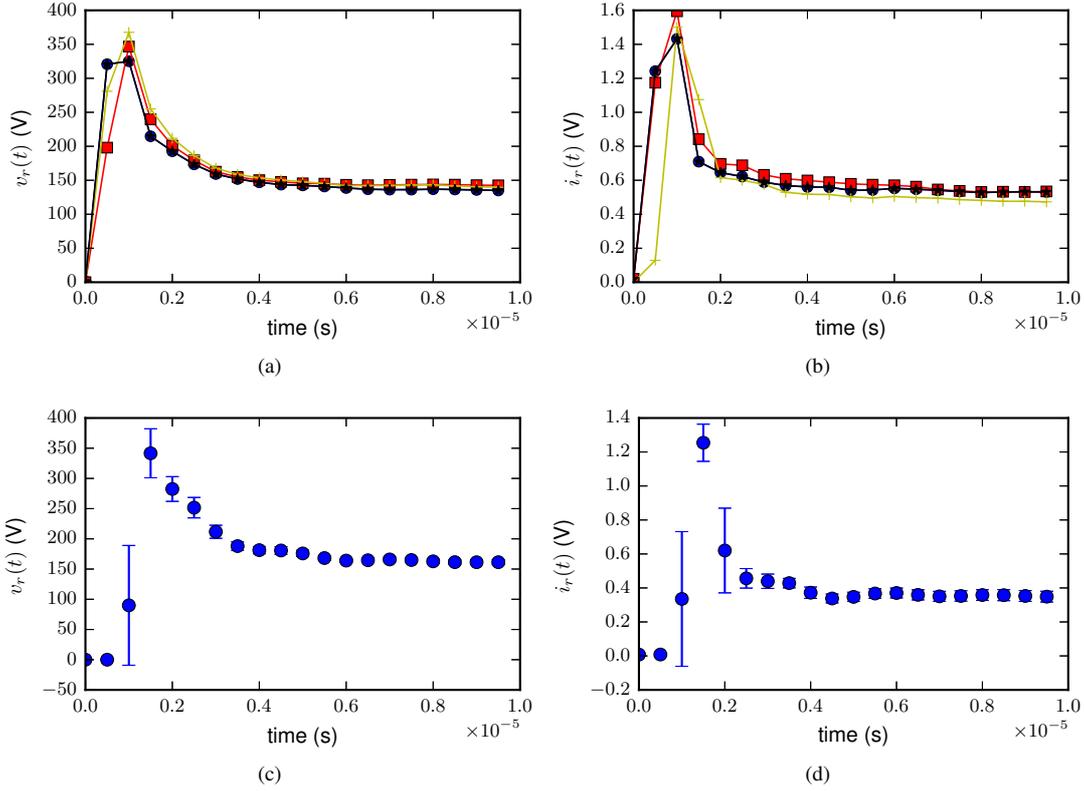


Fig. 5. Four acquired voltage (a) and current (b) signals of a GS topology composed by one ground rod. Mean and variance of discrete acquired values of voltage (c) and current (d), for thirty acquired signals.

of impedance signal $z(t)$, where $t = 0, \Delta t, 2\Delta t, \dots, n_t\Delta t$, where $n_t = 50000$ and $T = n_t\Delta t = 25$ ms. Let the DFT of the impedance signal be equal to

$$Z(n) = \sum_{k=0}^{N_f-1} z(k\Delta t) e^{-j\frac{2\pi}{N_f}nk} \quad (9)$$

for $n = 1, \dots, N_f$, where N_f is the total number of data points of impedance signal where the DFT is going to be applied. The features extracted from the impedance signal are the amplitude (referred here as F.1) and the phase angle (referred here as F.2). The features F.1 can be defined as $\mathbf{x} = [x_1, \dots, x_{N_f}]^T$, where $x_j = |Z(j)|$, and similarly the features F.2 can be defined as $\mathbf{x} = [x_1, \dots, x_{N_f}]^T$, where $x_j = \arg(Z(j))$, where $j = 1, \dots, N_f$.

The number of elements in $Z(n)$ is equal to N_f . The time covered by the N_f samples is equal to $t = 0, \Delta t, \dots, N_f\Delta t$ seconds (under the sample rate of 2MSa/s). As the values of N_f which gives the most informative Fourier coefficients are unknown, different values of N_f will be evaluated in the experimental part.

D. Pattern Recognition Subsystem

The objective here is to solve the Eq. (2) (the remain of section is similarly valid for Eq. (3)) by means of the pattern recognition framework. It means to find a mapping function (classification model):

$$h(\mathbf{x}, \zeta) : X \rightarrow \Theta, \quad (10)$$

that maps the input domain X into the output domain Θ (GS topologies). An element of X is represented by the vector $\mathbf{x} \in \mathbb{R}^{N_f}$, and the Θ is given by $\Theta = \{\Theta^1, \Theta^2, \dots, \Theta^C\}$, where C represents the number of classes (number of grounding topologies to discriminate). If $C = 2$ it is considered as a binary problem and if $C > 2$ it is considered as a multiclass problem. The major issue of pattern recognition framework is to select the parameters ζ of the classification model appropriately. This is done by training the model h with a set of N examples $\Phi = \{(\mathbf{x}_i, \Theta_i^c) | i = 1, \dots, N, c = 1, \dots, C\}$. The selection of the best value of ζ is done in the model learning phase.

There are different algorithms to compose the model $h(\cdot)$ by using the pattern recognition framework. However, the models used in this paper were limited to following classification models, Naives Bayes (NB), Decision Tree (DT), Adaptive Boosting (Adaboost) and Random Forest Classifiers (RF) classifiers. Each of the pattern classification models are described as follows.

1) *Naive Bayes*: Naive Bayes (NB) is a simple classification technique with many applications in the pattern recognition field. The objective of the classification procedure is to assign a feature \mathbf{x} to one of C predefined classes $\Theta = \{\Theta^1, \dots, \Theta^C\}$. Using the Bayesian framework, this problem can be solved by finding the solution of following equation:

$$p = \arg \max_{c=1, \dots, C} P(\mathbf{x} | \Theta^c) P(\Theta^c), \quad (11)$$

where p is the index class to be assigned, $P(\mathbf{x}|\Theta^c)$ is the conditional probability of sample \mathbf{x} given the class Θ^c , and $P(\Theta^c)$ is the *a priori* probability of class Θ^c .

In the NB classifier it is assumed the independence of input features. Assuming features independence, the term $P(\mathbf{x}|\Theta^c)P(\Theta^c)$ becomes $P(x_1|\Theta^c) \times P(x_2|\Theta^c) \times \dots \times P(x_n|\Theta^c) \times P(\Theta^c)$, and (11) can be written as

$$p = \arg \max_{c=1, \dots, C} \prod_{i=1}^n P(x_i|\Theta^c)P(\Theta^c). \quad (12)$$

This expression is referred to as Naive Bayes classifier and the predicted GS topology is given by Θ^p . The parameters ζ are *pdf* parameters choose to represent the $P(\mathbf{x}|\Theta^c)$ distribution. In this paper a Gaussian distribution was considered and the mean and standard deviation values are the parameters ζ to be tuned.

2) *Decision Tree*: The DT divides the feature space into sets of disjoint rectangular regions, defined as A_1, A_2, \dots, A_j , and learns a simple model for each these regions.

$$h(\mathbf{x}_i) = \sum_{i=1}^j c_j I(x_i \in A_j) \quad (13)$$

where $I(x_i \in A_j) = 1$, $I(x_i \notin A_j) = 0$ and c_j are the estimated values of the output class in region A_j . The parameters ζ (the disjoint regions) of DT classifier are going to be learned by using the C4.5 learning algorithm [24].

3) *Adaboost*: The AdaBoost (Adaptive Boosting) model, proposed in [23], is a pattern recognition model algorithm that combines weak classifiers to form a strong classifier. The output $h(\cdot)$ of Adaboost is given by

$$h(\mathbf{x}_i) = \text{sign} \left(\sum_{k=1}^T a_k h_k(\mathbf{x}_i, \theta_k) \right), \quad (14)$$

where h_k are the weak classifiers, T are the number of weak classifiers, and $\zeta = \{a_k, \theta_k\}_{k=1}^T$, are the set of parameters of the Adaboost model. After defining the number of weak classifiers, the learning step is responsible to tune the parameters ζ . There are plenty of algorithms used to learn the parameters ζ , the classical one is given in [23]. Recently, an approach based on the gradient descent algorithm was proposed in [27]. The Adaboost is primary designed to deal with binary classification problems, the extension of Adaboost for multiclass classification is done can be done using the max-wins rule [28].

4) *Random Forests*: The Random Forests (RF), proposed by [25], can be seen as a generalization of decision trees (DT). The RF grows T DT's, where for each tree the training samples and features are randomly assigned. This strategy wants to reduce the variance of the classifier and make it more robust. The output of the RF is given by

$$h(\mathbf{x}_i) = \text{mode} \{T_k(\mathbf{x}_i)\}_{k=1}^T, \quad (15)$$

where $T_k(\mathbf{x}_i)$ is the output of DT k for the input \mathbf{x}_i . The *majority vote* operand outputs the value that appears in the major number in $(T_k(\mathbf{x}_i))_{k=1}^T$. It means that means that forest chooses the classification having the most votes (over all the

trees in the forest). The number T of DT should be set. Then, each of parameters ζ of DT's can be learned by using the C4.5 learning algorithm [24].

III. EXPERIMENTAL RESULTS

In order to evaluate the proposed, a set of 26 controlled experiments, in distinct soils and in different locations were conducted. In all the soils, the configuration with one, two, three and four ground rods in straight line were mounted, and these configurations are represented by the following labels $\Theta^1, \Theta^2, \Theta^3$ and Θ^4 , where all the ground rods have 2.4 m length with 0.0150 m diameter, and for the configurations with two, three and four ground rods, the rods were spaced in intervals of 3 m. For each soil and each GS topology thirty $n_r = 30$ signals of voltage and current responses were acquired with the purpose to reduce the uncertainty associated with the transient data, as described in Section II-B. The total number of samples is equal to $N = n^o \text{ classes} \times n^o \text{ experiments} \times n_r = 4 \times 26 \times 30 = 3120$ samples. To avoid interference in results, the leads used in experimental setup were stretched along the soil to avoid coil formation and to avoid crossed leads.

In the experimental results, the prediction of exact and approximated GS topologies were evaluated (Eqs. (2) and (3)) by using the features described in Section II-C. In all the experiments, the predicted topology is the one given by Eq. (8). The influence of n_r on the accuracy of modeling Eqs. (2) and (3) is also discussed. Moreover, to show the superiority of the pattern classification models, described in Sec. II-D, over conventional or intuitive approaches, the following engineering approach was also evaluated in the experimental part. This approach is based on the comparison of the impedance curve under test with the existing ones of the respective topology to be evaluated. This was done by measuring the l^2 -norm distance between the test data \mathbf{x}_{test} and the train data of the respective topology to be evaluated. Define the $L_{\text{test}}(\Theta^j)$ as the sum of l^2 -norm between the test data sample \mathbf{x}_{test} and the train data samples of the topology Θ^j , defined as $\mathbf{x}_i^{\Theta^j}$:

$$L_{\text{test}}(\Theta^j, \mathbf{x}_{\text{test}}) = \sum_{i=1}^{25 \cdot n_r} \|\mathbf{x}_{\text{test}} - \mathbf{x}_i^{\Theta^j}\|_2 \quad (16)$$

where $25 \cdot n_r$ is the number of training data points for topology Θ^j . Then, the topology to be assigned to the test data sample \mathbf{x}_{test} is given by:

$$p = \arg \min_{c=1, \dots, C} L_{\text{test}}(\Theta^c, \mathbf{x}_{\text{test}}), \quad (17)$$

where, the predicted topology is given by Θ^p . This strategy assigns the \mathbf{x}_{test} sample to the topology who has the most similar samples. This approach is referred as l^2 -norm in the remaining of the paper.

A. Experimental Settings

In the experimental evaluation, all the models, NB, DT, Adaboost, and RF were evaluated in predicting the GS topology based on the input features described in Section II-C. The features extracted from the DFT of impedance signal (defined in Eq. (9)) are the amplitude F.1 and the phase angle F.2,

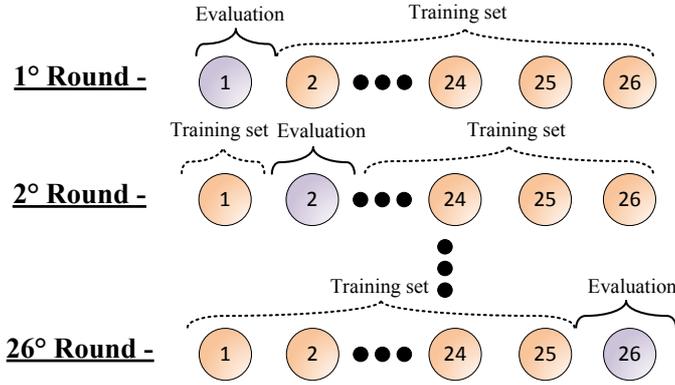


Fig. 6. Representation of the adopted training and test strategy.

respectively. The parameter N_f defines the number of data points used to compute the DFT. As discussed before, the value of N_f which gives the most accurate model is unknown, then different values of N_f will be evaluated in the experimental part. However, as it is known in advance, the transient part of the signal (up to $250 \mu\text{s}$) contains the majority information the GS topology, and high values of N_f will suppress the high frequency impedance signal, thus the range of $N_f = 26$ (the first $12 \mu\text{s}$), $N_f = 50$ (the first $25 \mu\text{s}$), $N_f = 76$ (the first $25 \mu\text{s}$), $N_f = 100$ (the first $50 \mu\text{s}$), $N_f = 250$ (the first $125 \mu\text{s}$), $N_f = 500$ (the first $250 \mu\text{s}$) are going to be evaluated in the experimental part.

The following methodology was applied to evaluate the proposed system in terms of its classification accuracy. From the 26 experiments, where each one has n_r collected data of voltage $v_r(t)$ and current $i_r(t)$ from each GS topology ($\Theta^1, \Theta^2, \Theta^3, \Theta^4$), 25 were used to train the pattern recognition model (in a total of $4 \cdot 25 \cdot n_r$ samples), and the remaining one, which is not part of the training set, was used to evaluate the prediction accuracy of the trained model ($4 \cdot n_r$ samples). The trained model was used to predict the four GS topologies in the test soil. This was repeated such that all the soils were used as the test set, so that the information of test set is not included in the training set. This approach simulates a practical application of the proposed system, where there is no information regarding the soil where the GS is going to be inspected. This approach is illustrated in Fig. 6.

Regarding the tuning of parameters of the pattern recognition models ζ , they were selected based on the same procedure as above; for the Adaboost the number of weak classifiers T , and the number of DT in the RF classifiers.

As discussed in Section II-B, to reduce the effect of the undersampling problem, each predicted GS topology is going to have n_r predictions, the selected one, defined as by Θ^* is given by Eq. (8), Section II-B.

The performance of each model is going to be evaluated based on its rate of accuracy (ACC).

$$\text{ACC} = \frac{\sum_{i=1}^N I(h(\mathbf{x}_i), y_i)}{N}, \quad (18)$$

where Λ^* is defined in (8) and $I(\cdot)$ is defined as:

$$I(h(\mathbf{x}_i), y_i) = \begin{cases} 1 & \text{if } h(\mathbf{x}_i) = y_i, \\ 0 & \text{if } h(\mathbf{x}_i) \neq y_i, \end{cases} \quad (19)$$

 TABLE I
 RATE OF ACCURACY OF ALL PATTERN RECOGNITION MODELS IN SOLVING EQ. (2) WITH DIFFERENT SCENARIOS OF INPUT FEATURES.

Features	N_f	l^2 -norm	NB	DT	Adaboost	RF
F.1	26	0.250	0.298	0.664	0.337	0.712
	50	0.260	0.308	0.635	0.356	0.731
	76	0.260	0.308	0.606	0.327	0.664
	100	0.260	0.308	0.606	0.327	0.702
	250	0.250	0.279	0.558	0.308	0.673
	500	0.250	0.289	0.519	0.317	0.664
F.2	26	0.404	0.394	0.500	0.385	0.615
	50	0.404	0.414	0.548	0.394	0.596
	76	0.433	0.423	0.539	0.385	0.567
	100	0.404	0.433	0.596	0.365	0.615
	250	0.442	0.433	0.539	0.442	0.635
	500	0.433	0.423	0.577	0.414	0.625
F.1+F.2	26	0.423	0.423	0.596	0.346	0.664
	50	0.404	0.462	0.539	0.414	0.577
	76	0.414	0.404	0.567	0.375	0.625
	100	0.414	0.414	0.558	0.414	0.587
	250	0.442	0.433	0.548	0.414	0.567
	500	0.452	0.423	0.539	0.433	0.596

The ACC measures the rate of correctness of the model and it is calculated as the sum of correct classifications divided by the total number of classifications.

B. Modeling the Exact GS Topology

The objective in this section is to evaluate the proposed system regarding its capability in modeling Eq. (2), i.e. predict the exact GS topology. To evaluate the accuracy of modeling Eq. (2), several classification models were trained according with the experimental settings (Section III-A). The results are given in terms of rate of accuracy and they are described in Table I. The left part of the Table I indicated the input features used in the modeling of Eq. (2), which are the F.1 (amplitude of DFT of the impedance signal), F.2 (phase angle of DFT of the impedance signal) and F.1+F.2 (both, amplitude and phase angle DFT of the impedance signal). The parameter N_f indicates the number of data points used to compute the DFT.

From the results, the best model is the RF classifier which reached a rate of accuracy of 0.73 when using the feature F.1 as input and a window of $N_f = 50$ for the DFT. This result reinforces the fact that the transient part (high frequency signal) contains the most information on the GS topology. On the other hand, the performance of l^2 -norm, NB, DT, Adaboost classifiers were not comparable with the RF classifier. As can be noticed, the performance of all classifiers tends to deteriorate with the increase of N_f ; large values of N_f suppresses the information contained in the high frequency signal, and this can be one reason that causes this performance deterioration.

In an overall analysis, the performance of the classifiers were not satisfactory. This is probably, due the fact that only few numbers of soils were used for training the pattern recognition classifiers (only 25 different soils) and with a large number of

TABLE II
RATE OF ACCURACY OF ALL PATTERN RECOGNITION MODELS IN SOLVING EQ. (3) WITH DIFFERENT SCENARIOS OF INPUT FEATURES.

Features	N_f	l^2 -norm	NB	DT	Adaboost	RF
F.1	26	0.500	0.539	0.865	0.856	0.885
	50	0.500	0.529	0.904	0.894	0.894
	76	0.500	0.529	0.904	0.894	0.894
	100	0.500	0.529	0.827	0.875	0.904
	250	0.500	0.529	0.817	0.856	0.885
	500	0.500	0.529	0.827	0.875	0.894
F.2	26	0.654	0.740	0.875	0.837	0.875
	50	0.673	0.721	0.837	0.846	0.856
	76	0.692	0.721	0.837	0.846	0.856
	100	0.702	0.731	0.837	0.837	0.865
	250	0.721	0.740	0.827	0.827	0.856
	500	0.721	0.721	0.875	0.837	0.856
F.1+F.2	26	0.664	0.548	0.865	0.894	0.885
	50	0.702	0.539	0.894	0.894	0.894
	76	0.702	0.539	0.894	0.894	0.894
	100	0.721	0.529	0.838	0.894	0.885
	250	0.721	0.529	0.827	0.875	0.894
	500	0.721	0.529	0.808	0.875	0.885

classes $C = 4$. To alleviate this issue, the next section will discuss the prediction of the approximated GS topology.

C. Modeling the Approximated GS Topology

The accuracy of Eq. (2) achieved only 0.73 (out of 26) in predicting the exact GS topology, probably motivated by the small number of soils to train the models, and the large number of classes $C = 4$. Differently of the previous approach, in this section the objective is to evaluate the equation Eq. (3), which will predict the approximated GS topology. By predicting the approximated GS topology, the number of classes is reduced to $C = 2$, then making the problem easier from the pattern recognition point of view. For that purpose, the GS topologies were grouped in two sets, $\Lambda^1 = \{\Theta^1, \Theta^2\}$ and $\Lambda^2 = \{\Theta^3, \Theta^4\}$. Then, the multiclass classification problem, was transformed in a binary classification problem, which is much easier to be solved. The results of predicting the approximated GS topology is shown in Table II

As can be seen in Table II, the Adaboost and RF classifiers have similar performance values, where both reached an accuracy of 0.904, which is satisfactory from a practical point of view. The feature F.1 provided the best prediction performance with the Adaboost and RF classifier and when $N_f = 50$ and $N_f = 100$, respectively.

D. Dependence of prediction performance on n_r

The objective of this section is to discuss the importance of Eq. (8), Section II-B, in improving the classification accuracy in the prediction of the exact and approximated GS topologies. For that purpose, the number of n_r was set to $1, \dots, 30$ and the accuracy was determined. For all values of n_r , the soil samples were selected randomly in 20 runs and the mean and standard deviation were considered for analysis.

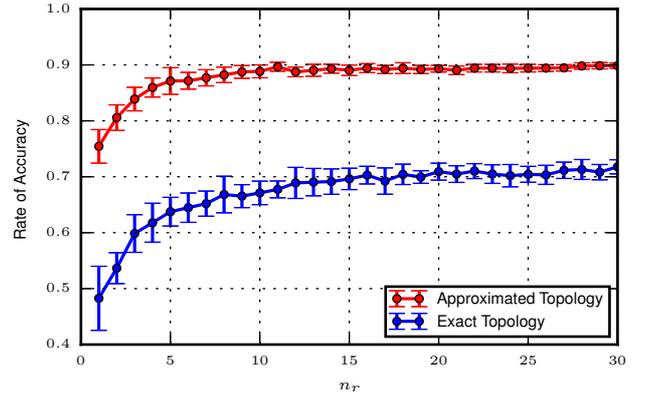


Fig. 7. Performance of RF classifier when using different values of n_r

The Fig. 7 shows the accuracy of RF classifier, with $N_f = 50$, regarding the number n_r to compose Eq. (8). As can be seen in Fig. 7, the proposed scheme reduces the effect of undersampling in the prediction performance with large values of n_r . In the case where $n_r = 1$, the average in prediction performance is around 0.5 for the case of prediction of exact GS topology and 0.75 for the case of prediction of approximated GS topology. These performances are improved with the increase of n_r , reaching its maximum value when $n_r = 30$, with the prediction of exact and approximated topology equal to 0.70 and 0.90, which considerably reduces the effect of the noisy due to the undersampling issue. Then, it is possible to conclude that the strategy proposed in Section II-B is efficient while dealing with the undersampling acquisition of voltage and current signals.

IV. DISCUSSION

The proposed approach is based on the pattern recognition framework, then it lacks the exact physical relation between the grounding impedance and the grounding topology, and conclusions can only be drawn based on available data; despite it is known in advance that the transient signal of grounding impedance contains information on ground rods. In this sense, experimental data provides strong evidence that grounding impedance can be employed to discriminate among one to four ground rods in straight line, when no information on soil is given. However, it is important to point out that limitation on acquisition data system, which collect part of transient signal with undersampling, plays a vital role in experimental results; Section II-B describes the adopted strategy to alleviate the undersampling problem on the classification results and Section III-D shows the effectiveness of the proposed strategy in the presented case study.

Moreover, the way the features \mathbf{x} are extracted from $Z(j\omega)$ have direct influence on the results, i.e. novel strategies for feature extraction can be adopted to minimize the issues related with intrinsic characteristics of GS, such as the frequency dependence of soil parameters and the soil variation along the grounding area. In the proposed system, the features are extracted based on the DFT of impedance signal, which has

show to perform well in the case study, even when dealing with a different number of soils.

V. CONCLUSIONS

In this work a system was proposed to predict the exact and approximated GS topology based on the pattern recognition framework. The proposed system is based in four main systems: excitation, data acquisition, feature extraction and pattern classification systems. In this work each of these components were detailed described.

The proposed system was evaluated in a set of controlled experiments conducted in 26 soils and with GS configurations with one, two, three and four rods in straight line. Three classification models, NB, DT, Adaboost and RF models were evaluated regarding its classification accuracy. The RF model reached the highest rate of accuracy, with the value of 0.73 for the exact topology and 0.90 for the approximated GS topology. The results in the case study provides strong evidence that different GS's can be discriminated using the information from grounding impedance, even with the undersampling issue of acquisition system. Furthermore, further investigation, in theoretical and practical level, should be performed to assure that the GS impedance can be used to discriminate for moderate difference among grounding systems (e.g. three interconnected rods, but with different topologies: triangle, in line; grounding systems composed by only (horizontal) conductors; etc.) or more complex GS topologies. In this context, the authors expects that by acquiring the exact transient signal it is going to be possible to discriminate among other different GS topologies, even for moderate difference among grounding systems or even more complex topologies.

From the pattern recognition point of view, the problem of predicting the exact GS topology can be seen as a multiclass problem with few numbers of samples and high dimensional data, which is very challenging. The accuracy of the proposed system can be further improved by getting experimental data in more soils and/or by improving the feature extraction step.

ACKNOWLEDGMENT

This work was supported by the Energy Company of Ceará (COELCE) project PD-0039-0045/2011. The authors would like to acknowledge the Ceará Foundation of Scientific and Technological Development (FUNCAP) and to the National Center of High Performance Processing of University Federal of Ceará (Cenapad-UFC).

REFERENCES

- [1] "IEEE recommended practice for grounding of industrial and commercial power systems," *IEEE Std 142-2007 (Revision of IEEE Std 142-1991)*, pp. 1–225, Nov 2007.
- [2] A. D. Conti and S. Visacro, "A simplified model to represent typical grounding configurations applied in medium-voltage and low-voltage distribution lines," in *IX International Symposium on Lightning Protection*, November 2007, pp. 1–6.
- [3] C. Mazzetti and G. M. Veca, "Impulse behavior of ground electrodes," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 9, pp. 3148–3156, Sept. 1983.
- [4] M. Ramamoorthy, M. M. B. Narayanan, S. Parameswaran, and D. Mukhedkar, "Transient performance of grounding grids," *IEEE Transactions on Power Delivery*, vol. 4, no. 4, pp. 2053–2059, Oct 1989.
- [5] L. Grcev and F. Dawalibi, "An electromagnetic model for transients in grounding systems," *IEEE Transactions on Power Delivery*, vol. 5, no. 4, pp. 1773–1781, Oct 1990.
- [6] A. F. Otero, J. Cidrás, and J. L. del Alamo, "Frequency-dependent grounding system calculation by means of a conventional nodal analysis technique," *IEEE Transactions on Power Delivery*, vol. 14, no. 3, pp. 873–878, Jul 1999.
- [7] L. Grcev and M. Popov, "On high-frequency circuit equivalents of a vertical ground rod," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 1598–1603, April 2005.
- [8] R. Alipio and S. Visacro, "Impulse efficiency of grounding electrodes: Effect of frequency-dependent soil parameters," *IEEE Transactions on Power Delivery*, vol. 29, no. 2, pp. 716–723, April 2014.
- [9] —, "Modeling the frequency dependence of electrical parameters of soil," *IEEE Transactions on Electromagnetic Compatibility*, vol. PP, no. 99, pp. 1–9, 2014.
- [10] J. C. Salari and C. Portela, "Grounding systems modeling including soil ionization," *IEEE Transactions on Power Delivery*, vol. 23, no. 4, pp. 1939–1945, October 2008.
- [11] —, "A methodology for electromagnetic transients calculation; an application for the calculation of lightning propagation in transmission lines," *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 527–536, January 2007.
- [12] J. C. Salari, "A methodology for computing transmission-line short circuits caused by direct and nearby ground lightning incidence-part ii: Application examples," *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 1586–1590, August 2014.
- [13] L. Grcev, "Impulse efficiency of grounding electrodes," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 441–451, January 2009.
- [14] —, "Modeling of grounding electrodes under lightning currents," *IEEE Transactions on Electromagnetic Compatibility*, vol. 51, no. 3, pp. 559–571, Aug 2009.
- [15] P. Yutthagowith, A. Ametani, N. Nagaoka, and Y. Baba, "Application of the partial element equivalent circuit method to analysis of transient potential rises in grounding systems," *IEEE Transactions on Electromagnetic Compatibility*, vol. 53, no. 3, pp. 726–736, Aug 2011.
- [16] S. Visacro, "A comprehensive approach to the grounding response to lightning currents," in *IEEE Transactions On Power Delivery*, January 2007, pp. 381–386.
- [17] L. Grcev and S. Grceva, "On hf circuit models of horizontal grounding electrodes," *IEEE Transactions on Electromagnetic Compatibility*, vol. 51, no. 3, pp. 873–875, Aug 2009.
- [18] A. M. Gaouda, S. H. Kanoun, M. M. A. Salama, and A. Y. Chikhani, "Pattern recognition applications for power system disturbance classification," *IEEE Transactions on Power Delivery*, vol. 17, no. 3, pp. 677–683, July 2002.
- [19] T. K. Abdel-Galil, M. Kamel, A. M. Youssef, E. F. El-Saadany, and M. M. A. Salama, "Power quality disturbance classification using the inductive inference approach," *IEEE Transactions on Power Delivery*, vol. 19, no. 4, pp. 1812–1818, October 2004.
- [20] N. W. A. Lidula and A. D. Rajapakse, "A pattern recognition approach for detecting power islands using transient signals - part i: Design and implementation," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 3070–3077, July 2010.
- [21] —, "A pattern-recognition approach for detecting power islands using transient signals - part ii: Performance evaluation," *IEEE Transactions on Power Delivery*, vol. 27, no. 3, pp. 1071–1080, July 2012.
- [22] F. E. Asimakopoulou, G. J. Tsekouras, I. F. Gonos, and I. A. Stathopoulos, "Estimation of seasonal variation of ground resistance using artificial neural networks," *Electric Power Systems Research*, vol. 94, pp. 113 – 121, 2013, lightning Protection of Advanced Energy Systems 7th Asia-Pacific International Conference on Lightning.
- [23] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.
- [24] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 1993.
- [25] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [26] M. S. Naidu and V. Kamaraju, *High Voltage Engineering*. 2009.
- [27] B. Babenko, M.-H. Yang, and S. Belongie, "A family of online boosting algorithms," in *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops) 2009*, September 2009, pp. 1346–1353.
- [28] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc. 2001.