Original papers

Automatic classification of plant electrophysiological responses to environmental stimuli using machine learning and interval arithmetic

Danillo Roberto Pereira[a], João Paulo Papa[b,*], Gustavo Francisco Rosalin Saraiva[a], Gustavo Maia Souza[c]

[a] Universidade do Oeste Paulista, Presidente Prudente, São Paulo, Brazil
[b] Universidade Estadual Paulista "Júlio de Mesquita Filho", Bauru, São Paulo, Brazil
[c] Universidade Federal de Pelotas, Pelotas, Rio Grande do Sul, Brazil

ABSTRACT

In plants, there are different types of electrical signals involving changes in membrane potentials that could encode electrical information related to physiological states when plants are stimulated by different environmental conditions. A previous study analyzing traits of the dynamics of whole plant low-voltage electrical showed, for instance, that some specific frequencies that can be observed on plants growing under undisturbed conditions disappear after stress-like environments, such as cold, low light and osmotic stimuli. In this paper, we propose to test different methods of automatic classification in order to identify when different environmental cues cause specific changes in the electrical signals of plants. In order to verify such hypothesis, we used machine learning algorithms (Artificial Neural Networks, Convolutional Neural Network, Optimum-Path Forest, k-Nearest Neighbors and Support Vector Machine) together Interval Arithmetic. The results indicated that Interval Arithmetic and supervised classifiers are more suitable than deep learning techniques, showing promising results towards such research area.

1. Introduction

Plants as sessile and modular organisms face the challenge to keep their stability growing in environments under constant changing (Souza et al., 2016). Since plants lack a central command to organize the environmental information gathered in each module (e.g., a branch root or a leaf), an efficient communication system has evolved in order to integrate local information (cell-to-cell communication) and to signalize through the plant body (long-distance communication) (Trewavas, 2003; Lüttge, 2012).

Long-distance communication, also referred to as systemic communication, can be triggered by different stimuli, such as biotic ones (e.g., systemic acquired resistance as a response to pathogens) or by abiotic stimuli (e.g., water deficit, heat, and salinity). Therefore, the ultimate goal of these systemic signaling is to activate response mechanisms in remote tissues, improving the ability of the whole plant to prepare its tissues to an upcoming challenge (Gilroy et al., 2014). Among the signals involved in long-distance communication, ROS (reactive oxygen species), calcium and electrical signals perform a central role (Baluska, 2016).

In plants, there are different types of electrical signals, which are

electrical activities involving changes in membrane potential, such as action potential (AP), variation potential (VP, or slow wave - SW), and system potential (Davies, 2006; Sukhova et al., 2017). APs are characterized by spike-like changes of the resting membrane potential and, independent of the stimulus strength, start propagating through the plant with a defined amplitude and velocity. Like in animals, APs seem to be all-or-nothing events (Fromm and Spanswick, 1993; Pyatygin et al., 2008). VPs differ from APs in various ways. VPs do not obey the all-or-nothing law, they are known as slow wave potentials (SWPs) with variable shape, amplitude and time frame. Moreover, the signals are related with the stimulus strength, and last for periods of 10 s up to 30 min (Zimmermann and Mithofer, 2013; Vodeneev et al., 2015). System potentials (SPs), in contrast to APs and VPs, reflect a systemic self-propagating hyperpolarization of the plasma membrane or depolarization of the apoplastic voltage. Like VPs, SPs have a magnitude and duration that are depended on the stimulus, but they are initiated via membrane hyperpolarization through the sustained activation of the proton pump. SPs are dependent on experimental conditions, and then they may occur under a very specific set of environmental conditions (Zimmermann and Mithofer, 2013; Choi et al., 2016). Furthermore, quite often all these signals are mixed altogether, which impairs a

proper signal analysis (van Bel et al., 2014; Saraiva et al., 2017). Strong evidences have demonstrated that bioelectrical signals play a central role in both cell-to-cell and long-distance communication in plants (Baluska et al., 2006; Zimmermann et al., 2009; van Bel et al., 2014), also supporting the ability to adjust their phenotypes to different environmental conditions (Fromm and Lautner, 2007; Gallè et al., 2015; Rìos-Rojas et al., 2014). For instance, Sukhov et al. (2014) and Magdalena et al. (2017) have demonstrated the role of electrical signals in the regulation of photosynthetic responses to different stimuli.

Very recently, Souza et al. (2017) proposed the concept of "plant electrome" based on the general proposition of "electrome" by De Loof (De Loof, 2016), describing the totality of the ionic currents in different scales of plant organization. By measuring low-voltage electrical signal using eletrophytography (EPG) (Debono, 2013; Souza et al., 2017), Souza et al. (2017) showed that different environmental stimuli could change some characteristics of the temporal dynamic of the electrical signaling, including the level of complexity. It was noticed that some specific frequencies, which were observed in non-stimulated plants, have disappeared after stimulation. Moreover, the environmental stimuli changed the type of color noise of the electrical signals. However, it was not clear if the different environmental cues (cold, low light, and osmotic stress) caused distinct effects on the plant signals (Souza et al., 2017).

Therefore, measuring the level of stress in plants is of crucial importance to a better understanding of their working mechanism. However, automatic plant stress identification by means of machine learning techniques has been considered recently only. Singh et al. (2016) presented an overview about machine learning tools and their applications in the context of biotic and abiotic stress traits classification. The main goal of such work is to guide the plant community when using machine learning techniques in the aforementioned situation. Two years earlier, Ma et al. (2014b) also considered a similar study, but in a more general way, and Ma et al. (2014a) employed machine learning to study stress-responsive transcriptomes in *Arabidopsis thaliana*. The experiments highlighted that such tools were able to outperform standard statistical approaches. Shaik and Ramakrishna (2014) used machine learning techniques to identify multiple stress conditions genes for broad resistance in rice, and Behmann et al. (2015) presented a review of different machine learning techniques applied for biotic stress identification in precision crop protection.

Chatterjee et al. (2014) established a relationship between the light stimulus and plant electrical response for different light stimuli intensity considering 19 different plants (17 *Zamioculcas zamiifolia* and 2 *Cucumis sativus plants*). The best results were obtained by Nonlinear Hammerstein-Wiener (NLHW) a good matching over others fitting methods. Later on, Chatterjee et al. (2015) classified three different types of stimuli (NaCl, H_2SO_4 and O_3) based on the response of electrical signals of tomatoes and cucumbers. The authors used different machine learning algorithms (FLDA - Fisher Linear Discriminant Analysis, QDA - Quadratic Discriminant Analysis, NB - Naive Bayes, and Mahalanobis Classifier) considering eleven features extracted from the electrical signal using linear and nonlinear methods. The best result was around 73.67% of recognition rate.

Chen et al. (2016) applied four classifiers (Template Matching, Artificial Neural Networks, Support Vector Machines and Deep Belief Networks) for the recognition of plant stimuli from electrical signals. The aforementioned work combined a waveform-based feature extractor and the Principal Component Analysis (PCA) approach, obtaining around 96% of recognition rate with Template Matching.

In this paper, we propose to use the concept of plant electrome (Souza et al., 2017) to automatically identify whether different environmental cues cause specific changes in the electrical signals of soybean plants. In order to verify such hypothesis, we considered using machine learning algorithms and arithmetic intervalar, a branch of mathematical tools that allows one to extend standard numbers to an interval representation. Therefore, the main contributions of this paper are:

- to use the plant electrome data as input for machine learning-based prediction of plant stress; and
- to employ deep learning techniques for plant stress identification.

The remainder of this paper is organized as follows. Sections 2 and 3 present the theoretical background and methodology used in this paper, as well as the results obtained using the proposed approach, respectively. Finally, Section 4 states conclusions and future works.

2. Materials and methods

2.1. Data acquisition

All datasets used herein to test the different methods of classification are part of the study published by Souza et al. (2017). The data consist of time series of low-voltage variation (ΔV in μV) measured in soybean plants subjected to different environmental stimuli: cold, low light and osmotic stress. The protocol of data acquisition was defined by Saraiva et al. (2017), using a signal amplifier (model MP36, Biopac Systems, US) inside a grounded Faraday cage. The measurements were carried out with one reference electrode attached to the grounded Faraday cage, and two electrodes inserted in the plants operating in a differential mode, where the instrumental amplifier cuts off the similar frequencies recorded in both electrodes. The sampling rate was 125 Hz with a high-pass filter settled to allow pass higher frequencies (>0.5 Hz), since the objective of that study was investigate the low-voltage noise that underlies the electrical signals (see more details in Saraiva et al. (2017)).

2.2. Datasets

The datasets described in the previous section were cropped to contain features per sample (signals obtained from the plants). Besides the large number of features, the signal is not so homogenous, therefore applying classical machine learning methods in the raw data is not advisable. To overcome this weakness, we applied some concepts of Arithmetic Intervalar to map raw data into lower-dimensional feature space.

In our work we consider four different datasets, as follows:

- **cold**: 67 signals obtained from plants in ideal conditions (without stress) and 76 signals obtained after cold stress.
- **low light**: 152 signals obtained in ideal conditions (without stress) and 118 signals after low light stress.
- **osmotic**: 123 signals obtained in ideal conditions (without stress) and 145 signals after osmotic stress.
- **all**: 342 signals obtained in ideal conditions (without stress), 76 signals after cold stress, 118 signals after low light stress, and 145 signals obtained after osmotic stress.[1]

Some examples of the signals from each class are depicted in Fig. 1.

2.3. Theoretical background

In this section, we present a brief theoretical background related to the machine learning and feature mapping techniques based on Interval Arithmetic used in this work.

2.3.1. Interval Arithmetic

The Interval Arithmetic (IA) was proposed by Moore in the 1960's, being the main idea to represent values as a range model instead of

---

[1] This dataset is a merge of the cold, low light and osmotic datasets. The main idea of this dataset is to verify whether the methods are able to differentiate the stress type or not.
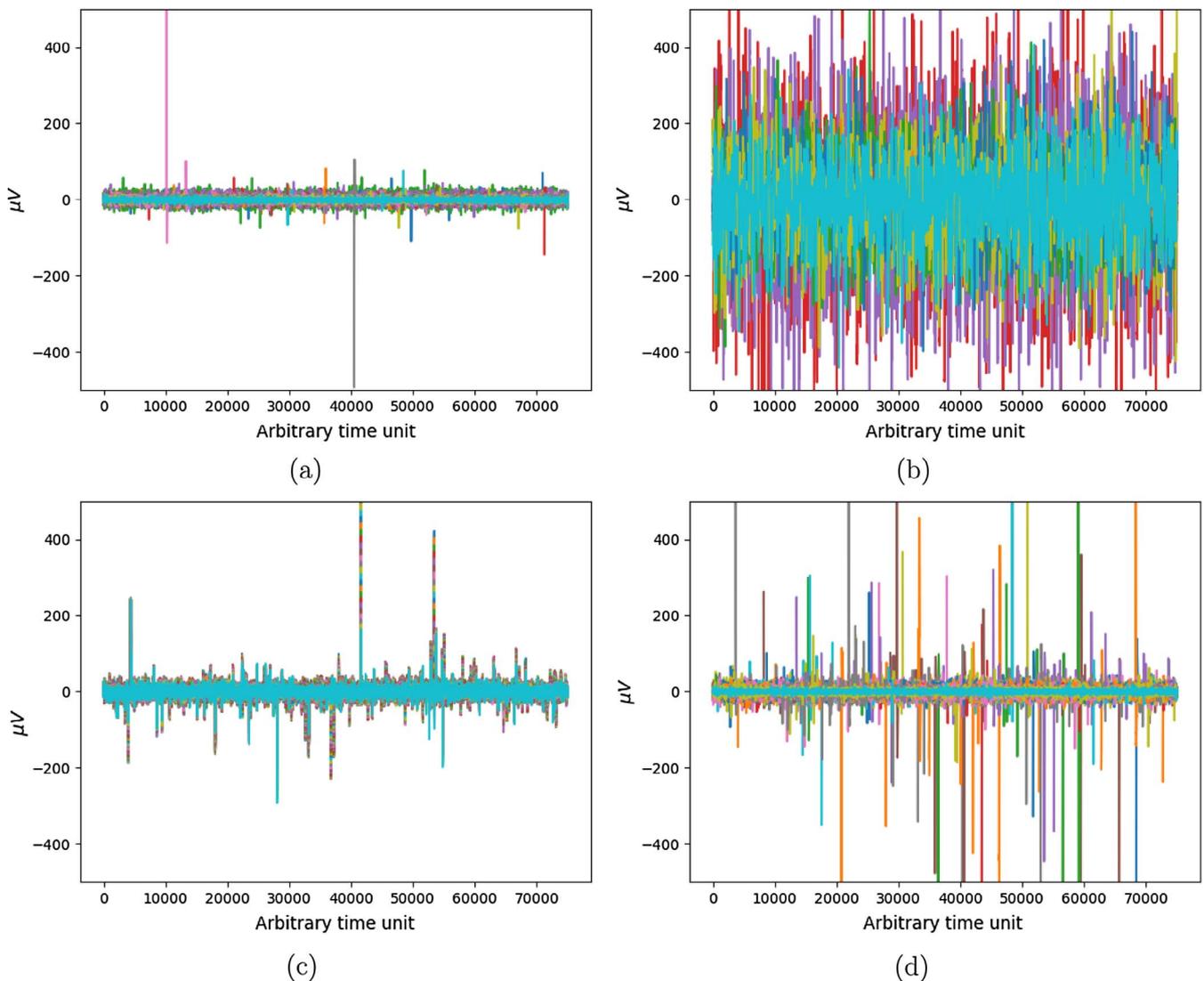
Fig. 1. Exemples of some signals obtained from the plants: (a) ideal condition, (b) after cold stress, (c) after low light, and (d) after osmotic stress.

single numbers. In such representation, an interval I is denoted by an non-empty real-valued range $[I_l, I_u]$, such that each interval encodes a subset of real numbers that satisfy the following condition:

$$I = \{x \in \Re | I_l \leqslant x \leqslant I_u\}. \tag{1}$$

As in the traditional arithmetic, the IA background defines a set of relations and operations (comparison, join, sum, and multiply) over the intervals (Moore, 1966; Moore and Yang, 1959; Moore, 1960, 1962, 1979). However, the Interval Arithmetic is more powerful than traditional arithmetic, since any real number x can be represented by the singular interval $[x,x]$. The Interval Arithmetic is a useful apparatus to provide representations of error bounds and uncertainty. This property is extremely interesting when working with the variations present in the EEG signals obtained from plants.

2.3.2. Machine learning techniques

2.3.2.1. Optimum-Path Forest.
The Optimum-Path Forest (OPF) classifier models the problem of pattern recognition as a graph partition task, in which a predefined set of samples from each class (*prototypes*) compete for a minimal path-cost attribution to the rest of the samples. Such process results in a collection of optimum-path trees rooted at the prototype nodes, building an optimum-path forest considering from all training samples. Test samples are classified through incrementally evaluating the optimum paths from the

prototypes, as though they were part of the forest, and assigning the labels of the most strongly connected roots. An OPF classifier can be designed as long as we use a smooth path-cost function (Falcão et al., 2004). Although there are two different versions of the supervised OPF classifier (Papa et al., 2009, 2017), in this paper we make use of the former and most widely used approach proposed by Papa et al. (2009) and further enhanced by Papa et al. (2012). The OPF classifier has been used in a number of applications in the last years, and particularly for leaf-based aquatic weed recognition (Pereira et al., 2012).

2.3.2.2. Multilayer perceptron.
An Artificial Neural Network with Multilayer Perceptons (ANN) is a feedforward neural network composed of several neuron layers aiming to solve multi-class problems (Haykin, 1999). The input to each layer is a weighted sum of the output from the previous layer, and the number of neurons in the first layer is equal to the number of features of the input, while the number of neurons in the last layer is equal to the number of classes. Such approaches can learn highly non-linear models that map inputs to their corresponding outputs.

2.3.2.3. k-Nearest neighbors.
The k-nearest neighbors (k-NN) is a simple but effective technique that works pretty well in many different applications (Coomans and Massart, 1982; Hall et al., 2008). In contrast to the OPF, the k-NN uses all training samples as prototypes,
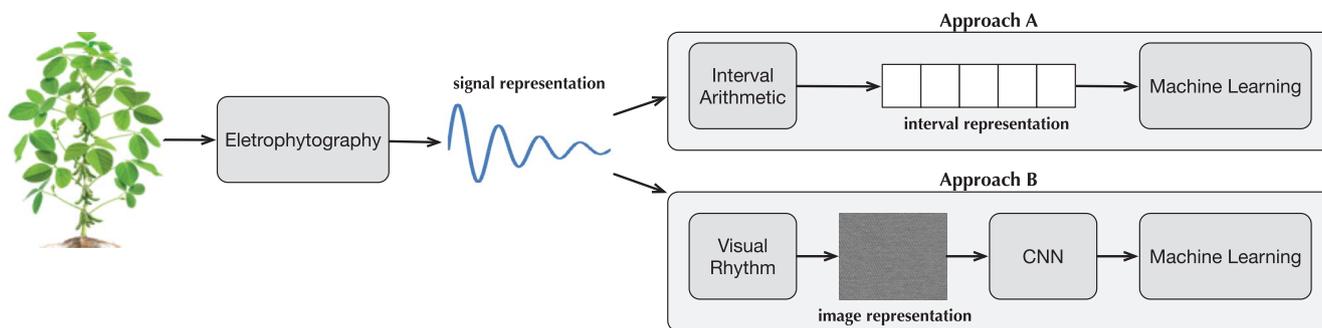
Fig. 2. Proposed pipeline for automatic plant stress identification.

and requires the input parameter k that establishes the number of neighbors that contribute to the classification of a given sample. For such purpose, a given test sample is labeled with the most frequent label in its k-neighborhood. Notice that for $k = 1$, the testing sample is assigned to the class of its closest training sample.

2.3.2.4. Support Vector Machines.
The Support Vector Machines (SVM) approach based on the principle of structural risk minimization (Vapnik, 1999; Schölkopf and Smola, 2002; Cortes and Vapnik, 1995), aiming at establishing an optimal discriminative function among two classes of patterns while accomplishing the trade-off between generalization and overfitting. The standard SVM training algorithm constructs the optimal hyperplane separating a two-class feature space (Vapnik, 1999). However, some problems may require a more robust approach. In order to extend from linear to nonlinear classification, the *kernel trick* is used (Schölkopf and Smola, 2002), where kernel functions nonlinearly map input data into higher-dimensional feature spaces in a computationally-efficient manner.

2.3.2.5. Convolutional Neural Networks.
Convolutional Neural Networks (CNN) are a special type of artificial neural networks that have been extensively used for unsupervised feature learning (Lecun et al., 1998). The main idea is to employ several layers of pooling and convolution operators in order to automatically extract features from input data that are invariant to some geometric operations (e.g., rotations and translations).

2.4. Methodology

In this work, we considered two distinct methodologies to learn features from the signal extracted from the plants: (i) the first one, hereinafter called approach A, is based on Interval Arithmetic (IA), and (ii) the second approach (approach B) is based on Visual Rhythm and CNNs. After extracting features, they are used to feed supervised machine learning techniques in order to classify whether the plant has

been affected by stress or not. Fig. 2 illustrates the aforementioned pipeline. The next sections describe in details the aforementioned approaches.

2.4.1. Approach A
As depicted in Fig. 2, approach A aims at decomposing the signal into an intervalar representation based on the theory presented in Section 2.3.1. Roughly speaking, the mapping process is straightforward and window-based. Let $S = \{s_1, s_2, ..., s_n\}$ be the discrete representation of the signal extracted from the soybean, where s_i stands for an energy acquisition at time step i. Additionally, let s be the window size, such that the whole signal S is equally divided into $m = \lfloor n/s \rfloor$ bins $b_k, k = 1, 2, ..., m$.

The ideia of the intervalar representation is to represent each bin b_k as a triplet $T_k = [min_k, avg_k, max_k]$, where min_k and max_k stand for the minimum and the maximum values present in b_k, and avg_k denotes for the average value of the numbers in b_k. Therefore, the signal S is finally represented as a set of m triplets

$$S' = \{[min_1, avg_1, max_1], [min_2, avg_2, max_2], ..., [min_m, avg_m, max_m]\},$$

which is further used to feed the supervised machine learning techniques. Fig. 3 illustrates the aforementioned procedure.

The proposed approach reduces considerably the number of features that describe the signal, which affects the computational cost as well. We performed tests with four different window sizes, say that 1000, 5000, 15,000 and 25,000, being the window size as of 15,000 the one that obtained the best results. Therefore, all experiments were conducted using such configuration. In this context, the number of features were reduced from to 15 (75,000/15,000 = 5 buckets, being each one represented by a triplet).

2.4.2. Approach B
The second approach aims at encoding the signal as an image by using the so-called Visual Rhythm (Almeida et al., 2015), which basically stacks the signal into rows in order to generate the image.
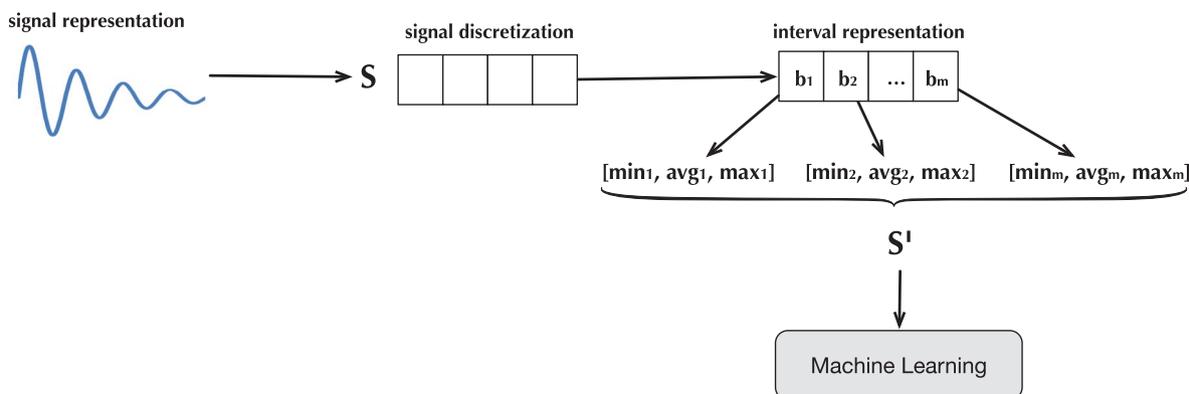

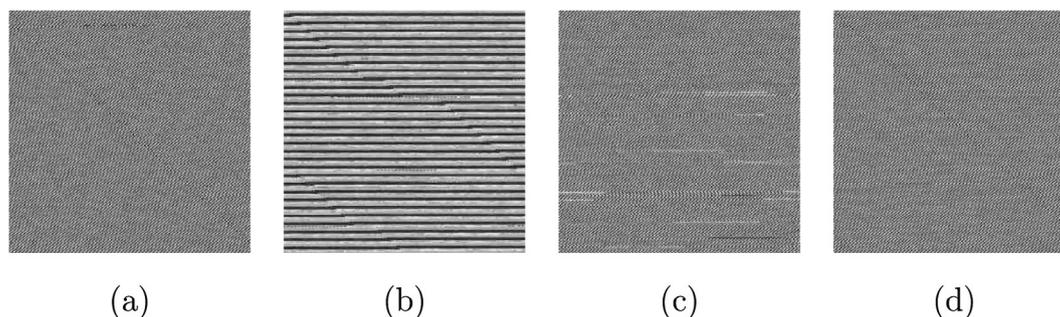**Fig. 3.** Approach A used for interval-based signal representation.

Fig. 4. Some examples of signals mapped into images under: (a) ideal conditions, (b) after cold stress, (c) after low light stress, and (d) after osmotic stress. The images were equalized for the sake of visualization purposes.

Therefore, each signal needs to be normalized into [0,1] for further being resized into a gray-scale squared matrix. Finally, one obtains a 273×273 image, were each signal value denotes a pixel in this image. Some examples of images mapped using this methodology are presented in Fig. 4. Clearly, one can observe the different patterns encoded by the images with respect to different stress conditions.

Later on, the images generated using the aforementioned pipeline are then employed to feed an CNN technique. In this work, we performed tests with two different classical architectures, as follows:

- **ImageNet:** composed of 5 convolution layers, 5 pooling layers and 2 normalization layers. It is also constituted by 5 ReLU layers among the convolution ones, 2 inner product layers, 2 dropout layers, 1 softmax loss layer and 1 accuracy layer for testing purposes.
- **Cifar-10:** a quick version is used, composed of 3 convolution layers and 3 pooling layers. It is also constituted by 3 ReLU layers among the convolution ones, 2 inner product layers, 1 softmax loss layer and 1 accuracy layer for testing intentions.

However, as the Cifar-10 model obtained the best results (significantly better than ImageNet), all experiments using CNN were realized using this configuration.

2.5. Statistical evaluation

In regard to the comparison assessment, we used an accuracy measure proposed by Papa et al. (2009), which is similar to the Kappa index (Cohen, 1960), but being more restrictive. If there are two classes, for example, with very different sizes and a classifier always assigns the label of the largest class, its accuracy will fall drastically due to the high error rate on the smallest class. The accuracy is measured by taking into account the classes may have different sizes in the test set. Also, we experiments were executed in 20 runs with different training and test sets for the further statistical evaluation by means of the Wilcoxon signed-rank test with significance of 0.05 (Wilcoxon, 1945).

2.6. Setting-up machine learning techniques

We considered the LibSVM library (Chang and Lin, 2011) for the implementation of SVM classifier. The hyperparameters C and σ were determined via a 5-fold cross-validation grid-search in the range $[2^{-15}, 2^{15}]$ with steps of 2, for each input parameter.

With respect to the ML classifier, we used the Fast Artificial Neural Network (FANN) library with two hidden layers consisting of eight neurons each (Nissen, 2003). The network was trained using the backpropagation algorithm (Haykin, 1999) with 70,000 epochs and the desired error as of 10^{-4}. The learning rate was set to 0.1.

In regard to k-NN technique, we defined k as the best value of a grid-search within the range $[1, \lfloor \frac{z}{5} \rfloor]$ with steps of two, where z is the number

of training samples. Finally, with respect to OPF, we used the LibOPF library.[2]

3. Results

In this section, we present the experimental results concerning the task of automatic identifying stress-like patterns before and after stress environment (i.e., cold, low light and osmotic). As aforementioned, we applied five different supervised classification algorithms: ANN, CNN, OPF, k-NN and SVM. Additionally, we considered the IA mapping experiment (procedure A) for ANN, OPF, k-NN, and SVM classifiers.

The main goal is to find an algorithm that can be successfully applied to learn a pattern even when trained on small datasets. Therefore, we partitioned the original dataset into training sets of increasing sizes, ranging from 10% to 90% of the original dataset. For each dataset, we executed each algorithm 20 times using a holdout configuration. Fig. 5 presents the results considering the "cold", "low light", "osmotic", and "all" datasets, respectively.

One can observe that good performances were obtained in the four datasets by all methods, except ANN that achieved the worst results in "cold" and "low light" datasets. As expected, all techniques were able to improve their results upon larger training sets, mainly CNNs, which are complex models and usually need a considerable amount of data for leaning (see Fig. 5d). Since "all" dataset comprises all stress-like patterns (i.e., it seems to pose a greater challenge), CNNs obtained the worst results over small training sets.

Table 1 presents the average results using 90% of the datasets for training purposes, being the most accurate techniques in bold according to the Wilcoxon statistical test. One can observe that SVM obtained the best results for all datasets, followed by k-NN that obtained statistically similar results in the "cold" and "osmotic" datasets. OPF classifier achieved good accuracies for three datasets, and CNN also obtained good results for "cold" and "low light" datasets, although it obtained the worst results for the "osmotic" dataset. Roughly speaking, one can observe that AI combined with supervised classifiers can provide better results than deep learning techniques. The recognition rates showed one can obtain promising results concerning the task of automatically identifying stress-like patterns.

Fig. 6 depicts the confusion matrices using 90% of the datasets for training purposes with SVM classifier. As one can observe, the signs of plants after cold stress have a peculiar pattern that is easier to differentiate than others stress conditions since it obtained better results. Additionally, the "low light" and "osmotic" datasets presented similar results. The experiment with "all" dataset showed that is possible to detect patterns among different stress signals. However, one can observe a confusion between "low light" and "no stress" conditions (Fig. 6d), which may deteriorate the final classifier's accuracy.

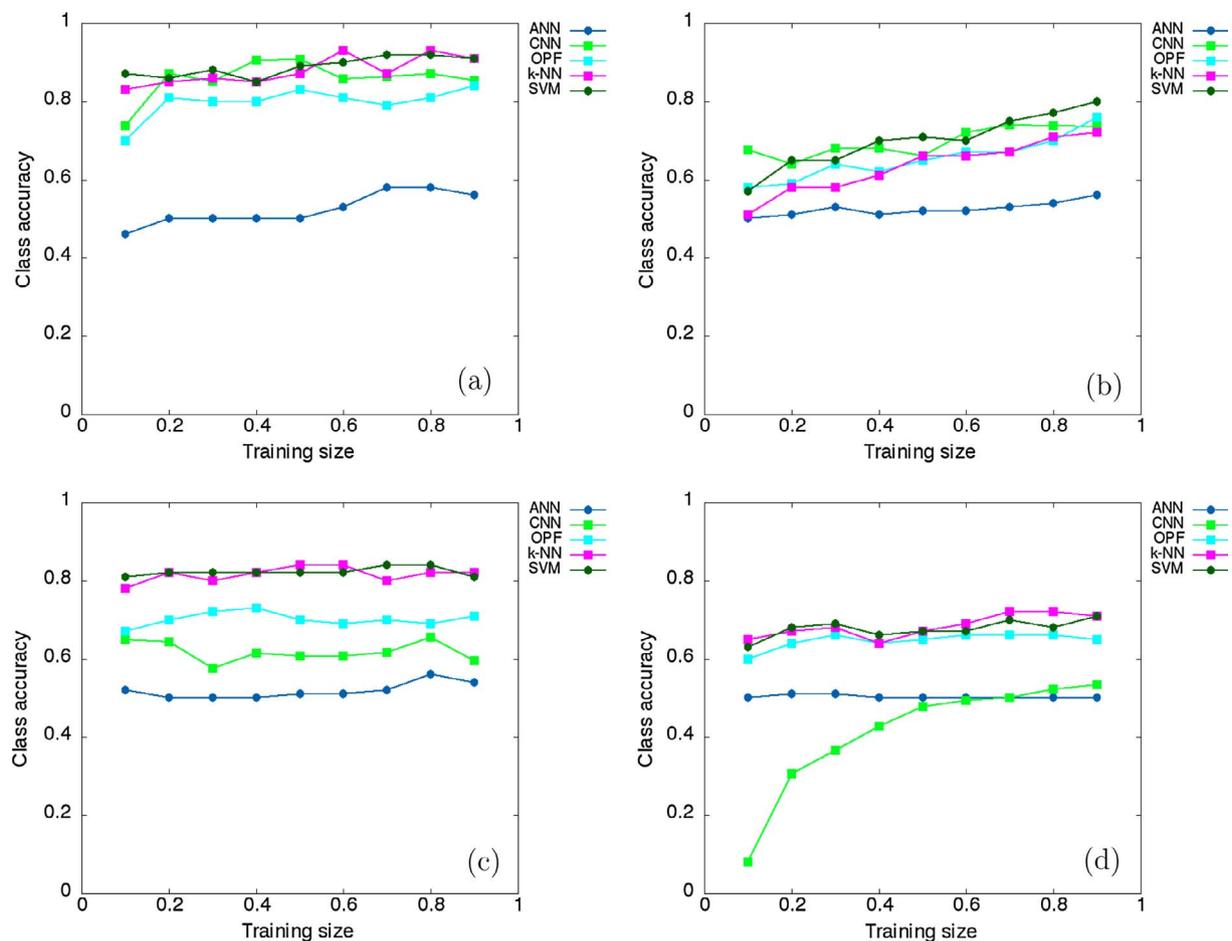
[2] https://github.com/jppbsi/LibOPF.

Fig. 5. Effect of different training set sizes for classification purposes over: (a) "cold", (b) "low light", (c) "osmotic", and (d) "all" datasets.

Table 1
Average recognition rates considering 90% of the samples for training purposes.

	Datasets			
	Cold	Low light	Osmotic	"all"
ANN	53.33 ± 0.00	58.57 ± 1.75	53.57 ± 0.00	63.11 ± 2.02
CNN	85.33 ± 9.07	73.57 ± 9.71	49.63 ± 6.22	53.33 ± 2.19
OPF	84.00 ± 5.33	77.86 ± 6.14	67.86 ± 6.39	65.23 ± 3.21
k-NN	**90. 67 ± 5. 33**	74.29 ± 5.25	**81. 43 ± 5. 25**	**71. 22 ± 0. 65**
SVM	**90. 67 ± 5. 33**	**80. 71 ± 7. 35**	80. 97 ± 5. 80	70. 98 ± 1. 01

4. Discussion

Different environmental cues, biotic or abiotic, trigger electrophysiological responses in plants, which are basically an ion imbalance across plasma membranes leading to a voltage transient (Maffei and Bossi, 2006; Huber and Bauerle, 2016). While VPs and SPs are mostly triggered by invasive injurious stimuli, such as wounding, APs are elicited as responses to non-invasive stimuli including irradiation, moisture, and temperature (Fromm and Lautner, 2007; Rìos-Rojas et al., 2014; Sukhov et al., 2017). Herein, the data analysed were prevenient from plants subjected to three abiotic stimuli (cold, osmotic, and low light) that affected the temporal dynamic of the low-voltage signals (Souza et al., 2017). The measured signals are supposed to be the resultant of all electrical activity of the cells around the electrodes, likely emerging from the superposition of cells APs propagated via plasmodesmata linking neighbouring cells (Volkov, 2012; Zimmermann and Mithofer, 2013). According to Debono (2013), such microvolted activity represents the resulting macroscopic currents locally sustained by plant receptor-channels acting through the different membrane compartments of synchronized protoneural networks, which was named "plant electrome" by Souza et al. (2017).

The previous spectral analysis (Souza et al., 2017) showed that before stimuli the runs exhibited three ranges of dominant frequencies (0.3–1.5, 3.8–5.7, and 9.3–11.1 Hz) and, after stimuli, only the lower frequencies remained. The β exponents calculated from power spectrum density function indicated that all time series showed long-range temporal correlation. However, β values from non-stimulated plants were lower, suggesting that the electrical signalling after plant stimulation showed more persistence (for $\beta \neq 0$ and $\beta \neq 2$), enabling long distance signalling (Souza et al., 2017). Besides some changes in the type of noise (reddened to black) and in the spectral analysis, it was observed an interesting appearance of "bursts of spikes" up to $500\,\mu V$ (considering the baseline signal around $10\,\mu V$) mostly in the runs scored after stimuli. The distribution of these higher voltage variations was different regarding the specific stimuli. While the spikes observed under low temperature showed an exponential distribution, under low light and, mainly, under osmotic stimulus they followed a power law, indicating that the spikes have no characteristic size (Souza et al., 2017). Those bursts of spikes were supposed as resulting from a self-organized collective behaviour among groups of cells Saraiva et al. (2017). The complex nature of these microvolted potential and their coherent responses to environmental cues suggests that H^+ electrogenic pumps, and others ionic receptors channels (e.g. Ca^{2+} and K^{1+}), are involved in such responses (Debono, 2013). Following a membrane electrical event, secondary messengers are initiated, leading to a downstream signalling cascade that induce many different metabolic changes, likely mediated by calcium waves and ROS (Maffei et al., 2007; Gilroy et al., 2014; Choi et al., 2016). Evidences have indicated
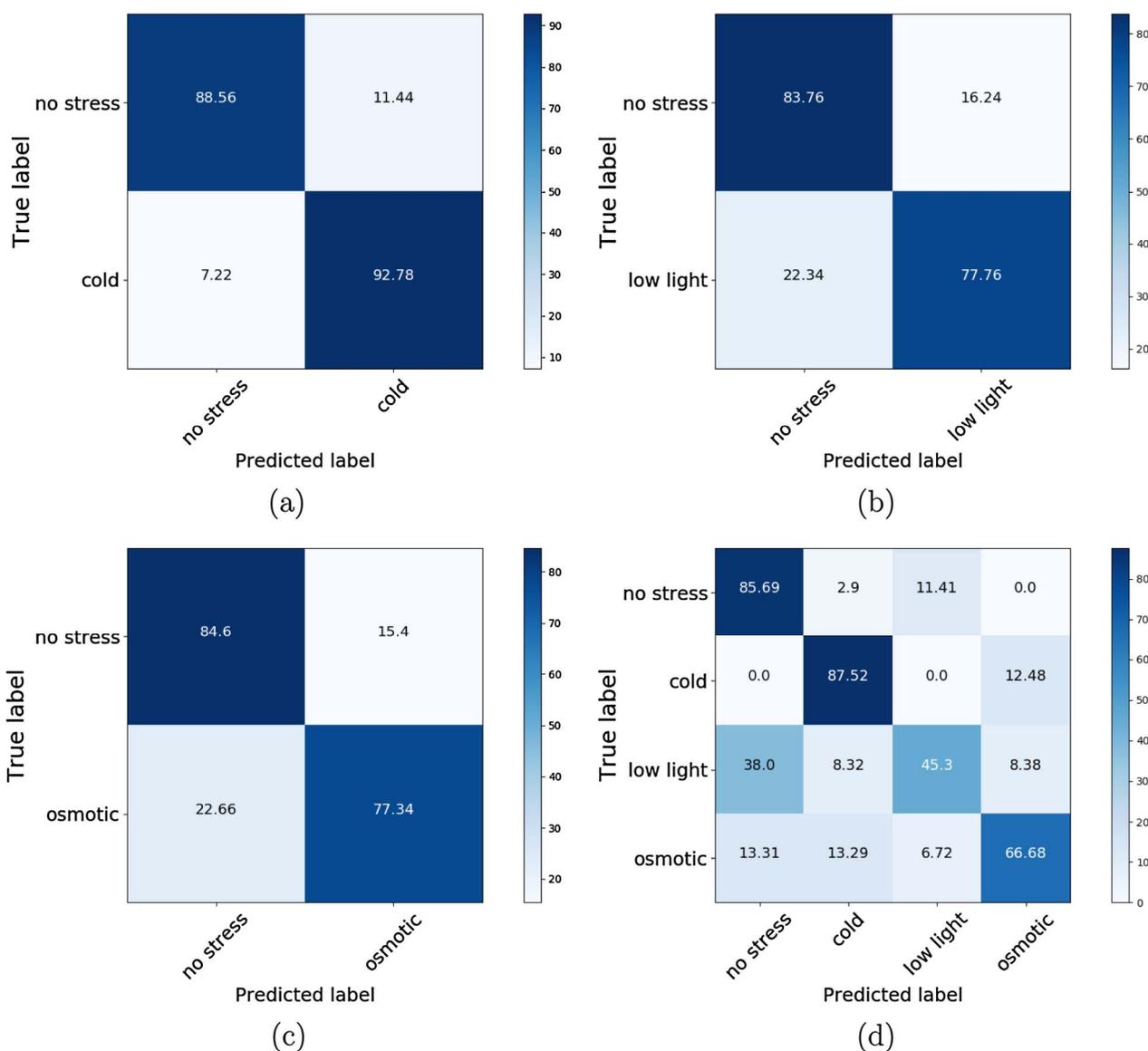
Fig. 6. Confusion matrices considering 90% for training purposes and SVM classifier over: (a) "cold", (b) "low light", (c) "osmotic", and (d) "all" datasets.

that electrical signalling regulates a fundamental physiological process, the well-known photosynthesis. Although the mechanisms of photosynthetic activation are still unclear, two different mechanisms have been proposed for the fast inactivation of photosynthesis. The first is calcium dependent, in which the increase of Ca^{2+} cytosolic can inactivate the Calvin-Benson cycle; and the second supposes that electrical activity changes the intra- and extracellular pH, decreasing the cytoplasmic pH by inactivation of H^+-ATPase, which reduces the photochemical yields (Sukhov, 2016).

Accordingly, our hypothesis is that the low voltage variations affected by different abiotic stimuli could be representative of the electrophysiological state of the plants as affected by specific stimulus allowing an algorithmic classification. In this work, we introduced two different methods for the automatic classification of plant stress based on the plant electrome. The first approach makes use of Interval Arithmetic to reduce the dimensionality of the input signal, and the second one employs deep learning for unsupervised feature learning. Chen et al. (2016) have studied the efficiency of some deep learning methods, such as ANN and SVM, comparing with template matching method to classify APs. In that study, after the elimination of some artifacts present in the raw signals, they extracted 19 features from the AP's signals, including time-domain, frequency-domain, statistics

characteristics, and nonlinear features. Then, by reducing the dimensionality with PCA, the different classifiers were tested against non-AP signals. Although the best performance have been reached with template matching algorithm (96.0% accuracy), ANN and SVM reached a maximum accuracy of 84.1% and 75.8%, respectively, which is pretty close to our best results (Table 1). However, differently from Chen et al. (2016) study, our classification algorithms were applied to, basically, continuous time series of raw data of low voltage variation, showing very complex dynamics (Saraiva et al., 2017; Souza et al., 2017). Chatterjee et al. (2015) have explored strategies for the classification of raw non-stationary plant electrical signal after different environmental stimuli (mV) by univariate and bivariate feature-based classification using five different discriminant analysis classifiers. The classification has reached a best average accuracy of 70% using variance and skewness as feature pairs, and an accuracy of 73.67% using variance and IQR as feature pairs in a diagquadratic classifier. Thus, as far as we know, this is the first time that microvolted runs are classified by machine learning techniques.

5. Conclusions

The proposed approaches herein achieved surprisingly good

classification performance on a relatively small set of training data, thus confirming the existence of a clear pattern captured by the plant electrome to identify potential plant stress condition. The experiments also suggested that Interval Arithmetic and supervised classifiers are more suitable than deep learning. Such achievement is very important since the plant stress could be identified before the plant has demonstrated physical injuries, such as leaf fall, and decreased productivity, among other factors. Our future works will be guided to develop better models that can distinguish all three stressful conditions at the same time.

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compag.2017.12.024.

References

Almeida, J.G., dos Santos, J.A., Alberton, B., Morellato, L.P.C., Torres, R.S., 2015. Phenological visual rhythms: Compact representations for fine-grained plant species identification. Pattern Recogn. Lett. http://dx.doi.org/10.1016/j.patrec.2015.11.028.

Baluska, F., 2016. Long-Distance Systemic Signalling and Communication in Plants. Springer International Publishing, Berlin.

Baluska, F., Mancuso, S., Volkmann, D., 2006. Springer International Publishing, Berlin.

Behmann, J., Mahlein, A.K., Rumpf, T., Römer, C., Plümer, L., 2015. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precision Agric. 16, 239–260.

van Bel, A.J.E., Furch, A.C.U., Will, T., Buxa, S.V., Musetti, R., Hafke, J.B., 2014. Spread the news: systemic dissemination and local impact of Ca^{2+} signals along the phloem pathway. J. Exp. Bot. 65, 1761. http://dx.doi.org/10.1093/jxb/ert425.

Chang, C.C., Lin, C.J., 2011. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

Chatterjee, S.K., Das, S., Maharatna, K., Masi, E., Santopolo, L., Mancuso, S., Vitaletti, A., 2015. Exploring strategies for classification of external stimuli using statistical features of the plant electrical response. J. R. Soc. Interface 12.

Chatterjee, S.K., Ghosh, S., Das, S., Manzella, V., Vitaletti, A., Masi, E., Santopolo, L., Mancuso, S., Maharatna, K., 2014. Forward and inverse modelling approaches for prediction of light stimulus from electrophysiological response in plants. Measurement.

Chen, Y., Zhao, D.J., Wang, Z.Y., Wang, Z.Y., Tang, G., Huang, L., 2016. Plant electrical signal classification based on waveform similarity. Algorithms 9http://dx.doi.org/10.3390/a9040070.. http://www.mdpi.com/1999-4893/9/4/70.

Choi, W., Hilleary, R., Swanson, S.J., Kim, S., Gilroy, S., 2016. Rapid, long-distance electrical and calcium signaling in plants. Annu. Rev. Plant Biol. 287–307.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Measur. 20, 37–46.

Coomans, D., Massart, D., 1982. Alternative k-nearest neighbour rules in supervised pattern recognition. Anal. Chim. Acta 136, 15–27.

Cortes, C., Vapnik, V., 1995. Support vector networks. Machine Learn. 20, 273–297.

Davies, E., 2006. Electrical Signals in Plants: Facts and Hypotheses. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 407–422. http://dx.doi.org/10.1007/978-3-540-37843-3_17.

De Loof, A., 2016. The cell self-generating electrome: The biophysical essence of the immaterial dimension of life. Commun. Integr. Biol. e1197446.

Debono, M., 2013. Dynamic protoneural networks in plants: A new approach of spontaneous extracellular potential variations. Plant Signall. Behav.

Falcão, A.X., Stolfi, J., Lotufo, R.A., 2004. The image foresting transform: Theory, algorithms, and applications. IEEE Trans. Pattern Anal. Mach. Intell. 26, 19–29.

Fromm, J., Lautner, S., 2007. Electrical signals and their physiological significance in plants. Plant, Cell Environ. 30, 249.

Fromm, J., Spanswick, R., 1993. Characteristics of action potentials in willow (salix viminalis l). J. Exp. Bot. 1119–1125.

Gallè, A., Lautner, S., Flexas, J., Fromm, J., 2015. Environmental stimuli and physiological responses: The current view on electrical signalling. Environ. Exp. Bot. 114, 15–21. http://dx.doi.org/10.1016/j.envexpbot.2014.06.013.

Hall, P., Park, B.U., Samworth, R.J., 2008. Choice of neighbor order in nearest-neighbor classification. Ann. Statist. 36, 2135–2152.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, second ed. Prentice Hall.

Huber, A.E., Bauerle, T.L., 2016. Long-distance plant signaling pathways in response to multiple stressors: the gap in knowledge. J. Exp. Bot. 67.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.

Lüttge, U., 2012. Modularity and emergence: biology's challenge in understanding life. Plant Biol. 14, 865–871. http://dx.doi.org/10.1111/j.1438-8677.2012.00659.x.

Magdalena, S.H., Maria, L., Stanisław, K., 2017. Electrical signaling, photosynthesis and systemic acquired acclimation. Front. Physiol. 8, 684.

Ma, C., Xin, M., Feldmann, K.A., Wang, X., 2014a. Machine learning–based differential network analysis: A study of stress-responsive transcriptomes in arabidopsis. Plant Cell Online 26, 520–537.

Ma, C., Zhang, H.H., Wang, X., 2014b. Machine learning for big data analytics in plants. Trends Plant Sci. 19, 798–808.

Maffei, E.M., Bossi, S., 2006. Electrophysiology and Plant Responses to Biotic Stress. Springer, Berlin.

Maffei, M.E., Mithofer, A., Boland, W., 2007. Before gene expression: early events in plant-insect interaction. Trends Plant Sci.

Moore, R.E., 1960. Book review: On Numerical Approximation (Proceedings of a Symposium) (R.E. Langer, ed.) 2, 49–50. http://dx.doi.org/10.1137/1002015.

Moore, R.E., 1962. Interval Arithmetic and Automatic Error Analysis in Digital Computing. Ph. D. dissertation. Department of Mathematics, Stanford University. Stanford, CA, USA. Also published as Applied Mathematics and Statistics Laboratories Technical Report No. 25.

Moore, R.E., 1966. Interval Analysis. Prentice-Hall, Englewood Cliff, New Jersey.

Moore, R.E., 1979. Methods and Applications of Interval Analysis. SIAM Studies in Applied Mathematics.

Moore, R.E., Yang, C.T., 1959. Interval Analysis I. Technical Document LMSD-285875. Lockheed Missiles and Space Division. Sunnyvale, CA, USA.

Nissen, S., 2003. Implementation of a Fast Artificial Neural Network Library (FANN). Department of Computer Science University of Copenhagen (DIKU). Software available at http://leenissen.dk/fann/.

Papa, J.P., Albuquerque, V.H.C., Falcão, A.X., Tavares, J.M.R.S., 2012. Efficient supervised Optimum-Path Forest classification for large datasets. Pattern Recogn. 45, 512–520.

Papa, J.P., Falcão, A.X., Suzuki, C.T.N., 2009. Supervised pattern classification based on Optimum-Path Forest. Int. J. Imaging Syst. Technol. 19, 120–131.

Papa, J.P., Fernandes, S.E.N., Falcão, A.X., 2017. Optimum-path forest based on k-connectivity: Theory and applications. Pattern Recogn. Lett. 87, 117–126.

Pereira, L.A.M., Nakamura, R.Y.M., Souza, G.F.S., Martins, D., Papa, J.P., 2012. Aquatic weed automatic classification using machine learning techniques. Comput. Electron. Agric. 87, 56–63.

Pyatygin, S.S., Opritov, V.A., Vodeneev, V.A., 2008. Signaling role of action potential in higher plants. Russ. J. Plant Physiol. 285–291.

Rios-Rojas, L., Tapia, F., Gurovich, L.A., 2014. Electrophysiological assessment of water stress in fruit-bearing woody plants. J. Plant Physiol. 171, 799–806. http://dx.doi.org/10.1016/j.jplph.2014.02.005. URL http://www.sciencedirect.com/science/article/pii/S0176161714000509 .

Gilroy, S., Suzuki, N., Miller, G., Choi, W.G., Toyota, M., Devireddy, A.R., Mittler, R., 2014. A tidal wave of signals: calcium and ROS at the forefront of rapid systemic signaling. Trends Plant Sci.

Saraiva, G.F.R., Ferreira, A.S., Souza, G.M., 2017. Osmotic stress decreases complexity underlying the electrophysiological dynamic in soybean. Plant Biol. 5, 702–708.

Schölkopf, B., Smola, A.J., 2002. Learning with Kernels. MIT Press, Cambridge, MA.

Shaik, R., Ramakrishna, W., 2014. Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. Plant Physiol. 164, 481–495.

Singh, A., Ganapathysubramanian, B., Singh, A.K., Sarkar, S., 2016. Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci. 21, 110–124.

Souza, G.M., Bertolli, S.C., Lüttge, U., 2016. Hierarchy and Information in a System Approach to Plant Biology: Explaining the Irreducibility in Plant Ecophysiology. Springer International Publishing, Cham, pp. 167–186. http://dx.doi.org/10.1007/978-3-319-25688-7_5.

Souza, G.M., Ferreira, A.S., Saraiva, G.F.R., Toledo, G.R.A., 2017. Plant electrome can be pushed toward a self-organized critical state by external cues: Evidences from a study with soybean seedlings subject to different environmental conditions. Plant Signal. Behav. 12 3, e1290040.

Sukhov, V., 2016. Electrical signals as mechanism of photosynthesis regulation in plants. Photosynth. Res.

Sukhov, V., Gaspirovich, V., Mysyagin, S., Vodeneev, V., 2017. High-temperature tolerance of photosynthesis can be linked to local electrical responses in leaves of pea. Front. Physiol.

Sukhov, V., Sherstneva, O., Surova, L., Katicheva, L., Vodeneev, V., 2014. Proton cellular influx as a probable mechanism of variation potential influence on photosynthesis in pea. Plant, Cell Environ. 2532–2541.

Sukhova, E., Akinchits, E., Sukhov, V., 2017. Mathematical models of electrical activity in plants. J. Membr. Biol.

Trewavas, A., 2003. Aspects of plant intelligence. Ann. Bot. 92, 1. http://dx.doi.org/10.1093/aob/mcg101.

Vapnik, V.N., 1999. An overview of statistical learning theory. IEEE Trans. Neural Netw. 10, 988–999.

Vodeneev, V., Akinchits, E., Sukhov, V., 2015. Variation potential in higher plants: Mechanisms of generation and propagation. Plant Sign. Behav.

Volkov, A., 2012. Plant Electrophysiology: Signaling and Responses. Plant Electrophysiology. Springer Berlin Heidelberg URL https://books.google.com.br/books?id=OjlffVX665EC .

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bull. 1, 80–83.

Zimmermann, M.R., Maischak, H., Mithofer, A., Boland, W., Felle, H.H., 2009. System potentials, a novel electrical long-distance apoplastic signal in plants, induced by wounding. Plant Physiol.

Zimmermann, M.R., Mithofer, A., 2013. Long-distance Systemic Signaling and Communication in Plants. Springer.