Computers and Electronics in Agriculture 145 (2018) 35–42

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

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

ARTICLE INFO

Keywords:
Plant stress
Optimum-Path Forest
Convolutional Neural Networks
Interval Arithmetic

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

* Corresponding author.
E-mail addresses: danilopereira@unoeste.br (D.R. Pereira), papa@fc.unesp.br (J.P. Papa), gustavosaraiva88@gmail.com (G.F.R. Saraiva), gumaia.gms@gmail.com (G.M. Souza).

https://doi.org/10.1016/j.compag.2017.12.024
Received 15 August 2017; Received in revised form 12 December 2017; Accepted 15 December 2017
0168-1699/

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.