

RESEARCH

Open Access



Early detection of dark-affected plant mechanical responses using enhanced electrical signals

Hongping Li^{1,2}, Nikou Fotouhi³, Fan Liu^{1*}, Hongchao Ji^{2*} and Qian Wu^{2*}

Abstract

Background Mechanical damage to plants triggers local and systemic electrical signals that are eventually decoded into plant defense responses. These responses are constantly affected by other environmental stimuli in nature, for instance, light fluctuation. In recent years, studies on decoding plant electrical signals powered by various machine learning models are increasing in a sense of early prediction or detection of different environmental stresses that threaten plant growth or crop yields. However, the main bottleneck is the low-throughput nature of plant electrical signals, making it challenging to obtain a substantial amount of training data. Consequently, training these models with small datasets often leads to unsatisfactory performance.

Results In the present work, we set out to decode wound-induced electrical signals (also termed slow wave potentials, SWPs) from plants that are deprived of light to different extents. Using non-invasive electrophysiology, we separately collected sets of local and distal SWPs from the treated plants. Then, we proposed a workflow based on few-shot learning to automatically identify SWPs. This workflow incorporates data preprocessing, feature extraction, data augmentation and classifier training. We established the integral and the first-order derivative as features for efficiently classifying SWPs. We then proposed an Adversarial Autoencoder (AAE) structure to augment the SWP samples. Combining them, the Random Forest classifier allowed remarkable classification accuracies of 0.99 for both local and systemic SWPs. In addition, in comparison to two other reported methods, our proposed AAE structure enabled better classification results using our tested features and classifiers.

Conclusions The results of this study establish new features for efficiently classifying wound-induced electrical signals, which allow for distinguishing dark-affected local and systemic plant wound responses. We also propose a new data augmentation structure to generate virtual plant electrical signals. The methods proposed in this study could be further applied to build models for crop plants using electrical signals as inputs, and also to process other small-scale signals.

Keywords Environmental stress, Plant electrical signal, Machine learning, Classification, Data augmentation

*Correspondence:

Fan Liu

liufan@tyut.edu.cn

Hongchao Ji

jihongchao@caas.cn

Qian Wu

wuqian01@caas.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

In nature, plants are constantly challenged by various environmental stimuli. Sub-optimal conditions like drought, salinity, rainfall, wind, light irradiation, as well as attacks by pathogens or insects cause severe damage to crop plants all over the world every year, which eventually leads to a great threat to food security [1, 2]. To deal with stressing factors, plants have evolved sophisticated strategies to defend themselves, in the meantime, maintaining their overall fitness under unfavored circumstances. Out of all these environmental stimuli, mechanical damages elicited by wounding of chewing insects not only break down tissue integrity and dampen crop growth, but also facilitate pathogen infections and greatly reduce the yield [3]. Therefore, an increasing number of studies were carried out to decipher how plants coordinate themselves in response to mechanical damages within the spot of injury as well as throughout the bodies, aiming at designing stress resilient plants.

As early as last decades, it was already observed that damages to plants triggered electrical signals that propagated to a distance away from the wounds [4]. These kinds of signals exist widespread across plant kingdoms, for instance, in the wounded tomato plants [5], the excised cucumber hypocotyls [6] and the crushed or chewed *Arabidopsis* leaves [7]. Interestingly, in these cases, the electrical signals typified the signature of the variation potentials (VPs), which has a steep depolarization phase and a slow repolarization phase. Besides VPs, damage also triggers action potentials (APs) and systemic potentials (SPs) in various plant species. For example, APs and SPs were detected on the stems and leaves of *Vicia faba* and *Hordeum vulgare*, respectively, upon herbivory [8]. However, the biological relevance of the electrical signals was largely unknown until the characterization of clade three *Arabidopsis* glutamate receptor-like (*GLRs*) genes [7]. Upon leaf wounding, *Arabidopsis* plants with impaired function of *GLR3.3*, *GLR3.6* and *GLR3.1* genes failed to propagate electrical signals to the distal intact leaves. Accordingly, defense responses mediated by the defense hormone jasmonate were largely reduced [7].

Wound-induced electrical signals are also termed as slow wave potentials (SWPs), in light of the different features in comparison with other electrical reactions. With the identification of more components in wound-induced electrical signaling, SWPs were further characterized and decoded using multiple parameters. Duration of SWPs was established as a major parameter in indicating the levels of defense activation in a couple of loss-of-function mutants. Unlike *glr* mutants, in *Arabidopsis* H^+ -ATPase *AHA1* mutants, the longer duration of SWPs corresponded to a stronger defense response [9].

In another *Arabidopsis* mutants defective in the mechanosensitive channels (MSLs), besides the SWP duration, SWP repolarization maxima was another key feature corresponded to the defense activation [10]. Interestingly, in the rice *glr3.4* loss-of-function mutants, peak amplitude of the SWP determined the defense levels in response to wounding [11]. Moreover, in two *irregular xylem* (*irx*) mutants *irx3* and *irx5* showing altered systemic defense responses, wound-induced SWPs displayed several different characteristics compared to wild-type plants, including the detected depolarization spikes prior to the main SWP signal, the slope of the principal depolarization and also the velocity of SWPs [12]. Therefore, the features of SWPs are tightly correlated with the defense activities in plants and are potentialized in indicating how plants react to wound stimuli under different circumstances. However, little information is available in the direction of decoding SWP features, which not only limits our understanding on how plants utilize SWPs to initiate systemic defense responses, but also restricts the potential application of SWPs in indicating the variable mechanical responses of plants when they are challenged by biotic or abiotic stresses.

In recent years, plant electrical signals are emerging as stimuli-specific phenotypes to differentiate plants from a wide range of environmental stimuli [13]. Moreover, as electrical signals are often produced rapidly upon stimuli, detection and classification of plant electrical signals also allow early diagnosis of the health status of plants before visible symptoms [14, 15]. In particular, machine learning-coupled strategies showed great power in analyzing and classifying plant electrical signals under various conditions. For example, Chatterjee et al. [16] extracted 11 statistical features from the electrical signals of tomato plants under different stimuli (NaCl , H_2SO_4 and O_3) and then classified them using several machine learning algorithms, i.e. Fisher's linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), naive Bayes classifier, and Mahalanobis classifier. The best accuracy of 73.67% was archived for the multiple classifications. More recently, an improved accuracy of above 90% was reached when four different curve fitting methods (Polynomial, Gaussian, Fourier and Exponential) were used to obtain the coefficients of the fitted model as features in combination with LDA and QDA as the classifiers [17]. In addition, Qin et al. [18] proposed a CNN structure to extract features and classify electrical signals of wheat plants under salt stress, reaching a classification accuracy of 92.3%. Following this, they further introduced a 1D-CNN-LSTM model to classify the electrical signals of wheat plants under serial concentrations of NaCl [19]. In another study, an interval algorithm was used to reduce the dimension of the input electrical signal in

combination with the SVM classifier [20]. More recently, to classify various nutrient deficiencies in tomato plants, Intrinsic Mode Functions (IMFs) of electrical signals from tomato plants cultured under various nutrient deficiencies (Ca, Fe, Mn and N) were extracted using Empirical Mode Decomposition (EMD). An average accuracy rate of 98.5% was reached using descriptive statistics and the bi-level measurements of IMF groups as features [21]. Despite the growing number of studies on decoding plant electrical signals, the successful utilization of plant electrical signals to predict how plants react to various stress conditions is still constrained by the low throughput of signals that can be used to generate reliable models.

In this study, we sought to determine if wound-induced local and systemic SWPs could be classified when plants are deprived of light to different extents. On the one hand, light qualities are crucial in determining plant growth [22]. On the other hand, geographical differences at different altitudes render plants with variable access to sunlight. Light fluctuation also occurs when smaller plants are shaded by the taller ones grown under high densities [23]. Whether plants grown under less sunlight react differently to mechanical stress compared to those receiving more sunlight remains uncharacterized. To address these questions, we simulated the variable light access in nature by treating the *Arabidopsis* plants under dark conditions to different extents, took advantage of the non-invasive electrophysiology to measure SWPs in the wounded and the distal connected leaves [24], extracted features from the traces and finally classified them using improved machine learning models. Notably, we also proposed a new network structure to augment the electrical signals, which showed a great performance in boosting the classification accuracy.

In summary, the main contributions of our present work are:

- 1) Our work established SWP as an effective parameter that can be utilized to differentiate plant local and systemic wound responses when they are deprived of light at different levels.
- 2) New features were established for efficiently classifying plant electrical signals. In our case, the electrical signals were collected from wounded plants that were pretreated under darkness to different extents.
- 3) An improved Adversarial Autoencoder (AAE) structure was proposed to augment the electrical signals.

The following sections of the manuscript were organized as below: (1) Materials and Methods. In this section, the pipeline for the data acquisition, preprocessing, feature extraction, data augmentation and classification models were described in detail. (2) Results. This section

presented the classification results of SWPs as well as performance evaluation of our proposed data augmentation structure. (3) Discussion. The results and the limitations of the present work were discussed in this Sect. (4) Conclusion. This section summarized the overall findings and potential future applications.

Materials and methods

Plant growth conditions

5-week-old *Arabidopsis thaliana* Colombia-0 (Col-0) plants were used as a model in this study. Individual plants were grown in the growth room at 21 °C under $150 \mu\text{E m}^{-2} \text{s}^{-1}$ light (10 h light, 14 h dark, 70% humidity) until 5-week-old with expanded rosettes. Then the plants were subjected to darkness treatment before measuring wound-induced electrical signals. Normal light condition: 10 h light/14 h night. Two extended darkness conditions were introduced: (1) short extended darkness (SED, 10 h light/18 h night); (2) long extended darkness (LED: 10 h light/40 h night). The above three conditions are summarized in Fig. 1A.

Experimental setup and data acquisition

As shown in Fig. 1A, measurements of wound-induced SWPs were carried out in a Faraday cage to avoid interference. Two Ag/AgCl recording electrodes (e1, e2) were placed on the petioles of leaf 8 (wounded) and leaf 13 (systemic), respectively. One drop of KCl/agar solution was added in between the electrode and the leaf petiole to maintain the connection. Another Ag/AgCl electrode was placed in soil as a reference. Upon crushing leaf 8 (wounded), the signals from both e1 and e2 were collected at a frequency of 100 Hz, amplified through a high-impedance signal amplifier (FD223a, WPI), and finally visualized using LabScribe3 software (iWorx System, Inc., Dover, NH). In this study, we collected SWPs from 20 plants under normal light conditions, 20 plants treated with SED and 15 plants treated with LED. For each plant, leaf 8 and leaf 13 SWPs were separately measured. In Table 1, the total number of the measured plants under different light conditions is listed.

Overall framework of data analysis

Figure 1B shows the overall workflow for analyzing the SWPs in this study. It contains the following procedures: (1) data preprocessing; (2) obtaining of characteristic equations; (3) data augmentation; (4) classification using different models.

Data preprocessing

Signal extraction

It was well established in the previous study [7] that when a local leaf (usually leaf 8) of *Arabidopsis* plants

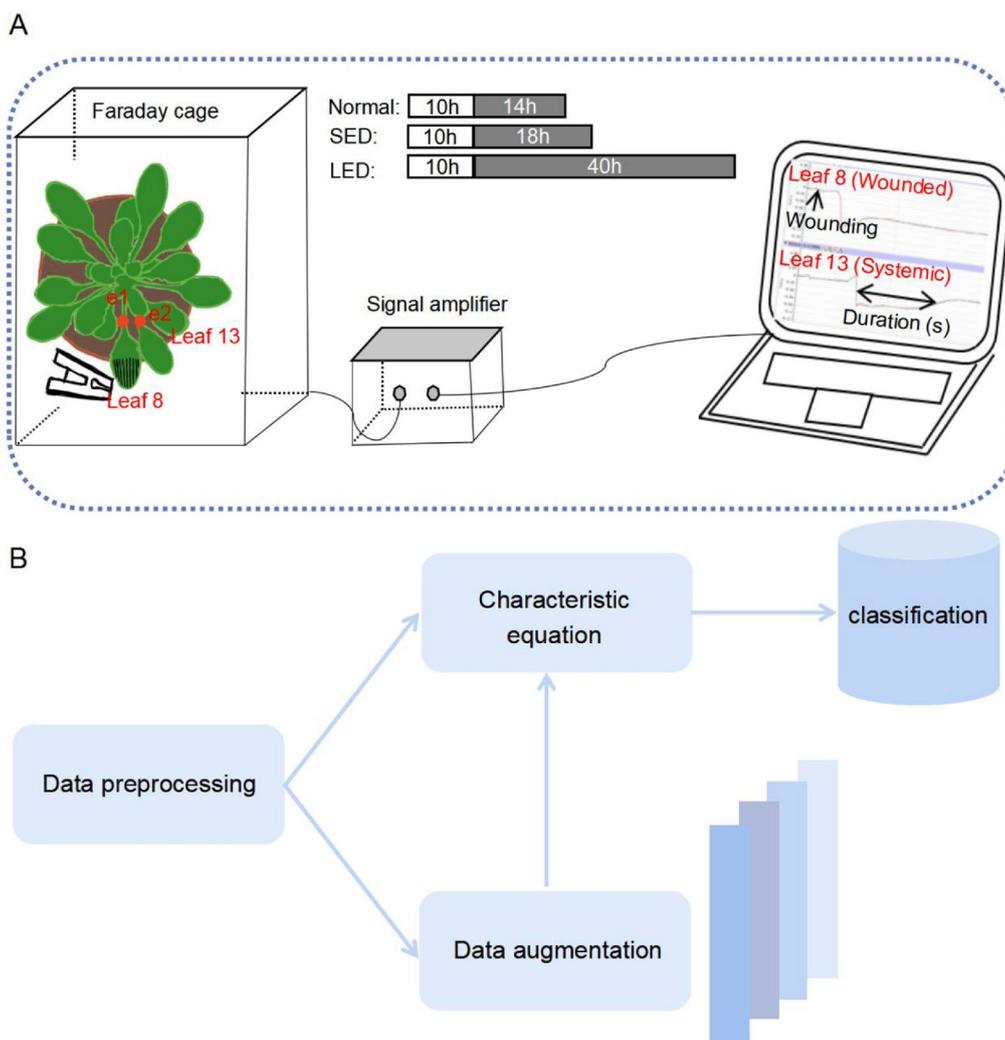


Fig. 1 Pipeline for the experimental setup and the data analysis. **A** Overview of the data acquisition system. To avoid external interference, 5-week-old *Arabidopsis* plants were treated under different light conditions before measurement in a Faraday cage. Light conditions: Normal: 10 h light/14 h night; SED: short extended darkness, 10 h light/18 h night; LED: long extended darkness, 10 h light/40 h night. Leaf 8 was wounded with forceps and the electrical signals were measured by two electrodes (e1, e2) that were placed on the petioles of leaf 8 and leaf 13, respectively. The signals were amplified through a signal amplifier and then converted for visualization on the laptop as shown. **B** Overall workflow for classifying the electrical signals. Following preprocessing, characteristics were extracted from the leaf 8 and leaf 13 electrical signals, respectively. Machine learning-based classification was performed using either the extracted characteristics from original samples or extended samples upon data augmentation. The colored bars next to “Data augmentation” represent the structural network for data augmentation

is wounded, SWPs could be generated in this leaf and then successfully transmitted to the distal part of plants. Leaf 13, which is distal to leaf 8, is normally measured to assess the systemic defense responses reflected by the SWP signals. Therefore, in our study, we also focused on the SWPs from this pair of leaves to estimate the local and systemic wound responses of the plants. As SWPs in the local leaf typically did not recover to the baseline during recording, and the time point for wounding the plants varies from sample to sample, to facilitate comparable features of SWPs under different conditions, we

extracted electrical signals from each sample following the same criteria. Specifically, electrical signals used for further analysis were extracted in between the “start” and “stop” window of the raw sample (as shown in Fig. 2A). Start: the time point when wounding was applied. Stop: the time point when the repolarization phase recovers to half of the maximum depolarization. For each sample, all the data points in this window were extracted and represented the useful signals. In the example trace in Fig. 2A, there are 12,800 data points after the signal extraction (between “start” and “stop”).

Table 1 Number of samples measured before treating with different light conditions

Leaves	Treatments	Number of samples
Leaf 8	Normal	20
	SED	20
	LED	15
Leaf 13	Normal	20
	SED	20
	LED	15

Electrical signals from both leaf 8 (wounded) and leaf 13 (systemic) were collected

Local mean compression

The process of local mean compression was performed on the extracted leaf 8 and leaf 13 signals for a specific purpose. The main objective was to reduce the amount of data while preserving the essential features of the original signals. Specifically, the signals were first divided into smaller segments. Then the mean values were calculated for each segment and were used to represent the entire segment. The compressed signal will have fewer data points compared to the original signal while the mean values of each segment still retain the essential features of the original signal. In this study, as shown in Fig. 2B, we set a fixed-size sliding window with a width of 100 data points. Following this, the original 12,800 data points were converted to a duration of 128 s (12,800/100=128). The same treatment was applied to all the samples.

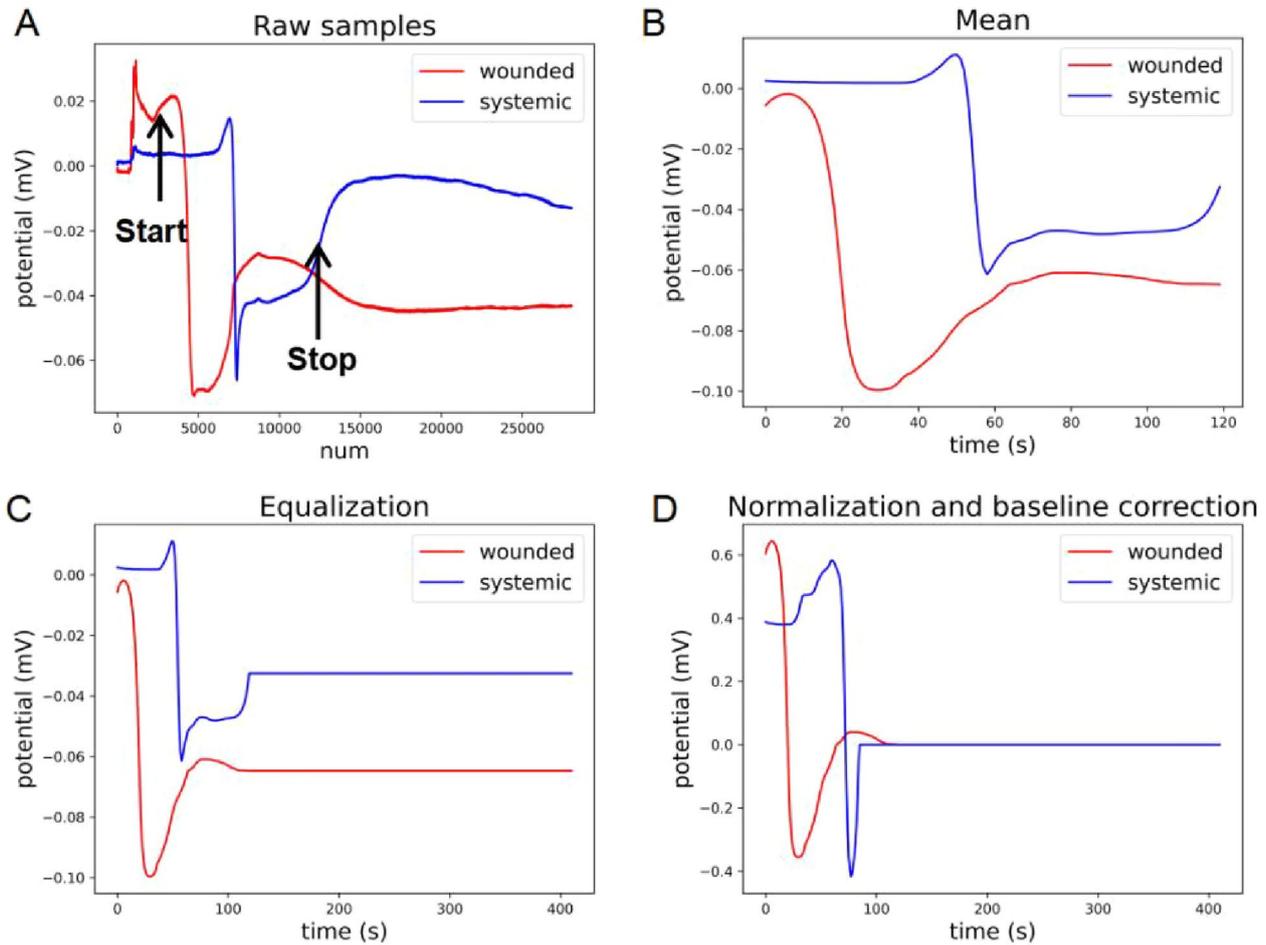


Fig. 2 Data preprocessing. **A** The electrical signal between the “start” and “stop” window was extracted from each raw recording as the useful signal. The horizontal axis shows the number (num) of the data points within the sample. **B** The values of every 100 data points in the window of **A** were averaged and the resulting mean trace is shown. The horizontal axis was converted to show the duration (s) within the window. **C** The mean trace in **B** was equalized with a length of 411 s by filling the sample with the value of the last data point. **D** Normalization and alignment of the equalized trace in **C**

Padding

Many machine learning models, especially those based on neural networks, require inputs of fixed dimensions. In this study, the duration of electrical signals can differ from one sample to another. Padding [25] was performed to fix the inputs into the same dimension. To do so, the samples were all filled with the values of their last data points to ensure an equal duration of 411 s, which corresponded to the sample with longest duration (Fig. 2C).

Normalization and baseline correction

We normalized each sample of all datasets to $[-1, 1]$ by Eq. (1), which allows variables with uncertainties to become more comparable. Normalization will make the model more data-sensitive which improves the accuracy of classification [26].

$$Y_i = \frac{X_{ij} - X_{i,min}}{X_{i,max} - X_{i,min}} \quad (i = 1, 2, \dots, n; \quad j = 1, 2, \dots, 411) \quad (1)$$

where Y_i is the i -th normalized sample, $X_{i,j}$ is the j -th datapoint of the i -th sample, $X_{i,min}$ is the minimum datapoint of the i -th sample, $X_{i,max}$ is the maximum datapoint of the i -th sample, i is the number of samples and j is the number of datapoints.

Next, to reduce the impact of the filled values on the following calculations, they were aligned to the amplitude at "0". The details of this section are shown in Fig. 2D.

Feature extraction

Time domain characteristics

To classify wound-induced Leaf 8 and Leaf 13 SWPs from plants that were treated with three light conditions (Normal, SED and LED), we started with extracting 12 time-domain features [27] from the SWP signals (Additional file 1: Table S1). Maximum and minimum respectively represent the maximum and minimum potential values in a sample. The mean is computed by dividing the overall potential values by the number of data points in a sample. The variance reflects the degree of fluctuation of the signal and is calculated by the distance between the potential value and the mean value at each time point. The standard deviation is the root of the variance. The skewness measures the skewed direction and degree of signal distribution. The kurtosis measures the sharpness of the signal distribution. The mean square is defined as the average value of each potential value squared. The root mean square is the root of the mean square value. The area represents the overall trend of a useful signal over the entire period. The declining slope represents the slope of the signal from the first maximum value to the minimum value. The rising slope represents the slope of the signal from the minimum value to the next maximum

value. The peak represents the difference between the maximum and the minimum peak values. The equations for calculating the above features are computed below:

$$\begin{aligned} \text{Maximum} &= \max(x_i) & \text{Minumum} &= \min(x_i) & (2) \\ \text{Mean} &:= \frac{1}{n} \sum_{i=1}^n x_i & \text{Variance} &: \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\ \text{Standard deviation} &: S = \sqrt{\sigma^2} & \text{Kurtosis} &: K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2]^2} \\ \text{Mean square root} &: x_{rms} & \text{Area} &= \sum_{i=1}^n | \frac{(f(x_i) + f(x_{i-1})) * (x_{i+1} - x_i)}{2} | \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n |x_i|^2} & \text{Decline slope} &= \frac{\text{maximum} - \text{minimum}}{-\Delta t} & \text{Risingslope} &= \frac{\text{maximum} - \text{minimum}}{\Delta t} \\ \text{Amplitude} &= \max(x_i) - \min(x_i) & \text{Skewness} &= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2]^{3/2}} \end{aligned}$$

Where x_i represents the i -th datapoint in a sample, $i = 1, 2, \dots, 411$.

Derivative and integral transformation

Plant electrical signals are kinds of time series, and are non-stationary and nonlinear. As a mathematical method, derivative transformation calculates the derivative of a signal and can be used to isolate specific features in a signal, such as edges, spikes, and changes in slopes [28]. The integral of a signal represents the accumulated change of the signal over time [29]. As another mathematical method, integral transformation calculates the integral of a signal. It is advantageous in signal smoothing, noise eliminating, and information extracting from the overall shape and trends of signals. Both derivative and integral transformation facilitate analysis and characterization of the accumulated change in a signal over time. The formulas for computing the first-order derivative and the integral are given in Eq. (3). The first-order derivative (deriv_1st) and the integral of a representative leaf 8 (wounded) SWP under normal condition are shown in Additional file 1: Fig. S1.

$$\begin{aligned} f'(x_i) &= \frac{1}{2} \left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i} + \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \right) s \\ &= \frac{(f(x_{i+1}) + f(x_i)) * (x_{i+1} - x_i)}{2} \quad (i = 1, 2, \dots, 411.) \end{aligned} \quad (3)$$

where x_i represents the i -th datapoint in a sample, $i = 1, 2, \dots, 411$.

Data augmentation

The core concept of few-shot learning lies in addressing the issue of limited samples, which often impacts classification accuracy due to experimental constraints [30]. To enhance the model's generalization ability and subsequently improve classification accuracy, one approach is to expand the number of electrical signal samples by generating augmented data within the training sets. In

our study, sample augmentation for plant electrical signals based on Adversarial Autoencoder (AAE) [31] was proposed. AAE was shown as a powerful tool in several aspects like semi-supervised classification, unsupervised clustering, image generation and data visualization [32]. However, its application in augmenting plant electrical signal samples has not been explored until now. This innovative use of AAE presents a significant opportunity to bolster few-shot learning capabilities in the context of plant electrical signal analysis.

Architecture for the data augmentation

The autoencoder (AE) [33] is composed of an encoder and a decoder. The encoder extracts key informative features from high-dimensional original data and maps them to a low-dimensional latent representation. Then the decoder maps the latent representation back to the original high-dimensional input. AAE adds a discriminator to the AE. After the original data is mapped to the latent space, the data conforming to a specific distribution is input. The discriminator is used to determine whether the mapped data space follows a specific data distribution pattern.

As shown in Fig. 3, the entire network structure is composed of the autoencoder (upper) and the discriminator (lower). The training process is divided into two stages: in the sample reconstruction stage, the original samples were input as one-dimensional data. The autoencoder was renewed by Stochastic Gradient Descent (SGD) [34] to minimize the reconstruction loss function. Here, we proposed to use the L1 loss function given by Eq. (5). Regularization [35] is accomplished using the aggregated posterior $q(z)$ to match any prior $p(z)$ in the regularization phase. The encoding function $q(z|x)$ on the autoencoder defines the hidden encoding $q(z)$, and the vector aggregated posterior distribution is computed as: $q(z) = \int_x q(z|x)p_d(x)dx$. $P(z)$ is the data that conforms to a certain distribution. Here we set it as the data that conforms to Gaussian distribution, which alternately allows updating the discriminant network and generator. We proposed to use Generative Adversarial Network [36] to complete the network training. The idea of using it is shown below:

$$\min_G \max_D E_{Z \sim P_{data}} [\log D(p(z)) + \log(1 - D(G(q(z))))] \tag{4}$$

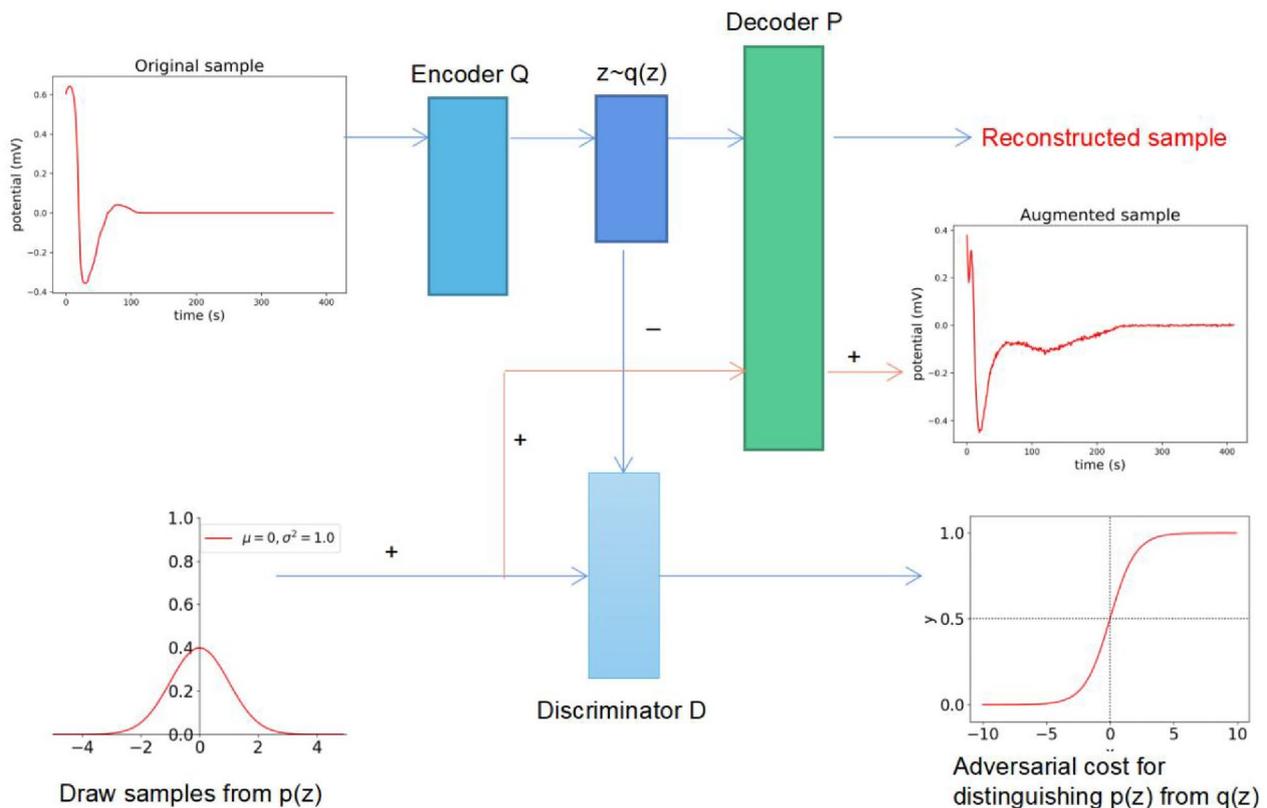


Fig. 3 Pipeline of the data augmentation. '+' represents the virtual vector, '-' represents the real vector. The pipeline consists of Encoder Q, Decoder P and Discriminator D. The curve in the lower left corner is a Gaussian distribution plot. The curve in the lower right corner shows a sigmoid function

where $p(z)$ is the real Gaussian data distribution. $G(q(z))$ is the virtual data distribution generated according to the input samples. $D(p(z))$ represents the probability of D judging the real data distribution. $D(G(q(z)))$ represents the probability of D judging the distribution of virtual data.

$$L1_loss : L1(\hat{y} - y) = \sum_{i=0}^n |\hat{y}^{(i)} - y^{(i)}| \quad (5)$$

where \hat{y} represents the predicted value, y represents the real value, and i represents the sample dimension.

The encoder, decoder, and the discriminator are all composed of fully-connected layers. The encoder contains 5 linear layers. Considering that each sample has 411 data points, the number of input neurons was set to 411, and the number of output neurons is 50. Corresponding to the encoder, the decoder also consists of 5 linear layers, with 50 input neurons and 411 output neurons. The discriminator consists of five linear layers with 50 input neurons and 1 neuron as a discriminative output. Detailed network architecture is shown in Additional file 1: Fig. S2. Dropout [37] was set to 0.2 for the linear layer to prevent overfitting. Relu activation function [38] was used to better mine sample features. fivefold cross-validation was used to split the dataset. The remaining parameters were set as follows: seed=10, epoch=200, reconstruction learning rate reg_lr=0.0001, and generation learning rate gen_lr=0.0005.

Evaluation criterion for the data augmentation

To evaluate the generative AAE, the mean values of each dimension from all samples were used as the template, and the similarity between the original sample and the augmented sample was compared using the following four indicators:

Euclidean distance (ED) measures the distance between two points in a multidimensional space.

$$ED = \frac{\sum_{h=1}^m \sqrt{\sum_{i=1}^n (x_i - y_i)^2}}{m} \quad (6)$$

Pearson correlation coefficients (PCC) measures the levels of correlation between x , y variables.

$$PCC = \frac{1}{m} \left(\sum_{h=1}^m \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \right)$$

$$\bar{x} = \sum_{i=1}^n x_i \bar{y} = \sum_{i=1}^n y_i \quad (7)$$

Cosine similarity (CS) indicates the similarity between two vectors by measuring the cosine of the angle.

$$\cos\theta = \frac{1}{m} \left(\sum_{h=1}^m \frac{\sum_{i=1}^n (x_i * y_i)}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \right) \quad (8)$$

where m is the number of samples, n is the number of dimensions, x_i represents the dimension of the original sample or the augmented sample, y_i represents the dimension of the template.

If the similarity between the original sample and the template is close to that between the augmented sample and the template, the augmented sample is considered reliable.

Classification

Classifier

Support Vector Machine (SVM) [39]. SVM is a supervised learning algorithm used for classification and regression tasks. It is a powerful and versatile classifier that finds the optimal hyperplane to separate data into different classes. In the SVM model, C is the regularization parameter, which is crucial in avoiding overfitting or underfitting issues. In our case, the value of C was set to 1 to balance the model's complexity and fitting capability. To deal with non-linearly separable data, we used the 'RBF' (Radial Basis Function) kernel as the kernel function. In the RBF kernel function, gamma is an important parameter that determines the impact of each sample on the decision boundary. We chose to automatically adjust the gamma value to determine the optimal decision boundary based on the characteristics of the data. By optimizing the values of C and the parameters for the kernel function, SVM is expected to achieve a good performance in classification problems, in our case, in classifying wound-induced SWPs from plants treated with different light conditions.

K-Nearest Neighbor (KNN) [40]. The KNN algorithm follows a specific classification process. First, it calculates the distances between each testing sample and all other samples in the dataset. Then the distances are sorted, and the k samples with the shortest distances to the testing sample are selected. Next, the categories of these k-nearest samples are determined. Finally, the testing sample is classified based on the category that is most frequently represented among the k-chosen neighbors. By considering multiple nearest neighbors, the KNN algorithm leverages the collective information from neighboring samples to make a more informative and reliable classification decision. This approach enhances the accuracy and robustness of the classification results. In our classification process, we tested different values of k to assess

the impact on classification accuracy. After evaluation, we determined to set the value of k to 5, which resulted in the best classification accuracy.

Random Forest [41]. Random Forest is a commonly used supervised learning algorithm composing of multiple decision trees. The "random" refers to two key components. First, when constructing each decision tree, a random subset of samples is selected from the original dataset using sampling with replacement. Second, a random subset of features is selected as input variables, which helps to increase the diversity among decision trees and reduce the risk of overfitting. By setting `random_state=0`, the results of running the Random Forest algorithm are reproducible. In other words, the same random seed will lead to the same random sampling and feature selection results, therefore ensuring consistency and comparability of the output results. The classification result of Random Forest is determined by a voting mechanism. After the samples are classified by each decision tree, the final classification is determined based on the voting results of all decision trees. This ensemble learning approach helps to improve the accuracy and robustness of the Random Forest algorithm.

Muti-Layer Perception (MLP) [42]. Multi-layer perceptron (MLP) is a classical neural network model. It usually consists of an input layer, a hidden layer and an output layer. The different layers are fully connected. In our study, we incorporated this deep learning model, in addition to the other machine learning models described above, to evaluate their performance in SWP classification. The number of hidden layer neurons is set to 10. Relu is used as the activation function between layers. The MLP model is ended up with Softmax layer, which outputs the confidence of each category. `Learning_rate=0.01`.

Evaluation criterion for the classifiers

To visualize the classification results, we introduced confusion matrix with the horizontal axis representing the true label and the vertical axis representing the predicted label.

The confusion matrix contains four pieces of information: True Positive (TP), which means the actual positive samples are predicted to be positive; True Negative (TN), which means the actual negative samples are predicted to be negative; False Positive (FP), which means the actual samples are negative but predicted to be positive; False Negative (FN), which means the actual samples are positive but predicted to be negative.

The following four indicators were used to evaluate the classification effect:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} & \text{Recall} &= \frac{TP}{TP+FN} \\ \text{F1 score} &= 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} & \text{ACC} &= \frac{TP+TN}{TP+TN+FP+FN} \end{aligned} \quad (9)$$

Precision represents the proportion of samples that are actually positive among the samples that are predicted to be positive. The recall is determined as the ratio between the number of positive samples that are correctly classified as positive to the total number of positive samples. The recall measures the model's ability to detect positive samples. The higher the recall, the more positive samples detected. F1 score is calculated as a weighted average of precision and recall. Greater F1 score reflects robust model. Given that our task is a multi-classification, we used macro-averaging method that first calculated the precision, recall, and F1 score for each positive example and then averaged all the results. Accuracy (ACC) represents the proportion of the correctly classified samples out of the total number of samples.

Results

Leaf 8 and leaf 13 SWPs separately evaluate the local and systemic wound responses for a plant. In this study, plants were pre-treated under either normal or extended darkness conditions (SED and LED) before leaf 8 and leaf 13 SWPs were measured upon wounding leaf 8. Then leaf 8 and leaf 13 signals were separately classified using various models. In this section, we presented the triple classification results for leaf 8 (wounded) and leaf 13 (systemic) SWPs, respectively. To improve the classification accuracy, we introduced a new data augmentation network to extend the sample size for the electrical signals. The performance of this method was evaluated in comparison to another two widely used methods.

Triple classification of wound-elicited SWPs under extended darkness using original data

After performing the data preprocessing procedures shown in **Materials and Methods "Data preprocessing"** on leaf 8 (wounded) and leaf 13 (systemic) SWP samples respectively, we extracted a group of 12 time-domain characteristics, the first-order derivative and the integral features from the traces. The features were then subjected to classification using SVM, KNN and Random Forest, and MLP classifiers, respectively. Triple classification results for leaf 8 (wounded) and leaf 13 (systemic) SWPs are shown in Additional file 1: Tables S2, S3, respectively. Table 2 shows the best classification results. In the case of leaf 8 (wounded) SWPs, an accuracy of 0.89 was reached using the integral as the feature and SVM as the classifier. For classifying leaf 13 (systemic) SWPs, the SVM classifier showed the best performance and generated a classification accuracy of 0.8 when using the integral as

Table 2 The best classification results for leaf 8 (wounded) and leaf 13 (systemic) SWPs

Leaves	Features	Classifiers	Accuracy	Recall	Precision	F1 score
Leaf 8	Integral	SVM	0.89	0.89	0.90	0.89
Leaf 13	Integral	SVM	0.80	0.80	0.80	0.80

the feature. Details of the classification are shown in the confusion matrices in Additional file 1: Fig. S3. Six samples for leaf 8 SWPs were misclassified (Additional file 1: Fig. S3A) while eleven samples were misclassified for leaf 13 SWPs (Additional file 1: Fig. S3B).

Evaluation of the augmented samples

To improve the classification accuracy for the misclassified leaf 8 and leaf 13 samples as shown in Additional file 1: Fig. S3, we proposed to use a data augmentation strategy to generate augmented leaf 8 and leaf 13 samples for training the models. After preprocessing the samples as shown in **Materials and Methods** “Data preprocessing”, the samples were augmented using AAE as described in the methods. Four evaluation indicators described in **Materials and Methods** “Evaluation criterion for the data augmentation” were used to evaluate the similarities between the original and the augmented samples. The mean values of each dimension of the original sample under Normal, SED and LED conditions were calculated and used as templates. For evaluation, we used the ED, CS and PCC parameters to compare the correlation among the templates, the original samples and the augmented samples. As shown in Table 3, for both leaf 8 and leaf 13 SWPs, the Euclidean distance between the augmented sample and the template, and that between the original sample and the template, is very close. In addition, the PCC and CS values are close to 0.8, indicating that the templates are strongly correlated to all the samples. Together, our results demonstrated that the augmented samples are very similar to the original samples, supporting the effectiveness of our proposed data augmentation model.

Triple classification of wound-elicited SWPs under extended darkness upon data augmentation

Based on the above model, 50 augmented samples were respectively generated for leaf 8 and leaf 13 SWPs under normal, SED and LED conditions. Similarly, we extracted the features and reclassified the leaf 8 and leaf 13 SWPs using the above four classifiers with the same settings as for the original data. For all of the datasets, the leave-one-out method was used to divide the training and testing sets. Upon data augmentation, the classification results were generally improved in all the combinations of different features and classifiers as reflected by the four parameters (Additional file 1: Table S4 and Table S5). We noted that upon adding the augmented samples, Random Forest showed the best performance in classifying either leaf 8 or leaf 13 SWPs using their deriv_1st as features (Table 4).

Specifically, the classification accuracy for leaf 8 (wounded) SWPs was increased to 0.99 when using the deriv_1st as the feature and Random Forest as the classifier (Table 4). Moreover, the scores for other evaluation indicators (Recall, Precision, F1-score) were also greatly boosted (Table 4). Detailed classification results are shown in the confusion matrix in Additional file 1: Fig. S4. Originally, 13 samples were misclassified (Additional file 1: Fig. S4A), whereas, upon data augmentation, only two samples were incorrectly classified under SED and LED conditions (Additional file 1: Fig. S4B).

For leaf 13 (systemic) SWPs, as shown in Table 4, the best classification accuracy was also greatly elevated to 0.99. Similar to leaf 8 SWPs classification, Random Forest again showed the best performance when using the deriv_1st as the feature. Detailed information for the

Table 3 Quality evaluation of the augmented samples for leaf 8 (wounded) and leaf 13 (systemic) SWPs

Leaves	Indicators	Augmented Normal_template	Original Normal_template	Augmented SED_template	Original SED_template	Augmented LED_template	Original LED_template
Leaf 8	ED	11.51	14.21	14.99	14.39	20.06	13.09
	CS	0.83	0.75	0.83	0.75	0.94	0.88
	PCC	0.82	0.74	0.81	0.72	0.93	0.87
Leaf 13	ED	17.76	17.81	23.16	24.82	23.82	16.43
	CS	0.84	0.78	0.94	0.82	0.83	0.79
	PCC	0.84	0.77	0.89	0.78	0.79	0.79

Table 4 The best classification results for leaf 8 (wounded) and leaf 13 (systemic) SWPs upon data augmentation

Leaves	Features	Classifiers	Accuracy	Recall	Precision	F1 score
Leaf 8	Deriv_1st	Random forest	0.76/0.99	0.76/0.99	0.77/0.99	0.76/0.99
Leaf 13	Deriv_1st	Random forest	0.71/0.99	0.69/0.99	0.75/0.99	0.70/0.99

The results before augmentation are shown before slash

classification was shown in Additional file 1: Fig. S4. Before data augmentation, seven samples under normal condition, two samples under SED condition and seven samples under LED condition were misclassified (Additional file 1: Fig. S4C) whereas after data augmentation, only one sample under normal condition and one sample under SED condition were misclassified (Additional file 1: Fig. S4D).

Performance comparison of the different augmentation methods

Finally, to evaluate if our proposed augmentation method was superior to other reported models, we compared the performance of AAE to the Generative Adversarial Network (GAN) [36] and the Variational Autoencoder (VAE) models [43] in augmenting SWPs. The GAN method was proposed to enrich the electrical signals from wheat seedlings under salt tolerance. By introducing a certain amount of noise into the model, a regularization effect was produced and the performance of the model was improved. The classification accuracy was increased from 80.77 to 92.31% [18]. Variational Auto-Encoder (VAE) is a generative algorithm which is highly versatile and its function is not affected by the format of data [44]. However, it has not been applied in generating plant electrical signals before.

To facilitate comparison, the same signal preprocessing, feature extraction and classifying procedures were used before deploying the GAN and VAE methods. The dataset was divided into training set and testing set by

fivefold cross-validation. All experiments were done by CPU through python3.7 under win10 operating system. In the case of using GAN to generate SWPs, we reused the same settings reported in [18], to compare its performance in augmenting and classifying our electrical signals with our proposed AAE. The number of input and output neurons was set to 411 based on the model proposed in [18]. In the case of using VAE to generate SWPs, as it was not reported in augmenting electrical signals before, the parameters were set the same as those for AAE. Specifically, the encoder consists of five linear layers with each layer containing 411, 200, 100, 100, and 50 neurons, respectively. The decoder consists of five linear layers with each layer containing 50, 100, 100, 200, and 411 neurons, respectively.

As shown in Table 5, in comparison to GAN and VAE, our proposed AAE method showed the best performance in combination with the tested classifiers to classify leaf 8 and leaf 13 SWPs. Altogether, the results strongly suggest that our method is an effective tool for augmenting SWPs in the testing sets, which further allowed us to improve the performance of the classifiers.

Discussion

Plants can generate electrical signals when subjected to various environmental stimuli or grown at different developmental stages [45, 46]. Particularly, in recent years, wound-induced electrical signals, widely known as slow wave potentials (SWPs), are receiving great

Table 5 Comparison of the classification accuracies (ACC (%)) for leaf 8 (wounded) and leaf 13 (systemic) SWPs using three different augmentation methods

Leaves	Time-domain (ACC)			Deriv_1st (ACC)			Integral (ACC)		
	AAE	VAE	GAN	AAE	VAE	GAN	AAE	VAE	GAN
Leaf 8	82.44	79.51	84.39	95.61	89.75	67.32	95.61	89.26	83.90
	80.98	82.43	81.95	90.24	87.31	69.93	84.88	85.85	86.83
	80.49	81.95	81.95	99.02	90.24	80.49	91.22	88.78	84.88
Leaf 13	87.80	64.88	76.10	93.66	89.27	76.59	97.07	91.71	84.39
	88.78	77.56	76.59	86.83	88.78	71.22	91.21	91.21	84.39
	88.78	84.39	80.49	99.02	90.73	70.73	96.10	93.65	83.90

The best results are underlined and highlighted in bold

attention as they were found to tightly connect with plant defense activities in response to herbivory or mechanical damage [45]. Therefore, decoding SWPs under different conditions will potentially enable differentiating whether plants could react differently to mechanical stress when grown under a varying environment. In this work, we measured local (leaf 8) and systemic (leaf 13) wound-induced SWPs from *Arabidopsis* plants that were kept in darkness to different extents (normal, short-extended darkness and long-extended darkness) and classified the signals in combination with machine learning models. Extended darkness has a great impact on plant growth [23]. Moreover, in a recent study by Fotouhi et al. [47], it was shown that extended darkness led to a failure in generating SWPs in an *aca10 aca12* mutant, which subsequently enhanced the susceptibility to herbivory. By extracting different features from SWPs and applying different machine learning models, we managed to classify leaf 8 and leaf 13 SWPs separately from plants treated under darkness to different extents (SED and LED) compared to normal light condition, with accuracies above 0.8. To further improve the performance, we introduced an AAE method to augment our obtained SWP samples. The classification accuracies were greatly increased up to more than 0.99 for both leaf 8 and leaf 13 SWPs. Our work established SWPs as an effective proxy that can be applied to differentiate plant local and systemic wound responses when they are deprived of light to different extents. This is actually in line with the genetic evidence that extended darkness reshapes plant wound responses, for example, in the *aca* mutants [47]. Therefore, our proposed methods for classifying SWPs under different light conditions could be further applied in the early detection of various wound responses in different genetic mutants when they are subjected to different environmental stimuli.

From the perspective of methodology, we introduced to use the integral and the *deriv_1st* as new features for plant electrical signals. Compared to the canonical duration and amplitude features used to quantify electrical signals [45], our established parameters also effectively allowed distinguishing plant wound response-SWPs affected by different levels of darkness. However, accurate and reliable classification of plant electrical signals under different conditions is still counteracted by the limited number of samples. Although data augmentation has been implemented as a useful tool to improve the performance of models [48, 49], its application in augmenting plant electrical signals was poorly investigated. In our work, we proposed to use AAE method to generate virtual SWP samples for training the machine learning models. This method has not been reported in augmenting plant electrical signals before. We compared

the classification results using data obtained by three different data extension methods, out of which, our proposed AAE model obtained the best result for classifying SWPs. This is probably because the AAE idea is to establish a reversible mapping from the real text distribution to an unknown distribution. This unknown distribution could be set following a Gaussian distribution, avoiding problems like text discreteness and non-differentiability. Given the powerful performance of the AAE model, it could also be helpful for future research to convert other small-scale datasets to large-scale that are needed for improving the performance of machine learning models.

While AAE serves as an effective and commonly employed data augmentation strategy for time series, which is in the context of plant electrical signals, it has several constraints which may cause limitation of this work. First, AAE, like any generative model, can be prone to overfitting especially when the training data is limited. It makes the model under the risk of missing to generalizing well to new or unseen data. Second, interpreting the latent space of AAE can be challenging. It might be unclear what specific features are encoded in the latent variables.

Conclusions

To summarize, using non-invasive electrophysiology, we generated three sets of wound-induced local and systemic electrical signals from plants treated with three light conditions (Normal, SED and LED). To develop new methods that could classify electrical signals more efficiently and precisely, we extracted different features from our generated signals and classified them using different machine learning models. We established the integral and *deriv_1st* as useful features for leaf 8 and leaf 13 SWPs classification under the three light conditions. In cooperation with our new proposed AAE method to augment the limited number of SWP samples, the Random Forest classifier showed the best performance in distinguishing both leaf 8 and leaf 13 SWPs with an accuracy of 0.99. In nature, plants are under constant mechanical stresses and their access to light is affected by various factors. Therefore, our model could also be a useful tool for crop plants in predicting their wound responses, even more broadly, other stress responses reflected by plant electrical signals.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-024-01169-4>.

Additional file 1: Fig S1. The first-order derivative (*deriv_1st*) (A) and the integral (B) features of an example trace from a preprocessed leaf 8 (wounded) SWP. **Fig S2.** Network structures of the Q Encoder, P Decoder

and D Discriminator. The circles filled with blue color represent neurons.

Fig S3. Confusion matrices for prediction results using SVM classifier. **Fig S4.** Confusion matrices for prediction results using Random Forest classifier. **Table S1.** Time-domain features used in this study. **Table S2.** Triple classification results for leaf 8 (wounded) SWPs. **Table S3.** Triple classification results for leaf 13 (systemic) SWPs. **Table S4.** Classification results for leaf 8 (wounded) SWPs upon data augmentation. **Table S5.** Classification results for leaf 13 (systemic) SWPs upon data augmentation.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (32200243) to Qian Wu.

Author contributions

QW designed the study. NF performed the experiments of collecting electrical signals. HPL carried out the computational section and prepared the figures and tables. QW, HPL and HCJ wrote the manuscript. All authors reviewed the manuscript.

Funding

This work is supported by the National Natural Science Foundation of China (32200243) to Qian Wu.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Computer Science and Technology (College of Data Science), Taiyuan University of Technology, Jinzhong 030600, Shanxi, China. ²Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, Guangdong, China. ³Desai Sethi Urology Institute, Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL 33136, USA.

Received: 2 August 2023 Accepted: 28 February 2024

Published online: 26 March 2024

References

- Dhankher OP, Foyer CH. Climate resilient crops for improving global food security and safety. *Plant Cell Environ*. 2018;41(5):877–84. <https://doi.org/10.1111/pce.13207>.
- Rivero RM, Mittler R, Blumwald E, Zandalinas SI. Developing climate-resilient crops: improving plant tolerance to stress combination. *Plant J*. 2022;109(2):373–89. <https://doi.org/10.1111/tpj.15483>.
- Erb M, Reymond P. Molecular interactions between plants and insect herbivores. *Annu Rev Plant Biol*. 2019;70:527–57.
- Boari F, Malone M. Wound-induced hydraulic signals: survey of occurrence in a range of species. *J Exp Bot*. 1993;44(4):741–6. <https://doi.org/10.1093/jxb/44.4.741>.
- Rhodes JD, Thain JF, Wildon DC. The pathway for systemic electrical signal conduction in the wounded tomato plant. *Planta*. 1996;200(1):50–7.
- Stahlberg R, Cosgrove DJ. Comparison of electric and growth responses to excision in cucumber and pea seedlings. II. Long-distance effects are caused by the release of xylem pressure. *Plant Cell Environ*. 1995;18(1):33–41.
- Mousavi SA, Chauvin A, Pascaud F, Kellenberger S, Farmer EE. *GLUTAMATE RECEPTOR-LIKE* genes mediate leaf-to-leaf wound signalling. *Nature*. 2013;500(7463):422–6.
- Zimmermann MR, Mithöfer A, Will T, Felle HH, Furch AC. Herbivore-triggered electrophysiological reactions: candidates for systemic signals in higher plants and the challenge of their identification. *Plant Physiol*. 2016;170(4):2407–19. <https://doi.org/10.1104/pp.15.01736>.
- Kumari A, Chételat A, Nguyen CT, Farmer EE. *Arabidopsis* H⁺-ATPase AHA1 controls slow wave potential duration and wound-response jasmonate pathway activation. *PNAS*. 2019;116(40):20226–31. <https://doi.org/10.1073/pnas.1907379116>.
- Moe-Lange J, Gappel NM, Machado M, Wudick MM, Sies CS, Schott-Verdugo SN, Bonus M, Mishra S, Hartwig T, Bezruczyk M, Basu D. Interdependence of a mechanosensitive anion channel and glutamate receptors in distal wound signaling. *Sci Adv*. 2021;7(37):4298.
- Yu B, Wu Q, Li X, Zeng R, Min Q, Huang J. *GLUTAMATE RECEPTOR*-like gene *OsGLR3.4* is required for plant growth and systemic wound signaling in rice (*Oryza sativa*). *New Phytol*. 2022;233(3):1238–56. <https://doi.org/10.1111/nph.17859>.
- Kurenda A, Nguyen CT, Chételat A, Stolz S, Farmer EE. Insect-damaged *Arabidopsis* moves like wounded *Mimosa pudica*. *PNAS*. 2019;116(51):26066–71. <https://doi.org/10.1073/pnas.1912386116>.
- Li JH, Fan LF, Zhao DJ, Zhou Q, Yao JP, Wang ZY, Huang L. Plant electrical signals: a multidisciplinary challenge. *J Plant Physiol*. 2021;261:153418.
- Tran D, Camps C. Early diagnosis of iron deficiency in commercial tomato crop using electrical signals. *Front Sustain Food S*. 2021;5:631529. <https://doi.org/10.1016/j.jplph.2021.153418>.
- Najdenovska E, Dutoit F, Tran D, Plummer C, Wallbridge N, Camps C, Raileanu LE. Classification of plant electrophysiology signals for detection of spider mites infestation in tomatoes. *Appl Sci*. 2021;11(4):1414. <https://doi.org/10.3390/app11041414>.
- Chatterjee SK, Das S, Maharatna K, Masi E, Santopolo L, Mancuso S, Vitaletti A. Exploring strategies for classification of external stimuli using statistical features of the plant electrical response. *J R Soc Interface*. 2015;12(104):20141225. <https://doi.org/10.1098/rsif.2014.1225>.
- Chatterjee SK, Malik O, Gupta S. Chemical sensing employing plant electrical signal response-classification of stimuli using curve fitting coefficients as features. *Biosensors*. 2018;8(3):83. <https://doi.org/10.3390/bios803083>.
- Qin XH, Wang ZY, Yao JP, Zhou Q, Zhao PF, Wang ZY, Huang L. Using a one-dimensional convolutional neural network with a conditional generative adversarial network to classify plant electrical signals. *Comput Electron Agr*. 2020;174:105464. <https://doi.org/10.1016/j.compag.2020.105464>.
- Yao JP, Wang ZY, de Oliveira RF, Wang ZY, Huang L. A deep learning method for the long-term prediction of plant electrical signals under salt stress to identify salt tolerance. *Comput Electron Agr*. 2021;190:106435. <https://doi.org/10.1016/j.compag.2021.106435>.
- Pereira DR, Papa JP, Saraiva GF, Souza GM. Automatic classification of plant electrophysiological responses to environmental stimuli using machine learning and interval arithmetic. *Comput Electron Agr*. 2018;145:35–42. <https://doi.org/10.1016/j.compag.2017.12.024>.
- Sai K, Sood N, Saini I. Classification of various nutrient deficiencies in tomato plants through electrophysiological signal decomposition and sample space reduction. *Plant Physiol Bioch*. 2022;186:266–78. <https://doi.org/10.1016/j.plaphy.2022.07.022>.
- Kami C, Lorrain S, Hornitschek P, Fankhauser C. Light-regulated plant growth and development. *Curr Top Dev Biol*. 2010;91:29–66. [https://doi.org/10.1016/S0070-2153\(10\)91002-8](https://doi.org/10.1016/S0070-2153(10)91002-8).
- Pierik R, Ballaré CL. Control of plant growth and defense by photoreceptors: from mechanisms to opportunities in agriculture. *Mol Plant*. 2021;14(1):61–76. <https://doi.org/10.1016/j.molp.2020.11.021>.
- Mousavi SA, Nguyen CT, Farmer EE, Kellenberger S. Measuring surface potential changes on leaves. *Nat Protoc*. 2014;9(8):1997–2004.
- Dwarampudi M, Reddy NV. Effects of padding on LSTMs and CNNs. *arXiv Preprint*. 2019. <https://doi.org/10.48550/arXiv.1903.07288>.

26. Singh D, Singh B. Investigating the impact of data normalization on classification performance. *Appl Soft Comput*. 2020;97:105524. <https://doi.org/10.1016/j.asoc.2019.105524>.
27. Tkach D, Huang H, Kuiken TA. Study of stability of time-domain features for electromyographic pattern recognition. *J Neuroeng Rehabil*. 2010;7(1):1–3.
28. Liu Y, Lin J. A general-purpose signal processing algorithm for biological profiles using only first-order derivative information. *BMC Bioinform*. 2019;20(1):1–7.
29. Huk AC, Shadlen MN. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J Neurosci*. 2005;25(45):10420–36. <https://doi.org/10.1523/JNEUROSCI.4684-04.2005>.
30. Bandara K, Hewamalage H, Liu YH, Kang Y, Bergmeir C. Improving the accuracy of global forecasting models using time series data augmentation. *Pattern Recognit*. 2021;120:108148. <https://doi.org/10.1016/j.patcog.2021.108148>.
31. Makhzani A, Shlens J, Jaitly N, Goodfellow I, Frey B. Adversarial autoencoders. *arXiv Preprint*. 2015. <https://doi.org/10.48550/arXiv.1511.05644>.
32. Ge P, Ren CX, Dai DQ, Feng J, Yan S. Dual adversarial autoencoders for clustering. *IEEE T Neur Net Lear*. 2019;31(4):1417–24.
33. Bank D, Koenigstein N, Giryres R. Autoencoders. *arXiv Preprint*. 2020. <https://doi.org/10.48550/arXiv.2003.05991>.
34. Robbins H, Monro S. A stochastic approximation method. *Ann Math Stat*. 1951. <https://doi.org/10.1214/aoms/1177729586>.
35. Poggio T, Torre V, Koch C. Computational vision and regularization theory. *Read Comput Vision*. 1987. <https://doi.org/10.1016/B978-0-08-051581-6.50061-1>.
36. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. *Commun Acm*. 2020;63(11):139–44.
37. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929–58.
38. Krizhevsky A, Hinton G. Convolutional deep belief networks on cifar-10. Unpublished manuscript. 2010; 40(7): 1–9.
39. Vapnik VN, Lerner AY. Recognition of patterns with help of generalized portraits. *Avtomat i Telemekh*. 1963;24(6):774–80.
40. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE T Inform Theory*. 1967;13(1):21–7.
41. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
42. Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N. Multilayer perceptron and neural networks. *WSEAS Trans Circuits Syst*. 2009;8(7):579–88.
43. Kingma DP, Welling M. Auto-encoding variational bayes. *arXiv Preprint*. 2013. <https://doi.org/10.48550/arXiv.1312.6114>.
44. Sun J, Wang X, Xiong N, Shao J. Learning sparse representation with variational auto-encoder for anomaly detection. *IEEE Access*. 2018;6:33353–61.
45. Farmer EE, Gao YQ, Lenzoni G, Wolfender JL, Wu Q. Wound-and mechnostimulated electrical signals control hormone responses. *New Phytol*. 2020;227(4):1037–50. <https://doi.org/10.1111/nph.16646>.
46. Damineli DSC, Portes MT, Feijó JA. Electrifying rhythms in plant cells. *Curr Opin Cell Biol*. 2022;77:102113. <https://doi.org/10.1016/j.cub.2022.102113>.
47. Fotouhi N, Fischer-Stettler M, Lenzoni G, Stolz S, Glauser G, Zeeman SC, Farmer EE. ACA pumps maintain leaf excitability during herbivore onslaught. *Curr Biol*. 2022;32(11):2517–28. <https://doi.org/10.1016/j.cub.2022.03.059>.
48. Oussidi A, Elhassouny A. Deep generative models: Survey. *ISCV*. 2018; 1–8.
49. Iglesias G, Talavera E, González-Prieto Á, Mozo A, Gómez-Canaval S. Data augmentation techniques in time series domain: a survey and taxonomy. *Neural Comput Appl*. 2023;35(14):10123–45.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.