Stimuli Classification with Electrophysiology and Impedance of Natural Plants: Comparing Discriminant Analysis and Deep Learning Methods

Removed for double-anonymous peer review process.

May 2022

Abstract. The physiology of living organisms, such as natural plants, is complex and particularly difficult to understand on a macroscopic, organsim-holistic level. Among the many options to study plant physiology, electrophysiology and tissue electrical impedance are arguably simple measurement techniques to gather plant-level information. Despite the many possible uses, our research is exclusively driven by the idea of phytosensing, that is, using living plants as sensors. As ready-to-use plant-level physiological models are not available, we consider the plant as a blackbox and apply statistics and machine learning to automatically interpret measured signals. In simple plant experiments, we expose *Zamioculcas zamiifolia* and *Solanum lycopersicum* (tomato) to four different stimuli: wind, heat, red and blue light. We measure electrophysiological and tissue impedance signals. Given these signals, we evaluate a large variety of methods from statistical discriminant analysis and from deep learning for the classification problem of determining the correct stimulus to which the plant was exposed. We identify a set of methods that successfully classify stimuli with good accuracy without a clear winner. The statistical approach is competitive, partially depending on data availability for the machine learning approach. Our extensive results show the feasibility of the blackbox approach and can be used in future research to select appropriate classifier techniques for a given use case. In our own future research, we will exploit these methods to drive a phytosensing approach for air pollution monitoring in urban areas.

Keywords: electrophysiology, tissue electrical impedance, phytosensing, machine learning, classification, discriminant analysis, artifical neural networks, time series classification

1. Introduction

One key to understand living organisms and life as such is to analyze their physiology. A good understanding of plant physiology, for example, can help to grow crop plants more efficiently and healthier [1]. Our main motivation, however, is driven by an engineering perspective as we ultimately want to use measurements of plant physiology to implement a so-called phytosensing approach [2, 3]. Phytosensing is the idea of using natural living plants as sensors (more details below). In plant science one would hope for fundamental insights once plant physiology is fully understood. As in all natural

sciences, an essential step of the experimental approach is measurement. There are many options to measure a plant's physiology. Here, we focus on electrophysiology and tissue electrical impedance for a number of reasons. Plants generate different types of extracellular electrical responses in reaction to environmental stresses [4, 5].

In electrophysiology a differential electrical potential of the plant is measured [6]. Changes in a plant's electrical potential can be detected for different stimuli, such as heat, light, humidity, CO_2-level, herbivory, and salt stress [2, 6]. The measured electric potential in the plant is subject to an uneven distribution of ions that are also actively and passively transported through the plant (see Sec. 2 for a few more details). This method can be intrusive as electrodes are inserted into the plant's tissue (as done here). Less intrusive techniques are also established by recording surface potentials of plant leaves [7].

For tissue electrical impedance a small alternating current is applied to probe, in our case, the plant tissue's resistance and reactance characteristics [8]. An advantages of both electrophysiology and impedance is their relatively simple application because only two electrodes and simple electronics are required to measure and process the signal. The measured electric potential is a macroscopic feature of the plant across several cubic centimeters of tissue. This macroscopic property complicates the interpretation of the signal. While previous experiments have shown that there are clearly measurable effects, such as a reaction of the plant to light, the scientific interpretation of the obtained data in terms of the underlying molecular mechanisms is still challenging [9]. Hence, the interpretation of data obtained by plant electrophysiology is complex and the respective research is still in its infancy. Here, we assume the plant to be a blackbox and we automatically interpret the electrophysiological or impedance signal using methods of statistics and machine learning.

Essential for our approach is to gather training data in controlled plant experiments that can explicitly be labeled by a respectively applied stimulus (e.g., heat or light color). The required amount of data and plant experiments is generally unknown but especially for methods of deep learning more data usually means more accuracy [10]. This poses challenges for plant science due to high costs of plant experiments (cf. [11]). Although we gather data from many plant experiments, at least a few of our tested methods may still suffer from lack of data (also see discussion in Sec. 4.3).

In previous work, an energy-efficient sensor node called *PhytoNode* was designed for plant electrophysiology (not used here). It can run classification models on board and interact, for example, with handheld devices of close-by citizens via Bluetooth Low Energy [12]. In this previous work, *Zamioculcas Zamiifolia* plants were exposed to four stimuli: wind, heat, blue light, and red light for data collection. Ten different artificial neural network (ANN) classifiers were trained based on the machine learning framework by Fawaz *et al.* [13]. The classifiers based on Convolutional Neural Networks (CNNs) achieved the best accuracies. The binary classifier (ANN architecture: Residual Network) achieved 98.1% accuracy and the five-class network (ANN architecture: Encoder) achieved up to 86.0% accuracy. The best classifiers were run on the embedded

hardware and performed on-board classifications, which were disseminated to a mobile phone. The main goal is to find accurate energy-efficient stimuli classifiers that can run on the PhytoNodes. Therefore, we extend the previous work by training five discriminant analysis classifiers (linear, quadratic, linear naive Bayes, quadratic naive Bayes, and Mahalanobis). We also retrain the ten deep learning classifiers on the raw electrophysiology data of only stimulus application intervals (see Sec. 4.3). We compare between all these approaches in terms of stimulus classification accuracy.

Our main motivation for this research is to exploit our results presented here to implement a phytosensing network. In the EU-funded project 'WatchPlant' [14, 15] (2021-2024), phytosensing is used to measure environmental features via living natural plants. The researchers of this project develop a self-powered, energy-efficient sensor network of electronic devices attached to plants [12, 16]. This forms a so-called bio-hybrid system of strongly coupled technological devices and living organisms. Such a bio-hybrid network can implement, for example, long-term monitoring and prediction of environmental conditions (e.g., air pollution) in urban settings. By measuring the physiology of natural plants in cities, we may be able to infer causing stimuli (e.g., particulate matter, increased ozone concentrations). Measured stress reactions in plants over periods of weeks or months in a given urban area may even allow to be correlated with health data of citizens. A distributed bio-hybrid system powered by electrophysiology, tissue impedance, and possibly additional means of measuring plant physiology may be used in the future to implement an early warning system about urban health hazards.

2. Related Work

Research on electrical activity in plants started in 1873 by John-Scott [17]. He conducted experiments on the carnivorous Venus flytrap (*Dionaea muscipula*), which is now a well-known example of action potentials [18]. Once an insect touches the sensitive hair of the trap leaves, action potentials are triggered, which gives an electrical impulse that leads to one of the fastest movements known in plants (100 ms), that is, the trap leaves close quickly. Electrical signals allow for rapid transmission of information compared, for example, to chemical signals. They are responsible for triggering plant-specific processes, including respiration, water uptake, leaf movement, biotic stress response, turgor pressure reduction, or photosynthesis changes. They also form the initial plant reaction to external stimuli (e.g., environmental changes) [19], and allow for internal communication between plant cells or organs. Depending on the species, plants have a resting potential between 50 mV and 300 mV [20]. This is caused by an unbalanced distribution of K^+, Na^+, Ca^{2+}, and Cl^- ions separated by a plasma membrane. However, the membrane has a variable permeability that may lead to variations in the potential. This can occur through active (e.g., H^+-ATPase) or passive (e.g., diffusion) transport mechanisms.

As discussed above, raw measured biopotential time series are not easily

interpretable. However, we are able to automatically interpret those signals by exploiting methods of pattern recognition and Time Series Classification (TSC) [21, 22]. In the common approach of discriminant analysis, discriminative statistical features (e.g., variance, skewness) are calculated to form a feature space that is used to distinguish between different classes of stimuli. For example, Najdenovska *et al.* [23] recorded the biopotential response of tomato plants before and after infestation by spider mites. They extracted up to 34 different features from measured time series and classified them with 80% accuracy using the gradient-boosted tree algorithm in a binary classification problem.

Chatterjee *et al.* [3] implemented a similar approach by exposing tomato plants to various stimuli. They measured the biopotentials of tomato plants in response to sodium chloride (5 ml and 10 ml, 3 mol), sulfuric acid (5 ml, 0.05 mol), and ozone (16 ppm/min). The authors extracted eleven features from the data and used them to train five discriminant analysis classifiers, including linear, quadratic, naive Bayes, and Mahalanobis classifiers. They trained classifiers for different classification problems (e.g., one-vs-one and one-vs-rest) with varying numbers of features. Using quadratic discriminant analysis, they achieved up to 95% accuracy for a binary classification of sulfuric acid and ozone. Our discriminant analysis approach for classification is based on this work.

In another approach, Reissig *et al.* [24] classified the ripening stage of tomatoes based on the electrical activity of the fruit. They classified the ripening stages based on three levels: mature green, breaker ($< 10\%$ of the fruit surface is not green), and light red. They used principal component analysis based on features gained by approximate entropy, Fourier transformation, power spectral density, and wavelets. The gathered principal components were used to train decision trees, support vector machines, Gaussian processes, k-neighbors, random forests, Gaussian naive Bayes, and dummy classifiers. The dummy classifier serves as baseline for the aforementioned classifiers as it uses several unintelligent ways to classify, for instance, assigning all new samples to the class with the highest prior probability. The three-class support vector machine classifier achieved the highest accuracy of 74.4%.

Another approach for such classification problems is training ANN classifiers. Generally, one has two options. The ANN classifiers can either be trained using the raw biopotential data or statistical features can be extracted that serve as input to the ANN. Xiao *et al.* [25] used raw biopotential data and investigated the salt tolerance of wheat seedlings. As mentioned above, training ANNs requires a large amount of data for training the weights [10], which can be challenging in plant science [11]. Facing this challenge, Xiao *et al.* used a conditional generative adversarial network to generate 200 artificial samples based on 127 natural plant experiments. Using this data they trained a one-dimensional CNN that achieved 92.3% accuracy. In another example, Pereira *et al.* [26] exposed soybean plants to three different stimuli: cold, low light, and osmotic stress. They followed two approaches to train five machine learning classifiers: ANNs, CNNs, optimum-path forest, k-nearest neighbors, and support vector machine. In the

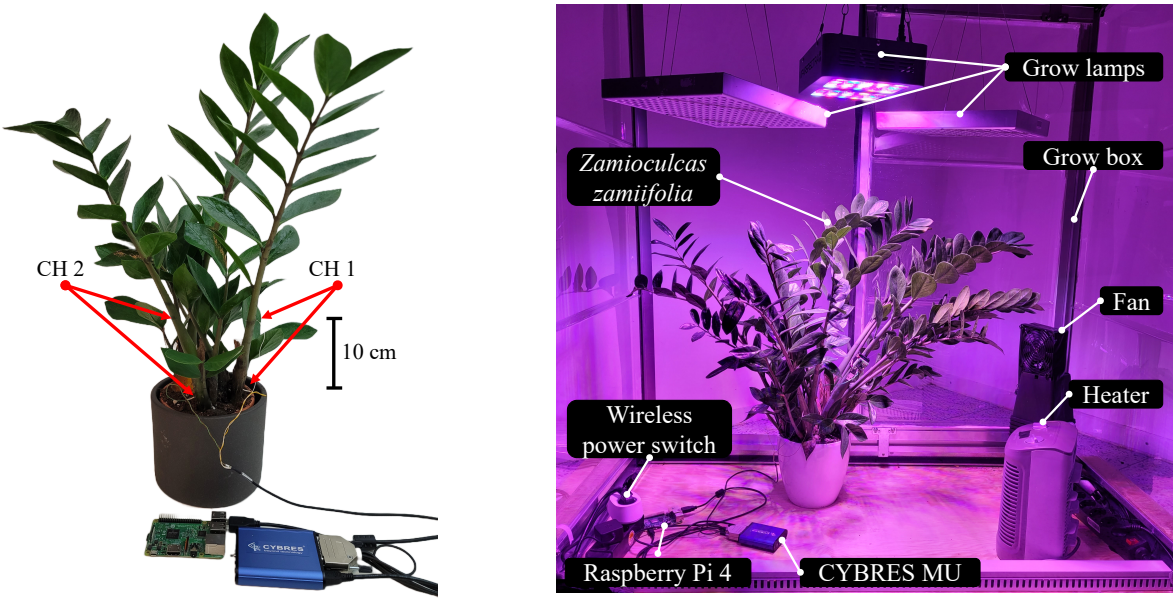
first approach (called interval arithmetic), they separate the raw data into intervals of certain length. They calculate each interval's minimum, average, and maximum as features. They obtain a chronological array of triplets as a training dataset. In the second approach, they used visual rhythm encoding to generate images from the raw data that were then used to train CNN classifiers. The authors reached accuracies of up to 90.7% with support vector machines in a binary class classification and up to 71.2% in multi-class classifications with k-nearest neighbors. Interestingly, the authors concluded that deep learning techniques are less efficient (i.e., lower accuracies).

3. Methods

Our main goal is to interpret natural plant responses to various environmental stimuli. Therefore, we attach our biopotential sensors to *Zamioculcas zamiifolia* and impedance sensors to *Solanum lycopersicum* in controlled experiment setups, as described in Sec. 3.1. We expose the plants to a single stimulus at a time in multiple data collection plant experiments (see Sec. 3.2). We use the collected data to train stimulus classifiers following two different approaches: discriminant analysis (see Sec. 3.3) and deep learning (see Sec. 3.4).

3.1. Bio-Hybrid experiment setups

In our first indoor experimental setup, we measure the biopotential of the 'ZZ' plant (*Zamioculcas zamiifolia*, family Araceae), which is a stemless forest plant with dark green glossy leaves [27] (see Fig. 1a). The 'ZZ' plant shows fast electrical responses to various stimuli and is robust to grow in technical laboratories without excessive maintenance (i.e., non-greenhouse condition). We minimized external environmental influences by conducting our experiments in a grow box of size 120 cm × 60 cm × 160 cm (L × W × H), see Fig. 1b and fully controlled the light conditions (no natural light). In our setup we can apply four different stimuli: wind, heat, red light, and blue light. The wind stimulus is generated by a fan placed 60 cm from the plant pot and at a height of 30 cm. We placed a heater (ROWENTA Mod.S02220F0) at a distance of 10 cm from the plant pot to increase the temperature inside the growbox. For both red and blue light stimuli, we use a total of six 45 W 'Erligpowht' LED grow lamps and place them around the plant. An 'Erligpowht' grow lamp contains 225 LEDs (165 red and 60 blue) with peak emissions λ_{max} at wavelengths 650 nm and 465 nm, respectively. We concealed the blue LEDs of three lamps for the red light stimulus. Similarly, we concealed the red LEDs of the remaining three lamps for the blue light stimulus. Two lamps are placed 30 cm above the plant. Two each are placed on both sides 30 cm from the plant. In summary, we have one blue and one red lamp on top and the two sides each. To minimize human intervention, we use programmable TP-Link HS 100 power sockets to automate the experiments. The electrical activity of the plant is recorded using the CYBRES phytosensor [28] in combination with a Raspberry Pi 4, which stores and uploads the



(a) (b)

Figure 1: The experimental setup for biopotential measurements. (a) Electrode placement of the two biopotential channels (CH1 and CH2) of the CYBRES phytosensor on the *Zamioculcas zamiifolia*. (b) The complete experimental setup with the CYBRES phytosensor, attached to ZZ plant, grow lamps, a fan, a heater, a wireless power switch, a Raspberry Pi 4, and a power supply.

measured data. We insert the two silver electrodes (Ag99) of each biopotential channel of the CYBRES phytosensor 10 cm apart into the lower end of one thickened petiole of the ZZ plant, see Fig. 1a

In our second experimental setup, we measure the tissue impedance of tomato plants (*Solanum lycopersicum*, family Solanaceae) in response to light (combination of red/blue LEDs as used in the first experimental setup for biopotential measurements, placed above the plants at about 0.5 meter). We apply the tissue impedance in the stem areas which is an example of an electrochemical sap flow measurement. Given that sap flow depends on many different parameters (e.g., irrigation cycles, leaf transpiration, cyclic hour-day rhythms, etc. [29]), we automatically trigger the environmental stimuli (e.g., light and irrigation) at fixed times with one experiment per day. Here, we use the two impedance channels of the CYBRES phytosensor to measure the sap flow of the plant [30], see Fig. 2. The electrodes of each channel are inserted 2 mm into phloem/xylem vascular tissue [31] of the low stem area at distances of about 1 cm between electrodes respectively, see Fig. 2. Typical tissue RMS-impedance is between 40 and 60 kΩ under these conditions. Due to electrochemical measurements, the variation of penetration depth and position is reflected in a variation of initial impedance in the prestimulus area. Therefore, all calculations should be performed in relation to this signal level.

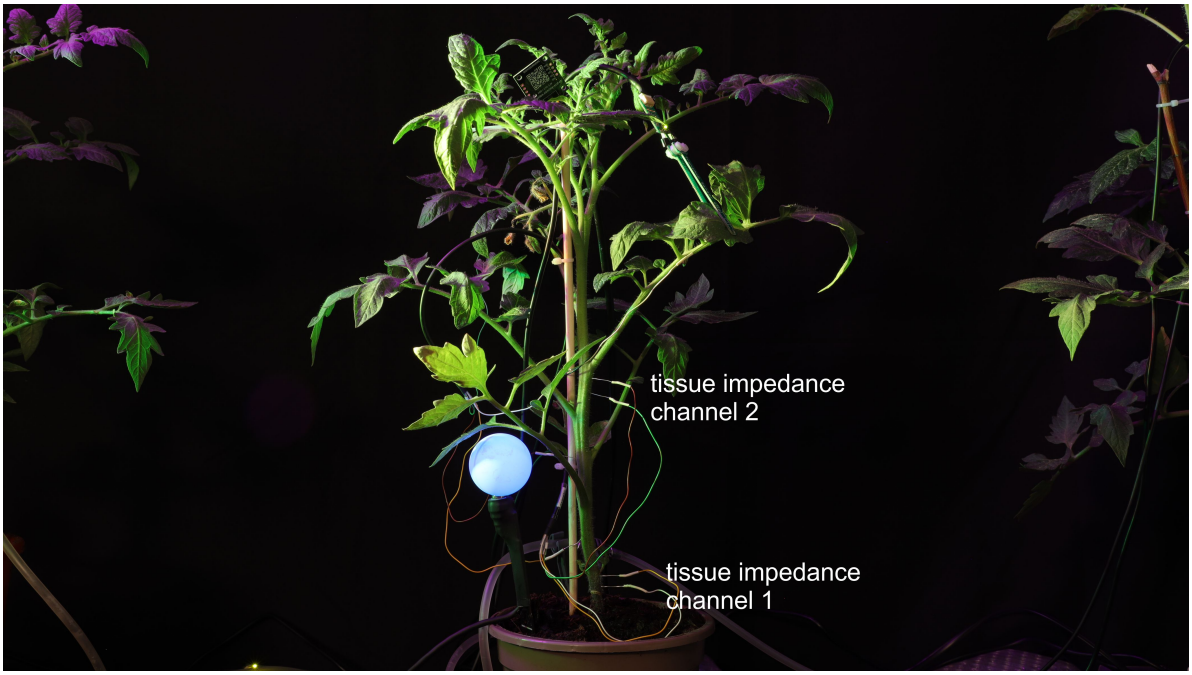


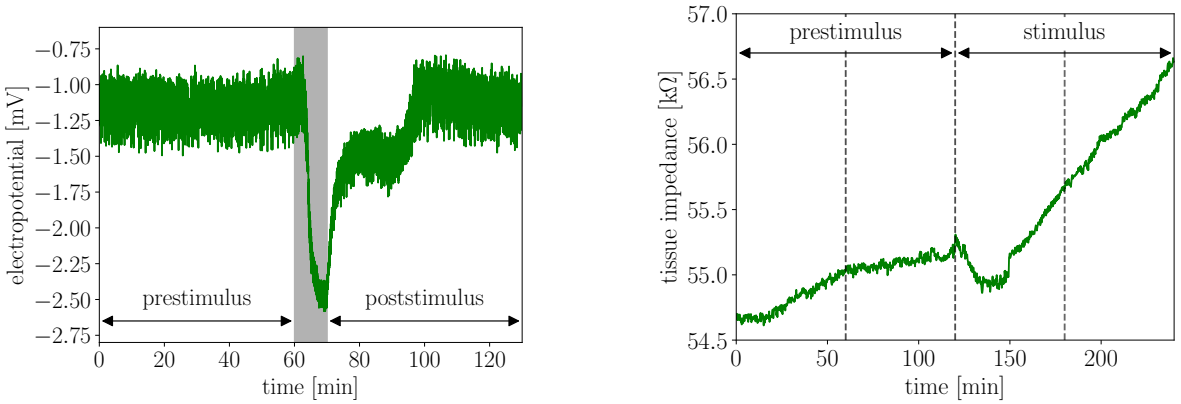
Figure 2: Setup with tissue impedance measurements (two independent channels) with tomato plants *Solanum lycopersicum*.

3.2. Data Collection

We use the experiment setups as described in Sec. 3.1 to run our experiments for automatic data collection. We run two types of data collection experiments: biopotential and impedance experiments. The collected data is labeled according to the applied stimulus and used to train and evaluate both the discriminant and deep learning classifiers. Given the labeled data, we can directly use the raw data or explore stimulus-specific features that lead to optimal class separation.

All biopotential experiments follow the same experimental protocol and are divided into three phases. Following the nomenclature of Chatterjee et al. [3], an experiment starts with a *prestimulus* phase of 60 minutes followed by a *stimulus* application phase of 10 minutes, and the experiment is concluded with a *poststimulus* phase of 60 minutes (see Fig. 3a).No stimulus is applied during either *prestimulus* and *poststimulus* phase to allow a stable initial plant state before the next stimulus application. The CYBRES phytosensor measures the plant's biopotential responses at a frequency of 0.58 Hz.

We did 544 experiments with wind stimulus, 504 experiments with heat stimulus, 134 experiments with blue light stimulus, and 138 experiments with red light stimulus. These result in a dataset of 1320 univariate time series of approximately 4500 samples each. We aim to train classifiers that are capable of identifying whether any of the previously mentioned stimuli is currently applied to the plant. Therefore, our classifiers should also be trained on biopotential measurements where no stimulus is applied. For this reason, we extract a total of 544 twenty minute intervals from randomly selected



(a) (b)

Figure 3: Example measurements of the biopotential and tissue impedance where the respective stimulus application time is marked with a gray background. (a) Example measurement of the biopotential change caused by a wind excitation using our experimental protocol. (b) Example measurement of the tissue impedance change caused by a light stimulus using our experimental protocol. The black dotted lines indicate interval boundaries which are used to differentiate between the two classes light and no stimulus, see Sec. 3.3.

prestimulus phases, which corresponds to the largest number of experiments from one class (i.e., wind stimulus). Now, our final dataset is composed of 1864 univariate time series.

Since tissue impedance has a slower reaction time than electrical potentials [32], we implement a different experiment protocol for impedance experiments with longer stimulus phase. The protocol is divided into a 120 min prestimulus phase followed by a 120 min stimulus application phase (see Fig. 3b). We conducted 20 experiments where we applied a light stimulus. The tissue impedance is sampled with a frequency of $\approx$ 0.08 Hz. From these experiments, we obtain 40 univariate time series of approximately 1250 samples, because of the two impedance channels of the MU. We use the collected impedance data to train classifiers that are capable of distinguishing between only two classes: *no stimulus* and *light*. Therefore, from each experiment, we extract the 120 min prestimulus intervals and label them as no stimulus. This results in a dataset of 80 univariate time series.

3.3. Discriminant Analysis Classification

To perform the discriminant analysis, statistical features are first calculated from the time series and then used for classification. In general, we are interested in classifying the plant responses during active stimulus application phases. That's why, for biopotential based classification, we extract the features for the first 340 samples ($\approx$ 9.8 min) of the stimulus phase and use the last 340 samples of the prestimulus phase for background

subtraction (see Fig. 3a). We follow the same procedure for each stimulus class (wind, temperature, blue light, and red light). In the case of no stimulus class, we split the 20 min time series into two halves. Similar to the other classes, we now extract the last 340 samples from the first half for background subtraction and the first 340 samples from the second half for classification.

In preprocessing the impedance dataset, we are again only interested in the change caused by a stimulus. For this reason, we extract the features for the first 295 samples ($\approx$ 60 min) of the stimulus phase and use the last 295 samples of the prestimulus phase for the background subtraction for the class light, see dashed lines in Fig. 3b. The time series of the nostimulus dataset no stimulus is again split in half and processed as for the nostimulus class in the biopotential measurements. We use the same division of data into test and training in both classification approaches. Here, 70% of all data is used for training, while the remaining 30% is used for testing. Since the datasets per stimulus are of different sizes, the respective proportions in the testing and training sets are kept constant.

Discriminant analysis begins by computing nine statistical features from each univariate time series. We use a subset of the statistical features defined by Chatterjee *et al.* [3] including the mean (μ), variance (σ^2), skewness (γ), kurtosis (β), interquartile range (IQR), Hjorth mobility (HM), Hjorth complexity (HC), wavelet packet entropy (WPE), and Average Spectral Power (ASP). The first four moments of distributions μ, σ^2, γ, β describe basic properties of a distribution [21]. The first moment μ is the expected value or mean. The second moment σ^2 gives the dispersion of the data around the mean. The third moment γ indicates the asymmetry around the mean. β is the fourth moment and indicates whether the data have a heavy or light tail compared to a normal distribution. These moments of distributions are calculated as follows:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2,\ \gamma = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^3,\ \beta = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^4, \quad (1)$$

where x_i represents the samples and n the total number of samples. IQR is another descriptive characteristic that describes the dispersion for the middle 50% of the data that lie between the first quartile Q_1 and the third quartile Q_3 of the observed data [33]. It is particularly useful for skewed distributions because it is unaffected by outliers. The IQR is defined as

$$IQR = Q_3 - Q_1. \quad (2)$$

Hjorth parameters [34, 35] are commonly used to analyze signals in the time domain and for feature extraction in electroencephalography (EEG) signals. The parameters are based on the variance σ_1^2 of the signal and its first σ_2^2 and second σ_3^2 derivatives. For discrete series [36], the variances can be calculated as

$$\sigma_1^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \sigma_2^2 = \frac{1}{n-1}\sum_{i=2}^{n}\left(x_i - x_{i-1}\right)^2, \quad \sigma_3^2 = \frac{1}{n-2}\sum_{i=2}^{n-1}\left(x_{i+1} - 2x_i + x_{i-1}\right)^2. \quad (3)$$

Based on the these variances, HM and HC can be calculated as follows

$$HM = \frac{\sigma_2}{\sigma_1}, \qquad HC = \sqrt{\frac{\sigma_3^2}{\sigma_2^2} - \frac{\sigma_2^2}{\sigma_1^2}}. \tag{4}$$

Wavelets are used to decompose signals arising from multi-scale processes such as climatology or neuroscience [37]. These wavelets are decaying wave like oscillations with zero mean that enable us to analyze complex signals in the frequency and time domain in different resolution. For instance, we get a high frequency and low time resolution in lower frequency ranges since the time signal barley changes. While increasing the frequency we increase the time resolution and decrease the frequency resolution since changes in the time domain occur more often. Through the discrete wavelet transformation, we obtain wavelet coefficients $\mathcal{W}_\psi(f)(j,k)$ that are defined by the convolution of the signal f with a wavelet ψ:

$$\mathcal{W}_\psi(f)(j,k) = \langle f, \psi_{j,k} \rangle \quad \text{where} \quad \psi_{j,k}(t) = \frac{1}{a^j}\psi\left(\frac{t - kb}{a^j}\right). \tag{5}$$

ψ is also called mother wavelet while $\psi_{j,k}$ represents the wavelet family. It is a set of scaled and translated versions of the mother wavelet ψ. We use the Daubechies 1 mother wavelet in our approach. a denotes a scaling while b is used to translate the wavelet across the signal. j and k denote the scaling and translating index while moving a scaled wavelet through the signal. Based on the gained wavelet coefficients and the Shannon entropy we are able to calculate the wavelet entropy. A low WPE indicate more organized data while a high WPE indicate disordered data and is defined as

$$WPE = -\sum_{\mathcal{W} \in \mathcal{W}_\psi} \mathcal{W}^2 \log\left(\mathcal{W}^2\right). \tag{6}$$

The average spectral power (ASP) is the integral over the power spectral density (PSD), also called the power spectrum. The PSD is a measure of the energy variation in a signal distributed over the measured frequencies. We calculate the PSD using the estimate provided by Welch [38]. The signal is first divided into k overlapping segments, then a Hann window is applied to each segment. We use a segment size of 50 samples in our approach, with 25 samples overlapping between segments. After windowing, the periodogram of each segment is calculated using the discrete Fourier transform. Then, all k periodograms are averaged. The resulting function $PSD(f)$ is the power spectral density as a function of frequency f. Finally, the discrete ASP is defined as the sum of all PSDs:

$$ASP = \sum_{f \in F} PSD(f), \tag{7}$$

where F denotes all frequencies contained in the signal.

After feature calculation, we apply background subtraction by subtracting each prestimulus time series feature from the stimulus time series features to analyze only the change in signal. The resulting incremental features $\mathbf{x'}_i \in \mathbb{R}^n$ of the test and training

set are min-max normalized based on the training data where $i \in [\mu, \gamma, \beta, \text{IQR}, \text{HM}, \text{HC}, \text{ASP}, \text{WPE}]$ and $j \in [1, ..., n]$:

$$x'_{i,j} = \frac{x_{i,j} - \min(\mathbf{x}_{\mathbf{i},\text{train}})}{\max(\mathbf{x}_{\mathbf{i},\text{train}}) - \min(\mathbf{x}_{\mathbf{i},\text{train}})} \,. \tag{8}$$

The normalized training set is now used to train five discriminant analysis classifiers. These include Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Naive Bayes LDA and QDA, and Mahalanobis. We use the implemented classifier of the Matlab framework *classify*‡. In this classification approach, we assume that each class can be approximated by a Gaussian distribution [21]. We start with estimating the mean $\mu = E[\mathbf{x}]$ and the covariance matrix $\Sigma = E\left[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T\right]$ to build the probability density function $p(\mathbf{x}|\theta)$ of each class based on the known observations $\mathbf{x}$ in a l-dimensional feature space:

$$p(\mathbf{x}|\theta) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right) \quad \text{with } \theta = \{\mu, \Sigma\} \tag{9}$$

Based on the probability density function $p(\mathbf{x}|\theta)$ an the prior probability $P(\theta)$ we assign new observations $\mathbf{x}$ to the class $\hat{\theta}$ with the highest a posterior $P(\theta|\mathbf{x})$ probability.

$$\hat{\theta} = \underset{\theta}{\text{argmax}}\ P(\theta|\mathbf{x}) = \underset{\theta}{\text{argmax}}\ p(\mathbf{x}|\theta)P(\theta) \tag{10}$$

All mentioned classifiers follow a similar methodology, but differ in the estimated covariance matrix. In LDA, we assume that all classes have the same covariance matrix Σ. However, since the covariances of the classes vary, the pooled covariance matrix of all classes is determined. In contrast, the quadratic and Mahalanobis classifiers use stratified covariance estimation. However, the prior probabilities are not considered by the Mahalanobis classifier. In both Naive Bayes classifiers, we assume independent features whereby only the diagonal entries of the covariance matrix are estimated.

We train all five classifiers in three different settings with an increasing number of classes. The first setting involves a binary classification between no stimulus and wind. We increase the complexity to a three-class classification by adding the heat class. The final setting includes all stimuli, consisting of wind, heat, red light, blue light, and no stimulus. If we use all nine available features, the difficulty of the classification problem could increase significantly. The classification might be inefficient and computationally demanding. In the worst case, the classification accuracy may decrease as the dimensionality of the feature space increases with each additional feature. To counter this problem, Sequential Forward Selection (SFS) is applied before each training session. SFS is a greedy algorithm that starts with an empty feature set. It sequentially expands the feature set with individual features based on the training set and accuracy improvement. SFS terminates when no improvement in accuracy can be

‡ https://de.mathworks.com/help/stats/classify.html

achieved by adding more features. We use the Matlab framework *sequentialfs*§ for SFS. All collected datasets and trained classifiers, as well as the code for the experiment and data collection setup, are available online [39].

3.4. Deep Learning-based Classification

Given the success of deep learning in recent years in many fields, such as image processing [40], natural language processing [41], drug discovery [42], and also time series prediction [13], we want to compare deep learning approaches with current state-of-the-art discriminant analysis in the application of stimulus classification. Therefore, we are using a deep learning framework for TSC developed by Fawaz *et al.* [13]. This framework includes ten different deep end-to-end classifiers, which take an electrophysiological plant signal time series as an input and output a likelihood for each possible class, i.e., applied external stimulus. All approaches are implemented using the machine learning library Tensorflow 2.0‖ (with Python 3). We trained the networks on a Nvidia DGX A100 with eight cores (two in use). Similar to the discriminant analysis approach, we are interested in classifying the plant's reaction to the applied stimulus. Therefore, we extract the first 340 samples ($\approx$ 9.8 min) of the stimulus phase. Next, we normalize them to remove any shifts or data scaling due to the inherent bias of various plants [43] and electrode placement. We apply the Z-score normalization: $\boldsymbol{x}_{norm} = (\boldsymbol{x} - \mu_x)/\sigma_x$, with μ_x the mean and σ_x the standard deviation of the extracted series $\boldsymbol{x}$.

The deep learning framework contains three different fundamental structures: ANNs, CNNs, and Echo State Networks (ESN). In the following, we explain the general idea of each fundamental structure and the remarkable differences in each implementation.

The ANN structure is the simplest and most traditional architecture for deep learning. The architecture consists of artificial neurons, which are stacked up to several layers connected in series. Each neuron has a non-linear activation function and receives the output of all neurons from the previous layer (fully connected) as input. The first layer receives one time series as one input vector¶. The final layer contains a softmax activation function for providing a likelihood distribution over the available classes. The cross-entropy loss function is used in all approaches, if not stated otherwise, because it is a good measure of how distributions differ. The foremost disadvantage of this architecture in TSC tasks is that the architecture does not incorporate spatial invariance [13], meaning that each time step of the time series has its own weights and is considered individually, leading to loss of temporal information, i.e., initially the network does not know that time step $t + 1$ comes after time step t. The sole implementation of this approach is the Multi-Layer Perceptron (MLP) [44]. Each layer is preceded by a

§ https://de.mathworks.com/help/stats/sequentialfs.html
‖ https://www.tensorflow.org; A list of further required Python 3 packages can be found at [39].
¶ A detailed structure of the architecture, e.g., dimensions of layers, of the various implementations can be found at [39].

dropout operation which helps to prevent overfitting [45]. In this implementation, the AdaDelta optimizer is used.

CNNs are common used architectures because of their success in many other domains, such as image [46] or natural language processing [41]. In the context of TSC tasks, convolutional layers (main ingredient of CNNs) can be considered as filters, which learn invariant discriminative features of time series [13]. In comparison to the aforementioned ANNs, the weight sharing property of convolutional layers allows for invariance because multiple time steps share the same weights, i.e., the network *knows* the coherence of the time steps initially. The general CNN structure has two stages: first learning a feature representation using convolutional layers, which can be augmented to convolutional blocks by adding pooling operations for further invariance. In the second stage, the output of the convolutional layers or blocks are forwarded to one or more fully connected layers for classifying the learned features. Finally, similar to the ANN approach, the last layer usually contains a softmax activation function that provides a likelihood distribution over the available classes.

The first implementation of the framework using this architecture is the Time Convolutional Neural Network (TCNN) [47] with two convolutional blocks. It is the only implementation, which does not use a softmax activation function but a sigmoid activation function and thus a mean squared error loss function (with Adam as an optimizer). The second implementation is a Fully Convolutional Neuronal Network (FCN) [44], which has three convolutional blocks where a batch normalization operation instead of pooling operation is applied [48]. Furthermore, a global average pooling layer instead of fully connected layers is used. The subsequent implementation is the Encoder [49] with three convolutional blocks, where an instance normalization operation [50] and a dropout operation are applied in each block. The major difference compared to the FCN is that an attention layer replaces the global average pooling layer [51]. The implementation of Time Le-Net (t-LeNet) [52] follows the structure of a classical CNN. In this approach additional preprocessing of the time series is done, that is, Window Warping (WW) and Window Slicing (WS), for data augmentation. As WW and WS are also applied during inference, the classification result is a majority vote over all window slices generated from one time series. The Multi-scale Convolutional Neural Network (MCNN) [53] also has a traditional CNN structure. The main difference lies in the heavy data preprocessing. First WS is used for data augmentation. Second, the slices are transformed using identity mapping, down sampling, and smoothing before being passed to the network. Similar to the t-LeNet for classification, a majority voting over all window slices of the time series is applied. The Multi-Channel Deep Convolutional Neural Network (MCDCNN) [54] is originally implemented for multivariate time series, as it splits them into multiple univariate time series. The network collapses to a traditional CNN structure in our application because we only consider univariate time series. It is the only approach using a stochastic gradient descent as an optimizer paired with the categorical cross-entropy loss function. The implemented Residual Network (ResNet) [44] has the deepest architecture with nine convolutional blocks. To prevent

the vanishing gradient problem skip connections are added every three convolutional blocks. The vanishing gradient problem is caused by the deep architecture, as during backpropagation the gradient for updating the weights becomes close to zero and therefore the weights do not change during training. The last implementation using the CNN structure is the Inception Neuronal Network [55]. It consists of six inception modules, which are characterized by multiple convolutions in parallel – following the idea of making the network wider, not deeper – instead of the classical convolutional blocks. In addition, similar to the ResNet, skip connections are added every three blocks.

Finally, we want to present the Echo State Network (ESN) which is a special form of recurrent neuronal networks designed for time series prediction [13]. As recurrent neuronal networks are hard to train due to the vanishing gradient problem and no convergence because of cyclic dependencies, in ESNs, only the output weights are changed during learning. The hidden layers, also called reservoir, and input weights are randomly initialized and kept constant. The framework implements the Time Warping Invariant Echo State Network (TWIESN) [56].

We train all previously mentioned deep classifiers for the same three classification settings as described in Sec. 3.3, that are: no stimulus and wind; no stimulus, wind and heat; and no stimulus, wind, heat, red light, and blue light. For each classification setting we do five repetitions (iterations) to minimize the influence of the random initialization of the weights. Similar to the discriminant analysis we split the data into 70 % training and 30 % testing sets. Given a time series of plant biopotential measurements as input, the networks are trained to find the likelihood within the stimulus classes mentioned above. We use three metrics to evaluate the performance of the ten deep classifiers: accuracy, precision, and recall. The accuracy is the ratio of correct classified samples and is defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{11}$$

with TP true positives, TN true negatives, FP false positives, and FN false negatives. With similar nomenclature we can define precision, which describes the proportion of actually correct classified positive samples, as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{12}$$

and recall, which describes the proportion of actual correctly identified positive samples,

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{13}$$

The precision and recall is calculated per class and then averaged over all classes.

4. Results

In this section, we report results from both the discriminant analysis classifiers (see Sec. 4.1) and the deep learning time series classifiers (see Sec. 4.2), followed by a discussion (see Sec. 4.3).

4.1. Discriminant Analysis

The nine computed statistical features (as described in Sec. 3.3) of each class are presented in violin plots in Fig. 4.[+] Each violin is divided into two halves. The left side shows the distribution of the normalized training set, while the right side reflects the normalized test set. Especially in classification tasks, we can use these plots to get a first impression of a possible separability of the data. On the one hand, high classification accuracy can be expected if the distributions of the classes do not overlap on the y-axis. On the other hand, high accuracy can be achieved if the training and test data of one class do not differ significantly, that is, both subsets represent the same data distribution.

This is confirmed for the binary classification of no stimulus (first violins in Fig. 4) and wind (second violins in Fig. 4). Any one of the features mean, variance, or ASP is required to achieve 100% accuracy with each of the five classifiers (see Table 1). Regardless of the classifier, we also achieve accuracies above 99% with the other features: kutosis, IQR, and WPE. The lowest accuracy of 89.5% is achieved with feature HC in combination with LDA, naive Bayes LDA, and Mahalanobis.

As more classes are added, the classification task becomes more difficult. This leads to a slight decrease in the highest accuracy to 99.8% when only one feature is used in the case of three classes. This is achieved when the mean is used in combination with the Mahalanobis classifier. However, accuracies above 99% are achieved with QDA, naive Bayes QDA, and Mahalanobis using the WPE and the mean (Table 1). Already in Fig. 4, a good separation of the three classes with WPE (Wentropy) or mean can be seen, since most of the data does not overlap on the vertical axis. Other promising features are variance, kurtosis, and IQR, as they achieve over 90% accuracy with all classifiers. The feature HC performs the worst, achieving an accuracy of ≈60% with any classifier.

In a next step, we investigate whether higher accuracies can be achieved if more than one feature is used. We select features using SFS as described in Sec. 3.3. The accuracy of QDA and Mahalanobis increased to 100% with a feature combination of mean and WPE. Adding ASP as a third feature, LDA also achieves 100% accuracy. Naive Bayes LDA achieves 99.6% accuracy with variance, HM, and WPE, while naive Bayes QDA achieves 99.8% with variance and WPE.

The highest accuracy of all classifiers using one feature (variance) is decreased to ≈90% in our five-class case. The lowest accuracy of ≈50% with all classifiers is achieved with the feature skewness. Again, we use SFS to increase the accuracy of the classifiers. In order to get an intuitive understanding of the task difficulty and the quality of the classifiers, we show the classification bounds and test data for all five classes and each classifier after adding the second feature in Fig. 5. According to SFS, the highest accuracy of 99.1% is achieved with QDA and all features except skewness. The second

[+] Violin plots combine boxplots with smoothed histograms and visualize more features of the data's distribution [57].

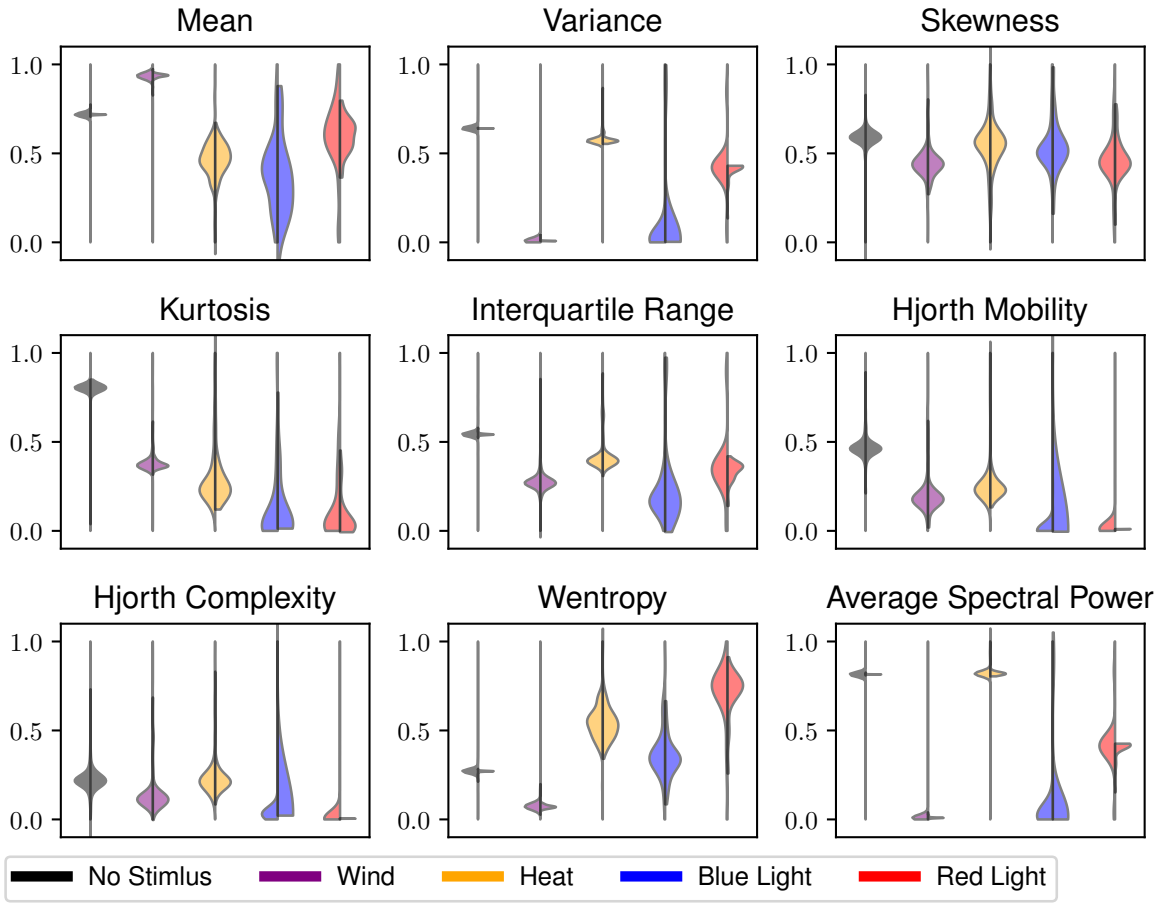


Figure 4: Biopotential dataset, violin plots of the distribution of all nine normalized features: mean μ, variance σ^2, skewness γ, kurtosis β, interquartile range (IQR), Hjorth mobility (HM), Hjorth complexity (HC), wavelet packet entropy (WPE, Wentropy), and average spectral power (ASP). Each violin represents one of the classes, from left to right: no stimulus, wind, temperature, blue light or red light. The left half of each violin shows the distribution of the training set, the right side that of the test set.

highest accuracy of 98.9% is achieved with the Mahalanobis classifier and the features mean, variance, skewness, WPE, and ASP. After adding IQR and Hjorth mobility to the last set of features, we obtain the highest accuracy of 98.6% with LDA. The naive Bayes classifiers require the fewest features but have the lowest accuracy. Naive QDA requires only the variance and mean to achieve its highest accuracy of 98.4%. After adding HM, we obtain the required feature subset of the naive Bayes LDA classifier to achieve an accuracy of 97.8%

We conclude this study on discriminant analysis with the results for the impedance dataset. The nine computed statistical features of the two classes light and no stimulus based on the impedance dataset are shown in Fig. 6. In this binary classification, we achieve a maximum accuracy of 96% with each classifier when ASP or kurtosis is taken as a single feature (see Table 2). The accuracies of QDA and Mahalanobis can be

Table 1: Biopotential dataset, accuracies for all five discriminant analysis classifiers in percent, first with two (wind & no stimulus), then three (+ temperature) and finally all five (+ red & blue light) classes. In addition, the subset of features selected by SFS. The order of features shown corresponds to that of SFS as it added features sequentially.

classifier	setting	accuracy [%]	feature set
LDA	two classes	100.0	[ASP]
	three classes	100.0	[WPE, μ, ASP]
	five classes	98.6	[σ^2, WPE, ASP, IQR, HM, μ]
Naive Bayes LDA	two classes	100.0	[ASP]
	three classes	99.6	[WPE, σ^2, HM]
	five classes	97.8	[σ^2, μ, HM]
QDA	two classes	100.0	[ASP]
	three classes	100.0	[WPE, μ]
	five classes	99.1	[σ^2, HM, ASP, IQR, β, HC, WPE, μ]
Naive Bayes QDA	two classes	100.0	[ASP]
	three classes	99.8	[WPE, σ^2]
	five classes	98.4	[σ^2, μ]
Mahalanobis	two classes	100.0	[ASP]
	three classes	100.0	[μ, WPE]
	five classes	98.9	[σ^2, μ, WPE, ASP]

Table 2: Impedance dataset, accuracies for all five discriminant analysis classifiers in percent and the subset of features selected by SFS. The order of features given is that of SFS as it adds features sequentially.

classifier	accuracy [%]	feature set
LDA	96.0	[ASP]
Naive Bayes LDA	96.0	[ASP]
QDA	100.0	[WPE, μ]
Naive Bayes QDA	88.0	[WPE]
Mahalanobis	100.0	[WPE, μ]

increased to 100% by SFS when the features WPE and μ are used for classification. However, SFS decreased the accuracy of the naive Bayes QDA when WPE was chosen instead of ASP or Kurtosis. This is because SFS selects the subset of features based on the training data.

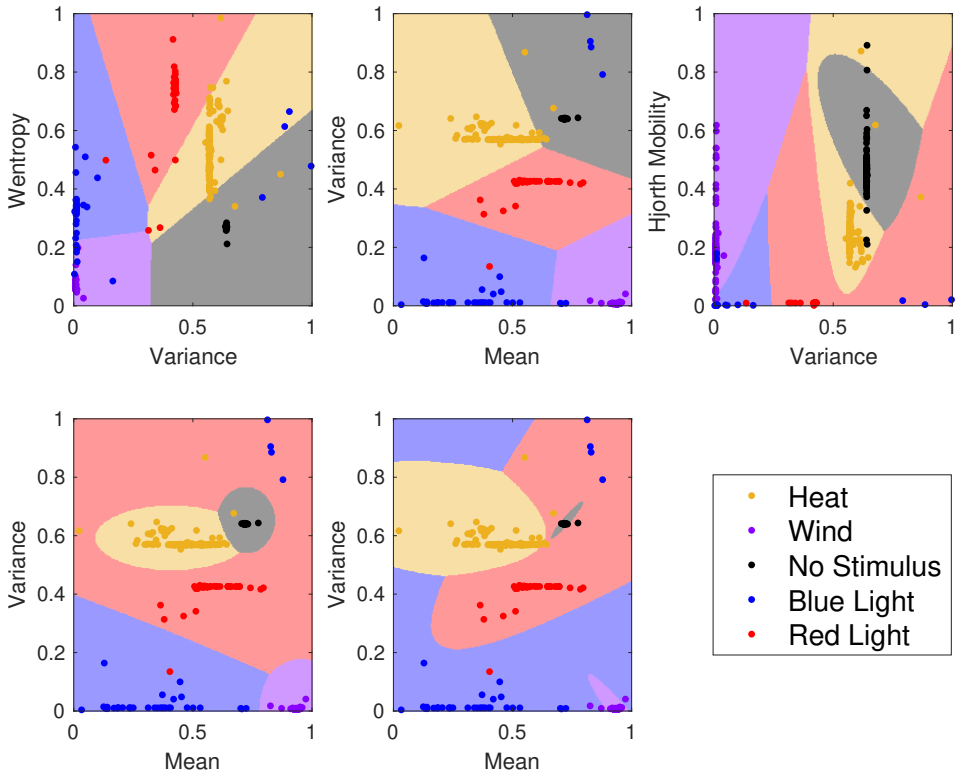


Figure 5: Biopotential dataset, class ranges and test set samples of the five-class classifiers using the first two features selected with SFS, upper row from left to right: LDA, naive Bayes LDA, QDA; lower row from left to right: naive Bayes QDA and Mahalanobis.

4.2. Deep Learning-based Classification

We trained ten deep end-to-end classifiers as described in Sec. 3.4 on the biopotential measurement dataset. We did five iterations per classifier to minimize the influence of weight initializations. The averaged results for accuracy, precision, and recall for nine out of ten classifiers can be seen in Table 3. The MCNN could not learn an appropriate representation of the data, so we do not consider it in further analysis.

For the two-class classification task, the Inception model has the highest accuracy with 89.7%, followed by the ResNet (89.4%) and the FCN (89.3%). The TWIESN performs worst (57.8%) in differentiating wind and no stimulus. The average accuracy over all classifiers is 80.8%. The precision is within 1% deviation from the accuracy for all classifiers except for MCDCNN ($\Delta = 1.2\%$) for this setting. As described in Sec. 3.4 the recall is calculated using the true positives and false negatives per class. Only looking at one class at a time implies that there are only positive samples, and thus there are no true and false negatives. Hence, in a perfectly balanced classification problem, i.e., in our case, the recall is similar to the accuracy.

By adding the heat stimulus, we have a three-class classification task, where the Inception model again has the highest accuracy with 92.2%, followed by the ResNet

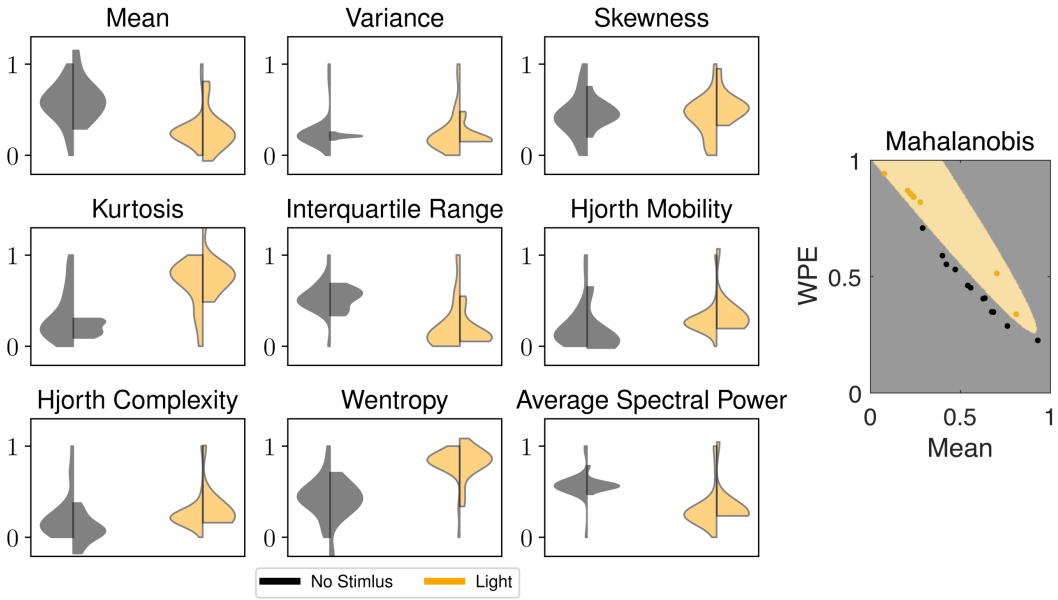


Figure 6: Impedance dataset, violin plots of the distribution of all nine normalized features. Each violin represents one of two classes: no stimulus or light. The left half of each violin shows the distribution of the training set, while the right side shows the distribution of the test set. On the far right, the class ranges and test set samples of the Mahalanobis classifier using the features mean and WPE.

(92.0%) and the FCN (91.4%). The TWIESN performs worst with 53.2%. Besides the three best models, the Encoder and the MCDCNN achieve higher accuracies compared to the two-class setting, and the average accuracy of all approaches drops slightly to 80.3%. The precision in the three-class classification task of all classifiers is within 1 % deviation from the accuracy MCDCNN ($\Delta = 1.4\%$) and TWIESN ($\Delta = 2.6\%$). The third class, heat stimulus, is slightly smaller than the other two classes, but the average recall for all classifiers is within 1 % deviation from the accuracy.

Finally, we present the five-class classification task by adding the blue light and red light stimulus. In this setting, the ResNet and the FCN have the highest accuracy with 83.5%, followed by the Inception model (83.0%). Similar to the previous settings, the TWIESN has the lowest accuracy (43.2%), leading to an average accuracy of 70.1% for all approaches. Due to the unbalanced classes, the precision deviates significantly from the accuracy. The largest deviation can be observed in the TWIESN with $\Delta = 14.3\%$ followed by the Inception model ($\Delta = 10.0\%$) and TCNN ($\Delta = 9.5\%$). The unbalanced dataset has a similar effect on the recall, where now the deviation from accuracy is, on average $\Delta = 12.9\%$, with t-LeNet having the highest deviation ($\Delta = 16.8\%$) and MLP having the lowest deviation ($\Delta = 9.7\%$).

Further, we provide a confusion matrix for each setting, see Table 4, which summarizes the classification results from all classifiers. The individual confusion matrices for each classifier can be found in additional material [39]. These tables reveal

Table 3: Biopotential dataset, average accuracies, precisions and recalls for all ten deep classifiers in percent. We averaged over five iterations to minimize the influence of the random initialization of the weights. We present three different settings: two classes (no and wind stimulus); three classes (no, wind, and temperature stimulus); and five classes (no, wind, temperature, red light, and blue light stimulus).

	two classes			three classes			five classes		
classifier	accuracy [%]	precision [%]	recall [%]	accuracy [%]	precision [%]	recall [%]	accuracy [%]	precision [%]	recall [%]
MLP	73.2	73.3	73.2	68.6	68.7	68.6	57.4	50.8	47.8
TCNN	80.5	80.6	80.5	75.7	76.1	75.9	62.5	53.0	50.3
FCN	89.3	89.3	89.3	91.4	91.6	91.6	83.5	75.6	71.3
Encoder	86.6	86.6	86.6	89.6	89.8	89.8	80.0	74.5	67.0
t-LeNet	82.5	83.1	82.5	80.8	81.5	81.0	70.3	62.5	53.5
MCDCNN	78.4	79.6	78.4	78.7	80.1	79.0	67.5	59.0	55.6
ResNet	89.4	89.5	89.4	92.0	92.2	92.2	83.5	74.9	70.2
Inception	89.7	89.8	89.7	92.2	92.3	92.3	83.0	73.0	68.6
TWIESN	57.8	57.9	57.8	53.2	55.8	53.2	43.2	29.0	30.3

among which stimuli the classifiers have difficulties discriminating. The general trend in the three-class classification setting is that all classifiers have problems correctly discriminating between wind and no stimulus, while differentiating between heat and wind seems less difficult. The classifiers are best in distinguishing between heat and no stimulus. We see similar trends for the five-class classification task compared to the three-class setting for the stimuli of wind, heat, and no stimulus. Furthermore, we notice that the blue and red light stimuli are often misclassified as wind.

4.3. Discussion

With our experiments we have found that both the ZZ plant and tomato exhibits electrical responses to our applied stimuli of wind, heat, blue light, and red light. Both the changes in biopotential and impedance exhibit stimulus-specific characteristics that allow us to train classifiers that identify the respective stimulus with high accuracy. Although both methods, discriminant analysis and deep learning, achieve high accuracies, the discriminant analysis significantly outperforms the deep learning approach. One of the main differences between the two approaches is the preprocessing

Table 4: Biopotential dataset, the average confusion matrix over all classifiers for three different settings: wind and no stimulus (left); wind, heat, and no stimulus (middle); and wind, heat, red light, blue light and no stimulus (right). The different stimuli are: no is no stimulus, wind is wind stimulus, heat is heat stimulus, blue is blue light stimulus, and red is red light stimulus. The rows are the true values, which sum up to 163 for no stimulus, 163 for wind stimulus, 151 for heat stimulus, 40 for blue light stimulus, and 41 for red light stimulus. The columns are the assigned values by the classifiers.

stimuli	no	wind	no	wind	heat	no	wind	heat	blue	red
no	126.1	36.9	126.5	32.8	3.7	121.5	30.4	3.6	3.6	3.8
wind	29.2	133.8	25.5	125.0	12.5	28.7	118.5	12.9	1.5	1.4
heat			8.9	10.8	131.3	10.4	10.2	129.9	0.4	0.1
blue						8.3	13.1	1.6	10.6	6.3
red						8.7	14.1	2.0	5.5	10.7

of the data. In discriminant analysis, statistical properties are first calculated from the raw signals and then classified assuming a Gaussian distribution. In contrast, the raw data is only Z-score normalized before presented to the neural networks increasing the task difficulty probably considerably. Depending on their architecture, the networks directly classify the time series (ANN), autonomously learn a representation of the time series before classification (CNN), or randomly map the time series to another dimension before classification (ESN). The property of autonomously learning a representation of the data is advantageous because less a-priori knowledge and engineering for finding appropriate features is required. Therefore, the deep learning method can readily be adapted to other classification tasks without the overhead of reconsidering which features need to be selected.

For the binary classification task, the discriminant analysis achieves an average accuracy of 100% while the deep learning approach only achieves an average accuracy of 80.8%. On average, more samples of the no stimulus class (36.9) are incorrectly assigned than samples of the wind class (29.2).

Next, we discuss the three-class task. By adding the heat stimulus, we add 504 time series to the dataset. The dataset sizes are now unequal but can still be consider balanced. To achieve 100% accuracy, we need at least the two features. The deep learning approach achieves a lower average accuracy of 80.3%. Nevertheless, the accuracy increases for the increased training data for the FCN, Encoder, MCDCNN, ResNet, and Inception model the accuracy increases. This counterintuitive effect by the training data getting bigger while the number of trainable parameters in the classifiers increases less. The confusion matrices indicate that the classifiers struggle more to discriminate wind from no stimulus than wind from temperature. This may be caused

by the shapes of the different biopotential signals.

A total of 272 samples is added to our dataset by including the two remaining stimuli: red light (134 samples) and blue light (138). Now we have an unbalanced dataset. The discriminant analysis is largely unaffected by unbalanced datasets (red and blue light correctly classified). On average, only five samples of the blue light stimulus class and one sample of the red light stimulus class are misclassified. We again observe a decrease in average accuracy to 98.56%. The average accuracy of the deep learning approach decreases to 70.1%. Possibly it is affected by the unbalanced dataset as seen by the drop in precision (average 61.4%) and recall (average 57.2%). This is also confirmed by the confusion matrices for all classifiers. As mentioned in Sec. 4.2, the confusion matrix shows that most red and blue light stimuli are classified as wind stimulus. This causes the decrease in precision and recall (about 20% compared to the three-class task).

Another issue with the deep learning approach to TSC is the lack of sufficient training data. In our experiments, we obtained approximately 1300 time series as a training dataset which is probably not sufficient for networks with larger numbers of trainable weights. In previous work [12], the same dataset was used. However, the time window used for classification was slightly larger (about 11.5 min compared to our 9.8 min), including about 1.5 min from the poststimulus phase. The larger time window allowed for more information, including a part of the steep negative slope at the beginning of the poststimulus phase, see Fig. 3a, resulting in better performance (on average, 11.3% for two classes, 7.5% for three classes, and 6.5% for five classes). The results suggest that with more information, i.e., more samples or longer time windows, the performance of the classifiers can be increased. However, we shortened the time window length because we wanted to train classifiers that only rely on electrophysiological changes during the stimulus application period. An advantage of deep-learning approaches is that they do not require additional information for background subtraction.

In addition to the biopotential experiments, we also attempted training classifiers for tissue impedance to test whether it can also be a valid method for stimuli classification. We chose to train discriminant analysis classifiers because they showed better classification accuracies than the deep learning classifiers during our biopotential experiments. The average accuracy of all classifiers in combination with SFS is 96%. QDA and Mahalanobis achieve an accuracy of 100% with the features mean and WPE. This suggests that the tissue impedance data also encode information about externally imposed stimuli and is a promising candidate for further investigation.

5. Conclusion and future work

Our objective was to compare methods of discriminant analysis and deep learning for the application of classifying electrophysiological and impedance signals of natural plants. We have studied the efficiency and accuracy of these methods. We gathered the required

data in many plant experiments with *Zamioculcas zamiifolia* (ZZ plant) and *Solanum lycopersicum* (tomato) measuring the electrophysological and tissue impedance response for five and two different stimuli respectively in a controlled environment.

Using 1864 biopotential time series, we trained five discriminant analysis classifiers and ten classifiers based on deep learning. We have achieved high accuracies for many of them depending on the difficulty of the classification task. We trained the classifiers for three different tasks. The discriminant analysis classifiers outperformed the classifiers based on deep learning for all tasks. The first task was a binary classification between wind and no stimulus that allowed for accuracies of up to 100%. The second task was a three-class case with wind, heat, and no stimulus. Also for that task we achieved 100% accuracy. The third task includes all five stimuli including red and blue light, where the maximum accuracy was slightly reduced to 99.1%.

Based on the 40 impedance time series, we trained five discriminant analysis classifiers for a single binary classifcation task (light stimulus, no stimulus). With QDA and Mahalanobis we achieved accuracies of up to 100%.

Given these results and also comparing them to previous work [12] where good results with deep learning methods for a qualitatively different dataset were achieved, we conclude that the decision of what classification technique to use needs to be done case to case. Depending on the amount of available data and the feature selection, the simple statisitical methods can outperform the deep learning techniques. The comparison is arguably unfair because for the statistical approach we used feature-based inputs while the deep learning approach was required to operate on the raw data. With increasingly more data being available, the deep learning methods may increase their accuracy and outperform the statistical approach. Data collection in plant experiments is comparatively expensive and, hence, data availability is challenging. Methods of data augmentation for plant data may be a promising option. While here we have used simple stimuli that are readily available, we will study other options to test classification techniques and phytosensing for monitoring urban areas and especially air pollution. In our future work, we want to focus on applications for phytosensing within our project. While here we have used simple stimuli that are readily available, we will study other options to test classification techniques and phytosensing for monitoring urban areas and especially air pollution.

Acknowledgments

Removed for double-anonymous peer review process.

References

[1] Umberto Garlando, Lee Bar-On, Paolo Motto Ros, Alessandro Sanginario, Sebastian Peradotto, Yosi Shacham-Diamand, Adi Avni, Maurizio Martina, and Danilo Demarchi. Towards optimal green plant irrigation: Watering and body electrical impedance. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2020.

[2] Alexander G Volkov. *Plant electrophysiology: methods and cell electrophysiology*. Springer Science & Business Media, 2012.

[3] Shre Kumar Chatterjee, Saptarshi Das, Koushik Maharatna, Elisa Masi, Luisa Santopolo, Stefano Mancuso, and Andrea Vitaletti. Exploring strategies for classification of external stimuli using statistical features of the plant electrical response. *Journal of the Royal Society Interface*, 12(104):20141225, 2015.

[4] Won-Gyu Choi, Masatsugu Toyota, Su-Hwa Kim, Richard Hilleary, and Simon Gilroy. Salt stress-induced Ca^{2+} waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proceedings of the National Academy of Sciences*, 111(17):6497–6502, 2014.

[5] Edward E Farmer. Surface-to-air signals. *Nature*, 411(6839):854–856, 2001.

[6] Patricio Oyarce and Luis Gurovich. Electrical signals in avocado trees. *Plant Signaling & Behavior*, 5(1):34–41, 2010.

[7] Seyed A.R. Mousavi, Chi Tam Nguyen, Edward E Farmer, and Stephan Kellenberger. Measuring surface potential changes on leaves. *Nature Protocols*, 9(8):1997–2004, 2014.

[8] Ildikó Jócsák, György Végvári, and Eszter Vozáry. Electrical impedance measurement on plants: a review with some insights to other fields. *Theoretical and Experimental Plant Physiology*, 31(3):359–375, 2019.

[9] Xiaofei Yan, Zhongyi Wang, Lan Huang, Cheng Wang, Ruifeng Hou, Zhilong Xu, and Xiaojun Qiao. Research progress on electrical signals in higher plants. *Progress in Natural Science*, 19:531–541, 5 2009.

[10] Martin Anthony, Peter L Bartlett, Peter L Bartlett, et al. *Neural network learning: Theoretical foundations*, volume 9. cambridge university press Cambridge, 1999.

[11] Quan Huu Cap, Hiroyuki Uga, Satoshi Kagiwada, and Hitoshi Iyatomi. Leafgan: An effective data augmentation method for practical plant disease diagnosis. *IEEE Transactions on Automation Science and Engineering*, 19(2):1258–1267, 2022.

[12] Eduard Buss, Tim-Lucas Rabbel, Viktor Horvat, Marko Krizmancic, Stjepan Bogdan, Mostafa Wahby, and Heiko Hamann. Phytonodes for environmental monitoring: Stimulus classification based on natural plant signals in an interactive energy-efficient bio-hybrid system. In *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*, GoodIT'21, pages 258–264, 2022.

[13] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

[14] Laura García-Carmona, Stjepan Bogdan, Antonio Diaz-Espejo, Mikolaj Dobielewski, Heiko Hamann, Virginia Hernandez-Santana, Andreas Kernbach, Serge Kernbach, Alfredo Quijano-López, Niclas Roxhed, Babak Salamat, and Mostafa Wahby. Biohybrid systems for environmental intelligence on living plants: Watchplant project. In *Proceedings of the Conference on Information Technology for Social Good*, GoodIT'21, pages 210–215, 2021.

[15] Heiko Hamann, Stjepan Bogdan, Antonio Diaz-Espejo, Laura García-Carmona, Virginia Hernandez-Santana, Serge Kernbach, Andreas Kernbach, Alfredo Quijano-López, Babak Salamat, and Mostafa Wahby. Watchplant: Networked bio-hybrid systems for pollution monitoring of urban areas. In *ALIFE 2022: The 2022 Conference on Artificial Life*, volume ALIFE 2021: The 2021 Conference on Artificial Life, 07 2021. pp. 37.

[16] Marko Križmančić, Tim-Lucas Rabbel, Eduard Buss, Mostafa Wahby, Heiko Hamann, and Stjepan Bogdan. Distributed connectivity control in bio-hybrid wireless sensor networks. In *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*, GoodIT '22, pages 250–257, 2022.

[17] John Scott Burdon-Sanderson. I. note on the electrical phenomena which accompany irritation of the leaf of dionaea muscipula. *Proceedings of the Royal Society of London*, 21(139-147):495–496, 1873.

[18] Yoël Forterre, Jan Skotheim, Jacques Dumais, and Lakshminarayanan Mahadevan. How the venus

flytrap snaps. *Nature*, 433:421–5, 2005.

[19] Jin Hai Li, Li Feng Fan, Dong Jie Zhao, Qiao Zhou, Jie Peng Yao, Zhong Yi Wang, and Lan Huang. Plant electrical signals: A multidisciplinary challenge. *Journal of Plant Physiology*, 261:153418, 2021.

[20] David Roux, Alexandre Catrain, Sébastien Lallechere, and Jean Christophe Joly. Sunflower exposed to high-intensity microwave-frequency electromagnetic field: Electrophysiological response requires a mechanical injury to initiate. *Plant Signaling & Behavior*, 10:e972787–1–e972787–6, 2015.

[21] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern recognition*. Academic Press, 4 edition, 2009.

[22] Pimwadee Chaovalit, Aryya Gangopadhyay, George Karabatis, and Zhiyuan Chen. Discrete wavelet transform-based time series analysis and mining. *ACM Comput. Surv.*, 43(2), 2011.

[23] Elena Najdenovska, Fabien Dutoit, Daniel Tran, Carrol Plummer, Nigel Wallbridge, Cédric Camps, and Laura Elena Raileanu. Classification of plant electrophysiology signals for detection of spider mites infestation in tomatoes. *Applied Sciences*, 11(4), 2021.

[24] Gabriela Niemeyer Reissig, Thiago Francisco de Carvalho Oliveira, Ádrya Vanessa Lira Costa, André Geremia Parise, Danillo Roberto Pereira, and Gustavo Maia Souza. Machine learning for automatic classification of tomato ripening stages using electrophysiological recordings. *Frontiers in Sustainable Food Systems*, 5, 2021.

[25] Xiao-Huang Qin, Zi-Yang Wang, Jie-Peng Yao, Qiao Zhou, Peng-Fei Zhao, Zhong-Yi Wang, and Lan Huang. Using a one-dimensional convolutional neural network with a conditional generative adversarial network to classify plant electrical signals. *Computers and Electronics in Agriculture*, 174:105464, 2020.

[26] Danillo Roberto Pereira, João Paulo Papa, Gustavo Francisco Rosalin Saraiva, and Gustavo Maia Souza. Automatic classification of plant electrophysiological responses to environmental stimuli using machine learning and interval arithmetic. *Computers and Electronics in Agriculture*, 145:35–42, 2018.

[27] Leelawadee Thongkham and Lop Phavaphutanon. Effect of position and size of leaflets on rooting and rhizome formation of ZZ plant (Zamioculcas zamiifolia (lodd.) engl.) leaflet cuttings. *Agriculture and Natural Resources*, 52:246–249, 2018.

[28] Serge Kernbach. *Differential Impedance Spectrometer for electrochemical and electrophysiological analysis of fluids and organic tissues. Handbook and User Manual*. CYBRES GmbH, Stuttgart, 2022.

[29] J.E. Fernández, Francisco Alcon, Antonio Diaz-Espejo, Virginia Hernández-Santana, and María Cuevas. Water productivity and economic analyses for super high density olive orchards. *Acta Horticulturae*, pages 395–402, 2022.

[30] Serge Kernbach. Device for measuring the plant physiology and electrophysiology. *IJUS*, 12-13(4):138, 2016. Pre-print 10.48550/arXiv.2206.10459, 2022.

[31] Gen Sakurai and Stanley J. Miklavcic. On the efficacy of water transport in leaves. a coupled xylem-phloem model of water and solute transport. *Frontiers in Plant Science*, 12, 2021.

[32] Vilma Kisnieriene, Indre Lapeikaite, Vilmantas Pupkis, and Mary Jane Beilby. Modeling the action potential in characeae nitellopsis obtusa: Effect of saline stress. *Frontiers in Plant Science*, 10, 2019.

[33] Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä, and Ludolf Erwin Meester. *A Modern Introduction to Probability and Statistics*. Springer London, 2005.

[34] Bo Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1970.

[35] Bo Hjorth. Time domain descriptors and their relation to a particular model for generation of EEG activity. *CEAN-Computerized EEG Analysis*, pages 3–8, 1975.

[36] Dimitris Kugiumtzis and Alkiviadis Tsimpiris. Measures of analysis of time series (mats): A matlab toolkit for computation of multiple measures on time series data bases. *Journal of*

Statistical Software, 33(5):1–30, 2010.

[37] Steven L. Brunton and J. Nathan Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

[38] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[39] Anonymous. Stimuli Classification with Electrophysiology and Impedance of Natural Plants: Comparing Discriminant Analysis and Deep Learning Methods, October 2022.

[40] Licheng Jiao and Jin Zhao. A survey on the new generation of deep learning in image processing. *IEEE Access*, 7:172231–172263, 2019.

[41] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *IEEE Computational intelligence magazine*, 13(3):55–75, 2018.

[42] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[43] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(3):1–31, 2013.

[44] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, pages 1578–1585. IEEE, 2017.

[45] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[46] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[47] Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1):162–169, 2017.

[48] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[49] Joan Serrà, Santiago Pascual, and Alexandros Karatzoglou. Towards a universal neural network encoder for time series. In *CCIA*, pages 120–129, 2018.

[50] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[51] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[52] Arthur Le Guennec, Simon Malinowski, and Romain Tavenard. Data augmentation for time series classification using convolutional neural networks. In *ECML/PKDD workshop on advanced analytics and learning on temporal data*, 2016.

[53] Zhicheng Cui, Wenlin Chen, and Yixin Chen. Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995*, 2016.

[54] Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J Leon Zhao. Time series classification using multi-channels deep convolutional neural networks. In *International conference on web-age information management*, pages 298–310. Springer, 2014.

[55] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[56] Pattreeya Tanisaro and Gunther Heidemann. Time series classification using time warping invariant echo state networks. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 831–836. IEEE, 2016.

[57] Jerry L. Hintze and Ray D. Nelson. Violin plots: A box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.