

Article

Comparing Machine Learning Methods for Classifying Plant Drought Stress from Leaf Reflectance Spectra in *Arabidopsis thaliana*

Ana Barradas¹, Pedro M.P. Correia^{1,2} , Sara Silva³, Pedro Mariano⁴ , Margarida Calejo Pires¹, Ana Rita Matos^{1,2} , Anabela Bernardes da Silva^{1,2}  and Jorge Marques da Silva^{1,2,*} 

¹ BioISI—Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal; arbarradas@fc.ul.pt (A.B.); pmpcorreia@fc.ul.pt (P.M.P.C.); mmpires@fc.ul.pt (M.C.P.); armatos@fc.ul.pt (A.R.M.); arsilva@fc.ul.pt (A.B.d.S.)

² Departamento de Biologia Vegetal, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

³ LASIGE, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal; sgsilva@fc.ul.pt

⁴ Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico, 2695-066 Bobadela, Portugal; plmariano@fc.ul.pt

* Correspondence: jmlsilva@fc.ul.pt; Tel.: +351-2175-00000

Abstract: Plant breeders and plant physiologists are deeply committed to high throughput plant phenotyping for drought tolerance. A combination of artificial intelligence with reflectance spectroscopy was tested, as a non-invasive method, for the automatic classification of plant drought stress. *Arabidopsis thaliana* plants (ecotype Col-0) were subjected to different levels of slowly imposed dehydration (S0, control; S1, moderate stress; S2, severe stress). The reflectance spectra of fully expanded leaves were recorded with an Ocean Optics USB4000 spectrometer and the soil water content (SWC, %) of each pot was determined. The entire data set of the reflectance spectra (intensity vs. wavelength) was given to different machine learning (ML) algorithms, namely decision trees, random forests and extreme gradient boosting. The performance of different methods in classifying the plants in one of the three drought stress classes (S0, S1 and S2) was measured and compared. All algorithms produced very high evaluation scores (F1 > 90%) and agree on the features with the highest discriminative power (reflectance at ~670 nm). Random forests was the best performing method and the most robust to random sampling of training data, with an average F1-score of 0.96 ± 0.05 . This classification method is a promising tool to detect plant physiological responses to drought using high-throughput pipelines.

Keywords: *Arabidopsis thaliana*; water stress; soil water content; reflectance spectra; plant phenotyping; artificial intelligence; machine learning; decision trees; random forests; extreme gradient boosting



Citation: Barradas, A.; Correia, P.M.P.; Silva, S.; Mariano, P.; Pires, M.C.; Matos, A.R.; da Silva, A.B.; Marques da Silva, J. Comparing Machine Learning Methods for Classifying Plant Drought Stress from Leaf Reflectance Spectra in *Arabidopsis thaliana*. *Appl. Sci.* **2021**, *11*, 6392. <https://doi.org/10.3390/app11146392>

Academic Editor: Giuseppe Perna

Received: 8 May 2021

Accepted: 4 July 2021

Published: 11 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Climate change, scarcity of resources and growing world population increasingly jeopardize global food and nutritional security [1]. To respond to this challenge, we need to accelerate plant breeding, provide new cultivars adapted to the changing environmental conditions, and optimize agricultural practices, decreasing the ecological footprint of food production [2]. The emergence of high throughput plant phenotyping and precision agriculture as key technologies to foster plant breeding and sustainable food production [3,4] requires the development of non-invasive techniques enacted to collect structural and functional information from plants [5]. Additionally, the collected information (images and/or spectra) must be processed by automated high-throughput pipelines, and translated to biologically relevant information, including their water status, since it is expected to be critically affected by recurrent drought episodes [6]. This challenge has renewed the interest in the application of in vivo spectroscopic techniques in plant sciences, including reflectance spectroscopy. In fact, the use of reflectance spectroscopy to infer the constitutive

and functional characteristics of plant tissues, from leaves [7] to fruits [8], is well established. Since the pioneering works that introduced reflectance spectroscopy in plant sciences [9] a multitude of reflectance indexes have been suggested to represent specific features of the plant material. Among the currently most used indexes, the Normalized Difference Vegetation Index (NDVI) [10] is related to the biochemical composition and structure of the plant tissue analysed, being used as an indicator of biomass, since it is strongly correlated with the amount of chlorophyll [11]. The Photosynthetic Reflectance Index (PRI) [12], in turn, is seen as a functional indicator, as it is related to the epoxidation state of the photosynthetic antenna xanthophylls [13], a very dynamic process connected with instant plant photosynthetic efficiency [14]. Reflectance indexes were developed to extract relevant information from reflectance spectra. However, increasing computational power and the developments in machine learning (ML) enacted the use of the entire data set of information present on reflectance spectra.

Spectral reflectance within the visible region (400 to 750 nm) derives from the absorption of light by the leaf pigments (mostly chlorophylls and carotenoids), which absorb distinct wavelength bands. This selective absorbance creates reflectance signatures that depict numerous physiological processes in plants, such as senescence [15], nutrition status [16], and stress [17], among others, therefore holding important phenotypic information. The ability to use these large collections of reflectance-based information increases the chances of finding meaningful signals which contribute to a better interpretation of the data, allowing for the development of more accurate prediction models. To date, several machine learning algorithms have been used to analyze and understand plant spectral data [18–20], however, there is no standard consensus on the best technique to use as this depends on a number of factors, such as the type of problem, type and size of the data, and type of output, for example. Therefore, a comparative approach is a conventional method to assess an algorithm's suitability to solve a particular problem.

In our study, we are interested in distinguishing between water-stressed and unstressed plants, and also between the level of stress (e.g., moderate or severe). Decision tree learning methods are a class of supervised learning that work well on simple and discrete classification [21]. Classification trees act like functions that map the input features to the different classes [22], and an individual decision tree is the simplest model of this type. Multiple estimators can also be built under ensemble tree-based methods, called bagging [23] and boosting [24], of which random forests [25] and extreme gradient boosting (XGBoost) [26] are the most typical examples, where multiple decision trees improve over the predictive capacity of single trees. In this work, we explore the ability of these three machine learning algorithms to classify the water stress of the model plant *Arabidopsis thaliana*, based on leaf reflectance spectra. Beginning with the simplest method that returns interpretable models (DT), we then move to a more complex method which implies the setting of more parameters, returning models that are not interpretable any longer (RF), and we finally test another complex method that requires much more computational time (XGBoost).

2. Materials and Methods

2.1. Plant Growth Conditions and Water Stress Imposition

A. thaliana ecotype Columbia (Col-0) seeds were surface sterilized and stratified for three days at 4 °C. Plants were grown in a controlled environment chamber (Fitoclima 5000 EH, Aralab, Rio de Mouro, Portugal) in 200 cm³ pots containing the same amount of a mixture of horticultural substrate (Compo Sana Universal, Compo Sana) and vermiculite (1:1), placed in trays. The light, an average photosynthetic photon flux density (PPFD) of 180 $\mu\text{mol m}^{-2} \text{s}^{-1}$, was provided by fluorescent lamps (Osram Lumilux L 58W/840 cool white lamps) with a photoperiod of 14 h light/10 h dark, under controlled temperature (23/18 °C day/night) and 50% relative humidity. Watering was provided by sub-irrigation, and the trays were frequently rotated and displaced in the chamber. Twenty-one days after sowing (DAS) plants with similar sizes (one per pot) were randomly assigned to

well-watered controls (SWC ~100%) (S0) and progressive drought stress by withholding watering, reaching $46.10 \pm 3.29\%$ soil water content after 7 days (S1) and $33.17 \pm 2.50\%$ after 12 days (S2) of water stress, adapted from the protocol in [27]. Soil water content (SWC) was determined by weighing the pots at 21 DAS (S0), 28 DAS (S1) and 33 DAS (S2) and calculated according to (1):

$$\text{SWC} = (\text{SFW}/\text{SFC}) \times 100, \quad (1)$$

where SWC is the soil water content expressed as the ratio between the soil fresh weight (SFW, pot weight at the sampling time) and the soil weight at field capacity (SFC, pot weight before the imposition of the stress). The statistical significance of SWC variation was tested by a Kruskal–Wallis test ($p < 0.05$), and multiple pairwise-comparison between groups tested by a pairwise Wilcoxon test with corrections for multiple testing by the Benjamini–Hochberg method ($p < 0.05$).

2.2. Spectrophotometric Determination of Leaf Pigments

At the end of the experiment, photosynthetic pigments were extracted from leaf discs with ethanol, 95% (v/v), in the dark, at 4 °C, for 24 h. To quantify differences in anthocyanin absorption, HCl 1% (v/v) was added to individual photosynthetic pigment extracts after extraction. Spectrophotometric absorption measurements were taken from 400 nm to 700 nm (Helios Beta UV/VIS Spectrometer, Thermo Scientific, Waltham, MA, USA). Chlorophyll a (Chl.a), chlorophyll b (Chl.b), total chlorophyll (Chl.a+b) and total carotenoids (C.x+c) were quantified according to Lichtenthaler [28]:

$$\text{Chl.a} = 13.36A_{664} - 5.19A_{648} \quad (2)$$

$$\text{Chl.b} = 27.43A_{648} - 8.12A_{664} \quad (3)$$

$$\text{Chl.a+b} = 5.24A_{664} + 22.24A_{648} \quad (4)$$

$$\text{C.x+c} = (1000A_{470} - 2.13\text{Chla} - 97.64\text{Chlb})/209 \quad (5)$$

The statistical significance of pigments content variation between stress conditions was tested by a Mann–Whitney U test ($p < 0.05$).

2.3. Spectral Reflectance Measurements

In order to obtain the reflectance spectra of the leaves we started with the characterization of the illumination, acquiring the source emission spectrum:

Set-up (Figure 1):

1. All the spectra were acquired in a dark chamber, to avoid the influence of ambient light
2. Four light sources were fixed in the four top corners of the chamber, providing uniform illumination of the sample area: Two LEDs Lexman14/100W warm white and two LEDs Derlight 6/28W Blue/Red, covering the visible and near-infrared spectral bands (400–750 nm) without significant temperature increase on the test chamber.
3. The spectra were obtained after the lamp's warming-up time.

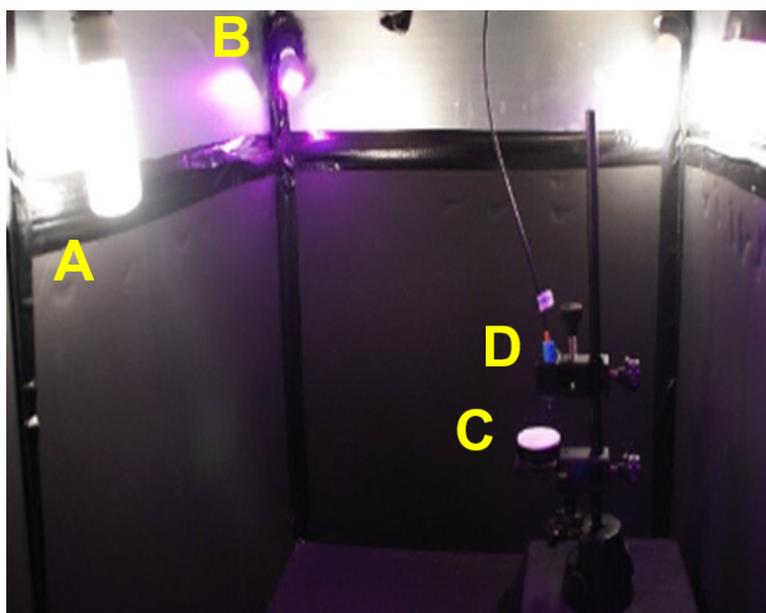


Figure 1. Set-up for spectra acquisition: (A) white LED; (B) blue/red LED; (C) white diffuse standard; (D) fiber tip.

Before the acquisition of the spectra reflected by the plant leaves, for each plant batch, the spectrum of 2 diffuse reflectance standards was acquired. The spectrum of the dark standard (SRS-02, Labsphere Inc., North Sutton, NH, USA) was used as indicative of the noise in all the acquired spectra, such as ambient light in the chamber or signal noise. The spectrum of the white standard (SRS-099, Labsphere Inc.) was used as indicative of the lamp spectra and allowed for the normalization of the plant spectra.

For the acquisition of the reflectance spectra, either for the diffuse reflectance standards, or for the plant leaves, the following steps were taken:

Procedure:

1. The lamps were turned on and allowed to stabilize (reduced warm up time for LED light sources).
2. The surface of the standard to be analyzed was positioned in the test chamber, facing the light sources.
3. The collecting tip of the optical fibre of the spectrometer was directed normal to the diffusing surface, at a fixed distance.
4. A sample of the light reflected by the standard diffusing surface was collected by the fibre in a quasi-normal geometry.
5. The raw spectrum of the collected light was processed with the software Ocean Optics SpectraSuite; with the following parameters: integration time: 500 msec; average: 1; smoothing: 5; gain level: 1. It was saved as an ASCII data set of wavelength vs. intensity.
6. After the standard spectra acquisition, one plant pot at a time was taken to the chamber, replacing this standard, in a way that the surface of the leaves, approximately flat, kept the same distance to the optical fibre tip.
7. For each plant, spectra were acquired with the following parameters: integration time: 1000 msec; average: 1; smoothing: 5; gain level: 1

This procedure was applied to the control plants batch (S0) and to the plants with moderate (S1) and severe stress (S2) caused by withholding watering for 7 days (S1, measured at 28 DAS) and 12 days (S2, measured at 33 DAS). Plant spectra data were normalized to the white standard (SRS-099 spectra), as in (6):

$$R(\lambda) = [I(\lambda) - D(\lambda)]/[W(\lambda) - D(\lambda)], \quad (6)$$

where $R(\lambda)$ is the sampled plant reflectance at each wavelength of the spectra, $I(\lambda)$ is the intensity of the collected light reflected by the sample, for each wavelength, and $D(\lambda)$ and $W(\lambda)$ are the intensity of the light reflected by the dark or white standard, respectively, for each wavelength. These parameters were computed for the same acquisition conditions.

2.4. Data Set

Each instance, or observation, in the data set consists of 1729 readings of radiation reflection in the wavelength range of 400–750 nm taken at 0.21–0.22 nm intervals (predictor variables) describing 179 *A. thaliana* plants grouped into the ground truth categories (dependent variables). The class distribution is as follows: class S0—120 observations (all the plants well-watered, from 21, 28 and 33 DAS); class S1—29 observations (water deficit for seven days, from 28 DAS); and class S2—30 observations (water deficit for 12 days, from 33 DAS). All the reflectance values were used as input features to the machine learning algorithms, with no feature selection.

2.5. Machine Learning Algorithms

In order to perform the classification task, i.e., to discriminate between the three types of plant water stress status (S0, S1, and S2), a set of tree-based supervised machine learning methods was used to construct the predictive models: decision trees (DT), random forests (RF), and extreme gradient boosting or XGBoost (XGB). Tree-based classification models are widely used learning methods with good predictive power and the ability to provide insights into the underlying structure of the data [29]. Tree-building algorithms grow classification trees by recursive binary splitting of the features into subsets with similar response values, represented by nodes, which are further partitioned into more subsets until no more splits are possible or a stop condition is reached. Ultimately, the most frequent response class in each terminal node of the tree is derived from all observations in that same node to predict a categorical outcome. The DT algorithm is one of the simplest tree-structured classifiers, which outputs a single mapping of binary decisions [22]. Single trees are easy to interpret and suitable for a wide range of classification problems involving non-linear decision making. However, due to the greedy nature of their learning process, tree structure may become unstable with minimal data set changes.

Tree instability can be mitigated by employing ensemble methods, such as RF [25] and XGB [30]. Ensemble techniques are able to achieve higher predictive power than single trees produced by DT, but their classification models are harder to interpret. Ensemble learning combines several classifiers and makes predictions by aggregating opinions over the ensemble of models [31,32], creating a stronger learner with smaller variance. In the case of RF, each tree of the ensemble is built on only a subset of the data, both in terms of observations and in terms of features. When making predictions on unseen data, RF combines the predictions of all the trees in the ensemble, normally as a simple majority voting. In the case of XGB, the trees that compose the model are added sequentially, where each tree attempts at correcting the mistakes made by the previous ones. The trees are attributed weights, optimized by gradient descent to minimize the overall error of the combined predictions. XGB uses regularization to reduce overfitting.

2.6. Parameter Estimation

All computations were performed in Python 3.8.5 using the Scikit Learn library [33], including the Scikit Learn wrapper interface for XGBoost, XGBClassifier. Due to the imbalanced nature of the data set, a stratified k-fold cross-validation approach was used for the search, to ensure that the training and test sets preserved the original proportion of samples in each class. The same seeded partition, consisting of 10 folds, was used for all machine learning algorithms. Hyperparameters for each method were chosen by testing several parameter combinations using the GridSearchCV module and adopting the combination that gave the best results on the test data, as returned by the 'best_params'

attribute. For the most computationally expensive algorithm, XGB, parameters were searched and fine-tuned in a stepwise manner by varying one or two parameters at a time.

2.7. Model Robustness Evaluation

To assess the spread of the dispersion of the predictions for each final classifier, we repeated the data splitting and model fitting 1000 times by varying the random state inside the cross-validator and the classifier. Different evaluation metrics, including balanced accuracy and F1, were used to assess model performance across model fit iterations, and an overall estimate of each metric was obtained by averaging over all predictions made on each test set. For DT, one example tree that achieved median performance according to the evaluation metrics was chosen for illustration purposes.

As stated in [34], “Any modeling decisions based upon experiments on the training set, even cross validation estimates, are suspect, until independently verified.” Therefore, a hold-out data set was removed from the main data set even before parameter estimation, consisting of approximately 10% of observations chosen in a random stratified manner. Kept aside until this point, it was then used to provide an external evaluation of each final classifier, only to make sure that the obtained scores were within the estimate obtained from the 1000 trials.

2.8. Feature Importance

In order to gain insight into the data and the focus given by each machine learning algorithm to the different features, for each of the one thousand trials we computed the relevance and the frequency of the variables employed by each model. Feature importance was obtained with Scikit Learn’s ‘feature_importances_’ property, which returns the relative importance scores for each input variable. For DT, this is calculated as the improvement in performance in a node, weighted by the proportion of samples in that same node, whereas for the ensemble methods RF and XGB, the feature importance is averaged across all trees within the model. These scores represent the usefulness of each variable in predicting each class label and provide information on their function in the predictive model. For each feature with a relative importance score different from zero, we also counted its frequency across the trials and represented their distribution as histograms.

3. Results

The visual appearance of well-watered controls (S0) and progressively drought stressed plants by withholding watering for 7 days (S1) and 12 days (S2) is shown in Figure 2a. Moderate changes were visible at S1 and much more pronounced changes at S2. Well-watered pots (S0) presented SWC ~100%, reaching $46.10 \pm 3.29\%$ soil water content after 7 days (S1) and $33.17 \pm 2.50\%$ after 12 days (S2) of water stress (Figure 2b). Significant differences in leaf pigment composition were found between S0 and S2 (Figure 2c). Water stress for 12 days (S2) decreased significantly the content of Chl.a, from 30.57 ± 2.07 to $23.01 \pm 2.11 \mu\text{g cm}^{-2}$ and Chl.b, from 5.38 ± 0.41 to $4.34 \pm 0.49 \mu\text{g cm}^{-2}$ ($p < 0.05$). as well as the ratio between Chl.a and Chl.b (from 5.69 ± 0.12 to 5.33 ± 0.15) and between Chl.a+b and C.x+c (from 4.99 ± 0.13 and 4.22 ± 0.12) ($p < 0.05$). The absorption spectra of photosynthetic pigments and anthocyanins are shown in Figure 2d. Total photosynthetic pigments show a peak in the blue region of the spectrum (~430–460 nm) and a peak on the red region of the spectrum (650–670 nm). A shoulder is observed centered at the wavelength of 470 nm. Anthocyanins show a significant absorption band between 400 and 450 nm. A significant overlap is visible in Figure 2e, showing leaf reflectance spectra measured in S0, S1 and S2 plants.

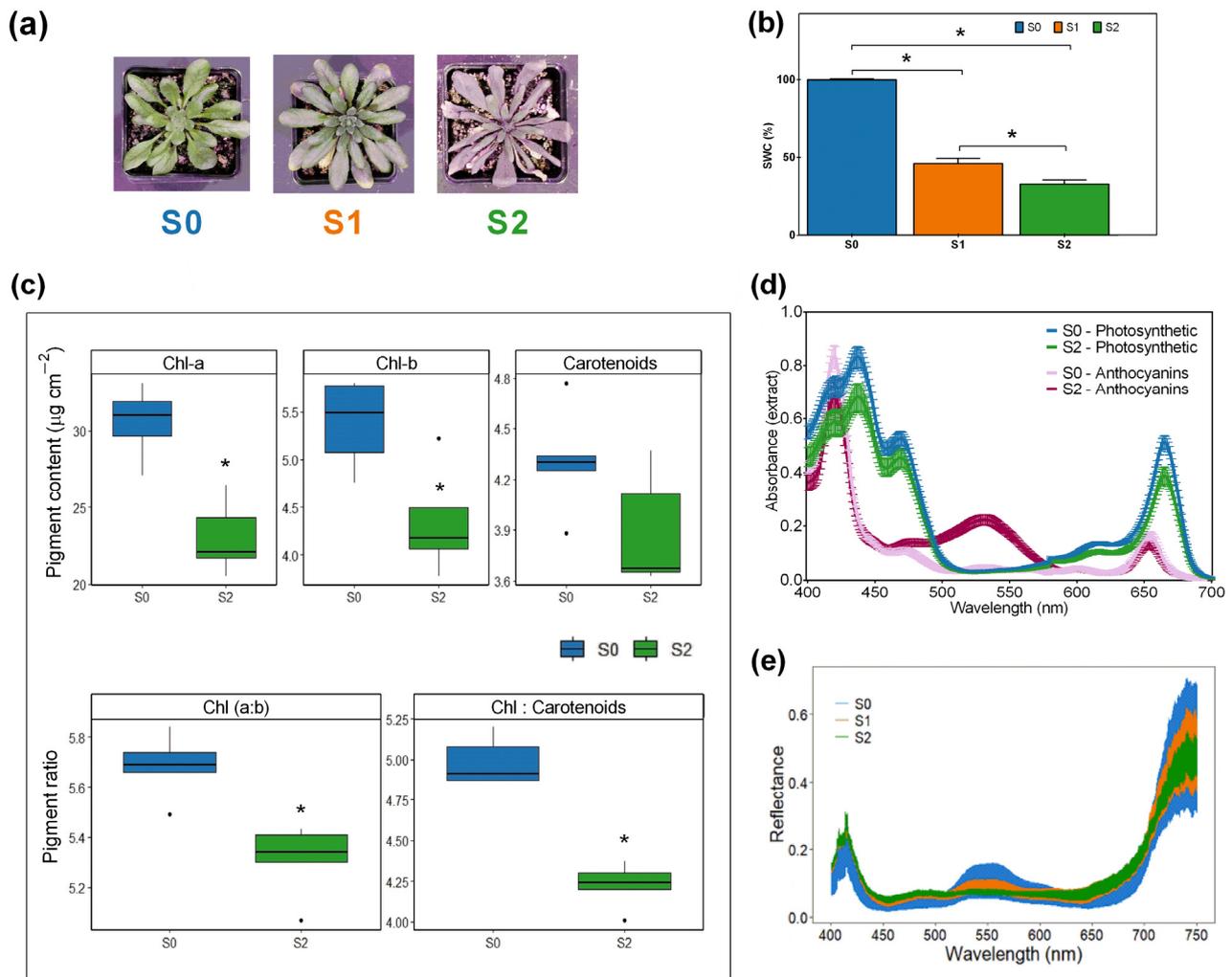


Figure 2. (a) Representative images of *A. thaliana* ecotype Columbia (Col-0) plants under control (S0), moderate stress (S1) and severe stress (S2); (b) soil water content at S0, S1 and S2 conditions (mean \pm S.D.; asterisks represent significant differences ($p < 0.05$)); (c) leaf pigment composition per leaf area (chlorophyll a (Chl.a), chlorophyll b (Chl.b), Chl.a/Chl.b, total carotenoids (Cx+c) and Chl.a+b/Cx+c (box plots show median values and interquartile range (IQR), asterisks represent significant differences ($p < 0.05$)); (d) in vitro absorbance spectra of total photosynthetic pigments and anthocyanins; (e) in vivo plant reflectance curves at each wavelength of the spectra at S0, S1, S2.

To establish machine learning classifiers with high predictive capability of water stress in *A. thaliana*, models from three algorithms (DT, RF, and XGB) were developed with hyperparameter search and tuning based on stratified 10-fold cross-validation. The best parameter combinations found for each classifier on the development data sets are listed in Table 1.

Table 2 shows the predictive performance of each estimator as a result of the randomly generated independent runs, each using different train/test data splits and classifier seeds. Figure 3 illustrates the dispersion of these results in a boxplot. As evaluation metrics, we have used both training and test accuracy, and also the balanced test accuracy and the test F1-score. Balanced accuracy represents the raw accuracy where each sample is weighted according to the inverse prevalence of its true class.

Table 1. Estimated parameters obtained by grid search for each machine learning algorithm.

Model Parameters	ML Algorithm		
	Decision Trees	Random Forests	XGBoost
Criterion to measure the quality of a split	entropy	entropy	-
Maximum depth of the tree	3	5	4
Minimum number of samples required to be at a leaf node	1	1	-
Minimum number of samples required to split an internal node	2	2	-
Number of features to consider when looking for the best split	-	log2	-
Number of trees in the forest	-	20	-
Number of boosting stages	-	-	50
Learning rate	-	-	0.1
Minimum sum of instance weight (hessian) needed in a child	-	-	0
Minimum loss reduction required to make a further partition on a leaf node of the tree	-	-	0
Fraction of samples to be used for fitting the individual base learners	-	-	0.8
Subsample ratio of columns when constructing each tree	-	-	0.55

Table 2. Evaluation metrics and scores for each learning algorithm. Values are mean ± standard deviation. Coincidentally, test F1 scores are the same as test accuracy scores.

Learning Algorithm	Training Accuracy	Test Accuracy	Balanced Test Accuracy	Test F1-Score
Decision trees	0.99 ± 0.005	0.93 ± 0.06	0.88 ± 0.12	0.93 ± 0.06
Random forests	1.00 ± 0.002	0.96 ± 0.05	0.94 ± 0.09	0.96 ± 0.05
XGBoost	1.00 ± 0.0	0.96 ± 0.05	0.92 ± 0.10	0.96 ± 0.05

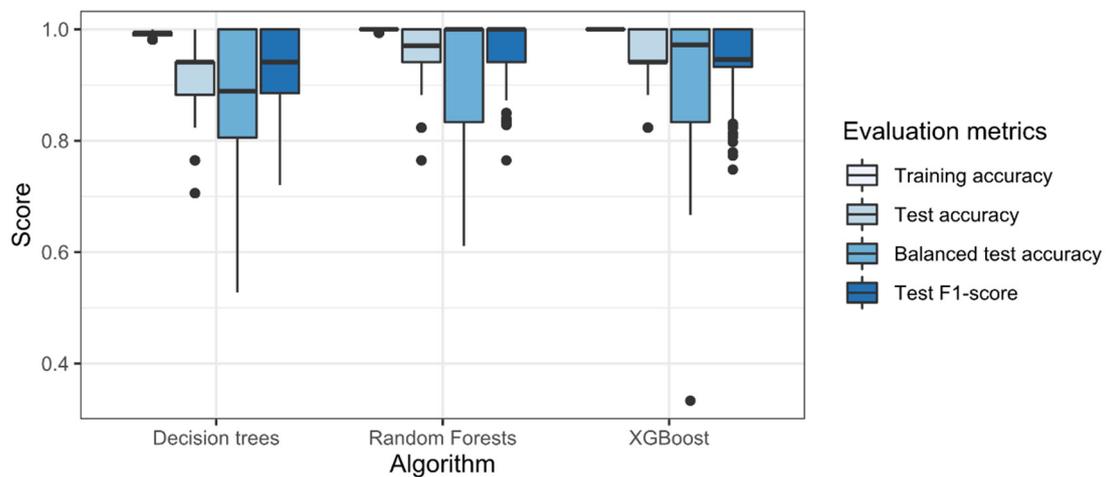


Figure 3. Boxplots of classification evaluation metrics.

All algorithms were able to induce models that achieve high average scores, but the ensemble methods performed better than single trees. DT attained not only the lowest average scores, but also the highest variance (balanced test accuracy 0.88 ± 0.12 , test F1 0.93 ± 0.06), while RF and XGB attained balanced test accuracies of 0.94 ± 0.09 and 0.92 ± 0.10 , respectively, and test F1-score of 0.96 ± 0.05 for both. The lower dispersion of RF and XGB results suggests that these methods have a higher robustness to random selections of features, samples and splits of the data. The boxplot reveals that, in terms of median, RF is the clear winner. Overall, RF reached the best performance with the lowest variance, proving to be a good choice for the present data set.

We have applied the Wilcoxon rank sum statistical test with significance threshold at 0.05 and Bonferroni–Holm correction for multiple testing. All the pairwise comparisons between the three ML methods revealed statistically significant differences, regardless of the evaluation metric considered.

In addition to the evaluation scores, model behavior can also be assessed in more detail through confusion matrices. These provide more information regarding class attribution of each sample by the classifier and report on the frequency of the misclassifications.

Figure 4 shows the confusion matrices computed for the selected models in each algorithm. In accordance with the model evaluation results, RF and DT are the algorithms with the lowest (6%) and highest (12%) misclassification rates per class, respectively. The matrices indicate that all algorithms are able to correctly distinguish severely water-stressed plants (S2) from the well-watered controls (S0) but moderately-stressed plants (S1) are mislabeled as controls (and vice versa) more often than with S2. In comparison, S2 is more often misclassified as S1, especially in DT.

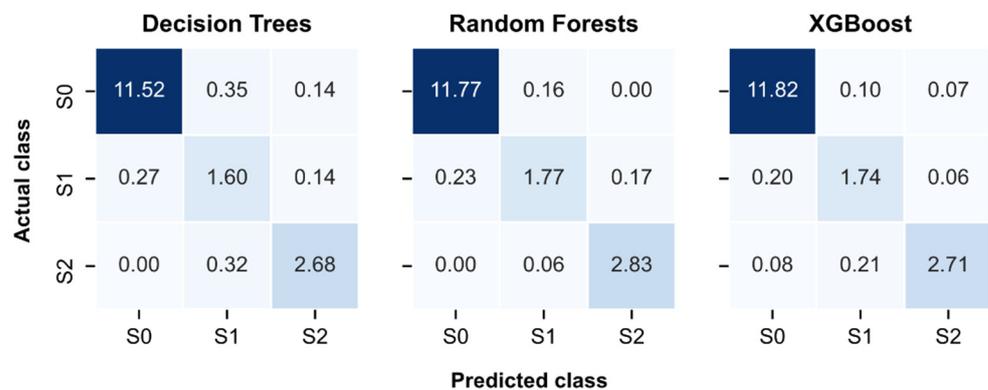


Figure 4. Average confusion matrices obtained from all test folds across multiple runs of the best cross-validated model with one thousand different seeds in each machine learning algorithm. The blue colors represent a heatmap of the predicted class against the actual class in the confusion matrices, where darker shades indicate larger values and lighter shades represent smaller values.

Due to their simplicity and ease of interpretation, in Figure 5 we illustrate a decision tree of a DT classifier, obtained from the model random runs. The tree shows the binary decisions at each split as if/then rules, based on the features chosen by the model in order to place subsets of the data in different nodes, according to their class label.

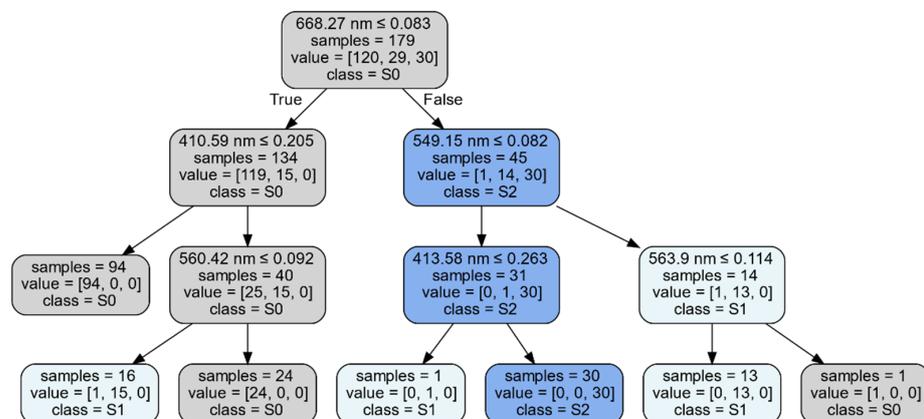


Figure 5. Decision tree to classify the SWC in *A. thaliana* representing a median-ranked classifier according to the evaluation metrics. The features for the decision rules are expressed as wavelengths and the colors indicate the different water stress levels (S0, S1, S2) in the plants. The numbers of samples falling on each node are based on the entire dataset.

For DT and RF, the samples tend to be first split based on the longest wavelengths (at around 670 nm) and, according to the illustrative tree, this feature is used to differentiate between the most extreme *A. thaliana* classes, i.e., the well-watered (S0) and the severely water-stressed (S2) conditions (Figure 4). As for XGB, samples tend to be first split on the wavelengths around 400 nm globally, but assessing the generation of separate trees for each class in the XGB algorithm, it is possible to see that first splits for S0 fall within the 600 nm region; for class S1 the region around 400 nm is preferably used; and for class S2 first splits between the 400 and 600 nm regions are evenly distributed (data not shown). It is also within class S2 that we find the highest number of decision stumps, indicating that the majority of predictions for this class are based on a single splitting variable involving spectral reflectance values at either the violet or the orange spectral colors.

The importance of the root node as a good predictor is confirmed by the highest importance ranking plots (Figure 6), where variables comprising reflectance at wavelengths between 663.34 and 677.51 nm are the top features selected for all algorithms. It is also possible to see that DT explore a narrower feature space within reflectance values at wavelengths around 600 nm, and therefore their most important features include more reflectance values at around 400 or 500 nm. Conversely, RF and XGB are able to use a broader feature space around 600 nm, and therefore attribute less importance to the reflectance values at around 400 and 500 nm.

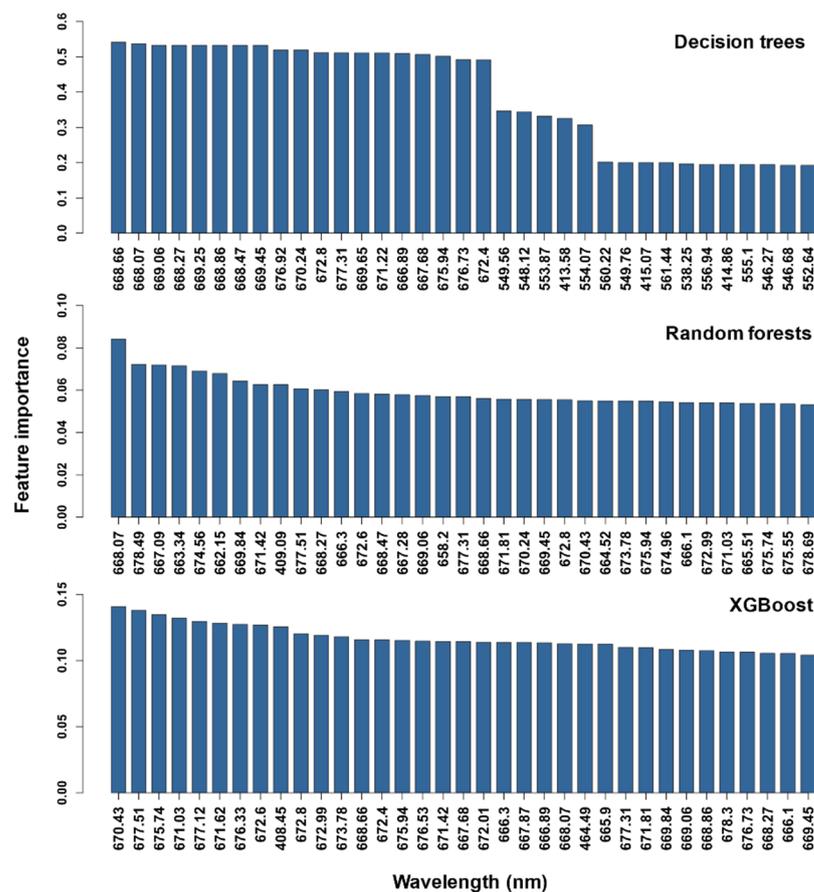


Figure 6. Features with the highest predictive importance for each machine learning algorithm.

Interestingly, we find that the most important features do not correspond to the features that each algorithm uses the most (Figure 7). For example, all trees produced by the DT algorithm in the model evaluation step have used features around 600 nm for the first split, however, features within 400–425 nm and 525–575 nm ranges have been used the most as decision rules throughout the trial. A similar trend is seen for XGB, whereas for RF we observe an incredibly high use of variables around 550 nm which were hardly

ever used as first splits (around 1%) in the model evaluation experiment. This observation may be related with the ability to explore more decision rules as the trees get bigger, since RF has the highest maximum depth in the optimized parameters ($\text{max_depth} = 5$), when compared to the other algorithms.

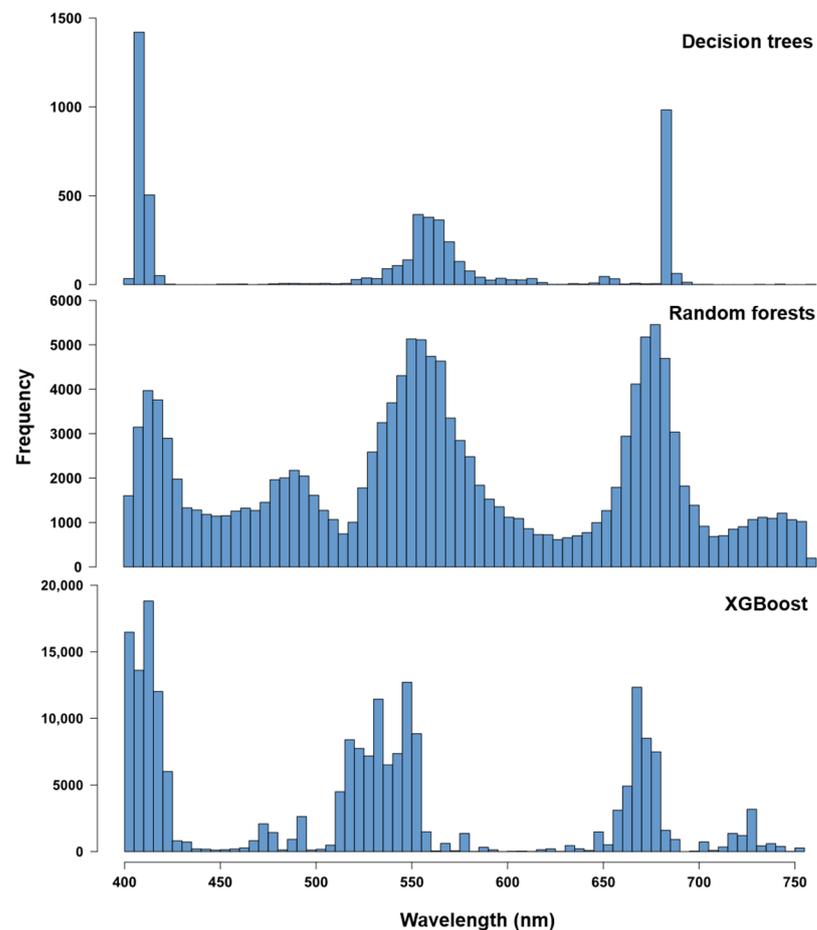


Figure 7. Histograms representing the most used features for each machine learning algorithm.

4. Discussion

The early selection of relevant plant phenotypes is key to the advance of plant breeding under a rapidly changing environment. We induced drought stress in *A. thaliana* plants by withholding watering for 7 days (S1, moderate stress) and 12 days (S2, severe stress), leading to differences in plants' aspect (Figure 2a) and SWC ($46.10 \pm 3.29\%$ and $33.17 \pm 2.50\%$, respectively) (Figure 2b). To classify *A. thaliana* drought stress severity we developed three machine learning models using leaf reflectance spectra. Even though a substantial overlap of leaf reflectance spectra was observed (Figure 2), all classifiers produced very high evaluation scores ($F1 > 90\%$) and we assessed their relative performance through stochastic variation.

When passing a thousand random seeds to each model's split function (which separates the train and test data) and the classifier's internal random number generator, all the scores showed a tendency to decrease. As expected, DT showed the most pronounced drop, owing to the algorithm's unstable response to small variations in the training set [35], while RF and XGB appeared to be more robust to this kind of random fluctuation as their prediction scores remained above 90% (Table 2), with RF displaying higher medians and less variance (Figure 3). It may be surprising that XGB, being known for its extreme ability to fit the data, did not surpass RF. However, XGB is also known for its tendency to overfit, which has to be counteracted by a careful setting of its parameters. For difficult data, it is

normally worth using XGBoost, but for easy data (such as ours apparently is, considering the quality of the achieved results) there is no advantage, and a more straightforward ensemble such as random forest tends to generalize better.

The models were able to adequately distinguish between unstressed and water stressed plants, however, unstressed plants were more often incorrectly classified as moderately stressed by DT. Discriminating between plants exposed to moderate and severe stress should be considered a harder task for the classifier given that although statistically different, these groups contain a slight overlap in their SWC levels (at around 39% SWC) (Figure 2b). However, moderately stressed plants were incorrectly classified as unstressed plants most of the time in all algorithms, in spite of severely stressed plants being indeed much more mistaken with moderately stressed ones (Figure 4). An important factor to consider in these predictions is the biological aging of the plants, which may introduce confounding effects related with pigment changes during senescence, which will affect leaf reflectance values [36]. In fact, unstressed plants in our experimental setup include individuals with different DAS (21, 28, and 33), which may explain the moderate stress misclassification as unstressed.

By virtue of its simplicity, DT was able to use the fewest number of features, and it was no surprise that the ensemble techniques (RF and XGB) were able to explore more features within the visible spectrum (Figure 7). From the initial high dimension feature space, we observe that all algorithms based their predictions on three main wavelength ranges within the reflectance spectrum, falling within the violet (400–425 nm), green (525–570 nm), and red (650 nm) regions of visible light. In addition, RF also considered a blue-green transition at around 475–500 nm. This narrowing of the initial feature space is useful in the development of future predictors as it will potentially contribute to a more reliable classification with a lesser chance of overfitting, given that noisy information will no longer be input to the model. We also found that reflectance values at wavelengths around 670 nm were particularly important, as they were observed to have very high discriminative power in all three algorithms (Figure 6). In fact, this particular wavelength has been shown to be the most responsive to different kinds of stress in diverse plant species [37]. Besides this specific wavelength, we also found that reflectance values based on wavelengths around 400 nm are also very selective, indicating a possible relation with water stress in particular.

Albeit other spectroscopic techniques, such as Raman spectroscopy, are able to provide detailed information about chemical structure of materials, due to its simplicity, reflectance spectroscopy is generally used in biological materials [38]. Reflectance depends on leaf surface properties and internal structure and the concentration and distribution of biochemical components, and, in the visible domain, reflectance depends mainly on the presence of photosynthetic pigments and defense chromophores [38]. Water also deeply influences the reflectance spectra of leaves, but mostly in the shortwave infrared region of the electromagnetic spectrum ([39,40] and references therein). Several studies indicate that leaf reflectance alteration by stress, at the 400–750 nm wavelength range, is related to changes in chlorophyll and anthocyanin concentrations [41,42]. Photosynthetic pigments composition (chlorophyll and carotenoids) showed significant changes in drought stressed *Arabidopsis thaliana* (Figure 2c). Also the absorption spectra of photosynthetic pigments and anthocyanins presented differences between S0 and S2 plants (Figure 2d). Therefore, in this work, the different reflectance bands that presented significant discriminant power may be attributed to different biological chromophores.

The increase of reflectance at wavelengths around 670 nm and higher is probably due to the sharp decrease of chlorophyll a and b absorbance in that spectral band (Figure 2c). Porra et al. [43] reported maximum absorption peaks of Chl. a dissolved in diethyl ether of 660.8 and 429.0 nm (Soret), and 643.0 nm and 454.0 nm (Soret) for Chl.b. Lichtenthaler [28] reported maximum absorption peaks of Chl. a dissolved in ethanol (95%) at 664 and 431 nm (Soret) and 648 and 463 nm (Soret) for Ch. b. In our work, photosynthetic pigments were also dissolved in ethanol 95% (v/v) and the absorption peaks of total pigments extract were at 437 nm (Soret) and 665 nm (Figure 2d). However, in vivo chlorophylls are

bound to an ensemble of different proteins to form the structurally complex photosystems, and the spectral characteristics of protein-bound chlorophyll molecules are modulated by interaction with their environment [44], presenting shifts on absorption peaks that may explain the effect we observed at 670 nm. Albeit in a previous work, with a similar experimental setup, we did not find a significant decrease in the chlorophyll content of drought stressed *Arabidopsis* plants [27], in this work, where we extended the drought stress for 5 additional days, significant differences were found between S0 and S2 for both Chl. a and Chl. b (Figure 2c). Moreover, significant differences were also observed on the Chl.a/Chl.b and on the Chl.a+b/Cx+c ratios (Figure 2c) contributing to changes on the photosynthetic pigments' absorption spectra and, most probably, on the leaf reflectance spectra. The significant changes observed in the anthocyanins' absorption spectrum (Figure 2d) is in line with the massive increase of anthocyanins content previously reported [27].

Anthocyanins are a class of different non-photosynthetic water-soluble pigments derived from flavonoids via the shikimic acid pathway [17]. Their biosynthesis is elicited by a number of abiotic stresses, including drought stress [45]. They may function as a screen to increase photoprotection and as a pool of antioxidant metabolites [46]. In vivo, anthocyanins present strong absorption between 450 and 600 nm [47], suggesting that they may be responsible for the high frequency observed between 500 and 600 nm in Figure 7. In fact, our results show a significant increase on the anthocyanins absorption spectra, from S0 to S2, on that wavelength range (Figure 2d). Furthermore, associated with anthocyanin biosynthesis, other chromophores, absorbing at wavelengths lower than 440 nm, are frequently accumulated in the leaf cell vacuoles, as is the case of UV-absorbing flavonols and other phenolics [48], possibly contributing to the high frequencies observed in the vicinity of 400 nm in Figure 7. Indeed, we also found significant differences on the absorption spectrum of anthocyanins around 400 nm, but in this case a decrease in absorption was found at S2, when compared to S0 (Figure 2d).

Additionally, the carotenoid content is known to change in drought stressed plants, in particular the content of violaxanthin, antheraxanthin and zeaxanthin, the three xanthophylls involved in the xanthophyll cycle [49]. Our results show some decrease of carotenoids content between S0 and S2 (Figure 2c). In vitro, carotenoids (xanthophylls and carotenes) show three distinctive absorption peaks, ranging from 416/440/465 nm for violaxanthin to 426/449/476 nm for β -carotene [50,51]. Therefore, in vivo, carotenoids and chlorophylls have overlapping absorption bands in the blue range, making it difficult to distinguish between them. However, there is an absorption peak at 520 nm which can be attributed to carotenoids [52]. However, only on XGBoost this wavelength showed a significant discriminant power, suggesting that alterations in carotenoids composition played a minor role in our results. Actually, we did not find such 520 nm absorption peak in our results, albeit the shoulder of the chlorophylls Soret band observed in the total photosynthetic pigments extract at 470 nm might be attributed to the presence of carotenoids (Figure 2d).

By using a non-invasive methodology, our study shows that leaf spectral readings are good indicators of soil water content in the plant model *A. thaliana*. We are aware that the translation of knowledge from *Arabidopsis* to agricultural species may bring new challenges, as always happens when knowledge is translated from model organisms to commercially valued organisms. Nevertheless, we produced discriminative models with a good potential to detect naturally occurring drought stress, and based on the results obtained on our data set, RF was the best performing algorithm. We also identified characteristics that are important for stress phenotyping in plants, which can be applied before crop growth cycles are affected, contributing to a more cost-effective management of plant breeding.

Author Contributions: Conceptualization, J.M.d.S.; methodology, A.B., P.M.P.C., S.S., M.C.P., A.R.M. and A.B.d.S.; software, P.M. and A.B.; formal analysis, A.B.; investigation, P.M.P.C.; writing—original draft preparation, J.M.d.S., S.S., M.C.P. and A.B.; writing—review and editing, J.M.d.S., P.M.P.C., S.S., A.B. and A.R.M.; visualization, A.B.; supervision, J.M.d.S. and S.S.; project administration, J.M.d.S.; funding acquisition, J.M.d.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FCT research units BioISI (UID/MULTI/04046/2019) and LASIGE (UIDB/00408/2020 and UIDP/00408/2020), and R&D project INTERPHENO (PTDC/ASP-PLA/28726/2017).

Acknowledgments: We acknowledge the support of André Cartaxana on the experimental work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Pretty, J.; Robinson, S.; Thomas, S.M.; Toulmin, C. Food security: The challenge of feeding 9 billion people. *Science* **2010**, *327*, 812–818. [[CrossRef](#)] [[PubMed](#)]
2. Henry, R.J. Innovations in plant genetics adapting agriculture to climate change. *Curr. Opin. Plant Biol.* **2019**, *56*, 168–173. [[CrossRef](#)]
3. Afonnikov, D.; Genaev, M.; Doroshkov, A.; Komyshev, E.; Pshenichnikova, T. Methods of high-throughput plant phenotyping for large-scale breeding and genetic experiments. *Russ. J. Genet.* **2016**, *52*, 688–701. [[CrossRef](#)]
4. Gebbers, R.; Adamchuk, V.I. Precision agriculture and food security. *Science* **2010**, *327*, 828–831. [[CrossRef](#)]
5. Das Choudhury, S.; Samal, A.; Awada, T. Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* **2019**, *10*, 508. [[CrossRef](#)]
6. Costa, J.M.; Marques da Silva, J.; Pinheiro, C.; Bar'on, M.; Mylona, P.; Centritto, M.; Haworth, M.; Loreto, F.; Uzilday, B.; Turkan, I.; et al. Opportunities and limitations of crop phenotyping in southern european countries. *Front. Plant Sci.* **2019**, *10*, 1125. [[CrossRef](#)] [[PubMed](#)]
7. Ulissi, V.; Antonucci, F.; Benincasa, P.; Farneselli, M.; Tosti, G.; Guiducci, M.; Tei, F.; Costa, C.; Pallottino, F.; Pari, L.; et al. Nitrogen concentration estimation in tomato leaves by vis-nir non-destructive spectroscopy. *Sensors* **2011**, *11*, 6411–6424. [[CrossRef](#)] [[PubMed](#)]
8. Cavaco, A.M.; Pinto, P.; Antunes, M.D.; Marques da Silva, J.; Guerra, R. 'Rocha' pear firmness predicted by a vis/nir segmented model. *Postharvest Biol. Technol.* **2009**, *51*, 311–319. [[CrossRef](#)]
9. Shull, C.A. A spectrophotometric study of reflection of light from leaf surfaces. *Bot. Gaz.* **1929**, *87*, 583–607. [[CrossRef](#)]
10. Rouse, J.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the great plains with erts. *NASA Spec. Publ.* **1974**, *351*, 309.
11. Vieira, S.; Utkin, A.B.; Lavrov, A.; Santos, N.M.; Vilar, R.; da Silva, J.M.; Cartaxana, P. Effects of intertidal microphytobenthos migration on biomass determination via laser-induced fluorescence. *Mar. Ecol. Prog. Ser.* **2011**, *432*, 45–52. [[CrossRef](#)]
12. Gamon, J.; Penuelas, J.; Field, C. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* **1992**, *41*, 35–44. [[CrossRef](#)]
13. Gamon, J.; Kovalchuck, O.; Wong, C.; Harris, A.; Garrity, S. Monitoring seasonal and diurnal changes in photosynthetic pigments with automated pri and ndvi sensors. *Biogeosciences* **2015**, *12*, 4149–4159. [[CrossRef](#)]
14. Jahns, P.; Holzwarth, A.R. The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochim. Biophys. Acta Bioenerg.* **2012**, *1817*, 182–193. [[CrossRef](#)]
15. Biswal, B. Carotenoid catabolism during leaf senescence and its control by light. *J. Photochem. Photobiol. B Biol.* **1995**, *30*, 3–13. [[CrossRef](#)]
16. Zhai, Y.; Cui, L.; Zhou, X.; Gao, Y.; Fei, T.; Gao, W. Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: Comparison of partial least-square regression and support vector machine regression methods. *Int. J. Remote Sens.* **2013**, *34*, 2502–2518.
17. Chalker-Scott, L. Environmental significance of anthocyanins in plant stress responses. *Photochem. Photobiol.* **1999**, *70*, 1–9. [[CrossRef](#)]
18. Mahlein, A.-K.; Steiner, U.; Dehne, H.-W.; Oerke, E.-C. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. *Precis. Agric.* **2010**, *11*, 413–431. [[CrossRef](#)]
19. Behmann, J.; Mahlein, A.-K.; Rumpf, T.; Römer, C.; Plümer, L. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precis. Agric.* **2015**, *16*, 239–260. [[CrossRef](#)]
20. van Dijk, A.D.J.; Kootstra, G.; Kruijer, W.; de Ridder, D. Machine learning in plant science and plant breeding. *IScience* **2020**, *24*, 101890.
21. Maimon, O.Z.; Rokach, L. *Data Mining with Decision Trees: Theory and Applications*; World Scientific Publishing: Hackensack, NJ, USA, 2014; Volume 81.
22. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. CART. In *Classification and Regression Trees*; Wadsworth and Brooks/Cole: Monterey, CA, USA, 1984.

23. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
24. Friedman, J. *Stochastic Gradient Boosting*. Department of Statistics; Technical Report; Stanford University: San Francisco, CA, USA, 1999.
25. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
26. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
27. Gameiro, C.; Utkin, A.; Cartaxana, P.; da Silva, J.M.; Matos, A. The use of laser induced chlorophyll fluorescence (LIF) as a fast and nondestructive method to investigate water deficit in arabidopsis. *Agric. Water Manag.* **2016**, *164*, 127–136. [[CrossRef](#)]
28. Lichtenthaler, H.K. Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods Enzymol* **1987**, *148*, 350–382.
29. Lin, N.; Noe, D.; He, X. Tree-based methods and their applications. In *Springer Handbook of Engineering Statistics*; Springer: London, UK, 2006; pp. 551–570.
30. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13–17 August 2016; pp. 785–794.
31. Bauer, E.; Kohavi, R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Mach. Learn.* **1999**, *36*, 105–139. [[CrossRef](#)]
32. Dietterich, T.G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.* **2000**, *40*, 139–157. [[CrossRef](#)]
33. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
34. Rao, R.B.; Fung, G. On the dangers of cross-validation. An experimental evaluation. In Proceedings of the SIAM International Conference on Data Mining, SDM 2008, Atlanta, GA, USA, 24–26 April 2008; pp. 588–596.
35. Kotsiantis, S.B. Decision trees: A recent overview. *Artif. Intell. Rev.* **2013**, *39*, 261–283. [[CrossRef](#)]
36. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [[CrossRef](#)]
37. Carter, G.A. Responses of leaf spectral reflectance to plant stress. *Am. J. Bot.* **1993**, *80*, 239–243. [[CrossRef](#)]
38. Katsoulas, N.; Elvanidi, A.; Ferentinos, K.P.; Kacira, M.; Bartzanas, T.; Kittas, C. Crop reflectance monitoring as a tool for water stress detection in greenhouses: A review. *Biosyst. Eng.* **2016**, *151*, 374–398. [[CrossRef](#)]
39. Seelig, H.D.; Hoehn, A.; Stodieck, L.S.; Klaus, D.M.; Adams, W.W., III; Emery, W.J. The assessment of leaf water content using leaf reflectance ratios in the visible, near-, and short-wave-infrared. *Int. J. Remote Sens.* **2008**, *29*, 3701–3713. [[CrossRef](#)]
40. Easterday, K.; Kislik, C.; Dawson, T.E.; Hogan, S.; Kelly, M. Remotely Sensed Water Limitation in Vegetation: Insights from an Experiment with Unmanned Aerial Vehicles (UAVs). *Remote Sens.* **2019**, *11*, 1853. [[CrossRef](#)]
41. Carter, G.A.; Estep, L. General spectral characteristics of leaf reflectance responses to plant stress and their manifestation at the landscape scale. In *From Laboratory Spectroscopy to Remotely Sensed Spectra of Terrestrial Ecosystems*; Springer: Dordrecht, The Netherlands, 2002; pp. 271–293.
42. Gitelson, A.A.; Chivkunova, O.B.; Merzlyak, M.N. Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *Am. J. Bot.* **2009**, *96*, 1861–1868. [[CrossRef](#)] [[PubMed](#)]
43. Porra, R.; Thompson, W.; Kriedemann, P. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: Verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta Bioenerg.* **1989**, *975*, 384–394. [[CrossRef](#)]
44. Zucchelli, G.; Jennings, R.C.; Garlaschi, F.M.; Cinque, G.; Bassi, R.; Cremonesi, O. The calculated in vitro and in vivo chlorophyll a absorption bandshape. *Biophys. J.* **2002**, *82*, 378–390. [[CrossRef](#)]
45. Cirillo, V.; D’Amelia, V.; Esposito, M.; Amitrano, C.; Carillo, P.; Carputo, D.; Maggio, A. Anthocyanins are key regulators of drought stress tolerance in tobacco. *Biology* **2021**, *10*, 139. [[CrossRef](#)]
46. Zhang, Q.; Zhai, J.; Shao, L.; Lin, W.; Peng, C. Accumulation of anthocyanins: An adaptation strategy of mikania micrantha to low temperature in winter. *Front. Plant Sci.* **2019**, *10*, 1049. [[CrossRef](#)]
47. Merzlyak, M.N.; Chivkunova, O.B.; Solovchenko, A.E.; Naqvi, K.R. Light absorption by anthocyanins in juvenile, stressed, and senescing leaves. *J. Exp. Bot.* **2008**, *59*, 3903–3911. [[CrossRef](#)]
48. Cerovic, Z.; Ounis, A.; Cartelat, A.; Latouche, G.; Goulas, Y.; Meyer, S.; Moya, I. The use of chlorophyll fluorescence excitation spectra for the non-destructive in situ assessment of uv-absorbing compounds in leaves. *Plant Cell Environ.* **2002**, *25*, 1663–1676. [[CrossRef](#)]
49. Latowski, D.; Kuczyńska, P.; Strzałka, K. Xanthophyll cycle—A mechanism protecting plants against oxidative stress. *Redox Rep.* **2011**, *16*, 78–90. [[CrossRef](#)] [[PubMed](#)]
50. Ruban, A.V.; Pascal, A.A.; Robert, B.; Horton, P. Configuration and dynamics of xanthophylls in light-harvesting antennae of higher plants: Spectroscopic analysis of isolated light-harvesting complex of photosystem II and thylakoid membranes. *J. Biol. Chem.* **2001**, *276*, 24862–24870. [[CrossRef](#)] [[PubMed](#)]
51. Zaghoudi, K.; Ngomo, O.; Vanderesse, R.; Arnoux, P.; Myrzakhmetov, B.; Frochot, C.; Guiavarc’h, Y. Extraction, identification and photophysical characterization of persimmon (*Diospyros kaki* L.) carotenoids. *Foods* **2017**, *6*, 4. [[CrossRef](#)] [[PubMed](#)]
52. Zur, Y.; Gitelson, A.; Chivkunova, O.; Merzlyak, M. The spectral contribution of carotenoids to light absorption and reflectance in green leaves. In Proceedings of the 2nd International Conference Geospatial Information in Agriculture and Forestry, Buena Vista, FL, USA, 10–12 January 2000; Volume 2.