Fig. 1. Schematic representation of the experimental setup with plants of both groups monitored using bioimpedance measurement. The portable impedance analyzer Analog Discovery 2 is connected to each plant through a multiplexer, synchronized by a Python script controlling the channel opening by an Arduino board.

Natick, MA, USA). To reduce the dataset to a smaller set of interpretable variables, the bioimpedance data were fitted by the MATLAB Zfit script [26] with the single dispersion Cole electrical model [23] shown in Fig. s2 (see Supplementary material). Here, the best fitting was calculated using the real and the imaginary components of the bioimpedance spectra as input data and iterating different combinations of nine selected optimal initial conditions with a reasonable magnitude for the considered dataset ranges, evaluating the results in terms of root-mean-square error (RMSE). As result of the fitting, the equivalent circuit parameters (i.e., the series resistor R_s, parallel resistor R_p, magnitude component of the constant phase element CPE_t, and phase angle component of the constant phase element CPE_p,) were extracted for each measurement. One-way ANOVA was performed to evaluate statistical differences among each circuit parameter for class discrimination (i.e., between control and iron-stressed plants). The Tukey's honestly significant difference (HSD) method was performed, and the differences were calculated for an appropriate level of interaction ($p \leq 0.05$). Results were reported as the mean and standard deviation (SD) of the mean. The possible linear dependency between the obtained equivalent circuit parameters and the greenhouse environmental conditions during the experiment was investigated by studying the correlation of each circuit component with respect to RH, T, and light, calculating Pearson's correlation coefficient for each relationship [27].

D. Data Processing and Classification

The dataset to train the models was composed of the equivalent circuit parameters extracted for the measurement acquired in the temporal interval between the 5th of August and the 27th of August, labeling the data according to the respective group of the plant, either control or stress. In addition, the measurements collected from the iron-stressed plants were further divided into two classes, with data from the 5th of August to the 16th of August labeled as in early stress while data referring to the measurements of the subsequent days (i.e., until 27th of August) indicated as late stress. Afterward, the obtained dataset was partitioned with an 80:20 ratio into train and test sets, to train and validate various machine learning classification models, including discriminant analysis (DA), kernel methods (KM), k-nearest neighbors (KNN), linear models (LM), naive Bayes classifier (NBC), multilayer perceptron (MLP), support vector machines (SVM), and decision trees (DT), using MATLAB's automatic toolbox *fitcauto* that allows the evaluation of the performance of different classification algorithms by automatically optimizing their hyperparameters. The different type of machine learning algorithms was employed due to their different characteristics, to evaluate the one with better performance in discriminating the fruit ripening. LR utilizes a logistic function to model a binary dependent variable, assuming a linear relationship between log-odds and predictors. DT are nonparametric, creating simple decision rules from data features, providing easily interpretable models but susceptible to overfitting. NBC assumes conditional independence of features given the class variable, offering speed and effectiveness in generating simple models. KNN assigns class membership based on similarity among samples, calculating distances to k nearest neighbors. SVM draw hyperplanes in feature space to maximize class separation. MLP is a simple feed-forward neural network capable of learning nonlinear models with at least three layers and nonlinear activation functions [28]. The robustness of the trained model was assessed by a subsequent 10 000-round bootstrapping validation phase together with a tenfold cross validation, evaluated on an equally distributed subset of data, and the accuracy of the obtained models was evaluated in terms of F_1-score.

III. RESULTS AND DISCUSSION

The eight plants, divided between four plants exposed to the complete nutrient solution (i.e., control) and four plants with