Title:

Multiclass classification of environmental chemical stimuli from unbalanced plant electrophysiological data

Published year: 2023

Link: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0285321

Main points to notice:

- Less accuracy

Our target:

- Increase accuracy

Summary:

This paper presents a multi-class classification approach for identifying different chemical stimuli applied to plants based on their electrical signal responses. The main findings are:

• Higher order statistical features were extracted from the plant electrical signals recorded under three chemical stimuli - sulfuric acid, ozone and sodium chloride.

• The dataset was highly imbalanced due to the varying lengths of experiments for different stimuli, with ozone accounting for majority of the data.

• Monte Carlo under-sampling was applied to the majority classes to create balanced subsets for classification using 8 different algorithms.

• Classification performance was evaluated using accuracy, balanced accuracy, F1-score and Matthews correlation coefficient to account for the imbalanced data.

• Amongst the classifiers, adaptive boosting and random forest (gini) performed the best, achieving a balanced accuracy of 0.71 using the original 15 statistical features.

• Principal component analysis was also performed to reduce the dimensionality to 7 components. However, this did not significantly improve classification performance for most classifiers.

• Multivariate analysis of variance testing showed that the difference in classification performance between the original 15

features and 7 principal components was statistically significant for multilayer perceptron classifier but not for others.

In summary, the study demonstrated that adaptive boosting and random forest classifiers performed the best for identifying chemical stimuli applied to plants based on their electrical responses, with the original 15 statistical features yielding better results than the reduced principal components for most classifiers.