DEPARTMENT OF LIFE SCIENCES

Broccoli Stress Detection and Shelf Life Prediction Based on Spectrometry and Machine Learning

AUTHOR:

XIAOSHENG LUO

CID:

01627437

SUPERVISOR:

Dr. OLIVER WINDRAM

August 27, 2019

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Research at Imperial College London

Formatted in the style of Methods in Ecology and Evolution

Submitted for the MRes in Computational Methods in Ecology and Evolution

Declaration

The images of the broccoli head on the conveyor belt used to construct and training the neural network was originally provided by Nathan E. Barlow (Imperial College London), and then the optimized image video was collected by the author and the superviosr, Dr. Oliver Windram ((Imperial College London). Besides, Dr. Oliver Windram was mainly responsible for shaping the direction in this project. Acquisition of experimental data, data cleaning, data analysis, method modification, model training and tuning and writing were exclusively performed by the author himself.

Abstract

Keywords:

Word count: 2191.

Contents

1 INTRODUCTION

2 MATERIALS AND METHODS

2.1 Data Collection

Broccolis were grown in Control Environment room and greenhouse. The normal growth temperature is controlled at 23 °C and the humidity is controlled at 60 percent, with long daylight (16 hours illumination, 8 hours darkness) treatment, and water is poured every 3 days from the tray to the soil. The whole growth cycle of broccoli takes about three months, during which it needs to be transferred to a suitable pot according to the size of broccoli.

Table 1: Apparatus used in the experiment.

Item	Description
Camera	Ximea 1.3 MP NIR Enhanced Camera MQ013RG-ON
Machine vision lens	MVL12M23-12 mm EFL, f/1.4, for 2/3" " C-Mount Format Cameras, with Lock.
Band pass filter	FEL0800: Ø25.0 mm Premium Longpass Filter, Cut-On Wavelength: 800 nm. FBH520: Ø25.0 mm Hard-Coated Bandpass Filters, Blocking Regions (OD >5): 200 - 485 nm, 556 - 1200 nm. FBH650: Ø25.0 mm Hard-Coated Bandpass Filters, Blocking Regions (OD >5), 200 - 611 nm, 690 - 1200 nm. FBH850: Ø25.0 mm Hard-Coated Bandpass Filters, Blocking Regions (OD >5), 200 - 805 nm, 896 - 1200 nm.
LED controller	Intelligent LED Solutions 12-Channel Light Controller
LED	12 Die LED array Full Spectrum 360-955nm

During stress treatment, broccolis were randomly divided into four groups with 8 individuals in each group. They were treated under control, heat stress (27°C), drought stress (without watering) and combination of heat and drought stress. Leaf reflectance spectroscopy data was collected by the Ocean Optics FLAME-S-XR1 spectrophotometer in a complete dark room, while the spectral image was collected by

Ximea cameras with a specific bandpass filter (FEL0800, FBH650-40) (Table 1) under the illumination of corresponding wavelength of the LED lamp, four days of data were collected, until the leaves show a distinct dehydration drooping phenotype. And in the open-air greenhouse, spectral images are collected in a grow tent.

Broccoli heads for shelf life prediction come from POLLYBELL FARMS LTD.. They are divided into two groups, 18 in each, one of which is stored in cold storage for a period of time, and the other is harvested freshly. They were placed naturally at room temperature and spectral image data were collected every day through the cameras with bandpass filter (FBH520-40, FBH650-40,FBH850-40) until they decay significantly.

2.2 Machine Learning

Machine learning generally includes several steps in practical operation, such as data collection and preprocessing, model selection, training, evaluation and repeatedly fine-tuning until a good prediction effect is achieved.(Figure 2) In the data preprocessing stage, Z-score standardization is applied to simplify the calculation and the categorical data is one hot enconded. In the strategy of training algorithms, firstly, logistic regression, support vector machine, random forest and XGBoost algorithm were trained to fit the raw data, and obtained the baseline score, and then the performance of the models were optimized by feature engineering and parameters adjustment. Most of the code used in this process is based on the API provided by sklearn(Pedregosa et al., 2011).

2.2.1 Model Fitting

Logistic regression: the binomial logistic regression model is a classification model, which is represented by the conditional probability distribution $P(Y|X)$, in the form of parameterized logistic distribution. Here, the value of X is a real number, and the random variable Y takes a value of 0 or 1, then we estiamte the model parameters by supervised learning. The binomial logistic regression model is the conditional probability distribution as follows:

$$P(Y=1|x)=\frac{\exp(w\cdot x)}{1+\exp(w\cdot x)}$$

$$P(Y=0|x)=\frac{1}{1+\exp(w\cdot x)}$$

Here, x is the input vector, w is the weight vector and Y is the output vector, $Y\in\{0,1\}$, $x=(x^{(1)},x^{(2)},...,x^{(n)},1)^T$, $w=(w^{(1)},w^{(2)},...,w^{(n)},b)^T$. By comparing the probability of $P(Y=1|x)$ and $P(Y=0|x)$ can finally determine the catagory. The cost function of logistic regression can be derived by the method of maximum likelihood estimation, which is known as the average of cross-entropy loss. Meanwhile, in order to avoid over-fitting, the L1 or L2 regularization terms was added during the optimization process.

SVM: the main idea of SVM is to find the decision boundary with the largest classification interval between two different catagories, which means the closest data point to the surface, also known as the support vectors, determines the margin of the classifier. For simple linear separability problems, it can be described as an optimization problem by mathematical formulas as follows:

$$\max_{w,b}\left[\min_{x_i}\frac{y_i\,(w\cdot x_i+b)}{\|w\|}\right]$$

The minimized item represents the distance from the support vectors to the decision boundary with sign, known as geometry margin. By scaling w and b so that (x_j,y_j) as the point to get the minimum value, $y_j\left(\omega^T x_j+b\right)=1$, so the other sample points are naturally greater than or equal to 1. Derived all the way and we can finally got a methematical optimization problem:

$$\begin{aligned}&\min_{w,b}\tfrac{1}{2}\|w\|^2\\&\text{s.t. }\ y_i\left(w^{\mathrm{T}}x_i+b\right)\geqslant 1,\quad i=1,2,\ldots,m\end{aligned}$$

Futher, in order to allow the SVM to ignore some noise, a slack variable ($\xi_i \geq 0$) is introduced to allow some wrong classification, that is, allow some data points' functional margin less than 1, correspondingly, a penalty term is needed to add to the objective function to limit the slack variable, and here is the basic linear separable SVM:

$$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{m}\xi_i$$
$$\text{s.t. } y_i\left(w^T x_i + b\right) \geq 1 - \xi_i (i = 1, 2, \dots m)$$
$$\xi_i \geq 0 \quad (i = 1, 2, \dots m)$$

Finally, the problem can be solved by Lagrange Duality and SMO algorithm(Platt, 1998). In addition, kernel mapping has also been tried to verify the linear separability of data.

Ensemble methods: Random Forests and XGBoost(Chen and Guestrin, 2016)(see SI 6for details of the algorithm.), both are based on decision tree model, they use the methods of bagging and boosting respectively, which can help to prevent high variance and high bias. Random forest mainly consists of two stochastic processes, random sampling of samples and features to construct many decision trees that are independent of each other. The final prediction results are summarized by voting strategy. As for XGBoost, it's an algorithm developed from gradient boosted decision trees and designed for speed and performance. It's widely used in many competitions and achieved good grades.

In the process of training machine learning algorithms, when multi-classification is performed on logistic regression and support vector machines, "one to the rest" strategy is applied. All the models are validated by 6-fold cross-validation, and ROC_AUC is used as the metric of model evaluation.

2.2.2 Feature Engineering

Generally, data and features determine the upper bound of machine learning, whereas models and algorithms only approximate this upper bound. The purpose of feature engineering is to extract effective features and remove redundant featuresfrom the original data. It basically includes feature extraction, feature construction and feature selection. Separately,featureextraction mainly uses dimension reduction methods such as PCA and LDA. Feature construction is to construct various indices based on previous spectral studies. Feature selection methods can be roughly divided into three types:

- Filter: scoring each feature according to divergence, correlation, etc., and then set a threshold for selection feature.
- Embedded: use some machine learning algorithms and models to train and get the coefficients of each feature, select features according to the coefficient, kind of similar to the filter method, but models are trained to determine the pros and cons of features. Specifically, multi-method ensemble selcetion (Feilhauer et al., 2015)was modified from the regression problem and later adapted to the classification problem.
- Wrapper: recursive elimination feature method, due to the high computational complexity and the long execution time of the algorithm, it is not adopted here.

In addition, in order to find a suitable bandpass filter for the camera, a search box with specific bandwidth was used to repeatedly and randomly select features, and the importance of features is sorted by simple ANOVA and tukeyHSD significance test. Finally, the graph is plotted by accumulating significant ($p < 0.05$) bandwith features.

2.3 Computer Vision

The project involves computer vision tasks such as image classification, segmentation, alignment, and image feature extraction like color histogram. Specifically, image alignment was performed by classic Scale-invariant feature transform (SIFT) (Lowe et al., 1999), Image classification and segmentation are mainly accomplished by the transfer learning of convolution neural network (Figure 1). Image classification was

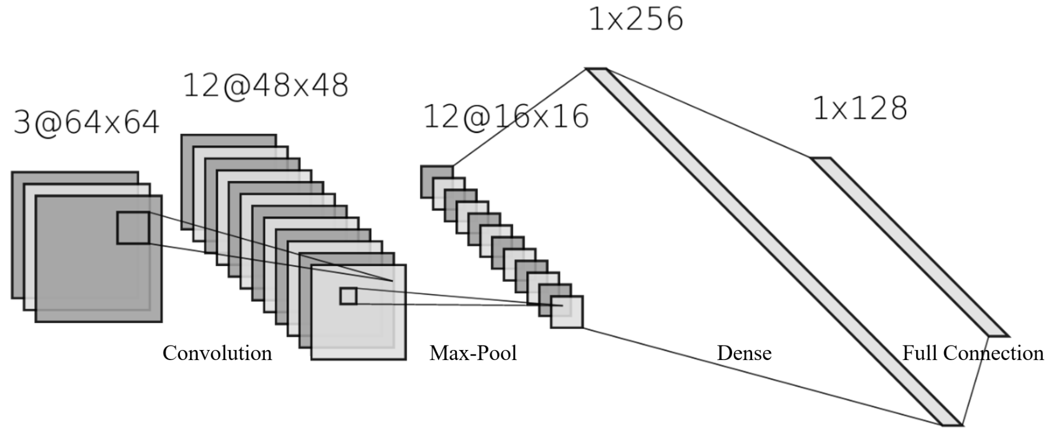

Figure 1: Structure of a simple convolution neural network
An image was taken as input (for example, a RGB image, normally three channels), then through the calculation with the multiple kernels' parameters and activation functions (usually ReLu) in convolution layer, and the downsampling process in pooling layer, can achieve the purpose of weight sharing and parameter reduction. Finally,the results are expanded and classifiedby the fully connected layer and the softmax function.In the figure, the number in front of @ is the number of channels, and the back is the height and width of pixels.

implemented by ResNeXt (Xie et al., 2016), by UC San Diego and Facebook AI Research, while Image segmentation was implemented by Unet (Ronneberger et al., 2015) and modified Unet.The training was conducted at P1000 on the HPC of Imperial College London, optimizer for the neural network is Adam, the cost function is cross-entropy and cyclical learning rates (Smith, 2017) was used, Most of the code is based on the API provided by Pytorch, fastai library (Howard et al., 2018), and opencv.

2.4 Workflow

The project mainly includes two parts (Figure 2).The laboratory part is to grow broccolis under control conditions and then perform individual and combined stress treatments, collect spectral images and leaf reflectance spectrum data to explore the signals that can effective distinguish among them and construct a robust machine learning classifer. The application part is to construct a detection system which can predict the shelf life of broccoli on the conveyor belt through computer vision methods and spectral images under specific bandwidth, which are selected based on the results obtained in the laboratory.

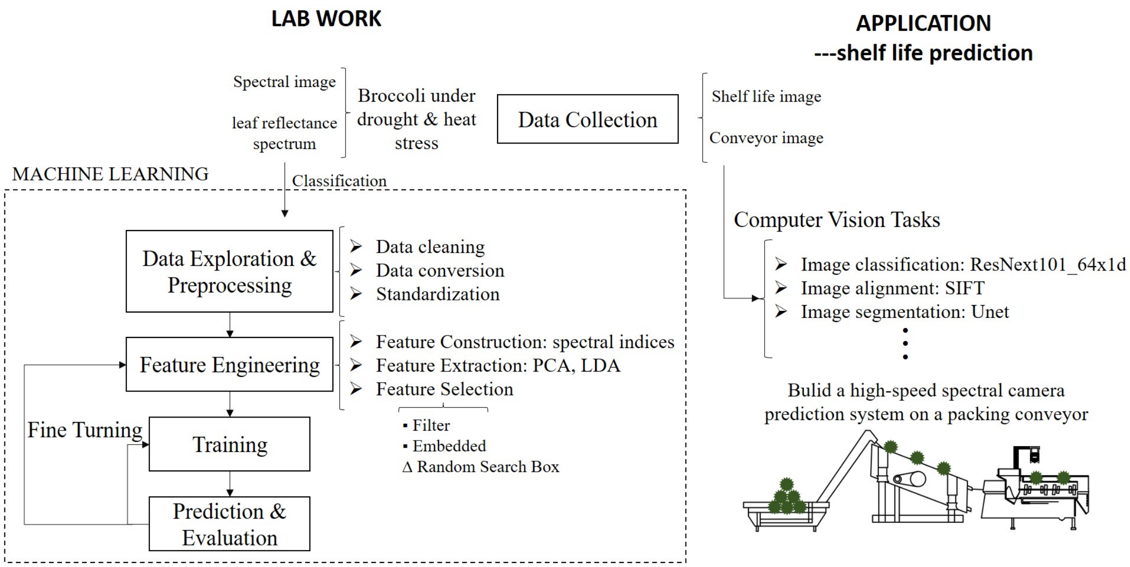

Figure 2: Project workflow

3 RESULTS

3.1 Data

In order to understand the upper limit of the classifier and ensure the treatment effect, we select the data of the day when the broccoli just appeared phenotype under the stress (Figure 3). Under the control and heat conditions, the broccolis have no obvious phenotype, while under drought and combined stress treatment, the broccolis leaves are a little drooping due to dehydration, and the combined stress is slightly more obvious than drought.

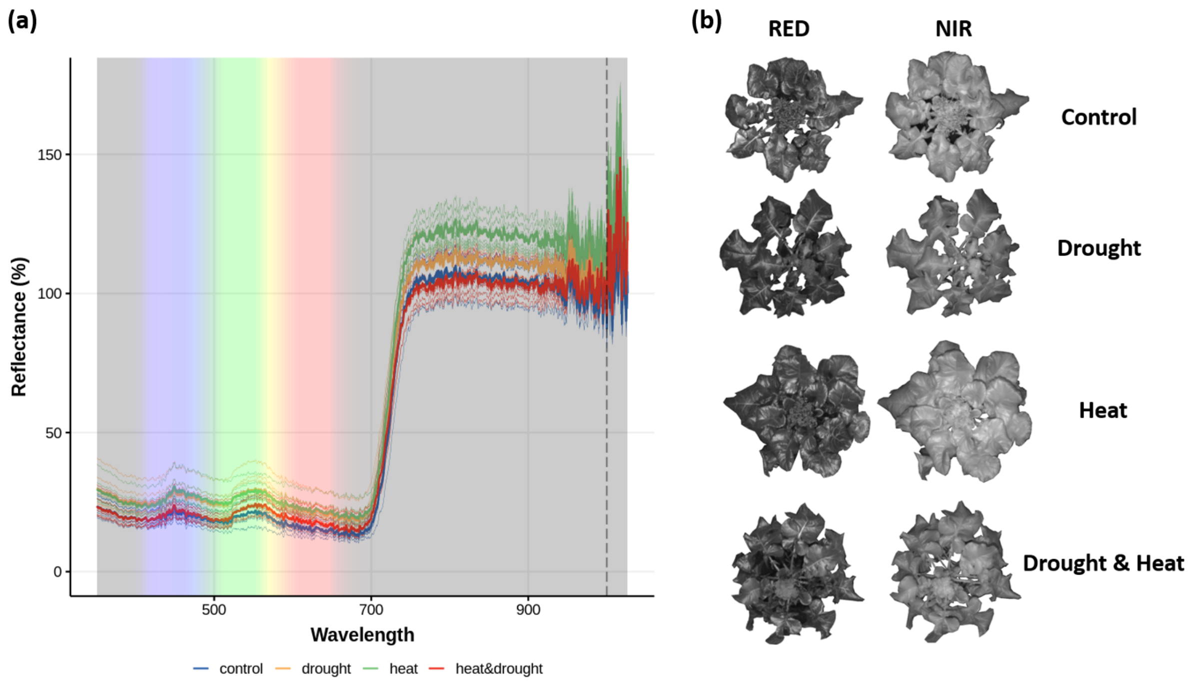

Figure 3: Spectral data of broccoli under heat and drought stress
(a) is the hyperspectral data of broccoli leaves detected by spectrophotometer in dark environment, the vertical axis represent their relative reflectivity.The thin line is averaged by the hyperspectral scan of all the leaves of each sample, and the thick line is the average of all samples in different treatments. The right side of the dashed line was discarded is subsequent processing due to abnormal signal fluctuation. (b) is the spectral images taken by the camera with red bandpass filter (CWL = 650 nm, FWHM = 40 nm) and near infrared bandpass filter (> 800 nm), under the illumination of the corresponding band of LED.

On the other hand, in the hyperspectrainl data of the leaves (Figure 3 (a)), it is difficult to get a distinct discriminant pattern from the perspective of data distribution, because the samples of different treatments are cross-covered. However, the overall trend of the broccoli leaves reflectance spectrum can still be clearly seen. There are small peaks at the blue and green-yellow junctions in the visible region, and it is well known that a samll valley in the red band and strong reflection rate shifting in the near infrared band. In particular, the average reflectance of the heat treatment appears to be significantly higher than other treatments.

3.2 Vegetation Indices

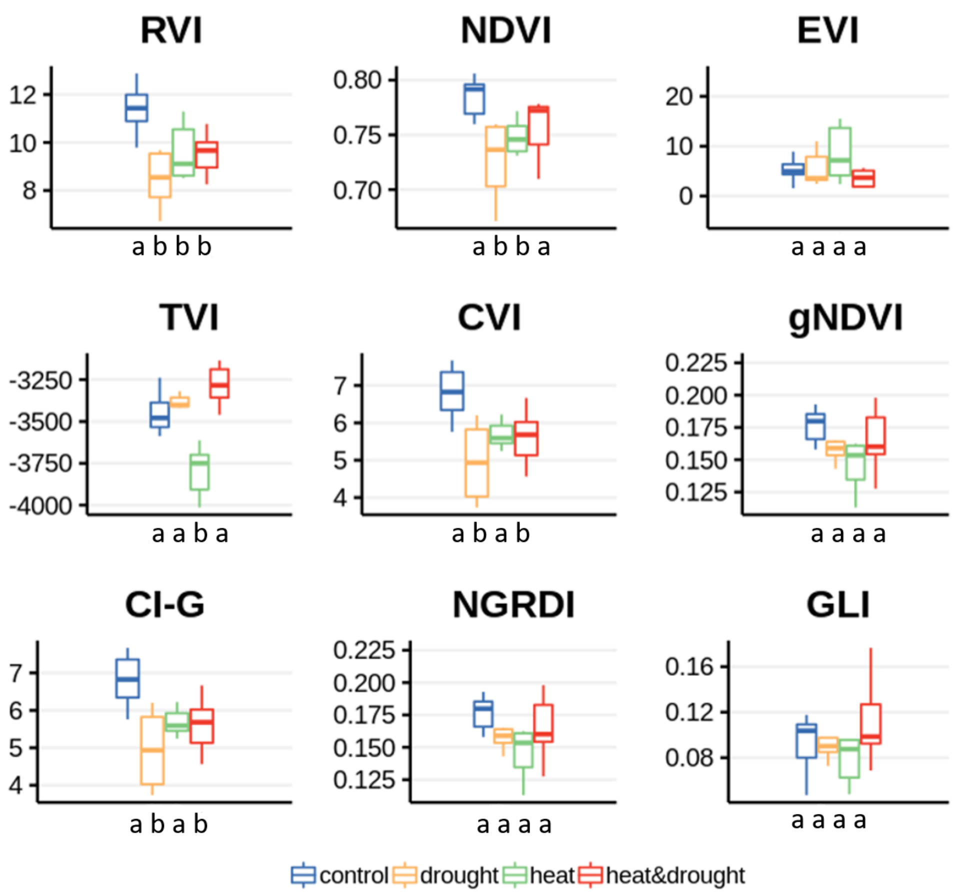

Figure 4: Various vegetation indices under different stresses
Each sanned data was ragarded as a data point, the equation for calculating vegetation index refers to Tabel**??**.Normality test by shapiro test, and homogeneity of variance test was implemented by Barlett test before using ANOVA and Tukey's HSD for post-hoc analysis. Different letters under the x-axis represent sigificant differences ($p < 0.05$).

The spectral indices related to vegetation cover and chlorophyll content in different remote sensing were calculated under differnt stresses (Figure 4). We use each sanned data as one data point for detailed significance test. he results show that sigificant differences between the two treatments can't be simply obtained from a single. The results with significant differences also show different forms of band feature calculation methods, which is difficult to establish a unified pattern, suggesting that more complex models and feature construction methods are needed.

3.3 Dimensionality Reduction

There may be muti-collinearity between hyperspectrainl features, that is variable may be correlated. Meanwhile, too many variables may hinder the pattern for model fitting, and it may also involve a lot of redundant Information. herefore, dimensionality reduction was used to reduce variable, speed up computation and extract effective information hidden in the data.

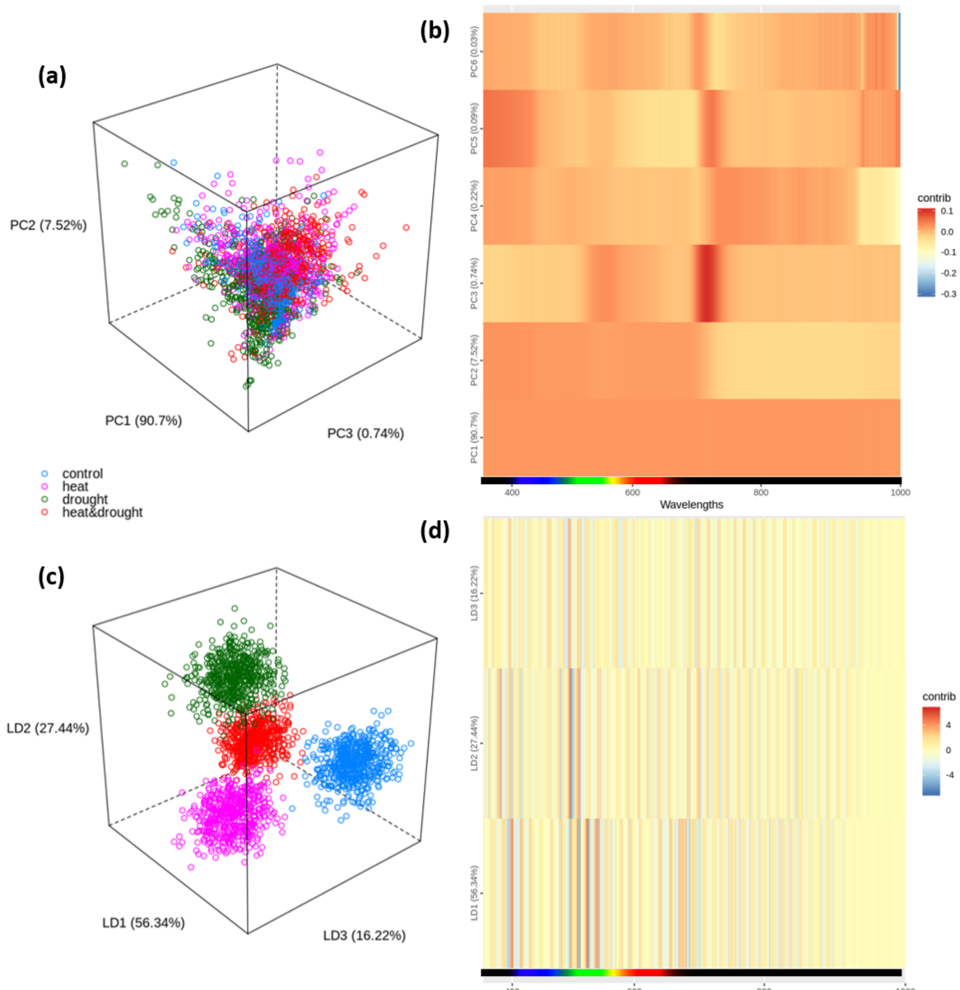

Figure 5: Feature dimensionality reduction by PCA and LDA

The results of unsupervised dimensionality reduction PCA show that, when the variables are mapped to the linear-independent direction of maximum variance, clustering between different treatment is not effective. Surprisingly, PC1 explains 90.7% of the variance, and the contribution of each wavelength seem to contribute equally to it, possibly due to the systematic errors. In PC2, the visible light region contributes a large variance, and there are two specific narrow bands in PC3 that are positively correlated with it. On the other hand, LDA, the supervised dimension reduction method, can completely separate the different treatments. In these components, the green (around 520nm) and red edge (around 680nm) wavelength may be important to separate each stress treatment in the principal components.

3.4 Feature Selcetion

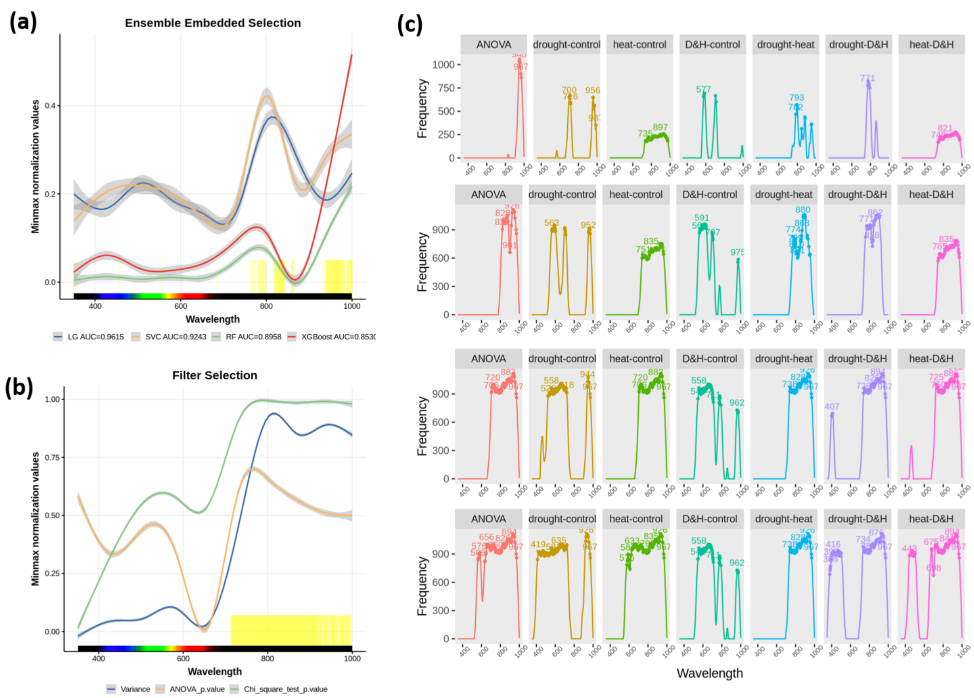

Figure 6: Feature Selection

3.5 Models

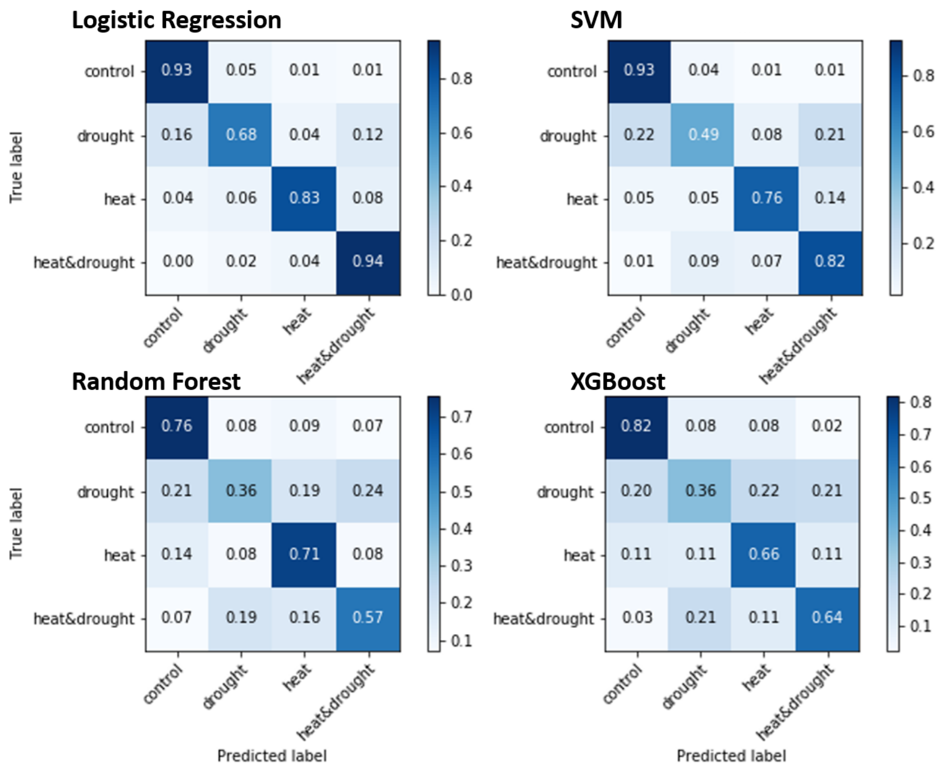

Figure 7: confusion matrix

3.6 Shelf Life Prediction

3.6.1

4 DISCUSSION

References

Chen, T. and Guestrin, C. (2016), Xgboost: A scalable tree boosting system, *in* 'Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining', ACM, pp. 785–794.

Feilhauer, H., Asner, G. P. and Martin, R. E. (2015), 'Multi-method ensemble selection of spectral bands related to leaf biochemistry', *Remote Sensing of Environment* **164**, 57–65.

Howard, J. et al. (2018), 'fastai', https://github.com/fastai/fastai.

Lowe, D. G. et al. (1999), Object recognition from local scale-invariant features., *in* 'iccv', Vol. 99, pp. 1150–1157.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* **12**, 2825–2830.

Platt, J. C. (1998), Sequential minimal optimization: A fast algorithm for training support vector machines, Technical report, ADVANCES IN KERNEL METHODS - SUPPORT VECTOR LEARNING.

Ronneberger, O., Fischer, P. and Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, *in* 'International Conference on Medical image computing and computer-assisted intervention', Springer, pp. 234–241.

Smith, L. N. (2017), Cyclical learning rates for training neural networks, *in* '2017 IEEE Winter Conference on Applications of Computer Vision (WACV)', IEEE, pp. 464–472.

Xie, S., Girshick, R., Dollár, P., Tu, Z. and He, K. (2016), 'Aggregated residual transformations for deep neural networks', *arXiv preprint arXiv:1611.05431* .

5 Supplementary Information

5.1 Experiment Apparatus