

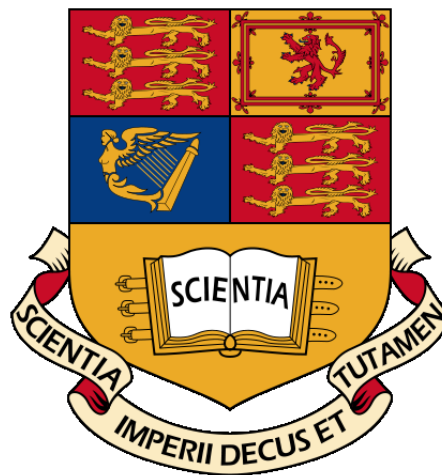
Classificaition of complex stresses in plants using machine learning

by

Abigail Baines (CID: 00742373)

Supervisor : Oliver Windram, Imperial College London,
o.windram@imperial.ac.uk

Department of Life Sciences
Imperial College London
London SL5 7PY
United Kingdom



Thesis submitted as part of the requirements for the award of the
MRes in Computational Methods in Ecology and Evolution, Imperial
College London, 2017-2018

Abstract

Determining the incidence and types of plant stresses has been of interest for many years, particularly due to their potential impact on agricultural production. Despite this, current methods are very qualitative, and rely on how the researcher interprets the type and extent of the stress. This is not only time consuming, but also makes comparing results from different studies open to speculation about how each researcher identified the stresses. With this in mind, our aim for this project is to see whether we can use machine learning to classify and quantify plant stress(es) from multi-spectral leaf images. We aim to do this by exposing *Arabidopsis thaliana* to a pathogen (*Botrytis cinerea*) and an abiotic stress. We will then subsequently monitor their spectral response following the stress exposure to determine what spectral features might evolve in the plant leaves to distinguish between the stressed (both multi-level and single level) and unstressed (control) plants.

Keywords: Digital Pathology, Random Forests, Python, Disease detection, Multi-spectral features, Machine Learning.

1 Introduction

Recently, the field of digital pathology has exploded; searching digital pathology on Web of Science returns over 3,000 hits, and 2013 onwards returns over 250 papers per annum. Despite this, plant pathology remains broadly unexplored using this technology.

The study of plant pathology has always received much attention, due to its link with agricultural yield and crop productivity (Donatelli et al. 2017; Waller et al. 2005). Plant pathology is similar to other streams of pathology; susceptibility is the antithesis of resistance, with resistance being that a plant can overcome or exclude pathogenic effects/organisms. The cycle of resistance and susceptibility is powered by natural selection; plants become able to attack invading pathogens via their immune response, then pathogens evolve so that they can suppress this triggered immunity, resulting in the plants being susceptible to attack once again (Lapin and Van den Ackerveken 2013; Pel and Pieterse 2013; Zheng et al. 2012). Artificial selection in plants has routinely been used in agriculture, in order to introduce resistance to specific diseases into crops (Dennis et al. 2008).

Despite the existence of numerous control measures for many crop diseases (Wood 1996), those that still manage to impact upon crop production can be devastating (McCook 2006; Woodham-Smith 1962). This is further complicated by the fact that different pathogen races and variable environmental conditions can radically affect the degree of disease severity (Dangl and Jones 2006; Suzuki and Mittler 2006). Past methods in detecting plant disease rely on experts simply observing the plant (Singh and Misra 2017); despite the rise in the use of machine learning for classification, many still rely on manual detection in this field. Manual detection, despite being easily applied in the field, is highly qualitative, and therefore is subject to interpretation. It is also noted that manual detection cannot quantify the level of disease severity, and it cannot detect low disease severity accurately (Lowe, Harrison and French 2017). Our project aims to detect presence/absence of disease, and severity of disease in plants.

The main aim of this project is to use machine learning to classify instances of complex stress exposure in plants (*Arabidopsis thaliana*). We aim to do this by performing 3 fully replicated experiments to capture multi-spectral data from different stress exposures. The stresses the plants will be exposed to are UV stress and *Botrytis cinerea*, a necrotrophic fungus that affects many plant species, and has been used widely in plant biology and pest management studies. The data collected will then be used to create an image classifier which will subsequently be tested on additional data samples. We also aim to build a fully functional web application to allow other pathologists to use this classifier in future research. We anticipate that this project will be successful due to previous work done using machine learning and image analysis, and that this project could be vastly influential to other areas utilizing image analysis techniques.

If our project was to be successful, we anticipate that it could be used to aid artificial selection, by choosing the most resistant individuals to breed from based on quantitative results, and also

precision agriculture. If diseases could be identified to specific individuals rather than whole plots, pesticides could be applied more precisely, minimizing costs to the farmers, and also benefiting ecosystems by reducing mass pesticide application.

2 Materials & Methods

The proposed methods are to gather multi-spectral image data from *Arabidopsis thaliana*, with treatments applied as laid out in Table 1.

Images will be taken of each treatment type and used to build a classifier to distinguish treatments. We will begin by using a simple random forest approach to classify known multi-spectral and texture features within the image set. These classifiers will be used to predict object features such as leaves and disease lesions and to distinguish these objects across treatments. This approach will allow us to produce single classifiers within each experimental replicate and to test its robustness across remaining replicates. We will use quantifiable metrics to score disease. These will include but are not limited to, the size of disease lesions between treatments and quantitative differences in the colour of leaves. For example the degree of chlorosis can be used to estimate and compare levels of disease severity between treatments. In addition to the complex stress trial planned for this project we will also process additional image data available from the lab members of diseased plants with different genetic backgrounds. In doing so we will look to quantify different levels of disease susceptibility in plants.

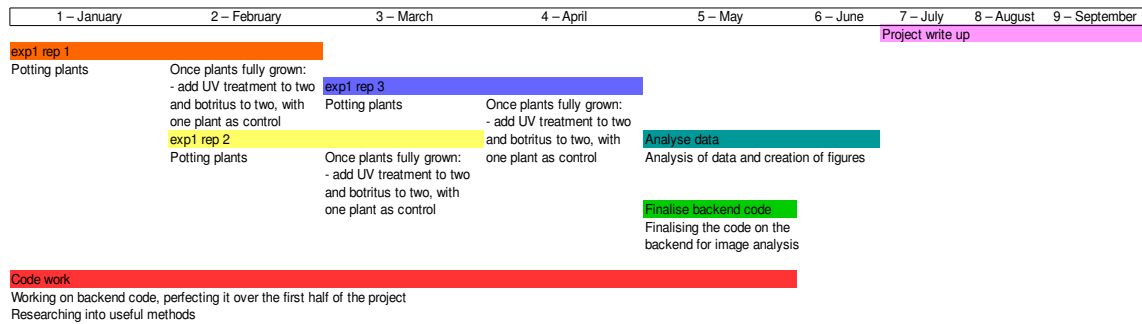
The pipeline outlined above will be written in python, with some code being translated from existing code written in matlab and R. We will also use packages such as opencv and ilastik to further build the pipeline, as well as to begin developing the web application.

The pipeline is highly adaptable and broadly applicable to a wide range of pathologies. To make this approach more broadly applicable to the plant pathology community we aim to develop a simple web application that new users can use to build classifiers to distinguish and quantify pathologies of their choice.

Table 1: Treatment types

	UV Treatment	
<i>Botrytis cinerea</i>	++	+-
treatment	-+	--

3 Gantt chart



4 Budget

The anticipated budget for this project is as follows:

- template for frontend webpage – £70
- Software and VPN – £60
- Backend management system – £80

- Sundries and consumables – for plants £300
 - Travelling costs – £100
 - Miscellaneous – £100
- Supervisor Oliver Windram has seen and accepted budget.

References

- Dangl, Jeffery L. and Jonathan D. G. Jones (2006). “The plant immune system”. In: *Nature* 444.7117, p. 323. ISSN: 1476-4687. DOI: 10.1038/nature05286. URL: <https://www.nature.com/articles/nature05286> (visited on 26/11/2017).
- Dennis, Elizabeth S et al. (2008). “Genetic contributions to agricultural sustainability”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1491, pp. 591–609. ISSN: 0962-8436. DOI: 10.1098/rstb.2007.2172. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2610172/> (visited on 21/11/2017).
- Donatelli, M. et al. (2017). “Modelling the impacts of pests and diseases on agricultural systems”. In: *Agricultural Systems* 155 (Supplement C), pp. 213–224. ISSN: 0308-521X. DOI: 10.1016/j.agsy.2017.01.019. URL: <http://www.sciencedirect.com/science/article/pii/S0308521X1730104X> (visited on 12/11/2017).
- Lapin, Dmitry and Guido Van den Ackerveken (2013). “Susceptibility to plant disease: more than a failure of host immunity”. In: *Trends in Plant Science* 18.10, pp. 546–554. ISSN: 1360-1385. DOI: 10.1016/j.tplants.2013.05.005. URL: <http://www.sciencedirect.com/science/article/pii/S1360138513001064> (visited on 21/11/2017).
- Lowe, Amy, Nicola Harrison and Andrew P. French (2017). “Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress”. In: *Plant Methods* 13, p. 80. ISSN: 1746-4811. DOI: 10.1186/s13007-017-0233-z. URL: <https://doi.org/10.1186/s13007-017-0233-z> (visited on 21/11/2017).
- McCook, Stuart (2006). “Global rust belt: *Hemileia vastatrix* and the ecological integration of world coffee production since 1850”. In: *Journal of Global History* 1.2, pp. 177–195. ISSN: 1740-0236, 1740-0228. DOI: 10.1017/S174002280600012X. URL: <https://www.cambridge.org/core/journals/journal-of-global-history/article/global-rust-belt-hemileia-vastatrix-and-the-ecological-integration-of-world-coffee-production-since-1850/F20E753BE062F107002E5D3F099D0AF5> (visited on 12/11/2017).
- Pel, Michiel J. C. and Corné M. J. Pieterse (2013). “Microbial recognition and evasion of host immunity”. In: *Journal of Experimental Botany* 64.5, pp. 1237–1248. ISSN: 0022-0957. DOI: 10.1093/jxb/ers262. URL: <https://academic.oup.com/jxb/article/64/5/1237/630577> (visited on 21/11/2017).
- Singh, Vijai and A. K. Misra (2017). “Detection of plant leaf diseases using image segmentation and soft computing techniques”. In: *Information Processing in Agriculture* 4.1, pp. 41–49. ISSN: 2214-3173. DOI: 10.1016/j.inpa.2016.10.005. URL: <http://www.sciencedirect.com/science/article/pii/S2214317316300154> (visited on 11/11/2017).
- Suzuki, Nobuhiro and Ron Mittler (2006). “Reactive oxygen species and temperature stresses: A delicate balance between signaling and destruction”. In: *Physiologia Plantarum* 126.1, pp. 45–51. ISSN: 1399-3054. DOI: 10.1111/j.0031-9317.2005.00582.x. URL: <http://onlinelibrary.wiley.com/doi/10.1111/j.0031-9317.2005.00582.x/abstract> (visited on 26/11/2017).
- Waller, Frank et al. (2005). “The endophytic fungus *Piriformospora indica* reprograms barley to salt-stress tolerance, disease resistance, and higher yield”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.38, pp. 13386–13391. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0504423102. URL: <http://www.pnas.org/content/102/38/13386> (visited on 12/11/2017).
- Wood, R. K. S. (1996). “Sustainable agriculture: the role of plant pathology”. In: *Canadian Journal of Plant Pathology* 18.2, pp. 141–144. ISSN: 0706-0661. DOI: 10.1080/07060669609500638. URL: <http://dx.doi.org/10.1080/07060669609500638> (visited on 12/11/2017).

- Woodham-Smith, C. (1962). "The Great Hunger: Ireland 1845-9." In: *The Great Hunger: Ireland 1845-9*. URL: <https://www.cabdirect.org/cabdirect/abstract/19631403835> (visited on 12/11/2017).
- Zheng, Xiao-yu et al. (2012). "Coronatine Promotes *Pseudomonas syringae* Virulence in Plants by Activating a Signaling Cascade that Inhibits Salicylic Acid Accumulation". In: *Cell Host & Microbe* 11.6, pp. 587–596. ISSN: 1931-3128. DOI: 10.1016/j.chom.2012.04.014. URL: <http://www.sciencedirect.com/science/article/pii/S1931312812001606> (visited on 21/11/2017).