
UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

# Computer imaging strategies for feature-rich quantitative analysis of plant pathogens

Master Thesis
EmPLANT Master Programme in Plant Breeding
2019 – 2021

Student:
Eric Javier Prendes Rodríguez

Supervised by:
Dr. Ana María Fita Fernández and Dr. Salim Bourras
Universitat Politècnica de València and
Sveriges lantbruksuniversitet

September 2021

emPLANT
master to sucseed

SLU

**ABSTRACT**

Accurate phenotypic detection, and quantification of plant pathogens is an important challenge in plant sciences. Thanks to the rapid development of scalable digital technologies, and more specifically computer imaging, new and precise methods are emerging to tackle this challenge with application both in model organisms as well as agricultural crops. This constitutes an asset for researchers and for breeders, who often must rely on time-consuming manual phenotyping or scoring by eye. Currently there are multiple tools that can be used for this purpose, two of them being ImageJ and PlantCV, both of which have become well recognized, scientific-grade software packages. The first, is possibly the most used image analysis software in biological sciences, and the latter is a more recent yet very powerful OpenCV-derived image processing toolkit for plant phenotyping analysis. Despite the availability of these tools as opensource software, their use and deployment are mainly restricted by the challenge of obtaining data that is at the same time insightful, and biologically meaningful to infer useful and correct results. In this work, we present several pipelines and methodologies for image-based biological data extraction and analysis in two fungal pathogens, *Neonectria ditissima*, a tree pathogen, and *Blumeria graminis* f. sp*. tritici*, the causal agent of the wheat powdery mildew disease. This work relies of two training datasets generated from *in vitro* and *in planta* assays and applied to both ImageJ and PlantCV. These approaches are here technically assessed for their usability in extracting feature-rich biological information, and the generated data is evaluated for its suitability for quantitative analyses ranging from basic fungal growth dynamics to complex QTL mapping. As a result, a workflow for the image-based quantitative analysis of *in planta* inoculation assay of wheat leaves with *Blumeria graminis* f. sp*. tritici* using PlantCV and Python was developed, being obtained six different disease related features. Also, a pipeline for the extraction of 50 different image feature from *Neonectria ditissima* growing essays images was implemented into the ImageJ software. This allowed for the generation of a profile of the different isolates based on the extracted features, which shows the potential of these approach for its use in image-based profiling applications.

**Resumen**

La detección fenotípica precisa y la cuantificación de fitopatógenos es un desafío importante en las ciencias vegetales. Gracias al rápido desarrollo de las tecnologías

digitales y, más específicamente, a las técnicas de imágenes por computadora, están surgiendo métodos nuevos y precisos para abordar este desafío con aplicaciones tanto en organismos modelo como en cultivos agrícolas. Esto constituye un recurso muy valioso para los investigadores y para los mejoradores, que a menudo deben depender del fenotipado manual y el puntaje visual, técnicas las cuales consumen mucho tiempo. Actualmente existen múltiples herramientas que se pueden utilizar para este propósito, dos de ellas son ImageJ y PlantCV, las cuales se han convertido en paquetes de software de grado científico bien reconocidos. El primero es posiblemente el software de análisis de imágenes más utilizado en las ciencias biológicas, y el segundo es un conjunto de herramientas de procesamiento de imágenes derivadas de OpenCV que es más reciente, pero con muchas potencialidades para el análisis cuantitativo y el fenotipado digital en plantas. A pesar de la disponibilidad de estas herramientas como software de código abierto, su uso e implementación está restringido principalmente por el desafío que supone obtener datos que sean al mismo tiempo intuitivos y biológicamente significativos para inferir resultados útiles y correctos. En este trabajo, presentamos varias metodologías para la extracción y análisis de datos biológicos basados en imágenes de dos patógenos fúngicos, *Neonectria ditissima*, un patógeno de árboles, y *Blumeria graminis* f.sp. *tritici*, el agente causal de la enfermedad del mildiú polvoriento del trigo. Este trabajo se basa en dos conjuntos de datos de entrenamiento generados a partir de ensayos *in vitro* e *in planta*, y usando tanto ImageJ como PlantCV. Las metodologías propuestas se evalúan técnicamente para su usabilidad en la extracción de información biológica rica en características, y los datos generados se evalúan para determinar su idoneidad para análisis cuantitativos que van desde la dinámica básica del crecimiento de hongos hasta el mapeo de QTL complejos. Como resultado, se obtuvo e implementó una metodología para el análisis cuantitativo basado en imágenes de ensayos de inoculación en planta de hojas de trigo con *Blumeria graminis* f. sp. tritici utilizando PlantCV y Python, logrando ser extraídas seis caracteres diferentes relacionadas con la enfermedad. Además, se implementó en el software ImageJ un procedimiento para la extracción de 50 parámetros visuales de diferentes imágenes de ensayos de crecimiento en *Neonectria ditissima*. Esto permitió la generación de un perfil de los diferentes aislamientos basado en los parámetros extraídos, lo que muestra el potencial de este enfoque para su uso en aplicaciones de creación de perfiles basados en imágenes.

**Resum**

La detecció fenotípica precisa i la quantificació de patògens vegetals són un repte important en ciències de les plantes. Gràcies al ràpid desenvolupament de tecnologies

digitals escalables i, més específicament, per imatge per ordinador, sorgeixen mètodes nous i precisos per fer front a aquest repte tant amb aplicacions en organismes model com en cultius agrícoles. Això constitueix un actiu per als investigadors i per als criadors, que sovint han de confiar en un fenotipat manual que consumeix molt de temps o en la puntuació a l'ull. Actualment hi ha diverses eines que es poden utilitzar per a aquest propòsit, dues d'elles són ImageJ i PlantCV, totes dues convertides en paquets de programa de reconegut prestigi científic. El primer, és possiblement el programa d'anàlisi d'imatges més utilitzat en ciències biològiques, i el segon és un conjunt d'eines de processament d'imatges derivat d'OpenCV més recent però molt potent per a l'anàlisi del fenotipatge de plantes. Tot i la disponibilitat d'aquestes eines com a programa de codi obert, el seu ús i desplegament es veuen principalment restringits pel repte d'obtenir dades que siguin alhora intel·ligents i biològicament significatives per inferir resultats útils i correctes. En aquest treball, presentem diversos fluxes de treball i metodologies per a l'extracció i anàlisi de dades biològiques basades en imatges en dos fongs patògens, Neonectria ditissima i Blumeria graminis f. sp. Tritici ( l'agent causal de la malaltia del míldiu en blat). Aquest treball es basa en dos conjunts de dades de formació generats a partir d'assaigs in vitro i en planta, i aplicats tant a ImageJ com a PlantCV. Aquests enfocaments s'avaluen tècnicament per la seva utilitat en l'extracció d'informació biològica rica en característiques, i les dades generades s'avaluen per la seva idoneïtat per a anàlisis quantitatives que van des de la dinàmica bàsica de creixement dels fongs fins al mapatge QTL complex. Com a resultat, es va obtenir i implementar una metodologia per a l'anàlisi quantitativa basat en imatges d'assajos d'inoculació en planta de fulles de blat amb Blumeria graminis f. sp. tritici utilitzant PlantCV i Python, aconseguint ser extretes 6 caràcters diferents relacionades amb la malaltia. A més, es va implementar en el programari ImageJ un procediment per a l'extracció de 50 paràmetres visuals de diferents imatges d'assajos de creixement en Neonectria ditissima. Això va permetre la generació d'un perfil dels diferents aïllaments basat en els paràmetres extrets, el que mostra el potencial d'aquest enfocament per al seu ús en aplicacions de creació de perfils basats en imatges.

**ACKNOWLEDGEMENTS**

I would like to express my profound acknowledgment to Dr. Salim Bourras and Dr. Ana María Fita Fernández for their support and advice along every step of my project and for giving me the opportunity to perform my thesis with them. Special thanks to Marion Müller and to Kerstin Dalman for kindly providing the data upon which I've worked.

In addition, I express my deepest appreciation to the EmPLANT Master Program coordinators in France and Spain for their help during the two years of my degree. Thanks to all my teachers and instructors, for sharing their valuable knowledge and experience during this time.

I would like also to thank my family and friends for being an endless source of inspiration and support even in the distance, and last but not least, thanks to my dear Dominika Otto for his moral support and company through ups and downs.

# Table of Contents

## INTRODUCTION

The advent, large-scale adoption, and mass-utilization of Next Generation Sequencing (NGS) technologies has led to an exponential accumulation of genomic and genotypic data. Nowadays there is a gap between the amount and pace of generation of genotypic vs phenotypic information, which has become the limiting factor when it comes to understanding the genetic basis of complex traits (i.e. linking genotypes to phenotypes). Overcoming this challenge, known as the "phenotypical bottleneck", is of paramount importance to improve the efficiency of molecular breeding (Zhao *et al.*, 2019). Recent advances in disciplines such as electronics, robotics, and computer sciences led to the development of a large set of tools with which to carry out quantitative phenotypic analyzes with high resolution and in a non-destructive way (Kumar, Kumar and Pratap, 2015; Tardieu *et al.*, 2017). However, the challenge remains to develop and implement new analytical tools that allow transforming this data into knowledge that can be used to infer genetic traits and breeding value (Pieruschka and Schurr, 2019).

Among the types of data most used in phenotyping, digital images are perhaps the most cost-effective and popular mean to generate massive amounts of phenomic information. This is further facilitated by the availability of a large number of software tools that are at our disposal to analyze plant traits (Lobet, 2017).

The current work focuses on the development of digital imaging approaches for quantitative analyses of two plant pathogens *Blumeria graminis* f. sp. *tritici*, causal agent of the wheat powdery mildew disease, and *Neonectria ditissima*, the causal agent of the European apple canker. In the following section we will briefly introduce the history and biology of these two pathogens as well as the current state-of-the art regarding the use of digital images in plant science and its main challenges.

### *Blumeria graminis* f. sp*. tritici.* Biology and control

All powdery mildews are ascomycete fungi that belongs to the *Erysiphaceae* family of the *Erysiphales* order, which consists of 16 genera and about 650 species worldwide (Braun *et al.*, 2002). Its host range is strictly limited to angiosperms, since it has never been reported to infect either gymnosperms or other types of organisms (Amano, 1986) being species *Blumeria graminis* (DC.) Speer the single causal agent of powdery mildew disease of the cereals.

Based on the strict host specialization, Marchal (1902) proposed a classification for *B. graminis* based on seven *formae speciales* (ff.spp.) within the same species, although it

was proposed later to be extended to eight by Oku, Yamashita, Doi, & Nishihara (1985). However, it has been demonstrated since Eshed (1970) that the specialization could not be strict in some cases, given that some isolates of the ff.spp. *hordei*, *tritici* and *avenae* are compatible with wild grasses belonging to different genera in several tribes.

This is a pathogen with some characteristics that promote rapid spread and adaptation, including a short life cycle, presence of airborne spores that can be easily transported over long distances, and the potential for sexual recombination for generating new virulent races (Kang *et al.*, 2020). This pathogen provides two types of pathogenic inoculum for infection, asexual conidia and sexual ascospores, the latter being released from sexual/conservation structure known as chasmothecia, which enables the pathogen to survive in the absence of a living crop. Both conidia and ascospores are infectious and become ready to germinate when there is enough humidity in the environment, and it is also known that mild temperatures ranging from 10 to 22 °C can further enhance the infection process (Jankovics *et al.*, 2015).

According to Dean et al., (2012), powdery mildew, caused by *Blumeria graminis* f. sp. *tritici*, ranks sixth out of 10 most important fungal diseases in plant pathology. It is also considered to cause the eighth highest yield loss by plant diseases worldwide (Savary *et al.*, 2019), since it can occur year-round in the most important crop producing regions in the world, recording production losses of up to 35%, 62% and 40% in places like Russia, Brazil and China, respectively (Mehta, 2014).

Powdery mildew can be controlled by systemic fungicides as seed treatments (Reis *et al.*, 2016) and the use of biological control methods (Köhl *et al.*, 2019), but these methods have historically remained limited in success. Actually, the primary control method of powdery mildew in wheat is the use of resistant cultivars. Thus, the selection pressure applied by commonly grown cultivars carrying major resistance genes causes virulence shifts in the pathogen (Niewoehner and Leath, 1998) which requires continued breeding efforts in this sense.

These efforts led to the over 200 resistance genes and quantitative traits loci (QTL) for resistance (Kang *et al.*, 2020), thus empowering the use of marker-assisted selection for the breeding of powdery mildew resistant wheat (Nakaya and Isobe, 2012; Karbarz *et al.*, 2020).

Despite this, breeding for *B. graminis* resistance using conventional crossing still presents several challenges. One of them being the problematic migration of exotic resistance

alleles by means of conventional crossing in wheat, which might come with associated pleiotropic effects, because of deleterious genes that are associated with an inferior performance in yield and quality (Johnson, 1992; Groos *et al.*, 2003). This highlights the importance of obtaining reliable phenotyping data for identifying true trait-associated markers, a process that when done by visual scoring is very subjective and error-prone, especially in the case of large screenings (Poland and Nelson, 2011).

### *Neonectria ditissima*. Biology and control

*Neonectria ditissima* is a filamentous fungus and the causative agent of the disease known as the European canker in apple, although it can also infect a wide range of hardwood tree species (McCracken *et al.*, 2003; Florez *et al.*, 2020). From a taxonomic perspective, the apple canker fungus belongs to a well-defined taxonomic group possessing a perfect (ascosporic) and a *Cylindrocarpon* imperfect (conidial) state (Castlebury, Rossman and Hyten, 2006; Chaverri *et al.*, 2011). This classification is based on sexual and asexual reproduction morphotypes, as well as molecular phylogenetics studies, but still the scientific name of the apple canker has change repeatedly through the years (Tulasne and Tulasne, 1865; Hartig, 1889; Rossman and Palm-Hernández, 2008).

*N. ditissima* has a wide host range that includes agronomically important species such as apple (*Malus*) and pear (*Pyrus*), as well as numerous broad-leaved trees such as *Alnus*, *Betula*, *Crataegus*, *Fagus*, *Fraxinus*, *Ilex*, *Juglans*, *Populus*, *Quercus*, *Ribes*, *Salix*, *Tilia* and *Ulmus* (Flack and Swinburne, 1968; Farr *et al.*, 1989). The infection of apple trees is more damaging than pear trees (Weber, 2014), as apple canker is responsible for important yield losses through direct infection of the wood and pre- and post-harvest rotting of the fruits (Xu and Robinson, 2010).*N. ditissima* is difficult to control and represents a major challenges to the profitability of apple production n in Europe, and many other parts of the world where this disease has also been recorded (Latorre *et al.*, 2002; Plante, Hamelin and Bernier, 2002; Campos *et al.*, 2017).

This pathogen produces two types of conidia in both nature and lab conditions, both of them are released passively by rain splash (Booth, 1966). Its ascospores are ellipsoidal and one-septate, and are released and dispersed to the air after being explosively discharge by the asci (Weber, 2014). It is thought that the relative contribution of conidia and ascospores to the dispersion of the disease varies with the climatic conditions, where ascospores are associated to maritime environments (Graf, 1976) while conidia a found more in arid ones (Dubin and English, 1975). The main sign of an infection by *N. ditissima* is a pale brown epidermis, that turned pale brown while the bark tissue becomes necrotic

and gradually dries out. These symptoms are most frequently seen in spring from flowering time onwards, although they may appear at any time of year at temperatures above freezing. Within a few weeks, white or pale yellow superficial conidial pustules (sporodochia) representing the *Cylindrocarpon* state develop on the surface of the dead bark (Weber, 2014).

The control of the apple canker is done mainly by the repeated and thorough pruning of canker wounds, combined with the application of fungicides (Cooke, 1999; Palm, 2009). As the diseases spreads, and fungicide efficiency being limited, much work is been done on practical ways to improve sanitation measures in nurseries (Weber, 2014).

This is a well characterized pathogen from the physiological and genetic standpoint, with the mechanism of infection of the woody tissues and the fruit having being elucidated (Gelain *et al.*, 2020) as well as its genetics and the genetics of the resistance to the disease (Gómez-Cortecero, Harrison and Armitage, 2015; Bus *et al.*, 2019). Regarding this, there are substantial variations among apple varieties in their susceptibility to this disease, variation that seems to be based on partial resistance (Garkava-Gustavsson *et al.*, 2013). So far, monogenic resistance to canker has not yet been found, but the existence of differential resistance between apple cultivars allows for pyramiding resistance genes into new cultivars, which offers a promising strategy of control (Weber, 2014). However, since the breeding cycle of the apple is very long (30 – 50 years), the success of any program of breeding for resistance to *N. ditissima* requires foresight, and in this regard genomic studies and in vitro analyses could provide important information to unravel durable sources or resistance and aid durable disease management strategies.

**Digital imaging in plant pathology and in microbiology: main challenges**

Among the types of data most used in plant science, digital images are perhaps the most cost-effective and popular mean to generate massive amounts of phenomics information.. In plant pathology, leaves are probably the most common organ targeted by imaging strategies , partly because of how relatively easy these images can be acquired (e.g compared to roots), the low level of dimensional complexity (i.e. 2D images are sufficient),and finally because leaves are the organ where the symptoms of many plant pathogens are the most easily distinguishable..

When it comes to plant leaf diseases detection and analysis, many methods and approaches have been described and discussed (Raichaudhuri and Sharma, 2016). In Gavhale and Gawande (2014), for example, a summary of different image processing

techniques has been discussed for detecting leaf diseases for several plants. Detection methods which use texture features such as BPNN, SVM, k-means clustering, and SGDM are discussed in detail. While there is a variety of methods and approaches to tackle the problem, the general steps that needs to be undergone for detecting and analyzing plant diseases in leaf images are mainly four: (i), image acquisition (ii) image pre-processing, (iii) feature extraction, and last (iv) classification (Ngugi, Abelwahab and Abo-Zahhad, 2021)

Two of the most used tools in this regard are ImageJ and PlantCV (Abràmoff, Magalhães and Ram, 2004; Fahlgren *et al.*, 2015; Gehan *et al.*, 2017). The strengths of the first lie it's multipurpose image analysis capabilities, in addition to a graphical user interface (GUI) that facilitates its use (to non programers) with an extensive number of plugins that extend its functionalities (Collins, 2007). , The second (PlantCV) is a suite of basic tools that the community could build upon (i.e. requires a higher level of user expertise). It is written in the Python, a high-level scripting language widely used in the scientific community including bioinformatics. Thus, PlantCV offer a high level of customization and versatility based on the many tools available from the Python scientific computing community(Gehan *et al.*, 2017).

Digital images also constitute an important pillar in microbiology, where multiple processing and analysis tools and approaches has been created in the last decades (Puchkov, 2016). Among its most immediate applications at the macroscopic level we can cite colony counting and identification, as well as applications in physiology, biochemistry, and molecular biology (Dörge, Carstensen and Frisvad, 2000; Marotz, Lübbert and Eisenbeiß, 2001; Yang *et al.*, 2001; Ogawa *et al.*, 2012).

Despite all this, many challenges remain for computer vision in plant science and microbiology that restricts their use and deployment. One of them is the challenge of obtaining data that is at the same time insightful, and biologically meaningful to infer useful and usable results (Li *et al.*, 2020). Another challenge is the need to analyze images at a very high throughput but still in a robust and accurate manner in order to cope with the increasing amount of data being generated by the "omics" disciplines (Scharr *et al.*, 2016).

**Image segmentation and feature extraction: main methods and approaches.**

As mentioned before, image segmentation, or pixel classification, is usually one of the first steps in any biological computer image analysis, and also one of the most critical ones, since an accurate and efficient image segmentation can rapidly and accurately obtain phenotypic traits (Hamuda, Glavin and Jones, 2016). In plant disease recognition applications, segmentation is a twofold process. Segmentation is first done to isolate the

leaf, fruit, or flower from the background. A second segmentation is then done to isolate healthy tissue from diseased tissue. However, classifying pixels has several limitations. One of the main ones being the difficulty to obtain consistent image data, since there is an intrinsic heterogeneity in images datasets (i.e., shadows, color distribution) (Lee *et al.*, 2018).

Historically, thresholding algorithms have been the most common method to segment images in plant sciences. However, these can rapidly show their limits when two or more visually similar phenotypic classes must be distinguished as these methods perform poorly when it comes to colors and are very susceptible to variations and inconsistencies that may be present in large sets of images. This becomes exceedingly more complex in anopen field because of light variations, plant movements, crop heterogeneity,surface reflectivity (accumulation of moisture), and many other factors (Tsaftaris, Minervini and Scharr, 2016).

Methods that utilize machine learning techniques are a promising approach to tackle these problems (Minervini, Abdelsamea and Tsaftaris, 2014; Singh *et al.*, 2016; Pound *et al.*, 2017). Consequently, a naive Bayes classifier tool has been integrated to PlantCV for machine-learning based pixel classification since version 2.0 (Abbasi and Fahlgren, 2017). Likewise, ImageJ now also includes the Trainable Weka Segmentation, a machine learning tool aimed at pixel segmentation for microscopy images (Arganda-Carreras *et al.*, 2017). In addition, we also have Ilastik, an easy-to-use interactive tool with the ambitious objective to bring machine-learning-based (bio) image analysis to users (with no previous knowledge in machine learning or programming) to apply basic image segmentation and classification (Berg *et al.*, 2019).

These methods are also being used to automatically classify the output of image analyses, since they can be trained to categorize features associated with each disease to be recognized. The trained algorithm can then be used to recognize features from new images captured from the field. Classification deals with matching a given input feature vector with one of the distinct classes learned during training (Ngugi, Abelwahab and Abo-Zahhad, 2021)

In this work we present different approaches for the extraction of biological data wheat leaves infected with the fungal pathogen *Blumeria graminis* f. sp*. tritici*, the causal agent of the wheat powdery mildew disease. We use the PlantCV image analysis tool for pixel classification and data extraction, and we analyze the suitability of the extracted data for its use in genetic analysis as well as compare it with the data obtained using a similar approach in two other image analysis tools, namely ImageJ and ilastik.

**OBJECTIVES**

1. To develop an image-based quantitative analysis workflow of lab-controlled infections ofwheat leaves with *Blumeria graminis* f. sp*. tritici* using PlantCV and Python*.* This will include:
    a. Training a naïve Bayes classifier using a sample dataset as an input
    b. Image segmentation into three categories using the trained classifier
    c. Post-processing of the segmented image and feature extraction
    d. Qualitative and quantitative analyses of the output
2. To develop an image-based feature rich quantitative analysis workflow of *Neonectria ditissima* isolates grown in three different types of culture media. This will include:
    a. Implementation of a multiple feature extraction pipeline from images of *N. ditissima* isolates growth assays
    b. Qualitative and quantitative analysis of the output
    c. Generation of an imaged-based colony profile

## MATERIALS AND METHODS

As a guideline for the development of the approaches presented in this work, a modified version of the general workflow described in Raichaudhuri and Sharma (2016), shown in Fig. 1 below was used.

The second step of this workflow is usually image pre-processing, a step which was removed from our *B. graminis* f. sp*. tritici* quantitative analysis approach. This is becasue the use of a machine learning method like naïve Bayes classification, combined with the characteristics of the image set that was used (same camera, same settings, same background, etc.), made it unnecessary to carry any further procedure to improve image quality after the image acquisition stage. The classification step was also not included in the developed approach since the main goal of the workflow is to extract features and to generate an output that could be further analyzed.

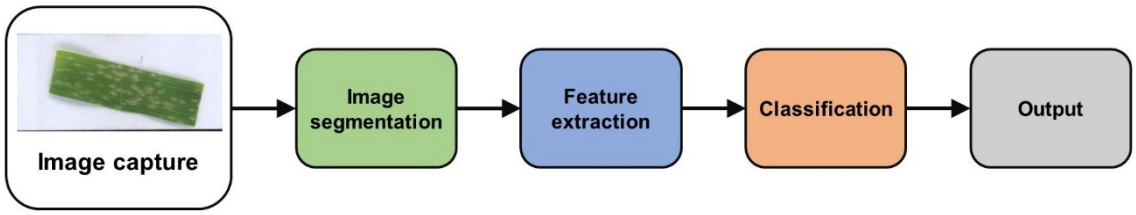



***Figure 1****. General workflow used as a guideline for the development of our approaches. Image capture involves the collection of photographic information with the use of a camera. Segmentation is generally done twice in plant disease recognition applications, first for isolating the plant material and second to isolate healthy tissue from diseased tissue. The use of the naïve Bayes Classifier allows for segmenting the image in just one step.*

For the *B. graminis* analysis, infected wheat leaves images were obtained from Marion Müller (personal data) and generated as previously described by Müller *et al.* (2021). *Neonectria ditissima* growing assay images were kindly provided by Kerstin Dalman, Department of Molecular Sciences, SLU. Image analyses were performed with the use of PlantCV v3.12.0, OpenCV v3.4.14, ImageJ v1.53c and iLastik v1.3.3 (Collins, 2007; Gehan *et al.*, 2017; Berg *et al.*, 2019). Statistical analyses and graphics were done using RStudio version 1.2.5001 with R version 3.6.1 and the following packages: ggplot 2.2.1, ggthemes 3.4.0, reshape2 1.4.3, plyr 1.8.4, car 3.0-11 as well as R base functions (R Core Team, 2021).

### *Blumeria graminis* f. sp*. tritici in planta* assay images analysis

#### Naïve bayes classifier training

The naïve Bayes classifier embedded in PlantCV was trained to generate predictors in form of Probability Distribution Functions (PDFs) that will allow to classify/segment the images.

During the training phase, a script included in PlantCV called “plantcv-train.py” was used. This script converts pixel RGB values for each input class into the hue, saturation, and value (HSV) color space. Kernel density estimation (KDE) is used to calculate a probability density function (PDF) from a vector of values for each HSV channel from each class. The output PDFs are used to parameterize the naive Bayes classifier function. To generate a training subset a set of images was picked randomly from the dataset using a command shell script. The training set size was chosen after analyzing the resulting Probability Distribution Functions (PDFs) for several set sizes and choosing a size and training set that renders PDFs curves with well-defined peaks and little overlapping.

The input for the training algorithm consisted of a tab-delimited table where each column is a class (minimum two) and each cell is a comma-separated list of RGB pixel values from the column class was created using the Pixel Inspection tool on ImageJ, being the proposed classes “background”, “healthy tissue/plant” and “infected tissue”. Once the training table is generated, it is used as input data for the ‘plantcv-train.py’ script to generate PDFs for each class. The optimal number of pixels per class for the training set was also determined after analyzing the resulting PDFs for several number of pixels.

**Image segmentation, postprocessing and feature extraction**

Images were initially segmented using the trained classifier and the three classes described above. For visualization purposes the possibility of showing a false color segmented version of the input image was added to the workflow.

To enhance the segmentation, a postprocessing stage was introduced by using PlantCV noise reduction tool (*fill* function) and by cropping the edges of the leaves where most of the false positive pixels (pixels misclassified as “infected tissue”) were found. Feature extraction was then performed using some of PlantCV included functions and some OpenCV functions. The extracted features proposed for the developed approach where: number of colonial formations, ratio “Disease tissue/healthy tissue”, total diseased area, average colony formations area, total colony formation perimeter, and average colony formation perimeter.

A similar approach was implemented with the use of the Trainable Weka Segmentation tool plugin of ImageJ (Arganda-Carreras et al., 2017) and also the iLastik software in order to compare the quality of the image segmentation without the postprocessing across the three tools. For this purpose, an example set of 20 images was taken from the dataset and was subjected to segmentation using the three tools. The effectiveness of the output was

later qualitatively and quantitatively assessed by measuring the error percentage of false positives pixels (pixels classified as “infected”) in completely healthy images.

### Output analysis

The developed workflow was implemented into a Python script and executed on a dataset comprising 9646 images (Marion Müller, University of Zurich, personal data). The output was stored in a .csv file and analyzed statistically from the descriptive point of view for normality and outliers. For both analyses, data was appropriately log-transformed, and a Q-Q (Figure 7) plot and a Kolmogorov-Smirnov normality test was used for qualitative and quantitative normality evaluation while outliers were detected with IQR. The presence of false positives (pixels wrongly classified as disease) in the approach was evaluated by implementing it on a set of images comprising only healthy plants and obtaining the percentage of false positive pixels. This was done also using the Trainable Weka Segmentation tool and Ilastik to compare the tendency of each approach to render false positives.

### *Neonectria ditissima in vitro* images: quantitative analysis workflow

For the analysis of *N. ditissima in vitro* images, the general guideline for the development of the approach was the same as mentioned for the previous analysis (Raichaudhuri and Sharma, 2016). The dataset consisted of images of 39 isolates of *N. ditissima* from which two replicates of each were grown *in vitro* using three different culture media and photographed at five different timepoints (Figure 2 – Kerstin Dalman, Department of

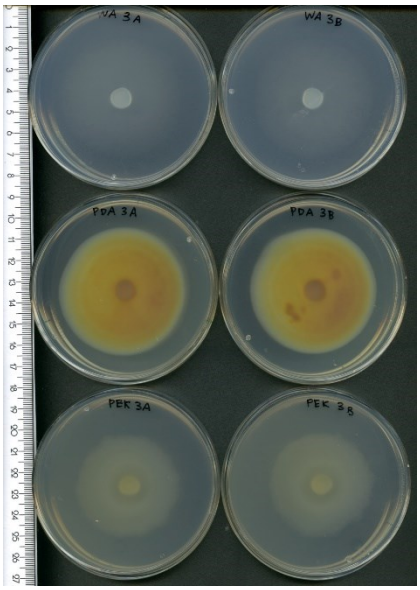



***Figure 2:*** *Example image from the dataset utilized for the analysis. Two replicas of each isolate were grown on each culture media and pictures were taken at five different moments. Image by Kerstin Dalman, Department of Molecular Sciences, SLU*

Molecular Sciences, SLU). The media utilized were potato dextrose agar (PDA), water agar (WA) and apple sap amended water agar (ASAWA), described by Amponsah, Walter and Scheper (2014)..

**Image segmentation and ROI determination**

Due to the characteristics of the images, and the relatively small size of the dataset (e.g. compared to the wheat powdery mildew dataset), ROIs where manually selected and the selection stored in multiple files for its use in the analyses. The selection was done using the circle tool on ImageJ given that the shape of the colonies was consistent amongst the isolates and the culture media.

**Features extraction**

For the feature extraction, a putative set of image parameters were first reviewed for its use in this analysis. The features ranged from the typical variables used in image analysis like intensity to RGB composition to more specific ones like Haralick's texture parameters and roughness (Haralick, Dinstein and Shanmugam, 1973; Puchkov, 2016). Its possible biological significance was discussed and pipelines for their retrieval was developed using ImageJ included scripting language and multiple plugins. Images were split for each color channel and each one of the selected features was extracted from each individual channel. Growth area was measured for each time point in all the isolates and it was used for measuring the fungal pathogen growth dynamic. Mycelium growth areas were square-root transformed and growth rate was estimated as the slope of a linear regression for changes of area through time combining data for all replicates for a given isolate as describe by Estrada *et al.* (2014)

**Output analysis**

The workflow was implemented on an ImageJ script and executed in the whole dataset. The output was stored in a .csv file and was then analyzed statistically. The analyzes focused on describing the variability of the different features across the isolates and across the culture media, to characterize differences in such variability across media, and to identify the media substrate with the highest impact on such variability.

**Generation of an image-based colony profile**

After assessing the impact of the type of media on colony image features a profile was made based on clustering and grouping methodologies. For this purpose, and to make all the variables comparable, the data was scaled first. Then a distance matrix was first obtained by computing the dissimilarities amongst the isolates using the Euclidian method,

and then this matrix was used to cluster the isolates in groups. The clustering was done using hierarchical clustering analysis, which is an analysis based on distance matrix renders and generates a dendrogram with the relationships that could exist between the clusters in the dataset. Ward's method was use as the linkage method for this analysis

## RESULTS

### *Blumeria graminis* f. sp. *tritici* infection assays: images analysis workflow

A diagram representing the developed PlantCV image analysis workflow is shown in Figure 3. It comprises the following steps: first, raw image acquisition from inoculated plants taken 7 days post inoculation (dpi) followed by sampling and training of the classifier used for

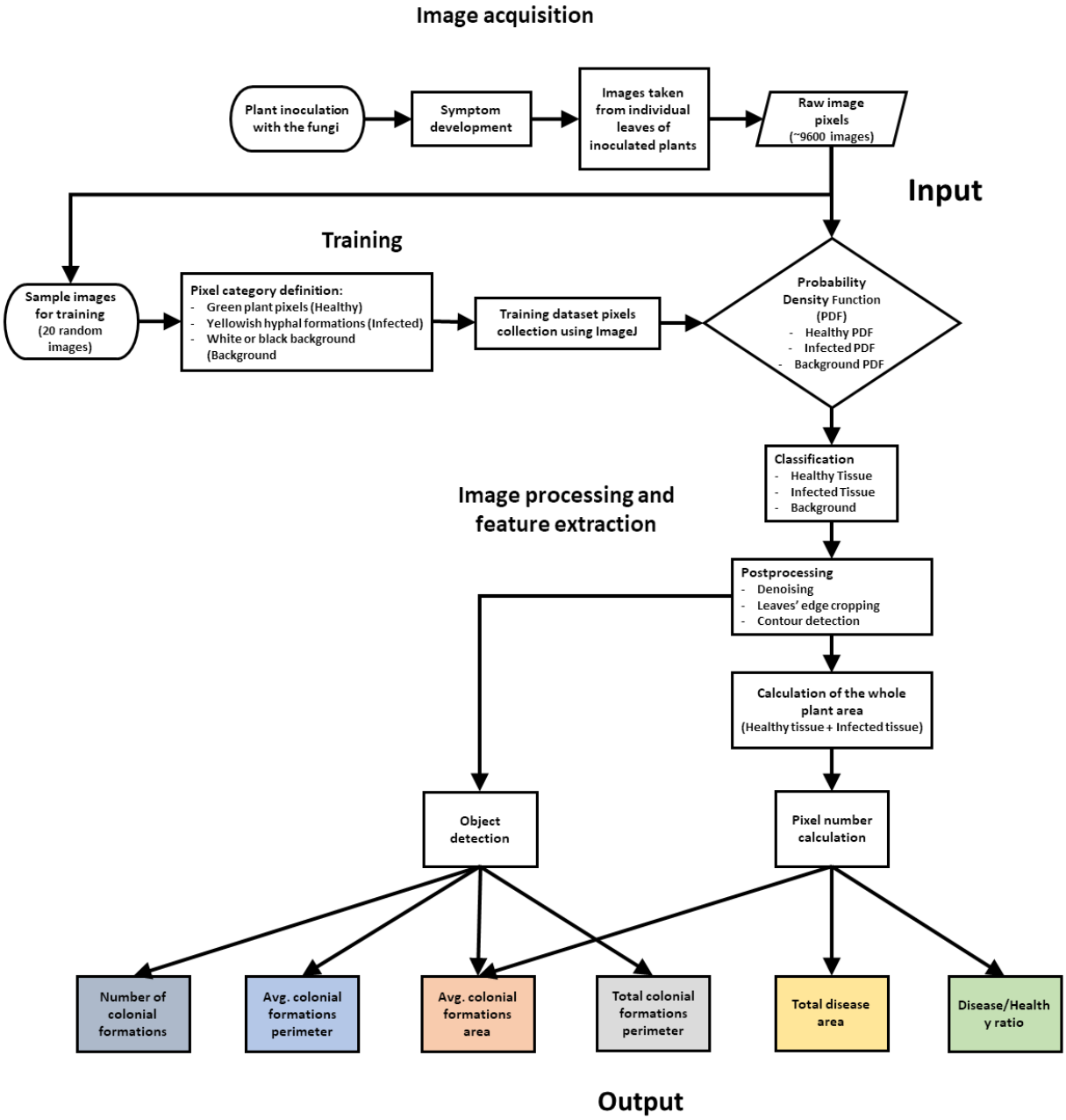



***Figure 3**. Image-based quantitative analysis workflow diagram. Plants were inoculated with the fungi. In the image acquisition section, individual leaves where photograph using the same background and camera at 7 dpi. For the training of the classifier, sample pixels of each the proposed class were taken. Three pixel-classifiers were defined and RGB information of 180 pixels for each class were collected to build the training dataset. The probability distribution function for each category was calculated based on that training dataset. In the testing session, each pixel from an input image was calculated and classified into each class. After postprocessing, the output produced six different metrics that would be utilized in further statistical analyses.*

segmentation. Then the segmentation stage using as an input the dataset and the PDFs generated in the training step. A post-processing step follows in which the segmentation is improved by denoising and edge-cropping, and in which contour and object detection functions are applied. Finally, features are extracted obtaining as an output a .csv file containing feature values and the sample ID for further analyses.

**Naïve Bayes classifier training**

For the training stage, after comparing datasets with 1, 5, 10, 20 and 40 images and 20, 40, 80, 120 and 180 pixels per class, a dataset of 20 images and 180 sample pixels per class was selected based on the characteristics of the PDFs it yielded (Figure 4). The latter showed well defined and non-overlapping probability curves for each one of the classes. The PDFs values resulting from this training dataset were stored in a text file to be used in the segmentation step.

**Image segmentation, postprocessing and feature extraction**

The output of the segmentation step for each image is a binary image for each class where the pixels are white if the probability of the pixel to be in the given class was the highest of all classes, or black if otherwise (Figure 5A). Segmentation/classification was quantitatively and qualitatively assessed, (i) by visual inspection of the false color segmented images generated from the binary images representing each class, and (ii) by the empirical discrepancy method of analyzing the number of misclassified pixels in relation to a reference segmentation. The reference segmentation used in this case was a sub-set of 20 completely healthy images and the average percentage of pixels misclassified as “diseased” was of 3.4% of the total when not using any post-processing, compared with the 7.3% when using iLastik and 4.8% when using ImageJ.

After postprocessing (Figure 5B) the misclassified pixels average percentage decreased down to 1.7% of the total plant tissue and with nearly zero misclassified “background” pixels. The postprocessing included first a denoise step using PlantCV’s *fill* function selectively applied to each class, to remove artifacts and glares that could be misclassified as “background” and to eliminate cluster of pixels in the leaf that were being misclassified as “infected”. In addition to this, and given that there was an abundance of misclassified pixels in the edges of the leaves(especially in the regions where the leaf was cut and necrosis starts to appear), a procedure for cropping edges of the leaves was implemented as a second postprocessing step.

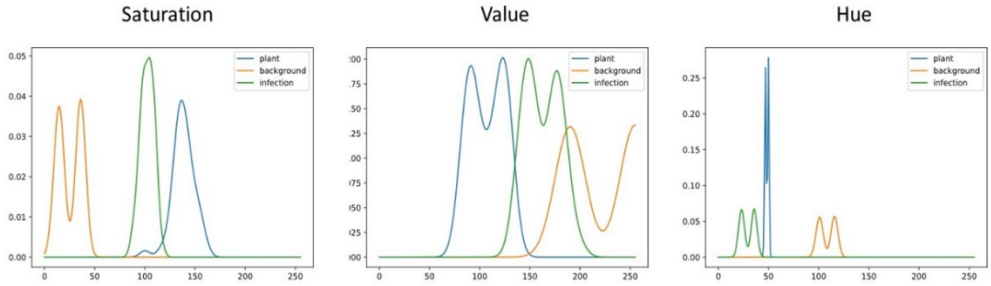
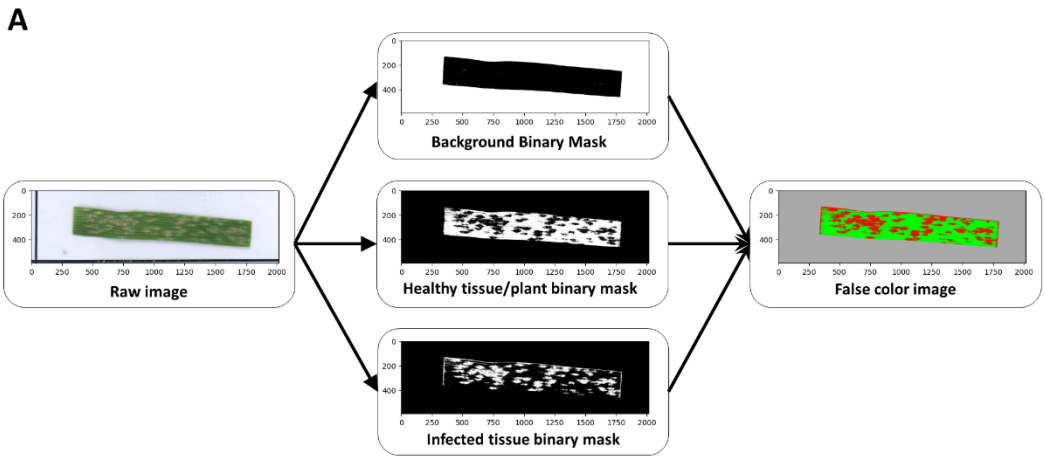



***Figure 4.*** *Classifier training output. Probability Distribution Functions (PDFs) obtained from the chosen training dataset of 20 images with 180 images per class. Each curve is the probability distribution for the saturation, value, or hue of a pixel for each of the given classes.*



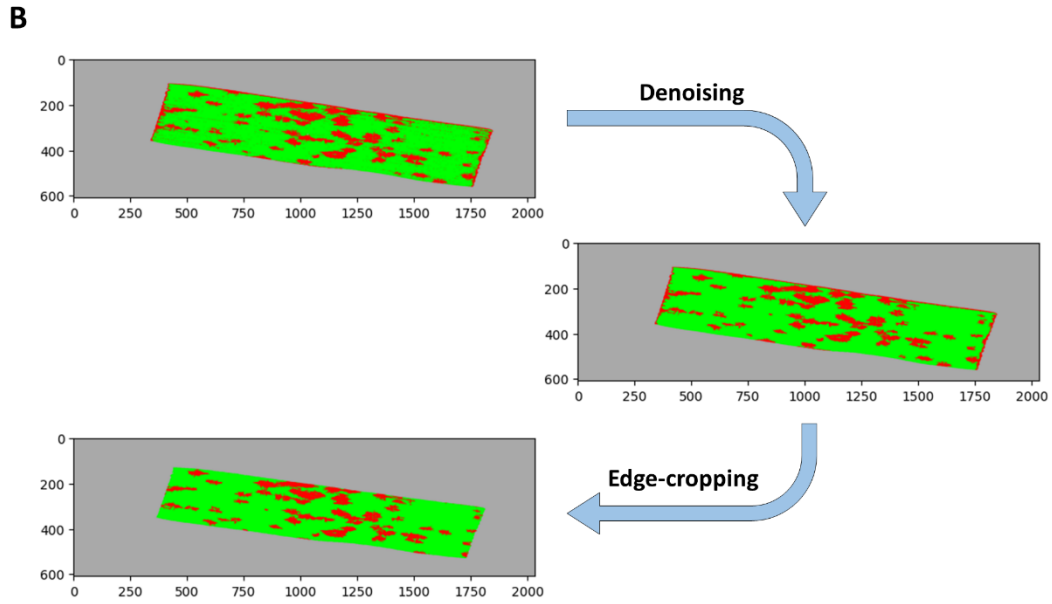




***Figure 5****. Image segmentation and postprocessing. A, Obtention of binary images representing each of the classes and generation of a false color segmented image with grey pixels representing the background, green ones the healthy plant tissue and red ones the infected tissue. B, Postprocessing stage is comprised of a denoising step first and followed by the trimming of the edges of the leaf where many false positive pixels can be found*

For the final step in the postprocessing, object contour detection was implemented (Figure 6). Since many of the colonies in the images are overlapping each other, the detected contours were those of contiguous colonies rather than of individual colonies. PlantCV function “cany_edge_detect” was used for detecting those contours and obtain a mask containing them. After this OpenCV function “dilate” was used to fill the gaps in the edge lines and optimize the resulting mask.

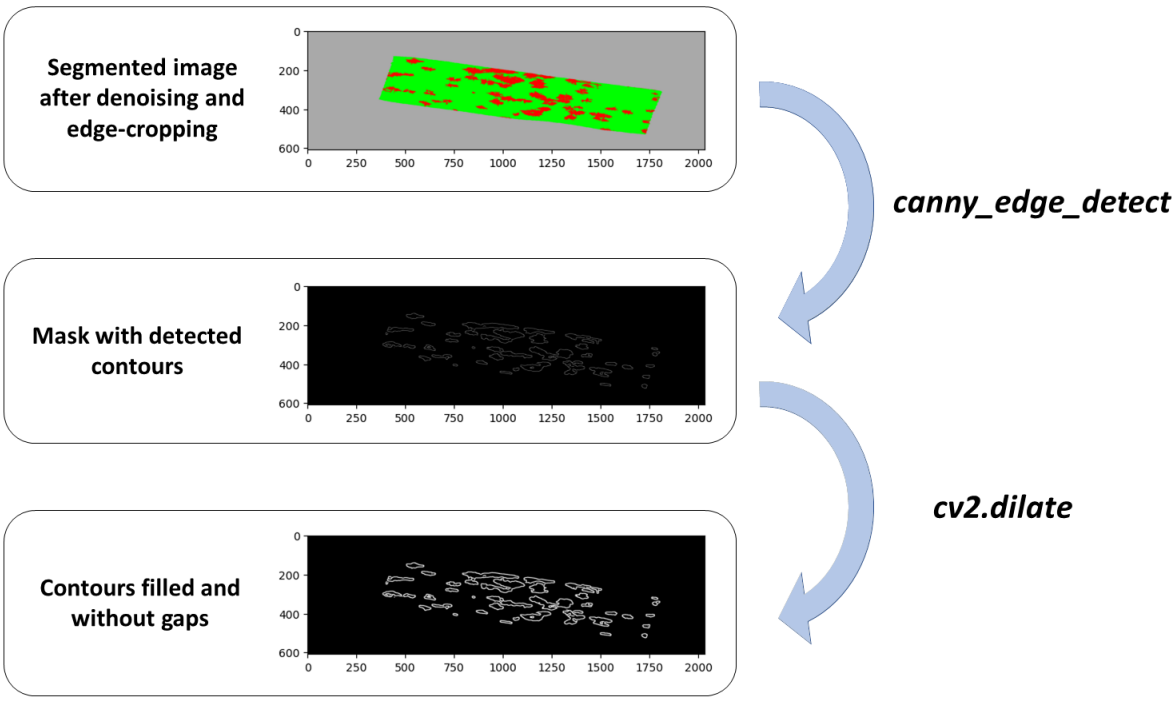



***Figure 6**. Final postprocessing step, object contour detection. This step was implemented by using the segmented image as an input and applying first the “canny_edge_detect” function in PlantCV to detect the contours and create a binary mask with them and then applying OpenCV “dilate” function to fill any probable gap on them.*

To calculate of the whole leaf area, the sum of the pixels belonging to the “healthy/plant” class and the “infected” class measured as in the previous step before feature extraction was obtained. The pixel count of the binary masks of each class was used directly for extracting the features “Total disease area” and ratio “Infected tissue/healthy tissue”. Altogether with the pixel number count, a series of OpenCV functions were applied to count the number of contour/objects in the image and to calculate the perimeter and the area of each (Figure 7). This allowed the extraction of four more features: “Number of objects/colonial formations”, “Total object/colonial formation perimeter”, “Average object/colony formation perimeter”, and “Average object/colony formation area”.

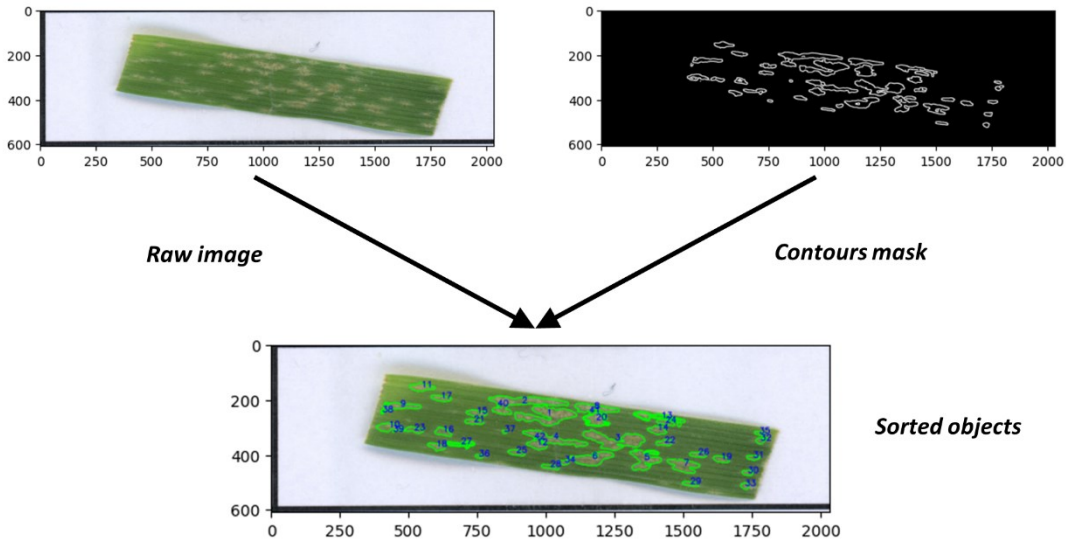



***Figure 7:*** *Sorting and counting of objects. By using the raw image and the contour mask generated previously it was possible to sort and count the objects in the original image, which allowed for the retrieval of features from the colonial formations*

## Output analysis

The generated output data was submitted to several analyses for evaluating its quality. Original plots showed a highly skewed data at first, skewness which was partially corrected with the log transformation. Although the Kolmogorov-Smirnov normality test showed that no feature follows a normal distribution, Q-Q plots showed for all the parameters a range of values for which normality is apparently true (Figure 8). The percentage of outliers for each variable/feature ranged from 1.3% of all the data points extracted in the feature "Number of colonial formations" to 12% in "Ratio Disease/Healthy".

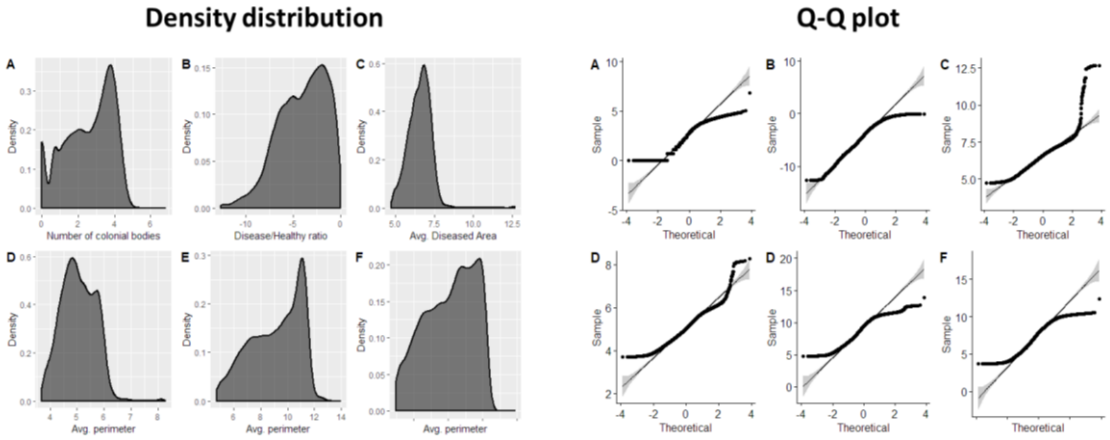


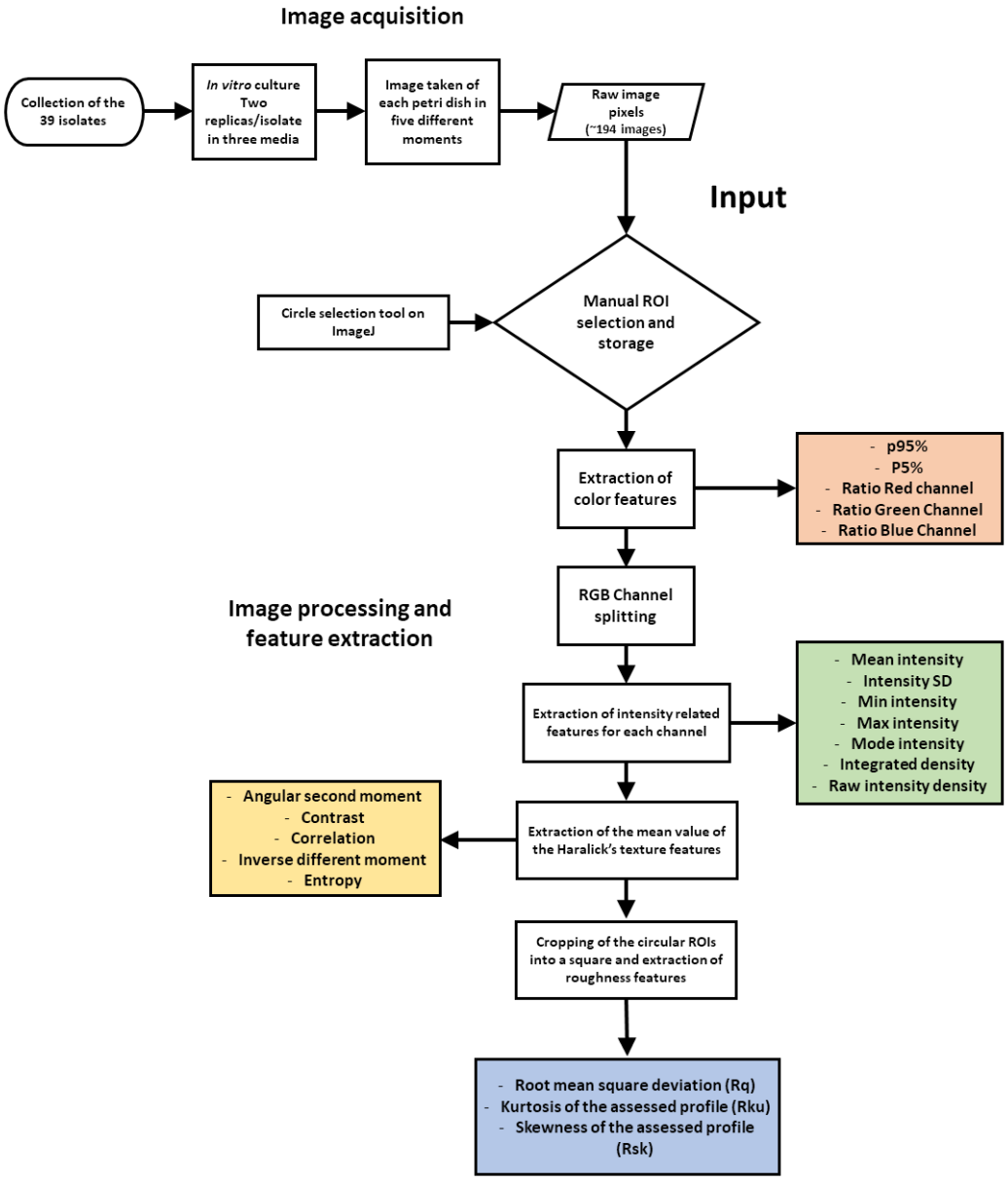

***Figure 8:*** *Normality evaluation of the output data. Plots representing the density distribution of the six output variables in the right and Q-Q plots showing the correlation between the distribution of these variables and a normal distribution. A. Number of colonial formations B. Ratio "Diseased/Healthy" C. Avg. colonial formation area D. Avg. colonial formations perimeter E. Total infected area D. Total colonial formations perimeter*

### *Neonectria ditissima* in *vitro* images quantitative analysis workflow

A flow diagram representing the developed ImageJ image analysis pipeline is shown on Figure 9. In this workflow, features were extracted at different moments as the pipeline progresses. The pipeline was implemented in a series of scripts written in ImageJ Macro Language (Mutterer and Rasband, 2012).

### Features extraction

The extracted parameters were included into four groups: color features, intensity related




features, Haralick's texture features, and roughness features. Except for the color features (which was the first parameter extracted), the features of the other three groups were

retrieved for each individual RGB channel. The result was a total of 50 variables that were extracted for each image and the data was stored in a .csv file. The growth rate was also obtained as the values of the slopes of a linear regression for changes of area through time. The r- values rangeed from 0.9212 to 0.9999 and have a median of 0.9914 which means that the linear regression fitted the data.

**Output analysis and profiling**

To visualize how dispersed and variable was each of the features amongst the three different culture media, the means and the standard deviations of all the variables where obtained. Mean values and standard deviation differ between treatments and between analyzed channels, being standard deviation values were always bigger higher in isolates grown on PDA media (Figure 10).

For the image-based profiling, the distant matrix was computed and the values where used for rendering a dendrogram and a heatmap representing the relationships that could exist amongst the clusters formed in the dataset (Figure 11).

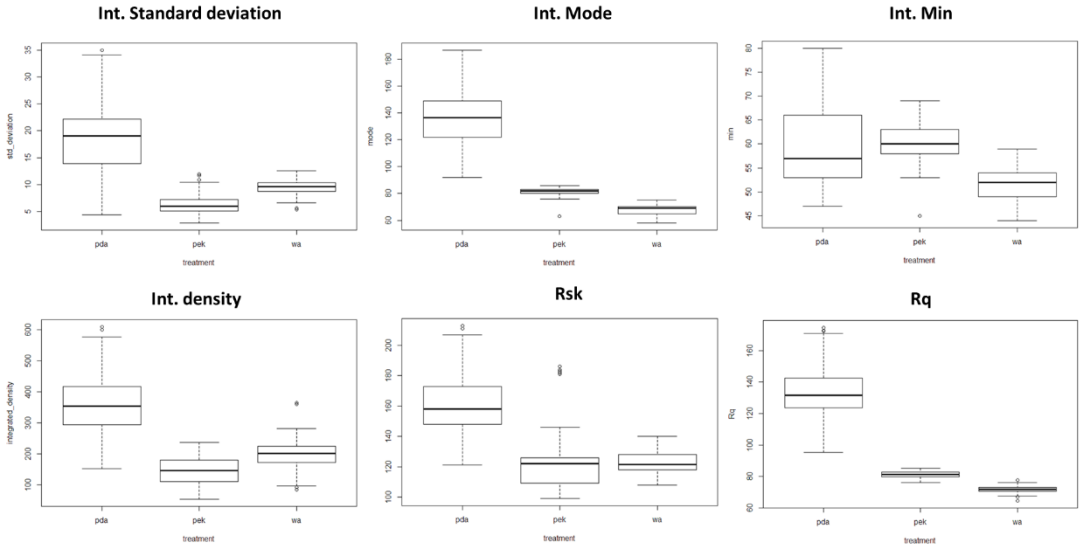



***Figure 10:*** *Boxplots representing the characteristics of the extracted data for six example features. PDA is the media that shows a higher SD in these features and in the ones not displayed here.*

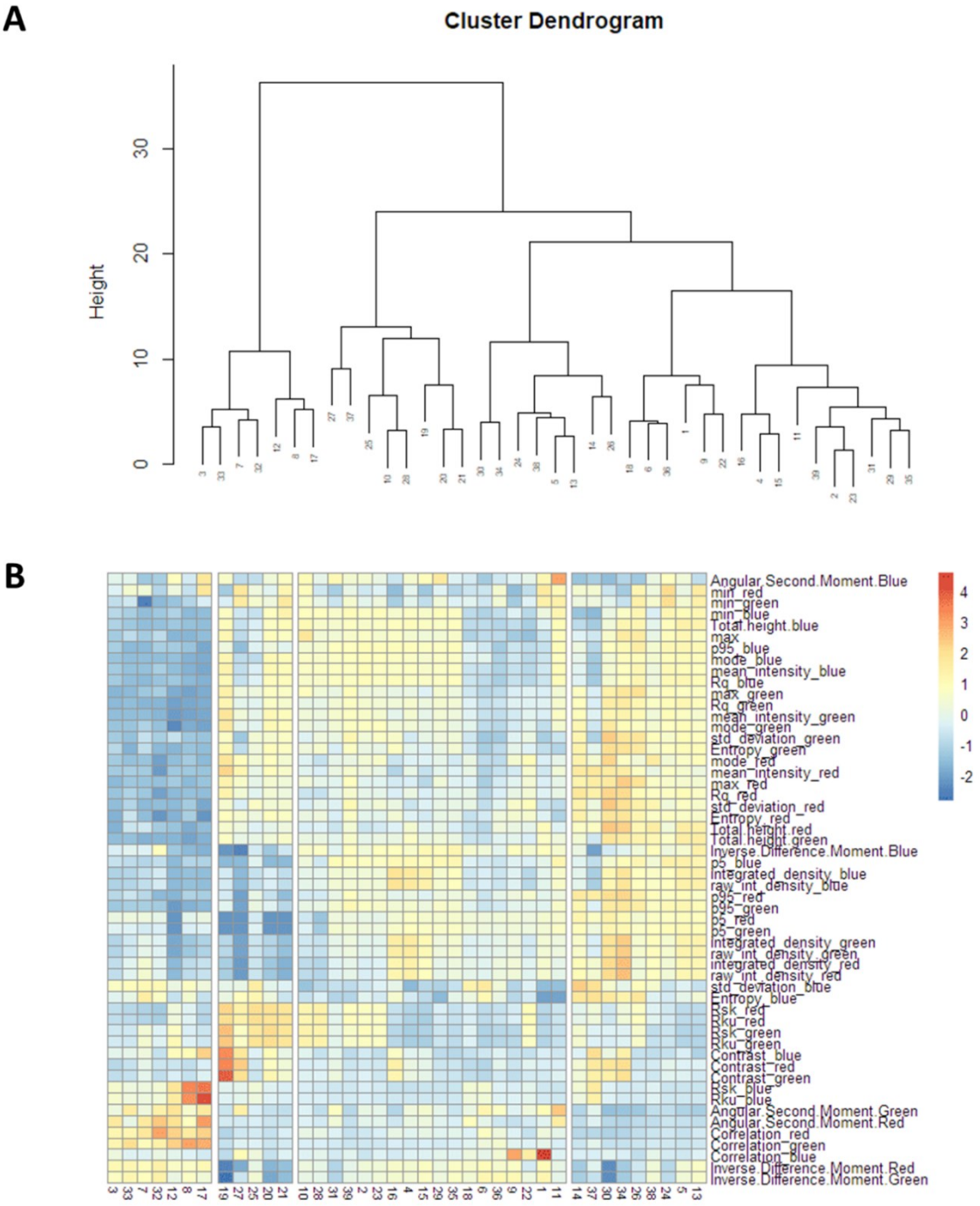



***Figure 11.*** *Clusters and image base profiling of the extracted features. A. Dendrogram representing the relationship between the isolates based on the distance between the extracted features. B. Representation of the distance and the clusters formed as a heat map.*

## DISCUSSION

### *Blumeria graminis* f. sp*. tritici in planta* assay images analysis workflow

The use of a naïve Bayes classifier as the main approach for pixel classification allowed a simpler segmentation step in this analysis, since this is generally done twofold in plant disease recognition applications, and the use of this approach allows for segmenting the image in just one step (Gavhale and Gawande, 2014). The PDFs used for the classifier showing non-overlapping curves demonstrate the consistence of the training outcome and its success in differentiating between the three defined classes. For this type of approach, it is very important to adequately capture the variation in the image dataset for each class to improve pixel classification (Gehan *et al.*, 2017). In a very consistent dataset, the number of sample images needed for generating the training text file is low. The evaluation of the classifier via PDFs curves is a crucial step in this sense, since it allows for defining a correct sample size for training,and it to forecast the quality of the pixel classification. The success of the segmentation is visible in the false color images generated by the PlantCV image analysis workflow, since they showed a convincing segmentation of the images before and after the postprocessing, that could also overcomethe false positives, artifacts, and other irregularities in the images (Figure 12A). Another advantage of this approach is that it enables the use of pipelines that are both simpler (fewer steps) and more flexible especially when comparing to the thresholding approaches, which are currently the most commonly used segmentation methods (Lomte and Janwale, 2017). A machine learning approach can likely be used for a variety of applications, such as identifying a plant under variable lighting conditions or quantifying specific areas of stress on a plant. To summarize, the naïve Bayes approach has several advantages, first, it allows for the simultaneous segmentation of two or more classes, second, since it is a probabilistic method, the segmentation is more robust across images than when using a thresholding method, and third, it replaces the multiple steps required in threshold-based pipelines, reducing its complexity.

When comparing the outcome of the segmentation without postprocessing between the three software tools (using images of non-infected plants as a control) we could the accuracy of the classification was visually clear and better witha PlantCV, and generally with a more defined background and no clusters of misclassified pixels (Figure 12B). When quantifying these differences, the PlantCV workflow showed an error margin of 3.4% of

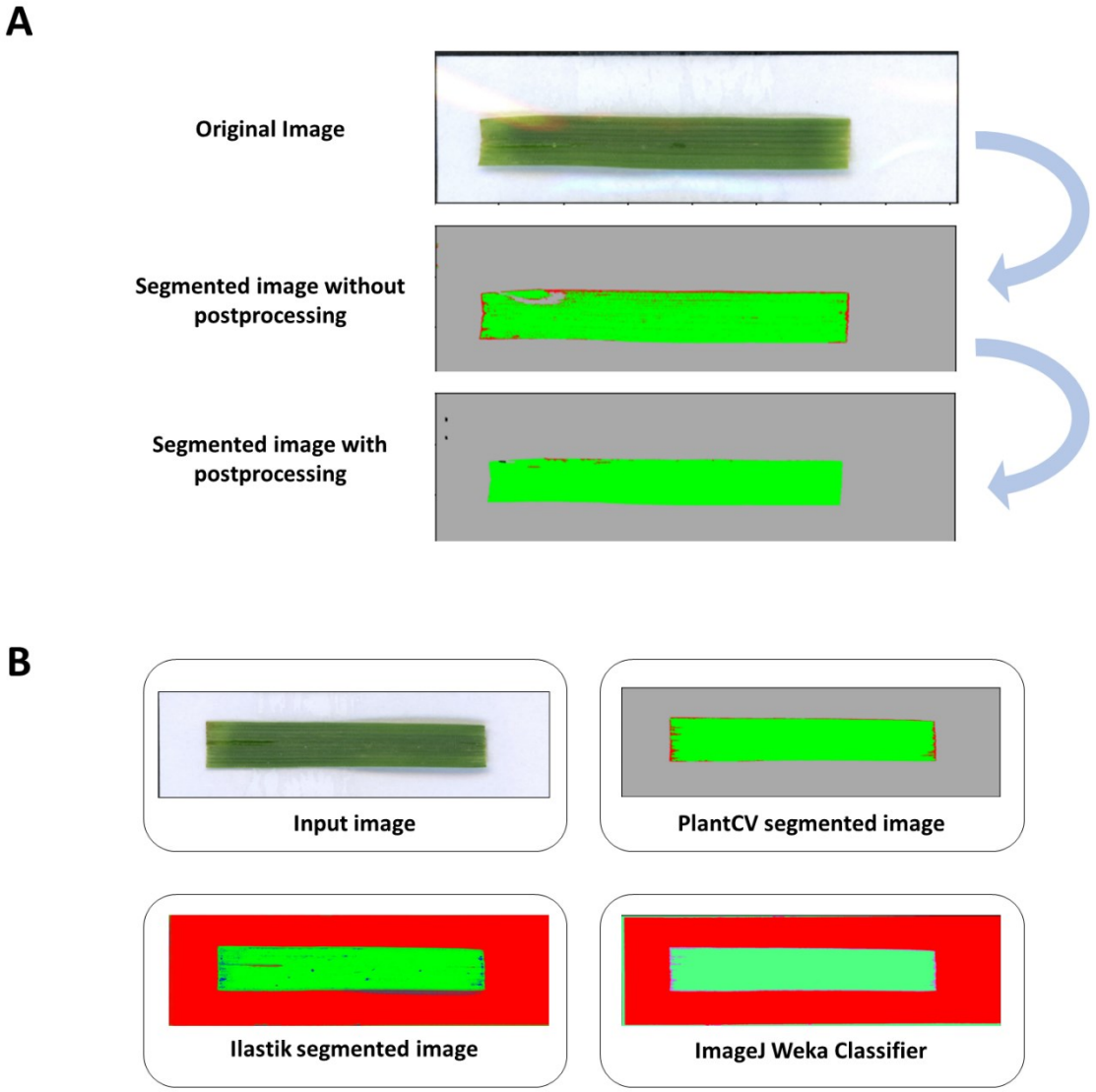



***Figure 12**. False color images representing the outcome of the segmentation. A. An example of a highly irregular image from the dataset. The presence of the yellow flare in the image introduces some noise than nonetheless can be handled with the postprocessing at a very good extent. B. Comparison of the segmentation output without postprocessing in the same image using three different software. Note that in PlantCV the background is better segmented, generally with less false both healthy and infected pixels*

pixels being selected as false positive, compared with the 7.3% when using iLastik and 4.8% when using ImageJ. In the case of Ilastik, this is a tool that makes use of neural networks optimized for segmenting, classifying, tracking, and counting of cells in microscopy images, although it is intended as a general purpose tool (Berg *et al.*, 2019). This platform is aimed to be "user-friendly" and although it produced robust and quick results it lacks the options for fine tuning that we find in PlantCV as well as the relative simplicity of the naïve Bayes approach, especially when compared with the complexities of a neural network. The Trainable Weka Segmentation plugin of ImageJ on the other hand, is a very complete software that contains a collection of machine learning tools that, although remaining "user-friendly", does allows for more control during both the training

and the classification step. However, just as Ilastik, its mainly thought for being used with microscopy applications, and the type of machine learning tools they use generally require more computing power, which translate in longer processing times, something critical when we are in presence of large datasets or real time applications.

It was possible to further decrease the value of average false positive pixels in the PlantCV workflow down to 1.7% using postprocessing. In this case an initial step of denoising was followed by edge-cropping. These two steps were easily implemented in the workflow thanks to the fact is is written in Python and the versatility that this represents. In particular,this does not constraint the utilities and packages that can be used to those of PlantCV, as it is possible to use any other Python package at the same time, which was the case for the edge cropping step, in which some base python functions where used. The flexibility provided by the Python programming language was also clear in the last step of postprocessing, in which functions from Python's OpenCV (Bradski, 2000) package were used to identify the individual colonies and enumerate them, a critical step for retrieving information from each individual colonial formation. Since PlantCV branched from OpenCV, with the latter being already rich in packages (Fahlgren *et al.*, 2015), the combination of both revealed very powerful yet easy to implement..

For integration with genotypic data, like in QTL analysis for example, phenotypical data containing information on trait segregation must be obtained by paying special attention to the experimental design and the method by which the data is retrieved. Inadequate quantification of a trait often affects the data quality in a negative manner. Noisy images highly affect the dataset and could bias the results. Data normality tests and outlier detection is necessary to improve the data quality (Rahaman *et al.*, 2015). In standard interval mapping (IM) of quantitative trait loci (QTL), the QTL effect is described by a normal mixture model. When this assumption of normality is violated, the most adopted strategy is to use the previous model after data transformation. However, an appropriate transformation may not exist or may be difficult to find and this approach can also raise interpretation issues (Fernandes, Pacheco and Penha-Gonçalves, 2007). The output analysis of our workflow showed that the features were not normally distributed for some ranges and were generally skewed. This is not uncommon given that many phenotypic traits are not normally distributed (Rodo *et al.*, 2006) and the assumption of normality can occult important characteristics of the model. Many proposals have been made in order to cope with this, like the use of nonparametric interval mapping based on the Kruskal-Wallis test statistic (Broman *et al.*, 2003) and the replacement of the assumption of normality by

a weaker assumption that the quantitative variable has a “smooth” density that may be skewed (Dalla Valle, 2004).

Previous studies have attempted an image based scoringof the *B. graminis* - wheat pathosystem (Li, Gao and Shen, 2010; Yuan *et al.*, 2014; Awad *et al.*, 2015; Majumdar *et al.*, 2015). Many of these approaches, however, are characterized by having as a main goal the detection of the disease and by relying on small datasets or satellite images. While they can have specific uses, they are not ideal for high throughput phenotyping platforms, in which fast and reliable data extraction is needed and dealing with huge datasets is necessary to complement the enormous amount of genomic data available. On the other hand, in Lück *et al.*, (2020) and in Zhang *et al.*, (2019), platforms for high throughout phenotyping for wheat are described. The first one, called “Macrobot”, scores the visible powdery mildew disease symptoms, typically 5-7 days post-inoculation (dpi), in a highly automated manner. The system can precisely and reproducibly quantify the percentage of the infected leaf area with a theoretical throughput of up to 10000 individual samples per day, making it appropriate for phenotyping of large germplasm collections and crossing populations. The second one is more general purpose oriented and allows for the phenotyping of several traits like plant morphology, greenness, leaf area, amongst others. Both platforms software is meant to be implemented in specific hardware that is an indissociable part of the platform per se. This makes widespread use difficult, due to the obligation of having to acquire both the software and the hardware, the latter being the critical point when it comes to availability and price. An approach like the one describe in this work, implemented in Python using free resources like PlantCV and OpenCV, have a great versatility since it offers the advantage of being capable of running on a larger range of systems and hardware, and this enables its implementation on “in-house” developed hardware that could be based on affordable technologies, like for example Raspberry Pi microcomputers (Gehan and Kellogg, 2017; Tovar *et al.*, 2018).

### *Neonectria ditissima* in *vitro* images quantitative analysis workflow

Manual image selection might have several drawbacks, one of them is that the use of ImageJ “Circle” tool might cause an oversimplification of the real colony shape not considering the variation in the colony borders, which could be an interesting feature to extract. However, across all images of the dataset the shape of the colonies was consistently circular and there ws no important variations in their shape across the different isolates. For this reasons the features selected to be extracted where non-shape related pictures, for which an approximated selection/segmentation of the colony is enough (Puchkov, 2016). During the development of the pipeline, several post-processing

modifications were made to the manual segmentation to obtain more accurate results for some features. The best example of this is the reshaping of the circular ROIs into square ROIs that was made before the extraction of the roughness variable (Figure 13). This was implemented due to the inability of the roughness calculation plugin of ImageJ for handling circular ROIs (Chinga and Dougherty, 2002)

When comparing the Standard Deviation values between treatments, growth on the PDA substrate gave the always higher values. This points to the fact that PDA generates more disperse and variable growth dynamics, and hence, the visual variability of isolates grown

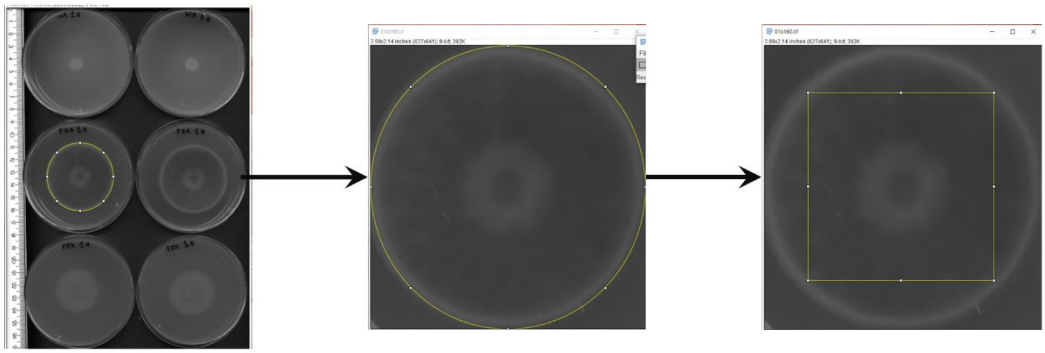


***Figure 13**: ROI reshaping step previous to the extraction of the roughness variable using the Roughness ImageJ Plugin by Chinga and Dougherty (2002)*

on this media is higher. This is valid to almost all the extracted features, with exceptions in some of the Haralick's texture features. It seems that the biggest variation amongst the isolates is displayed when grown on PDA media, something that was assessable visually by noting the variation in the image dataset with visual inspection, but that here is represented with descriptive statistics. The idea of the PDF media being the most suitable one for image-based in vitro analysis of *N. ditissima* isolates colonies is reinforced also after analyzing how the features behave amongst the different isolates (Figure 14). For most of the variables, when grown on PDA, variables are more uniform when comparing the replicates of each isolate and there is less overlap and an apparently higher variability when comparing different isolates, which is not true for most of the features when the other two media are analyzed. Based on this, PDA assay images where the one chosen to perform the cluster analysis because the data extracted from these assays have better quality and seems to be more informative due to the higher variability it showed.

Euclidian distances were used as the distance matrix computation method, since our main purpose is to look at the absolute distance between our isolates given the extracted features and no correlations that might be very distant according to these features were searched. We were trying to quantify the dissimilarity between isolates based on the extracted features, and to group them afterwards. The dendrogram resulting from this cluster analysis

shows a hierarchical representation of the distances between the isolates. It shows a different number of clusters depending on the height its analyzed. By representing this analysis in a heatmap it is possible to visually assess how similar are the different variables and which ones have more weight for the grouping of the isolates. In this case we can see that some variables like Rsk, Contrast, Correlation, and others, are very different within a same group, which mean they could be less informative variables and bring some noise to the analysis.

One of the multiple applications of a multiple feature extraction workflow like the one described here is Computer Based Image Retrieval (CBIR), an approach in which the use

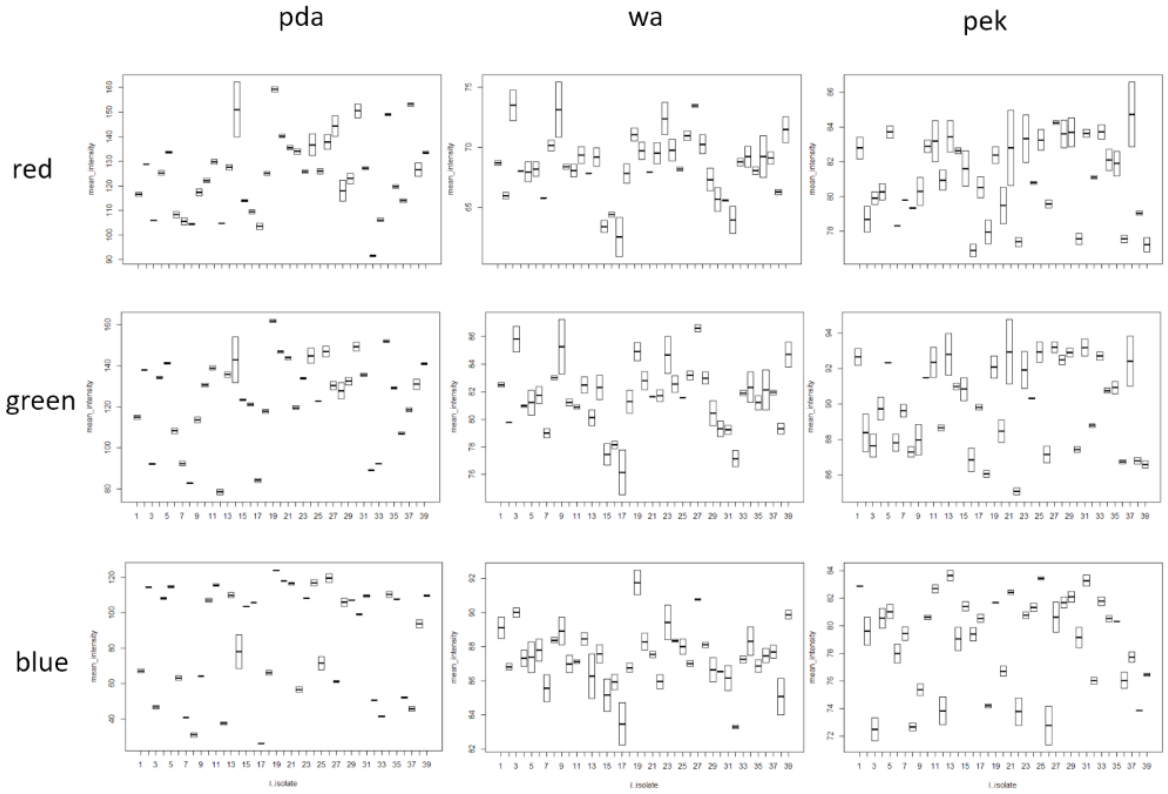



***Figure 14**: Boxplots representing the variation of the Mean Intensity feature amongst the different isolates and the three culture media utilized. In the PDA grown isolates the variation within the same isolates (variation amongst the replicas) is lower, something that does not happens when evaluating the other two culture media.*

of a profile based on visual features is used for the classification and retrieval of images in a dataset. In this approach the extraction of features is the most critical steps, since them should be informative enough to create a profile that's representative of the dataset (Choras, 2007).

## CONCLUSIONS

A workflow for the image-based quantitative analysis of *in planta* inoculation assay of wheat leaves with *Blumeria graminis* f. sp*. tritici* using PlantCV and Python was developed. For the pixel segmentation stage, a naïve Bayes classifier was first trained and then executed over a dataset of over 9600 images. The segmentation yielded satisfactory results on separating the pixels in three classes with the aid of postprocessing. Six different disease related features were obtained as an output and were described statistically. Since it is Python based, this workflow could be use in multiple applications and implemented across multiple hardware platforms.

A workflow for an image-based feature rich quantitative analysis from isolates of *Neonectria ditissima* grown in three different types of culture media was also developed as a part of this work. A pipeline for the extraction of 50 different image feature was implemented into the ImageJ software. These features where then analyzed from the quantitative and qualitative standpoint, showing to be more informative and variable when the isolates were grown on PDA media. Finally, a profile of the different isolates based on the extracted features was generated, which shows the potential of these approach for its use in image-based profiling applications.

**BIBLIOGRAPHY**


Abbasi, A. and Fahlgren, N. (2017) “Naïve Bayes pixel-level plant segmentation,” *2016 IEEE Western New York Image and Signal Processing Workshop, WNYISPW 2016*. doi: 10.1109/WNYIPW.2016.7904790.

Abràmoff, M. D., Magalhães, P. J. and Ram, S. J. (2004) “Image processing with imageJ,” *Biophotonics International*, 11(7), pp. 36–41. doi: 10.1201/9781420005615.ax4.

Amano, K. (1986) *Host Range and Geographical Distribution of the Powdery Mildew Fungi*. Tokyo: Japan Scientific Societies Press.

Amponsah, N. T., Walter, M. and Scheper, R. W. A. (2014) “Agar media for isolation of Neonectria ditissima from symptomatic and asymptomatic apple tissues and production of infective conidia,” *New Zealand Plant Protection*, 67(i), pp. 116–122. doi: 10.30843/nzpp.2014.67.5741.

Arganda-Carreras, I. *et al.* (2017) “Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification,” *Bioinformatics*, 33(15), pp. 2424–2426. doi: 10.1093/bioinformatics/btx180.

Awad, Y. M. *et al.* (2015) “Early detection of powdery mildew disease in wheat (Triticum aestivum L.) using thermal imaging technique,” *Advances in Intelligent Systems and Computing*, 323(November), pp. 755–765. doi: 10.1007/978-3-319-11310-4_66.

Berg, S. *et al.* (2019) “Ilastik: Interactive Machine Learning for (Bio)Image Analysis,” *Nature Methods*. Springer US, 16(12), pp. 1226–1232. doi: 10.1038/s41592-019-0582-9.

Booth, C. (1966) “The genus Cylindrocaipon,” *Mycological papers*, (104).

Bradski, G. (2000) “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*.

Braun, U. *et al.* (2002) “The taxonomy of the powdery mildew fungi,” in Bélanger, R. . et al. (eds.) *The Powdery Mildews: A Comprehensive Treatise*. St. Paul, USA: APS Press, pp. 13–55.

Broman, K. W. *et al.* (2003) “R/qtl: QTL mapping in experimental crosses,” *Bioinformatics*, 19(7), pp. 889–890. doi: 10.1093/bioinformatics/btg112.

Bus, V. G. M. *et al.* (2019) “Genetic mapping of the European canker (Neonectria ditissima) resistance locus Rnd1 from Malus ‘Robusta 5,’” *Tree Genetics and Genomes*. Tree Genetics & Genomes, 15(2), pp. 1–13. doi: 10.1007/s11295-019-1332-y.

Campos, J. da S. *et al.* (2017) “European apple canker: morphophysiological variability and pathogenicity in isolates of Neonectria ditissima in southern Brazil,” *Ciência Rural*, 47(5),

pp. 1–9. doi: 10.1590/0103-8478cr20160288.

Castlebury, L. A., Rossman, A. Y. and Hyten, A. S. (2006) “Phylogenetic relationships of Neonectria/Cylindrocarpon on Fagus in North America,” *Canadian Journal of Botany*, 84(9), pp. 1417–1433. doi: 10.1139/B06-105.

Chaverri, P. *et al.* (2011) “Delimitation of Neonectria and Cylindrocarpon (Nectriaceae, Hypocreales, Ascomycota) and related genera with Cylindrocarpon-like anamorphs,” *Studies in Mycology*. CBS-KNAW Fungal Biodiversity Centre, 68, pp. 57–78. doi: 10.3114/sim.2011.68.03.

Chinga, G. and Dougherty, B. (2002) *Roughness calcularion*. Available at: https://imagej.nih.gov/ij/plugins/roughness.html.

Choras, R. S. (2007) “Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems,” *International Journal of Biology and Biomedical Engineering*, 1(1), pp. 6–15.

Collins, T. J. (2007) “ImageJ for microscopy,” *BioTechniques*, 43(1S), pp. S25–S30. doi: 10.2144/000112517.

Cooke, L. R. (1999) “The influence of fungicide sprays on infection of apple cv. Bramley’s seedling by Nectria galligena,” *European Journal of Plant Pathology*, 105(8), pp. 783–790. doi: 10.1023/A:1008778900607.

Dalla Valle, A. (2004) “The skew-normal distribution,” in *Skew-elliptical distributions and their applications*. Chapman and Hall/CRC, pp. 16–38.

Dean, R. *et al.* (2012) “The Top 10 fungal pathogens in molecular plant pathology,” *Molecular Plant Pathology*, 13(4), pp. 414–430. doi: 10.1111/j.1364-3703.2011.00783.x.

Dörge, T., Carstensen, J. M. and Frisvad, J. C. (2000) “Direct identification of pure Penicillium species using image analysis,” *Journal of Microbiological Methods*, 41(2), pp. 121–133. doi: 10.1016/S0167-7012(00)00142-1.

Dubin, H. J. and English, H. (1975) “Epidemiology of European apple canker in California,” *Phytopathology*, 65(5), pp. 542–550.

Eshed, N. (1970) “Host Ranges and Interrelations of Erysiphe graminis hordei, E. graminis tritici, and E. graminis avenae,” *Phytopathology*, p. 628. doi: 10.1094/phyto-60-628.

Estrada, C. *et al.* (2014) “Fungal endophyte effects on leaf chemistry alter the invitro growth rates of leaf-cutting ants’ fungal mutualist, Leucocoprinus gongylophorus,” *Fungal Ecology*. Elsevier Ltd, 8(1), pp. 37–45. doi: 10.1016/j.funeco.2013.12.009.

Fahlgren, N. *et al.* (2015) “A versatile phenotyping system and analytics platform reveals

diverse temporal responses to water availability in Setaria," *Molecular Plant*. Elsevier Ltd, 8(10), pp. 1520–1535. doi: 10.1016/j.molp.2015.06.005.

Farr, D. F. *et al.* (1989) *Fungi on plants and plant products in the United States*. APS Press.

Fernandes, E., Pacheco, A. and Penha-Gonçalves, C. (2007) "Mapping of quantitative trait loci using the skew-normal distribution.," *Journal of Zhejiang University. Science. B.*, 8(11), pp. 792–801. doi: 10.1631/jzus.2007.B0792.

Flack, N. . and Swinburne, T. R. (1968) "Host range of Nectria galligena Bres. and the pathogenicity of some Northern Ireland isolates," *Transactions of the British Mycological Society*, (2), pp. 185–192.

Florez, L. M. *et al.* (2020) "Reference genes for gene expression analysis in the fungal pathogen Neonectria ditissima and their use demonstrating expression upregulation of candidate virulence genes," *PLoS ONE*, 15(11 November), pp. 1–21. doi: 10.1371/journal.pone.0238157.

Garkava-Gustavsson, L. *et al.* (2013) "Screening of apple cultivars for resistance to european canker, Neonectria ditissima," *Acta Horticulturae*, 976(February), pp. 529–536. doi: 10.17660/ActaHortic.2013.976.75.

Gavhale, M. K. R. and Gawande, P. U. (2014) "An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques," *IOSR Journal of Computer Engineering*, 16(1), pp. 10–16. doi: 10.9790/0661-16151016.

Gehan, M. A. *et al.* (2017) "PlantCV v2: Image analysis software for high-throughput plant phenotyping," *PeerJ*, 2017(12), pp. 1–23. doi: 10.7717/peerj.4088.

Gehan, M. A. and Kellogg, E. A. (2017) "High-throughput phenotyping," *American Journal of Botany*, 104(4), pp. 505–508. doi: 10.3732/ajb.1700044.

Gelain, J. *et al.* (2020) "Neonectria ditissima physiological traits and susceptibility of 'Gala' and 'Eva' detached apple fruit," *Tropical Plant Pathology*, 45(1), pp. 25–33. doi: 10.1007/s40858-019-00314-y.

Gómez-Cortecero, A., Harrison, R. J. and Armitage, A. D. (2015) "Draft genome sequence of a European isolate of the apple canker pathogen Neonectria ditissima," *Genome Announcements*, 3(6). doi: 10.1128/genomeA.01243-15.

Graf, H. (1976) "Die Biologie des Obstbaumkrebses (Nectria galligena Bres.) als Grundlage seiner gezielten Bekämpfung," *Mitt d Obstbauversuchsringes d Alten Landes*, 31, pp. 68–78.

Groos, C. *et al.* (2003) "Genetic analysis of grain protein-content, grain yield and thousand-

kernel weight in bread wheat," *Theoretical and Applied Genetics*, 106(6), pp. 1032–1040. doi: 10.1007/s00122-002-1111-1.

Hamuda, E., Glavin, M. and Jones, E. (2016) "A survey of image processing techniques for plant extraction and segmentation in the field," *Computers and Electronics in Agriculture*. Elsevier B.V., 125, pp. 184–199. doi: 10.1016/j.compag.2016.04.024.

Haralick, R. M., Dinstein, I. and Shanmugam, K. (1973) "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6), pp. 610–621. doi: 10.1109/TSMC.1973.4309314.

Hartig, R. (1889) "Erkrankungen durch atmosphärische Einflüsse. § 24.," in *Lehrbuch der Baumkrankheiten*. Berlin, Heidelberg: Springer, pp. 253–286.

Jankovics, T. *et al.* (2015) "New insights into the life cycle of the wheat powdery mildew: Direct observation of ascosporic infection in Blumeria graminis f. sp. tritici," *Phytopathology*, 105(6), pp. 797–804. doi: 10.1094/PHYTO-10-14-0268-R.

Johnson, R. (1992) "Past, present and future opportunities in breeding for disease resistance, with examples from wheat," *Euphytica*, (63), pp. 3–22.

Jupyter, P. *et al.* (2018) "Environments for Science At Scale," (Scipy), pp. 113–120.

Kang, Y. *et al.* (2020) "Mechanisms of powdery mildew resistance of wheat – a review of molecular breeding," *Plant Pathology*, 69(4), pp. 601–617. doi: 10.1111/ppa.13166.

Karbarz, M. *et al.* (2020) "Quantitative trait loci mapping of adult-plant resistance to powdery mildew in triticale," *Annals of Applied Biology*, 177(2), pp. 223–231. doi: 10.1111/aab.12613.

Köhl, J. *et al.* (2019) "Stepwise screening of candidate antagonists for biological control of Blumeria graminis f. sp. tritici," *Biological Control*. Elsevier, 136(February), p. 104008. doi: 10.1016/j.biocontrol.2019.104008.

Kumar, J., Kumar, S. and Pratap, A. (2015) *Phenomics in crop plants: Trends, options and limitations*, *Phenomics in Crop Plants: Trends, Options and Limitations*. doi: 10.1007/978-81-322-2226-2.

Latorre, B. A. *et al.* (2002) "The effect of temperature and wetness duration on infection and a warning system for European canker (Nectria galligena) of apple in Chile," *Crop Protection*, 21(4), pp. 285–291. doi: 10.1016/S0261-2194(01)00099-0.

Lee, U. *et al.* (2018) "An automated, high-throughput plant phenotyping system using machine learning-based plant segmentation and image Lee, U., Chang, S., Putra, G. A., Kim, H., & Kim, D. H. (2018). An automated, high-throughput plant phenotyping system

using machine learning-," *PLoS ONE*, pp. 1–17.

Li, J., Gao, L. and Shen, Z. (2010) "Extraction and analysis of digital images feature of three kinds of wheat diseases," *Proceedings - 2010 3rd International Congress on Image and Signal Processing, CISP 2010*, 6, pp. 2543–2548. doi: 10.1109/CISP.2010.5646912.

Li, Z. *et al.* (2020) "A review of computer vision technologies for plant phenotyping," *Computers and Electronics in Agriculture*. Elsevier, 176(August), p. 105672. doi: 10.1016/j.compag.2020.105672.

Lobet, G. (2017) "Image Analysis in Plant Sciences: Publish Then Perish," *Trends in Plant Science*. Elsevier Ltd, 22(7), pp. 559–566. doi: 10.1016/j.tplants.2017.05.002.

Lomte, S. S. and Janwale, A. P. (2017) "Plant Leaves Image Segmentation Techniques: A Review," *Article in International Journal of Computer Sciences and Engineering*, 5(5), pp. 147–150. Available at: www.ijcseonline.org.

Lück, S. *et al.* (2020) "'Macrobot': An automated segmentation-based system for powdery mildew disease quantification," *Plant Phenomics*, 2020. doi: 10.34133/2020/5839856.

Majumdar, D. *et al.* (2015) "An integrated digital image analysis system for detection, recognition and diagnosis of disease in wheat leaves," *ACM International Conference Proceeding Series*, 10-13-August-2015, pp. 400–405. doi: 10.1145/2791405.2791474.

Marchal, E. (1902) "De la specialisation du paristisme chez l'Erysiphe graminis," *Compt. Rend. Acad. Sci. Paris*, (135), pp. 210–212.

Marotz, J., Lübbert, C. and Eisenbeiß, W. (2001) "Effective object recognition for automated counting of colonies in Petri dishes (automated colony counting)," *Computer Methods and Programs in Biomedicine*, 66(2–3), pp. 183–198. doi: 10.1016/S0169-2607(00)00128-0.

McCracken, A. R. *et al.* (2003) "Relative significance of nursery infections and orchard inoculum in the development and spread of apple canker (Nectria galligena) in young orchards," *Plant Pathology*, 52(5), pp. 553–566. doi: 10.1046/j.1365-3059.2003.00924.x.

Mehta, N. (2014) "Epidemiology and Forecasting for the Management of Rapeseed-Mustard Diseases," *J Mycol Plant Pathol*, 44(2), pp. 131–147. Available at: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L365849697%5Cnhttp://www.sfd.si/modules/catalog/products/prodfile/fv_4_2012_copy1.pdf.

Minervini, M., Abdelsamea, M. M. and Tsaftaris, S. A. (2014) "Image-based plant phenotyping with incremental learning and active contours," *Ecological Informatics*. Elsevier B.V., 23, pp. 35–48. doi: 10.1016/j.ecoinf.2013.07.004.

Müller, M. C. *et al.* (2021) “Host adaptation through hybridization: Genome analysis of triticale powdery mildew reveals unique combination of lineage-specific effectors,” *bioRxiv*, p. 2021.05.06.442769. Available at: https://www.biorxiv.org/content/10.1101/2021.05.06.442769v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.05.06.442769v1.abstract.

Mutterer, J. and Rasband, W. (2012) “ImageJ Macro Language Programmer ’ s Reference Guide v1.46d,” *RSB Homepage*, pp. 1–45.

Nakaya, A. and Isobe, S. N. (2012) “Will genomic selection be a practical method for plant breeding?,” *Annals of Botany*, (110), pp. 1303–1316.

Ngugi, L. C., Abelwahab, M. and Abo-Zahhad, M. (2021) “Recent advances in image processing techniques for automated leaf pest and disease recognition – A review,” *Information Processing in Agriculture*. China Agricultural University, 8(1), pp. 27–51. doi: 10.1016/j.inpa.2020.04.004.

Niewoehner, A. S. and Leath, S. (1998) “Virulence of Blumeria graminis f. sp. tritici on winter wheat in the Eastern United States,” *Plant Disease*, 82(1), pp. 64–68. doi: 10.1094/PDIS.1998.82.1.64.

Ogawa, H. *et al.* (2012) “Noise-free accurate count of microbial colonies by time-lapse shadow image analysis,” *Journal of Microbiological Methods*. Elsevier B.V., 91(3), pp. 420–428. doi: 10.1016/j.mimet.2012.09.028.

Oku, T. *et al.* (1985) “Host range and forma specialis of cocksfoot powdery mildew fungus (Erysiphe graminis DC) found in Japan,” *Ann. Phytopathol. Soc. Jpn*, (51), pp. 613–615.

Palm, G. (2009) “Untersuchungen zur Bekämpfung des Obstbaumkrebses (Nectria galligena, Bres.),” *Mitteilungen des Obstbauversuchsringes des Alten Landes*, (64), pp. 180–185.

Pieruschka, R. and Schurr, U. (2019) “Plant Phenotyping: Past, Present, and Future,” *Plant Phenomics*, 2019, pp. 1–6. doi: 10.34133/2019/7507131.

Plante, F., Hamelin, R. C. and Bernier, L. (2002) “A comparative study of genetic diversity of populations of Nectria galligena and N. Coccinea var. Faginata in North America,” *Mycological Research*, 106(2), pp. 183–193. doi: 10.1017/S0953756201005329.

Poland, J. A. and Nelson, R. J. (2011) “In the eye of the beholder: The effect of rater variability and different rating scales on QTL mapping,” *Phytopathology*, 101(2), pp. 290–298. doi: 10.1094/PHYTO-03-10-0087.

Pound, M. P. *et al.* (2017) “Deep machine learning provides state-of-the-art performance

in image-based plant phenotyping," *GigaScience*, 6(10), pp. 1–10. doi: 10.1093/gigascience/gix083.

Puchkov, E. (2016) "Image Analysis in Microbiology: A Review," *Journal of Computer and Communications*, 04(15), pp. 8–32. doi: 10.4236/jcc.2016.415002.

R Core Team (2021) "R: A language and environment for statistical computing." Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.r-project.org/.

Rahaman, M. M. *et al.* (2015) "Advanced phenotyping and phenotype data analysis for the study of plant growth and development," *Frontiers in Plant Science*, 6(AUG), pp. 1–15. doi: 10.3389/fpls.2015.00619.

Raichaudhuri, R. and Sharma, R. (2016) "On Analysis of Wheat Leaf Infection by Using Image Processing," in Kacprzyk, J. (ed.) *Proceedings of the International Conference on Data Engineering and Communication Technology*. Springer, pp. 569–577.

Reis, M. M. *et al.* (2016) "Analysis of reference evapotranspiration of Janaúba, Minas Gerais," *Idesia*, 34(4), pp. 55–62. doi: 10.4067/S0718-34292016005000017.

Rodo, J. *et al.* (2006) "MHC Class II Molecules Control Murine B Cell Responsiveness to Lipopolysaccharide Stimulation," *The Journal of Immunology*, 177(7), pp. 4620–4626. doi: 10.4049/jimmunol.177.7.4620.

Rossman, A. Y. and Palm-Hernández, M. E. (2008) "Systematics of plant pathogenic fungi: Why it matters," *Plant Disease*, 92(10), pp. 1376–1386. doi: 10.1094/PDIS-92-10-1376.

Savary, S. *et al.* (2019) "The global burden of pathogens and pests on major food crops," *Nature Ecology and Evolution*, 3(3), pp. 430–439. doi: 10.1038/s41559-018-0793-y.

Scharr, H. *et al.* (2016) "Special issue on computer vision and image analysis in plant phenotyping," *Machine Vision and Applications*. Springer Berlin Heidelberg, 27(5), pp. 607–609. doi: 10.1007/s00138-016-0787-1.

Singh, A. *et al.* (2016) "Machine Learning for High-Throughput Stress Phenotyping in Plants," *Trends in Plant Science*. Elsevier Ltd, 21(2), pp. 110–124. doi: 10.1016/j.tplants.2015.10.015.

Tardieu, F. *et al.* (2017) "Plant Phenomics, From Sensors to Knowledge," *Current Biology*, 27(15), pp. R770–R783. doi: 10.1016/j.cub.2017.05.055.

Tovar, J. C. *et al.* (2018) "Raspberry Pi–powered imaging for plant phenotyping," *Applications in Plant Sciences*, 6(3), pp. 1–12. doi: 10.1002/aps3.1031.

Tsaftaris, S. A., Minervini, M. and Scharr, H. (2016) "Machine Learning for Plant Phenotyping Needs Image Processing," *Trends in Plant Science*, 21(12), pp. 989–991. doi:

10.1016/j.tplants.2016.10.002.

Tulasne, L. R. and Tulasne, C. (1865) “Selecta fungorum carpologia III,” *Paris Museum*, pp. 1–221.

Weber, R. W. S. (2014) “Biologie und Kontrolle des Obstbaumkrebs-Erregers Neonectria ditissima (Syn. N. galligena) aus der Perspektive Nordwesteuropas,” *Erwerbs-Obstbau*, 56(3), pp. 95–107. doi: 10.1007/s10341-014-0210-x.

Xu, X. M. and Robinson, J. D. (2010) “Effects of fruit maturity and wetness on the infection of apple fruit by Neonectria galligena,” *Plant Pathology*, 59(3), pp. 542–547. doi: 10.1111/j.1365-3059.2009.02232.x.

Yang, K. *et al.* (2001) “Strain selection of Metarrhizium anisopliae by image analysis of colony morphology for consistency of steroid biotransformation,” *Biotechnology and Bioengineering*, 75(1), pp. 53–62. doi: 10.1002/bit.1164.

Yuan, L. *et al.* (2014) “Damage mapping of powdery mildew in winter wheat with high-resolution satellite image,” *Remote Sensing*, 6(5), pp. 3611–3623. doi: 10.3390/rs6053611.

Zhang, C. *et al.* (2019) “Development of an automated highthroughput phenotyping system for wheat evaluation in a controlled environment,” *Transactions of the ASABE*, 62(1), pp. 61–74. doi: 10.13031/trans.12856.

Zhao, C. *et al.* (2019) “Crop phenomics: Current status and perspectives,” *Frontiers in Plant Science*, 10(June). doi: 10.3389/fpls.2019.00714.