Geo-information Science and Remote Sensing

Thesis Report GIRS-2018-09

Exploring machine learning for disease assessment from high-resolution UAV imagery

Francisco Arias Rojas

April, 10 2018

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Exploring machine learning for disease assessment from high-resolution UAV imagery

Francisco Arias Rojas

Reg. No.: 830409018010

Supervisors:

dr. ir. Lammert Kooistra

MSc. Marston Domingues Franceschini

A thesis submitted in partial fulfilment of the degree of Master of Science

at Wageningen University and Research Centre,

The Netherlands.

April 2018

Wageningen, the Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2018-09
Wageningen University and Research
Laboratory of Geo-Information Science and Remote Sensing

Preface

Hyperspectral imagery combined with UAV platforms have been extensively explored over the past years as a promising approach to acquire large amount of data as a key factor for plant disease assessment. In that sense, advanced and complex analysis need to be implemented to generate reliable information that could be related to vegetation properties and therefore for the detection, identification and prediction of plant health status. Hence, this study gave me the opportunity to explore the scope and capability of new technologies, methods and applications such as hyperspectral technology, machine learning algorithms, data analysis techniques, programming, among others, to handle, explore and analyze high resolution UAV imagery and related it to plant disease assessment. Moreover, despite the complexity of the analysis and the challenge of its interpretability, several related works, online tutorials, together with online discussion groups, helped me to overcome my limitations.

This study would not be possible without the supervision of dr. ir. Lammert Kooistra and MSc. Marston Domingues Franceschini. My special thanks to dr. ir. Kooistra for his kind cooperation, guidance and substantial support during this study. Many thanks to MSc. Marston Domingues Franceschini, my second supervisor, for all his feedbacks, recommendations, and especially for kindly providing the data that made possible this thesis, such as the preprocessed hyperspectral images, and the ground disease assessment dataset. I would like to thank for all the recommendations and feedbacks of dr. Devis Tuia and MSc. Benjamin Kellenberger during the development of the predictive model.

This study was supported by Wageningen University & Research, which provided the work space and software. This thesis would not be possible without the financial support of the national graduate scholarship program for studying abroad “Don Carlos Antonio Lopez” that gave me the opportunity to study in Wageningen University & Research.

I would like to express my gratitude to my parents Santiago Arias and Antonieta Rojas, who supported me whenever I needed to finish this thesis. Finally, I would also like to thank my girlfriend Valentina Bedoya, who always supported me with kindness and tolerance throughout this thesis.

Table of Contents

List of Figures

List of tables

Abbreviations

DAP	Days after planting
EPPO	European and Mediterranean Plant Protection Organization
EDA	Exploratory Data Analysis
FPI	Fabry-Perot interferometer
FP	False Positive
FN	False Negative
FWHM	Full width at half maximum
FAO	Food and Agriculture Organization of the United Nations
GIS	Geographic Information System
HYMSY	Hyperspectral Mapping System
INS	Inertial navigation system
MLP	Multilayer Perceptron
NDVI	Normalize Difference Vegetation Index
NIR	Near Infrared
nm	Nano meters
OVR	One-versus-Rest
OSAV	Optimized soil-adjusted vegetation index
PCA	Principle Component Analysis
RBF	Radial Basis Function kernel
SAM	Spectral Angle Mapper
SVM	Support Vector Machines
SWIR	Shortwave Infrared
SVI	Spectral Vegetation Index
SU	Sampling Unit
TP	True Positive
TN	True Negative
UAV	Unmanned Aerial Vehicle
VIS	Visible Region
WDVI	Weighted difference vegetation index

Summary

The late blight disease caused by the fungus-like *Phytophthora infestans* is considered one of the most important and devastating diseases, reducing the potato crop production and affecting the agricultural economy and food security worldwide. Moreover, one of the main characteristics of the late blight disease, is that it spreads extremely quickly during growing season, especially in those production systems where no chemicals are applied in the field, such as those organic farming systems, where production tend to be lower comparing to conventional practices due the constrains on the use of fertilizers and pesticides. The need of real time, location-precise, non-invasive methods to assess late blight disease could provide reliable information for the crop management community to establish suitable solutions for this agricultural practice.

The extensive exploration and use of remote sensing over the past years to retrieve physical variables to estimate vegetation properties, has contributed to the development and improvement of different applications of remotely sensed observations to provide spatial, temporal and non-invasive field-based information relative to plant condition, and consequently the detection, identification, quantification and prediction of crop stress, representing one of the major contributions of this science to agriculture. However, this progress in the remote sensing domain and sensor technologies in combination with UAV platforms, brings together the generation of large amount of data that requires of complex analysis to find underlying information that in most cases statistical approaches are sometimes limited to reach.

Progress in computer science allowed the development and design of new approaches such as machine learning algorithms to analyze in depth the full spectrum offered by the hyperspectral technology to retrieve spectral responses related to plant health status. Even though, most of the studies related to plant disease assessment were applied under controlled conditions (laboratory or greenhouses), some authors have been working intensively to transfer results obtained under controlled environmental conditions to experiments performed at field level. Hence, studies for monitoring fields and plots could provide new insights to better understand the complex host-pathogen relation with the disease spread and its spatial distribution.

Therefore, this study focused to explore the potential use of machine learning techniques for the assessment of late blight disease in an organic potato production system from high-resolution UAV imagery. Moreover, this study explored which discriminant function (linear – non-linear) of a support vector machine classifier provided the best solution and performance to predict late blight disease. The experiment was conducted over eight experimental plots, where two different production systems were evaluated through visual assessments (ground truth), together with the acquisition of high-resolution UAV imagery. The combination of both datasets was used as input to train and evaluate the selected predictive model.

A support vector machine learning algorithm was used to classify three different disease severity classes, which due to the disease evolution over the growing season, labeled classes were represented by an unbalanced overall distribution. As expected, a high correlation between the acquired spectral features was found, therefore during the model selection procedure, a preprocessing feature selection technique such as the PCA was incorporated within the pipeline, to evaluated if the new

uncorrelated components contribute for the selection of the best model. Results obtained from the grid search procedure, showed that the incorporation of the first five new components contributed to the selection of the model with the best classification performance, yielding in a balanced accuracy of 82%. In addition, both linear and non-linear SVM discriminant function were tested, to explore which extension of the classifier was more suitable to predict late blight disease. As most of the researches reviewed for this study, among the methods evaluated, a radial basis kernelized SVM algorithm was selected during the model selection procedure.

This study concludes, based on the literature reviewed and the results obtained, that the use of hyperspectral sensors in combination with machine learning techniques such as the support vector machine, has the potential to monitor crop health status. However, it is still necessary to consider factors such as environmental conditions during flight acquisition, image preprocessing and the selection of well-known specific features for the development of automated systems that can provide timely, non-invasive, and reliable information to forecast temporal and spatial disease spread, information that can be use by the crop protection community.

1 INTRODUCTION

1.1 CONTEXT AND BACKGROUND

Progress within the remote sensing domain along with advances in computer science and sensor technologies, have served as basis for the development of tools and methods for the assessment of plant diseases (Ashourloo et al., 2016; Mahlein, 2016; Zhang et al., 2003). It is well known, that plant diseases reduce crop production affecting the agricultural economy and food security worldwide (Rahman et al., 2017). Therefore, researchers, together with the precision crop community, have been working intensively to harness the potential of the enormous amount and quality of spectral data that has increased over the past years (Behmann et al., 2015). Furthermore, major improvements in crop monitoring relay on the potential of remotely sensed observations providing spatial, temporal and non-invasive field-based information relative to plant condition and, consequently, detection of crop disease stress at early stages which potentially represents one of the major contributions of remote sensing to agriculture (Garcia-Ruiz et al., 2013; Mahlein et al., 2012; Zhang et al., 2003).

In addition, the extensive use of remote sensing observations due to its capability to record spectral data at different scales to estimate vegetation properties, allowed the implementation of different platforms and sensors in the crop monitoring field (Behmann et al., 2015; Sankaran et al., 2010; Verrelst et al., 2015). However, even though sensors do not measure the physiological parameters directly but records the sum of reflectance attributes of various plant characteristics and measurements conditions, it has been proven that specific regions in the electromagnetic spectrum from visible to near infrared serve as basis to decision support systems in agriculture (Baranowski et al., 2015; Garcia-Ruiz et al., 2013). Therefore, there is a growing tendency to combine Unmanned Aerial Vehicles (UAVs) platforms equipped with hyperspectral cameras to generate site-specific, non-invasive, high spectral, high spatial and temporal information needed for reliable crop disease detection and assessment methods (Garcia-Ruiz et al., 2013; Yeh et al., 2013).

Potato (*Solanum spp.)* constitutes one of the leading non-grain commodities in the global food system with a production area of approximately 19 million/ha worldwide in 2014 (Food and Agriculture Organization of the United Nations [FAO], 2014). Meanwhile, average potato yields in organic farming systems are lower comparing to conventional practices due to the constrains on the use of fertilizers and pesticides (Franceschini et al., 2017; Lammerts van Bueren et al., 2008). This year to year variation can be caused by the fungal pathogen *Phytophthora infestans* which causes late blight disease. One of the main characteristic of late blight disease is that it spreads quickly during growing season, and variations in timing and severity can result in devastating yields losses (Gebru et al., 2017; Lammerts van Bueren et al., 2008; Zhang et al., 2003).

According to Franceschini et al. (2017), a limited number of studies have evaluated the benefits from continuous monitoring focusing on crop management and stress detection in organic farming systems. In that sense, the combination of hyperspectral technology and UAV platforms measurements seems to be a promising approach to acquire timely information as a key factor for disease assessment. This timely information requires complex analyses to be performed in order to find underlying information that traditional statistical approaches are limited to reach (Mucherino et al., 2009). Therefore, many

data mining techniques have been developed and designed to offer new opportunities to solve specific problems in many research fields (Behmann et al., 2015; Mucherino et al., 2009). For instance, in the field of agriculture, different automatic classification and prediction approaches have been utilized for vegetation parameter estimation, disease identification, classification, quantification and prediction (Ashourloo et al., 2016; Garcia-Ruiz et al., 2013; Mucherino et al., 2009; Zhang et al., 2003). Although, the analysis and interpretation of the outputs obtained could be challenging.

1.2 Problem Definition

Spectral vegetation indices (SVIs) constitute one of the most traditional variable estimation approaches since their simplicity (Verrelst et al., 2015). On the other hand, it is well known that despite all the successful applications of remote sensing technologies, conventional imaging is not always suitable for detection or assessment of intrinsic physical or chemical properties (Park & Lu, 2015). Thus, hyperspectral images seem to be one of the most common techniques used in the exploration of new approaches and methods in the field of plant disease assessment (Dale et al., 2013).

A plant which is infected, reacts to protect itself, and this physiological reaction caused by the disease, will lead to a decrease of the spectral reflectance values in the near infrared (NIR) range, due to the decreasing chlorophyll content and internal structure changes (Martinelli et al., 2015; Zhang et al., 2003). This absorption of light variations between diseased and non-diseased plants could allow to identify the stress severity of green vegetation with high potential for the evaluation of plant health (Franceschini et al., 2017; Martinelli et al., 2015; Zhang et al., 2003). Therefore, spectral responses of vegetation are considered to be the basis for remote sensing for a timely data collection of spectral information, concerning the assessment of stress factors of an organic potatoes system and their impact on the crop development (Franceschini et al., 2017).

Even though multiple approaches are available to extract physical variables to estimate and monitor vegetation properties based on remote sensing, there are several intrinsic and extrinsic factors which make these methods too error prone (Barbedo, 2016; Thomas et al., 2018; Verrelst et al., 2015). Some of the main challenges in the implementation of advanced retrieval methods are related to the heterogeneity of elements of the background, capture environmental conditions, disease intrinsic characteristics, and the presence of more than one pathogen at the same time, which implies that the methods relies on very weak differences to discriminate among them (Barbedo, 2016).

Progress in computer science allows to explore methods to analyze the full potential of hyperspectral optical data (Dale et al., 2013). Even though, vegetation indices are still a simple and robust indicators for assessment of crop health, advance retrieval methods allows to explore the full spectrum offered by hyperspectral imagery (Behmann et al., 2014). Several studies using hyperspectral technology for disease detection , identification and classification have been carried out over the past years (Thomas et al., 2018). Most of these studies were applied under controlled conditions (laboratory or greenhouses) at a leaf scale, which were focused to explored the plant internal changes and its relation with spectral information, while others authors started to explore the disease detection and its spatial distribution working at canopy scale, combining different platforms to acquired spectral data, encountering several technical and external conditions setups that needs to be considered, in order to provide reliable information for the crop management community.

The need of real time, location-precise, non-invasive methods for predicting late blight disease in organic farming systems seems to be an essential research topic that could contribute to the establishment of suitable solutions of these agricultural practices (Franceschini et al., 2017; Jiang et al., 2012; Lammerts van Bueren et al., 2008). Therefore, this research aims to explore the combination of UAV hyperspectral data and advanced data analysis methods to assess plant disease.

1.3 RESEARCH OBJECTIVE AND RESEARCH QUESTIONS

This research aimed to explore machine learning techniques for assessment of late blight disease in an organic potato production system from high-resolution UAV imagery.

Based on the overall objective, the following research questions had been designed:

RQ1) Can high-resolution UAV imagery provide information which allows to predict late blight disease in organic potato production?

RQ2) Which supervised machine learning (linear- non-linear) technique is suitable to predict late blight plant disease?

RQ3) Does the prediction capability of the model increase with increasing disease severity?

1.4 OUTLINE

This thesis consists of 7 chapters. Chapter 1 outlined the content and background of the thesis research as well as the research question on which the thesis is build. The following chapter describes in detail the late blight disease and its implications in a potato crop, it also reviews the concept of remote sensing and its purposes and possibilities in the disease assessment domain, and the use of data mining techniques to analyze the data acquired using remote sensing technologies. Chapter 3 describes the study area and the data acquisition methods used in the growing season of 2016. Chapter 4 provides an overview of the process and methods used to prepare the dataset used to d develop and implement the selected model. The final outputs are presented in chapter 5 and will be discussed based on existing literature in chapter 6. The last chapter 7 summarizes the conclusions of this thesis research.

2 LITERATURE REVIEW

2.1 LATE BLIGHT DISEASE IN POTATO CROPS

The late blight disease is caused by the oomycete *Phytophthora infestans* (Mont.) de Bary (Olanya et al., 2015); it is one the main diseases that affects potato production worldwide (Anderson et al., 2004) and the cause of the use of agrochemicals and copper fungicides in organic farming years ago, which had been restricted by the European Commission (European Commission, 2002). Besides, there is public resistance against the use of ecologically disapproving chemicals, resulting in a pressure for prioritization new tools to control this disease (Haverkort et al., 2008). The late blight incidence is accentuated in those areas with periods of high relative humidity and low temperatures. In favorable environmental conditions for the pathogen it can spread very rapidly and provoke severe crop losses, causing the death of leaves, stems and tubers of the plant (Stephan et al., 2005).

The late blight pathogen caused the Irish potato famine in the mid-19th century, which left more than a million deaths and generated the migration of more than 1.5 million people from different parts of the world, particularly the United States (Ristaino & Johnston, 1999). Late blight disease was first reported in the eastern United States just prior to be reported in Europe.

Phytophthora infestans is responsible for yield losses of $6-7 billion annually in potato, and crop losses up to 100% in tomato (Nowicki et al., 2012) and has been considered a risk to global food security (Cooke et al., 2012). Due to this severe economic impact many efforts have been put into planning agronomic strategies to raise potato yields by delaying the infection and decelerating the epidemic, implementing new cultural practices, treating seed tubers to improve the crop and growing resistant varieties (Haverkort et al., 2008). Yields are usually low when an early attack occurs, therefore national measures had been taken (i.e. in the Netherlands), which force to kill the potato vine when more than 7% of the leaf surface is affected (Lammerts van Bueren et al., 2008).

The factors that contribute to the severity of the damage are varying but the main one is the intensity and speed with which this pathogen destroys large areas of the crop and the relationship between genetically resistant/susceptible cultivars to the different stages of the infection through which these crops are exposed (Fabre et al., 2012). According to Bock et al. (2010), it is crucial in plant disease assessment domain to describe concepts and their interpretation at the time to estimate or measure plant disease symptoms. Therefore, within this study disease severity is define as the area of the sampling unit (SU) showing symptoms of disease which is expressed as a percentage (Bock et al., 2010; European and Mediterranean Plant Protection Organization [EPPO], 2008).

2.2 SYMPTOMS AND LIFE CYCLE OF LATE BLIGHT DISEASE

One of the first symptoms observed in the field are small spots on the lower leaves, which vary in color from green light to dark, which may be irregular or circular, soaked with water. Under conditions of high humidity, the spots enlarge rapidly and take on a brown color and well-defined edges and the growth of the sporangiophores and sporangia of the pathogen is observed. The part closest to the ground will wither. The speed with which this oomycete destroys the fields is extremely fast if the

humidity conditions are optimal (Fry, 2008). However, in dry conditions the infection rate decreases, the existing lesions stop growing, blacken and the sporangiophores are not observed, however if the humidity increases the pathogen is reactivated, developing the disease rapidly. It is important to point out that the tubers can present infection because of the spores that fall to the ground, these present red to violet depressions that can penetrate the tissue and continue to grow even after the harvest, leading to a rot of the tubers (Agrios, 2005).

The lifecycle of *Phytophthora infestans* comprises of three sub-cycles (*Figure 1*). The first cycle starts at the infected plant, which will generate sporangia with spores on the leaf surface and produced spores will germinate on another plant. This cycle takes about 5 days to complete if the growing conditions are optimal. This optimal condition generate spores that germinate in 2 or 3 hours at 20°C and high humidity causing the infection in the potato plant (Skelsey et al., 2009). The second cycle occurs in the plant tuber, causing the sporulation in young plants and spores that germinate in another plant.

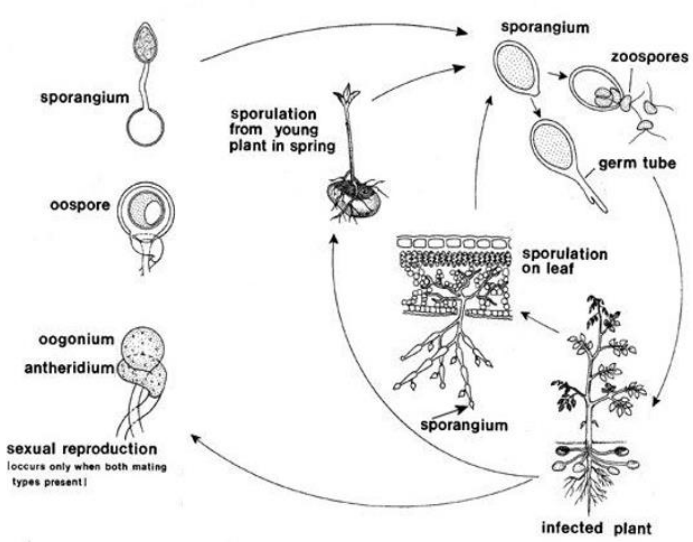


Figure 1. Disease cycle of *Phytophthora infestans* (source: Schumann and D'Arcy 2000)

The third cycle goes from an infected plant to a sexual reproduction where an oospore is generated creating a sporangium with a spore that will germinate in another plant, this cycle can occur only if both mating type of the pathogen (A1 and A2) are present. This last cycle causes the disease in the growing season (Fry, 2008) due to spores remain in good conditions after the winter. Independently of the cycle both produce spores and sporangia and infect plants by germination through a leaf colonization causing the lesions (Wiik, 2014).

2.3 REMOTE SENSING AND DISEASE ASSESSMENT

Remote sensing is defined as a science of obtaining information through the detection of the physical characteristics of an area or object, by measuring its reflected and emitted radiation at a distance from it (Campbell, 2002). This characteristic allowed the exploration of the use of remote sensing in many domains during the past decades. The importance of spatial and temporal information specially in the

agricultural sector was realized when the first aerial photograph was used for mapping soil resources (Seelan et al., 2003).

One of the main advantages besides the capability to acquire vast amounts of spectral information that can be related to the physical and physiological characteristic of vegetation, is that remote sensing imagery nowadays can be acquired through a range of different platforms which makes them a perfect tool to timely monitor vast areas of agricultural landscapes (Lillesand et al., 2015; Sankaran et al., 2010).

The basic principle behind the use of the spectral information is that plants absorb and reflect light based on their structure and plant pigments. The reflectance in the different wavelengths can be related to the leaf pigments (visible - VIS), cell structure or biomass (near-infrared - NIR) and water content (shortwave infrared - SWIR) (*Figure 2*) (Clevers & Kooistra, 2012). Therefore each plant has a unique spectral signature based on their health status which allows the assessment of stress factors and their impacts over vegetation that could cause the reduction of crop production (Muhammed, 2005; Rahman et al., 2017; Ustin et al., 2004).

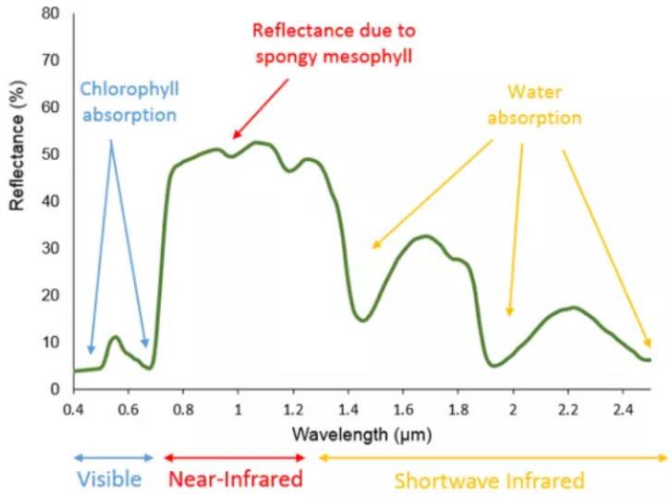


Figure 2. Typical spectral response characteristics of green vegetation (Humboldt State, 2017)

As described previously, a plant which is infected, reacts with protection mechanisms, and this physiological reaction caused by disease, could lead to suboptimal growth, producing a decrease of the spectral reflectance values in the near infrared (NIR) range, due to the decreasing chlorophyll content and internal structure changes (Martinelli et al., 2015; Zhang et al., 2003). Thus, the use of remote sensing as a key technology for real-time detection and diagnosis of crop status is being implement by the precision crop community and researchers worldwide (Behmann et al., 2014).

Hyperspectral images also known as hypercubes consist in a three-dimensional data sets that holds within light intensity measurements (Dale et al., 2013). Hyperspectral imaging in combination with UAV platforms offer high potential as a non-invasive, timely, diagnostic tool for disease detection, classification, quantification and prediction (Mahlein et al., 2013). This innovative sensor technology may allow an objective and automatic assessment of plant disease severity in combination with

advance data analysis methods that can process the vast amount of data that can be acquired (Mahlein et al., 2012).

In most of the studies where hyperspectral imaging is used, the spectral signature of plants affected by a disease is compared to the spectral response of healthy plants. For example, Rumpf et al. (2010) discriminated three different disease from non-disease sugar beet leaves using spectral vegetation indices and spectral values from a non-imaging spectroradiometer. Wang et al. (2008) were capable to predict *Phytophthora infestan*s infections on a tomato plants using regions of hyperspectral signatures, and Bravo et al. (2003) used field hyperspectral images for an early detection of yellow rust in wheat.

A recent study explored the effects of disease symptoms over automatic prediction methods and estimating disease severity. As the disease severity increased, Ashourloo et al. (2016) compared the scattering patterns of spectral vegetation indices (SVI) with advanced methods using the full spectrum. Their results showed that disease symptoms adversely affected the performance of SVI in estimating disease severity, while machine learning techniques applied were less affected by disease symptoms (Ashourloo et al., 2016). Zhang et al. (2003) studied late blight in tomatoes and evaluate the capability of hyperspectral imagery to monitor crop disease at field scale. Within the research, they found that NIR spectral region was much more valuable than VIS range to detect disease.

Therefore, based on the data that can be acquired by hyperspectral sensors, advance techniques to analyze data, and the use of geographic information systems (GIS), reliable and accurate estimates of disease can be calculated to generate application maps that can improve the management of agricultural fields (Mahlein et al., 2012).

2.4 MACHINE LEARNING AND DISEASE ASSESSMENT

Due the fact that hyperspectral technology provides the capability to acquired large amount of data, complex analyzes were explored to find hidden regularities in the growing volumes of data that could contribute to the establishment of suitable solutions to the crop management community (Franceschini et al., 2017; Jiang et al., 2012).

Machine learning is defined as a multidisciplinary approach to data analysis. This data mining technique also known as predictive analysis or statistical learning, is a research field that falls in the intersection of statistics, computer science and artificial intelligence (Mitchell, 2006; Mucherino et al., 2009; Muller & Guido, 2016). Nowadays, the extensive use in real-world problems such as speech recognition, computer vision, and many others data-intensive sciences, is due the fact that machine learning methods serve as a support for the scientific discovery process (Mitchell, 2006).

In addition, machine learning techniques aim to discover patterns by looking at a combination of features instead of analyzing each feature separately (Behmann et al., 2015; Singh et al., 2016). There are two main categories of machine learning which depends on if target variables are labeled or not. Unsupervised learning tries to find unknown patterns from unlabeled training data. In contrast, when target variables of the training data are labeled, the identification of patterns are related to the label

(Behmann et al., 2015). The key ability of this methods is their capability to generalize patterns from available data, which allows to develop robust and flexible prediction models (Singh et al., 2016).

Even though, there are very few reported activities on the use of advance methods for the prediction of plant stresses (Singh et al., 2016), many applications of machine learning within plant disease assessment domain reviewed for this study involve tasks that can be implement through a supervised learning set up, where target variables admit only discrete values (Maglogiannis, 2007). The success of classification learning rely on the quality of the data provided as input, the selection of the algorithm, and the metric used to evaluate performances among others (Hall, 1999).

Many authors have been explored different machine learning approaches to provide new insights into the complex host-pathogen systems using hyperspectral information. For instance, Rahman et al. (2017) compared different machine learning algorithms like support vector machines (SVM), random forest (RF) and multilayer perceptron (MLP) to classify healthy and non-healthy plant leaves of sorghum, citrus and cabbage. They concluded that the use of color-based features, together with the use of a non-parametric classifier such as the random forest, provided a good combination of procedures to distinguish between healthy and infected plant leaves. Additionally, Garcia-Ruiz et al. (2013), compared two different aerial imaging platforms for the identification of Huanglongbing (HLB) or citrus disease. Both UAV, and aircraft based spectral features were used with different classification algorithms, such as linear discriminant analysis, quadratic discriminant analysis, linear and non-linear support vector machines. In this study, authors found accuracies in the range of 67 % - 85%, where a kernelized support vector machines outperformed the other classifiers.

Other authors like Zhang et al. (2003), explore the use of several techniques that includes the minimum noise fraction, pure pixels endmember selection and spectral angle mapping classifier for the identification of late blight disease in a tomato crop. The concluded that the use of spectra-based classification approach has the potential to identify crop disease. Moreover, Maldonado et al. (2013) explored the use of different feature selection techniques and compared them using a support vector machine learning algorithm. Their proposed approach, based on successive holdout steps, reached a better performance, in comparison with the conventional recursive feature elimination technique.

Most of the literature reviewed for this study, suggest that the capability of machine learning to retrieve information from the full spectrum provided by hyperspectral technology allows to explore the effects of disease over vegetation and therefore it implementation to improve plant disease assessments strategies that can be use by the precision crop community.

3 MATERIALS

3.1 STUDY AREA

The dataset used in this study was acquired on eight experimental plots cultivated with potato, during the growing season of 2016. These plots were located within a strip cropping experiment (51.9917°N, 5.66332°E; WGS84) started in 2014, at the Droevendaal experimental farm of the Wageningen University, The Netherlands. The annual precipitation in Wageningen ranges from 765 ± 130 mm and the annual evapotranspiration is on average 525 ± 50 mm; the average annual air temperature is around 11°C (Surmann, 2015).

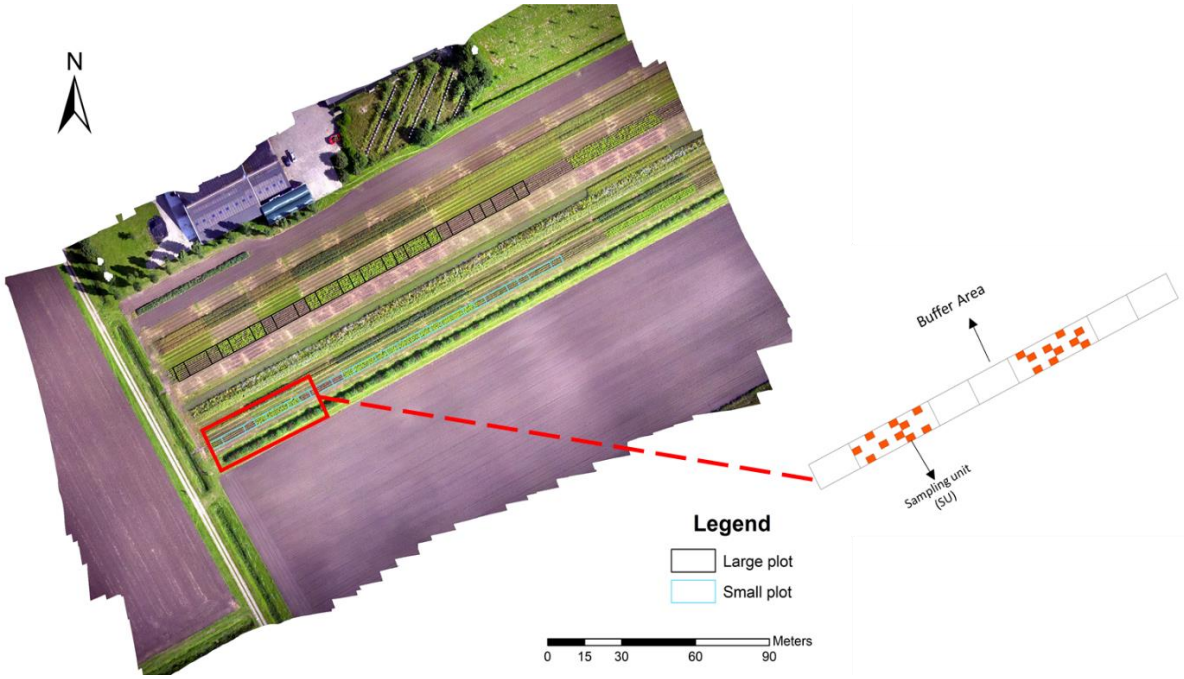


Figure 3. Experimental plots arrangement during 2016 growing season (True color composite). Small plots (in pale blue) and large plots (in black). Right image describes Sampling Units (in orange) within the first two small plots.

3.2 DATA DESCRIPTION

This section mainly describes all the procedures conducted to acquire the data used to build the final dataset. It contains a brief description of the field experimental design, the methodology used to assess late blight disease, the sensors used to acquire spectral information, and the selection criteria of the ground assessments and UAV flight data used in this study.

3.2.1 Experimental design in growing season 2016

Plots were arranged in strips along the experimental site and measured 3 m x 10 m (small plots) or 6 m x 10 m (large plots) (*Figure 3*). Buffer areas, measuring 3 m x 5 m or 6 m x 5 m, were placed before and after each plot, in the same strip, to mitigate border effects along the experiment (*Figure 3*). Two different cultivation methods were compared: (a) plots in which a variety mixture of three different

cultivars, with different degrees of resistance (from low to high) to late blight, were iterated in each crop row, referred as mixed crop system; and (b) plots cultivated with only one potato variety considered susceptible to late blight, named non-mixed system. The experiment followed a generalized randomized block design, with three blocks and two replicates of each treatment (i.e., cultivation methods) in each block (*Figure 3*), although only the first two blocks of the trial (eight plots) were followed over the season.

3.2.2 Ground disease assessment during growing season 2016

Late blight incidence and severity was assessed at a given number of sampling units (SU) per plot (i.e., ten SU for smaller plots and only one SU for larger plots), each one measuring 0.75 by 1 m (*Figure 3*) at regular intervals (i.e., from 3 to 5 days between assessments) after the disease was first detected in the field (*Figure 4*). Disease assessments consisted of visual evaluations according to methodology described by the European and Mediterranean Plant Protection Organization (European and Mediterranean Plant Protection Organization [EPPO], 2008) (Appendix B).

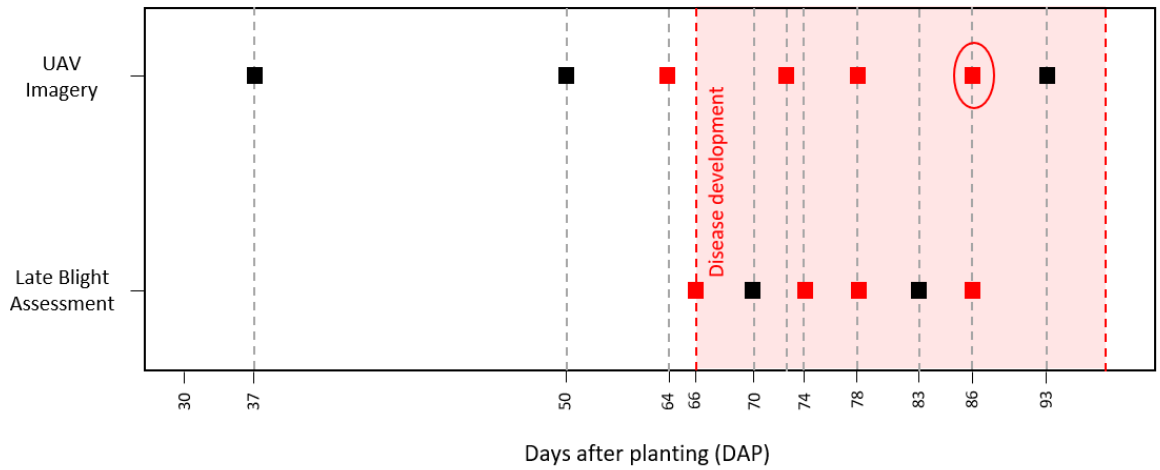


Figure 4. Data acquisition during the growing season 2016. Ground and UAV flight assessments used within this study (red squares). In circle the UAV-HYMSY acquisition is indicated.

For this study, data available from ground assessments [66,74,78 and 86 days after planting (DAP)] were used (*Figure 4*). These assessments were selected based on the availability of UAV flight assessment spectral data from the same date or couple of days before.

3.2.3 UAV-VHR acquisition assessments

A lightweight hyperspectral frame camera (Rikola Ltd., Oulu, Finland) based on a Fabry-Perot interferometer (FPI) (Honkavaara et al., 2013; Roosjen et al., 2017) was used to acquire data in the study area during the 2016 season. The camera was set to register 16 spectral bands over the interval between 600 and 900 nm, with full width at half maximum (FWHM) varying between 10 and 20 nm. Due to intrinsic sensor characteristics, images corresponding to different wavelengths were not registered at the same time, since changes in the wavelength measured depend on internal mechanical adjustment of the system.

Along the field experiment on growing season 2016, seven acquisition flights were conducted as shown in *Figure 4*. Only flights acquisitions with ground assessment data available were used [64,73,78 and 86 DAP] (*Figure 4*). Although the Rikola camera, described above, was used in most of the flights during the growing season, it is important to mention that on acquisition DAP [86 DAP], the Hyperspectral Mapping System (HYMSY) was used instead.

The HYMSY used on acquisition 6 consists of a custom UAV-based pushbroom imaging system. Comprises a custom spectrometer with a range 450-950nm, FWHM 9nm (PhotonFocus SM2-D1312 camera – PhotonFocus AG, Lachen, SZ, Switzerland – with a Specim ImSpector V10 2/3 spectrograph – Specim, Spectral Imaging Ltd, Oulu, Finland), a photogrammetric camera (Panasonic GX1 16 MP – Panasonic Corp., Osaka, Japan – with 14 mm pancake lens), an integrated GPS and inertial navigation system (INS) unit (XSens MTi-G-700 – Xsens Technologies BV, Enschede, Netherlands), together with synchronization and data sink elements (Franceschini et al., 2017; Suomalainen et al., 2014). Bands wavelength from HYMSY where selected base on the bands wavelengths used with the Rikola camera (*Table 2*).

4 METHODOLOGY

The proposed methodology for this study is depicted in *Figure 5*. It can be divided in two main stages: a) the data preparation and b) the modelling disease assessment. These stages are described in detail in the following sections.

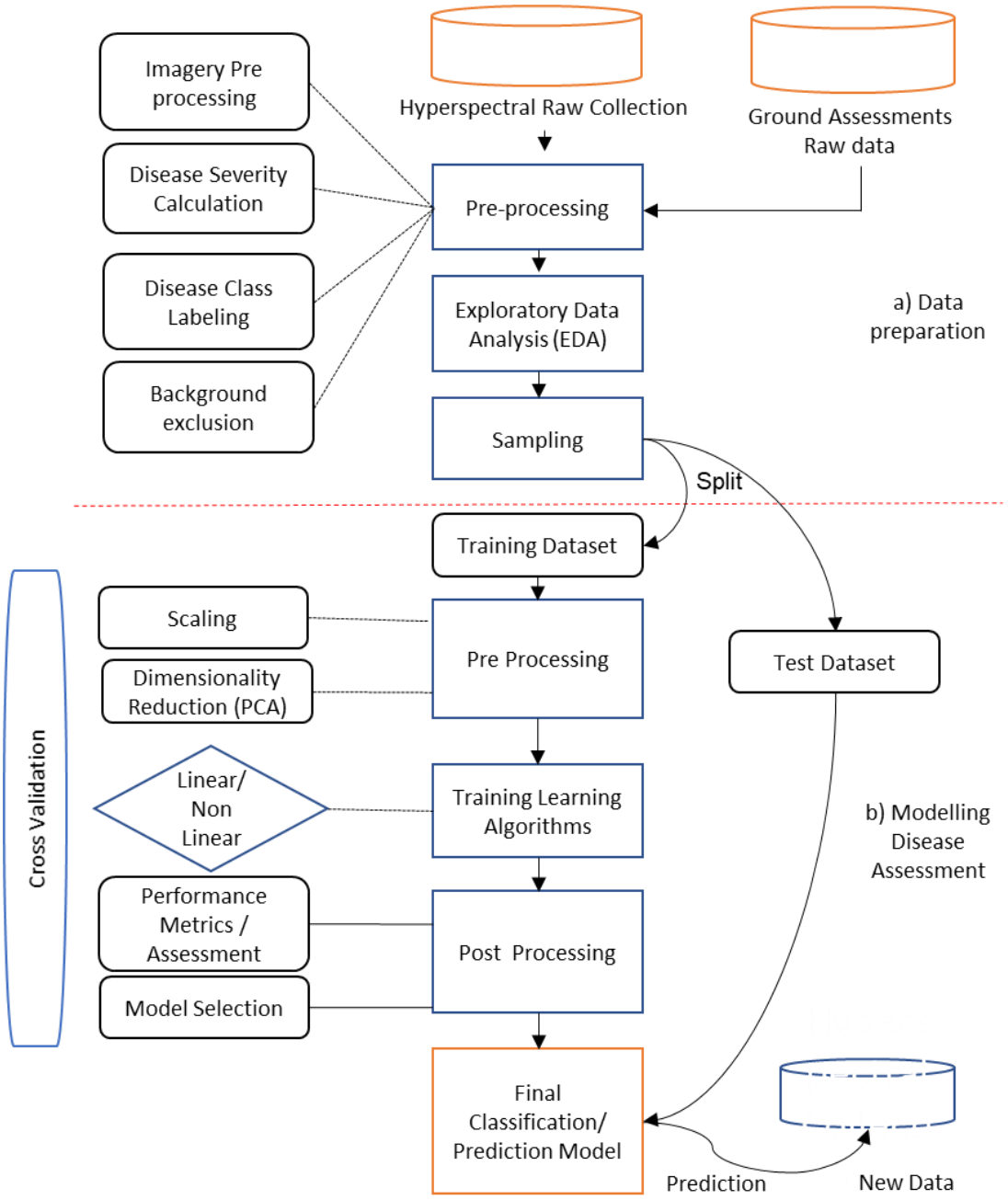


Figure 5. Flowchart of the research approach to meet the proposed research objectives.

4.1 DATA PREPARATION

In most field work experiments data is generally incomplete and is likely to contain redundant information or errors (Alpaydin, 2014). Therefore, preprocessing as an important phase in data mining process aims to transform raw data into an understandable format to improve the performance of models that can be sensitive to the predictors (Ye, 2015).

Data preparation is commonly described as a cumbersome and tedious process, even though it implementation before data analysis is considered a crucial step (Rahman et al., 2017). In this research, data preparation consisted mainly of a group of pre-processing steps together with some data exploration techniques that aimed to prepare, build and understand the final dataset used as input for the development and analysis of the selected model.

This stage is divided in two main activities. The first one so-called preprocessing consisted in a series of procedures that aimed to prepare and combine different datasets acquired along the field experiment into a single dataset which was used as input of the selected disease model. Within preprocessing phase image pre-processing, spectral information extraction at SU scale, background signal exclusion, disease severity calculation at sampling unit (SU) and labeling of different disease classes are considered as the main activities performed to generate the desired dataset.

Once the desired dataset was obtained from the preprocessing stage, data exploration analysis (EDA) took place. This phase aimed to gain the necessary confidence in the dataset before deploying machine learning algorithms. EDA techniques used in this study helped to maximize insight into the dataset, therefore is considered a key activity of this study.

Finally, sampling of the final dataset was carried out. This stage is considered as a crucial step that allowed to evaluate the generalization capability and performance of the machine learning model. Spatial location of experimental plots was considered in the proposed sampling method used within this study.

4.1.1 Preprocessing

Hyperspectral image preprocessing was already conducted before this research (Appendix A). Two spectral bands (**B5**= 682.31nm, **B7**= 696.66nm) were excluded due to technical problems during flight acquisition. Therefore, a total of 14 bands acquired with the Rikola camera (Table *1*) were selected as features to build the input dataset for the analysis. As described in section 3.2.3, assessment 6 [86 DAP] (Figure *4*) used the HYMSY system, where the selection of bands was based on bands acquired with the Rikola camera (Table *2*). After the band selection procedure, sampling unit polygon layers were used to extract pixels for each acquisition image since SU was used as the scale at which this study focused on.

Table 1. Rikola spectral band definitions selected for this study

Band	Wavelength (nm)	Band	Wavelength (nm)
B0	609.08	B9	735.17
B1	615.42	B10	764.28
B2	624.96	B11	793.58
B3	631.33	B12	832.95
B4	663.35	B13	862.71
B6	692.27	B14	872.67
B8	715.87	B15	887.65

The availability of ground assessment data allowed to calculate disease severity relating the counting and area of lesions per plant caused by late blight disease with the percentage of disease severity at SU scale accordingly to the key assessment of (European and Mediterranean Plant Protection Organization [EPPO], 2008) (Appendix B). Hence, sampling units were labeled according to ranges of disease severities selected for this study as shown in Table *3*. These ranges were selected based on Dutch late blight regulations which forces every potato grower to kill the haulm when 7% of leaf surface is visually affected (Hoofdproductschap Akkerbouw, 2003; Lammerts van Bueren et al., 2008).

Table 2. HYMSY spectral bands selected based on Rikola camera acquisitions

Band	Wavelength (nm)	Band	Wavelength (nm)
B32	610	B57	735
B33	615	B63	765
B35	625	B69	795
B36	630	B77	835
B43	665	B82	860
B48	690	B84	870
B53	715	B88	890

Table 3. Disease severity range and class label

% Disease Severity	Class Name	Class Label
0	None Disease	0
0 – 7	Up to 7%	1
> 7	Above 7%	2

Once the datasets of disease severity and spectral information at SU were obtained from previous steps, both datasets were merged based on the assessment date, plot number, plot size and sampling unit information of each observation. As a result, a total of 336 sampling units distributed in plots along the selected assessments dates were obtained (Table *4*)

Table 4. Total of sampling units (SU) per assessment

	Plot size	Number of Plots	Number of SU evaluated per Plot	Total of SU
Assessment 1	small	8	10	80
	large	8	1	8
Assessment 3	small	8	10	80
	large	0	1	0
Assessment 4	small	8	10	80
	large	8	1	8
Assessment 6	small	8	10	80
	large	0	1	0
Total		48		336

To include in the analysis pixels covering vegetation regions, a simple NDVI (Normalized difference vegetation index) based segmentation approach was used, where a NDVI threshold for each assessment was selected as optimal to exclude soil (non-vegetation) and excessive unwanted non-canopy pixels from each SU evaluated for this study.

As the first step of the NDVI based segmentation approach proposed to exclude background signal, calculation of the NDVI (*Equation 1*) for each assessment date was performed, followed by a manual selection of a NDVI threshold derived from histogram of NDVI values and visual assessment.

$$\text{NDVI} = \frac{Nir_{793.586\text{nm}} - R_{663.353\text{nm}}}{Nir_{793.586\text{nm}} + R_{663.353\text{nm}}} \quad \text{(Equation 1)}$$

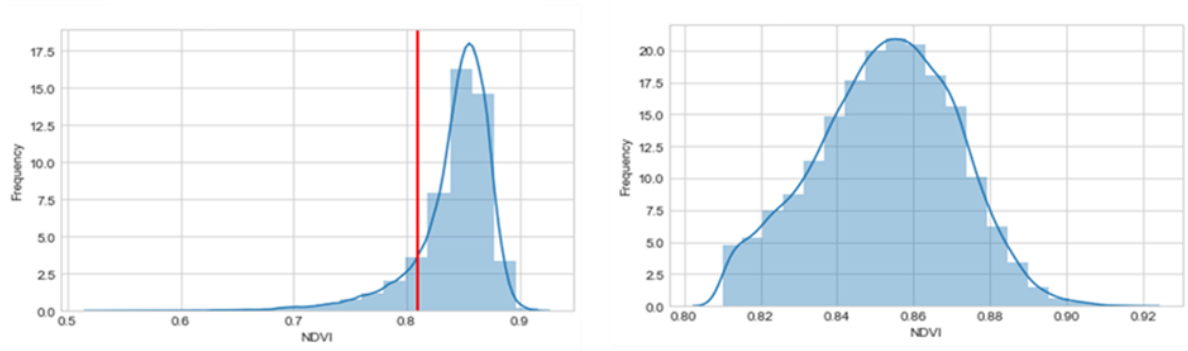
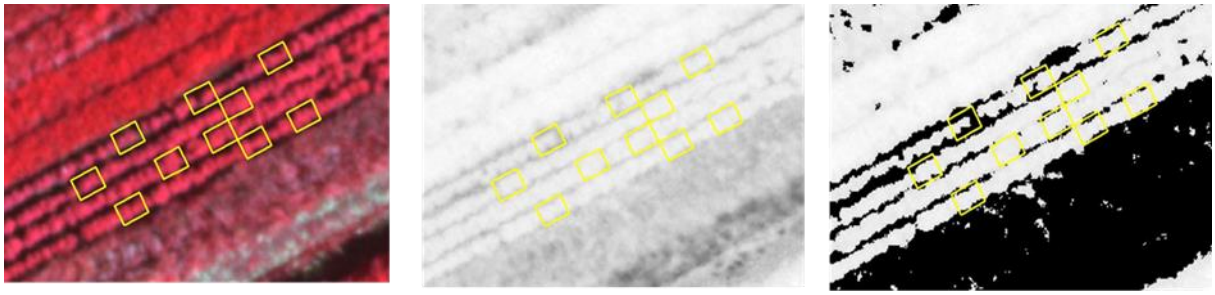


Figure 6. Normalized difference vegetation index (NDVI) histogram for Assessment 1. Figure (a) depict NDVI distribution and threshold selected (red line). Figure (b) NDVI-threshold selected values distribution (vegetation).

After calculation of the NDVI (Equation 1), histograms of NDVI distribution together with NDVI percentiles calculation for each assessment were used to select an optimal threshold to exclude background signal as described above. As an example of the outputs obtained within this stage, distribution of NDVI values for all sampling units (SU) evaluated on assessment 1 can be observed in *Figure 6a*. After the selection of the NDVI threshold, a new distribution of NDVI values is shown in *Figure 6b*, where pixels related to non-vegetation regions were removed.

The NDVI threshold was then evaluated with a visual assessment of images to determine if exclusion of soil and unwanted pixels was successfully accomplished (Figure *7*). The threshold value selected for this example was (NDVI > 0.81) (Table *5*).



a) b) c)

Figure 7.Background removal visual evaluation for Assessment 1. Figure (a) False color composite image. Figure (b) NDVI grey-scale image. Figure (c) NDVI- Assessment 1 threshold mask. Yellow squares depict sampling units (SU).

The background removal procedure was applied for each assessment, and as a result Table *5* shows the NDVI threshold used for each one. It can also be observed that NDVI thresholds decrease over time, this can be attributed to both physiological impact of late blight disease over crop health status and natural vegetation senescence, indicating a possible reduction in canopy structure.

Table 5. NDVI thresholds selected per assessment

Assessment	**NDVI threshold**
Assessment 1 – DAP 66	> 0.81
Assessment 3 – DAP 74	> 0.80
Assessment 4 – DAP 78	> 0.76
Assessment 6 – DAP 86	> 0.69

As mentioned previously, this study was focused to work at sampling unit scale described in section 3.2.2. Therefore, after background exclusion an average of all pixels within each SU for all features were calculated and related to the disease severity label. In addition, it was proposed to calculate and add as features the standard deviation, maximum and minimum value of features to capture variation of spectral information of each sample unit.

Finally, after each step of this stage was performed a total of 320 sampling units (SU) with 17 features constituted the final dataset (Appendix D). This dataset was the one explored and used as input for the selected predictive model.

4.1.2 Exploratory data analysis

As part of data preparation, exploratory data analysis (EDA) described as an approach to summarize the main characteristics of the data through visualization was performed (Morgenthaler, 2009). This stage aimed to gain the necessary confidence in the dataset before deploying machine learning

algorithms, maximize insight into the dataset, identify the most important features, determine relationships among features, inspect class balance distribution of target variables among others (Cox, 2017; Muller & Guido, 2016).

In this study, the exploratory data analysis of the disease dataset started with the inspection of the disease severity evolution along ground disease assessments. Distribution of disease severity per assessment, disease class balance distribution and the number of sampling per treatment system affected are considered as the main outputs that helped to understand the data structure of the disease dataset.

In addition, as part of EDA spectral features acquired were also explored. Spectral responses of mixed and non-mixed affected crops were analyzed to observed if variation of spectral signatures could be related to disease severity and crop susceptibility, which could provide discriminatory information. As part of the spectral feature analysis, a correlation matrix of features was elaborated to visualize the presence of highly correlated features that can have a negative impact on the model performance.

4.1.3 Sampling

As mentioned in section 2.4, one of the main requirements of machine learning is to build models with high generalization capabilities (Reitermanová, 2010). Using the same data to build and evaluate the model has been recognized as generating over optimistic outcomes, mainly because the model can remember the training dataset, therefore always predict the correct target in the training dataset (Faraway, 2016; Muller & Guido, 2016).

There are several sampling methods which were developed for various types of datasets (Picard & Berk, 2017). These sampling methods aims to contribute in the development of models with a high ability to generalize the extracted knowledge from the learning process into new unseen data (May et al., 2010; Reitermanová, 2010).

Due the size of the disease dataset, part of the data was used to train and validate the supervised machine learning model, called training data or training set, while the rest of the data was used to evaluate the performance of the selected model, called test data or test set. In addition, spatial context of experimental plots (Figure *8*) was considered to split the dataset, this criterion was proposed to ensure that the model could be evaluated with a complete unseen data avoiding collinearity of feature information between training and test set.

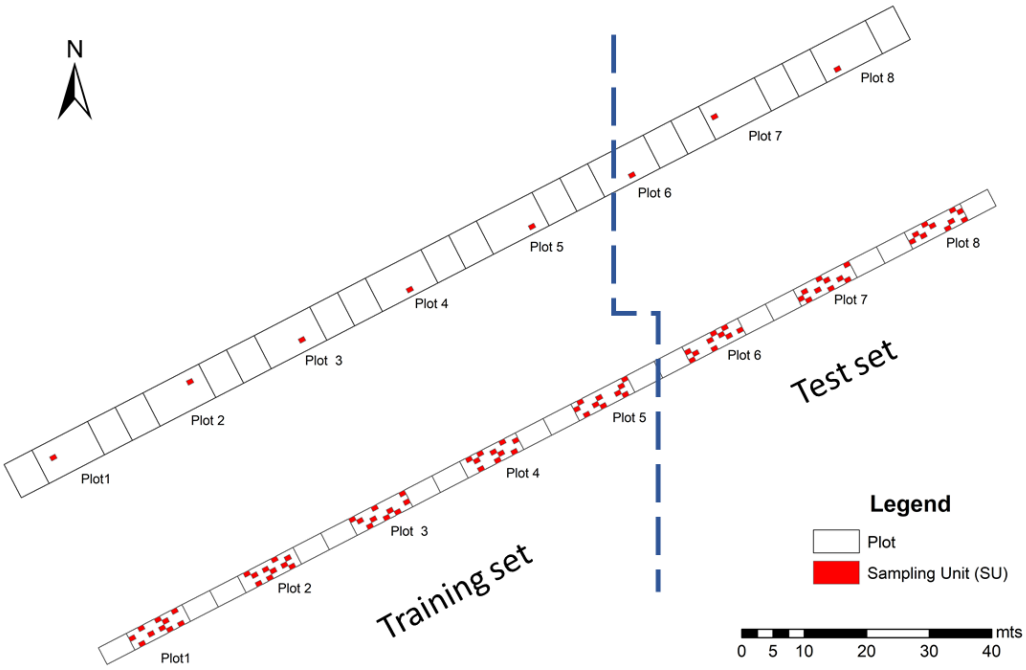


Figure 8. Map of sampling method used to split dataset

As a result, around 60% of the dataset (194 SU) were used as training dataset corresponding to all sampling units evaluated along the experiment within plots 1 through 5 , this includes small and large plots (Figure *8*), while the remaining 126 SU constituted the test set used to evaluate the model performance.

4.2 MODELLING DISEASE ASSESSMENT

This stage was mainly focused in the selection, implementation and evaluation of the model. Therefore, it included steps like, scaling of data, dimensionality reduction, hyper-parameter optimization, training and evaluation of the performance of the model.

4.2.1 Support Vector Machines (SVM)

Support Vector Machine learning algorithm (SVM) is one of the most powerful classification technique based on a structural risk minimization derived from the statistical learning theory developed by Vapnik and collaborators (Härdle, 2011). SVMs were first successfully applied to classification tasks where the main objective was focused to find hidden patterns within data with regard to label classes (Behmann et al., 2015; Härdle, 2011; Rumpf et al., 2010). This algorithm can be effective with both linear and non-linear discrimination function to separate different classes (Sathyanarayana & Amarappa, 2014). In this section, a linear and non-linear extension of SVM which were implemented and evaluated in this study, is described in detail.

4.2.1.1 Linear SVM

To describe the linear SVMs, an assumption that the data is linearly separable is needed. The model of an SVM basically consist on a subset of samples called support vectors, together they define a

hyperplane that separates the classes in the feature space, while it position is defined by the principle of maximum margin (Sathyanarayana & Amarappa, 2014).

To describe a linear function, it is better to start with a binary classification task. Let for example denote X as the input variables and Y be the space of output variables. The basic idea here is to find a linear function of the inputs $x \in X$ *such that* $f(x) < 0$ whenever label $Y = -1$ and $f(x) \geq 0$ otherwise. This can be expressed by the decision function in the space X as:

$$f(x) = sgn(w^T x) + b \quad \text{(Equation 2)}$$

Where $\vec{w}$ is the normal vector and the bias b (*Equation 2*) (*Figure 9*). Depending on the side of the hyperplane where the sample falls in, the sign of the Equation 2 will be assigned. As explained by Rumpf et al. (2010), the optimal hyperplane is the one that maximizes the distance between the hyperplane and the margins, where margin is defined as the minimal Euclidean distance between any training sample and the separating hyperplane (*Figure 9*). The training samples with the minimal distance to the hyperplane are called support vector which define the hyperplane (Abe, 2010; Härdle, 2011; Rumpf et al., 2010).

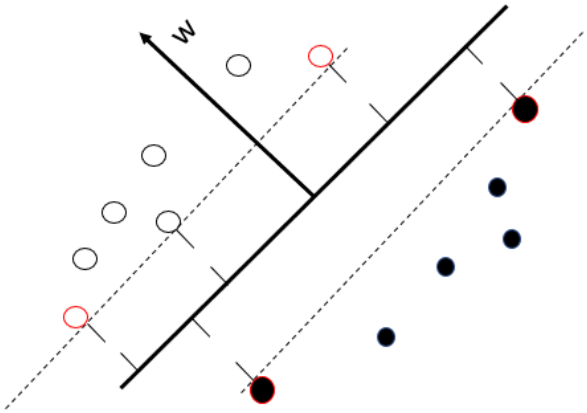


Figure 9. Linear SVM classifier is defined by the normal vector w of a hyperplane (solid line) and an offset b. Margins (dashed lines) of a linear classifier is the minimal distance of any training point to the hyperplane. Circles with red outline are the support vectors for each class.

Since the samples are assumed to be linearly separable, any f can be used to split the data into a canonical hyperplane, requiring the scaling of w and adjusting the threshold b to be such that the closest points to the hyperplane satisfy *Equation 3*. Here the norm of $\vec{w}$ is equal to the inverse of the distance of the closest samples of each class to the hyperplane.

$$|\langle \vec{w}, \vec{x}_i \rangle + b| = 1 \quad \text{(Equation 3)}$$

Hence, the optimal hyperplane can be formulated as a quadratic optimization problem as can be observed in *Equation 4*:

$$\min_{w \in R^m, b \in \mathbb{R}} \quad \tau(\vec{w}) = \frac{1}{2} \ |\vec{w}|^2 \text{ (Equation 4)}$$

$$\text{Constrained to} \longrightarrow \quad y_i \ .(\langle \vec{w}, \vec{x} \rangle + b) \geq 1 \quad \forall_i = 1, \dots \dots, n$$

This constraint guarantees that the function $f(\vec{x}_i)$ obtain 1 for each $y_i \in \{+1\}$ and -1 for each $y_i \in \{-1\}$, so both classes can be separate (Rumpf et al., 2010; Sathyanarayana & Amarappa, 2014). Instead of the optimization problem described above, SVM can be formulated to learn a linear classifier by solving an optimization problem introducing the Lagrange multipliers.

4.2.1.2 Non-linear SVM

In contrast to linear SVM assumptions, most real-world dataset will not be linearly separable. Thus, a kernel function is used to extend the linear SVM to handle overlapping classes and non-linear discriminant functions (Härdle, 2011; Mukherjee & Pal, 2005).

The idea behind this so-called kernel methods is to preprocess the data by some non-linear mapping, which constructs hyperplanes in a high dimensional space. The main constraint with explicitly using the mapping to construct a feature space is that the resulting space can be extremely high-dimensional (Härdle, 2011).

Therefore, the kernel trick which maps inputs into a high dimensional space, can be applied since all feature vectors only occurs in dot products (Equation 5) (Rumpf et al., 2010). The kernel trick contains an implicit mapping in the feature space without explicitly computing the mapping space, represented by φ , and can still solve the problem in a huge feature space (Härdle, 2011; Mukherjee & Pal, 2005; Rumpf et al., 2010). The $(\vec{x}, \vec{x}_i)$ so called dot product, can then substitute by the kernel, as shown in the Equation 5.

$$k(\vec{x}, \vec{x}_i) = (\varphi(\vec{x}).\varphi(\vec{x}_i)) \text{(Equation 5)}$$

Radial Basis Function Kernel (RBF)

Even though the most appropriate kernel depends mainly on the problem at hand for this study the Gaussian kernel as an example of Radial basis function kernel was selected.

$$k(\vec{x}, \vec{x}_i) = exp(-\frac{|\vec{x}+\vec{x}_i|^2}{2\sigma^2}) \text{(Equation 6)}$$

The support vector $\vec{x}$ (Equation 6) will be the center of the RBF and σ will determine the area of influence this support vector has over the data space. A larger value of σ will give a smoother decision surface and more regular decision boundary (Girma, 2009; Rumpf et al., 2010). A large σ will also allow a support vector to have a strong influence over a larger area (Girma, 2009).

4.2.1.3 Multi-class classification

Most of the supervised classification methods were developed mainly for binary classification tasks, but some of them can be extended to handle multi-class classification effectively. For this study the one-versus-rest (OVR) technique was selected. For this extension, a binary model is trained for each class which tries to separate that class from all the other classes, resulting in as many binary models as there are classes (Muller & Guido, 2016). The binary classifier that has the highest score on its class on a test point is the class label that is returned as the prediction.

4.2.2 Data transformation - Scaling of data

Many machine learning algorithms works properly when the range of features are scaled. Scaling the input data before applying SVM is considered a common practice as a preprocessing step, because this algorithm tries to find the best separating hyperplane that maximizes between the hyperplane and the margins, therefore, if one of the features has greater numerical ranges, the distance could be influenced by this specific feature (Hsu & Chang, 2008; Muller & Guido, 2016).

Feature standardization ensures that values of each feature have zero mean and the variance is 1, bringing all features to the same magnitude, hence it refers to centering the distribution of the data (Muller & Guido, 2016). It is important to consider, that it is crucial to apply the same transformation method to scale both the training and test set, in order to have the desired effect of scaling the data (Hsu & Chang, 2008; Muller & Guido, 2016).

In this study, the StandardScaler tool from the scikit-learn package was used for this procedure. According to Muller and Guido (2016), this function ensures that for each feature the mean is 0 and unit-variance.

4.2.3 Dimensionality reduction

The amount of information provided by hyperspectral imagery enhance the ability to classify and recognize materials (Muller & Guido, 2016; Preet et al., 2015). On the other hand, high dimensionality has several obstacles that need to be considered. For example, it has been observed that highly correlated features have a negative impact on classification performance (Preet et al., 2015).

The principal component analysis (PCA) is an unsupervised algorithm, described as one of the most common feature selection techniques which involves linear transformation of the original data. This method rotates the data such that the generated features so-called components are statistically uncorrelated keeping the maximum amount of information (Gilbertson & van Niekerk, 2017; Muller & Guido, 2016; Preet et al., 2015).

Even though the idea to explore the full potential of spectral information acquired during UAV flights on 2016 season is proposed in this study, the spectral information along the spectrum regions used from RED to NIR are highly correlated, feature correlation will be covered in section 5.1.2.2. Therefore, as a preprocessing step before training the models, it was tested if applying a principal component

analysis (PCA) as a feature selection technique, improved the performance of the selected models. Scikit-learn package has the *PCA* tool that was used in this study.

4.2.4 Cross-validation

Due the amount of labeled data used to feed the selected models, a cross-validation technique was used to tune the parameters to achieve the best performance. Cross-validation is defined as a statistical method of evaluating and comparing learning algorithms generalization performance by splitting data into a training and test set (Muller & Guido, 2016).

In cross-validation, the dataset is divided repeatedly, therefore several models need to be trained. This technique ensures that training and validation sets cross-over such that each sample is validated against (Refaeilzadeh et al., 2009). *Stratified k-Fold cross-validation* is implemented in scikit-learn for classification task by default, where the data is divided in such way that the proportions between classes remains the same in each fold as they are in the whole dataset (Muller & Guido, 2016; Refaeilzadeh et al., 2009).

4.2.5 Hyper parameter optimization

Often in machine learning, hyperparameter optimization also known as model selection, is described as a step that aims to improve model's generalization performance by choosing a set of parameters with the objective of optimizing a measure of the algorithm on a validation dataset (Bergstra & Yoshua, 2012). For this stage cross validation described in section 4.2.4, was used to estimate the generalization performance.

The use of hyperparameter optimization to configure learning algorithms can have varying effects on the resulting model and its performance (Claesen & De Moor, 2015). Therefore, several approaches are being tested in the research community specially when the number of hyperparameters is large (Bergstra & Yoshua, 2012; Claesen & De Moor, 2015).

Finding values of the important parameters is commonly performed manually , via rules-of-thumb or by testing of hyperparameters on a grid (Bergstra & Yoshua, 2012; Muller & Guido, 2016). Grid search requires to select a set of values for each parameter, then it will test all possible combination of these values until the best combination is achieved and one model reaches the best performance according to the performance metric selected (Muller & Guido, 2016). Because it considered as a common task, scikit-learn package provides a *GridSearchCV* class that was used in this study.

Since this study was focused to explore which type of SVM was more suitable to classify and generalize late blight disease, several parameters (*Table 6*) were combined and evaluated to find the best combination, based on the balanced accuracy metric to test the performance of each trained model. This procedure allowed to evaluate which kernel (linear-non-linear) was able to better classify and generalize the target classes. As mentioned in section 4.2.3, a PCA preprocessing step was also included, to evaluate if its inclusion contributed to the performance of the best model.

Table 6 . Grid of parameters used for hyper parameter tuning

Preprocessing		N° of components	Algorithm	Kernel	Decision function	Class Weight	C [range]	Gamma [range]
PCA	Yes	[1-6]	SVM	Linear	OVR	Balanced	[0.001-1000]	---
						None	[0.001-1000]	---
				RBF		Balanced	[0.001-1000]	[0.001-1000]
						None	[0.001-1000]	[0.001-1000]
	No	---		Linear		Balanced	[0.001-1000]	---
						None	[0.001-1000]	---
				RBF		Balanced	[0.001-1000]	[0.001-1000]
						None	[0.001-1000]	[0.001-1000]

--- Not applicable

4.2.6 Evaluation metrics

The choice of an appropriate evaluation metric is highly task-specific, hence selection should be based on the problem domain and objectives (Mokhtarian, 2017). Performance metric plays a fundamental role in assessing the quality of learning techniques and to achieve the optimal algorithm during selection and implementation of the model either for classification or prediction tasks (Ferri et al., 2009; Hossin & Sulaiman, 2015).

Hossin & Sulaiman (2015), states that "evaluation metrics are mainly employed in three different evaluation applications during the modeling process". Firstly, as an evaluator for the selection of the model. Secondly, commonly used to measure the performance of the trained classifier when tested with unseen data. And thirdly, to identify and select the best solution among all generate solutions during training. In that sense, a series of metrics were used within this study along different stages of model development and implementation that will be explained below.

4.2.6.1 Confusion matrix

In machine learning, one of the most comprehensive ways to evaluate a classification task is by using a confusion matrix (Bekkar et al., 2013; Muller & Guido, 2016). As an example, for a binary classification task, the output of confusion matrix is a two-by-two matrix as shown in Table *7* , where rows correspond to the true classes and the columns to the predicted classes. Performance of classification algorithms is evaluated using data within this matrix, where entries on the main diagonal correspond to correct classifications, while off-diagonal entries correspond to mistakenly classifications (Muller & Guido, 2016).

Table 7. Confusion matrix for binary classification task

	Predicted Positive	**Predicted Negative**
Actual Positive	TP (number of True Positive)	FN (number of False Negative)
Actual Negative	FP (number of False Negative)	TN (number of True Negatives)

4.2.6.2 Overall accuracy metric

The accuracy metric is known as the most common and simplest performance metric for a classifier evaluation, it basically evaluates the overall effectiveness of an algorithm (Bekkar et al., 2013). This metric is defined as the degree of correct classification of a model and it is calculated from the principal diagonal entries of the confusion matrix as observed in Equation 5, where the overall accuracy is the proportion of correctly classified observations divided by the total number of observations (Ferri et al., 2009; Mosley, 2013).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(Equation 5)}$$

4.2.6.3 Precision and Recall

The sensitivity, true positive rate or recall, is a metric that measures the proportion of positives examples that were correctly identified as such, quantifying the capability of the model to avoid the false negative (Equation 6).

$$Recall = \frac{TP}{TP + FN} \quad \text{(Equation 6)}$$

The precision or positive predicted value metric, measures how many of the samples predicted as positives are correctly predicted (Equation 7). It is commonly use when the main objective is to limit the number of false positives.

$$Precision = \frac{TP}{TP+FP} \quad \text{(Equation 7)}$$

4.2.6.4 Balanced accuracy

Even though there is not yet a consensus on the implementation of the balanced accuracy for multi class tasks, authors like Brodersen et al. (2010) and Mosley (2013) describe the balanced accuracy as the recall for each class, averaged over the number of classes (Equation 8).

$$Balanced\ Accuracy = \frac{\sum_{l \in L} \frac{TP}{TP+FN}}{L} \quad \text{(Equation 8)}$$

As a performance metric, the balanced accuracy evaluates the quality of the model performance, which it is first measured for each class independently and then aggregated. In an unbalanced case scenario, if the overall accuracy described above is high because it is biased due the good performance of the model on the majority class, the use of the balanced accuracy metric will balance the score (Brodersen et al., 2010). In the Equation 8, 'l' is the label, while 'L' is the set of labels.

4.2.6.5 ROC curve and AUC

Receiver operating characteristic (ROC) have been used over the past years within machine learning community to visualize and evaluate the tradeoff between the true positive rate and the false positive rate algorithms performance (Fawcett, 2006). Therefore, the ROC curve is defined as a graphs plotted with the true positive rate on the y-axis against the specificity or false positive rate to create the curve, where the goal of a good model is to be at the upper left corner, which indicates low false positives , with a high recall (Cortes & Mohri, 2003; Fawcett, 2006).

To compare classifiers ROC can be reduce to a single scalar value called the area under the curve(AUC), which is defined as the area under the ROC curve that is related as a measure of the quality of the classification (Cortes & Mohri, 2003). According to Mosley (2013) "at perfect accuracy both ROC axis is maximized suggesting that larger areas are superior".

4.2.7 Software

This study required the use of different software along the methodology process. *Table 8* shows the list of the main software applied and the process in which they were used.

Table 8. Software used: name and actions

Software	**Process**
Spyder 3.2.4 (Python 2.7.14)	Data preparation. Exploratory data analysis. Model selection. Model Evaluation. Sckit-learn v0.19.1 Package (Machine Learning in python library)
Erdas Imagine 2015	Background exclusion. NDVI thresholds visual assessment
ArcGis 10.5	Visualization maps.

Scikit-learn v0.19.1 library for python

Scikit-learn is an open source project which contains many state-of-the-art machine learning algorithms (documentation can be found in [http://scikit-learn.org/stable/documentation.html]. It provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python (Muller & Guido, 2016).

All the tools available and needed within scikit-learn library were used to analyze, pre-process, prepare and process the available data, on the other hand it was also used to select and evaluate the model used to assess late blight disease in an organic potato field.

5 RESULTS

This chapter presents the main results obtained in this study. The outcomes are presented in the same order of the procedures performed and described in chapter 4. Firstly, results obtained from data preparation phase are reported, beginning with the preprocessed dataset that was explored and then used as input of the selected prediction model. Finally, major findings obtained during model selection, development and implementation are discussed.

5.1 DATA PREPARATION

In this section, key results obtained during the data preparation stage (*Figure 5*) are presented due to their importance and contribution in the generation of the disease dataset, development and deployment of the disease assessment predictive model.

5.1.1 Preprocessing

The number of sampling units that remained after the preprocessing procedures was reduced in comparison with the total of sampling units presented in Table *4*. A total of 16 sampling units of the assessment 6 (86 DAP) that belonged to the first and third of the small plots were excluded from the preprocessed dataset so-called disease dataset, because pixels within these sampling units presented lower values of NDVI than the selected threshold for this assessment date. As a result of this preprocessing procedures, a total of 320 samples constituted the disease dataset (*Table 9*) that served as input to the model (Appendix D).

Table 9. Number of sampling units (SU) per assessment - as part of the field observed disease dataset

	Plot size	Number of Plots	Number of SU evaluated per Plot	Total of SU
Assessment 1	small	8	10	80
	large	8	1	8
Assessment 3	small	8	10	80
	large	0	0	0
Assessment 4	small	8	10	80
	large	8	1	8
Assessment 6	small	8	10	64
	large	0	0	0
Total		48		320

5.1.2 Exploratory data Analysis (EDA)

Once the disease dataset was generated, exploratory data analysis took place. Within this section, major findings are presented that helped to better understand data structure, target distribution, and feature characteristics for the field disease evaluations and the flight acquisitions.

5.1.2.1 Disease severity evolution along growing season

Ground assessment of late blight disease symptoms along the field evaluations allowed the estimation of disease severity at sampling unit scale. From this estimation, the evolution of disease severity and the distribution of severity within each assessment can be observed *(Figure 10*).

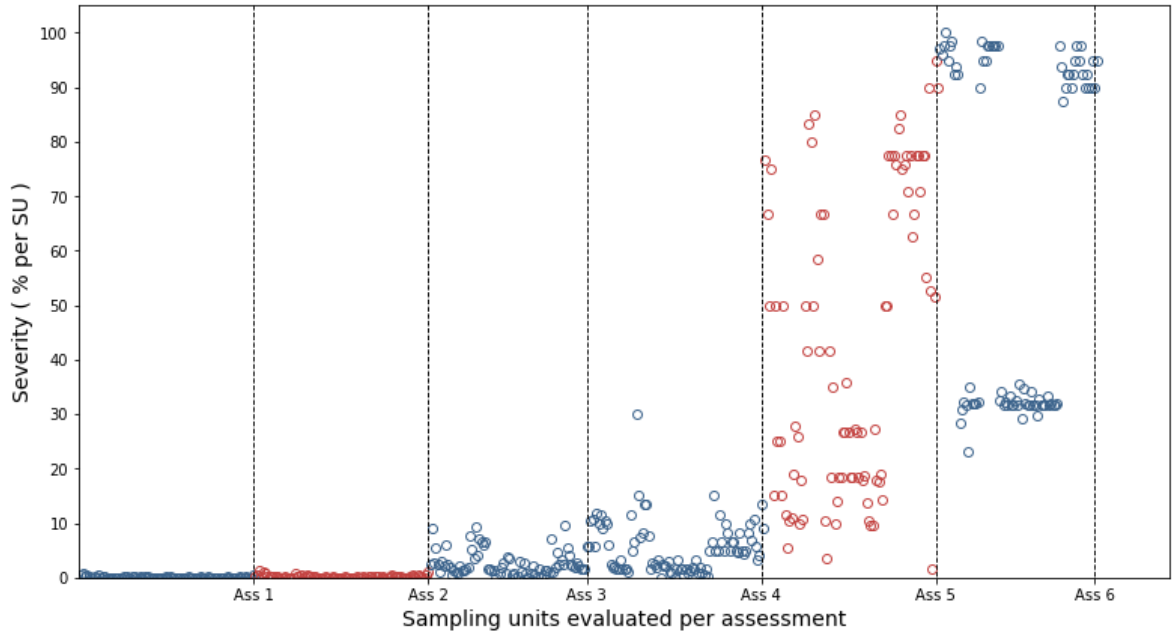


Figure 10. Disease severity at sampling unit (SU) level along the field assessments (plot 1 to plot 8 – small and large plots). In blue circles, are the sampling units within assessments dates selected for this study. In red circles, sampling units not used in this study.

It can be noticed how disease severity starts out with a low (0% - 0.8%) severity indicated by leaf area affected and progressively increases over time reaching almost 100% of disease severity in some of the SU evaluated on the last assessment. Thus, disease severity, which is strongly related with the development of disease symptoms over affected crops, had an unbalanced distribution of disease classes along the field experiment (*Figure 10*). For example, during the first assessment, sampling units evaluated were assigned mainly to the 'non-disease' class, without any sampling unit belonging to the 'above 7%' class. On the other hand, all sampling units evaluated on the last assessment date were assigned to the class 'above 7%', with no samples of 'non-disease' class captured on that date.

As disease severity levels increased over time, especially between the fourth (78 DAP) and the last assessment (86 DAP), a consistently decrease of canopy structure together with ground cover was detected after the fifth acquisition (78 DAP) towards the end of the growing season (*Figure 11*). Even though these changes can be attributed to both physiological impact of late blight disease over crop health status and natural vegetation senescence. *Figure 11* shows differences on crop structure between mixed and non-mixed system, especially for the last acquisition (86 DAP) in which treatments had more differences in comparison with other dates. As an example, it can be observed from *Figure 11* how the non-mixed system in which susceptible plant varieties were used, presented major changes in crop structure and ground coverage. In that sense, it is expected that non-disease and

infected SU within different treatment systems present variable reflectance pattern of the signal response, which is explored in the next section.

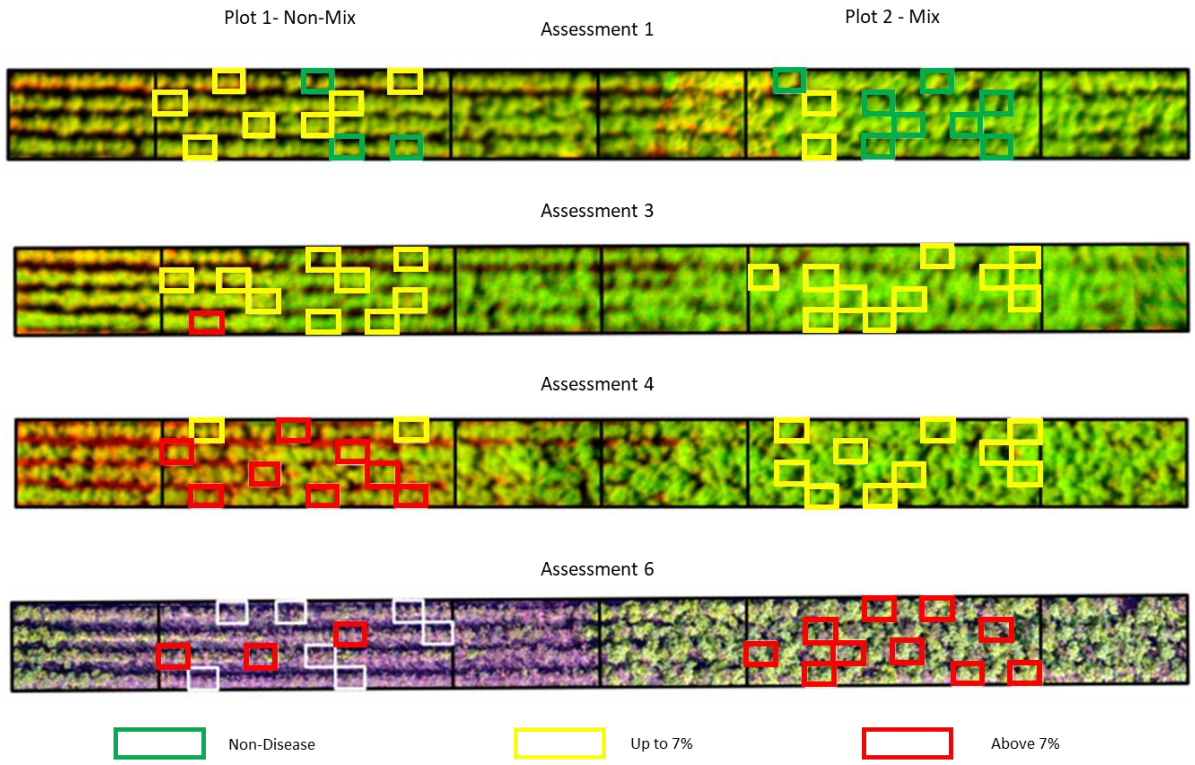


Figure 11. RGB of flight acquisitions of the first two experimental plots (located on west part of experiment) with non-mixed (Plot 1) and mixed system (Plot 2). Small rectangles indicate the evaluated sampling units.

Out of the total number of SUs evaluated in the first assessment, more than 70% of them were assigned as SU without disease, meanwhile only a small fraction of SUs evaluated in other assessments belonged to this class, consequently most of the SU related to this label were obtained during the first evaluation. On the other hand, the total (100%) of SU assessed on the last date were assigned as class 'more than 7%', hence this label was gathered mainly on the last evaluation performed. In addition, class 'up to 7%' was the only label collected from most of the assessments (*Figure 12*).

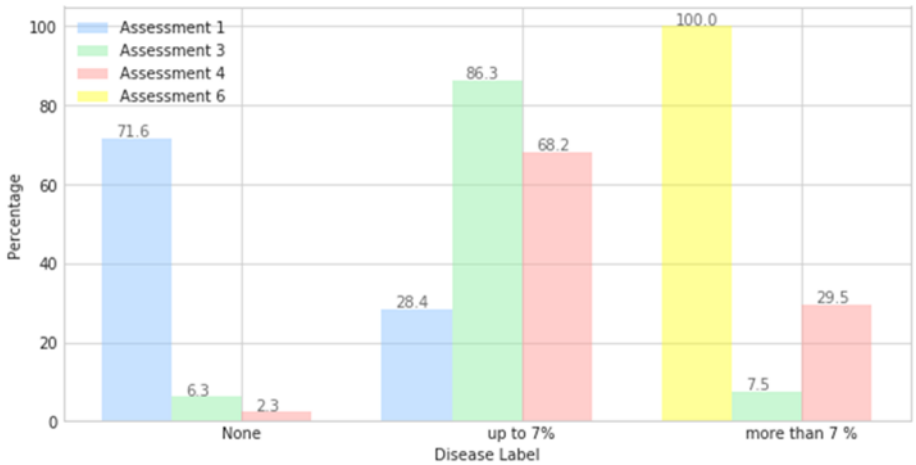


Figure 12. Percentage of disease classes per assessment

Due to disease severity distribution along the field experiment, as most of real-world datasets, labeled classes are represented by an unbalanced overall distribution. For example, a majority class can be observed in *Figure 13*, where almost half of the entire dataset is represented by the class named 'up to 7%'. Since most traditional learning techniques are usually biased to the majority class showing a poor prediction power for minority classes (Ortigosa-Hernández et al., 2017), overall class distribution within the disease dataset was taken into consideration especially for the selection of a suitable performance metric designed for unbalanced case scenarios as the one described in section 4.2.6.4, the so-called balanced accuracy.

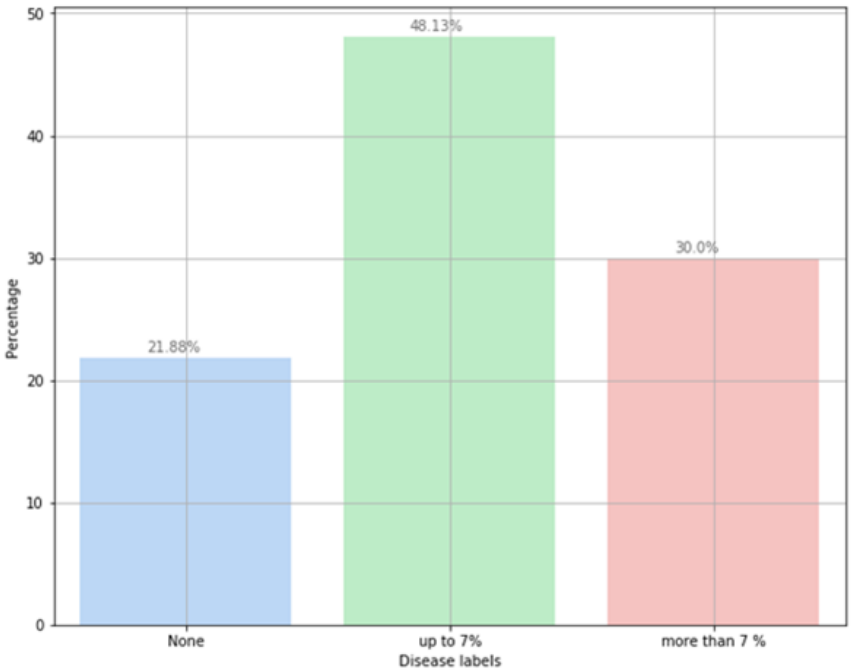


Figure 13. Disease dataset overall class distribution

Distribution of disease severity within each treatment system used in the experimental setup is shown in *Figure 14*. Around 80% of the SU evaluated along the experiment were affected by late blight disease. Out of this total, half of the SU affected corresponded to those with mixed varieties, meanwhile the remaining (124 SU) belonged to those cultivated with the non-mixed treatment (*Figure 14*). Even though the mixed system is composed of three different varieties concerning resistance to the pathogen, 41% of the total SU with disease severity above 7% are represented by SU with a mixed cultivation method. The crop treatment system has been taken into consideration in the next section, to explore if there were spectral features that could be distinguished and that could serve as discriminator between diseases classes.

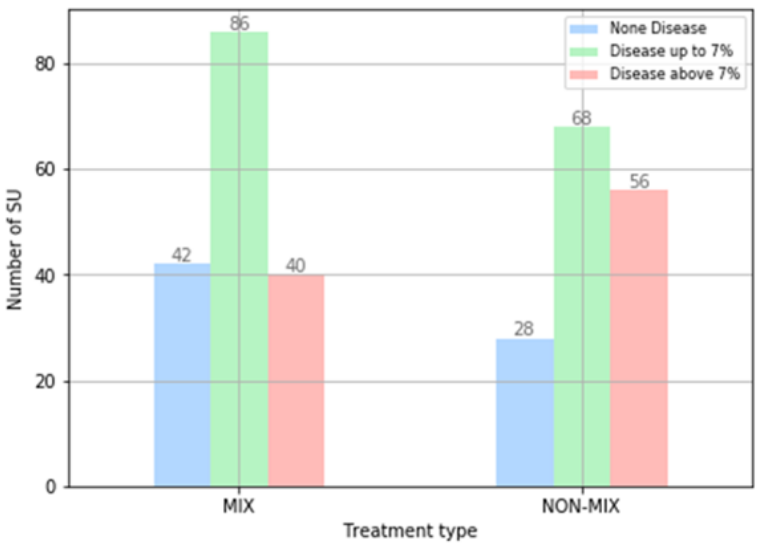


Figure 14. Number of sampling units per treatment type and disease class

5.1.2.2 Description of spectral response, crop treatment and disease evolution during growing season

Spectral measurements acquired at plot level during crop growth are summarized in *Figure 15*. Since each plant has a unique spectral signature based on their health status and intrinsic characteristics of each potato variety included within each cultivation method, it seemed important to explore and understand spectral responses of vegetation.

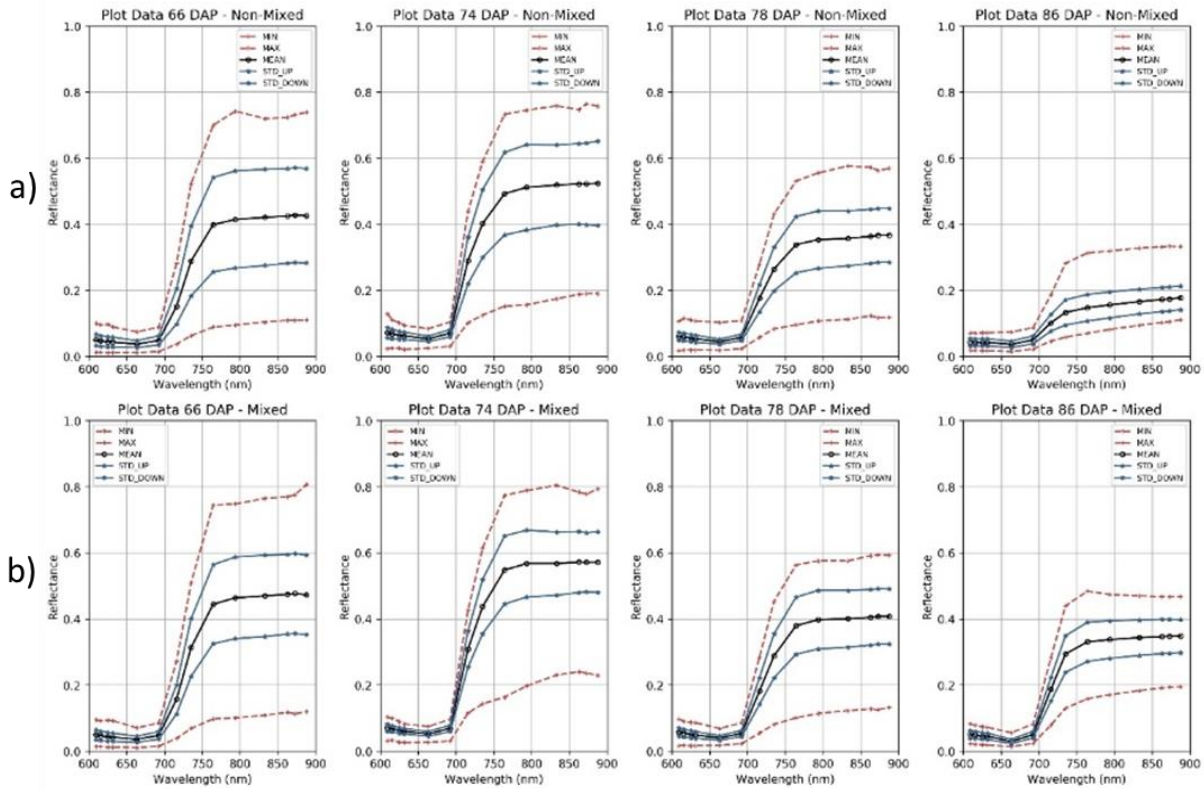


Figure 15. Description of the spectra acquired using the Rikola camera (66-78 DAP) and HYMSY camera (86 DAP) both under a UAV for each cultivation method (a) Non-Mix and b) Mix varieties) and acquisition date. Maximum and minimum spectral values in (red lines), standard deviation (blue lines), and mean spectra values in (black).

During the third acquisition date (64 DAP) which was related to the first assessment (66 DAP) as described in section 3.2.2, mean reflectance values in wavelengths affected by absorption from leaf pigments between (600 nm – 700 nm) has similar responses for both treatment types in the range of (0.035 – 0.047) reflectance values. On the other hand, mean reflectance values placed in the NIR range shows that the mixed system (*Figure 15b*) has higher values along this spectral region than non-mixed system. This difference could be mainly related to the canopy structure that was still developing, and the intrinsic characteristic of the different potato varieties used in each cultivation method.

The same differences between treatment systems were observed on the fourth acquisition (73 DAP), although mean spectral values on NIR range were higher than previous acquisition (*Figure 15*) mainly because vegetation cover reached its maximum as observed in (Figure *11*). It is important to mention that even though this acquisition provided higher reflectance values in comparison with other dates, more than 80% of SU evaluated on the third assessment (74 DAP), which was related to this acquisition, were assigned as class 'up to 7%' of disease severity (*Figure 12*).

From the fourth assessment (78 DAP) towards the end of the growing season, ground coverage together with mean spectral reflectance experienced a consistently decrease in comparison with previous dates, which can be related with the physiological impact of late blight disease and its evolution over time together with natural and intrinsic characteristics of different crop systems (*Figure 15*).

Comparison between spectral responses at SU scale per disease class and assessment is presented in *Figure 16*. No obvious differences were found between disease classes concerning spectral responses on the red region of the spectrum (609 nm - 692 nm). Furthermore, it is in the NIR range where differences between classes can be distinguished, especially among classes that represent SU without disease and those with disease severity over 7 % acquired of the last date (*Figure 16 c*). Therefore, spectral data acquired along acquisitions date seems to provide spectral information that could serve to discriminate disease severity classes.

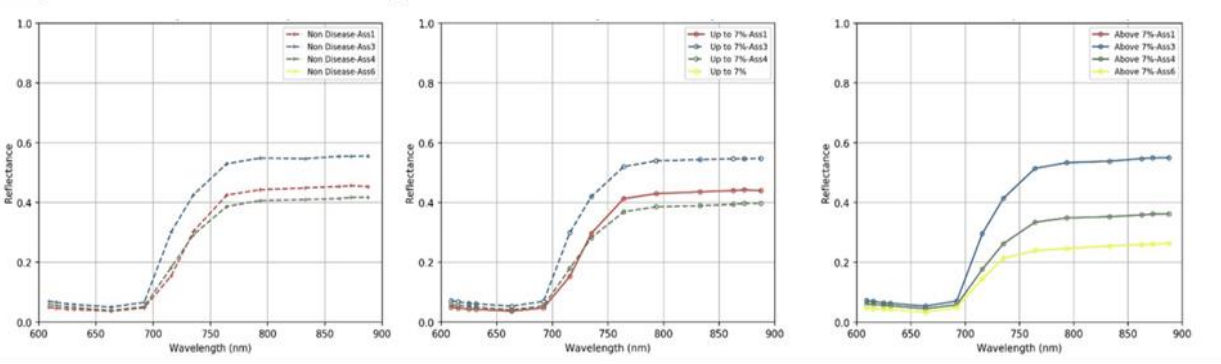


Figure 16. Mean spectral values per assessment and disease class. Figure (a) mean spectra of Non-disease class. In figure (b) mean spectra values of 'Up to 7%' disease class. Figure (c) depict mean spectral values for disease severity above 7% class.

It has been noted that highly correlated features have a negative impact on classification accuracy (Preet et al., 2015), therefore, a correlation matrix was calculated for all features within the spectral

range used (600nm – 900nm) in this study (*Figure 17*). As expected, high correlation was found between features specially in bands within NIR range. This collinearity was considered to test if applying preprocessing procedures like dimensionality reduction, could help to improve classification and generalization capabilities of the selected model.

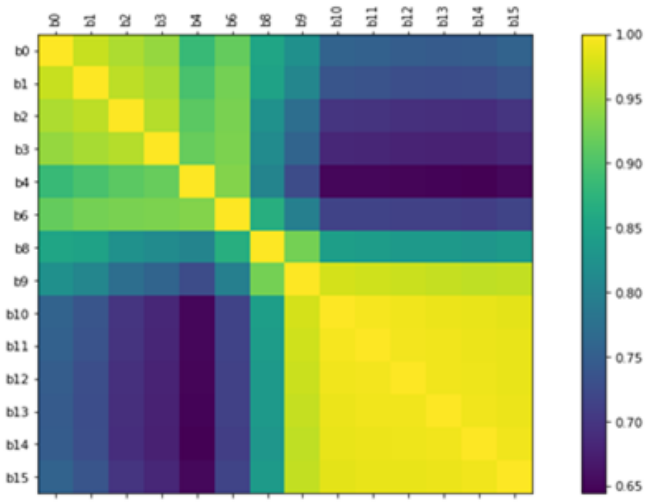


Figure 17. Disease dataset spectral features correlation matrix. Definition of bands used in the analysis can be observed in Table *2.*

As part of the data exploratory analysis, the first two principal components were calculated to visualize in a 2-dimensional scatter plot feature space distribution of observations of the disease dataset (*Figure 18*). It can be observed that there is a clear relation between the assessments date on which features were acquired, and their location within feature space, with some overlapping between the last two acquisitions where most of the SU affected by disease were gathered.

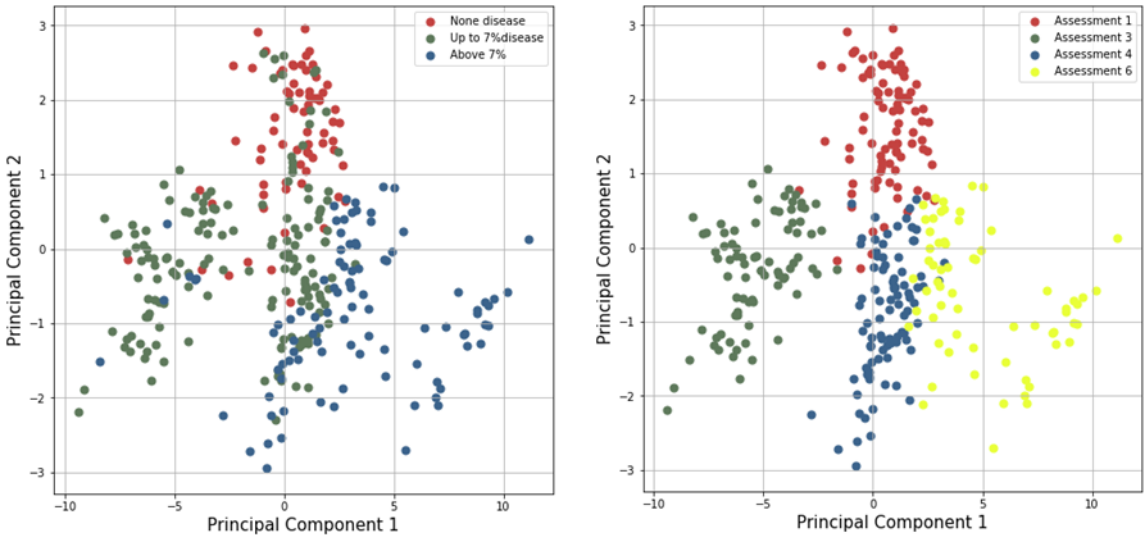


Figure 18. First two principal components of the disease dataset. Right image present features distribution over feature space with disease labels. Left image present features distribution with the assessment date as label.

5.2 MODELLING DISEASE ASSESSMENT

This section is intended to present main results obtained during the development and implementation of a support vector machine (SVM) learning algorithm which was built to predict late blight disease severity on an experimental organic potato crop. As described in section 2.4, the SVM learning algorithm has been successfully applied in various applications especially under controlled experimental setups. In that sense, this chapter tries to present obtained outcomes to verify whether the selected model can be applied with the dataset acquired during a field experiment and has a satisfying predictive performance.

5.2.1 Description of SVM model development and model selection

For many machine learning algorithms, representation of the data within model development is a crucial step. For instance, classifiers as SVM are known to be sensitive to the way features are scaled, hence preprocessing procedures as data transformation and dimensionality reduction already described, were tested to verify if they contributed to improve the model performance (Ben-Hur & Weston, 2010). Within this study, scaling of data, described in section 4.2.2, was performed. Even though the main idea was to use all features acquired during flights acquisitions to explore if the selected algorithm has the capability to find patterns within data to predict disease severity, PCA technique was tested as a dimensionality reduction preprocessing procedure due to high correlations of features discussed earlier.

Therefore, the development of the model required not only the application of the SVM, but the chaining together of the PCA and the selected machine learning algorithm, hence a pipeline was constructed together with the grid of parameters (Table *6*), which was used to test all possible combination of parameters values using the GridSearchCV tool together with the balanced accuracy metric, as the performance metric to find the best model as described in section 4.2.5. As a result, it was found that a kernelized SVM was able to better classify disease labels with the highest balanced accuracy of 82%, hereinafter model A (*Table 10*). An overview of trained models during the grid search procedure and associated performances can be observed in Appendix C.

Table 10. Best parameters set found on development set. Model A

Preprocessing		N° of components	Algorithm	Kernel	Decision function	Class Weight	C [range]	Gamma [range]
PCA	Yes	5	SVM*	RBF	OVR	Balanced	45	0.01

*Model A

Once the best parameters were obtained according to the performance metric used to reach the best score, a SVM model was re-built and trained on the full training dataset, based on the best parameters shown in *Table 10*. Moreover, it was also explored if the performance could be improved with a manual inspection and combination of those parameters, and for this step, overall accuracy was used to evaluate improvements. Results presented in *Table 11* indicate that it was possible to improve the overall accuracy in both training and test set, by reducing the number of components and decreasing the C and gamma parameter, hereinafter model B. Furthermore, the receiver operating characteristic

curve and area under the curve described in section 4.2.6, were calculated to visualize the quality of both models to discriminate each class from the others.

Table 11. Overall training and test accuracy performance of best parameters found for Model A and Model B

Preprocessing		N° of components	Algorithm	Kernel	Decision function	Class Weight	C [range]	Gamma [range]	OA Train	OA Test
PCA	Yes	5	SVM*	RBF	OVR	Balanced	45	0.01	0.80	0.77
PCA	Yes	4	SVM**	RBF	OVR	Balanced	35	0.02	0.81	0.81

*Model A; **Model B

Multiclass extension of SVM described in section 4.2.1.3, basically creates a binary model for each class which tries to separate one class from all the other classes. This allowed to calculate both the receiver operating characteristic and area under the curve per disease class, to visualize the quality of the models to classify each class from others. As described in section 4.2.6.5, the roc curve is defined by the true positive rate or recall on the X axis, and the false positive rate or false alarm ratio on the Y axis, representing relative trade-offs between true positives and false positives along different thresholds. It can be observed from *Figure 19*, that both models seem to achieve a good performance to discriminate both 'non-disease' class and disease class 'above 7 %'. On the other hand, they could not reach the same performance to correctly classify the class 'up to 7 %' from the other classes.

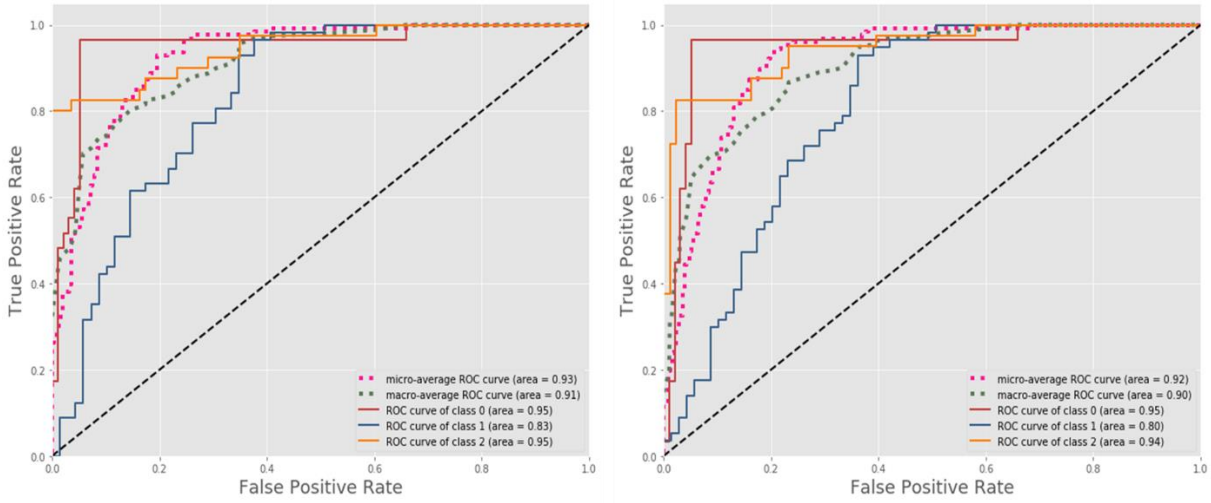


Figure 19. Receiver operating characteristic and area under the curve (Roc-auc) per disease label classification. Roc-auc curves of the best model based on the grid search cross validation (right), on the left Roc-auc curves of the manually improved model based on overall accuracy performance metric.

The quality of a model can be estimated from the roc space, in such a way that the closer the model to the upper left corner, the better it performs, representing a high sensitivity (low false negatives) and high specificity (low false positives). From *Table 12* it can be observed that both models have similar values of true positive rate (TPR) and false positive rate (FPR) for both disease classes, 'non-

disease' and 'above 7%', while Model B shows a higher value of true positive rate on the disease class 'up to 7%' than Model A. From these results it can be inferred that Model B has better capabilities to discriminate disease class 'up to 7%' than Model A.

Table 12.True positive rate and false positive rate per disease class of SVM models

	SVM* (C=45; gamma=0.01)		**SVM** (C=35; gamma=0.02)**	
Disease Class	**TPR**	**FPR**	**TPR**	**FPR**
Non-Disease	0.97	0.05	0.97	0.06
Up to 7%	0.65	0.12	0.74	0.12
Above 7%	0.83	0.17	0.83	0.10
Avg/total	0.81	0.11	0.84	0.09

TPR = True positive rate; FPR = False positive rate
*Model A; **Model B

A learning curve aims to visualize the training and validation score as a function of the number of training samples. This curve provides a quantitative view into how beneficial it will be to add training samples to the learning algorithm. Firstly, looking at the performance over the training dataset when there are a few observations, the model can fit them with high accuracy, but as new instances are added to the training set, the accuracy start to decrease until it reaches a plateau, approximately when the model is trained with 100 samples, at which point adding new instances to the training set does not make the accuracy much better or worse. Looking at the model performance on the validation data, it can be observed that when the model is trained with very few instances, it is incapable to generalize properly. Moreover, as more training samples are included in the model, the accuracy slowly increases, but it also reaches to a plateau very close to the training score curve. These learning curves could indicate an underfitting of the model reaching around 80% of overall accuracy in both curves (*Figure 20*).

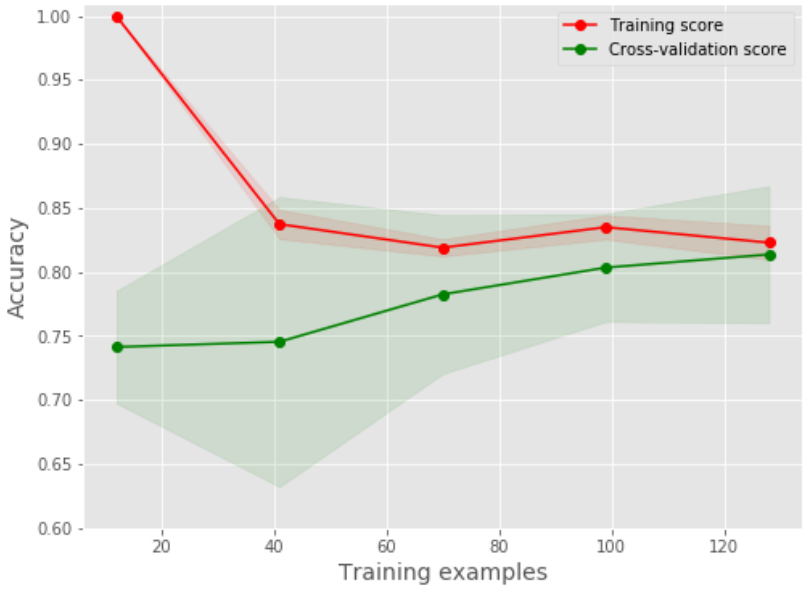


Figure 20. Learning curves of the selected SVM model B (RBF kernel, C= 35, gamma = 0,02).

5.2.2 Description of SVM predictive performance

The trained SVM model was then used to predict disease severity classes over the unseen dataset (126 SU) that was hold during the sampling phase, described in section 4.1.3. One of the most comprehensive ways to evaluate a classification and prediction task is by using a confusion matrix. Below, the confusion matrix is used to summarize the performance capability of the model in which predicted values are compared with actual labels (*Figure 21*).

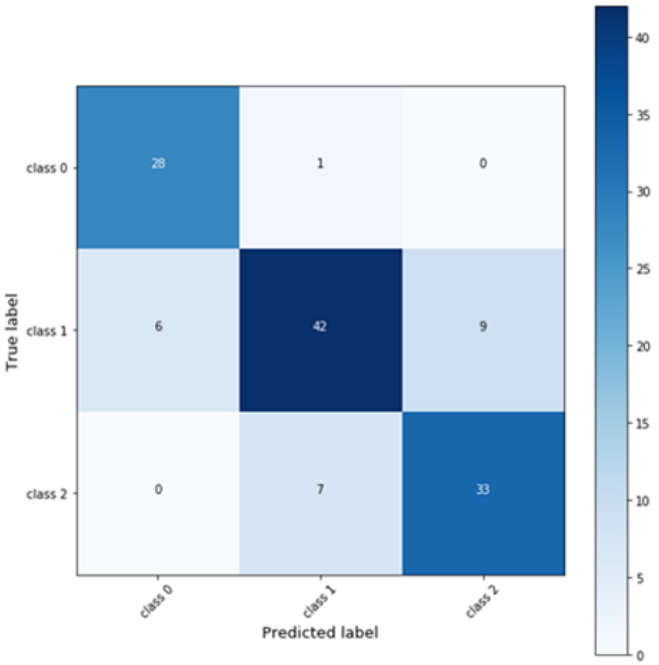


Figure 21. Model B SVM confusion matrix metric

As expected from previous analysis, Model B confusion matrix shows the capability of the model to correctly identify positives samples especially for those SUs without disease. In contrast, it seems difficult for the model to identify SUs labeled as 'up to 7%' as such, having SUs assigned to this class that were classified as other classes, for instance out of the 57 sampling units labeled as the class 'up to 7%', 15 of them were assigned to other classes. Regarding to the capability of the model to avoid false positives, it is clear to notice that none of the SU that were predicted as SU without disease were assigned to the class 'above 7%', while none of the SU predicted as 'above 7%' class were assigned as the class 'non-disease' (*Figure 21*). The predicted overall accuracy reached with the Model B was 81%, while the predicted balanced accuracy was 84% (Table *12*).

Table 13. SVM classification report

Classes	precision	recall	f1-score	support
Non-disease	0.82	0.97	0.89	29
Up to 7%	0.84	0.74	0.79	57
Above 7 %	0.79	0.82	0.80	40
avg/total	0.82	0.82	0.82	126

From confusion matrix technique, many performance metrics can be reduced to evaluate the performance of the selected algorithm. Hence, a classification report per label was elaborated to summarize firstly, the proportion of positive data points that were correctly considered as positives with respect to all positives classes, so-called recall or sensitivity, and secondly, the probability of a class that was predicted, and actually belongs to that class label, called positive predictive values or precision. It can be observed from *Table 13* that class without disease got the highest recall score indicating that fewer samples were misclassified, while class 'up to 7%' in contrast had more misclassification, although 84% of samples predicted as this label corresponded to that class.

A spatially explicit representation of results obtained with the predictive model are shown in *Figure 22*. The main objective was to compare the ground truth SU labels with those ones which were predicted by the model using the spatial context to visualize the potential and applicability of learning algorithms to contribute with the plant crop protection community. In that sense, *Figure 22* shows that, even though the SVM model had an overall good performance to generalize the learned data to unseen samples, is possible to visualize that in the first assessment all SU were predicted as 'non-disease class' (*Figure 22*) without capturing some SU that were assigned as 'up to 7%'. Similar results can be observed on the third assessment, while on the fourth assessment most of the SU predicted as 'above 7%' did not match with the actual disease label of the corresponded SU. In addition, it was not possible to find a relation between the crop system used within SU and the errors found on predictions of labels.

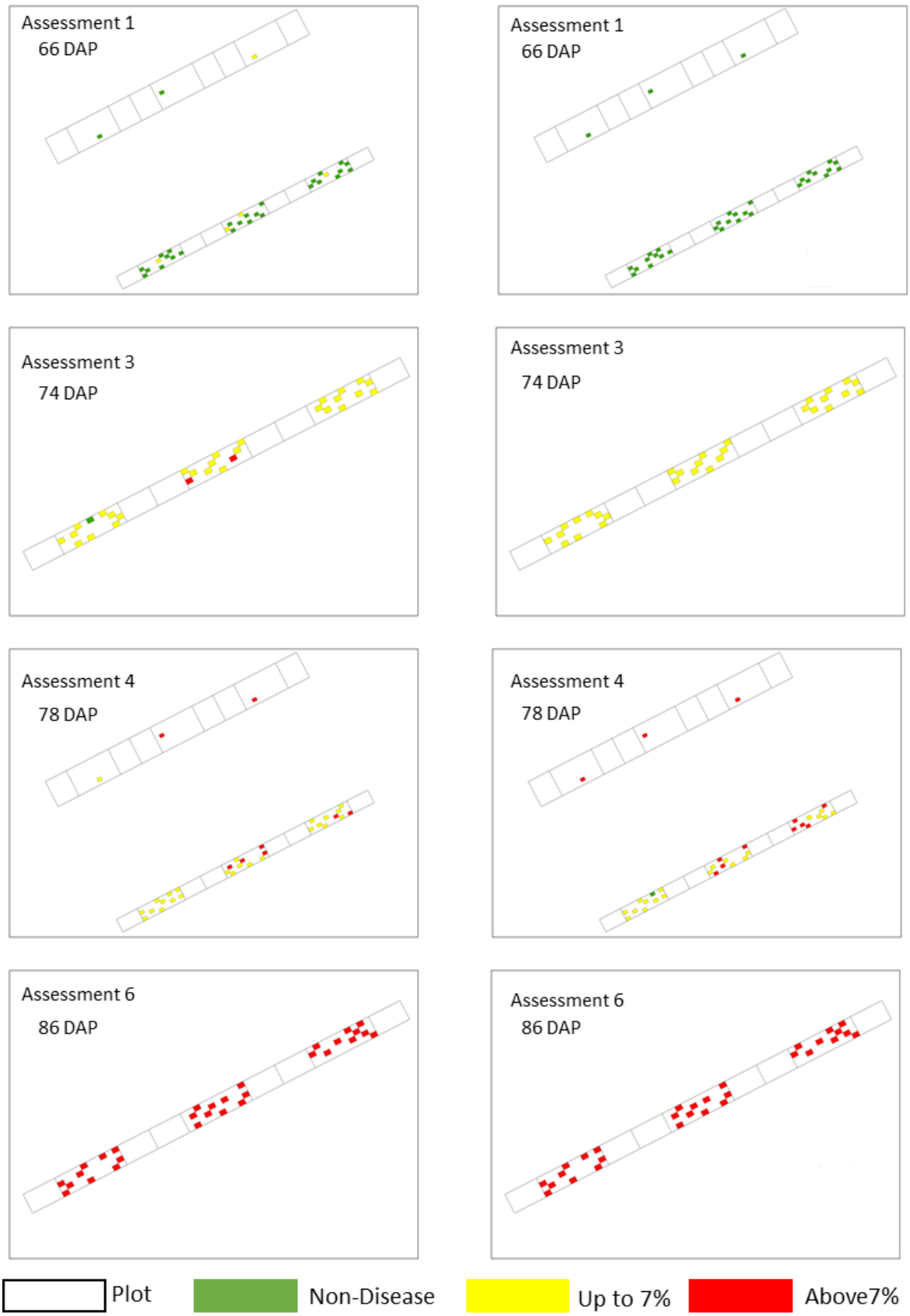


Figure 22. Comparison between grown truth labeled SU (left) and predicted SU labels (right) from the SVM model for SUs in the test plots at different assessment moments

6 DISCUSSION

The following chapter aims to discuss the main results obtained during the development and implementation of the disease predictive model generated in this study. The results are discussed in the same order as presented in chapter 5. In addition, based on an extensive literature review that contributed to answer the first research question of this study, the potential use of high-resolution UAV imagery in combination with advanced analysis techniques for plant disease assessment, is also discussed.

6.1 POTENTIAL USE OF HIGH-RESOLUTION UAV IMAGERY TO PREDICT PLANT DISEASE

This section discusses the hyperspectral sensor challenges to retrieve information that could be used to detect the physiological effects of late blight disease over a potato plant, in order to discriminate crop health status. Firstly, the discussion is focused on the use of hyperspectral technology for detection, identification, quantification and prediction of plant disease and secondly the complexity of the physiological impact of late blight disease is described.

Remote sensing described as an indirect assessment technique, has been extensively explored and used to provide information through the detection of the physical characteristics (spectral signatures) of vegetation. These characteristics can be related to their health status, based on the physiological processes and plant parameters differences, since most biotic stress factors affects both internal and external structures, which in turn has an effect on the optical properties of plants (Behmann et al., 2015). This relation between the spectral response of vegetation, which is influenced by the canopy structure and the health status of vegetation, allowed to investigate the potential use of remote sensing as a timely, location-precise, non-invasive method for the assessment of plant diseases (Franceschini et al., 2017; Jiang et al., 2012; Martinelli et al., 2015).

Results obtained during the exploratory data analysis (*Figure 16*) shows that a major reflectance difference between healthy and late blight affected sampling units can be observed in the range of 700nm and 900nm. These differences of spectral responses clearly provided information that allowed to discriminate healthy sampling units from the ones that were severe affected by the disease. Similarly, Zhang et al. (2003) explored the capability of applying hyperspectral remote sensing to monitor late blight disease in a tomato crop. They concluded that hyperspectral remote sensing bands mainly in the range of 700nm and 900nm, were more valuable and effective to detect tomatoes late blight disease. Moreover, the potential discrimination of healthy and infected tomatoes plants provided by spectral information, was possible for those plants with the higher disease levels, but not for the plants with lower severity levels. This was also observed in this study, where the discrimination of the class 'up to 7%' was less effective (*Figure 16*) in comparison with the other two classes.

Different types of sensors have been used to detect plant response to different diseases and disease severity levels (Martinelli et al., 2015). For instance, there is a growing tendency to use hyperspectral cameras combined with UAV platforms to generate high-resolution data needed for crop disease detection, identification, quantification and prediction (Garcia-Ruiz et al., 2013; Mahlein et al., 2012; Singh et al., 2016; Zhang et al., 2003). For example, Garcia-Ruiz et al. (2013), compared the use of a

hyperspectral imaging system attached to a UAV, with a similar imaging system (aircraft-based sensors), for the detection of Huanglongbing (HLB) disease affecting a citrus orchard. Authors found that at both spatial resolution (5.45 cm and 0.5 m per pixel), reflectance values at 710nm wavelengths were significantly different between healthy and HLB-infected trees. In contrast, the minimum differences identified between reflectance values were observed at 690nm.

Dissemination of late blight is mainly due to aerial dispersal of *Phytophthora infestans* sporangia from lesions on leaves and stems, or by oospores on the ground or plant material in the neighborhood (Olanya et al., 2015; Sugiura et al., 2016). This foliar and tuber disease component can destroy potato foliage within 7 to 10 days under favorable environmental conditions (Jiang et al., 2012; Olanya et al., 2015). A plant which is infected, reacts to protect itself, and this physiological reaction, will lead not only to internal structure changes, but also to external changes as the disease severity increases (Martinelli et al., 2015; Zweep, 2014).

During the host-pathogen interaction, different spectral regions of the electromagnetic spectrum can be related to specific changes in both internal and external structures of plants. For instance, thermography and fluorescence allow to detect plant stress even before symptoms are visible. Once the pathogen affects the photosynthesis apparatus, the visible region 500nm - 680nm can be used to detect changes in chlorophyll content. As soon the pathogen influences the cellular structure, variation in the near infrared wavelength are described (Thomas et al., 2018). Spectral resolution of hyperspectral sensors allowed the exploration to find specific reflectance patterns that are related to specific host-pathogens interactions.

For instance, at early stages, when visual symptoms cannot be observed yet, plants react by decreasing the photosynthesis rate, which could result in an increase of fluorescence, heat emission, and a decrease of chlorophyll content (Martinelli et al., 2015). For example, as observed from the third acquisition performed for this study (64 DAP)(*Figure 4*), canopy structure was still vigorous, specially within those plots with mixed production system (*Figure 11*). Two days later (66 DAP), the disease development was registered, when the first visual symptoms appeared. These first symptoms are generally observed in the lower leaves, where smalls, light to dark green with irregular shape lesions start to appear towards the margin of the leaf (Appendix B) (Tekos, 2010). As disease severity increases, the spots enlarge rapidly and take a brown color with well-defined edges until the lesions turn black as the affected leaves start rotting, while at canopy scale, disease impact can change canopy density and vegetation coverage (Martinelli et al., 2015; Sankaran et al., 2010; Tekos, 2010).

Since one of the main characteristic of late blight disease is that it spreads extremely rapidly during growing season, with devastating results within the second or third week after the first visible symptoms appears if no control measures are taken, especially in those production systems where no chemicals are applied in the field (Gebru et al., 2017; Lammerts van Bueren et al., 2008; Nowicki et al., 2012; Sugiura et al., 2016; Zhang et al., 2003), there is a need in the plant breeding and crop community, to explore alternative methods that could contribute to overcome the limitations encountered in conventional methods, such as the time-consuming and subjective labor, described for the visual assessment (Sugiura et al., 2016).

Even though hyperspectral sensors are widely explored and used to extract physical variables that can be related to vegetation properties to estimate the vegetation health status, there are many internal and external factors that make this method too error prone. Some of the main challenges in the implementation of this advanced retrieval method such as heterogeneity of elements of the background, features correlation, among others are discussed in the next section.

6.2 CHALLENGES IN THE IMPLEMENTATION OF HYPERSPECTRAL TECHNOLOGY FOR PLANT DISEASE ASSESSMENT

The vast amount of data that can be retrieved from hyperspectral imagery, requires the exploration and implementation of complex analysis to find underlaying patterns, that could be related to the disease characteristic and its progress over the time (Behmann et al., 2014; Jiang et al., 2012; Mahlein et al., 2013). However, high dimensionality of the hyperspectral data where wavelengths next to each other are highly correlated, could have a negative impact on classification and prediction performance (Mahlein et al., 2013; Preet et al., 2015). Hence, many approaches have been employed in hyperspectral data analysis to overcome the effect of collinearity between acquired features.

In this study, a high correlation was found between spectral features, especially between bands within NIR range. Therefore, a principal component analysis (PCA) was included within a pipeline, to evaluate if this preprocessing procedure contributed in the selection of the best model. The PCA method involves a linear transformation of the original data into the so-called components, which are statistically uncorrelated keeping the maximum amount of information (Gilbertson & van Niekerk, 2017; Muller & Guido, 2016; Preet et al., 2015). As a result, the best model identified for this study (*Table 10*) was obtained when the principal component analysis was introduced as a preprocessing step within the pipeline, where the first five components were selected to train the predictive model.

Although this dimensionality reduction technique has been extensively used, especially for classification tasks (Rivera-Caicedo et al., 2017), authors like Rivera- Caicedo et al. (2017), stated that, in principle, other techniques such as partial least squares (PLS) or canonical correlation analysis (CCA), could be used instead. However, they concluded that when combining dimensionality methods with non-linear machine learning algorithms, it is the principal components analysis method together with partial least squared, the top performing techniques in terms of accuracy. Moreover, authors like Kong et al. (2018) proposed the use of second derivative spectra and PCA loadings to select the optimal wavelengths to detect *sclerotinia sckeritiorum* on oilseed rape stems.

Alternatively, other feature selection techniques, in which the original data is not transformed, has been also used for the identification of an optimal subset of features (Bazi & Melgani, 2006; Kong et al., 2018; Maldonado et al., 2014). For instance, authors like Maldonado et al. (2014), proposed an embedded method for backward feature elimination, inspired in the recursive feature elimination procedure (RFE), which aims to exclude those features that do not contribute to the model performance. It was found that the proposed method outperformed other ranking techniques, such as the recursive feature elimination wrapped with a SVM.

Furthermore, Bazi and Melgani (2006), compared the performance of a SVM classifier trained with the entire set of features, to classify nine different types of land cover classes, with the performances of the models trained with an optimal subset of features, applying different feature selection methods, such as the recursive feature elimination (RFE) and the radius margin bound minimization method. This comparative study, found that the recursive feature elimination, together with the radius margin bound minimization feature selection method, showed relatively poor performances in terms of classification accuracy, in comparison with the SVM model trained with all features.

Implementation of hyperspectral technology also requires considering factors related to the heterogeneity of elements of the background, disease characteristic , illumination, the effect of the atmosphere, which forces the methods to rely on very tenuous differences to discriminate among them (Barbedo, 2016; Hy et al., 2011). Therefore, in this study a NDVI threshold approach was used to exclude soil (non-vegetation) and excessive unwanted non-canopy pixels. As a result, several pixels were removed from each sampling unit which were considered non-vegetation pixels (*Figure 7*), mostly from those sampling units evaluated on the last date, where the potato plants were severely affected by the disease (*Figure 11*) . However, it is known that this vegetation index is largely affected by background effects, therefore, several indices have been proposed to correct background effect such as the weighted difference vegetation index (WDVI), or the optimized soil-adjusted vegetation index (OSAVI)(Franceschini et al., 2017). Besides the use of vegetation indices, other techniques have been proposed to overcome this issue. For example, authors like Hy et al. (2011), investigated the use of machine vision techniques to develop a plant segmentation algorithm to detect individual plants under outdoor light condition imagery. Object oriented analysis has been explored and compared with per-pixel analysis techniques, where accuracy of vegetation parameters estimation and discrimination of above ground biomass were significantly higher when object-oriented techniques were applied (Addink et al., 2007).

Besides the limitation of the NDVI-threshold approach, Garcia-Ruiz et al. (2013) used a similar approach, with the exception that the analysis was performed for a citrus orchard. In their study, citrus trees were manually segmented using a 0.2 threshold value for NDVI such that the pixels covering the tree canopy were included, while pixels (non-vegetation) were excluded. They concluded that their approach successfully achieved the expected result. It is also important to mention, that as explained before, the occurrence of different degrees of severity level on each plant within each SU varied, therefore, the analysis should be also be evaluated at a pixel level scale, to test if the model capture the variability of information within each SU, which could improve the predictive performance (Behmann et al., 2014).

Although, this study considered soil-background affect, there are many external factors especially attributed with the use of UAV platforms and the environmental conditions during measurements that could influence the performance of the model. Factors like flight conditions, illumination conditions (specific illumination and sun-target-sensor geometry) and overcast conditions (cloud, wind) can affect the accuracy of the spectral information over time, these variables should be explored in more detail to understand their influences on the performance of the model. Hence, there still is a need to explore methods for preprocessing procedures that can be implemented and incorporated to automated systems (Thomas et al., 2018).

6.3 MODELLING DISEASE ASSESSMENT

A support vector machine (SVM) learning algorithm was selected to explore its capability to assess late blight disease. As recommended by Hsu et al. (2008), the grid search approach together with cross validation was implemented in this study to evaluate several parameters values combinations (*Table 6*). As a result of the procedure, it was found that applying a PCA as a preprocessing step together with a radial basis function (RBF) kernelized SVM, achieved the best performance, reaching 82% of balanced accuracy (*Table 10*).

One of the most important procedures that aims to improve a model's generalization performance is called hyperparameter optimization. Tuning hyperparameters to configure learning algorithms can influence the effects on the resulting performance of the model, hence, it is considered a best practice when a machine learning algorithm is under development (Hsu & Chang, 2008; Huang et al., 2017; Koch et al., 2017).

Even though, the combination of the manual and grid search is the most applied method for hyperparameter optimization, there are many advanced optimization methods that have been explored to improve the selection of optimal parameters (Bergstra & Yoshua, 2012). For instance, Bergstra and Yoshua (2012) proposed the use of random search technique as a substitute of grid search. They described the grid search approach as a naïve, ineffective and a poor choice for configuring algorithms, especially for new datasets.

However, there are some reasons that explain why the grid search method was implemented in this study despite other advanced methods. Firstly, even though the grid search is described as a computer intensive technique because it grows exponentially with the number of hyperparameters that will be evaluated (Koch et al., 2017), the time and computer resources required to find the best parameters due to the size of the disease data set, allowed its implementation. Secondly, the scikit-learn package, as implemented in this study, provides the GridSearchCV tool, which is simple to implement and takes advantage of the inherent ability of this approach to train and evaluate all models in parallel (Bergstra & Yoshua, 2012; Chih-Wei Hsu, Chih-Chung Chang, 2008; Koch et al., 2017; Muller & Guido, 2016). Finally, given the relatively low number of features used in this study (17 features), the grid search approach was considered a reliable approach for low dimensional spaces as described by Bergstra and Yoshua (2012).

Authors like Rumpf et al. (2010) applied the grid search approach to search for the best radial basis function and an appropriate factor for penalizing classification errors, to find the best classification method to discriminate, differentiate and identify diseases on sugar beet leaves, even before specific symptoms became visible. Authors concluded that early differentiation between healthy and inoculated plants can be achieved using kernelized radial basis function support vector machine, reaching accuracies between 65% - 90% depending on the type and stage of the disease.

Similarly, Kong et al. (2018) evaluated the performance of several discriminant models to detect *sclerotinia sckeritiorum* disease on oilseed rape stems. The optimal combination of parameters of the support vector machine with a radial basis function kernel evaluated by the authors, was obtained through a grid search approach, achieving overall accuracies over 90%, for both the training and test

set. In addition, Huang et al. (2017) despite to describe several currently techniques to select model parameters such as the gradient descent method, Bayesian method, artificial immune, genetic algorithm, among others, decided to implement the grid search cross validation method for optimal hyperparameter selection, to find the best classifier to detect sugarcane borer diseases.

In general, for most datasets only a few hyperparameters are relevant, but different hyperparameters are important on different datasets (Bergstra & Yoshua, 2012). This phenomenon described by Bergstra and Yoshua (2012), seems to reveal a major drawback in the procedures to derive the optimal hyperparameters, and use them with new datasets. In that sense, many efforts are being made for the development of new hyperparameter optimization algorithms that could allow the use of the trained model for broadly applications.

The Performance metric plays a fundamental role to achieve the optimal algorithm during the model selection (Brodersen et al., 2010; Ferri et al., 2009). For classification tasks, there are several metrics that have been explored and used over the years, however, some of them without a consensus and a clearly justified theoretical basis, especially for the multiclass imbalanced problem (Ferri et al., 2009; Mosley, 2013).

The accuracy is one of the most widely performance metric used for the assessment of supervised machine learnings algorithms, mainly because it is described as a metric that measures overall effectiveness of an algorithm (Bekkar et al., 2013). However, this metric is more appropriate for balanced datasets, because in some cases it can be biased to the majority class, therefore results can be overoptimistic. Many different measures have been proposed to overcome this limitation, yet most of the works reviewed for this study, relies on the accuracy to evaluate their classification and prediction performances (Bekkar et al., 2013; Mosley, 2013).

Due to the characteristics and structure found in the disease dataset, in which the three target classes have an overall unbalance distribution, the balanced accuracy metric was proposed as the measure to evaluate the models trained during the grid search cross validation procedure. This metric basically measures the performance of the classifier for each class separately, and then an average of the performances of classes is calculated. Hence, if the model performs well in all classes, the balanced accuracy will be equal to the conventional accuracy, but if the accuracy is biased towards the majority class, then the balanced accuracy will balance the score (Brodersen et al., 2010).

Authors like Mahlein et al. (2013) and Mosley (2013), explored and used the balanced accuracy as a performance. However, it is important to mention that using balanced accuracy, the metric is based on how well the model avoids the false negatives (recall), therefore it neglects how well the classifier is performing to avoid the false positives (precision), which means that the evaluation is based on the classification power of the algorithm instead of the prediction capabilities.

The improvement of the performance of the model that was found with the grid search procedure in this study (Model A), was evaluated using the overall accuracy. Furthermore, the ROC curve and the AUC were also calculated to visualize the capability of both models to discriminate each class from the other classes (*Figure 19*). Even though overall accuracy indicated that Model B outperformed in both, training and test set, a slightly higher AUC metrics were observed in Model A in comparison with

Model B, especially for the ROC curve for the class 'up to 7%'. The fact that both metrics evaluated different aspects, generated an interest of different authors to explore their relationship. Authors like Cortes and Mohri (2003), made a statistical analysis of this relationship and concluded that ''the average AUC is monotonically increasing as a function of the classification accuracy, but that the standard deviation for uneven distributions and higher error rates is noticeable. Thus, algorithms designed to minimize the error rate may not lead to the best possible AUC values". In that sense, those slightly differences between performances can be expected, and for this study, ROC curve was only used to visualize the capability of the model to discriminate one class from the other classes.

The learning curve of the selected model (Model B) was elaborated to visualize and describe the relation between the training size and the classifier performance. This relation seemed to be important to visualize, because it is known that achieving a good classification performance is usually related to the amount of available training data (Weiss & Battistin, 2014). At some point, it seems that the model cannot improve its classification performance as data is shown to the model, therefore, instead of adding more training samples, training the current learning algorithm with more features could be tested. Several approaches from different domains such as computer vision, deep learning and remote sensing, have been proposed in literature as alternatives to add relevant features that could help to improve the model performance. Some of these techniques are based on spectral vegetation indices (SVIs), spectral disease indices (SDIs), while others are based on the idea that colors hold a large part of relevant information, descriptors such as histogram of colors, histograms of gradients and histogram of edge orientations (Bileschi & Wolf, 2007; Gao et al., 2018; Mahlein et al., 2013; Rahman et al., 2017; Rey et al., 2017).

6.4 SVM PREDICTIVE PERFORMANCE

The kernelized SVM-RBF learning algorithm implemented in this study achieved a good performance identifying each class as such, especially for those classes where SUs were assigned as 'non-disease' and those with the highest disease severity so-called 'above 7%' (*Figure 21*). Therefore, the model has the capability to avoid false negatives (FN) for those classes with high accuracy, indicating the potential use of hyperspectral information to detect SUs severely affected and those without disease. Since one of the main characteristics of late blight disease is that it can disseminate over the field in a short period of time , around twenty days after the first symptoms appeared, as observed in this study (66 DAP – 86 DAP)(*Figure 11*) (Gebru et al., 2017; Lammerts van Bueren et al., 2008), an important requirement for a model, is to reach low false negative percentages, mainly because in practice high number of false negatives could led to inappropriate crop management decisions. Similar results were obtained by Garcia-Ruiz et al. (2013), where a SVM with kernel with 85% accuracy and 11% false negatives were obtained, they concluded that the use of UAV at low altitudes could become a promising tool for disease detection.

On the other hand, the predictive power of the model is explained through its capability to avoid false positives, the average positive predicted value (PPV) for the classes reached by the predicted algorithm was 82% (Appendix E). The capability to avoid false positives was less effective between the class "up to 7%' and the other classes. A predictive model requires low numbers of predicted false positives, since predicting a sampling unit as severely affected by the disease while it is not, could also led the farmers to take drastic measures to control the disease with unreliable information. There are

few reported activities that uses machine learning as a predictive tool of plant stress (Kaundal et al., 2006; Singh et al., 2016). Authors like Behmann et al. (2014) focused their work on the prediction of drought-induced stress, based on a pixel scale to predict local stress for each plant. Authors found that the SVM one-versus-all approach achieved only (66%) of accuracy, probably related to the low number of features. On the other hand, even though linear SVM did not have the best performance, it has a potential applicability, due to its capacity to handle vast amount of data, which is required when hyperspectral data is used. Other authors like Kaundal et al. (2006), explored the use of weather variables as predictors, to introduce a prediction approach based on support vector machines, for the forecasting of rice blast disease. They concluded that SVM-based regression provided a better description of the relation between weather variables and the disease level.

6.5 FURTHER RESEARCH

The importance of an automated systems that could provide reliable information for decision making is considered a relevant one the most relevant challenge for the precision crop community(Thomas et al., 2018). Many efforts have been made to better understand the relation between the vegetation characteristics and the spectral responses captured by different non-invasive optical sensors over the past years. The combination of high-resolution and UAV platforms explored in this study, allowed to understand the potential of these techniques, together with complex analysis such as machine learning, to explore the potential applicability to monitor plant stress, host-pathogen relation and the spatial spread of the late blight disease over the growing season.

However, there are several factors that need to be considered, in order to develop a reliable, timely and automated application that could be used by the crop protection community. For instance, environmental conditions of each acquisition can affect the intensity of reflected light, the selection of specific spectral ranges according to the host-pathogen system among others, are crucial steps to be incorporated for the development of better models. In that sense, authors like Thomas et al. (2018) propose that disease-specific spectral signatures seem to be transferable between different scales, this could allow to explore if hyperspectral signatures ranges acquired under controlled conditions, can be transferable to field-based experiments. In addition, Behmann et al. (2015) states the need to implement adaptive analysis pipelines within the development of reliable and easy to interpret outcomes.

7 CONCLUSION

The main objective of this research was to explore the potential use of machine learning techniques for the assessment of late blight disease in an organic potato production system from high-resolution UAV imagery. Moreover, this study explored which discriminant function (linear – non-linear) of a support vector machine classifier provided the best solution and performance to predict late blight disease. The experiment was conducted over eight experimental plots, where two different production systems were evaluated through visual assessments (ground truth), together with the acquisition of high-resolution UAV imagery. The combination of both datasets was used as input to train and evaluate the selected predictive model.

A support vector machine learning algorithm was used to classify three different disease severity classes. As expected, a high correlation between spectral features was found. Besides the use of all spectral features to train the model, the introduction of a PCA as a feature selection technique was also evaluated. Results obtained from the grid search procedure indicated that the incorporation of a the PCA technique within the pipeline, outperformed other models, therefore the best performance obtained during the grid search, reaching 82% of accuracy (balanced accuracy), was obtained when the SVM model was trained with the first five uncorrelated PCA components.

In addition, both a linear and non-linear SVM discriminant function was tested, to explore which extension of the classifier was more suitable to predict late blight disease. As most of the researches reviewed for this study, among the methods evaluated, a radial basis kernelized SVM algorithm was selected during the model selection procedure. The combination of grid-search and manual search methods to find the best combination of hyper-parameters of the model, allowed an improvement of (1% and 4%) on the overall accuracy in both train and test set respectively. However, there is a need to incorporate and select features based on specific characteristics of the physiological impact of the disease and its relation to the disease severity. This is required to provide more discriminatory information to the model that could turn into an improvement of the model performance.

The discrimination capability between classes evaluated with the receiver characteristic operating (ROC) curve, helped to visualize the quality of the trained model to discriminate the most severe affected sampling units from those without disease symptoms. As the disease severity increased, the model was able to separate with a good performance the class ‘above 7%’ from the other classes, while with those sampling units assigned to low disease severity values, the model did not reach the same performances. This capability can be related to the spectral responses of the sampling units affected by the disease, especially in the range of 700nm and 900nm, which provides discriminatory information due to the difference found in the spectral responses within this range between both, the healthy sampling units and the most affected ones.

Based on the literature reviewed and the results obtained in this study, aerial remote sensing with high spectral and spatial resolution has a great potential to generate information that can be used to assess late blight disease. The capability of the machine learning techniques such as the support vector machines in combination with the vast amount of data acquired with hyperspectral technology, serves

as a basis to the development of applications to monitor crop health status. However, it is still necessary to consider factors such as environmental conditions during flight acquisitions, image preprocessing and the selection of well-known specific features for the development of automated systems that can provide timely, non-invasive, and reliable information to forecast temporal and spatial disease spread, information that can be use by the crop protection community.

This study was focused on the exploration of the capability of a machine learning algorithm to predict late blight disease. There are several external factors that needs to be considered to better understand the host-pathogen relation and how this influence the spectral responses of vegetation. The incorporation of specific features, related to the disease severity could provide enough information to improve the classification and prediction performances, especially at lower disease severity levels. On the other hand, the need of timely information from the crop management community, requires the consideration of the applicability of the model since the beginning of the project.

8 REFERENCES

Abe, S. (2010). *Support Vector Machines for Pattern Classification*. *Pattern Recognition Letters* (Vol. 26). https://doi.org/10.1007/978-1-84996-098-4

Addink, E. A., de Jong, S. M., & Pebesma, E. J. (2007). The Importance of Scale in Object-based Mapping of Vegetation Parameters with Hyperspectral Imagery TT -. *Photogrammetric Engineering & Remote Sensing TA -*, *73*(8), 905–912.

Agrios, G. (2005). Plant pathology. *Molecular Plant Pathologyathology*, *15*(4), 315–318. https://doi.org/10.1111/mpp.12135

Alpaydin, E. (2014). *Introduction to Machine Learning*. The MIT Press.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, *19*(10), 535–544. https://doi.org/https://doi.org/10.1016/j.tree.2004.07.021

Ashourloo, D., Aghighi, H., Matkan, A. A., Mobasheri, M. R., & Rad, A. M. (2016). An Investigation Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(9), 4344–4351. https://doi.org/10.1109/JSTARS.2016.2575360

Baranowski, P., Jedryczka, M., Mazurek, W., Babula-Skowronska, D., Siedliska, A., & Kaczmarek, J. (2015). Hyperspectral and thermal imaging of oilseed rape (Brassica napus) response to fungal species of the genus Alternaria. *PLoS ONE*, *10*(3), 1–19. https://doi.org/10.1371/journal.pone.0122913

Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering*, *144*, 52–60. https://doi.org/10.1016/j.biosystemseng.2016.01.017

Bazi, Y., & Melgani, F. (2006). Toward an Optimal SVM Classification System for Hyperspectral Remote Sensing Images. *Geoscience and Remote Sensing, IEEE Transactions on*, *44*(11), 3374–3385. https://doi.org/10.1109/TGRS.2006.880628

Behmann, J., Mahlein, A. K., Rumpf, T., Römer, C., & Plümer, L. (2015). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture*, *16*(3), 239–260. https://doi.org/10.1007/s11119-014-9372-7

Behmann, J., Schmitter, P., Steinrücken, J., & Plümer, L. (2014). Ordinal classification for efficient plant stress prediction in hyperspectral data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *40*(7), 29–36. https://doi.org/10.5194/isprsarchives-XL-7-29-2014

Behmann, J., Steinrücken, J., & Plümer, L. (2014). Detection of early plant stress responses in hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *93*, 98–111. https://doi.org/10.1016/j.isprsjprs.2014.03.016

Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, *3*(10), 27–38. Retrieved from http://www.iiste.org/Journals/index.php/JIEA/article/view/7633

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. *Methods in Molecular Biology (Clifton, N.J.)*, *609*, 223–239. https://doi.org/10.1007/978-1-60327-241-4_13

Bergstra, J., & Yoshua, B. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, *13*, 281–305. https://doi.org/10.1162/153244303322533223

Bileschi, S., & Wolf, L. (2007). Image representations beyond histograms of gradients: The role of Gestalt descriptors. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2007.383122

Bock, C. H., Poole, G. H., Parker, P. E., & Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences*, *29*(2), 59–107. https://doi.org/10.1080/07352681003617285

Bravo, C., Moshou, D., West, J., McCartney, A., & Ramon, H. (2003). Early Disease Detection in Wheat Fields using Spectral Reflectance. *Biosystems Engineering*, *84*(2), 137–145. https://doi.org/https://doi.org/10.1016/S1537-5110(02)00269-6

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. *Proceedings - International Conference on Pattern Recognition*, 3121–3124. https://doi.org/10.1109/ICPR.2010.764

Campbell, J. B. (2002). Introduction To Remote Sensing, 341.

Chih-Wei Hsu, Chih-Chung Chang, and C.-J. L. (2008). A Practical Guide to Support Vector Classification. *BJU International*, *101*(1), 1396–1400. https://doi.org/10.1177/02632760022050997

Claesen, M., & De Moor, B. (2015). Hyperparameter Search in Machine Learning, 10–14. Retrieved from http://arxiv.org/abs/1502.02127

Clevers, J. G. P. W., & Kooistra, L. (2012). Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(2), 574–583. https://doi.org/10.1109/JSTARS.2011.2176468

Clevers, J. G. P. W., & Kooistra, L. (2012). Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(2), 574–583. https://doi.org/10.1109/JSTARS.2011.2176468

Cooke, D. E. L., Cano, L. M., Raffaele, S., Bain, R. A., Cooke, L. R., Etherington, G. J., … Kamoun, S. (2012). Genome Analyses of an Aggressive and Invasive Lineage of the Irish Potato Famine Pathogen. *PLOS Pathogens*, *8*(10), 1–14. https://doi.org/10.1371/journal.ppat.1002940

Cortes, C., & Mohri, M. (2003). AUC Optimization vs. Error Rate Minimization. *Nips*, 313–320. https://doi.org/10.1.1.9.3518

Cox, V. (2017). Translating Statistics to Make Decisions. https://doi.org/10.1007/978-1-4842-2256-0

Dale, L. M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., & Pierna, J. A. F. (2013). Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review. *Applied Spectroscopy Reviews*, *48*(2), 142–159. https://doi.org/10.1080/05704928.2012.705800

European and Mediterranean Plant Protection Organization [EPPO]. (2008). Phytophthora infestans on potato. *EPPO Bulletin*, *38*(3), 268–271. https://doi.org/10.1111/j.1365-2338.2008.01224.x

European Commission. (2002). EC No 473/2002. *Official Journal of the European Communities*, (473), 21–24.

Fabre, F., Rousseau, E., Mailleret, L., & Moury, B. (2012). Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytologist*, *193*(4), 1064–1075. https://doi.org/10.1111/j.1469-8137.2011.04019.x

Faraway, J. J. (2016). Does data splitting improve prediction? *Statistics and Computing*, *26*(1–2), 49–60. https://doi.org/10.1007/s11222-014-9522-9

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Ferri, C., Hernández-Orallo, J., & Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, *30*(1), 27–38. https://doi.org/10.1016/j.patrec.2008.08.010

Food and Agriculture Organization of the United Nations [FAO]. (2014). FAOSTAT. Retrieved September 20, 2017, from http://www.fao.org/faostat/en/#data/QC

Franceschini, M. H. D., Bartholomeus, H., van Apeldoorn, D., Suomalainen, J., & Kooistra, L. (2017). Intercomparison of unmanned aerial vehicle and ground-based narrow band spectrometers applied to crop trait monitoring in organic potato production. *Sensors (Switzerland)*, *17*(6). https://doi.org/10.3390/s17061428

Fry, W. (2008). Phytophthora infestans: The plant (and R gene) destroyer. *Molecular Plant Pathology*, *9*(3), 385–402. https://doi.org/10.1111/j.1364-3703.2007.00465.x

Gao, Q., Lim, S., & Jia, X. (2018). Hyperspectral Image Classification Using Convolutional Neural Networks and Multiple Feature Learning. *Remote Sensing*, *10*(2), 299. https://doi.org/10.3390/rs10020299

Garcia-Ruiz, F., Sankaran, S., Maja, J. M., Lee, W. S., Rasmussen, J., & Ehsani, R. (2013). Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Computers and Electronics in Agriculture*, *91*, 106–115. https://doi.org/10.1016/j.compag.2012.12.002

Gebru, H., Mohammed, A., Dechassa, N., & Belew, D. (2017). Assessment of production practices of smallholder potato (Solanum tuberosum L .) farmers in Wolaita zone , southern Ethiopia. *Agriculture & Food Security*, 1–11. https://doi.org/10.1186/s40066-017-0106-8

Gilbertson, J. K., & van Niekerk, A. (2017). Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Computers and Electronics in Agriculture*, *142*, 50–58. https://doi.org/10.1016/j.compag.2017.08.024

Girma, H. (2009). A Tutorial on Support Vector Machine. *Pattern Recognition*, 18. Retrieved from http://lab.fs.uni-lj.si/lasin/wp/IMIT_files/neural/doc/seminar6.pdf

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. The University of Waikato. Retrieved from https://www.lri.fr/~pierres/donn%E9es/save/these/articles/lpr-queue/hall99correlationbased.pdf

Härdle, W. K. (2011). *Handbooks of Computational Statistics*. *Methods*. https://doi.org/10.1007/978-3-642-16345-6

Harwin, S., Lucieer, A., & Osborn, J. (2015). The impact of the calibration method on the accuracy of point clouds derived using unmanned aerial vehicle multi-view stereopsis. *Remote Sensing*, *7*(9), 11933–11953. https://doi.org/10.3390/rs70911933

Haverkort, A. J., Boonekamp, P. M., Hutten, R., Jacobsen, E., Lotz, L. A. P., Kessel, G. J. T., … Van Der Vossen, E. A. G. (2008). Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification. *Potato Research*, *51*(1), 47–57. https://doi.org/10.1007/s11540-008-9089-y

Honkavaara, E., Rosnell, T., Oliveira, R., & Tommaselli, A. (2017). Band registration of tuneable frame format hyperspectral UAV imagers in complex scenes. *ISPRS Journal of Photogrammetry and Remote Sensing*, *134*, 96–109. https://doi.org/10.1016/j.isprsjprs.2017.10.014

Honkavaara, E., Saari, H., Kaivosoja, J., Pölönen, I., Hakala, T., Litkey, P., … Pesonen, L. (2013). Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture. *Remote Sensing*, *5*(10), 5006–5039. https://doi.org/10.3390/rs5105006

Hoofdproductschap Akkerbouw, sector A. (HPA). (2003). Ontheffingenbeleid bestrijding Phytophthora infestans. Retrieved from http://wetten.overheid.nl/BWBR0014851/2003-07-01

Hossin, M., & Sulaiman, M. N. (2015). a Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, *5*(2), 1–11. https://doi.org/10.5121/ijdkp.2015.5201

Huang, T., Yang, R., Huang, W., Huang, Y., & Qiao, X. (2017). Detecting sugarcane borer diseases using support vector machine. *Information Processing in Agriculture*, *5*(1), 74–82. https://doi.org/10.1016/j.inpa.2017.11.001

Humboldt State, U. (2017). GEOSPATIAL SCIENCE AT HUMBOLDT STATE. Retrieved February 22, 2018, from https://i1.wp.com/grindgis.com/wp-content/uploads/2017/05/clip_image001.png

Hy, J., Lf, T., & Zhu, H. (2011). Robust crop and weed segmentation under uncontrolled outdoor illumination. TT -. *Sensors (Basel, Switzerland) TA -*, *11*(6), 6270–6283.

Jiang, Y., Zebarth, B. J., Somers, G. H., Macleod, J. A., & Savard, M. M. (2012). *Sustainable Potato Production: Global Case Studies*. *Sustainable Potato Production: Global Case Studies*. https://doi.org/10.1007/978-94-007-4104-1

Kaundal, R., Kapoor, A. S., & Raghava, G. P. S. (2006). Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC Bioinformatics*, *7*, 485. https://doi.org/10.1186/1471-2105-7-485

Koch, P., Wujek, B., Golovidov, O., & Gardner, S. (2017). Automated Hyperparameter Tuning for Effective Machine Learning, (x), 1–23. Retrieved from http://support.sas.com/resources/papers/proceedings17/SAS0514-2017.pdf

Kong, W., Zhang, C., Huang, W., Liu, F., & He, Y. (2018). Application of Hyperspectral Imaging to Detect Sclerotinia sclerotiorum on Oilseed Rape Stems. *Sensors*, *18*(1), 123. https://doi.org/10.3390/s18010123

Lammerts van Bueren, E. ., Tiemens-Hulscher, M., & Struik, P. . (2008). Cisgenesis does not solve the late blight problem of organic potato production: Alternative breeding strategies. *Potato Research*, *51*(1), 89–99. https://doi.org/10.1007/s11540-008-9092-3

Lillesand, T., Kiefer, R., & Chipman, J. (2015). *Remote sensing and image interpretation* (7th ed.). John Wiley & Sons.

Maglogiannis, I. G. (2007). *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*. IOS Press. Retrieved from https://books.google.nl/books?id=vLiTXDHr_sYC

Mahlein, A.-K. (2016). Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease*, *100*(2), 241–251. https://doi.org/10.1094/PDIS-03-15-0340-FE

Mahlein, A. K., Oerke, E. C., Steiner, U., & Dehne, H. W. (2012). Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology*, *133*(1), 197–209. https://doi.org/10.1007/s10658-011-9878-z

Mahlein, A. K., Rumpf, T., Welke, P., Dehne, H. W., Plümer, L., Steiner, U., & Oerke, E. C. (2013). Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment*, *128*, 21–30. https://doi.org/10.1016/j.rse.2012.09.019

Mahlein, A. K., Steiner, U., Hillnhtter, C., Dehne, H., & Oerke, E. C. (2012). Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods*, *8*(1), 1–13. https://doi.org/10.1186/1746-4811-8-3

Maldonado, S., Weber, R., & Famili, F. (2014). Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines. *Information Sciences*, *286*, 228–246. https://doi.org/10.1016/j.ins.2014.07.015

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., … Dandekar, A. M. (2015). Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development*, *35*(1), 1–25. https://doi.org/10.1007/s13593-014-0246-1

May, R. J., Maier, H. R., & Dandy, G. C. (2010). Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Networks*, *23*(2), 283–294. https://doi.org/10.1016/j.neunet.2009.11.009

Mitchell, T. M. (2006). The Discipline of Machine Learning. *Machine Learning*, *17*, 1–7. https://doi.org/10.1080/026404199365326

Mokhtarian, P. (2017). Machine Learning Model Performance and Error Analysis. Retrieved October 11, 2017, from https://www.linkedin.com/pulse/machine-learning-model-performance-error-analysis-payam-mokhtarian

Morgenthaler, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 33–44. https://doi.org/10.1002/wics.2

Mosley, L. (2013). *A balanced approach to the multi-class imbalance problem*. Iowa State University. Retrieved from http://lib.dr.iastate.edu/etd/13537

Mucherino, A., Papajorgji, P. J., & Pardalos, P. M. (2009). *Data Mining in Agriculture*. *Media* (Vol. 34). https://doi.org/10.1007/978-0-387-88615-2

Muhammed, H. H. (2005). Hyperspectral crop reflectance data for characterising and estimating fungal disease severity in wheat. *Biosystems Engineering*, *91*(1), 9–20. https://doi.org/10.1016/j.biosystemseng.2005.02.007

Mukherjee, D. P., & Pal, S. (2005). *Support Vector Machines for Pattern Classification*. *Pattern Recognition Letters* (Vol. 26). https://doi.org/10.1007/1-84628-219-5

Muller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python*. (D. Schanafelt, Ed.) (First Edit). O'Reilly Media Inc. Retrieved from http://oreilly.com/catalog/errata.csp?isbn=9781449369415

Nowicki, M., Foolad, M. R., Nowakowska, M., & Kozik, E. U. (2012). Potato and Tomato Late Blight Caused by *Phytophthora infestans* : An Overview of Pathology and Resistance Breeding. *Plant Disease*, *96*(1), 4–17. https://doi.org/10.1094/PDIS-05-11-0458

Olanya, O. M., Larkin, R. P., & Honeycutt, C. W. (2015). Incidence of Phytophthora infestans (Mont.) de Bary on potato and tomato in Maine, 2006-2010. *Journal of Plant Protection Research*, *55*(1), 58–68. https://doi.org/10.1515/jppr-2015-0009

Ortigosa-Hernández, J., Inza, I., & Lozano, J. A. (2017). Measuring the class-imbalance extent of multi-class problems. *Pattern Recognition Letters*, *98*, 32–38. https://doi.org/10.1016/j.patrec.2017.08.002

Park, B., & Lu, R. (2015). *Hyperspectral in Food and Technology Imaging Agriculture*. *Hyperspectral Imaging Technology in Food and Agriculture*. https://doi.org/10.1007/978-1-4939-2836-1_13

Picard, R. R., & Berk, K. N. (2017). Data Splitting Author (s): Richard R . Picard and Kenneth N . Berk Published by : Taylor & Francis , Ltd . on behalf of the American Statistical Association Stable URL : http://www.jstor.org/stable/2684155 REFERENCES Linked references are available on J, *44*(2), 140–147.

Preet, P., Batra, S. S., & Jayadeva. (2015). Feature Selection for classification of hyperspectral data by minimizing a tight bound on the VC dimension. Retrieved from http://arxiv.org/abs/1509.08112

Rahman, H., Jabbar Ch, N., Manzoor, S., Najeed, F., Siddique, M., & Khan, R. (2017). A comparative analysis of machine learning approaches for plant disease identification. *International Quartely Journal Of Biological Sciences*, *4*(4), 120–126. Retrieved from www.als-journal.com/ ISSN 2310-5380/ August 2017

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. In L. LIU & M. T. ÖZSU (Eds.), *Encyclopedia of Database Systems* (pp. 532–538). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-39940-9_565

Reitermanová, Z. (2010). Data Splitting. *Week of Doctoral Students 2010 -- Proceedings of Contributed Papers*, 31–36. Retrieved from https://www.mff.cuni.cz/veda/konference/wds/proc/pdf10/WDS10_105_i1_Reitermanova.pdf

Rey, N., Volpi, M., Joost, S., & Tuia, D. (2017). Detecting animals in African Savanna with UAVs and the crowds. *Remote Sensing of Environment*, *200*(July), 341–351. https://doi.org/10.1016/j.rse.2017.08.026

Ristaino, J. B., & Johnston, S. A. (1999). Ecologically Based Approaches to Management of Phytophthora Blight on Bell Pepper. *Plant Disease*, *83*(12), 1080–1089. https://doi.org/10.1094/PDIS.1999.83.12.1080

Rivera-Caicedo, J. P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G., & Moreno, J. (2017). Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, *132*, 88–101. https://doi.org/10.1016/j.isprsjprs.2017.08.012

Roosjen, P. P. J., Suomalainen, J. M., Bartholomeus, H. M., Kooistra, L., & Clevers, J. G. P. W. (2017). Mapping reflectance anisotropy of a potato canopy using aerial images acquired with an unmanned aerial vehicle. *Remote Sensing*, *9*(5). https://doi.org/10.3390/rs9050417

Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., & Plümer, L. (2010). Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, *74*(1), 91–99. https://doi.org/10.1016/j.compag.2010.06.009

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, *72*(1), 1–13. https://doi.org/10.1016/j.compag.2010.02.007

Sathyanarayana, S., & Amarappa, S. (2014). Data classification using Support vector Machine (SVM), a simplified approaCH. *International Journal of Electronics and Computer S Cience Engineering, Volume 3, Number 4, ISSN- 2277-1956*, 435–445. Retrieved from http://www.ijecse.org/wp-content/uploads/2012/06/Volume-3Number-4PP-435-445x.pdf

Seelan, S. K., Laguette, S., Casady, G. M., & Seielstad, G. A. (2003). Remote sensing applications for precision agriculture: A learning community approach. *Remote Sensing of Environment*, *88*(1–2), 157–169. https://doi.org/10.1016/J.RSE.2003.04.007

Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, *21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015

Skelsey, P., Kessel, G. J. T., Rossing, W. A. H., & van der Werf, W. (2009). Parameterization and evaluation of a spatiotemporal model of the potato late blight pathosystem. *Phytopathology*, *99*(3), 290–300. https://doi.org/10.1094/PHYTO-99-3-0290

Stephan, D., Schmitt, A., Carvalho, S. M., Seddon, B., & Koch, E. (2005). Evaluation of biocontrol preparations and plant extracts for the control of Phytophthora infestans on potato leaves. *European Journal of Plant Pathology*, *112*(3), 235–246. https://doi.org/10.1007/s10658-005-2083-1

Sugiura, R., Tsuda, S., Tamiya, S., Itoh, A., Nishiwaki, K., Murakami, N., … Nuske, S. (2016). Field phenotyping system for the assessment of potato late blight resistance using RGB imagery from an unmanned aerial vehicle. *Biosystems Engineering*, *148*, 1–10. https://doi.org/10.1016/j.biosystemseng.2016.04.010

Suomalainen, J., Anders, N., Iqbal, S., Roerink, G., Franke, J., Wenting, P., … Kooistra, L. (2014). A lightweight hyperspectral mapping system and photogrammetric processing chain for unmanned aerial vehicles. *Remote Sensing*, *6*(11), 11013–11030. https://doi.org/10.3390/rs61111013

Surmann, K. (2015). *Exploring the use of Hyperspectral Imagery for Monitoring Complex Cropping Systems*. Wageningen University and Research Centre. Retrieved from http://edepot.wur.nl/396205

Tekos, P. (2010). *Investigation of potato late blight control strategies in southwestern Greece - Potential and suitability of a decision support system.* Wageningen Univertsity. Retrieved from http://edepot.wur.nl/137719

Thomas, S., Kuska, M. T., Bohnenkamp, D., Brugger, A., Alisaac, E., Wahabzada, M., … Mahlein, A. K. (2018). Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. *Journal of Plant Diseases and Protection*, *125*(1), 5–20. https://doi.org/10.1007/s41348-017-0124-6

Ustin, S. L., Roberts, D. A. R. A., Gamon, J. A., & Green, R. O. (2004). Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. *BioScience*, *54*(6), 523–534. https://doi.org/10.1641/0006-3568(2004)054[0523:UISTSE]2.0.CO;2

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 273–290. https://doi.org/10.1016/j.isprsjprs.2015.05.005

Weiss, G. M., & Battistin, A. (2014). Generating Well-Behaved Learning Curves : An Empirical Study. *Proceedings of the International Conference on Data Mining (DMIN)*, 1–4.

Wiik, L. (2014). Potato Late Blight and Tuber Yield: Results from 30 Years of Field Trials. *Potato Research*, *57*, 77–98.

Ye, J. (2015). *Using Machine Learning for Exploratory Data Analysis and Predicte Modelling*. University of Stavanger. Retrieved from http://hdl.handle.net/11250/299605

Yeh, Y.-H. F., Chung, W.-C., Liao, J.-Y., Chung, C.-L., Kuo, Y.-F., & Lin, T.-T. (2013). *A Comparison of Machine Learning Methods on Hyperspectral Plant Disease Assessments*. *IFAC Proceedings Volumes* (Vol. 46). IFAC. https://doi.org/http://dx.doi.org/10.3182/20130327-3-JP-3017.00081

Zhang, M., Qin, Z., Liu, X., & Ustin, S. L. (2003). Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *4*(4), 295–310. https://doi.org/10.1016/S0303-2434(03)00008-4

Zweep, M. (2014). *Investigation of the relation between foliage and tuber late blight resistance under laboratory and field conditions*. Wageningen University. Retrieved from http://edepot.wur.nl/330465

9 APPENDICES

APPENDIX A

STUDY AREA AND EXPERIMENTAL SET-UP (PERFORMED PREVIOUS THE STUDY)

Data acquisition was realized between the Spring and Summer of 2016 in an organic strip cropping experiment (51.9917°N, 5.66332°E; WGS84) started in 2014, at the Droevendaal experimental farm of the Wageningen University, The Netherlands. In this site, plots cultivated with potato were followed over time, mainly focusing on the assessment of late blight (*Phytophtora infestans*) development and general crop healthy status. Twelve plots measuring 3 by 10 m (small plots) were stablished in a strip along the field (*Figure 3*) while buffer areas, measuring 3 by 5 m, were placed before and after each plot, in the same strip, in order to avoid border effects along the experiment. The same experimental configuration, but with larger plots (with 6 by 10 m, and buffer areas with 6 by 5 m), was repeated in a nearby field.

Two different treatments were compared within the experiment: (a) plots in which a single cultivar susceptible to late blight was planted (non-mixed system); and (b) plots in which a mixture of three varieties with different degrees of resistance (from low to high, respectively) to late blight were iterated in each crop row (mixed system). Considering the treatments, the minimum comparable area between plots, besides individual plants, corresponded to that including three consecutive plants arranged in the same row. Based on that, each plot surface was divided in multiple sampling units measuring 0.75 by 1 m. These sampling units were then used during data acquisition at ground level and to extract spectral information from UAV-borne imagery. Late blight occurrence and severity was visually assessed periodically after the first symptoms of the disease were detected in the field (every 3 to 5 days), following the methodology described by the European and Mediterranean Plant Protection Organization (European and Mediterranean Plant Protection Organization [EPPO], 2008). For late blight assessment, four fixed sample units per plot were followed during the growing season (one per crop row), and other seven were randomly chosen before each assessment in order to better describe intra-plot variability over time (*Figure 3*).

The experiment followed a generalized randomized block design, with three blocks and two replicates of each treatment (i.e., cultivation methods) in each block. From the three blocks included in the experimental site, only the first two were followed in this study, since evaluating the overall performance of the cropping systems was not the main objective but rather exploit the variability present in the cultivated areas to evaluate the methods tested here.

UAV-BORNE HIGH RESOLUTION OPTICAL IMAGERY

Spectral data was acquired on six dates during the growing season in order to follow the dynamics of crop and disease development over time using mainly a lightweight hyperspectral frame camera (Rikola Ltd., Oulu, Finland). This Fabry-Perot interferometer (FPI) based camera (Honkavaara et al., 2013) was configured to register 16 narrow bands between 600 and 900 nm, chosen due to their

relevance to describe changes in leaf pigments (mainly chlorophylls) and leaf area (Clevers & Kooistra, 2012).

Due to intrinsic characteristics of the FPI system, images corresponding to different wavelengths were acquired sequentially, since changes in the wavelengths measured depend on internal camera adjustments. Consequently, a mismatch between images corresponding to different wavelengths in a given datacube occurred, an issue solved during photogrammetric processing in dedicated software (PhotoScan version 1.3, AgiSoft LLC, St. Petersburg, Russia). This procedure relied on the implementation of Structure from Motion (SfM) algorithm, with feature matching and self-calibrating bundle adjustment (Harwin, Lucieer, & Osborn, 2015)(Harwin et al., 2015). During image alignment and derivation of dense points cloud, imagery with full resolution was used (i.e., setting quality to 'high' and 'ultra-high' for these steps in the software processing chain, respectively). Optimization of retrieved camera positions and orientations was performed based on 4 to 8 ground control points (depending on the acquisition date, which had their coordinates registered using RTK-GPS. Before the optimization step, sparse point clouds were filtered based on residuals and reconstruction uncertainty (10% of points with the largest values were removed in each case), as performed by (Honkavaara, Rosnell, Oliveira, & Tommaselli, 2017). Dense point cloud depth filtering was set to 'mild' to preserved details in the final 3D reconstruction of the crop surface. Considering the approximate flight height of 80 m, a ground sampling distance between 0.04 and 0.05 m was achieved in the final orthorectified images.

Conversion of digital numbers (registered with 12-bit radiometric resolution) to radiance, in mW/ (m^2 str nm), was performed using proprietary software provided with the camera (HyperspectralImager version 2.0). Corrections (e.g., flat field and dark current compensation) were made based on factory calibration parameters and on images taken with the sensor lens completely covered (dark reference), before flight. Radiance values were converted into reflectance factor through the empirical line approach using images, also acquired before flight, of a Spectralon reference panel with 50% reflectance (LabSphere Inc., North Sutton, NH, USA), taken under the same general illumination conditions observed during data acquisition. In this case, after radiometric and geometric correction of the data acquired the final product used to extract spectral information for the sampling units correspond to the mosaic of images acquired over the study area in each date composed by the nearest spectra/pixel to the nadir field of view.

APPENDIX B

Key assessment of (European and Mediterranean Plant Protection Organization [EPPO], 2008)

Table 1 Relationship between percentage disease severity and disease assessments used in Europe

Phytophthora infestans – Late Blight symptoms	Severity (%)	ADAS*	PPS†	SAC‡	CIP§	Phytophthora infestans – Late Blight symptoms	Severity (%)	ADAS*	PPS†	SAC‡	CIP§
No infection visible	0.000	0	10	0	1	1–3 lesions per plants	1	–	9	–	–
1 lesion per plot	0.001	0.1	–	0.1	–	3–4 lesions per plant	–	0.7	–	0.8	–
2 lesions per plot	0.002	–	–	–	–	6 lesions per plant	–	0.8	–	–	–
3 or 4 lesions per plot	–	–	–	0.2	–	5–9 lesions per plant	–	–	–	0.9	–
5 lesions per plot	0.005	–	–	–	–	9 lesions per plant	–	0.9	–	–	–
5–8 lesions per plot	–	–	–	0.3	–	10 lesions per plant	2.5	1	8.5	1	2
6–10 lesions per plot	0.01	–	9.5	–	–	20 lesions per plant	5	–	8	–	–
9–16 lesions per plot	–	–	–	0.4	–	50 lesions per plant	–	5	–	5	–
17–31 lesions per plot	–	–	–	0.5	–	100 lesions per plant	–	10	–	–	–
25 lesions per plot	–	0.2	–	–	–	5% leaf area dead per plot	–	–	–	50	–
20–30 lesions per plot	0.1	–	–	–	–	10% leaf area dead per plot	10	–	7	–	3
50 lesions per plot	–	0.3	–	–	–	25% leaf area dead per plot	25	25	6	–	4
32–63 lesions per plot	–	–	–	0.6	–	50% leaf area dead per plot	50	50	4	–	5
75 lesions per plot	–	0.4	–	–	–	75% leaf area dead per plot	75	75	3	–	6
0.5 lesions per plant	0.5	–	–	–	–	90% leaf area dead per plot	90	–	2	–	7
1 lesion per plants	–	0.5	–	–	–	97.5% leaf area dead per plot	97.5	95	1	–	8
1–2 lesions per plants	–	–	–	0.7	–	All stems and leaves dead	100	100	0	–	9
2 lesions per plants	–	0.6	–	–	–						

*ADAS = Agricultural Development and Advisory Service.
†PPS = Plant Protection Service, Wageningen, NL.
‡SAC = Scottish Agricultural College.
§CIP = International Potato Center.

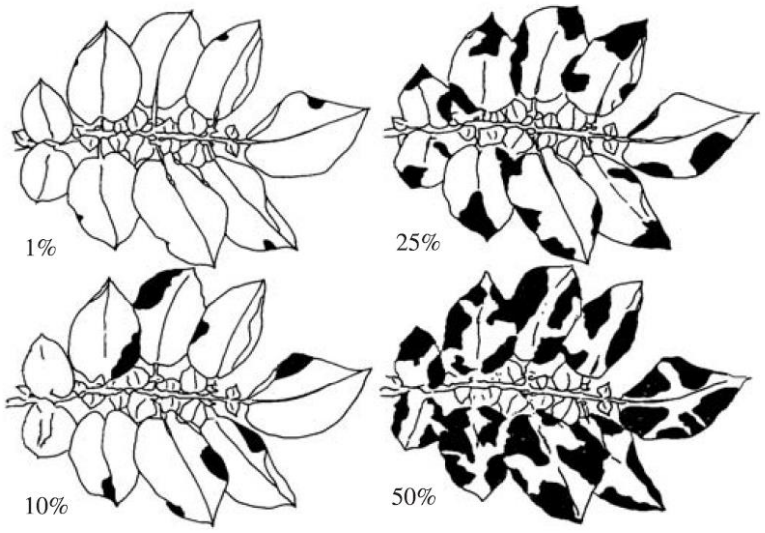


Fig. 1 Visual key for Phytophthora infestans on potato (% leaf area affected). Figure reproduced courtesy of Canada department of agriculture.

European and Mediterranean Plant Protection Organization
2008 OEPP/EPPO, Bulletin OEPP/EPPO Bulletin 38, 268–271

APPENDIX C GRID SEARCH MODEL PERFORMANCES (BALANCED ACCURACY METRIC)

	Mean_test	Mean_train	C	Class weight	Gamma	kernel	N° components
1	0.827	0.833	45	balanced	0.01	rbf	5
2	0.824	0.834	45	balanced	0.01	rbf	6
3	0.824	0.834	50	balanced	0.01	rbf	5
4	0.824	0.834	50	balanced	0.01	rbf	6
5	0.824	0.831	45	balanced	0.01	rbf	-
6	0.824	0.831	50	balanced	0.01	rbf	-
7	0.822	0.832	50	balanced	0.01	rbf	3
8	0.822	0.823	75	balanced	0.01	rbf	3
9	0.822	0.826	100	balanced	0.01	rbf	3
10	0.822	0.832	45	balanced	0.01	rbf	3
11	0.822	0.826	45	balanced	0.01	rbf	4
12	0.821	0.835	10	balanced	-	linear	-
13	0.821	0.832	1000	balanced	0.001	rbf	-
14	0.820	0.895	50	balanced	0.1	rbf	-
15	0.820	0.857	35	balanced	-	linear	-
16	0.820	0.860	45	balanced	-	linear	-
17	0.820	0.862	50	balanced	-	linear	-
18	0.818	0.827	1000	balanced	0.001	rbf	3
19	0.818	0.829	1	balanced	-	linear	4
20	0.818	0.821	1	balanced	-	linear	3
21	0.818	0.824	1000	balanced	0.001	rbf	6
22	0.816	0.828	100	balanced	-	linear	6
23	0.816	0.826	1000	balanced	0.001	rbf	4
24	0.816	0.850	25	balanced	-	linear	-
25	0.816	0.840	1000	balanced	0.01	rbf	3
26	0.815	0.835	75	balanced	0.01	rbf	5
27	0.815	0.836	15	balanced	-	linear	-
28	0.815	0.820	1	balanced	-	linear	5
29	0.815	0.824	1	balanced	-	linear	6
30	0.815	0.815	1	balanced	-	linear	-
31	0.814	0.828	35	balanced	0.01	rbf	5
32	0.814	0.824	10	balanced	0.01	rbf	6
33	0.813	0.836	1000	balanced	-	linear	6
34	0.813	0.824	50	balanced	0.01	rbf	4
35	0.813	0.828	75	balanced	-	linear	6
36	0.813	0.868	75	balanced	-	linear	-
37	0.812	0.840	100	balanced	0.01	rbf	6
38	0.812	0.816	45	balanced	-	linear	4
39	0.812	0.816	50	balanced	-	linear	4
40	0.811	0.824	100	balanced	0.001	rbf	5
41	0.811	0.874	75	balanced	0.1	rbf	3
42	0.811	0.887	100	balanced	0.1	rbf	3
43	0.810	0.824	10	balanced	0.01	rbf	5
44	0.810	0.826	10	balanced	0.01	rbf	-
45	0.810	0.834	25	balanced	-	linear	6
46	0.810	0.828	35	balanced	-	linear	6
47	0.809	0.824	75	balanced	0.01	rbf	4
48	0.809	0.834	75	balanced	0.01	rbf	6
49	0.809	0.829	15	balanced	0.01	rbf	3
50	0.809	0.885	50	balanced	0.1	rbf	6
51	0.809	0.892	45	balanced	0.1	rbf	-

52	0.808	0.834	35	balanced	0.01	rbf	6
53	0.808	0.833	35	balanced	0.01	rbf	-
54	0.808	0.823	25	balanced	0.01	rbf	5
55	0.808	0.826	15	balanced	0.01	rbf	6
56	0.808	0.824	100	balanced	0.001	rbf	6
57	0.808	0.826	100	balanced	0.001	rbf	-
58	0.807	0.834	100	balanced	0.01	rbf	5
59	0.807	0.832	100	balanced	0.01	rbf	4
60	0.807	0.828	45	balanced	-	linear	6
61	0.807	0.828	50	balanced	-	linear	6
62	0.807	0.871	100	balanced	-	linear	-
63	0.807	0.824	35	balanced	0.01	rbf	4
64	0.807	0.829	35	balanced	0.01	rbf	3
65	0.807	0.824	25	balanced	0.01	rbf	4
66	0.806	0.873	45	balanced	0.1	rbf	5
67	0.806	0.818	10	balanced	-	linear	3
68	0.806	0.818	15	balanced	-	linear	3
69	0.806	0.818	25	balanced	-	linear	3
70	0.806	0.818	35	balanced	-	linear	3
71	0.806	0.818	45	balanced	-	linear	3
72	0.806	0.818	50	balanced	-	linear	3
73	0.806	0.818	75	balanced	-	linear	3
74	0.806	0.818	100	balanced	-	linear	3
75	0.806	0.836	75	balanced	0.01	rbf	-
76	0.806	0.830	1000	balanced	0.001	rbf	5
77	0.806	0.816	75	balanced	-	linear	4
78	0.806	0.808	10	balanced	-	linear	4
79	0.806	0.810	15	balanced	-	linear	4
80	0.806	0.810	25	balanced	-	linear	4
81	0.806	0.812	35	balanced	-	linear	4
82	0.805	0.881	45	balanced	0.1	rbf	6
83	0.805	0.837	100	balanced	0.01	rbf	-
84	0.805	0.822	10	balanced	-	linear	6
85	0.804	0.817	75	balanced	0.001	rbf	5
86	0.803	0.822	15	balanced	0.01	rbf	4
87	0.803	0.823	25	balanced	0.01	rbf	3
88	0.803	0.896	75	balanced	0.1	rbf	6
89	0.803	0.847	10	balanced	0.1	rbf	3
90	0.802	0.869	45	balanced	0.1	rbf	3
91	0.802	0.875	50	balanced	0.1	rbf	3
92	0.802	0.826	10	balanced	0.01	rbf	3
93	0.802	0.825	25	balanced	0.01	rbf	6
94	0.802	0.827	25	balanced	0.01	rbf	-
95	0.801	0.831	15	balanced	-	linear	6
96	0.801	0.828	15	balanced	0.01	rbf	-
97	0.801	0.817	75	balanced	0.001	rbf	6
98	0.801	0.817	75	balanced	0.001	rbf	-
99	0.800	0.818	1000	balanced	-	linear	3
100	0.800	0.819	1	balanced	0.1	rbf	4
101	0.800	0.820	1	balanced	0.1	rbf	5
102	0.800	0.818	1	balanced	0.1	rbf	-
103	0.800	0.814	1	balanced	0.1	rbf	3

APPENDIX D – Disease Dataset

	Assess	Plot	Intra_plot	P_size	treat_x	DAP	b0	b1	b2	b3	b4	b6	b8	b9	b10	b11	b12	b13	b14	b15	severity_label	std	Maximum	Minimum
0	1	1	2	s	Non-mix	66	0.05444	0.0517	0.04878	0.0472	0.04	0.05249	0.17223	0.32613	0.44155	0.45644	0.46209	0.4681	0.47131	0.46983	1	0.20058	0.471313795	0.04000207
1	1	1	7	s	Non-mix	66	0.05497	0.0523	0.04902	0.04782	0.04043	0.05251	0.17149	0.3257	0.44446	0.46371	0.46796	0.47192	0.4731	0.47045	0	0.20216	0.473095191	0.04042998
2	1	1	9	s	Non-mix	66	0.05344	0.05053	0.04702	0.04592	0.03834	0.05123	0.1655	0.31262	0.42548	0.44	0.44641	0.45109	0.45592	0.45166	0	0.19322	0.455922591	0.03834141
3	1	1	14	s	Non-mix	66	0.05265	0.04978	0.04645	0.04558	0.03732	0.05009	0.16617	0.30874	0.4152	0.43263	0.43532	0.43905	0.44143	0.43958	1	0.18808	0.441431257	0.03731841
4	1	1	16	s	Non-mix	66	0.05432	0.0514	0.04757	0.04675	0.03867	0.05135	0.17186	0.32547	0.44182	0.45734	0.46284	0.4663	0.46861	0.46459	1	0.20028	0.468610829	0.03867217
5	1	1	21	s	Non-mix	66	0.05805	0.0541	0.05052	0.04913	0.03898	0.05245	0.17362	0.31816	0.42232	0.43658	0.44238	0.4423	0.44221	0.44099	1	0.18829	0.442381023	0.03898331
6	1	1	27	s	Non-mix	66	0.05474	0.05143	0.04793	0.04653	0.03779	0.05055	0.17204	0.32765	0.44566	0.45814	0.46277	0.46763	0.46677	0.46524	1	0.20078	0.467629298	0.0377911
7	1	1	33	s	Non-mix	66	0.0594	0.05609	0.05292	0.05102	0.04266	0.05567	0.17705	0.32367	0.43017	0.44266	0.44594	0.44541	0.45061	0.44573	1	0.18985	0.450612783	0.04266292
8	1	1	36	s	Non-mix	66	0.05871	0.0552	0.05203	0.05138	0.04072	0.05402	0.17675	0.32497	0.43731	0.44888	0.44846	0.45384	0.45696	0.45626	0	0.19361	0.456961065	0.04071631
9	1	1	39	s	Non-mix	66	0.0537	0.05015	0.04699	0.0466	0.0376	0.04992	0.16505	0.30605	0.41626	0.42726	0.43308	0.43567	0.43776	0.43518	1	0.18633	0.437757266	0.03760328
10	1	1	43	l	Non-mix	66	0.05313	0.05043	0.04657	0.04546	0.03618	0.04919	0.16385	0.29882	0.39541	0.40973	0.41238	0.41825	0.42013	0.41529	1	0.17747	0.420133479	0.0361789
11	1	2	3	s	Mix	66	0.05339	0.05026	0.0462	0.04429	0.03796	0.04993	0.16334	0.33674	0.4957	0.51503	0.52317	0.52416	0.52966	0.52443	1	0.22959	0.529660995	0.03795756
12	1	2	5	s	Mix	66	0.04727	0.0439	0.04065	0.03958	0.03336	0.04379	0.14568	0.29476	0.42739	0.44668	0.45157	0.45486	0.46088	0.45652	0	0.19864	0.460879304	0.03336447
13	1	2	9	s	Mix	66	0.04653	0.0438	0.04031	0.03926	0.03338	0.04482	0.1521	0.31068	0.44984	0.47303	0.47487	0.47724	0.47946	0.47775	0	0.20955	0.479457388	0.03338176
14	1	2	16	s	Mix	66	0.05888	0.05526	0.05052	0.04913	0.04107	0.05408	0.1887	0.38349	0.54832	0.57303	0.58027	0.58036	0.58425	0.58292	0	0.25466	0.584249525	0.04107396
15	1	2	18	s	Mix	66	0.05336	0.04957	0.04573	0.04383	0.03745	0.05005	0.17337	0.3564	0.51056	0.53435	0.53974	0.54083	0.547	0.54226	0	0.23805	0.547000562	0.0374468
16	1	2	23	s	Mix	66	0.0471	0.0442	0.04082	0.03982	0.03387	0.04408	0.14751	0.30515	0.44651	0.4653	0.46973	0.4745	0.47808	0.47493	1	0.20762	0.478084404	0.03386662
17	1	2	25	s	Mix	66	0.0474	0.04442	0.04085	0.03942	0.03329	0.0441	0.15051	0.30526	0.44244	0.46106	0.46663	0.46935	0.47318	0.46973	0	0.20539	0.473176852	0.03329354
18	1	2	29	s	Mix	66	0.04623	0.04322	0.0401	0.03927	0.03333	0.04475	0.15231	0.3139	0.45454	0.47649	0.48475	0.48731	0.48999	0.48645	0	0.21356	0.489993388	0.03333212
19	1	2	32	s	Mix	66	0.05085	0.04869	0.04516	0.04424	0.03656	0.04716	0.14918	0.29272	0.41823	0.43517	0.44053	0.44922	0.45316	0.45021	0	0.1925	0.453160977	0.03655685
20	1	2	37	s	Mix	66	0.05214	0.04884	0.04521	0.0442	0.03587	0.04783	0.16306	0.32301	0.45776	0.47716	0.48471	0.4896	0.49237	0.48791	0	0.21204	0.492366105	0.03587287
21	1	2	58	l	Mix	66	0.04877	0.04539	0.04248	0.04157	0.03446	0.04593	0.1421	0.26838	0.36403	0.37962	0.384	0.38818	0.39097	0.38934	0	0.16517	0.390966799	0.03446351
22	1	3	3	s	Non-mix	66	0.05069	0.04758	0.04454	0.04347	0.03724	0.05035	0.16798	0.33093	0.46362	0.48188	0.49153	0.49545	0.49714	0.49468	1	0.21486	0.497142057	0.03724095
23	1	3	6	s	Non-mix	66	0.04878	0.04557	0.04212	0.04127	0.03523	0.04836	0.15874	0.31391	0.43992	0.45858	0.46599	0.46987	0.47264	0.46894	0	0.20387	0.472637599	0.03522543
24	1	3	8	s	Non-mix	66	0.05068	0.04854	0.04537	0.04415	0.03824	0.04857	0.15217	0.30818	0.43779	0.4527	0.45807	0.45966	0.46574	0.4652	1	0.19987	0.465735697	0.03824009
25	1	3	14	s	Non-mix	66	0.05009	0.0472	0.04381	0.04292	0.03754	0.0507	0.17114	0.34276	0.48209	0.50311	0.50772	0.51372	0.51745	0.51041	1	0.22396	0.517445226	0.03754135
26	1	3	17	s	Non-mix	66	0.05111	0.04744	0.04417	0.04319	0.03733	0.05093	0.17523	0.35199	0.49712	0.51934	0.52521	0.5297	0.53223	0.52655	1	0.23155	0.532234351	0.03732825
27	1	3	20	s	Non-mix	66	0.0604	0.05726	0.05412	0.05284	0.04439	0.05649	0.18063	0.36536	0.51805	0.54291	0.5458	0.5501	0.5559	0.55746	0	0.23892	0.557455068	0.04438794
28	1	3	22	s	Non-mix	66	0.04638	0.04374	0.04086	0.04034	0.03533	0.04724	0.1567	0.3123	0.43445	0.45333	0.45944	0.46549	0.46574	0.46117	0	0.20157	0.465737071	0.03532986
29	1	3	27	l	Non-mix	66	0.05373	0.05061	0.0467	0.04607	0.03705	0.04685	0.15418	0.29901	0.41696	0.43512	0.44647	0.45326	0.45726	0.4545	0	0.19325	0.45726418	0.03704696
30	1	3	31	s	Non-mix	66	0.05061	0.04745	0.04393	0.04289	0.03511	0.0492	0.16342	0.31429	0.43011	0.44871	0.45713	0.46402	0.46564	0.45997	0	0.19912	0.465641687	0.03511023
31	1	3	35	s	Non-mix	66	0.04679	0.04431	0.04094	0.04024	0.03346	0.04586	0.15052	0.28877	0.39699	0.41335	0.42055	0.42469	0.42774	0.42331	0	0.18277	0.427743404	0.03345813
32	1	3	40	s	Non-mix	66	0.05392	0.05121	0.04659	0.04534	0.03675	0.04947	0.16302	0.31318	0.43208	0.45048	0.45527	0.45855	0.46171	0.45786	0	0.19718	0.461713807	0.0367462
33	1	4	3	s	Mix	66	0.04465	0.04303	0.04043	0.03941	0.0342	0.04473	0.14207	0.28671	0.40527	0.42388	0.43018	0.43078	0.42914	0.43146	0	0.1869	0.431459055	0.03420179
34	1	4	6	s	Mix	66	0.04787	0.04521	0.04271	0.04154	0.03583	0.0473	0.15197	0.30807	0.43395	0.45747	0.46198	0.46685	0.46275	0.46383	1	0.20171	0.466846543	0.03582914
35	1	4	9	s	Mix	66	0.04674	0.04442	0.04205	0.04111	0.03518	0.0466	0.14774	0.29448	0.41425	0.43126	0.43665	0.44182	0.43863	0.43596	0	0.18996	0.441818954	0.03517908
36	1	4	16	l	Mix	66	0.05016	0.0475	0.04394	0.04377	0.0357	0.04699	0.15211	0.29024	0.39647	0.40477	0.41044	0.41331	0.41341	0.40534	1	0.17637	0.413412544	0.03569536
37	1	4	17	s	Mix	66	0.05352	0.05108	0.04793	0.04656	0.04045	0.05394	0.17711	0.36689	0.52152	0.54555	0.55086	0.55529	0.55888	0.554	0	0.24283	0.558881741	0.04044569
38	1	4	23	s	Mix	66	0.04329	0.04162	0.03883	0.03874	0.03528	0.04719	0.15172	0.31101	0.43831	0.45833	0.46698	0.47085	0.47663	0.46867	1	0.20567	0.476633003	0.03528098
39	1	4	26	s	Mix	66	0.04843	0.04593	0.04236	0.04257	0.03636	0.05	0.16064	0.31902	0.44186	0.46175	0.4696	0.47391	0.47526	0.47097	0	0.20499	0.475262564	0.03636015
40	1	4	30	s	Mix	66	0.04342	0.0412	0.03789	0.03828	0.03389	0.04636	0.14696	0.29014	0.40286	0.42135	0.42638	0.43383	0.43567	0.43153	0	0.18736	0.435671569	0.03389036
41	1	4	32	s	Mix	66	0.04929	0.04712	0.04279	0.04161	0.03591	0.04807	0.1576	0.30769	0.42739	0.44671	0.45681	0.46283	0.46743	0.46164	1	0.19931	0.46742847	0.03591457
42	1	4	34	s	Mix	66	0.04765	0.04546	0.04087	0.03997	0.03292	0.04599	0.1575	0.3098	0.42746	0.44579	0.45369	0.46171	0.46302	0.45927	0	0.19932	0.463024549	0.03291948
43	1	4	38	s	Mix	66	0.04889	0.04582	0.042	0.04096	0.0335	0.04618	0.16219	0.31789	0.43966	0.4588	0.46824	0.47377	0.47969	0.47233	0	0.20554	0.479693469	0.03350126
44	1	5	2	s	Mix	66	0.04628	0.04397	0.04171	0.03994	0.03412	0.04475	0.13742	0.26726	0.37708	0.39469	0.39486	0.40946	0.406	0.40052	0	0.17285	0.409463395	0.0341175
45	1	5	5	s	Mix	66	0.04513	0.04227	0.03893	0.0382	0.03447	0.04372	0.16159	0.32463	0.47166	0.49566	0.50218	0.49753	0.50034	0.48945	0	0.21993	0.502182935	0.03446744
46	1	5	9	s	Mix	66	0.04857	0.04626	0.04212	0.03994	0.03509	0.04521	0.15329	0.31393	0.45999	0.48256	0.48933	0.49235	0.49833	0.49905	0	0.21614	0.499048059	0.03509401

47	1	5	14	s	Mix	66	0.06582	0.06465	0.06227	0.06201	0.04768	0.05627	0.18382	0.37457	0.54095	0.56076	0.56939	0.56799	0.57154	0.57409	1	0.24559	0.574086737	0.04767522
48	1	5	18	l	Mix	66	0.05211	0.04891	0.04446	0.04358	0.03586	0.04759	0.15582	0.30451	0.43185	0.44847	0.45112	0.45693	0.45826	0.4548	0	0.19687	0.458258499	0.03586226
49	1	5	18	s	Mix	66	0.05248	0.04951	0.04511	0.043	0.03736	0.04843	0.16202	0.33371	0.48867	0.50653	0.51461	0.51778	0.52069	0.52101	0	0.22661	0.52101222	0.03736237
50	1	5	21	s	Mix	66	0.04938	0.04644	0.04228	0.04196	0.03587	0.0487	0.16106	0.32268	0.45914	0.47702	0.47733	0.48525	0.48788	0.47973	0	0.21093	0.487880728	0.0358668
51	1	5	25	s	Mix	66	0.05444	0.05012	0.04567	0.04324	0.03874	0.04972	0.17193	0.35369	0.51866	0.53634	0.54264	0.54428	0.54499	0.54647	1	0.23934	0.546472948	0.03874353
52	1	5	29	s	Mix	66	0.04732	0.04524	0.04128	0.03915	0.0339	0.04435	0.15208	0.31695	0.45924	0.47771	0.48436	0.49114	0.49596	0.49317	0	0.21508	0.495964905	0.03390102
53	1	5	32	s	Mix	66	0.04783	0.04546	0.04107	0.04003	0.03386	0.04619	0.15972	0.31746	0.45788	0.47408	0.4745	0.48725	0.48988	0.48258	0	0.21167	0.489878152	0.03385923
54	1	5	40	s	Mix	66	0.04819	0.0458	0.04192	0.03967	0.03337	0.04338	0.15055	0.29787	0.42421	0.4416	0.45045	0.45904	0.46393	0.45896	0	0.19834	0.463925548	0.03336513
55	1	6	1	s	Mix	66	0.04487	0.04303	0.03909	0.03738	0.03285	0.04397	0.15062	0.29784	0.42373	0.44472	0.44947	0.44868	0.45534	0.44406	0	0.19652	0.455342262	0.032853
56	1	6	5	s	Mix	66	0.04526	0.04309	0.03972	0.03798	0.03253	0.04367	0.1473	0.28317	0.39296	0.41473	0.41902	0.42455	0.42651	0.41772	0	0.1826	0.426505821	0.03253433
57	1	6	12	s	Mix	66	0.05002	0.04697	0.04314	0.04101	0.0361	0.04896	0.16712	0.33715	0.47178	0.49197	0.49155	0.4928	0.49263	0.48911	0	0.216	0.492799909	0.03610444
58	1	6	16	l	Mix	66	0.04788	0.04509	0.04107	0.03955	0.03397	0.04564	0.14845	0.28774	0.39494	0.41221	0.41486	0.41852	0.42174	0.41976	0	0.18069	0.421740805	0.03396748
59	1	6	18	s	Mix	66	0.05274	0.04957	0.04536	0.04329	0.03761	0.05124	0.18649	0.37703	0.52847	0.55441	0.55595	0.55867	0.56255	0.54759	0	0.2455	0.562549565	0.03761227
60	1	6	20	s	Mix	66	0.05112	0.04802	0.04369	0.04204	0.03692	0.04975	0.17132	0.33836	0.46709	0.49115	0.49699	0.49771	0.49667	0.48759	0	0.21612	0.497711474	0.03692062
61	1	6	21	s	Mix	66	0.0492	0.04648	0.04214	0.04022	0.03418	0.04643	0.15947	0.31452	0.43792	0.45762	0.46572	0.47349	0.47677	0.47217	0	0.20475	0.476770778	0.03417519
62	1	6	25	s	Mix	66	0.05011	0.04751	0.04341	0.04129	0.03603	0.04815	0.16986	0.34791	0.48329	0.50509	0.51057	0.5193	0.52365	0.51397	1	0.22633	0.523652247	0.03602756
63	1	6	27	s	Mix	66	0.05175	0.04814	0.04383	0.04223	0.03629	0.04984	0.17164	0.35048	0.48614	0.50338	0.50926	0.51498	0.51552	0.51537	0	0.22495	0.51551954	0.03628951
64	1	6	36	s	Mix	66	0.04921	0.04675	0.04264	0.0407	0.03492	0.04752	0.16621	0.33208	0.46618	0.48809	0.49568	0.50284	0.50687	0.50297	0	0.21902	0.506869279	0.0349238
65	1	6	38	s	Mix	66	0.04932	0.04727	0.04296	0.04106	0.03453	0.04755	0.16903	0.33901	0.46818	0.48934	0.49565	0.50861	0.50895	0.50403	0	0.22	0.508948094	0.03452522
66	1	7	2	s	Non-mix	66	0.05172	0.04932	0.04508	0.04471	0.0388	0.05056	0.16743	0.32162	0.44029	0.45797	0.46474	0.46492	0.469	0.47034	0	0.20158	0.47033799	0.03880059
67	1	7	6	s	Non-mix	66	0.04976	0.04725	0.04383	0.04343	0.03799	0.05009	0.16085	0.31115	0.42425	0.44104	0.44828	0.45163	0.45263	0.45576	0	0.19468	0.455755744	0.0379948
68	1	7	9	s	Non-mix	66	0.04538	0.04299	0.03989	0.03913	0.03592	0.04733	0.15353	0.30453	0.43043	0.44784	0.45583	0.45874	0.45949	0.45836	0	0.19942	0.459491613	0.03591624
69	1	7	11	s	Non-mix	66	0.05224	0.04967	0.04615	0.04584	0.03989	0.05187	0.1663	0.32673	0.44741	0.46689	0.46909	0.4694	0.47315	0.47039	1	0.2036	0.473149775	0.03989309
70	1	7	14	s	Non-mix	66	0.05673	0.05443	0.0501	0.04958	0.04243	0.05639	0.18531	0.35672	0.48342	0.50364	0.51108	0.50962	0.5106	0.51093	0	0.22047	0.511084464	0.04243133
71	1	7	18	s	Non-mix	66	0.04958	0.04694	0.0436	0.04278	0.03823	0.05088	0.16715	0.3352	0.47093	0.49315	0.49643	0.49867	0.50058	0.49959	0	0.2179	0.500578144	0.03823026
72	1	7	22	s	Non-mix	66	0.04716	0.04463	0.04145	0.04123	0.03665	0.04752	0.15016	0.29932	0.4146	0.42851	0.43451	0.43429	0.43338	0.43464	0	0.18827	0.434643943	0.03664825
73	1	7	26	s	Non-mix	66	0.05314	0.05002	0.0461	0.04574	0.0403	0.05253	0.17323	0.34471	0.48186	0.499	0.50581	0.50561	0.50785	0.50524	0	0.22034	0.507851824	0.04030347
74	1	7	35	s	Non-mix	66	0.05344	0.05042	0.04646	0.04618	0.03943	0.05131	0.16376	0.31964	0.4402	0.45618	0.4633	0.46824	0.47469	0.47324	1	0.20182	0.474690486	0.03942704
75	1	7	40	s	Non-mix	66	0.05599	0.05299	0.04875	0.04832	0.04009	0.05365	0.16939	0.3222	0.43832	0.45383	0.46443	0.47151	0.47268	0.4749	0	0.20074	0.4749001	0.04009363
76	1	7	42	l	Non-mix	66	0.0554	0.05186	0.04855	0.04736	0.03891	0.05239	0.17521	0.34117	0.46565	0.48726	0.50528	0.51089	0.51352	0.51019	0	0.21889	0.51352369	0.03890885
77	1	8	7	s	Non-mix	66	0.06053	0.0569	0.05425	0.05426	0.04367	0.05207	0.17312	0.3337	0.46757	0.4821	0.49109	0.49004	0.49739	0.49681	0	0.21103	0.497389747	0.04367064
78	1	8	10	s	Non-mix	66	0.05923	0.05594	0.05152	0.0515	0.04339	0.05351	0.17444	0.33814	0.47326	0.4907	0.49923	0.49945	0.50205	0.50282	0	0.21501	0.502819833	0.04338968
79	1	8	11	s	Non-mix	66	0.05583	0.05335	0.04935	0.04834	0.04331	0.056	0.18079	0.35116	0.48832	0.50699	0.50498	0.51805	0.51692	0.50832	0	0.22187	0.518054184	0.04331191
80	1	8	13	l	Non-mix	66	0.04712	0.04436	0.04234	0.03988	0.03297	0.04361	0.1505	0.28605	0.39726	0.40558	0.40607	0.4127	0.41399	0.40767	1	0.17766	0.413985112	0.03297472
81	1	8	13	s	Non-mix	66	0.05338	0.05107	0.04649	0.04531	0.04013	0.05254	0.1736	0.33419	0.46462	0.47957	0.48033	0.48646	0.48903	0.48762	0	0.21057	0.489030537	0.04012802
82	1	8	18	s	Non-mix	66	0.06483	0.06142	0.05886	0.05958	0.04623	0.05639	0.18154	0.34452	0.47325	0.48624	0.49429	0.49045	0.49352	0.49731	0	0.20978	0.497309331	0.04622571
83	1	8	22	s	Non-mix	66	0.05079	0.04802	0.04428	0.04316	0.03763	0.04934	0.16155	0.31277	0.43124	0.44784	0.44762	0.4528	0.4543	0.45217	0	0.19556	0.45430207	0.03763454
84	1	8	25	s	Non-mix	66	0.05413	0.04439	0.03829	0.03759	0.03338	0.04301	0.15595	0.29777	0.41913	0.43237	0.4407	0.44093	0.44518	0.43855	1	0.19166	0.445177672	0.03337699
85	1	8	30	s	Non-mix	66	0.06079	0.05629	0.05228	0.05319	0.04292	0.05349	0.1736	0.33827	0.47262	0.48814	0.49681	0.49504	0.49836	0.50084	0	0.21347	0.5008379	0.04291659
86	1	8	34	s	Non-mix	66	0.04884	0.04611	0.04213	0.04212	0.03571	0.04696	0.14651	0.27374	0.38149	0.38924	0.39084	0.39692	0.40224	0.39525	0	0.16962	0.402235802	0.03571266
87	1	8	39	s	Non-mix	66	0.06435	0.06065	0.05755	0.05607	0.04256	0.05213	0.17296	0.32183	0.43888	0.45271	0.46081	0.46125	0.46213	0.46545	0	0.19516	0.465451208	0.04256291
88	3	1	2	s	Non-mix	74	0.08858	0.08484	0.07747	0.07557	0.06268	0.08584	0.3461	0.49123	0.61457	0.6315	0.64213	0.6388	0.65068	0.64974	2	0.27099	0.650675451	0.06267698
89	3	1	6	s	Non-mix	74	0.0796	0.07501	0.06601	0.06151	0.05482	0.07214	0.31797	0.45436	0.58597	0.60307	0.59978	0.59471	0.59568	0.613	1	0.25709	0.61300186	0.05481681
90	3	1	8	s	Non-mix	74	0.08496	0.07941	0.06958	0.06386	0.05642	0.07274	0.3244	0.45909	0.58157	0.59097	0.59053	0.58253	0.58679	0.60742	1	0.25175	0.607416556	0.05642462
91	3	1	14	s	Non-mix	74	0.08433	0.08054	0.07428	0.07179	0.05984	0.07853	0.33205	0.46974	0.58758	0.59988	0.60834	0.60507	0.60858	0.60971	1	0.25632	0.609712462	0.05983993
92	3	1	19	s	Non-mix	74	0.08715	0.08136	0.07394	0.06772	0.05709	0.07436	0.32892	0.46281	0.58454	0.59878	0.59367	0.5894	0.59086	0.59795	1	0.25181	0.598778655	0.05708849
93	3	1	21	s	Non-mix	74	0.09603	0.08973	0.08286	0.08066	0.06641	0.08772	0.36475	0.51326	0.62946	0.63831	0.65627	0.65003	0.65707	0.65571	1	0.27391	0.657067115	0.06640768
94	3	1	23	s	Non-mix	74	0.09352	0.08857	0.08049	0.07907	0.06516	0.08661	0.36134	0.50754	0.62826	0.645	0.6537	0.65322	0.654	0.6558	1	0.27475	0.655804232	0.06516392

95	3	1	27	s	Non-mix	74	0.09025	0.08508	0.07514	0.06898	0.06035	0.07847	0.34851	0.49124	0.62016	0.63853	0.62817	0.62511	0.62826	0.64492	1	0.26905	0.644915477	0.06034937
96	3	1	36	s	Non-mix	74	0.0889	0.08418	0.07228	0.06398	0.05691	0.07362	0.3219	0.45612	0.56971	0.5739	0.56817	0.56229	0.58355	0.60702	1	0.24501	0.607021933	0.05690978
97	3	1	39	s	Non-mix	74	0.08657	0.08024	0.07074	0.0647	0.05534	0.07392	0.31741	0.44805	0.56018	0.57448	0.56279	0.56306	0.56886	0.59195	1	0.24199	0.59194589	0.05533518
98	3	2	3	s	Mix	74	0.08061	0.075	0.0684	0.06486	0.05829	0.07701	0.34595	0.49322	0.62777	0.64383	0.64178	0.64031	0.63617	0.64068	1	0.27534	0.64382532	0.05829472
99	3	2	5	s	Mix	74	0.07716	0.07198	0.06532	0.06299	0.05555	0.07264	0.32757	0.46738	0.58945	0.60378	0.60166	0.60871	0.60079	0.60526	1	0.25907	0.608713514	0.05554822
100	3	2	14	s	Mix	74	0.07671	0.07225	0.06636	0.06335	0.05611	0.07474	0.33518	0.48119	0.60717	0.62259	0.62754	0.61847	0.61421	0.62291	1	0.26714	0.627542407	0.05610683
101	3	2	16	s	Mix	74	0.07986	0.07492	0.06908	0.06546	0.05852	0.07661	0.35376	0.5067	0.64283	0.6606	0.66218	0.65699	0.65089	0.65891	1	0.28359	0.66217886	0.05851727
102	3	2	20	s	Mix	74	0.08073	0.07606	0.06963	0.06696	0.05887	0.07873	0.35894	0.52106	0.66656	0.68452	0.68521	0.68335	0.67549	0.67564	1	0.29417	0.685206826	0.05887307
103	3	2	21	s	Mix	74	0.0769	0.07377	0.06699	0.06322	0.05685	0.07484	0.34008	0.49116	0.62196	0.63942	0.63929	0.63278	0.6324	0.63606	1	0.27405	0.639423035	0.05684806
104	3	2	23	s	Mix	74	0.07988	0.0761	0.06985	0.06617	0.05847	0.07735	0.35472	0.50872	0.64839	0.66663	0.66608	0.65888	0.65541	0.66228	1	0.28532	0.666633221	0.05847027
105	3	2	29	s	Mix	74	0.07636	0.07228	0.0671	0.06372	0.05699	0.07533	0.34149	0.48947	0.62821	0.64553	0.64859	0.64422	0.63774	0.63783	1	0.27714	0.64859135	0.05699286
106	3	2	37	s	Mix	74	0.0807	0.0754	0.0697	0.06528	0.05622	0.0753	0.34037	0.4812	0.60504	0.61898	0.61497	0.61471	0.613	0.62342	1	0.26431	0.623420188	0.0562239
107	3	2	40	s	Mix	74	0.07905	0.07526	0.06869	0.06424	0.0558	0.07389	0.34102	0.48663	0.61544	0.63447	0.63305	0.6332	0.62755	0.63201	1	0.27153	0.63447464	0.05580388
108	3	3	1	s	Non-mix	74	0.07181	0.06821	0.06301	0.06124	0.05068	0.06732	0.29187	0.40987	0.52247	0.53654	0.5335	0.5544	0.55796	0.55368	2	0.2323	0.557956512	0.05068212
109	3	3	3	s	Non-mix	74	0.06886	0.0656	0.06219	0.05928	0.04962	0.06646	0.28601	0.40295	0.51474	0.53199	0.52737	0.54886	0.55346	0.54886	1	0.2304	0.553463852	0.0496187
110	3	3	6	s	Non-mix	74	0.07154	0.06905	0.06348	0.06155	0.05101	0.06967	0.29555	0.41626	0.52065	0.54104	0.53738	0.56354	0.56736	0.57199	2	0.23568	0.571987406	0.05100737
111	3	3	17	s	Non-mix	74	0.06761	0.06391	0.05966	0.05831	0.0499	0.06685	0.30903	0.44217	0.55677	0.57104	0.57566	0.57791	0.57307	0.57171	1	0.2474	0.577911554	0.04989991
112	3	3	20	s	Non-mix	74	0.06587	0.06326	0.05875	0.05735	0.04876	0.06557	0.29397	0.41856	0.52768	0.54389	0.5461	0.54536	0.54436	0.54236	1	0.23363	0.546099058	0.04876033
113	3	3	22	s	Non-mix	74	0.06397	0.06165	0.05689	0.05554	0.04808	0.063	0.2802	0.39967	0.51146	0.52154	0.52828	0.53173	0.52627	0.51859	1	0.22529	0.531733908	0.04807812
114	3	3	28	s	Non-mix	74	0.07072	0.06725	0.06218	0.06099	0.05047	0.06808	0.32187	0.46179	0.579	0.59705	0.59827	0.61605	0.60907	0.61296	1	0.26134	0.616053142	0.05046991
115	3	3	33	s	Non-mix	74	0.07211	0.06919	0.06299	0.06157	0.05009	0.06735	0.29817	0.41981	0.51946	0.53573	0.53689	0.55773	0.55445	0.55449	2	0.23256	0.557726025	0.05009311
116	3	3	35	s	Non-mix	74	0.06706	0.06406	0.05881	0.05742	0.04772	0.06427	0.27657	0.39027	0.48828	0.50543	0.50441	0.52379	0.52019	0.52112	1	0.21854	0.523790555	0.0477173
117	3	3	40	s	Non-mix	74	0.07017	0.06738	0.0624	0.06059	0.04978	0.06776	0.28753	0.40038	0.49461	0.51459	0.51013	0.53257	0.53853	0.53877	1	0.22249	0.538768993	0.04977728
118	3	4	1	s	Mix	74	0.06617	0.06382	0.05974	0.0579	0.05044	0.06772	0.28254	0.40352	0.5094	0.52797	0.53632	0.53684	0.53802	0.54344	1	0.2283	0.543438959	0.05044344
119	3	4	3	s	Mix	74	0.06542	0.06202	0.05687	0.0549	0.0473	0.06196	0.28239	0.40587	0.51833	0.54175	0.53809	0.54222	0.54801	0.55397	1	0.23368	0.5539653	0.04729696
120	3	4	6	s	Mix	74	0.06722	0.06461	0.05818	0.05697	0.04858	0.06255	0.29177	0.41883	0.53626	0.55991	0.55337	0.5604	0.56389	0.56802	1	0.24093	0.568017984	0.04858109
121	3	4	15	s	Mix	74	0.06308	0.06136	0.05522	0.05375	0.04622	0.05862	0.27538	0.39796	0.50591	0.53036	0.52419	0.52795	0.53508	0.53666	1	0.22794	0.536655105	0.04621658
122	3	4	17	s	Mix	74	0.06749	0.06512	0.05916	0.05678	0.05054	0.06456	0.31032	0.44806	0.57918	0.60669	0.59775	0.60288	0.60435	0.60882	1	0.26121	0.608821139	0.05053821
123	3	4	19	s	Mix	74	0.06717	0.06477	0.0582	0.0564	0.04876	0.06381	0.3025	0.43242	0.55164	0.57335	0.56945	0.57104	0.57667	0.5779	1	0.24724	0.577899392	0.04875801
124	3	4	21	s	Mix	74	0.07006	0.06699	0.06197	0.06007	0.05121	0.06887	0.32352	0.46866	0.59526	0.61657	0.61681	0.6198	0.62213	0.62051	1	0.2676	0.622130153	0.05121472
125	3	4	28	s	Mix	74	0.06842	0.06552	0.05928	0.05692	0.04845	0.06272	0.29585	0.42201	0.52968	0.55567	0.55143	0.55517	0.55672	0.55958	1	0.23816	0.559579622	0.04844906
126	3	4	30	s	Mix	74	0.06355	0.0612	0.05576	0.05396	0.04625	0.05996	0.2778	0.39355	0.49186	0.5098	0.51092	0.51427	0.51865	0.51684	1	0.22002	0.518652781	0.04624948
127	3	4	34	s	Mix	74	0.07107	0.06829	0.0629	0.06009	0.04986	0.06333	0.29002	0.41154	0.52171	0.54152	0.53967	0.54386	0.54514	0.54663	0	0.23128	0.54663231	0.04985604
128	3	5	2	s	Mix	74	0.06268	0.06057	0.05535	0.05423	0.04566	0.06154	0.26524	0.3687	0.45696	0.47217	0.47421	0.48309	0.48549	0.48863	1	0.20381	0.488633964	0.04565793
129	3	5	9	s	Mix	74	0.06433	0.06177	0.05627	0.05489	0.04793	0.06329	0.29335	0.41166	0.52225	0.54209	0.53779	0.55135	0.55379	0.55231	1	0.23513	0.553787476	0.04792828
130	3	5	16	s	Mix	74	0.06567	0.06257	0.05744	0.05614	0.0476	0.0646	0.30433	0.42918	0.53781	0.55773	0.55376	0.56164	0.56562	0.5657	1	0.24157	0.565700587	0.04759501
131	3	5	18	s	Mix	74	0.06447	0.0623	0.05609	0.05513	0.04775	0.0637	0.30096	0.42692	0.53656	0.55557	0.55419	0.56001	0.56169	0.5619	1	0.24091	0.56189791	0.04774643
132	3	5	23	s	Mix	74	0.06477	0.06241	0.05609	0.05473	0.04701	0.06301	0.29637	0.42017	0.52658	0.54565	0.54254	0.55109	0.55292	0.55217	0	0.23629	0.552921636	0.04700986
133	3	5	25	s	Mix	74	0.06906	0.06561	0.05992	0.05789	0.05003	0.06686	0.31591	0.45088	0.56302	0.57947	0.57546	0.58084	0.58118	0.58166	1	0.25015	0.581660073	0.0500254
134	3	5	29	s	Mix	74	0.0661	0.06294	0.05697	0.05587	0.04811	0.0635	0.30401	0.43569	0.54438	0.56578	0.56177	0.57128	0.57268	0.57634	0	0.24569	0.576344619	0.04810909
135	3	5	32	s	Mix	74	0.06883	0.06539	0.05915	0.05777	0.04907	0.06554	0.30372	0.42662	0.5337	0.55449	0.5519	0.5572	0.56181	0.56439	1	0.23909	0.564385486	0.04906539
136	3	5	36	s	Mix	74	0.06587	0.06271	0.05731	0.05601	0.04833	0.06322	0.29492	0.41595	0.51987	0.53959	0.53861	0.54642	0.54712	0.54981	1	0.23345	0.549805112	0.0483284
137	3	5	40	s	Mix	74	0.067	0.06407	0.05904	0.05675	0.04687	0.06303	0.27995	0.39377	0.48686	0.50176	0.49953	0.50658	0.5074	0.50848	0	0.21473	0.508478308	0.04687222
138	3	6	3	s	Mix	74	0.0717	0.06958	0.0634	0.06269	0.05376	0.07286	0.3238	0.4552	0.55938	0.57778	0.57401	0.57875	0.57404	0.56596	1	0.24587	0.578750246	0.05376094
139	3	6	5	s	Mix	74	0.07017	0.06698	0.06133	0.06123	0.05191	0.07069	0.31779	0.43848	0.52869	0.55467	0.54832	0.56346	0.55619	0.55182	1	0.23661	0.563459518	0.05191219
140	3	6	9	s	Mix	74	0.07068	0.06814	0.06291	0.06208	0.05508	0.0728	0.32806	0.46303	0.56426	0.57922	0.57352	0.57179	0.56495	0.56339	1	0.24538	0.579224081	0.0550784
141	3	6	20	s	Mix	74	0.073	0.07005	0.06318	0.0628	0.05369	0.07236	0.33179	0.46911	0.56895	0.5895	0.58761	0.58951	0.5894	0.58279	1	0.25224	0.589507936	0.05368748
142	3	6	21	s	Mix	74	0.07415	0.07026	0.06448	0.0633	0.05225	0.07101	0.32601	0.46042	0.57058	0.58408	0.58134	0.59567	0.58965	0.58461	1	0.25175	0.595672353	0.05225071

143	3	6	23	s	Mix	74	0.07551	0.07186	0.06609	0.06428	0.05408	0.07442	0.34914	0.49032	0.6049	0.61419	0.60898	0.62015	0.6108	0.60487	1	0.26399	0.620152089	0.05407852
144	3	6	29	s	Mix	74	0.07399	0.07098	0.06466	0.06374	0.05451	0.07461	0.35013	0.49184	0.60148	0.6199	0.62524	0.62807	0.62468	0.61742	1	0.2684	0.628069583	0.05450828
145	3	6	34	s	Mix	74	0.07723	0.07194	0.06694	0.06498	0.05391	0.07453	0.34459	0.48229	0.5896	0.61046	0.60699	0.62065	0.61449	0.60842	1	0.26245	0.620653399	0.05390691
146	3	6	36	s	Mix	74	0.07912	0.07543	0.06874	0.06709	0.05631	0.07689	0.36097	0.50278	0.61726	0.63544	0.63371	0.64183	0.63584	0.63313	0	0.27319	0.64182527	0.05631194
147	3	6	38	s	Mix	74	0.07661	0.07216	0.06697	0.06472	0.05366	0.07401	0.34211	0.47398	0.5721	0.59338	0.59281	0.59987	0.59437	0.58683	1	0.25378	0.599865907	0.05365756
148	3	7	6	s	Non-mix	74	0.07339	0.06859	0.06351	0.06206	0.05082	0.0699	0.31076	0.43198	0.51995	0.54814	0.55544	0.55307	0.55597	0.55456	1	0.23472	0.555974713	0.05081747
149	3	7	11	s	Non-mix	74	0.07658	0.07349	0.068	0.06669	0.05348	0.0722	0.32485	0.45208	0.55189	0.57833	0.58149	0.5765	0.58095	0.5867	2	0.24645	0.586700434	0.05347659
150	3	7	14	s	Non-mix	74	0.08076	0.07566	0.07129	0.06882	0.05526	0.07529	0.34275	0.47232	0.5656	0.5988	0.59922	0.59503	0.59604	0.59939	1	0.25334	0.599386767	0.05526365
151	3	7	18	s	Non-mix	74	0.0706	0.06712	0.06231	0.06157	0.05267	0.0713	0.32774	0.46124	0.56659	0.59728	0.60281	0.6008	0.60253	0.60705	2	0.25812	0.607049855	0.05266633
152	3	7	22	s	Non-mix	74	0.08011	0.07438	0.06945	0.06823	0.05515	0.07374	0.33758	0.47141	0.5728	0.60082	0.60462	0.5974	0.59847	0.60411	1	0.25561	0.604623308	0.05514896
153	3	7	25	s	Non-mix	74	0.07824	0.07333	0.06762	0.06697	0.05483	0.07427	0.34475	0.48578	0.5908	0.6212	0.62591	0.61928	0.61979	0.62422	1	0.26601	0.62590533	0.05483053
154	3	7	29	s	Non-mix	74	0.07235	0.06762	0.06274	0.0619	0.05203	0.07046	0.32066	0.4466	0.53917	0.56884	0.57425	0.57355	0.56889	0.57163	1	0.24365	0.574250768	0.05202933
155	3	7	31	s	Non-mix	74	0.08558	0.08043	0.0741	0.07141	0.0583	0.07767	0.34844	0.48641	0.58152	0.61439	0.61673	0.6125	0.61176	0.61588	1	0.25971	0.61673182	0.05830291
156	3	7	36	s	Non-mix	74	0.0738	0.06924	0.0644	0.06371	0.05213	0.07116	0.3199	0.44638	0.54207	0.56921	0.57414	0.57237	0.57121	0.57907	1	0.24393	0.579068908	0.05212763
157	3	7	40	s	Non-mix	74	0.08561	0.08024	0.07402	0.07289	0.05835	0.07962	0.35573	0.49078	0.57907	0.61201	0.61502	0.61154	0.6119	0.62093	1	0.25947	0.620933008	0.05835441
158	3	8	5	s	Non-mix	74	0.07704	0.07273	0.06728	0.06646	0.05493	0.07674	0.34132	0.47019	0.55761	0.58503	0.59371	0.59252	0.59399	0.58986	1	0.25111	0.593993834	0.05493248
159	3	8	10	s	Non-mix	74	0.07694	0.07238	0.06675	0.0656	0.05444	0.07619	0.33755	0.46669	0.55034	0.57517	0.57747	0.58737	0.58904	0.58419	1	0.2475	0.589036033	0.05444145
160	3	8	13	s	Non-mix	74	0.08216	0.07745	0.07257	0.07112	0.0574	0.08114	0.35562	0.48752	0.57924	0.60679	0.611	0.60945	0.60669	0.60688	1	0.25771	0.610995738	0.05740239
161	3	8	18	s	Non-mix	74	0.07616	0.07091	0.06668	0.06559	0.05062	0.07325	0.31786	0.43138	0.50688	0.52579	0.53239	0.53406	0.53741	0.53868	1	0.22496	0.538681674	0.05062471
162	3	8	22	s	Non-mix	74	0.07964	0.07497	0.06911	0.06805	0.05545	0.07751	0.34581	0.47487	0.56436	0.59022	0.59404	0.59649	0.59155	0.59221	1	0.2517	0.596490545	0.05545428
163	3	8	25	s	Non-mix	74	0.07879	0.07355	0.06862	0.06749	0.05436	0.07681	0.33722	0.46166	0.54704	0.57409	0.57929	0.57458	0.5765	0.57391	1	0.24376	0.579288482	0.05436238
164	3	8	30	s	Non-mix	74	0.0769	0.07212	0.06624	0.06656	0.05274	0.0754	0.33785	0.46192	0.55018	0.572	0.57536	0.57808	0.57746	0.58133	1	0.24524	0.581327027	0.05274248
165	3	8	33	s	Non-mix	74	0.08059	0.07461	0.07005	0.06907	0.05455	0.07771	0.3304	0.4504	0.53054	0.55627	0.55952	0.56453	0.55829	0.55522	1	0.23509	0.564526775	0.05455101
166	3	8	36	s	Non-mix	74	0.08374	0.07852	0.07096	0.07164	0.05645	0.07873	0.34151	0.47121	0.55572	0.58283	0.58337	0.58963	0.58636	0.58136	1	0.24655	0.589629988	0.05644656
167	3	8	39	s	Non-mix	74	0.08427	0.07839	0.07203	0.07108	0.05605	0.07981	0.33747	0.46054	0.54447	0.56283	0.56602	0.57711	0.56698	0.56701	1	0.2388	0.57711283	0.05604954
168	4	1	2	s	Non-mix	78	0.06981	0.06796	0.0643	0.062	0.04844	0.06437	0.21245	0.30851	0.38154	0.40162	0.39881	0.40587	0.40368	0.40489	2	0.1633	0.405872353	0.04844381
169	4	1	6	s	Non-mix	78	0.06683	0.063	0.0586	0.05703	0.04527	0.05968	0.20846	0.30594	0.39342	0.40975	0.41024	0.41761	0.42067	0.421	2	0.17136	0.421004192	0.04527383
170	4	1	9	s	Non-mix	78	0.06985	0.06625	0.06163	0.06044	0.04732	0.06232	0.21208	0.30704	0.38475	0.40338	0.4048	0.41206	0.41408	0.41417	2	0.16689	0.414167574	0.04731605
171	4	1	14	s	Non-mix	78	0.06763	0.06451	0.06072	0.05934	0.0452	0.05968	0.20468	0.29549	0.37035	0.38802	0.39045	0.39468	0.39843	0.4011	2	0.16047	0.401100869	0.04519708
172	4	1	18	s	Non-mix	78	0.06618	0.06318	0.05849	0.05658	0.04366	0.05742	0.20862	0.30635	0.39146	0.40735	0.40713	0.41309	0.41182	0.4139	2	0.16955	0.413901946	0.04366139
173	4	1	21	s	Non-mix	78	0.07529	0.07173	0.06745	0.06502	0.04962	0.0646	0.22325	0.32196	0.40135	0.42062	0.42066	0.42296	0.4274	0.43301	2	0.17236	0.433010368	0.04962378
174	4	1	27	s	Non-mix	78	0.07794	0.07341	0.06835	0.06668	0.05073	0.0673	0.24203	0.35628	0.45441	0.46955	0.46719	0.47218	0.47287	0.47275	2	0.19429	0.472869962	0.05072973
175	4	1	32	s	Non-mix	78	0.06783	0.06573	0.06254	0.06054	0.0472	0.06118	0.20282	0.29669	0.3784	0.3973	0.39908	0.40182	0.40468	0.40806	1	0.16347	0.408063634	0.04719694
176	4	1	35	s	Non-mix	78	0.06231	0.06025	0.05681	0.05626	0.04524	0.05829	0.19195	0.28689	0.37828	0.39826	0.39823	0.40776	0.40736	0.4124	2	0.16649	0.41240206	0.04524276
177	4	1	39	s	Non-mix	78	0.06549	0.06289	0.05925	0.05809	0.04619	0.05985	0.20279	0.29809	0.38943	0.40442	0.41036	0.41665	0.41956	0.41864	1	0.17005	0.419560721	0.04619476
178	4	1	43	l	Non-mix	78	0.07416	0.07084	0.06519	0.06585	0.04742	0.06185	0.21768	0.3078	0.37972	0.39651	0.39229	0.39597	0.4005	0.3979	2	0.15996	0.400497387	0.04741688
179	4	2	3	s	Mix	78	0.05495	0.053	0.0485	0.0459	0.03787	0.05102	0.1841	0.29796	0.39856	0.41839	0.42248	0.42598	0.42885	0.42701	1	0.18013	0.428853072	0.03787151
180	4	2	5	s	Mix	78	0.05824	0.05602	0.05141	0.04901	0.04026	0.05385	0.18548	0.29455	0.38176	0.40114	0.40656	0.40934	0.41206	0.40995	1	0.17069	0.412057459	0.04026204
181	4	2	12	s	Mix	78	0.05214	0.05038	0.04647	0.04356	0.03626	0.04835	0.17499	0.28912	0.38956	0.41226	0.41512	0.41447	0.41695	0.41749	1	0.17682	0.417491204	0.03625847
182	4	2	16	s	Mix	78	0.05635	0.05445	0.05017	0.04728	0.03931	0.05272	0.18993	0.30884	0.40808	0.43039	0.43374	0.43639	0.43724	0.43651	1	0.1844	0.437242909	0.03930505
183	4	2	20	s	Mix	78	0.06004	0.05725	0.05276	0.04962	0.04175	0.05762	0.21348	0.34928	0.46563	0.48703	0.49012	0.49126	0.49381	0.49306	1	0.2102	0.493806005	0.04175475
184	4	2	24	s	Mix	78	0.05784	0.05563	0.05081	0.04832	0.04016	0.05302	0.19313	0.31186	0.41487	0.43297	0.4368	0.4341	0.43717	0.43841	1	0.18487	0.438414249	0.04015868
185	4	2	29	s	Mix	78	0.05212	0.05018	0.04729	0.04417	0.0376	0.05066	0.17502	0.28184	0.37357	0.39354	0.39695	0.3967	0.40183	0.40141	1	0.16819	0.401827203	0.03759869
186	4	2	32	s	Mix	78	0.05658	0.05455	0.04984	0.04746	0.03946	0.05149	0.18286	0.2904	0.38055	0.39852	0.40098	0.40519	0.40548	0.40607	1	0.16945	0.406070085	0.03946392
187	4	2	37	s	Mix	78	0.05894	0.05671	0.05211	0.04968	0.03943	0.05358	0.19496	0.31176	0.40712	0.42961	0.4317	0.43366	0.43479	0.43556	1	0.18278	0.435562546	0.03942858
188	4	2	40	s	Mix	78	0.05822	0.05669	0.05223	0.05009	0.04098	0.05463	0.1967	0.31576	0.41568	0.43775	0.44052	0.44409	0.44517	0.44678	1	0.1872	0.446781478	0.04098398
189	4	2	58	l	Mix	78	0.06292	0.06002	0.05481	0.05402	0.04352	0.05749	0.1853	0.28153	0.35375	0.36703	0.36742	0.3709	0.36689	0.37316	1	0.15095	0.373163848	0.04352355
190	4	3	1	s	Non-mix	78	0.0633	0.06084	0.05566	0.05314	0.04277	0.05881	0.19648	0.29805	0.37403	0.39082	0.39744	0.40358	0.40615	0.40652	2	0.16528	0.406520661	0.04277409

191	4	3	3	s	Non-mix	78	0.066	0.06359	0.0587	0.05555	0.04527	0.06158	0.20551	0.30805	0.38476	0.39995	0.4015	0.40941	0.41392	0.41183	2	0.16746	0.413918892	0.04527476
192	4	3	8	s	Non-mix	78	0.05812	0.0548	0.05235	0.0505	0.04113	0.0512	0.17905	0.28143	0.36382	0.37578	0.3831	0.38458	0.38929	0.38772	2	0.15972	0.389285033	0.04112866
193	4	3	15	s	Non-mix	78	0.06311	0.06078	0.05586	0.05301	0.04252	0.05789	0.19554	0.2941	0.37039	0.38521	0.38776	0.39509	0.39794	0.39362	2	0.16144	0.397937915	0.0425181
194	4	3	17	s	Non-mix	78	0.06043	0.05733	0.05328	0.0504	0.04096	0.05585	0.19505	0.2998	0.38304	0.39741	0.40001	0.40748	0.41004	0.40784	2	0.16873	0.410038999	0.04096372
195	4	3	20	s	Non-mix	78	0.05894	0.05614	0.05371	0.05247	0.04159	0.05177	0.16805	0.2656	0.35465	0.36963	0.3728	0.37556	0.37894	0.38264	2	0.15505	0.382636915	0.0415939
196	4	3	22	s	Non-mix	78	0.05759	0.05486	0.05112	0.04856	0.04028	0.05523	0.19186	0.30169	0.38944	0.40525	0.40647	0.4115	0.41123	0.41099	1	0.17192	0.411501474	0.04028402
197	4	3	27	l	Non-mix	78	0.05954	0.05723	0.05406	0.05326	0.04013	0.05049	0.16829	0.26213	0.34322	0.36056	0.36376	0.36702	0.37225	0.3739	2	0.15077	0.373895262	0.04012753
198	4	3	31	s	Non-mix	78	0.06564	0.06286	0.05777	0.05468	0.04321	0.0591	0.20373	0.31177	0.39347	0.41173	0.41401	0.42185	0.42273	0.42068	1	0.17313	0.422725452	0.04320757
199	4	3	35	s	Non-mix	78	0.06398	0.06173	0.05716	0.05396	0.04319	0.05857	0.20799	0.3222	0.40783	0.42439	0.42542	0.43041	0.43384	0.43054	2	0.17903	0.433836794	0.0431873
200	4	3	40	s	Non-mix	78	0.06047	0.05775	0.05305	0.05235	0.04109	0.05111	0.17625	0.27373	0.35875	0.37625	0.38146	0.38345	0.38747	0.3907	2	0.15851	0.390696143	0.04108966
201	4	4	1	s	Mix	78	0.06046	0.05777	0.05575	0.05363	0.04416	0.05669	0.18287	0.28625	0.37662	0.39272	0.39625	0.39433	0.39708	0.39954	1	0.1638	0.399543426	0.04415735
202	4	4	3	s	Mix	78	0.0565	0.05358	0.05142	0.04981	0.04191	0.05412	0.18098	0.28583	0.37899	0.39517	0.39982	0.39937	0.4013	0.40455	1	0.16731	0.404552272	0.0419099
203	4	4	10	s	Mix	78	0.06359	0.05991	0.05661	0.05446	0.04348	0.05875	0.20557	0.32335	0.42754	0.44463	0.44417	0.4477	0.44718	0.44774	1	0.18754	0.447738184	0.04347709
204	4	4	15	s	Mix	78	0.05593	0.0525	0.05032	0.04867	0.04014	0.05272	0.17969	0.28456	0.37413	0.39092	0.39345	0.39609	0.39696	0.3975	1	0.16549	0.397495762	0.0401428
205	4	4	16	l	Mix	78	0.06135	0.05783	0.05359	0.05192	0.04084	0.05488	0.18583	0.28541	0.36664	0.38296	0.38164	0.38531	0.38805	0.39242	1	0.15974	0.392421455	0.04084377
206	4	4	17	s	Mix	78	0.05723	0.05451	0.05151	0.04961	0.04159	0.05523	0.19565	0.31736	0.42448	0.44602	0.44604	0.44809	0.4493	0.45064	1	0.1902	0.450636925	0.04159171
207	4	4	19	s	Mix	78	0.05748	0.05462	0.05242	0.04991	0.04065	0.05426	0.19232	0.3061	0.40542	0.42457	0.42426	0.42592	0.42774	0.42714	1	0.17976	0.427737045	0.04065442
208	4	4	30	s	Mix	78	0.05519	0.05266	0.05017	0.04908	0.03924	0.05176	0.17494	0.27144	0.35315	0.36995	0.37343	0.37769	0.3826	0.38005	1	0.15661	0.382602925	0.0392383
209	4	4	32	s	Mix	78	0.05383	0.0511	0.04781	0.04601	0.03793	0.04915	0.17133	0.26753	0.3509	0.36817	0.37176	0.37742	0.38337	0.38373	1	0.15748	0.383734172	0.03792784
210	4	4	34	s	Mix	78	0.06049	0.05719	0.05242	0.05125	0.04079	0.05353	0.18964	0.29395	0.3829	0.40123	0.40465	0.40752	0.41362	0.41505	1	0.17039	0.415046793	0.04079452
211	4	4	36	s	Mix	78	0.06171	0.05864	0.05395	0.05287	0.04187	0.0553	0.19586	0.30518	0.39521	0.41526	0.4193	0.42234	0.42739	0.42958	0	0.17652	0.429582143	0.04186943
212	4	5	2	s	Mix	78	0.05111	0.04848	0.04494	0.04177	0.03486	0.04546	0.15127	0.2445	0.32955	0.34443	0.34779	0.35201	0.35807	0.35353	1	0.14693	0.358070551	0.03486336
213	4	5	5	s	Mix	78	0.05342	0.05056	0.04605	0.044	0.03646	0.0482	0.16884	0.27625	0.37404	0.39323	0.3986	0.3998	0.40459	0.40515	1	0.16938	0.405146073	0.03646353
214	4	5	9	s	Mix	78	0.05435	0.05093	0.04659	0.04393	0.03714	0.04968	0.17455	0.28698	0.39028	0.41008	0.41258	0.41599	0.42027	0.41925	1	0.17656	0.420272804	0.03714221
215	4	5	14	s	Mix	78	0.05499	0.05131	0.04695	0.04496	0.03669	0.04877	0.16883	0.27702	0.37425	0.39153	0.39485	0.39812	0.40503	0.40128	1	0.16817	0.405025071	0.03669413
216	4	5	18	l	Mix	78	0.05729	0.05454	0.05063	0.04858	0.03907	0.052	0.16906	0.26726	0.3477	0.36164	0.36081	0.36676	0.36792	0.36959	1	0.15133	0.369585823	0.03906933
217	4	5	18	s	Mix	78	0.05225	0.04881	0.04485	0.04233	0.03544	0.04716	0.16899	0.28088	0.3842	0.40524	0.40588	0.40879	0.41698	0.41298	1	0.17476	0.416982115	0.03544219
218	4	5	23	s	Mix	78	0.0551	0.05153	0.04694	0.04486	0.03594	0.04756	0.17073	0.28089	0.38144	0.40088	0.40389	0.40672	0.41016	0.40775	0	0.17199	0.410162617	0.0359354
219	4	5	25	s	Mix	78	0.05984	0.05573	0.05103	0.04868	0.03899	0.05243	0.19173	0.31851	0.43314	0.45277	0.45284	0.45819	0.46155	0.45763	1	0.19466	0.461547319	0.0389882
220	4	5	32	s	Mix	78	0.06224	0.0579	0.05271	0.0501	0.04034	0.05303	0.18761	0.30748	0.41515	0.43657	0.43825	0.4436	0.44692	0.44722	1	0.18675	0.447216449	0.04034495
221	4	5	38	s	Mix	78	0.0584	0.05499	0.05024	0.04725	0.0379	0.05111	0.17875	0.28953	0.3868	0.40549	0.40978	0.41438	0.41718	0.41861	1	0.17391	0.418608559	0.03789624
222	4	5	40	s	Mix	78	0.06043	0.05679	0.05232	0.04991	0.03937	0.05285	0.1826	0.29041	0.38105	0.40318	0.4063	0.4103	0.41214	0.41407	1	0.17093	0.414071885	0.03936859
223	4	6	1	s	Mix	78	0.05754	0.05498	0.05051	0.04941	0.03972	0.05328	0.17855	0.27808	0.35902	0.37515	0.37675	0.38395	0.38638	0.3862	1	0.15855	0.386379206	0.03972435
224	4	6	5	s	Mix	78	0.05795	0.05496	0.05133	0.04996	0.04118	0.05498	0.18325	0.28684	0.37208	0.39086	0.39315	0.39757	0.40012	0.40039	1	0.16513	0.400390229	0.04117676
225	4	6	13	s	Mix	78	0.06114	0.05728	0.05255	0.04919	0.0393	0.05489	0.19683	0.31241	0.40406	0.42462	0.42571	0.42853	0.43095	0.43047	1	0.18024	0.430945498	0.03929953
226	4	6	16	l	Mix	78	0.06249	0.05855	0.05416	0.05219	0.04171	0.05814	0.19106	0.2952	0.37111	0.39008	0.3872	0.39151	0.39579	0.39553	1	0.16209	0.395789071	0.04171424
227	4	6	18	s	Mix	78	0.06262	0.05894	0.05437	0.05172	0.04042	0.0566	0.20821	0.33648	0.44306	0.46133	0.46163	0.46792	0.47022	0.46966	1	0.19792	0.470219839	0.04042223
228	4	6	20	s	Mix	78	0.06099	0.05765	0.05326	0.05071	0.04103	0.05539	0.20079	0.32126	0.41845	0.43519	0.43511	0.43965	0.4447	0.44137	1	0.18553	0.444699378	0.04103462
229	4	6	21	s	Mix	78	0.05998	0.05625	0.05165	0.04902	0.03875	0.05383	0.18872	0.29577	0.37792	0.39756	0.40279	0.40874	0.41137	0.41085	1	0.16976	0.411372844	0.03875045
230	4	6	26	s	Mix	78	0.06213	0.05843	0.05294	0.05021	0.04069	0.05732	0.209	0.33591	0.44248	0.46027	0.46123	0.46877	0.47123	0.46888	1	0.19806	0.471234702	0.0406943
231	4	6	35	s	Mix	78	0.05909	0.0556	0.05106	0.04853	0.03933	0.05392	0.18888	0.29509	0.38412	0.40118	0.40595	0.41377	0.41758	0.41542	1	0.17214	0.417579336	0.03933286
232	4	6	38	s	Mix	78	0.06719	0.06253	0.05685	0.05378	0.04194	0.05886	0.21257	0.3302	0.42119	0.43866	0.43997	0.44793	0.45282	0.44764	1	0.18653	0.452821609	0.04194435
233	4	6	40	s	Mix	78	0.06395	0.05983	0.05401	0.05149	0.04029	0.05648	0.20309	0.31599	0.40751	0.42365	0.42419	0.43739	0.43906	0.43645	1	0.18123	0.439056727	0.04029047
234	4	7	2	s	Non-mix	78	0.06446	0.06242	0.0595	0.05696	0.04646	0.06147	0.19361	0.29964	0.38282	0.40031	0.40389	0.40876	0.41169	0.41595	1	0.1675	0.415945758	0.04645689
235	4	7	6	s	Non-mix	78	0.05795	0.05694	0.05363	0.05137	0.0432	0.05813	0.17934	0.28059	0.36006	0.3776	0.38196	0.38721	0.38987	0.3914	1	0.15899	0.391401092	0.04320312
236	4	7	9	s	Non-mix	78	0.06074	0.05887	0.05527	0.05284	0.04452	0.05906	0.19015	0.29868	0.38421	0.40215	0.40387	0.40999	0.41133	0.41416	1	0.16929	0.414158603	0.04451544
237	4	7	11	s	Non-mix	78	0.05927	0.05717	0.05415	0.05202	0.04425	0.05788	0.18264	0.28776	0.37273	0.38921	0.39217	0.39372	0.39726	0.39984	1	0.16326	0.399843605	0.04425425
238	4	7	14	s	Non-mix	78	0.06446	0.06232	0.05872	0.05669	0.04601	0.06157	0.19531	0.30461	0.38887	0.40705	0.40615	0.40948	0.41379	0.41559	1	0.16907	0.415587097	0.0460101

239	4	7	20	s	Non-mix	78	0.06002	0.05809	0.05493	0.05249	0.04359	0.05866	0.1904	0.30056	0.39351	0.41061	0.41023	0.41621	0.42104	0.42023	2	0.17323	0.421044103	0.04359363
240	4	7	22	s	Non-mix	78	0.05967	0.05837	0.05509	0.05315	0.04494	0.05899	0.18794	0.29631	0.38679	0.40058	0.40266	0.40433	0.40848	0.40531	2	0.1679	0.408479413	0.04494087
241	4	7	25	s	Non-mix	78	0.06372	0.06105	0.05777	0.05544	0.04553	0.06037	0.20015	0.31381	0.41098	0.42951	0.43148	0.43339	0.43704	0.43633	2	0.18052	0.437035592	0.04552849
242	4	7	34	s	Non-mix	78	0.06458	0.06233	0.05929	0.05737	0.04643	0.06061	0.19524	0.30343	0.39209	0.40943	0.41214	0.41446	0.41786	0.42173	1	0.17107	0.42173281	0.04642806
243	4	7	40	s	Non-mix	78	0.06778	0.0659	0.06216	0.05916	0.04697	0.06335	0.20215	0.30965	0.39751	0.41407	0.41275	0.42164	0.42343	0.42349	2	0.17193	0.423485362	0.046973
244	4	7	42	l	Non-mix	78	0.0691	0.0657	0.06309	0.06092	0.0468	0.0609	0.19853	0.29526	0.37285	0.3906	0.38716	0.38193	0.38745	0.39233	2	0.15731	0.392332162	0.04679617
245	4	8	4	s	Non-mix	78	0.06775	0.06604	0.06161	0.05856	0.04754	0.06324	0.1951	0.31043	0.4068	0.42357	0.42597	0.4334	0.43412	0.43292	1	0.17716	0.434118685	0.0475376
246	4	8	8	s	Non-mix	78	0.05825	0.05515	0.05084	0.04781	0.03865	0.05117	0.17419	0.27838	0.35919	0.37571	0.38161	0.38597	0.39071	0.39157	1	0.16035	0.391571259	0.03864738
247	4	8	10	s	Non-mix	78	0.0553	0.0525	0.04836	0.04562	0.03741	0.04939	0.16746	0.2648	0.34535	0.36214	0.36616	0.37257	0.37671	0.37835	2	0.15459	0.378349858	0.03740688
248	4	8	11	s	Non-mix	78	0.06377	0.06123	0.05699	0.05381	0.04375	0.05952	0.20054	0.31036	0.40248	0.41502	0.41304	0.41969	0.42221	0.41911	1	0.17393	0.422210501	0.04374833
249	4	8	13	l	Non-mix	78	0.06376	0.06133	0.05616	0.05352	0.04113	0.05326	0.17506	0.267	0.33674	0.34704	0.34562	0.35079	0.35148	0.35025	2	0.1418	0.351484797	0.04112766
250	4	8	13	s	Non-mix	78	0.06468	0.06186	0.05761	0.05471	0.0443	0.05933	0.19949	0.30887	0.40117	0.41342	0.40989	0.42089	0.42257	0.41896	1	0.17328	0.422567952	0.0442983
251	4	8	17	s	Non-mix	78	0.05347	0.05032	0.04616	0.0437	0.03567	0.04601	0.15767	0.25136	0.32978	0.34533	0.34773	0.35278	0.35867	0.35601	2	0.14669	0.358666874	0.03567404
252	4	8	25	s	Non-mix	78	0.05873	0.05611	0.0508	0.0482	0.03827	0.0494	0.17389	0.27961	0.36754	0.38279	0.38495	0.38687	0.38979	0.39246	1	0.16188	0.392457619	0.03827105
253	4	8	28	s	Non-mix	78	0.05727	0.05389	0.04921	0.04635	0.03743	0.04925	0.17079	0.27189	0.35446	0.37206	0.37485	0.37716	0.38346	0.38202	1	0.15766	0.383461869	0.03742784
254	4	8	32	s	Non-mix	78	0.06282	0.06195	0.0574	0.05532	0.04467	0.05966	0.18585	0.28026	0.36368	0.37991	0.38003	0.39003	0.39366	0.39155	1	0.15806	0.39366209	0.04467184
255	4	8	39	s	Non-mix	78	0.06139	0.05859	0.05345	0.04991	0.04017	0.05184	0.1753	0.27551	0.35589	0.37117	0.37689	0.37956	0.38519	0.38353	1	0.15658	0.385194784	0.04017055
256	6	1	11	s	Non-mix	86	0.03429	0.0334	0.03294	0.03276	0.02752	0.0403	0.09929	0.13945	0.15831	0.16788	0.1818	0.18934	0.19176	0.19617	2	0.07199	0.196174998	0.027525
257	6	1	14	s	Non-mix	86	0.0424	0.0414	0.04107	0.04013	0.02823	0.04703	0.12943	0.1755	0.18843	0.19067	0.19693	0.20103	0.20207	0.20217	2	0.07707	0.202166672	0.02823333
258	6	1	27	s	Non-mix	86	0.0295	0.0286	0.0281	0.0292	0.0242	0.036	0.0859	0.1166	0.1315	0.1453	0.1602	0.1697	0.1716	0.1772	2	0.06377	0.177200004	0.0242
259	6	2	3	s	Mix	86	0.05624	0.0532	0.04981	0.04871	0.03078	0.05443	0.21235	0.33123	0.37132	0.37621	0.37982	0.38124	0.38144	0.38068	2	0.16155	0.381438884	0.03078333
260	6	2	8	s	Mix	86	0.05463	0.05203	0.04911	0.04825	0.03193	0.05476	0.19013	0.28621	0.31717	0.3205	0.32506	0.32598	0.32643	0.32681	2	0.135	0.326813334	0.03192667
261	6	2	10	s	Mix	86	0.05802	0.05511	0.05214	0.05135	0.03439	0.05777	0.20811	0.31937	0.35772	0.36327	0.36969	0.37219	0.37192	0.37293	2	0.15499	0.372933333	0.03438889
262	6	2	11	s	Mix	86	0.04629	0.04351	0.04106	0.04045	0.02526	0.04409	0.18069	0.29072	0.32802	0.33583	0.34229	0.34549	0.34575	0.34677	2	0.14707	0.346768422	0.02525789
263	6	2	14	s	Mix	86	0.04898	0.04665	0.04381	0.04321	0.02859	0.04879	0.18515	0.28988	0.32846	0.33911	0.35041	0.35674	0.3586	0.36165	2	0.14938	0.361649995	0.028595
264	6	2	16	s	Mix	86	0.04899	0.04655	0.04403	0.04332	0.02856	0.04852	0.18736	0.29865	0.33733	0.34596	0.35282	0.35705	0.35808	0.3577	2	0.15071	0.358078951	0.02855789
265	6	2	23	s	Mix	86	0.04113	0.03868	0.03693	0.03628	0.02269	0.04097	0.16444	0.26696	0.30348	0.3114	0.32067	0.32513	0.32622	0.32837	2	0.13836	0.328372223	0.02269444
266	6	2	29	s	Mix	86	0.04282	0.04033	0.03807	0.03751	0.02429	0.04176	0.16677	0.27708	0.31786	0.32709	0.33466	0.3402	0.34091	0.34251	2	0.14481	0.342505558	0.02429444
267	6	2	35	s	Mix	86	0.04308	0.04052	0.03819	0.0375	0.02283	0.0403	0.15984	0.25418	0.28563	0.29102	0.29552	0.29803	0.29858	0.29928	2	0.12614	0.299279999	0.022825
268	6	2	37	s	Mix	86	0.05237	0.04977	0.04714	0.04661	0.02983	0.05053	0.17347	0.26139	0.2923	0.29868	0.30676	0.3108	0.31157	0.31202	2	0.12687	0.312021053	0.02983158
269	6	3	22	s	Non-mix	86	0.0406	0.0399	0.04	0.03985	0.03215	0.0467	0.11375	0.1522	0.1711	0.1825	0.1922	0.19885	0.20055	0.20435	2	0.07401	0.204350002	0.03215
270	6	4	3	s	Mix	86	0.05566	0.0528	0.05027	0.04967	0.03438	0.05568	0.20144	0.31267	0.35107	0.35967	0.36519	0.36824	0.36932	0.36925	2	0.15367	0.369324999	0.034385
271	6	4	8	s	Mix	86	0.05704	0.05432	0.05172	0.05107	0.03563	0.05716	0.18637	0.27592	0.30654	0.31352	0.31903	0.32183	0.32334	0.32515	2	0.13076	0.325149998	0.03562778
272	6	4	11	s	Mix	86	0.05126	0.04806	0.04544	0.04475	0.02876	0.04978	0.19862	0.31854	0.36007	0.36857	0.3711	0.37608	0.37734	0.37909	2	0.16019	0.379094442	0.02876111
273	6	4	14	s	Mix	86	0.05044	0.04791	0.04525	0.04454	0.03005	0.04932	0.17896	0.27781	0.31105	0.31927	0.32286	0.32728	0.32881	0.3305	2	0.1364	0.330500002	0.03004706
274	6	4	17	s	Mix	86	0.05011	0.04716	0.0447	0.04392	0.02922	0.04954	0.1888	0.30268	0.34319	0.35093	0.35765	0.36323	0.36452	0.36773	2	0.15342	0.367730002	0.029215
275	6	4	23	s	Mix	86	0.04853	0.04581	0.04301	0.04258	0.02722	0.04724	0.19198	0.30981	0.35099	0.35909	0.36477	0.36898	0.37063	0.3732	2	0.1576	0.373199999	0.02722222
276	6	4	30	s	Mix	86	0.04622	0.04368	0.0414	0.04065	0.02668	0.04401	0.16579	0.26181	0.29514	0.30262	0.30838	0.31273	0.31388	0.31591	2	0.13091	0.315911766	0.02667647
277	6	4	34	s	Mix	86	0.05182	0.04909	0.04607	0.04526	0.0276	0.04922	0.17947	0.27454	0.30514	0.30918	0.31327	0.31506	0.31556	0.31643	2	0.13122	0.31643333	0.0276
278	6	4	36	s	Mix	86	0.0561	0.05283	0.04911	0.04818	0.02958	0.05241	0.20079	0.31441	0.35167	0.35597	0.35961	0.36147	0.36163	0.36031	2	0.15219	0.361626321	0.02957895
279	6	4	39	s	Mix	86	0.05337	0.05039	0.04714	0.04634	0.02849	0.05078	0.19331	0.30263	0.33857	0.34279	0.34732	0.349	0.34878	0.34932	2	0.14699	0.349315792	0.02849474
280	6	5	2	s	Mix	86	0.05362	0.05105	0.04852	0.04792	0.03289	0.05398	0.18898	0.29283	0.32891	0.33502	0.34132	0.34488	0.34539	0.3454	2	0.14298	0.345399999	0.03288824
281	6	5	5	s	Mix	86	0.05213	0.04916	0.04646	0.04591	0.03038	0.05048	0.19291	0.30628	0.3449	0.35256	0.35891	0.36216	0.36275	0.36268	2	0.15246	0.362750001	0.030385
282	6	5	10	s	Mix	86	0.05098	0.04808	0.04519	0.04449	0.02903	0.05046	0.2013	0.33102	0.37554	0.37904	0.38462	0.38572	0.38564	0.38544	2	0.16553	0.385716667	0.02903333
283	6	5	13	s	Mix	86	0.04844	0.04553	0.04238	0.04162	0.02588	0.04633	0.19273	0.31519	0.356	0.36102	0.3666	0.3687	0.3688	0.36832	2	0.1583	0.368795	0.025885
284	6	5	18	s	Mix	86	0.04866	0.04545	0.04238	0.04168	0.02677	0.04637	0.18814	0.30607	0.34871	0.35551	0.36289	0.36594	0.36658	0.36675	2	0.15609	0.36675	0.02676667
285	6	5	22	s	Mix	86	0.04607	0.04356	0.04106	0.04055	0.02693	0.04536	0.17297	0.279	0.31915	0.3294	0.33926	0.346	0.34829	0.3508	2	0.14563	0.350795002	0.02693
286	6	5	25	s	Mix	86	0.05103	0.04777	0.04491	0.04403	0.027	0.04836	0.19872	0.32057	0.36152	0.36785	0.37452	0.37757	0.37801	0.37897	2	0.16117	0.378973683	0.027

287	6	5	34	s	Mix	86	0.0523	0.04938	0.04649	0.04573	0.02812	0.04912	0.19049	0.30023	0.33711	0.3433	0.34722	0.35033	0.35118	0.3514	2	0.14769	0.351400001	0.028125
288	6	5	38	s	Mix	86	0.05429	0.051	0.04769	0.04682	0.02725	0.05009	0.1943	0.30142	0.33642	0.34149	0.34346	0.34511	0.34455	0.34349	2	0.14529	0.345110001	0.027255
289	6	5	40	s	Mix	86	0.05082	0.04838	0.04573	0.04515	0.02884	0.0492	0.16292	0.24509	0.27311	0.27753	0.28104	0.28316	0.28347	0.28377	2	0.11548	0.283770002	0.028835
290	6	6	1	s	Mix	86	0.05915	0.05669	0.05423	0.0534	0.03647	0.05934	0.19167	0.28477	0.31434	0.32124	0.3262	0.32915	0.32932	0.32984	2	0.1332	0.329842108	0.03646842
291	6	6	5	s	Mix	86	0.05253	0.04961	0.04714	0.04666	0.03078	0.05264	0.18764	0.28973	0.32394	0.32769	0.33538	0.33806	0.33852	0.33862	2	0.14064	0.338622224	0.03077778
292	6	6	9	s	Mix	86	0.05308	0.05092	0.04829	0.04746	0.03285	0.05332	0.17669	0.26435	0.29531	0.30228	0.30938	0.31388	0.31494	0.31611	2	0.12767	0.316105269	0.03285263
293	6	6	12	s	Mix	86	0.05238	0.04948	0.04653	0.04573	0.02868	0.05128	0.19698	0.31146	0.35155	0.35886	0.36891	0.37447	0.37614	0.37758	2	0.15759	0.377582355	0.02868235
294	6	6	20	s	Mix	86	0.04931	0.04677	0.04382	0.04286	0.02624	0.04773	0.19511	0.31824	0.36181	0.37114	0.37912	0.38401	0.38518	0.38574	2	0.16381	0.385742104	0.02623684
295	6	6	21	s	Mix	86	0.04561	0.04329	0.04038	0.03966	0.02436	0.04327	0.16352	0.25276	0.28485	0.29299	0.30286	0.30877	0.30973	0.31132	2	0.12821	0.311315791	0.02436316
296	6	6	24	s	Mix	86	0.05221	0.04902	0.04559	0.0446	0.02645	0.04927	0.20796	0.33704	0.38179	0.38672	0.39084	0.39357	0.39434	0.39395	2	0.16929	0.394336842	0.02645263
297	6	6	35	s	Mix	86	0.04584	0.04364	0.04141	0.04096	0.02565	0.04459	0.16572	0.26246	0.2952	0.30204	0.30659	0.30991	0.30997	0.31047	2	0.12987	0.310468184	0.02565
298	6	6	38	s	Mix	86	0.05463	0.05105	0.04773	0.04724	0.02732	0.05064	0.20438	0.32025	0.35814	0.36377	0.36732	0.36873	0.36904	0.36769	2	0.15656	0.369044446	0.02732222
299	6	6	40	s	Mix	86	0.05166	0.04865	0.04554	0.04482	0.02726	0.04775	0.18748	0.29304	0.32934	0.33606	0.34006	0.3434	0.34399	0.34256	2	0.14442	0.343994445	0.02725556
300	6	7	2	s	Non-mix	86	0.0398	0.03898	0.03874	0.03848	0.03176	0.04543	0.1146	0.16264	0.18224	0.19069	0.20087	0.20704	0.20813	0.21172	2	0.07871	0.211720002	0.03176
301	6	7	6	s	Non-mix	86	0.04248	0.04154	0.04093	0.04054	0.03144	0.04768	0.13086	0.1903	0.21289	0.22051	0.23282	0.23887	0.24059	0.24445	2	0.09325	0.244450001	0.0314375
302	6	7	9	s	Non-mix	86	0.04801	0.04651	0.04532	0.04509	0.03448	0.05191	0.15083	0.22401	0.24959	0.25954	0.26916	0.27469	0.27672	0.27868	2	0.10887	0.27867857	0.03447857
303	6	7	14	s	Non-mix	86	0.0405	0.03955	0.039	0.03879	0.03032	0.0459	0.11349	0.15869	0.17717	0.18426	0.19595	0.20076	0.20099	0.2054	2	0.07573	0.205400002	0.03032
304	6	7	20	s	Non-mix	86	0.03705	0.03619	0.03554	0.03543	0.02805	0.04083	0.11322	0.1707	0.19173	0.19909	0.20738	0.21305	0.21515	0.21829	2	0.08397	0.218293334	0.02804667
305	6	7	22	s	Non-mix	86	0.03345	0.03292	0.03248	0.03195	0.02493	0.03662	0.0927	0.126	0.13992	0.14952	0.1577	0.16355	0.16457	0.17105	2	0.0614	0.171050001	0.024925
306	6	7	25	s	Non-mix	86	0.03051	0.02988	0.02972	0.02968	0.02309	0.03532	0.09032	0.13049	0.14837	0.15598	0.16711	0.17283	0.17354	0.17834	2	0.06655	0.178344443	0.02308889
307	6	7	27	s	Non-mix	86	0.04104	0.03979	0.03833	0.03838	0.02911	0.04365	0.12594	0.18786	0.21255	0.224	0.23542	0.2424	0.24423	0.24786	2	0.09566	0.247857895	0.02910526
308	6	7	33	s	Non-mix	86	0.04995	0.048	0.0468	0.04657	0.03502	0.05057	0.13423	0.18548	0.20313	0.21153	0.2206	0.2268	0.22623	0.22875	2	0.08476	0.228750003	0.03501667
309	6	7	40	s	Non-mix	86	0.03364	0.03285	0.03229	0.03236	0.02567	0.03727	0.10748	0.16614	0.18619	0.19458	0.20363	0.20782	0.20911	0.21187	2	0.08297	0.211872725	0.02567273
310	6	8	10	s	Non-mix	86	0.044	0.0425	0.0434	0.0425	0.034	0.0503	0.1286	0.1772	0.1987	0.2094	0.2224	0.2307	0.2311	0.2358	2	0.08716	0.235799998	0.034
311	6	8	13	s	Non-mix	86	0.03241	0.03149	0.03081	0.03065	0.02486	0.03592	0.09194	0.12843	0.14675	0.15902	0.17315	0.18324	0.18583	0.19118	2	0.06959	0.19118	0.02486
312	6	8	17	s	Non-mix	86	0.02847	0.0278	0.02805	0.0282	0.02365	0.0339	0.08187	0.1112	0.12737	0.1408	0.1555	0.16612	0.16783	0.17422	2	0.06233	0.174224999	0.02365
313	6	8	19	s	Non-mix	86	0.0304	0.0304	0.0305	0.0302	0.0239	0.0371	0.0835	0.1104	0.1278	0.1389	0.1549	0.165	0.1664	0.174	2	0.06093	0.173999995	0.0239
314	6	8	21	s	Non-mix	86	0.03488	0.0342	0.03345	0.03325	0.02632	0.03867	0.10111	0.14272	0.16079	0.17014	0.18151	0.18816	0.19025	0.19373	2	0.07192	0.193730001	0.02632
315	6	8	25	s	Non-mix	86	0.03278	0.03175	0.03102	0.03116	0.02398	0.03635	0.09529	0.13212	0.14915	0.15893	0.17243	0.18065	0.18314	0.18766	2	0.06886	0.187663634	0.02398182
316	6	8	28	s	Non-mix	86	0.03188	0.03123	0.03062	0.03035	0.02405	0.03587	0.08853	0.11863	0.13533	0.14482	0.1584	0.1669	0.16937	0.17282	2	0.06225	0.172816664	0.02405
317	6	8	32	s	Non-mix	86	0.03486	0.03439	0.03371	0.0335	0.02719	0.03841	0.09721	0.13609	0.15371	0.16379	0.17378	0.18116	0.18356	0.18568	2	0.0683	0.185678571	0.02719286
318	6	8	37	s	Non-mix	86	0.02192	0.02125	0.02117	0.02125	0.01815	0.0255	0.06215	0.08558	0.0998	0.1137	0.12955	0.14052	0.14468	0.1522	2	0.05381	0.152200002	0.01815
319	6	8	39	s	Non-mix	86	0.02717	0.02665	0.02602	0.02605	0.0211	0.03168	0.07422	0.10137	0.11603	0.12682	0.14045	0.14885	0.15148	0.15588	2	0.05568	0.155875001	0.0211

APPENDIX E

Confusion Matrix reduced metrics of model A and model B

Model A: SVM (C= 45; gamma =0,01; number of components = 5)

	TPR	TNR	PPV	NPV	FPR	FNR	FDR	ACC
Non-Disease	0.97	0.95	0.85	0.99	0.05	0.03	0.15	0.95
Up to 7%	0.65	0.88	0.82	0.75	0.12	0.35	0.18	0.78
Above 7%	0.83	0.83	0.69	0.91	0.17	0.18	0.31	0.83
Avg/total	0.81	0.89	0.79	0.88	0.11	0.19	0.21	0.85

Model B: SVM (C= 35; gamma =0,02; number of components = 4)

	TPR	TNR	PPV	NPV	FPR	FNR	FDR	ACC
Non-Disease	0.97	0.94	0.82	0.99	0.06	0.03	0.18	0.94
Up to 7%	0.74	0.88	0.84	0.80	0.12	0.26	0.16	0.82
Above 7%	0.83	0.90	0.79	0.92	0.10	0.18	0.21	0.87
Avg/total	0.84	0.91	0.82	0.90	0.09	0.16	0.18	0.88

Abbreviations

FP = False Positive
FN= False Negative
TP= True Positive
TN= True Negative
TPR = TP/(TP+FN) (Sensitivity, hit rate, recall, or true positive rate)
TNR = TN/(TN+FP) (Specificity or true negative rate)
PPV = TP/(TP+FP) (Precision or positive predictive value)
NPV = TN/(TN+FN) (Negative predictive value)
FPR = FP/(FP+TN) (Fall out or false positive rate)
FNR = FN/(TP+FN) (False negative rate)
ACC = (TP+TN)/(TP+FP+FN+TN) (Overall accuracy)