Geo-information Science and Remote Sensing

Thesis Report GIRS-2018-51

Emulating vision: an object-based image analysis approach to trait retrieval of potato plants

Jasper Siebring

November 19, 2018

WAGENINGEN
UNIVERSITY & RESEARCH

Emulating vision: an object-based image analysis approach to trait retrieval of potato plants

Jasper Siebring
Registration number 920110760010

Supervisors:
Lammert Kooistra (Wageningen University & Research)
Joao Pereira Valente (Wageningen University & Research)

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre, The Netherlands.

19th of November, 2018
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2018 -51
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Abstract

There is a growing demand in both food quality and quantity, but as of now, one-third of all food produced for human consumption is lost with pests and other pathogens accounting for roughly 40% of pre-harvest loss in potatoes. Pathogens in potato plants, like the Erwinia bacteria and the PVY^{NTN} virus for example, exhibit symptoms of varying severity that are not easily captured by pixel-based classes (as these ignore shape, texture, and context in general). The aim of this research is to develop an object-based image analysis (OBIA) method for trait retrieval of individual potato plants that maximizes information output from UAV-based RGB VHR imagery and its derivatives, to be used for disease detection of the *Solanum tuberosum*. The approach proposed can be decomposed in two steps: object-based approximation of potato plants using an optimized implementation of Large Scale Mean-Shift Segmentation (LSMSS), and classification of disease within said approximations using a set morphological features computed from their associative objects. The proposed approach was proven to be viable as the associative Random Forest model detected presence of the Erwinia bacteria and potato virus Y with a maximum F1 score of 0.75 and an average MCC score of 0.47.

Acknowledgements

This thesis has been quite the journey, one that allowed me to fully explore my interests in object-based image analysis, machine learning, and GEO-IT, whilst adhering to the environmental values Wageningen University & Research is known for throughout the world. The notion of object-based image analysis, or the bundling of pixels on some varying similarity, felt like a organic response to the ever-growing complexity of remote sensing imagery, and I am thankful that I was given the opportunity to contribute to its associative research. I would like to extend specific gratitude to my supervisors Lammert Kooistra and Joao Pereira Valente for their competent feedback, incredible patience, and mental support throughout.

Contents

1. Introduction 1
   1.1 General background 1
      1.1.1 Detecting pathogens with remote sensing 1
      1.1.2 Object-based image analysis 2
      1.1.3 Spectral depiction in remote sensing 2
   1.2 Problem definition 3
   1.3 Research objectives 4
   1.4 Reading guide 4
2. Related search 5
   2.1 Potato crop 5
      2.1.1 Origin and spread of the European potato 5
      2.1.2 General cycle and characteristics of the common potato plant 5
      2.1.3 Erwinia 7
      2.1.4 Potato Virus Y 7
      2.1.5 Summary and feature translation 8
   2.2 Object-based image analysis 10
      2.2.1 Rationale for object-based image analysis 10
      2.2.2 Principles of object-based image analysis 11
      2.2.3 Large-Scale Mean-Shift Segmentation 12
   2.3 Color and color transformations 14
      2.3.1 Color and color spaces 14
      2.3.2 Color distribution of objects 15
      2.3.3 Overview of color transformations 16
3. Methodology and data 17
   3.1 Study area and data 17
   3.2 Methodology 18
      3.2.1 Sample selection and manual segmentation 20
      3.2.2 Instantiating class data 20
      3.2.3 Evaluating model separability and generalizability 21
      3.2.4 Instantiating LSMSS input imagery 23
      3.2.5 Large Scale Mean-shift Segmentation and Naive Bayesian classification 24
      3.2.6 Feature extraction from class objects 25
      3.2.8 Random forest classification 27
4. Results 28
   4.1 Class polygons and class data 28

4.2 Separability and generalizability....30
4.3 LSMSS parametrization, instancing and evaluation....32
4.4 Expanding and evaluation of class objects....34
4.5 Random Forest Classification....36
5. Discussion....38
5.1 Modelling the human vision system....38
5.2 Data quality and usage....39
5.3 Method modifications and comparison....40
5.4 Modelling and classification....41
6. Conclusions and recommendations....42
Bibliography....44
Appendices....49
A. Separability comparison....49
B. Accuracy comparison....50
C. Python scripts....51
D. Detailed imagery....51

List of Abbreviations

GEOBIA	Geometric Object-Based Image Analysis
HVS	Human visible system
RS	Remote Sensing
RF	Random Forest
VHR	Very high resolution
VIS/NIR	Visible light/near infrared
FAO	Food and Agriculture Organization of the United Nations
UAV	Unmanned aerial vehicle
DSM	Digital Surface Model
DTM	Digital Terrain Model
RTK	Real-time kinematic
HLS	Hue, lightness, saturation
HSV	Hue, saturation, value
YUV	Luma (Y), chrominance(UV)
KDE	Kernel Density Estimation
GLCM	Gray level co-occurrence matrix
MCC	Matthew Correlation coefficient

1. Introduction

1.1 General background

There is a growing demand in both food quality and quantity (Savary et al., 2012). Not only is the world's population expected to grow to almost 10 billion by 2050, a parallel income growth in low- and middle-income countries is also hastening a dietary transition towards higher consumption of meat, fruits and vegetables. The FAO has projected that agriculture will have to produce almost 50 percent more food globally than it did in 2012 to meet this demand (Liu, 2017). Similar leaps in agricultural productivity have occurred but have come at heavy costs to the natural environment with a severe loss of biodiversity, rapid depletion of natural resources and an increase of global greenhouse gas emissions that ultimately contributes to global warming, the spread of transboundary pests, deforestation and desertification (FAO, 2011; Liu, 2017). Subsequently, maintaining the pace of production increases via high-input, resource-intensive farming systems may be more difficult than in the past (Liu, 2017). As of now, one-third of all food produced for human consumption is currently lost with pests and other pathogens accounting for roughly 40% of pre-harvest loss in potatoes (Savary et al., 2012). Reducing this loss would grossly lessen the need for production increases, making early and accurate detection of these pathogens (and subsequent management) a key factor in securing global food security (Savary et al., 2012).

1.1.1 Detecting pathogens with remote sensing

Currently there are several promising proximal sensor-based methods in early development that (indirectly) detect the spread of crop diseases, done by measuring either reflectance, temperature or fluorescence of the respective plant (Sankaran et al., 2010). Spectral signatures in particular clearly show certain chemical and morphological characteristics of vegetation that are indicative of disease (Xue & Su, 2017). These signatures are generally explored in a very 'pixel-centric' manner using either the spectral domain (as a spectral profile per pixel), the image domain (collective pixels representing one geometric scene) or in feature space (pixels displayed as points in n-dimensional space) (Landgrebe, 1999). There is an inherent trade-off here between spectral and spatial resolution, making subtle spatial patterns exclusive to imagery of coarse spectral resolution (and vice versa). The aforementioned practices are in support of visual assessment by experts, which is still considered common practice but also very time consuming and increases the chance of damage to the crops (Mahlein, 2016). Less-invasive sensor-based methods are found in the field of remote sensing, a widely documented and established field of science that acquires information about phenomena without any physical contact making it particularly useful for non-destructive and rapid characterization of vegetation (Li et al., 2014). Recent technical advancements in sensor sensitivity, material weight and costs, computational capacity and telemetry have enabled very-high-resolution (VHR) remote sensing using Unmanned Aerial Vehicles (UAVs) (Benediktsson et al., 2013). This higher resolution consequently not only refers to the increased spatial resolution, i.e. the pixel size, but also the spectral and temporal resolutions that enable more detailed and frequent acquisition, resolutions not possible with proximal sensing.

1.1.2 Object-based image analysis

Contrasting the aforementioned pixel-based approach, a more object-based approach came to the forefront around the year 2000 when the first commercial software packages started appearing that specifically delineated and analyzed *image-objects* instead of individual pixels. This area of research was referred to as *Object-Based Image Analysis* (OBIA) and was later expanded to *Geometric Object-Based Image Analysis* (GEOBIA) for earth surface-specific imagery as geographic space is considered intrinsic to these images (Gartner et al., 2010). The aforementioned *image-objects* represent distinguishable components of an image (a forest, a tree, a leaf, etc.) and are essentially pixels that are bundled due to varying levels of similarity or patterns in their spectral characteristics, shape or topology (not unlike our cognitive ability to recognize objects) (Gartner et al., 2010a; R. M. Haralick & Shapiro, 1985).

1.1.3 Spectral depiction in remote sensing

The aforementioned field of remote sensing employs systems that measure the reflected or emitted electromagnetic radiation in pre-determined parts of the electromagnetic spectrum, not unlike the human visual system (HVS). This spectral information is subsequently saved in an array of numbers, called *image* or *scene*, with the individual numbers called *pixels* (i.e. picture elements) arranged in a format more befitting its use case. Images depicting visible light, for example, are commonly encoded in the RGB color space, which is only one of many mathematical models that allow for reproducible definitions of 'color'. The concept of color herein is not an intrinsic property of any particular *scene*, but merely a perception of light limited by the HVS (Pascale, 2003). As color spaces allow us to change this definition, it has been employed in various use cases ranging from more intuitive color representations (e.g. XYZ, HSV) to emphasizing specific spectral responses (e.g. I1I2I3 for facial recognition), akin to spectral indices (García-Mateos et al., 2015).

1.2 Problem definition

The aforementioned leap in spatial, spectral and temporal resolution results in an increase of pixel variability (and ultimately overall data volume, complexity and required computing resources). This poses some challenges for image classification as opposed to the relatively trivial procedure of classifying more homogeneous pixel ranges that come with coarser resolutions (Blaschke et al., 2014). Pathogens in potato plants, like the Erwinia bacteria and the PVY^{NTN} virus for example, exhibit symptoms of varying severity that are not easily captured by pixel-based classes (as these ignore shape, color variability, and context in general). This means that the coinciding increase in *within-class* variance (multiple pixels detailing one plant) actually decreases the potential accuracy of purely pixel-based classification, this conflict is coined the H-resolution problem by Hay et al. (1996).

This is not to say that pixel-based disease detection is not possible, indicators in both the spectral domain and feature space have competently been exploited to do just that (see Kamp et al., 2016, and Mahlein, 2016). It is just that as UAV-based VHR imagery reaches a spatial resolution that effectively matches that of the objects of interest, e.g. subtle plant disease traits, arguments have been made that image analysis in these situations should forgo this pixel-based approach in favor of a more object-based approach that could potentially better capture these subtle classes (Blaschke et al., 2014; Gartner et al., 2010a; Weih & Riggan, 2010).

Additionally, imagery of said spatial resolution are still bound to lower spectral resolution which excludes several robust indices known to help segment, classify or otherwise characterize host vegetation (Xue & Su, 2017). Some RGB-counterparts to these indices do exist (Hamuda, Glavin, & Jones, 2016) and perform moderately well given homogenous backgrounds and dominantly green plant colors green, but these conditions are increasingly difficult to meet as spatial resolution increases (Hamuda et al., 2016; Michel, Youssefi, & Grizonnet, 2015).

Object-based classes via GEOBIA, incorporating morphological properties like shape and texture, might be able to capture more context-sensitive information and exploit the higher information content in VHR orthophotos, similar to manual assessment by experts in the field, which would allow for a more scalable disease detection method, ultimately enabling more site-specific crop management (Haralick et al., 1985; Mahlein, 2016). However as is the same for pixel-based classes (e.g. land covers in land cover classification), image objects are a matter of semantics and have to be regarded as a user-driven set of conditions (Blaschke et al., 2014). This is made all the more difficult by the fact that the conditions in this case are the spectral and optical symptoms of pathogens which are expected to vary greatly in scale, severity and visibility not only per pathogen but also per host, which in turn are influenced by environmental conditions (Sankaran etal., 2010).

1.3 Research objectives

The aim of this research is to develop an object-based method for trait retrieval of individual potato plants that maximizes information output from RGB VHR orthophotos and its derivatives, to be used for disease detection in sick or diseased plants. This is done in an effort to emulate the HVS of plant experts where they (initially) asses the health status of plants using only visually apparent traits, characterizing them through texture, shape, and other morphological features. This does require a thorough understanding of how, where and when specific viral or bacterial pathogens express themselves in potato plants and what morphological features could capture the respective symptoms. The aforementioned Erwinia bacteria and Potato virus Y (PVY^{NTN} strain) will hereby be selected as topic of research as they are currently one of the more prevalent pathogens affecting potato production (Kamp et al., 2016).

Trait retrieval is set to be performed on image objects that approximate individual plants, following an application of GEOBIA on VHR imagery depicting potato crops. The underlying segmentation and classification parts that characterize a typical GEOBIA workflow (Gartner, Meng, & Peterson, 2010b) requires the classes under consideration (e.g. plant components exhibiting disease traits) to be separable, either visually or in feature space. Accounting for the subtlety of the symptoms and also the possible ambiguity between the classes due to a limited spectral resolution, we will explore the application of color transformations following García-Mateos et al. (2015) and Hamuda et al. (2016).

Using the object-based approximations of healthy and diseased potato plans, we will compute various morphological features said to relate to pathogen traits and employ them in the modelling of disease in potato plants following Feng et al. (2015) and Barbedo (2013). In this exploratory part of the research, we will first determine the validity of this approach (i.e. disease detection without hyperspectral imagery) before quantifying what features (or composition thereof) are most important in the delineation of disease in potato plants.

As such, the following research questions will be addressed:

1. *What are the morphological signs of Erwinia and PVY^{NTN} infection in potato plants?*
2. *What feature space transformations of UAV-based orthophotos (RGB) can improve the separability of classes found in potato crop site scenes?*
3. *What morphological features, derived from object-based approximations of potato plant, best detect the presence and severity of Erwinia or PVYNTN infection?*

1.4 Reading guide

The next chapter (2) gives a theoretical background to the research in this thesis. The underlying literature review for the first section of this chapter focuses on the morphology of the potato plants found in the research area and the aforementioned pathogens that infect (some of) these plants. The second section briefly discusses the rationale for OBIA in the context of remote sensing imagery, and how it works conceptually, before detailing one notable implementation. The third section provides a small overview of color transformations relevant to crop imagery. Having given some necessary context, we will then detail our exhaustive methodology in steps before discussing their associative results.

2. Related search

2.1 Potato crop

This chapter will briefly describe the origin and initial spread of the common European potato crop before giving a generic description of its life cycle, and the visually apparent effects that two major potato pathogens (also found in the research area) can have on said life cycle, and thus their hosts. As implied, the emphasis throughout this chapter will be on finding *texture*, *color* and otherwise morphological traits that are highly descriptive of disease in the respective plant, giving us a sense of what plant experts are (initially) looking for during manual plant health assessment.

2.1.1 Origin and spread of the European potato

The plant that ultimately usurped the name 'potato' was a solanaceous (flowering) plant going by the Latin name *Solanum tuberosum esculentum,* given by Bauhin in 1596, a name essentially retained in the binomial Solanum tuberosum given by Linnaeous in his Species Plantarum (1753). Solanum tuberosum L. was[1] a wild potato plant from the mountainous regions of South America, said to have been cultivated by the Inca Indians around 200 B.C. from which it was introduced to Spain and subsequently Central Europe where it is now a staple food crop (Rich, 1983). As shown by its spread, the geographic range of the potato is almost worldwide which helped make it a major source of food in most countries. An estimated 300 million metric tons of potato is produced globally to feed more than a billion people worldwide, making it a critical crop for international food security, after maize, rice and wheat (Celis-Gamboa, 2002).

The popularity of the potato as a food crop can largely be attributed to the relatively high food value per unit of time, land and water, its natural resistance to environmental stresses, the balanced contents of calories and proteins in the tubers, and the uniquely high agronomical plasticity (Burton, 1989; Horton 1988). The latter has enabled breeding of new potato plants that better fit local needs (e.g. frost resistance or a particular taste) and has resulted in a wide variety of potato genotypes (Datiles & Acevedo-Rodríguez, 2014). These genotypes vary not only in morphological features (e.g. tuber shape and color), but also in their time of tuber initiation, percentage of dry matter, the rate of bulking, length of their growing season, response to environmental conditions and their degree of resistance to the many diseases which may infect the crop (Burton, 1989).

2.1.2 General cycle and characteristics of the common potato plant

The common potato plant is a herbaceous annual plant that can grow up to 0.4 - 1.4 meters tall and ranges from erect to fully prostrate (Canadian Food Inspection Agency, 2015). Development of the common potato starts from either a *mother* tuber or *true* seed, planted roughly 50 cm under topsoil.

[1] This original potato variety no longer exists but the Solanum genus has grown to include over 2000 'new' members including the current day common potato (*Solanum tuberosum*) and tomato (*Solanum lycopersicum*)

Crop potatoes are generally propagated vegetatively using mother tubers as these produce exact genetic clones of the mother plant, which is useful for mass production. Potato propagation with *true* seeds is generally reserved for plant breeding as these can produce offspring with different characteristics, facilitating hybridization. One or more main plant stems will originate from this mother tuber (*true* seed) that branch depending on the potato variety, physiological age of the tuber and environmental conditions. These exhibit varying hairiness and can be green, purple or a combination of the two. Branching generally terminates in an inflorescence (the flower part of the plant) that can also *set* fruit (green berries with *true* seeds) roughly 1-4 cm in diameter and spherical in shape.

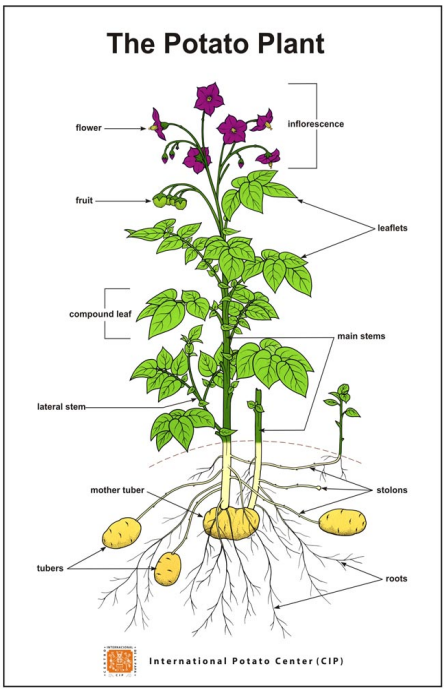


Figure 1 *Conceptual potato plant model*

The inflorescences contain up to 25 pentamerous flowers that are each 3-4 cm in diameter but can vary widely in color, ranging from white, lilac, pink, blue and purple (Agency, 2015; see figure 1). Vegetative growth continues through the development of axillary buds below said inflorescence, that will grow into a second-order stem that can also produce leaves and an inflorescence even allowing for a tertiary stem (Celis-Gamboa, 2002).

These branching habits are highly descriptive as they determine the total area of leaf area, life cycle duration and the total number and size of the tubers, flowers and seeds. The leafs themselves are pinnate with a single terminal leaflet and three or four pairs of ovoid leaflets with smaller ones in between. The leafs can range from 8-22 x 5-13 cm, exhibit varying hairiness akin to the stems, and are generally medium to dark green.

With the emergence of vegetation that allows for photosynthesis, the mother tuber is no longer the sole source of energy and concurrently initiates the stolons, i.e. the underground offshoots from the main stem that culminates in tuber formation. These are characterized by their long internodes that end in a ‘hook-like’ tip and, like the other stems, also branch and elongate under certain environmental conditions. Growth of this stolon network stops at some point, called cessation, which is when swelling occurs under the ‘hook’ that ultimately grows into tubers (the deposition of starch and protein). Concurrently, vegetation of most genotypes begin to perish some time after tuber initiation with the yellowing of the leaves (reduction of photosynthesis) and the hardening of the tuber skin that act as containers while the plant undergoes a process of senescence (Celis-Gamboa, 2002). The actual timespan of these processes vary considerably per genotype with some cultivars exhibiting the premature dropping of floral buds, male sterility, or a complete inability to set fruit (Agency, 2015).

Figure 2 From left to right: two Erwinia symptoms including soft rot potatoes and black leg symptom (see 2.1.3), and two PVY symptoms including stunting, necrosis, desiccation, necrosis streaks, and blotting (see 2.1.4),

2.1.3 Erwinia

First described in 1917 by Pérombelon, Erwinia is a genus that encapsulates all members of the *Enterobacteriaceae* family (I. K. Toth et al., 2011). Throughout the years however, Erwinia has seen numerous reclassifications due to nomenclatural difficulties (e.g. *E. stewartii* to *Pantoea stewartii*) and one big offshoot into a genus now known as Dickeya (Toth et al., 2011). We will be following the consensus that Erwinia still generally refers to soft rot pathogens including *Pectobacterium carotovora* ssp. *Atroseptica* (Eca), *P. carotovora* ssp. *Carotovora* (Ecc), *P. chrysanthemi (*Ech) and *Dickeya solani*. We are aware that current taxonomy technically sees them as separate genera but their subtle differences (e.g. host ranges, composition) do not express in ways relevant to our research, i.e. detectable by the naked eye, which is why we will be following our definition (for more information on taxonomy see: Graham, 1964, Dye, 1969, Hauben et al. 1998, Samsom et al., 2005).

As their name implies, soft rot pathogens are mainly known for causing tuber rot by producing high levels of exoenzymes that degrade cell walls and tissue (Toth et al., 2003). Some of these pathogens however also cause *black leg,* which are rot lesions spreading from the rotting mother tuber up to the stems (Pérombelon, 2002). This will subsequently cause growth stunting, yellowing and wilting of the leaves, and desiccation of stems and leaves (dependent on wet conditions). These pathogens infect host plants through natural pores or wounds (e.g. insect bites, mechanical damage) and can thus be spread through insects, surface water and even aerosols (des Essarts et al., 2016). After initial infection, the pathogens spread through the vascular vessels where they will remain until pathogen-specific environmental conditions (e.g. moisture, oxygen, temperature) allow for further development (Toth et al., 2003).

2.1.4 Potato Virus Y

First reported in 1913, Potato virus Y (PVY) is the type member of the genus Potyvirus with a host range limited to the Solanaceae (Singh et al., 2008) and is known as one of the most damaging of potato viruses (10-50% yield losses by quality degradation; worse if combined with other pathogens). The main source of PVY pathogens are infected seed tubers, where they can remain dormant until more favorable environmental changes occur that allow for development. The virus is then commonly (and continuously) spread through aphid vectors via non-persistent transmission (called primary infection), or occasionally via mechanical

transmission (or sap transmission) which requires a wound and direct contact between infected and healthy plants (Nolte, Alvarez, & Whitworth, 2002). The former transmission is referred to as non-persistent due to the relatively short period of time where the virus stays 'viable' (i.e. 2 hours). The virus is acquired by aphids when they are feeding on infected plants, where the virus particles attach to their mouth parts (or *stylets*). The now contaminated aphids can subsequently transmit the virus by feeding on healthy plants, effectively within seconds. This is why PVY can spread very rapidly, and if aphid populations are large enough, also very extensively (Nolte et al., 2002).

The severity of the disease symptoms depends on the PVY strain involved, host tolerance, time of infection, method of infection, and environmental factors, and consequently exhibit considerable variance in symptoms per strain. The common strains (PVY^{O}) causes discrete mottling and green mosaics on young leaflets, the PVY^{N} strain causes tobacco veinal necrosis (brown streaks across leaf veins), the PVY^{NTN} strain causes necrotic ringspots on tubers, and PVY^{C} strain causes stipple-streak in the leaves (Blancard et al., 2012). Most of these symptoms (and thus traits; see figure 2) initially express as discoloration in the leaves but will ultimately lead to leaf dropping due to necrosis and leaf wrinkling (palm tree look).

2.1.5 Summary and feature translation

Summarizing the last few paragraphs, there are several morphological traits of potato plants that hold information regarding their health (or lack thereof) and are technically visually apparent. We will use the concepts of morphology (the study of form and structures; specifically *shape*, *structure*, *color*, *texture* and *size*), and spatial and temporal detail (inspired by Strahler et al., 1986) to enable some basic categorization of said traits.

The majority of the pathogen traits express themselves as discoloration on the leaf level, which is ultimately a function of *color* and *texture* (at different scales). This is because necrosis, mosaics, and stipple-streak all affect the appearance, and consistency of the leaf surface (and collectively of the canopy) by introducing new patterns or colors (or by deteriorating existing ones). These symptoms can be quite subtle (especially necrosis at early stages) thus requiring considerable spatial detail. Once detected however, their effect are immediate apparent (i.e. disease presence) meaning temporal detail can be very low. Mottling and desiccation however affect the feel (or structure) of the surface, technically making it a function of both *texture* and *structure*. Both class of descriptors require some sense of neighborhood, inherently necessitating high spatial detail in multiple scales, but mottling and desiccation themselves only require low temporal detail as their presence alone has immediate repercussions.

Over time, disease-stricken potato plants (usually by Erwinia) will exhibit growth stunting subsequently affecting a myriad of plant components, and as such can be seen as a function of *shape*, *texture*, *structure*, and *size.* The 'blackening' of the stems is one such result (or rather origin) of growth stunting which could be captured by *texture* features but would require exorbitant spatial detail (or rather spectral) to distinguish these components.

Similarly, branching habits of potato plants would be highly descriptive (as functions of *structure)* but would necessitate exorbitant both spatial and temporal detail to evaluate growth trends over time. Morphological descriptors of *shape* and *size* could however exploit simple measurements like area or volume as proxies to assess growth rates (which would not require high spatial detail but is obviously benefited by higher temporal detail or some set growth norm). Similarly, *texture* and *structure* features of the canopy could evaluate whether stunting (or local/complete collapse) has occurred and would only require coarse spatial and temporal detail (as the scale is big).

It is worth emphasizing that even healthy potato plants will begin to perish some time after tuber initiation as it is 'only' an annual crop, the exact time varies by species. With this in mind, we hereby also mention the morphological features used in potato plant taxonomy following the Canadian Food Inspection Agency (2015). They distinguish potato plants by their (1) foliage, (2) inflorescence, (3) date of maturity, (4) tubers, and (5) sprout, of which the latter two cannot be approximated by visually apparent morphological features. The presence and state of inflorescences themselves do not seem to hold much information regarding disease status, only as a proxy for growth as it can indicate maturity by tuber initiation. The specific features (and their requirements) for both taxonomy and disease presence can be found in table 1 following this literature research.

Table 1 Feature list of disease-stricken potato plants and general taxonomy

FEATURE	LOCATION	INDICATOR	SPATIAL	TEMPORAL	DESCRIPTOR
Venial necrosis	*Leaf*	*PVY*	*Medium*	*Low*	*texture/color*
Stipple-streak	*Leaf*	*PVY*	*Medium*	*Low*	*texture/color*
Mottling	*Leaf*	*PVY*	*Medium*	*Low*	*texture/color /structure*
Mosaic	*Leaf*	*PVY*	*Medium*	*Low*	*texture/color*
Wilting	*Canopy*	*Erwinia*	*Low*	*Medium*	*shape/size /structure*
Desiccation	*Leaf*	*Erwinia*	*Medium*	*Low*	*texture/structure*
Black-leg	*Stem*	*Erwinia*	*High*	*Low*	*texture/color/ structure*
Yellowing	*Leaf*	*Erwinia*	*Medium*	*Low*	*color*
Stunting	*All*	*Erwinia/ PVY*	*Low*	*Medium*	*shape/structure /color/texture/size*
Branching	*Stem*	*Taxonomy*	*High*	*High*	*shape/size/structure*
Biomass	*Canopy*	*Taxonomy*	*Low*	*Low*	*shape/size*
Inflorescence	*Canopy*	*Taxonomy*	*Medium*	*Low*	*color*
Fruit	*Canopy*	*Taxonomy*	*High*	*Low*	*color*

2.2 Object-based image analysis

In this section we will briefly discuss the rationale for object-based image analysis (OBIA) in the context of remote sensing imagery and how it works conceptually, before detailing one notable implementation in Orfeo toolbox (an open-source software package offering remote sensing solutions). We would first like to acknowledge the vast body of work that competently explores the history of OBIA, relevant theoretical underpinnings, its nomenclature, several applications in remote sensing, and its advantage to pixel-based approaches (see Blaschke, 2010; Blaschke, 2014; Gartner et al., 2010a; Maillot et al., 2004; Weih et al., 2010). Particularly relevant examples include the arguments made by Blaschke (2014) for the acknowledgment of object-based image analysis of earth surface-specific imagery as its own paradigm, to be referred to as geometric object-based image analysis (GEOBIA). Whilst we subscribe to these arguments, we will be using the terms OBIA/GEOBIA interchangeably for the sake of simplicity, and similarly, will only draw from the aforementioned sources to convey the rationale and basic principles behind OBIA (allowing us to focus on one specific implementation).

2.2.1 Rationale for object-based image analysis

Remote sensing systems measure the reflected or emitted electromagnetic radiation in pre-determined parts of the electromagnetic spectrum, not unlike the human visual system (HVS). This spectral information is subsequently saved in an array of numbers, called *image* or *scene*, with the individual numbers called *pixels* (or picture elements) encoded in a format more interpretable to its (human) users (see section 2.3). Objects under consideration in these scenes (e.g. a house, tree or a river in a 1:3000 scale color orthophoto) are thus composed of individual pixels, forming essentially modeled representations of real-world objects (Strahler,. 1986). Depending on the object(s) under consideration (or classes), as these are entirely scale-dependent, an increase in spatial resolution (detail) leads to a coinciding increase in *within-class* variance as the objects would be composed of more pixels (containing brightness levels or 'color', see figure 3). This will ultimately decrease the potential accuracy of any purely pixel-based classification, especially for complex heterogeneous classes (see Weng, 2009, Marceau et al., 1990; Hay et al., 1994).

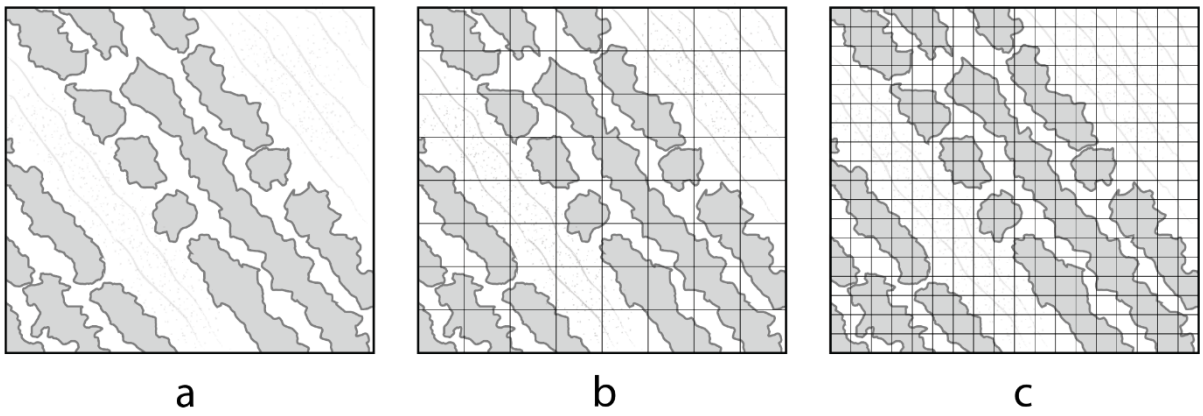


Figure 3 Illustration of the relationship between objects under consideration and spatial resolution: (a) low resolution: pixels significantly larger than objects, sub-pixel techniques needed (b) medium resolution: pixel and objects sizes are roughly of the same size, pixel-by-pixel techniques are appropriate (c) high resolution: pixels are significantly smaller than objects, regionalization of pixels into groups of pixels and finally objects is needed (see Blaschke, 2010)

As mentioned in section 1.1.1, numerous technical developments in remote sensing have made cheap (if not free) and readily available very high resolution (VHR) imagery commonplace, in which more of these complex classes can be (visually) distinguished (Gartner et al., 2010a). This steady climb in spatial resolution for remote sensing imagery, and the subsequent need for tools to deal with the added detail, parallels the development of proximal image analysis (preceding it by decades; Blaschke, 2010). From this older development cycle arose more advanced image processing tools that focused more on *image-texture* and *contextual* information, describing the association of neighboring pixel values via moving window or kernel methods, which was shown to improve classification results (Blaschke et al., 2014). Building on the collective of these concepts (e.g. segmentation, edge-detection, texture analysis), object-based methods were envisioned which allowed for the exploitation of scene characteristics other than color by bundling pixels on some varying similarity.

2.2.2 Principles of object-based image analysis

Summarizing a comprehensive documentation by Blaschke et al. (2014) and Gartner et al. (2010), object-based image analysis can now be considered a methodological framework for machine-based interpretation of complex classes within imagery, defined by their spectral, spatial, structural, and even hierarchical properties. In the context of remote sensing, these classes can be considered scene components that are distinguishable in an image (e.g. a house, tree or a river in a 1:3000 scale color orthophoto) using these properties.

The approximation of said classes by OBIA can be generalized in two (interrelated and iterative) steps: (1) image segmentation and (2) image classification. The former divides the image into meaningful homogenous segments using the aforementioned properties, whilst the latter classifies these segments to correspond to real-world objects (figure 4). The term *segmentation* refers to all procedures that iteratively build, modify, grow, cut or shrink objects until they represent nested or scaled representations of real world objects (or otherwise meaningful objects), whilst the term *classification* refers to all classification algorithms that use the information contained within each segment (e.g. *texture*, *shape*) to make this representation more explicit, employing their spectral and geometrical properties. This cyclical nature makes OBIA highly adaptive and open for the accommodation of different categories of target classes, from various domains, adhering to different semantics (Gartner et al., 2010a).

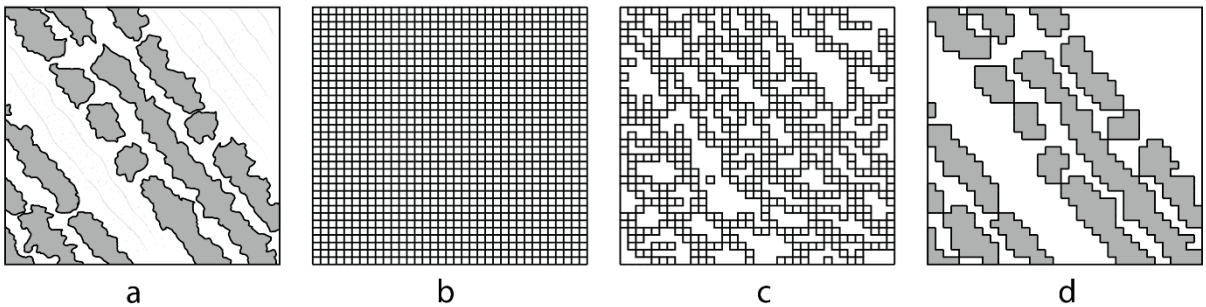


Figure 4 Abstraction of OBIA workflow: (a) initial scene containing very discrete and homogenous *vegetation* classes, (b) pixel-based representation of VHR scene, (c) one iteration of OBIA's bundling of pixels based on some varying similarity, (d) the resulting approximation of *vegetation* classes after *n* iterations and classification of classes found in *a* (colorized)

2.2.3 Large-Scale Mean-Shift Segmentation

Object-based image analysis has been implemented in various forms since its initial premiere as proprietary software package eCognition in 2000 (which arguably popularized the paradigm, see Blaschke, 2010; Knoth & Nüst, 2016). A growing number of these implementations consist of open source alternatives, reflecting the increasing interest of free and open-source software (FOSS) within the geospatial field (Clewley et al., 2014). The inherent plasticity of FOSS (due to various licenses, see FSF, 2014) and the cost savings on proprietary software licenses are common arguments listed to explain said popularity. With the increasing availability of VHR datasets (see section 1.1.1) and expansive computing resources through cloud computing platforms (e.g. Microsoft Azure, Amazon Web Services; see Clewley et al., 2014), these open source solutions would need to be scalable across these (or similar) services for them to be viable alternatives (both now and in the near future).

Clewley (2014) has provided an overview to a number of open source packages that can perform various parts of the generic OBIA process (see section 2.2.2) whilst adhering to the aforementioned scalability clause, some examples include InterIMAGE, SPRING, Ilwis, and Ilastik. We however would like to emphasize one notable OBIA implementation from said overview, namely Large Scale Mean-Shift Segmentation (LSMSS) as part of Orfeo Toolbox. This software package from the French Centre National d'Études Spatiales (CNES) provides a number of remote sensing solutions (e.g. SAR denoising, watershed segmentation) with intercompatibility to a myriad of other software packages and programming languages (QGIS, Monteverdi, Bash, Python, C++). LSMSS is notable as it guarantees stability regardless of scale and detail of the input imagery, and employs intuitive parameters (see Michel et al., 2015).

As the name implies, LSMSS is a segmentation tool that builds upon the mean-shift algorithm (see Zheng et al., 2009) which by itself is not a segmentation algorithm. It is a nonparametric method first introduced by Fukunaga and Hostetler in 1975 that estimates modes in a multivariate probability density function (Michel et al., 2015). This was later generalized by Sheng in 1995 who introduced a kernel function, and made spatially explicit by Commaniciu who applied the mode estimation to the joint spatial and spectral domain (Michel et al., 2015).

Given one VHR image as input, the process underlying LSMSS can be split up into 4 steps: filtering, clustering, merging (optional), and vectorization.

(1) The filtering step to the image is a simple iterative procedure that shifts each data point (pixel) to the weighted mean of data points in its neighborhood until it converges on the mode of the underlying probability density function (evaluated by *mode threshold*, i.e. the minimum distance from the current mode to the mean; can also be terminated by reaching a *maximum number of iterations*). The neighborhood in each iteration is determined by a *spectral* and *spatial range* with respect to the location and spectral values of the current mode, both expressed as Euclidean distances. The weights in each iteration are determined by the distance of the current mode estimation to all points in the neighborhood using a kernel density estimation function (LSMSS uses a uniform kernel). This procedure is applied to each pixel of the input image where the spectral (*n* dimensions, e.g. [*R,G,B*]) and spatial values ([*X,Y*]) of the estimated local mode are assigned to the same pixel in two separate output images, an filtered-image and an displacement-image.

(2) The clustering step largely rethreads the first step as it clusters all (neighboring) pixels together given the filtered-image (and optionally the displacement-image) using the same procedure (see figure 5) where it effectively relates pixel similarity to a set spectral and spatial neighborhood, outputting an label-image containing homogenous clusters of pixels that are considered to be similar. It is worth mentioning that because of initial spatial threshold and the various stopping criteria, each pixel only converges towards its own estimate of the mode which could differ slightly (even for neighboring pixels effectively belonging to the same *object*). As such, the label image could potentially have a lot of very small segments that do not correspond to any meaningful *objects* of the scene.

(3) The third (and optional) step iteratively merges the aforementioned 'orphan' segments in the label-image to the closest (spectrally; in terms of Euclidean distance) neighboring segments given a minimum region size (i.e. the Edison implementation of the Mean-Shift algorithm; see Michel et al., 2015), outputting a 'cleaned' label image. (4) The last step is the vectorization (i.e. conversion of raster graphics to vector graphics) of the segments found in the label image where they are assigned the mean and variance values of the pixels in the original input image, thus concluding segmentation (see Michel et al., 2015 for a complete outline of this procedure, including equations).

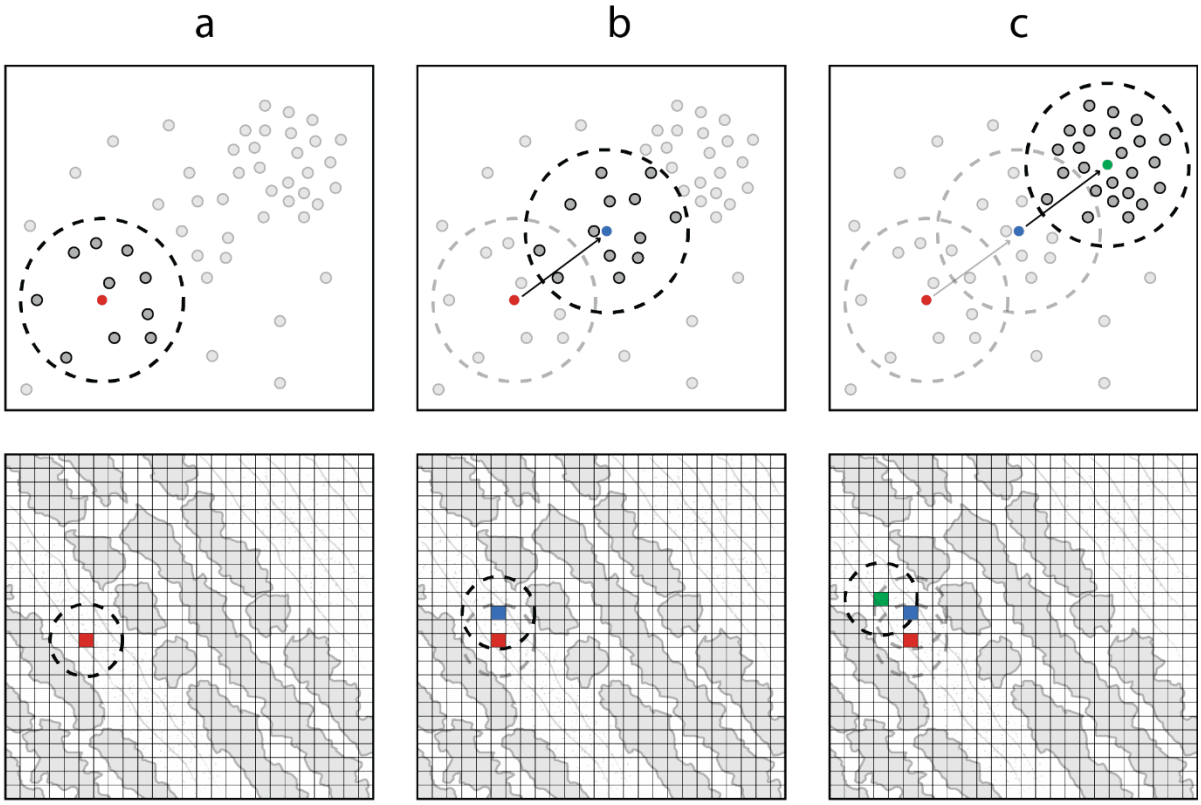


Figure 5 Two-dimensional mock-up of the mean-shift algorithm which is set to find the local mode for **every** pixel, effectively segmenting the image if performed on all pixels. Top row depicts feature space (associative circle represents spectral threshold), and the bottom row represents image space (associative circle represents spatial range): (a) The first iteration where the mean of a set pixels points to a new mode (blue), the set being composed of pixels that adhere to the spectral and spatial thresholds with respect to the initial pixel (red), (b) the second iteration where the mean of a new set (with respect to blue) is nearing another mode (green) , (c) the third iteration where the distance between the mean of the new set (with respect to green) and the local mode is close enough to indicate convergence (evaluated by *mode threshold*), which is when the algorithm is terminated and the red pixel is assigned the spectral values of green (its relative local mode)

2.3 Color and color transformations

As referenced in section 2.2.1, images are arrays of numbers containing spectral information that are *arranged* in formats best suited for their particular use-case. In this chapter we will provide an overview of such arrangements with use cases relevant to remote sensing, drawing from papers written by García-Mateos et al. (2015) and Hamuda et al. (2016). We will be highlighting arrangements shown to improve segmentation of vegetation, or that otherwise help characterize objects commonly found in VHR imagery (as befitting our own use case).

2.3.1 Color and color spaces

Following definitions in the aforementioned papers, the color of any particular pixel in an image can be observed as a stochastic event within the *n*-dimensional space defined by the color space used. Color spaces herein are mathematical models that allow for different (reproducible) definitions of 'color', which in of itself is not an intrinsic property of image object (thus pixels) but merely a perception of light limited by the HVS (Pascale, 2003).

As such, images produced for media consumption (i.e. information, entertainment) commonly only capture visible light as that reflects our own sensitivity to this particular section of the electromagnetic spectrum (specifically to red, green and blue; called trichromacy). If the particular use case for the images then equates to the consumption of media through some electronic medium, these images will commonly be encoded in the RGB color space as that offers easier compression (less data to be sent) and depiction across varying monitors (consistency). By employing color spaces, often using non-linear transformations from the initial RGB space, one can separate (and employ) various spectral characteristics in any one image (e.g. contrast and luminance as seen in figure 6).

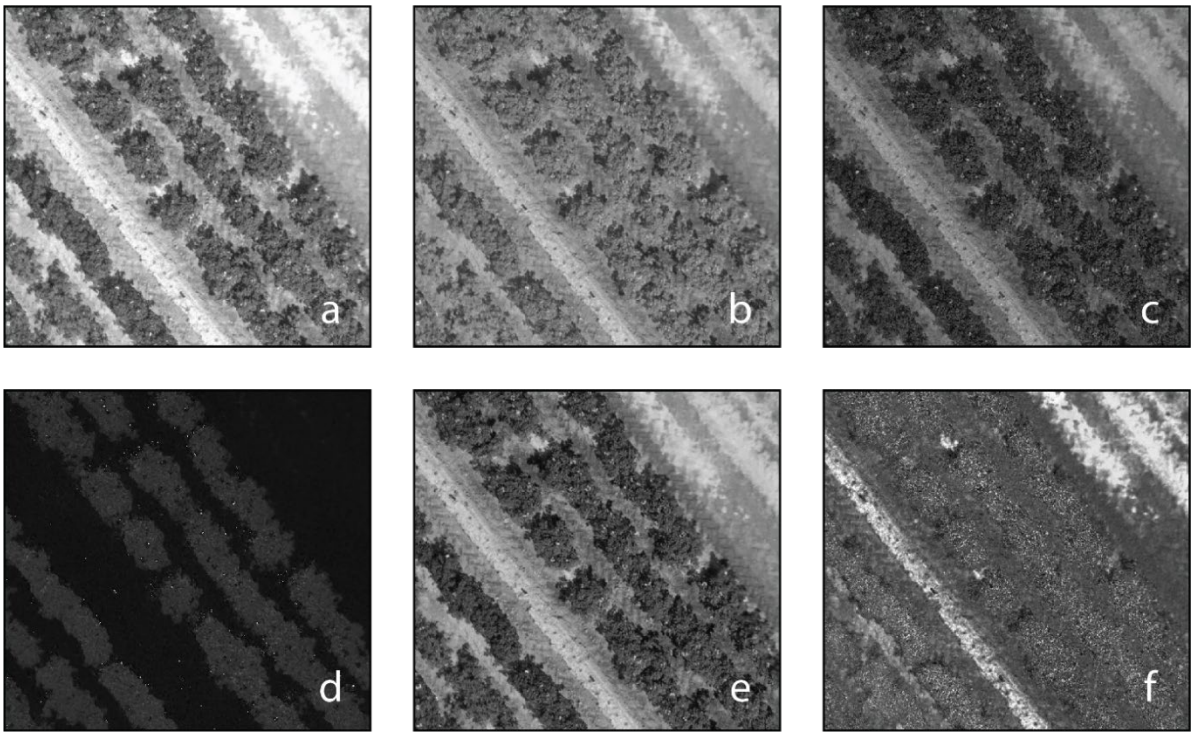


Figure 6 Six monochrome VHR images depicting a potato crop field: (a) red, (b) green, (c) blue, (d) hue, (e) value, (f) saturation. Observe how different facets of the image are emphasized (e.g. illumination in *e* and sheer color in *d*)

One notable example that exploits these characteristics is the color space HSV (figure 6), which expresses color by hue (color), saturation (amount of gray in said color), and value (the brightness or intensity of said color), by which it more closely resembles how humans perceive light, making for a more intuitive representation (Pascale, 2003). Other use cases of color spaces for remote sensing imagery include: compression/broadcasting (*YCbCr*), normalization (e.g. illumination; *normalized RGB*), enhancement (e.g. green tones; *excessive green*), uniformity (*L*u*v**), and segmentation using any appropriate color space that emphasizes the object(s) under consideration. The options regarding color spaces are largely dependent on the initial spectral range of the image (Hamuda et al., 2016). As the intuitive sense of color no longer holds up for hyperspectral imagery, transformations in these ranges commonly go by spectral indices (and are arguably its own paradigm due to the sheer difference in spatial and spectral detail, see section 1.1.1; García et al., 2015).

2.3.2 Color distribution of objects

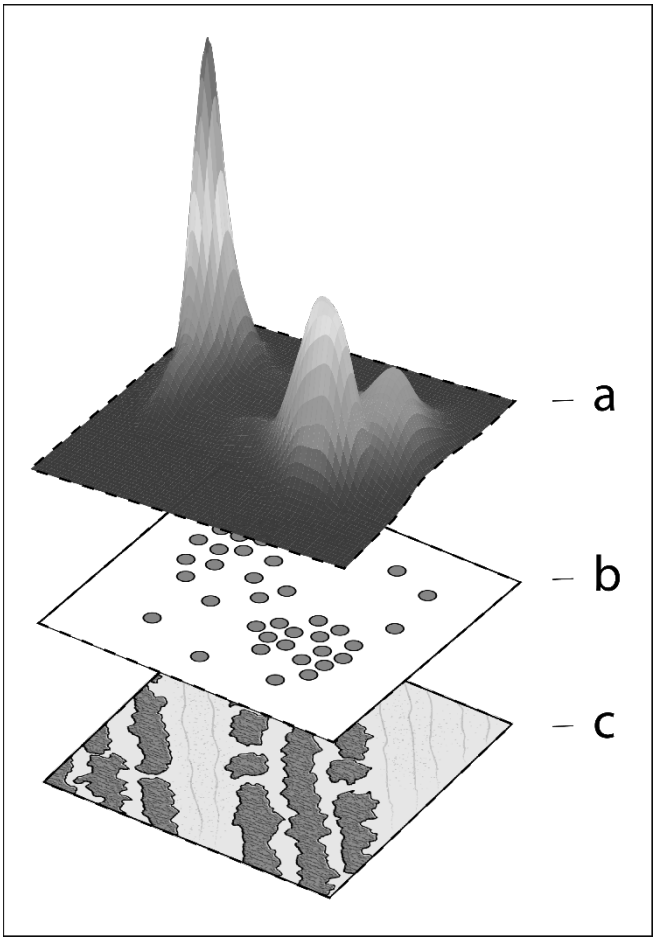


Figure 7 Illustration of underlying distributions for image objects: (c) the original image containing various crops, (b) feature space depiction of the pixel values in two dimensions, (a) underlying distribution of the pixels in *b* (approximated by kernel density estimation). Observe how *a* exhibits various local modes (peaks)

If you were to represent *objects* in the *scene* as classes (figure 7), there would exist a probability distribution for each particular class (García-Mateos et al., 2015). The color distribution of each class is then mainly[2] determined by their physiological properties, e.g. chlorophyll presence in vegetation that manifests as dominant green tones.

The appearance of the underlying distributions (i.e. shape, size, kurtosis) is then tied to the color space used. As such, the concept of *color space* can be redefined as the n-dimensional space induced by all possible tuples of that model. These tuples will commonly be composed of three or less parts due to the HVS (more specifically due to trichromacy, see Pascale, 2003; Sandwell, 2004 for the intricacies).

As an image can depict various objects in a scene, each having their own distribution, the choice of color spaces changes how these distributions relate to each other (e.g. varying overlap).

[2] Intrinsic and extrinsic factors like illumination will also influence the color distribution of classes

2.3.3 Overview of color transformations

The following color transformations, i.e. both color spaces and spectral indices, have been evaluated on their ability to emphasize various spectral characteristics by García-Mateos et al. (2015) and Hamuda et al. (2016), and subsequently filtered by their expected ability to do so for objects present in VHR remote sensing imagery. As with OBIA, we want to acknowledge the vast body of work that competently explores the complex paradigm of color and color spaces (notable examples being Pascale, 2003, and Shih et al., 2005), but do not feel that their subtleties express in ways relevant to this particular research (including their associative equations). What we do consider relevant are their respective purposes, the mention of the fact that they are commonly computed from the initial RGB color space using both linear and non-linear transformations, and references to competent works that describe them in more detail (all shown in table 2).

Table 2 Selected color transformations that emphasize varying characteristics of an image

ID	Description	Purpose	#	Reference
RGB	*Red (1), green (2), blue (3)*	*Compression (image)*	*4*	*García-Mateos et al. (2015)*
rgb	*Red (1), green (2), blue (3)*	*Normalization*	*3*	*Shih and Liu (2005)*
YUV	*Luminance (1) and chrominance (2, 3).*	*Compression (video)*	*3*	*García-Mateos et al. (2015)*
HSV	*Hue (1), saturation (2), value (3)*	*Intuitive color representation*	*3*	*García-Mateos et al. (2015)*
HLS	*Hue (1), brightness (2), purity (3)*	*Intuitive color representation*	*3*	*Valensei (1938)*
LAB	*Luminance (1) and chrominance (2, 3)*	*Uniform color representation*	*3*	*García-Mateos et al. (2015)*
LUV	*Luminance (1) and non-linear X,Y (2, 3)*	*Uniform color representation*	*3*	*García-Mateos et al. (2015)*
XYZ	*X (L/M wavelength), Y (illumination), Z (S wavelength)*	*Modeling the human retina (cones & rods)*	*3*	*CIE, (1931)*
I1I2I3	*Red (1), green (2), blue (3)*	*Decorrelate primary colors*	*3*	*Ohta et al . (1980)*
EXR	*Excess red index*	*Emphasize red tones*	*1*	*Meyer et al. (1998)*
EXB	*Excess blue index*	*Emphasize blue tones*	*1*	*Meyer et al. (1998)*
EXG	*Excess green index*	*Emphasize green tones*	*1*	*Meyer et al. (1998)*
CIVE	*Color Index of Vegetation Extraction*	*Separate vegetation from background*	*1*	*Kaaoka et al. (2003)*

3. Methodology and data

3.1 Study area and data

This research uses data that was originally made available for a joint experiment by Wageningen Plant Research and the Laboratory of Geo-information Science and remote Sensing of the Wageningen University. The dataset covers roughly 0.5 hectare of an experimental potato field in the south-east corner of the Dutch village Tollebeek (municipality of Noordoostpolder, province of Flevoland). A multitude of potato species (*Solanum tuberosum)* are cultivated in the research area including Vermont , Kondor, Lady Claire and Rosagold. Numerous plants harboring these potatoes in the research area are known to be infected with the Potato Y virus (specifically the PVY^{NTN} strain) and the Erwinia bacteria.

The dataset consists of the following (illustrated in figure 8):

1. RGB orthomosaic (byte) with a spatial resolution of 0.0083m (a);
2. Point shapefile (433 features) sampling diseased plants, localized using RTK sensors (b);
3. Polygon shapefile (7 features) describing potato species, amount, planting distance and known pathogens per field (c);
4. Digital Surface Model (DSM) derived from orthomosaic (d);

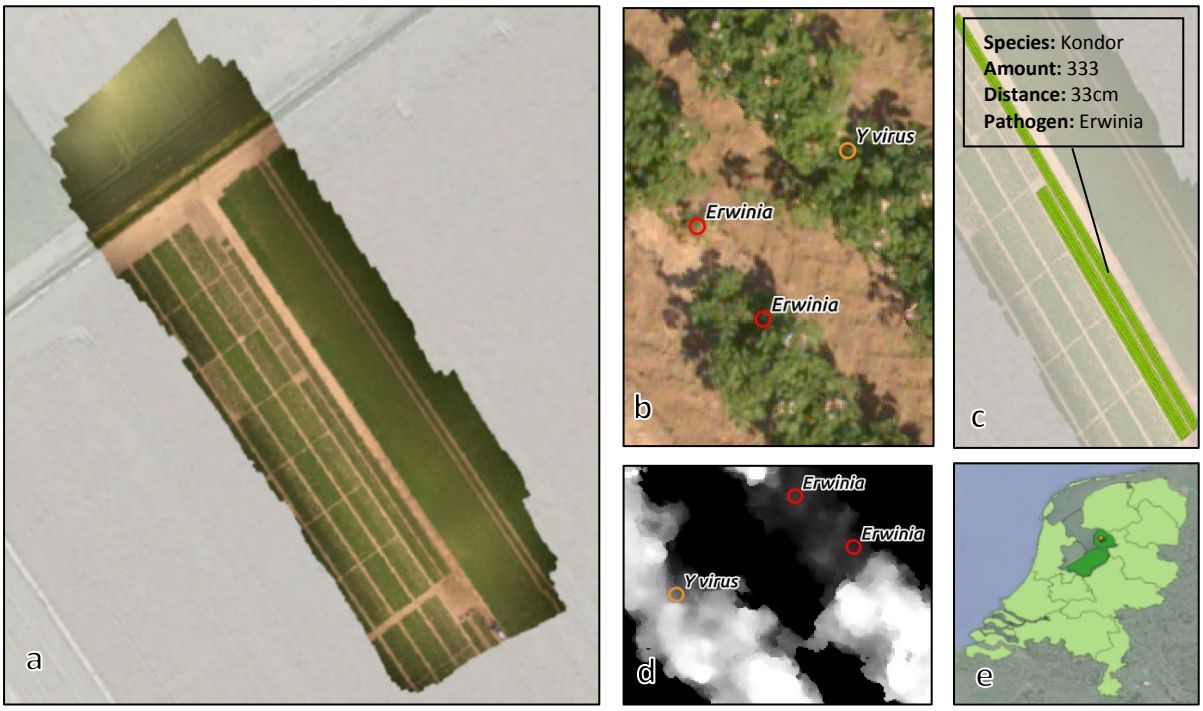


Figure 8 *(a) RGB orthomosaic from HYMSY camera under UAV acquired on* 19-6-2017*, (b) RTK located diseased plants, (c) potato field information, (d) thresholded depiction of DSM acquired on* 19-6-2017*, (e) location of experimental field near Tollebeek in the Netherlands*

The VHR (RGB) imagery was captured and processed by the Hyperspectral Mapping System (HYMSY, Suomalainen et al., 2014) on 19-6-2017 and 5-7-2017 whereas the ground samples and their associative RTK measurements were taken by field experts on 19-6-2017 and 27-6-2017. Initial inspection of the imagery shows that the spatial resolution (0.0083m) is high enough to distinguish objects roughly 15 cm in diameter on both dates, e.g. flowers inside potato plants, excluding individual leafs (that often occlude each other). Spatial and optical patterns can thus be seen (e.g. leaf necrosis) in objects of these sizes. Empirically, three classes can reliably be distinguished; inflorescence, green vegetation, and soil (now referred to as $flower$, veg and $soil$). Similarly, different potato plant types can be distinguished by their morphological features (specifically color and shape, see CFI, 2015 for taxonomy information).

Further inspection shows that the plants indicated as diseased by figure 8.*b* (from either PVY or Erwinia) show considerable variance in the severity of symptoms, to the point where supposed traits of diseased plants are not visible despite them being labeled as such. It is also important to mention that RTK referencing can have errors, both from technical sources (e.g. satellite clock offsets) or simple interpretation error (e.g. plants not being registered at the same locations consistently), resulting in small spatial offsets. Furthermore, the imagery's top-down perspective will not capture information under the canopy (e.g. stems with signs of *blackleg,* section 2.1.3). Lastly, the digital surface model is an approximation of the actual surface as it is derived from a process called *structure from motion* (Suomalainen et al., 2014). The overall model is accurate but is considerably smoother than the actual surface (e.g. tire tracks that are visible in the orthophotos but not in the DSM). Height patterns with bigger differences however can still be seen (e.g. canopy height).

3.2 Methodology

In short, we seek to model diseased potato plants with object-based features derived from VHR imagery that are visible with the naked eye (as to emulate manual plant assessment done with the HVS). This approach can roughly be split in two parts; object-based approximation of potato plants using LSMSS and classification of disease in said objects using their morphological traits. Our methodology for the former involves a continuation of Hernandez (2009) where several color spaces were evaluated by their ability to optimize segmentation of vegetation in crop site imagery using a naive Bayesian classifier and supervised sampling. We will largely follow the same approach but applied to segments (i.e. bundled pixels) instead of individual pixels, following LSMSS of VHR imagery where $soil$, $flower$ and veg are approximated as objects. We further modify their approach by adding more class-specific indices from Hamuda et al. (2016) that potentially improves classification, and employ kernel density estimation in an effort to account for limited sample sizes (expected for VHR imagery).

Similarly, our methodology for the latter is an adaptation of Feng et al. (2015) where texture analysis and Random Forest were used to classify land covers in an effort to overcome the limitations of low spectral resolution. Inspired by overviews given by Barbedo (2013) and Mahlein (2016), we will be expanding their feature list with several morphological descriptors (e.g. shape, texture, and size) expected to be indicative of Erwinia and PVY presence in potato plants (see section 2.1.5). The entire processing chain is written in Python, providing a common interface for Orfeo Toolbox and other software employed.

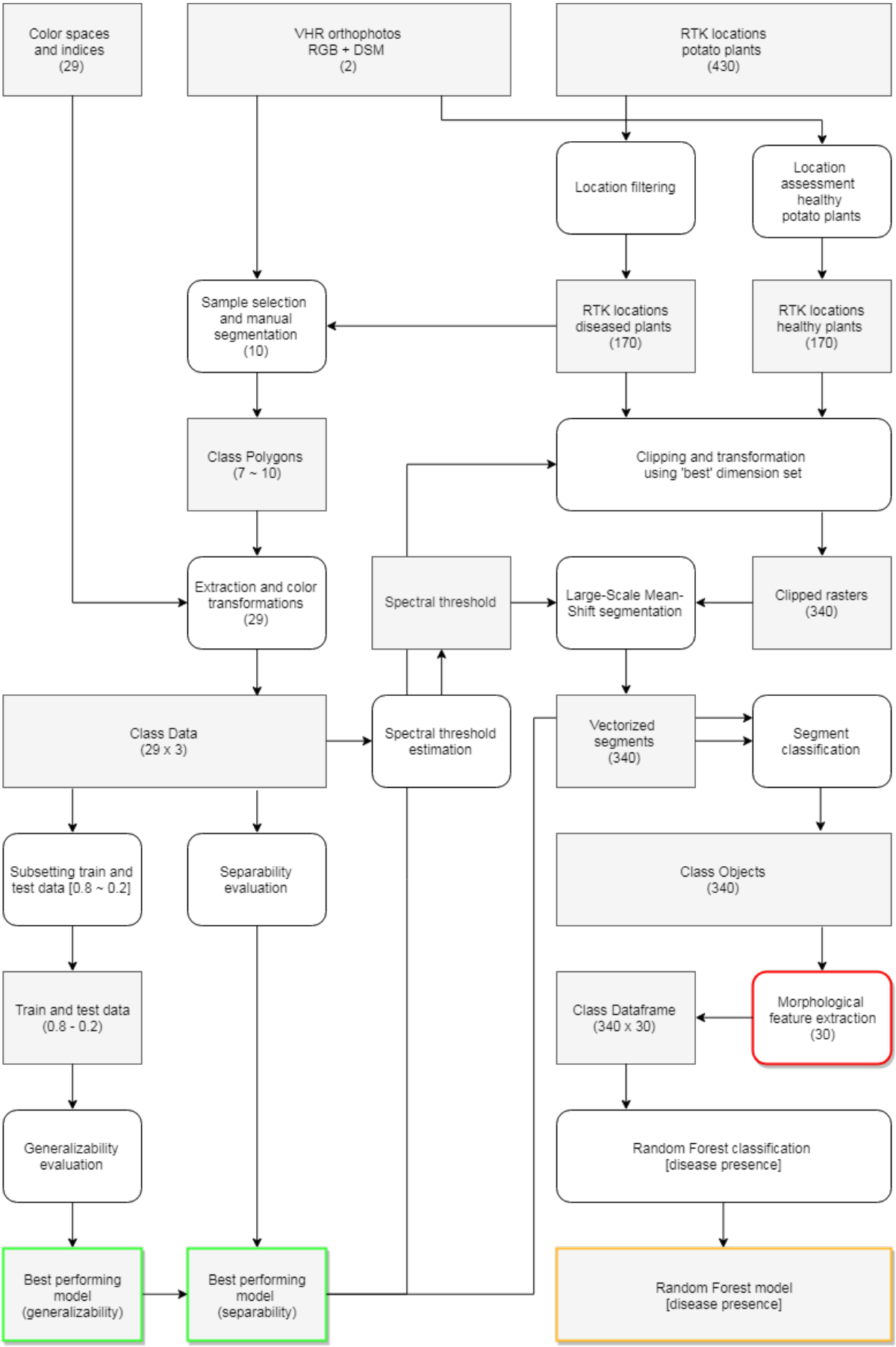


Figure 9 Overview of proposed methodology (orientated from top): grey squares represent results (or input), white rounded squares represent processes, and red, green and orange strokes represent steps directly related to research questions 1, 2 and 3 respectively

3.2.1 Sample selection and manual segmentation

Using the known locations of diseased plants in our scene (figure 8b), 10 observations will be selected on the criteria that they reflect similar variance found in the complete set (scene location, illumination, plant species, disease type, and severity). These points are subsequently buffered to 40 cm which should incorporate all possible sizes of potato plants found in the research area (see section 2.1.2). The resulting polygons are then used as outline for manual segmentation to be done in QGIS (GIS software), wherein polygons are drawn around the selected classes ($soil$, $flower$ and veg). We acknowledge that $flower$ might not hold relevant information regarding host health (see section 2.1.5), but accounting for said class is necessary as it could interfere with texture features to be computed from veg. Similarly, accounting for $soil$, $flower$ and veg allows us to validate any classification regarding said classes.

Figure 10 Illustration of the sampling scheme and manual segmentation of $soil$, $flower$ and veg (*class polygons)*

3.2.2 Instantiating class data

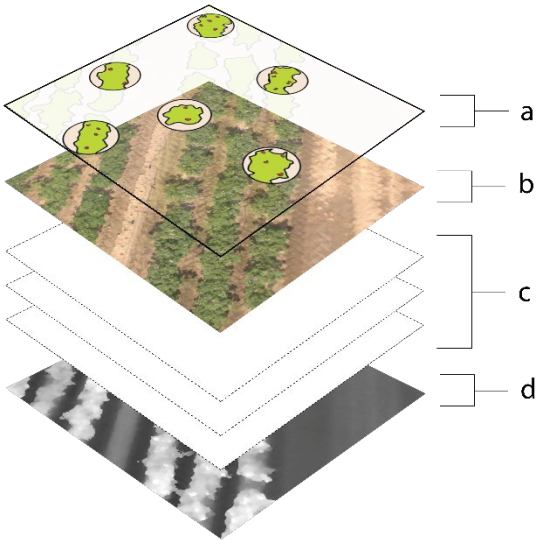


Figure 11 Extraction of *class data* visualized as isometric stack: (a) class masks, (b) RGB image, (c) RGB color space transformations, (d) Digital Surface Model. Observe how *class polygons* (*a)* consist of 3 classes (*veg, flower, soil)*

The *class polygons* derived from manual segmentation will be used as masks to extract the raster values from the VHR imagery (RGB + DSM) on both dates, giving us a range of pixel values per class (figure 11). We ultimately want to use this *class data* to classify new data (more specifically, the segments to be created by LSMSS given the complete VHR RGB image).

As discussed in section 2.3.2, each object (or *class)* in any particular scene has an underlying probability distribution within the *n*-dimensional space defined by the color space used (which also determines the shape and thus overlap between the classes under consideration). Given that we sampled for classes $soil$, $flower$ and veg, we can now define $p_{soil}(color)$, $p_{veg}(color)$, and $p_{flower}(color)$ as probability density functions (PDF) of color and height (now referred to as simply *color*).

We set out to approximate (or model) the underlying distributions by means of kernel density estimation (*KDE*) using the supervised samples (*class data)*, giving us $\hat{f}_{soil}(color)$, $\hat{f}_{veg}(color)$, and $\hat{f}_{flower}(color)$ and appropriate bandwidth sizes (as that highly influences the estimate by KDE, more so than the actual shape of the kernel, see Scipy, 2013). These models can subsequently be used to classify new data, effectively making them naive Bayesian classifiers (figure 12). Bandwidth selection will be done by employing Scott's rule of thumb, see Scott et al., 2004 and Scott, 2010 for its definition and comparison to other rules.

García-Mateos (2015) states that the amount of overlap in distributions is indicative of *separability* (i.e. the ability to separate the distributions of different classes in the color space) whilst the ability to correctly classify samples outside the training set is indicative of *generalizability*. Collectively, these attributes tells us something about the effectivity of the model (i.e. the color space used). We will be testing the following 29 color transformations (their products will be refered to as *dimensions* from now on) on these criteria, which were carefully selected from García-Mateos et al., (2015) and Hamuda et al., (2016) in section 2.3.3:

$$[HSV(3), HLS(2), YUV(3), LAB(3), LUV(3), XYZ(3), I1I2I3(3),\\ EXR(1), EXG(1), EXB(1), CIVE(1), RGB(3), DSM(1)]$$

The associatative transformations for these *dimensions* will be applied to the *class data*, and subsequently normalized across all axes to fit in one byte (0-255) to allow for easier parametrization of LSMSS (section 3.2.5).

3.2.3 Evaluating model separability and generalizability

With *class data* now composing of 29 dimensions, there are 80730 possible model sets (given a set limit of 5 dimensions to prevent redudant information) to be evaluated. We first measure the separability by applying the following algorithm (modified from García, 2015):

INPUT: C: class data
D: set of possible dimension sets (80730)

OUTPUT: M, S: minimal overlap, soil overlap, respectively
F, V: flower overlap, vegetation overlap, respectively

1. Let $C=\{c_{soil}, c_{flower}, c_{veg}\}$ and $c_{class}=\{c_{class}^1, \ldots, c_{class}^n\}$ with n being
2. the class-specific sample size
2. For each dimension set d in D:
3. Filter C using d, resulting in c
4. Compute KDE(c_{soil}), KDE(c_{flower}), KDE(c_{veg}),
5. Compute $P_{soil}(c)$ $P_{veg}(c)$, $P_{flower}(c)$ from class-specific KDEs
6. Compute P_{min} from min(P_{soil}, P_{flower}, P_{veg})
7. Integrate $\int_{soil}$, $\int_{flower}$, $\int_{veg}$, $\int_{min}$
8. Compute m by $(\int_{min}/\int_{soil}+\int_{flower}+\int_{veg})$
9. Compute s by $(\int_{soil}/\int_{soil}+\int_{flower}+\int_{veg})$
10. Compute f by $(\int_{flower}/\int_{soil}+\int_{flower}+\int_{veg})$
11. Compute v by $(\int_{veg}/\int_{soil}+\int_{flower}+\int_{veg})$

Algorithm 1 The proposed algorithm for evaluating class separability: Each set is used to filter *class data*, which is subsequently split into *veg data*, *soil data*, *flower data*, of which their underlying distributions are approximated via kernel density estimation. We then compute the probability values of these functions given *class data*, and compute the minimal probability values across all class probabilities. These are then integrated using the trapezoid rule, where the ratios of both individual and collective classes' distribution area from the total area gives us a sense of separability

In turn, the *generalizability* of each possible model will be evaluated using the following algorithm (again modified from García, 2015) which employs the Matthew Correlation Coefficient (MCC; a measure of binary classification accuracy incorporating both true and false positives and negatives, see Powers, 2007)*:*

```
INPUT:    R: train data (sampled from class data, 0.8 ratio)
          E: test data (sampled from class data, 0.2 ratio)
          D: set of possible dimension sets (80730)

OUTPUT:   S, F, V, M: MCC values for soil, flower, veg, and their mean

1.     Let R={R_soil, R_flower, R_veg} and R_veg={R_class^1, …, R_class^n} with n being
2.     the class-specific sample size
3.     For each dimension set d in D:
4.           Filter R and E using d, resulting in r and e
5.           Compute KDE(r_soil), KDE(r_flower), KDE(r_veg)
6.           Compute P_soil(e), P_flower(e), P_veg(e) from class-specific KDEs
7.           Compute class-specific confusion matrices
8.           Compute s, f, v from said matrices
9.           Compute m by weighted averaging of s,f,v (by class size)
```

Algorithm 2 The proposed algorithm for evaluating generalizability: Each set is used to filter *class data*, which is then split into *veg data*, *soil data*, *flower data*, of which their underlying distributions are approximated via kernel density estimation. The generalizability of these 'models' is then evaluated using the test data, from which class-specific confusion matrices are computed, and subsequently MCC values per class. The weighted average of these class-specific binary evaluators is then computed (using their respective class size) giving us an estimate on generalizability

The set of dimensions that are shown to be most separable and general (thus increasing the limited spectral information, expected to improve LSMSS classification) will be used to transform the input data for LSMSS, while their associatative model set will be used to classify the resulting segments to $soil$, $flower$ and veg. Our proposed method for evaluating class separability is also depicted in figure 12, which is illustrative of KDE's ability to both approximate class distributions, and classify new data employing said distributions.

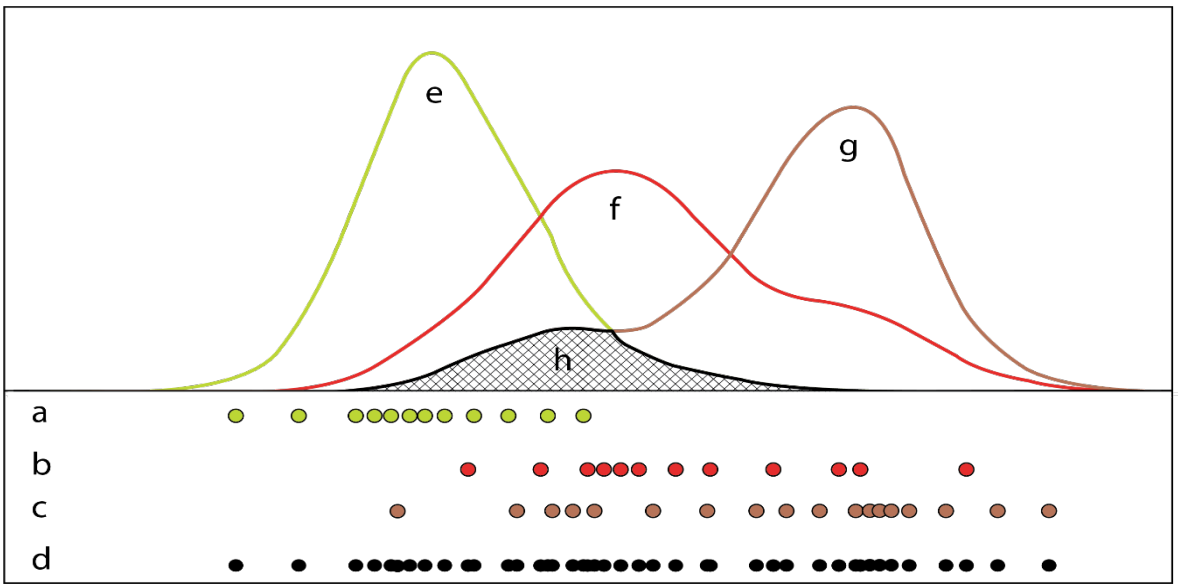


Figure 12 One-dimensional illustration of the approximation of class distributions using KDE, and the proposed method of evaluating overlap in said distributions: (a) *veg* subset of *class data*, (b) *flower* subset of *class data,* (c) *soil* subset of *class data,* (d) collective *class data* in set dimension, (e) approximated *veg* distribution, (f) approximated *flower* distribution, (g) approximated *soil* distribution, (h) minimal overlap across all distributions. Observe the various modes

3.2.4 Instantiating LSMSS input imagery

As mentioned in section 3.2, we have set out to approximate both healthy and diseased potato plants as object-based classes, from which we will evaluate whether their morphological traits allow us to classify disease presence therein. The object-based classes are to be established by classifying segments of VHR imagery, which is first transformed to the dimensions expected to increase the dissimilarity of all classes present in said imagery. To ensure that only individual plants are being approximated (as they exhibit varying overlap, see section 3.1) and to enable supervised sampling of known diseased plants, we will build the input imagery for this process around the known locations (430 in total) of diseased plants (see fig 8b).

In this respective data set, first all non-overlapping points are removed to make sure that all locations are present in both dates. Then a buffer of 40cm is applied which is subsequently used to filter out points that are too close to each other (i.e. disallowing redundant sampling) and to account for the spatial offset in RTK localization (see section 3.1), resulting in a total of 170 non-overlapping diseased plant locations. As field experts have inspected the entire field, we assume that plants that are not labeled as diseased can be considered healthy. Under that assumption (and using empirical knowledge gained in section 2.1), another 170 non-overlapping buffered points are created that represent healthy plants.

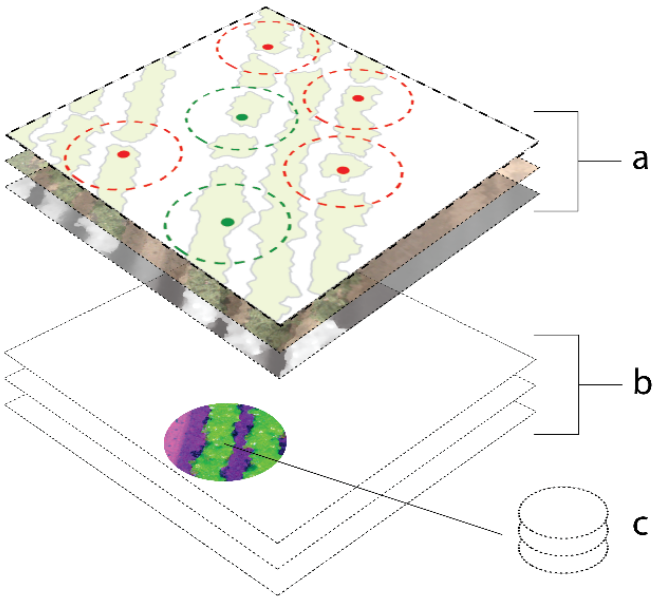


Figure 13 Creation of discrete *point images* visualized as isometric stack: (a) 340 selected locations of healthy and diseased plants used to clip the VHR imagery, (b) 340 discrete raster images, (c) Circle stack emphasizing that *b* is composed of *n* dimensions

Given the VHR imagery (fig. 12.*a*), 340 locations of healthy and diseased plants (fig. 12.*a*), and a selected dimension set, we essentially reapply the method described in section 3.2.2 with the distinction of only applying the transformations associated with the selected dimension set and saving them as 340 discrete raster images (fig. 14.*b*) instead of extracting their raw values.

We use the minima and maxima of the *class data* to normalize these images to the same byte range (0-255) for consistent LSMSS parametrization (even if *n* dimensions vary) and valid classification using said data. This gives us 340 images, each containing at least one instance of $soil$, $flower$ and veg, to be used as input for our proposed LSMSS-based pipeline which will approximate the real-world classes via objects.

3.2.5 Large Scale Mean-shift Segmentation and Naive Bayesian classification

As detailed in section 2.2.3 and Michel et al., (2015), LSMSS requires a spatial range and a spectral threshold to be set, parameters that essentially define when pixels should be considered similar. These parameters function as (Euclidean) thresholds that establish and iteratively expand segments as the associated pixels converge to their local modes (see section 2.2.3 for a more in depth explanation). As described in section 3.1, $soil$, $flower$ and veg all exhibit high variance which means multiple modes can be expected, even within individual classes, along with varying levels of class overlap (both are illustrated by figure 13). As we seek to approximate these classes by means of homogenous (non-overlapping) segments, these parameters need to be set to minimize this within-class variance whilst ensuring preservation of the most distinctive class modes (i.e. allowing for image smoothing that preserves class boundaries).

The *spatial range* will hereby simply be fixed to the smallest class under consideration, as the 'window' in each iteration (pixels that adhere to set spectral and spatial boundaries) needs to contain a minimum of two classes as to not under-segment (Michel et al., 2015). Given a spatial resolution of 0.0086m and the smallest class being $flower$ (which can encapsulated in a range of 8 cm, seen in section 3.1), we will set the spatial range to 10 pixels (0.08/ 0.0086 = 9.28). The *spectral threshold* however is estimated via the following algorithm (further illustrated by figure 13):

INPUT: B: class data in most separable dimension

OUTPUT: S: Euclidean threshold that preserves soil, flower, and veg

1. Let B={B_{soil}, B_{flower}, B_{veg}} and B_{class}={B_{class}^{1}, …, B_{class}^{n}} with *n* being
2. the class-specific sample size
3. Plot B_{soil}, B_{flower}, B_{veg} in n-dimensional feature space
4. Compute convex hull for B_{soil}, B_{flower}, B_{veg}
5. If $hull_{soil}$, $hull_{flower}$, $hull_{veg}$ overlap:
6. Remove points causing the overlap in smallest hull
7. Set S to minimal Euclidean distance between all classes

Algorithm 3 The proposed algorithm for estimating an appropriate spectral threshold: Class-specific convex hulls are calculated from given class data (filtered by set dimensions) giving us a measure of spectral overlap. If class hulls do not overlap, the spectral threshold can safely be set to the smallest Euclidean distance between the classes. If the class hulls do overlap, it means that the pixel values in these overlapping areas return equal or ambiguous probability values for multiple classes (i.e. no decision boundary can be established). These pixel sets are then discarded from the smallest class (less data evaluation), and their convex hulls recomputed. We then compute the smallest distance between the outer points that make up the actual class hulls, given us an appropriate spectral threshold

As shown in algorithm 3, we disregard the data preventing us to establish any spectral threshold, effectively equating them to outliers as we assume the spectral signatures of our classes to be fairly distinctive (or deterministic). This is done because if no 'safe' spectral threshold can be determined, LSMSS would only establish individual pixels as segments, effectively devolving our approach to a pixel-based classifier. The parameters that terminate the underlying LSMSS processes (maximum number of iterations, mode threshold) are kept to their default values (100, 0.9) as convergence is expected to occur long before they are needed because of the relatively low *spatial range* derived from $flower$.

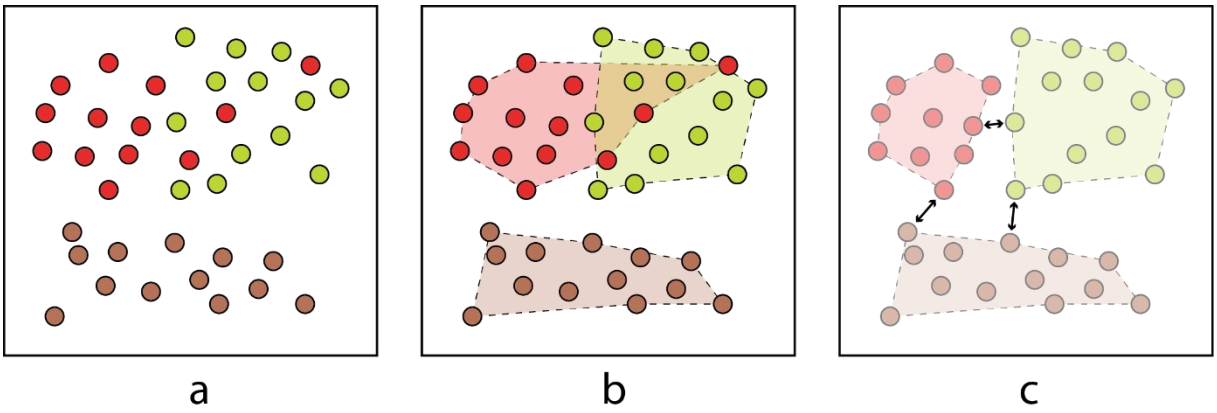


Figure 14 Illustrating the proposed method of spectral threshold estimation in 2D feature space: (a) *soil, veg, and flower* data in two dimensions, (b) calculated convex hulls per class, requiring removal as they overlap , (c) recalculated hulls, and the estimation of minimal Euclidean distance between classes (i.e. selecting the smallest arrow)

Having estimated appropriate parameters for LSMSS, we input the aforementioned 340 images (each containing at least one instance of $soil$, $flower$ or veg) to said tool giving us 340 segmented images (i.e. polygons containing the mean and variance of their associative pixels). We make explicit their relationship to the sampled real-world objects by classifying the segments by their mean values with the associated Naive Bayesian classifier created in section 3.2.2 (again, see figure 13). These 340 classified sets of segments (composed of $soil$, $flower$ and veg) will now be referred to as *class objects*.

The segmentation and classification preluding to the *class objects* are validated using the manually segmented *class polygons* (see section 3.2.1) as ground truth, which was also used to create the associative classifier. This approach is valid as only a subset of the *class data* was used to build said classifier, i.e. we are not testing and training on identical datasets. The *class objects* comprising the same space as the *class polygons* are first selected, from which class-specific confusion matrices, and finally Matthew Correlation Coefficient values, are computed.

3.2.6 Feature extraction from class objects

Following our findings regarding the morphological traits of diseased potato plants (table 1), we set out to model this knowledge into quantifiably features that, if extracted from the 340 (now validated) class objects, can be used to model disease said to be present in these objects. The following four morphological feature sets were hereby deemed appropriate as their derivatives are frequently employed to detect, quantify, and classify plant diseases (see Barbedo, 2013): *size, shape, texture, structure*. Before we can compute these feature sets, we first ensure the ability to do so by removing segments too small for texture analysis (minimum of 9 pixels) or that are invalid by definition ($flower$ outside veg). The resulting gaps will be filled using class values of neighboring segments, after which we add multiple known attributes to all segments (disease status, plant type, row ID, disease type).

We will be employing the gray-level co-occurrence matrix (GLCM) for texture analysis, as it is still one of the most popular methods to do so (Blaschke, 2014; Su et al., 2008), also in plant disease detection (Barbedo, 2013). These matrices are two-dimensional histograms of gray levels in an image, that indicate the probability of pixel pairs to co-occur in a given direction and at certain lag distances in said image.

As the pathogens are expected to induce color deterioration and growth stunting, the GLCM matrices will be computed from both DSM and hue (depicts color, section 2.3.1), using the 340 *class objects* as masks (only *veg* and *flower)*, thus giving us *texture* and *structure.* They will be computed for four directions (0, 45, 90, 135), to be subsequently summed to achieve directional invariance, and two lag sizes (1 and 5 pixels, 0.0086m) to capture subtle disease patterns. From these matrices, the following texture statistics will be computed: *contrast (CON), dissimilarity (DIS)*, *homogeneity (HOM)*, *angular second moment (ASM)*, *energy (ENG)*, and *correlation (COR),* largely following Feng et al. (2015) and Su et al., (2008) (see the respective papers for the equations and detailed definitions).

Furthermore, we mainly seek to quantify *shape* and *size* as it is expected to reflect growth stunting (or lack thereof). *Size* will be quantified by simply computing the volume in each *class object (*only using *veg* and *flower* components) using the DSM. *Shape* however will be quantified by three features; *perimeter*, *aspect ratio*, *solidity*, inspired by overviews given by Barbedo (2013) and Blaschke (2010). See OpenCV (2018) and appendix IV for the implementation. Lastly, we count the number of *flower* parts present in each *class object* to confirm their irrelevance with regards to disease presence. Summarizing, we will compute the following feature list (30) for all 340 *class objects*, to be referred to as *class dataframe:*

[*aspect ratio, solidity, perimeter, volume, net area, # flowers,* ASM_{DSM_1}, ASM_{DSM_5}, ASM_{hue_1}, ASM_{hue_5}, HOM_{DSM_1}, HOM_{DSM_5}, HOM_{hue_1}, HOM_{hue_5}, CON_{DSM_1}, CON_{DSM_5}, CON_{hue_1}, CON_{hue_5}, DIS_{DSM_1}, DIS_{DSM_5}, DIS_{hue_1}, DIS_{hue_5}, ENG_{DSM_1}, ENG_{DSM_5}, ENG_{hue_1}, ENG_{hue_5}, COR_{hue_1}, COR_{hue_5}, COR_{DSM_5}, COR_{DSM_5}]

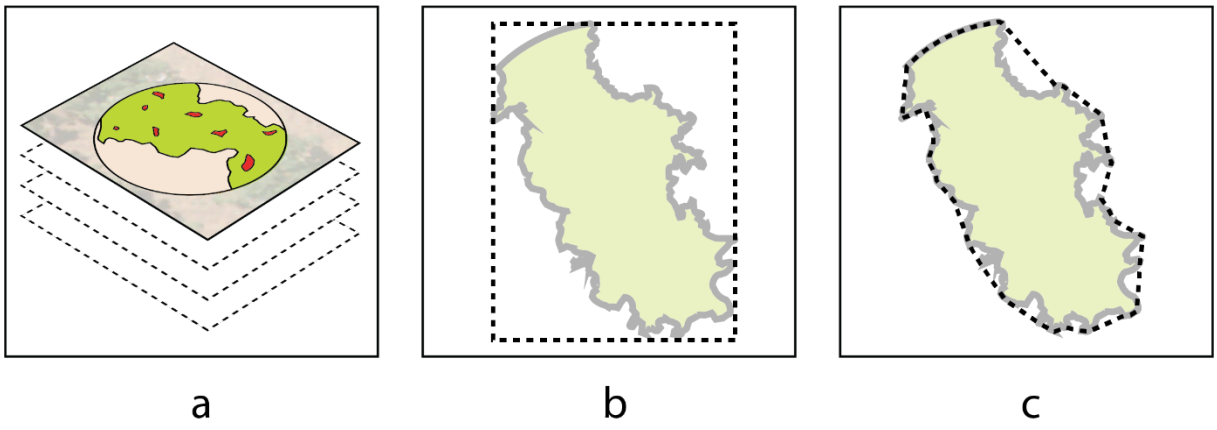


Figure 15 Mock-up of the 340 class objects (a), each containing one instance of *veg, flower,* or *soil*, and two examples features: (b) aspect ratio, or the ratio of width to height of the object extent, and (c) solidity, or the ratio of object area to its convex hull

3.2.8 Random forest classification

Given our *class dataframe* (340 *class objects* with 30 *features*), we will be training a random forest model following Feng et al. (2015) on disease status. First we split said data in training and test subsets, to be used in the training of a base model. This base model uses default parameters (meaning it employs bootstrapping, does not add class weights or feature limits, and is composed of 10 trees), and is only used to iteratively determine the optimal *hyperparameters* (i.e. meta settings) via a process called *gridsearch* (Koehrsen, 2018). This process essentially randomizes model parameters between given ranges, creates a model on each parameter set, tests said model given testing data and a set performance evaluator (e.g. *gini*), ultimately giving us an estimate on the 'best' parameters for a RF model given our data.

Once the optimal parameters have been established (number of trees, bootstrapping, class weights, criterion, maximum tree depth, maximum features, minimal sample requirement per leaf, minimal sample requirement per split, minimal impurity decrease), we employ n-fold cross validation (3) in the training of our definitive model to account for overfitting and redundant covariates. This optimized *hypermodel* will then be evaluated using the previously split testing data, and the following metrics: precision, recall, f1-score, and Matthew correlation coefficient. Precision tells us about the proportion of positive identifications that were actually correct, whilst recall (or sensitivity) tells us about the proportion of actual positives that were identified correctly. F1-score is a weighted ratio of both precision and recall, as they do evaluate different model characteristics irrespective of each other. Support is the number of elements in each class (i.e. *diseased* or *healthy*), and Matthew Correlation Coefficient is an overall evaluate of performance for binary classifiers (which incorporates both true and false positives and negatives, giving an overall sense of model performance between -1 and 1).

The same metrics will be computed for the base model and a dummy model (purely random classifier), giving us a sense whether our models can outperform pure chance, and evaluate whether disease assessment using our feature set is even possible. Lastly, we will compute the variable importance of covariates in our hypermodel to determine what features contain the most information on disease status (thus answering our final research question.

4. Results

In this chapter we will provide an overview of relevant results following our methodology, i.e. we will generally only visualize data if provides us with new insights. All plots, tables and corresponding parent datasets can be reinstated by running the Python scripts given in appendix IV (with a link to the required resources that are hosted online).

4.1 Class polygons and class data

We have sampled 10 locations from our pool of 430 plants (section 3.2.1) that reflect the variance found in the complete set (with regards to date, illumination, and plant type), which were subsequently used to create class-specific masks. The 7 masks associated with the first observation were then used to extract color and height data, of which the former is used to calculate 29 color transformations (4 are shown in figure 18). This data totaled 43287 vectors (or *pixels*) of which 68.1% are labeled veg, 29.6% are $soil$, and 2.3% are $flower$.

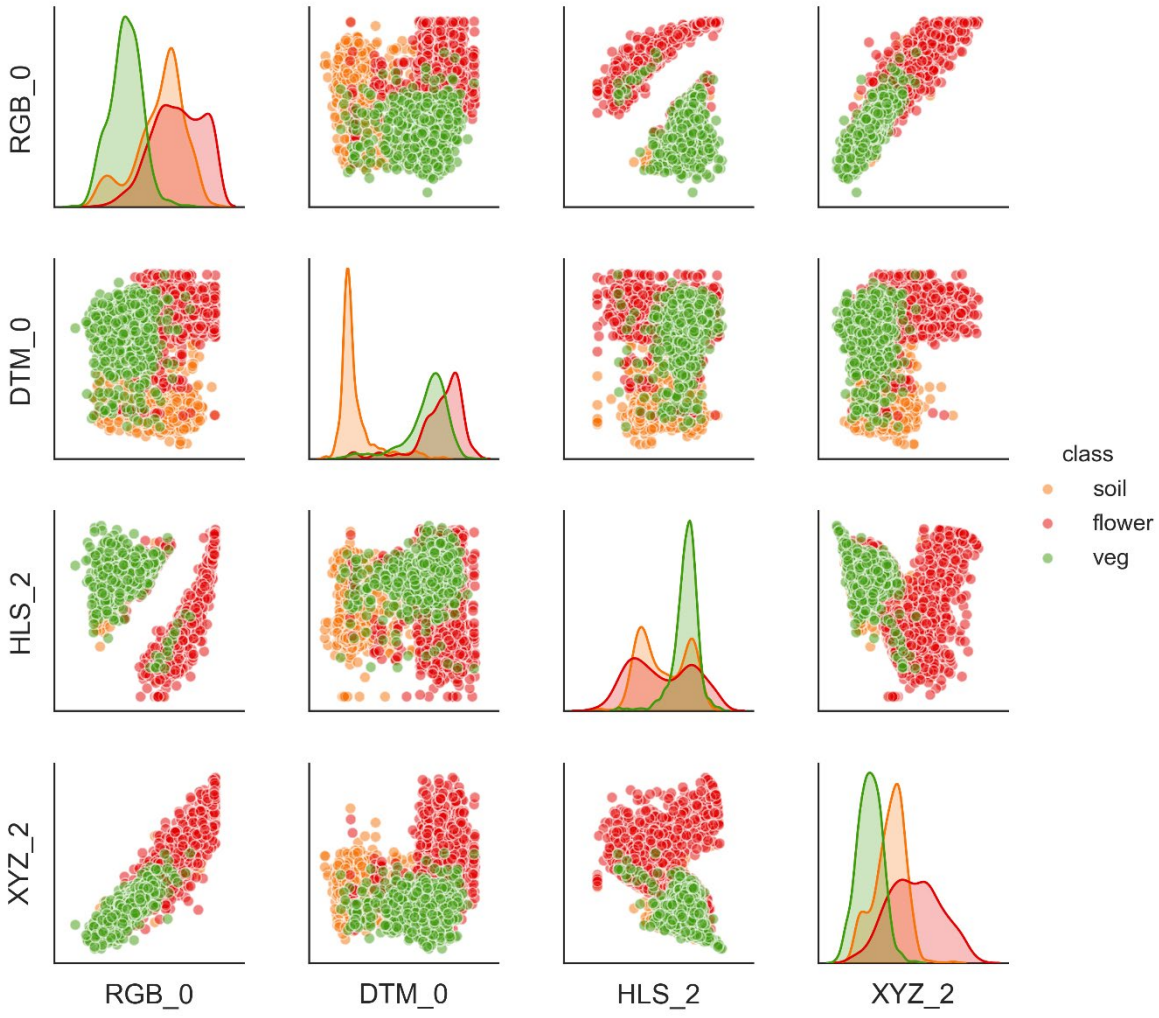


Figure 16 Sampled scatter matrix of 4 dimensions (all in 0-255 range) with sample class data (995 samples which adheres to the smallest class *flower):* only 4 dimensions are depicted out of a possible 30; diagonal axis depicts the underlying class distributions approximated via KDE

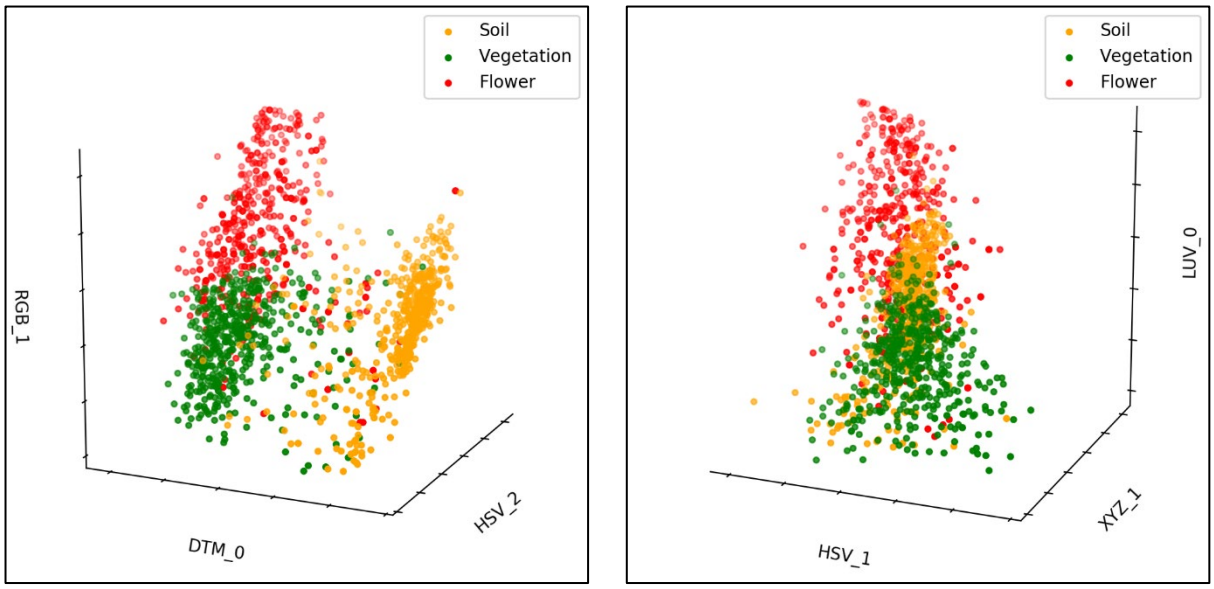


Figure 17 Two sampled scatterplots (500 points per class) in three dimensions (all in 0-255 range): (a) [*RGB_1, DSM_1, HSV_2*], and (b) [*HSV_1, XYZ_1, LUV_0*], notice how each plot exhibits a varying overlap between the sampled classes

The diagonal axis on figure 18 confirms the expected variance of the underlying class distributions' shapes when viewed through different color spaces (thus with varying class separability). Similarly, it shows that YUV_1 on its own will not be able to distinguish the classes as their distributions almost completely overlap. Notably, dimension DSM_0 does illustrate a clear difference between $soil$ and $veg/flower$. Overall veg and $soil$ do appear to exhibits single modes in all depictions but $flower$ appears to have multiple which can be indicative of either faulty sampling, class definitions that are too broad, or even erroneous bandwidth selection resulting in *under-* or *oversmoothing* in classes.

The other plots illustrate correlation between the dimensions (again, only notable subsets are shown) which coincidentally also depicts the classes' separability. As all dimensions were computed from only 3 bands, the majority of the products are highly (inter)correlated with one notable example of this given by the scatterplot [XYZ_2, RGB_0] depicted in figure 19. This does vary considerably by the composition of each set, as are their collective shapes and overlap (only observed empirically at this point) meaning different sets could emphasize different parts of the data. Looking at figures 18 and 19, classes appear to be better separable in higher dimensions but that is highly dependent on the set composition (see [DSM_0 RGB_0 HSV_0] in figure 19 and [DSM_0 RGB_0] in figure 18).

4.2 Separability and generalizability

Following our methodology, we evaluated the *separability* and *generalizability* of all possible models (i.e. dimension sets). As shown in figure 20, it can occur that particular dimensions on their own do not allow for adequate class distinction. They might however emphasize particular characteristics in the classes that improve separability if combined with other dimensions. This particular phenomena can be seen in table 3 where dimension YUV_2 is used in a set that gives us the second best separability (with regards to minimum overlap across all approximated class distributions).

Table 3 Notable separability results, only the five best performing sets for limits [5, 3, 1] are shown: the colors indicate the lowest overlap in each class (orange = *soil*, red = *flower*, green = *veg*)

Dimension set	MIN	SOIL	FLOWER	VEG	MAX
DSM_0+EXB_0+I23_2+HSV_1+RGB_2	0.00150	0.87284	0.02607	0.10109	0.00738
DSM_0+EXB_0+YUV_2+I23_1+HSV_1	0.00151	0.89697	0.03895	0.06409	0.00742
DSM_0+EXB_0+I23_1+I23_2+HSV_1	0.00151	0.88885	0.04074	0.07041	0.00784
DSM_0+I23_1+I23_2+LUV_1+LUV_2	0.00153	0.85819	0.01586	0.12595	0.00798
DSM_0+EXB_0+YUV_2+I23_2+HSV_1	0.00153	0.89661	0.03992	0.06347	0.00762
RGB_0+DSM_0+HSV_2	0.00349	0.93227	0.04438	0.02336	0.01176
RGB_1+DSM_0+HSV_2	0.00372	0.17926	0.05609	0.76464	0.02215
DSM_0+HSV_0+HSV_2	0.00526	0.63854	0.10304	0.25842	0.02576
RGB_0+HSV_2+EXB_0	0.00979	0.91542	0.06204	0.02254	0.05498
RGB_1+HSV_2+HLS_2	0.01307	0.11714	0.04382	0.83904	0.03103
DSM_0	0.02073	0.38014	0.30617	0.31369	0.22062
HSV_0	0.04865	0.44117	0.30198	0.25685	0.24214
LAB_1	0.05197	0.34822	0.32853	0.32325	0.27767
YUV_2	0.06396	0.31585	0.33266	0.35149	0.26678
HLS_2	0.14207	0.32771	0.23658	0.43571	0.28933

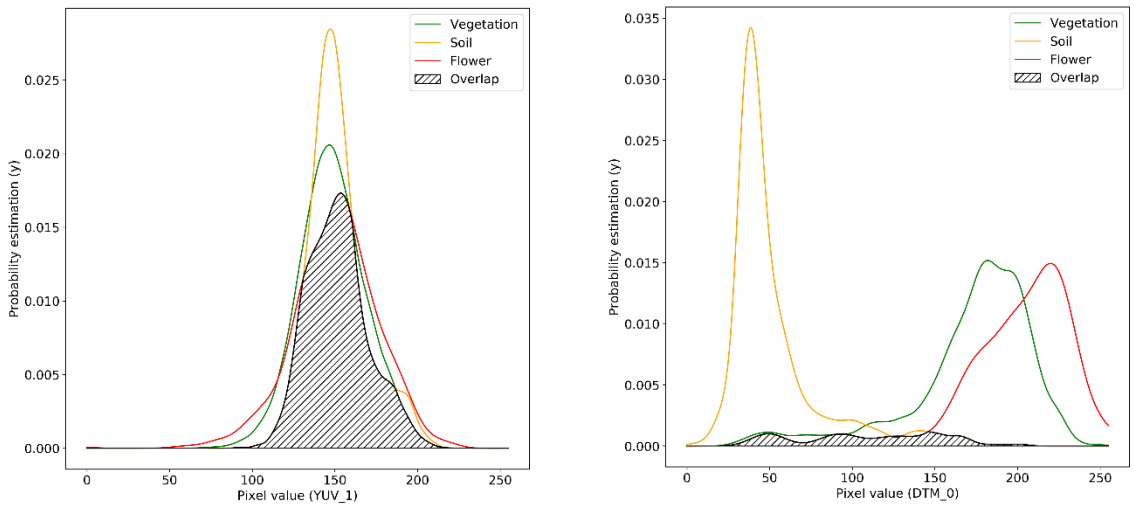


Figure 18 Two plots of probability density functions depicting both the best and worst performing one-dimensional dimension set with regards to *separability*: DSM_0 exhibits high separability across all classes but does show considerable overlap in *veg and flower* (intuitive result). YUV_1 barely separates the classes (gray area)

Other notable outcomes shown in table 3 is visible in the one-dimensional sets, where DSM_0 and HSV_0 exhibit the highest variability for the selected classes. This effectively confirms the validity of our *separability* index as these outcomes are very intuitive. Since $soil$ is always going to be physically lower than veg or $flower$, it makes a lot of sense that DSM_0 exhibits a high separability index. Similarly and as shown in section 2.3.1 , HSV_0 depicts the *hue* of color allowing distinction between colors like green, white, and brown (coincidentally the dominant colors in our classes; effectively reducing 3 dimensions to 1). As the number of dimensions grow (limited to 5), dimensions are starting to surface that do not hold a lot of information on their own (as discussed in the first paragraph). The most separable set of dimensions (regarding overlap across all classes) are $[DSM_0, EXB_0, I23_2, HSV_1, RGB_2]$ which represent height, 'blue-ness', 'green-ness', and color saturation respectively.

This set, as others in table 3, offers a high overall class separability but with considerable differences in the overlap for individual classes (e.g. the first row in table 3). Ultimately, a high soil overlap for example only means that this particular class comprises most of the area given the collective approximations of all underlying class distributions. As these functions are given equally sized class data (see section 3.2.3) , this either means that either a particular set of dimensions is sensitive to this class (e.g. DSM_0 and $soil$), or that the class definitions themselves are not distinctive enough in this said set. Accounting for this possible bias and evaluating each model's potential as classifier (as separability does not necessarily equal good classifiers), we evaluate the generalizability of the same models (of which notable results are depicted in table 4). The same patterns are visible here: better results in higher dimensions, $soil$ is best classified by height, $flower$ is best classified by 'blue-ness', veg is best classified by 'green-ness'. Slightly different compositions do surface here though, confirming that separability indeed does not necessarily equal good classification.

Table 4 Notable generalizability results, only the five best performing sets for limits [5, 3, 1] are shown: the MCC(μ*) column depicts the weighted mean, computed from the class sizes and class-specific MCC results

Dimension set	MCC(*veg*)	MCC(*flower*)	MCC(*soil*)	MCC(μ*)
DSM_0+HLS_2+HSV_0+LUV_1+LUV_2	0.860262	0.649457	0.883537	0.51032
DSM_0+YUV_2+HLS_2+HSV_0+LUV_2	0.857509	0.63696	0.881851	0.50556
DSM_0+I23_1+I23_2+LUV_1+LUV_2	0.845499	0.625602	0.869054	0.49758
DSM_0+YUV_2+HLS_2+HSV_0+YUV_1	0.854386	0.604864	0.881494	0.49469
DSM_0+EXB_0+I23_1+I23_2+LUV_1	0.84328	0.617831	0.866718	0.49426
DSM_0+HSV_0+HLS_1	0.834355	0.537583	0.871446	0.46829
DSM_0+YUV_0+HSV_0	0.835072	0.535683	0.868721	0.46711
RGB_1+DSM_0+HSV_2	0.838721	0.523977	0.873642	0.46484
DSM_0+HSV_0+HSV_2	0.832676	0.524994	0.867187	0.46301
RGB_0+DSM_0+HSV_2	0.775362	0.55155	0.814206	0.45313
DSM_0	0.6276	0.19633	0.83016	0.32548
RGB_2	0.57866	0.43062	0.50568	0.32044
XYZ_2	0.57112	0.42312	0.49702	0.31516
HSV_0	0.70743	0.13215	0.79449	0.30469

As we have sorted table 3 by weighted MCC, giving us the best classification results across all classes, the best model [*DSM_0 HLS_2 HSV_0 LUV_1 LUV_2]* appears to best classify *veg* and *soil* with a decrease in *flower* (remember that MCC's range is -1 to 1 and depicts binary classification, i.e. *flower* vs *others*). The same set is given if you sort the results by FLOWER (the smallest class both spatially and spectrally due to limited samples) making it an appropriate choice.

4.3 LSMSS parametrization, instancing and evaluation

We use the best set of dimensions (with regards to separability and generalizability) and their associated classifier (along with the training and test dataset) to initialize our proposed LSMSS pipeline. As detailed in section 3.2.5, we first estimate the spectral threshold from the training dataset which, although fairly low, still exhibits overlap (table 3) meaning that some data had to be discarded for a threshold to be determined (i.e. data without definitive decision boundaries). After said removal, the spectral threshold was estimated at 37.98.

We then first validated our object-based approach to classification by applying it to the original 10 locations, classifying the resulting objects using the associated model, and recomputing the MCC values using the same *class polygons* used in extraction (remember that the associated model was built on a subset of this data, allowing validation). As shown by table 5, the classification results using objects are very similar but are lower overall. *Flower* in particular appears to be misclassified more frequently but we would argue that this is merely due to the increased unit size (bundled pixels) and ultimately, the small size of *flower* (both spatially and in terms of samples). As each class is evaluated per base unit, i.e. bundled pixels of which *flower* already has few (and arguably not enough samples), missclassification will influence any classification metric much more. It however is still the best performing classifier expected to be general enough to be performed on new data.

Following section 3.2.3, we then gave 340 locations of known healthy and diseased plants to our LSMSS pipeline (using the aforementioned set of dimensions and the associated classifier) resulting in 340 approximated plant objects. Two examples of all LSMSS' products are depicted in figure 22, of which *b* and *c* make it clear that our initial estimation of the spectral threshold was valid as all classes appear much more homogenic without losing their minimal spectral boundaries. (*d)* also shows that convergence was found quickly, indicating a valid spatial range (also adds to reliability of segment values as mostly local data is used. Cleaning of the plant objects was occasionally deemed necessary, again indicating either faulty sampling, or indefinite classes (limited information).

Table 5 Two-part table with: overlap results for the selected dimension set, and the classification results for both the initial *class polygons* (CP) and the *class objects* (CO), of which the latter has a lower score overall, mainly in flower

	MIN	**SOIL**	**FLOWER**	**VEG**	**MAX**
DSM_0+HLS_2+HSV_0+LUV_1+LUV_2	0.00249	0.3215	0.01236	0.66614	0.01251
	UNIT	**MCC(s)**	**MCC(*f*)**	**MCC(*v*)**	**MCC(*μ*)**
DSM_0+HLS_2+HSV_0+LUV_1+LUV_2	CP	0.88354	0.64946	0.86026	0.51032
DSM_0+HLS_2+HSV_0+LUV_1+LUV_2	CO	0.86854	0.59616	0.84917	0.48259

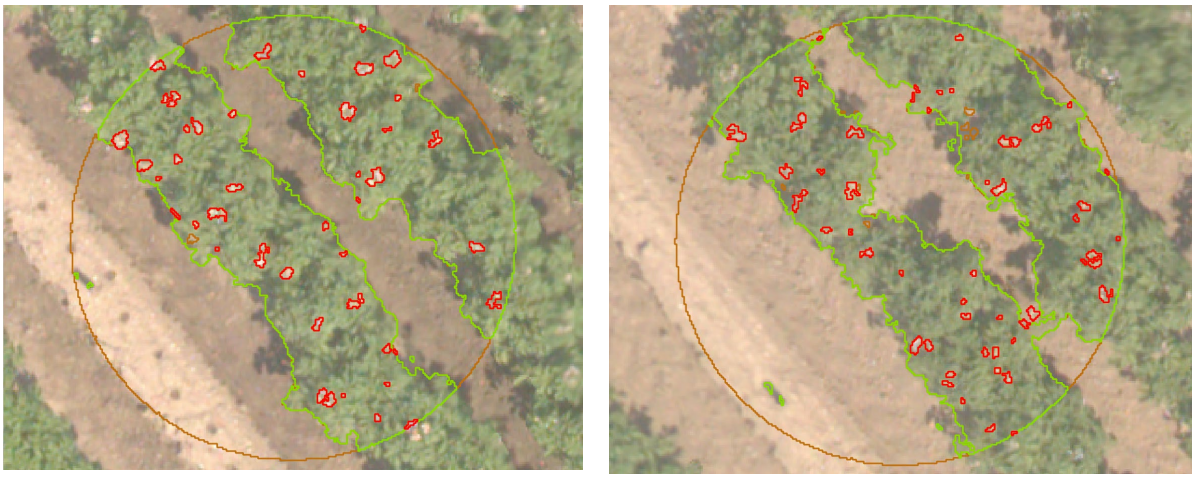


Figure 19 Two examples of approximated plant objects following LSMSS' pipeline given the 10 initial sample locations, to be evaluated by the manually segmented class polygons to validate OBIA method. As suggested by the table 3, *veg* and *soil* show accuracte segmentation with *DSM_0* largely accounting for shadow. Flowers are occasionally misclassified

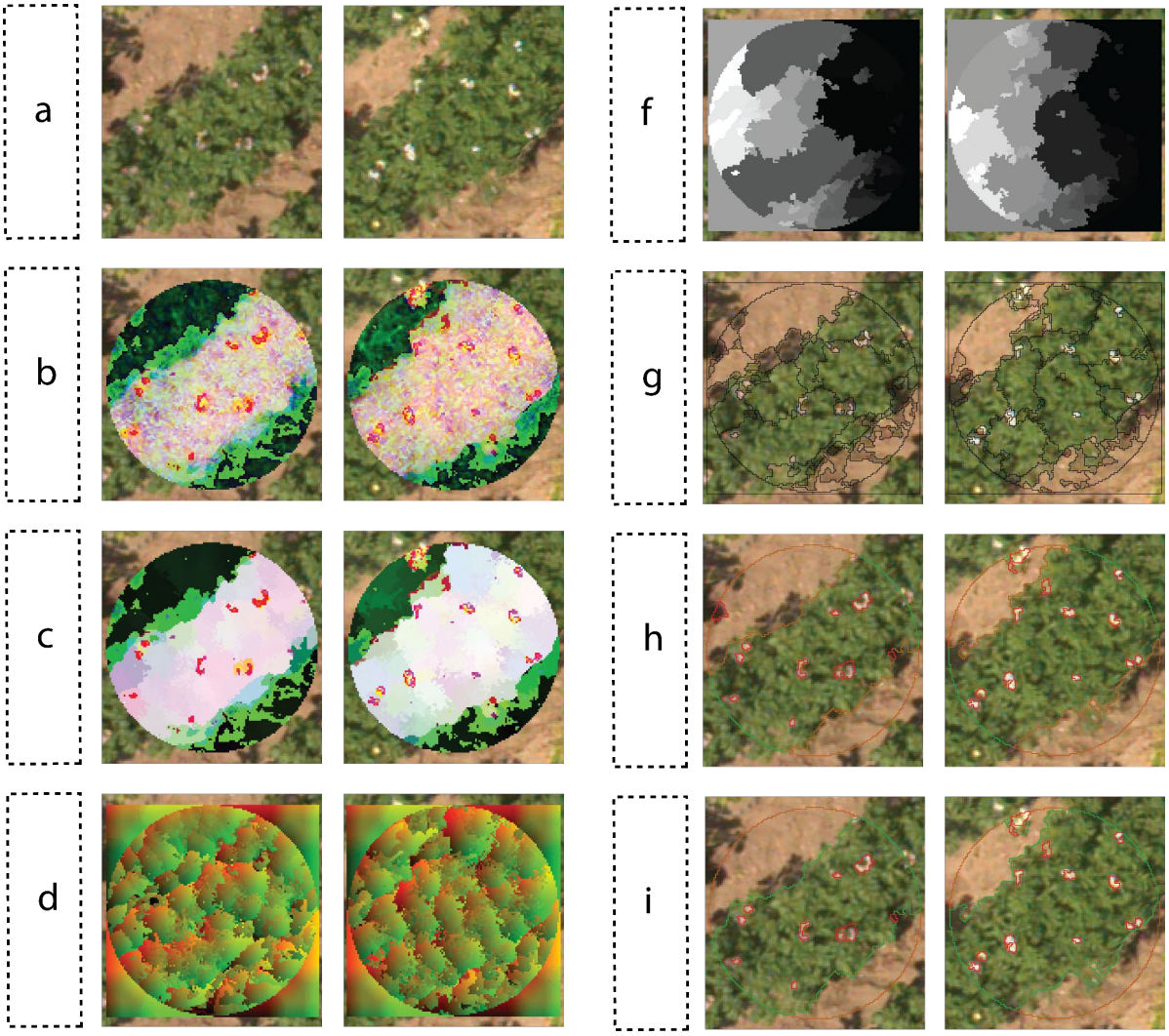


Figure 20 Two examples of LSMSS' products given class points (340 in total) : (a) location, (b) clipped buffers given 'best' dimensions, (c) filtered clips, (d) displacement images, (e) labeled images 1, (f) labeled images 2, (g) vectorized images, (h) classified vectors, (i) cleaned vectors.

4.4 Expanding and evaluation of class objects

Given the 340 preliminary class objects, we compute an exhaustive feature list that further approximates the real-world objects (i.e. potato plants) using their color, texture, shape, and volumetric characteristics. The means for all these variables in *healthy* and *diseased* groups can be found in table 5, together with T test results to evaluate how reliably different these groups are from each other (i.e. testing the null hypothesis). Some noteworthy differences are sizes (e.g. sick plants exhibit smaller volumes), and color contrasts (diseased plants appear to be bleaker). The majority of the differences are significant as seen in table 6.

Table 6 Mean feature list for diseased and healthy objects: T tests are performed on every feature and displayed with their associated P values (indicating significance). Significant T values are highlighted with a bold font (P <= 0.05)

	Healthy (μ)	**Diseased (μ)**	**T values**	**P values**
Volume	72.7668	61.566	**7.44281**	0
# Flowers	7.64118	6.38235	2.78653	0.00564
Aspect Ratio	1.06597	1.06297	0.39755	0.69125
Solidity	0.05259	0.06375	**-2.217**	0.02733
Net Area (m2)	0.50749	0.46247	**6.11125**	0
Perimeter (m)	6.22416	5.7683	**3.29756**	0.00108
Contrast 1 (DSM)	111.887	140.36	**-6.0246**	0
Dissimilarity 1 (DSM)	5.00174	5.56444	**-5.9483**	0
Homogeneity 1 (DSM)	0.33563	0.32129	**4.3921**	0.00002
ASM 1 (DSM)	0.001	0.00082	**5.26375**	0
Energy 1 (DSM)	0.03113	0.02835	**5.40393**	0
Correlation 1 (DSM)	0.9672	0.96978	-2.6944	0.0074
Contrast 5 (DSM)	759.653	933.711	**-5.9451**	0
Dissimilarity 5 (DSM)	17.7592	19.8636	**-6.1313**	0
Homogeneity 5 (DSM)	0.07983	0.07286	**4.44554**	0.00001
ASM 5 (DSM)	0.00033	0.00029	**4.0238**	0.00007
Energy 5 (DSM)	0.01786	0.01686	**4.04216**	0.00007
Correlation 5 (DSM)	0.75588	0.78134	**-4.2657**	0.00003
Contrast 1 (Hue)	30.6429	31.2285	-0.5783	0.56345
Dissimilarity 1(Hue)	3.92445	3.97107	-0.7314	0.46502
Homogeneity 1 (Hue)	0.24755	0.24455	0.84087	0.40102
ASM 1 (Hue)	0.00406	0.00368	**3.21612**	0.00143
Energy 1 (Hue)	0.06311	0.06009	**3.37492**	0.00082
Correlation 1 (Hue)	0.64758	0.68589	**-4.8532**	0
Contrast 5 (Hue)	67.2367	71.6527	**-2.4429**	0.01508
Dissimilarity 5 (Hue)	6.08534	6.30297	**-2.7502**	0.00628
Homogeneity 5 (Hue)	0.16461	0.15893	**2.96378**	0.00326
ASM 5 (Hue)	0.00313	0.00277	**4.29112**	0.00002
Energy 5 (Hue)	0.05551	0.05213	**4.48815**	0.00001
Correlation 5 (Hue)	0.19399	0.2515	**-5.3375**	0

As shown by the diagonal axis of figure 23, the individual features mostly exhibit one mode with considerable overlap between diseased and healthy points across all features (slightly lower texture correlation and dissimilarity in hue and DSM respectively). The scatter plots also exhibit considerable overlap but with varying 'centers' for the two class clusters (still only evaluated empirically) and various instances of outliers. The shape of the clusters tells us something about potential relations and/or interactions between the various features, of which the feature sets for *shape* and *size* for instance, imply considerable correlation (unsurprising as they depict varying facets of the same information). As mentioned in section 3.1, the majority of the plants said to be diseased expressed this in ways not immediately visible in the available imagery (empirically speaking). This is why we relish the sight of various outliers (that vary both in location and amount) and slight offsets of the two clusters in various plots, as that implies that some information regarding disease presence indeed has been captured by our morphological features (or a combination thereof).

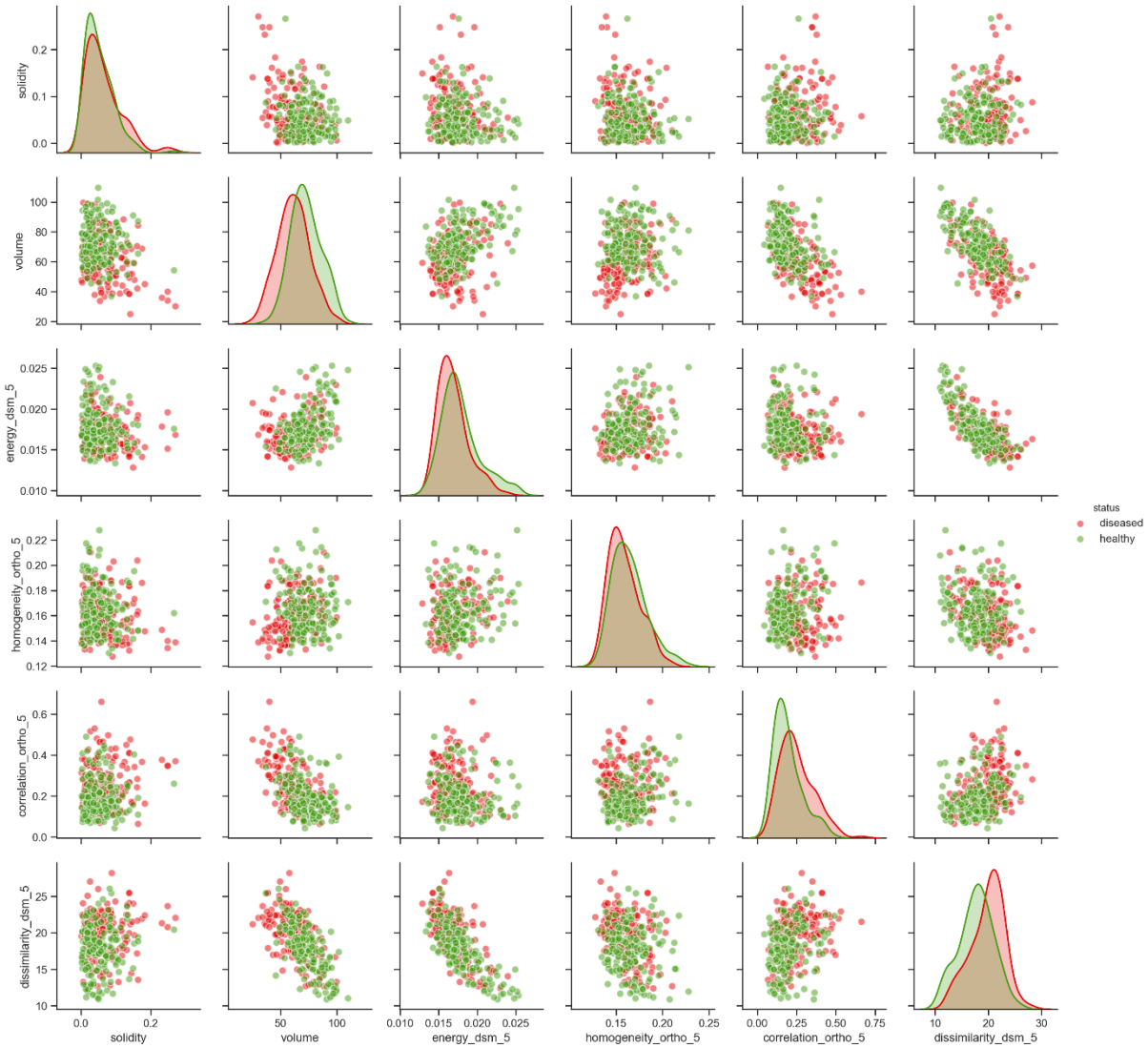


Figure 21 Scatter matrix depicting 6 cross scattered dimensions (of which there are 30 in total) with approximated distributions of data in said dimensions along the diagonal axis. Empirically, one can see that *diseased* vectors tend to cluster around the edge of the collective point cloud. Overlap is present in all dimensions but varies considerably

4.5 Random Forest Classification

Following our proposed methodology, we seek to model (and thus exploit) the information supposedly contained within the established *class objects* (that were expanded with several morphological features, and subsequently referred to as *class dataframe*) regarding disease status using random forest. The initial base model was initialized with 10 trees, bootstrapping, no maximum tree depth, a minimal sample size for leafs of 1, a minimal sample size of 2 for potential splits, which is all to be evaluated using the 'gini' criteria (a measure of information content, alternative to the 'entropy' measure).

Despite the base model' purpose (i.e. evaluating initial parameters and/or features to be tweaked to maximize performance in subsequent models), the initial results already show promising results (table 7). This table displays several model performance metrics, all calculated from underlying confusion matrices (giving us *true positives, true negatives, false positives,* and *false negatives*), presented and explained in section 3.2.8. Going by the MCC values alone, it is evident that classification of disease using the proposed feature set is possible, which is further supported as it structurally outperforms a dummy model (which equates classification to flipping a coin). Our optimized *hypermodel* only manages to slightly improve upon the initial base model, by a factor of 0.2, which is arguably unsurprising as their hyperparameters only differ in tree amount (10 -> 100) and maximum feature size (all -> 20).

We would also argue however that the fact that considerably more trees are employed also holds information, as it could imply that the feature set exhibits minute differences important to disease classification which cannot be captured by low tree amounts. It could also simply mean that the two feature datasets (healthy and diseased) are not 'clean', resulting in arbitrary leaf splits giving a false sense of information gain. The latter statement seems more plausible as some features were also considered to be redundant (as the suggested feature limit was set to 20). Other machine learning algorithms might have *overfitted* this potentially ´dirty' data, but RF's inherent averaging of trees and our n-fold cross validation should account for any such overfitting.

Table 7 Results of several Random Forest models: (base) initial model with default parameters, (dummy) purely random model, used as a base case, (hyper) improved base model by tweaking the hyperparameters

	precision	*recall*	*f1-score*	*support*	*mcc*	
healthy	0.69	0.77	0.73	35	..	
diseased	0.72	0.64	0.68	33	..	*base*
avg/total	0.71	0.71	0.70	68	0.41	
healthy	0.53	0.57	0.55	35	..	
diseased	0.50	0.45	0.48	33	..	*dummy*
avg/total	0.51	0.51	0.51	68	0.03	
healthy	0.73	0.77	0.75	35	..	
diseased	0.74	0.70	0.72	33	..	*hyper*
avg/total	0.74	0.74	0.73	68	0.47	

4.5.1 Variable importance

Further evaluating the optimized hypermodel, we have computed the variable importance of the used morphological features (figure 24) for the associative random forest, which should tell what features are most important for disease detection in potato plants. What is immediately apparent though is the considerably low importance values for all individual features (in the range of two decimals where commonly single decimals can be expected). This means that the fact that most expressive features are derivatives from DSM might not actually equate to much, given the relativity.

We can say with some certainty (by empirical evaluation and performed validation) that the *class objects* give appropriate approximations of the real-world potato plants, and that the resulting *class dataframe* is both exhaustive in its sample size (making it a reflection of the actual variance for the plants themselves, e.g. illumination) and balanced in its (perceived) representation of both diseased and healthy potato plants. This is leading us to suggest that the modelled disease traits are either considerably subtle, its data uncertain, or both, all of which are supported by section 3.1 where we mentioned that plants that are said to be diseased exhibit barely any empirical visibility, the considerable overlap of the clusters in figure 23, and the number of required trees for a relatively small feature set. The model itself performs well, which is indicative of deterministic properties in the collective feature set.

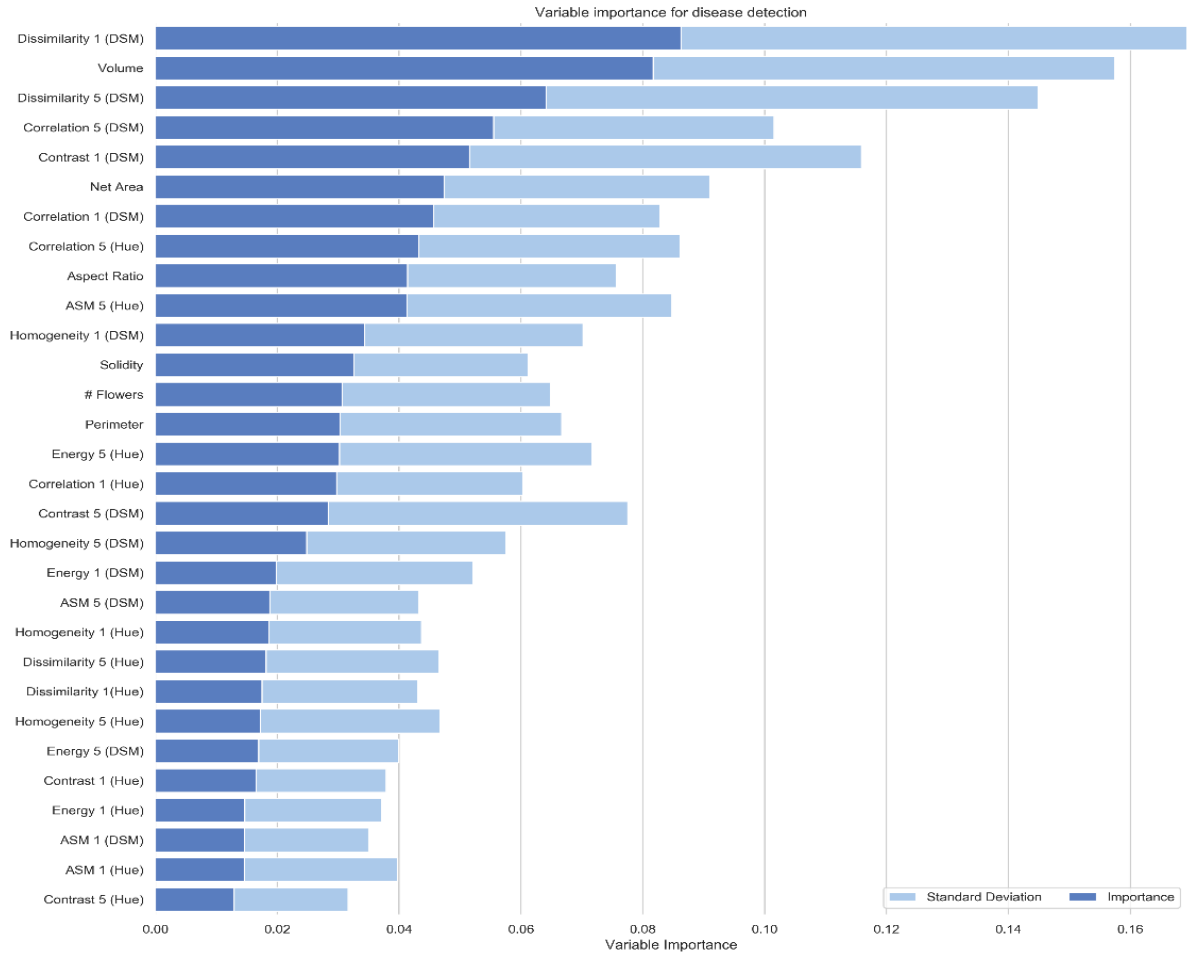


Figure 22 Variable importance plots for the base model and hyper model

5. Discussion

5.1 Modelling the human vision system

As we have mentioned in section 1.1.1, early and accurate detection of potato pathogens is a key factor in securing global food security but would necessitate highly accurate and regularly updated geo-spatial information, an endeavor that cannot be met by advancement of sensor technology alone (Gartner et al., 2010a). In an effort to explore the viability of object-based disease modelling, we have employed OBIA, specifically LSMSS with supervised class modelling, to approximate and model diseased potato plants present in VHR RGB UAV-based imagery. This section also states that, although promising hyperspectral proximal sensing methods are being explored, initial disease detection of crops is still primarily performed manually by plant experts. It is in this context that we related our modelling approach to the emulation of HVS, but how fitting is this relation?

Many studies have already compared OBIA methods with human interpretation of high resolution imagery (T. Blaschke, 2010), which was only recently made possible by technical advancements in both remote sensing and GIS (arguably also in computer science with solutions such as FOSS OBIA implementations and scalable computing clusters). Some notable parallels to the HVS drawn by these studies include: objects characterization by semantic knowledge, and scalable segmentation making object hierarchy explicit (D. Opitz, 2010; E. Weinke, S. Lang, 2010; M. Ehlers, 2010). Arguably the biggest discrepancy here is the fact that HVS decomposes images whereas OBIA builds them up (section 2.2.2; Gartner et al., 2010a; Wetheimer., 1925), which could prove to be problematic when delineating (sub)objects with more conceptual boundaries.

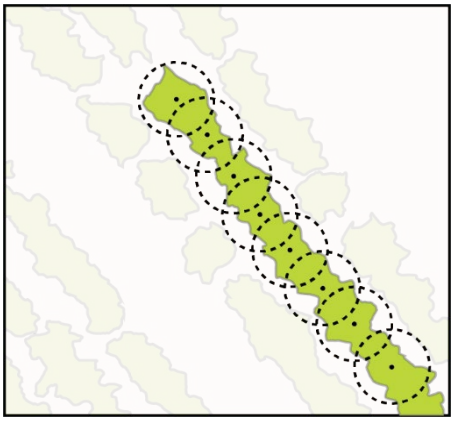


Figure 25 Visualization of ambiguous class boundaries, points indicate *true* origin

A more intuitive way to think about this problem is to consider an orchard, which is delineated on an aerial-photograph with ease, because of the specific arrangement of fruit trees in a matrix of grass (Lang, 2010). It exhibits a different kind of homogeneity, namely, regularity in structure (repetitive patterns) or conformity (i.e. constancy) in change, i.e. patterns that are readily delineated by humans (that factor in proximity, continuation, and other gestalt principles; see Wertheimer., 1925) but are hard to grasp by a machine. The potato plants in our VHR imagery exhibit a similar homogeneity as they are planted in parallel rows, roughly 33cm away from each other.

Over time, the individual plants have grown to physically overlap making for ambiguous class boundaries. Due to the principles of good continuation and good form respected by the HVS (gestalt principles; see Lang, 2010) however, human interpreters can imagine their boundaries as the plants exhibit recurring patterns (e.g. volume, center of mass). We effectively circumvented this problem as locations of (diseased) plants were known (fig 8.b), as required by our supervised classification scheme, but we do want to emphasize that similarly detailed location data is not commonly available.

5.2 Data quality and usage

We have performed manual segmentation to the best of our abilities (e.g. shifting local contrasts to better approximate class boundaries), but could not guarantee 'clean' supervised samples (i.e. completely separable class distributions) as seen in table 3. Although this is arguably reflecting real use cases that could employ flawed sampling schemes, our decision to use a naïve classifier in combinations with faulty sampling could result in ambiguous class definitions (or if they are already too broad to begin with) that hinder classification. We have subsequently used these *class polygons* to extract color and height information (including various color spaces often resulting from non-linear transformations) into *class data*, which was normalized across all channels to fit in one byte (the format used for the original RGB bands). Although done to enable easier LSMSS parametrization, several channels in the *class data* did exhibit varying levels of skewness and kurtosis (when initially evaluated), which is why normalization could have resulted in some loss of detail.

Furthermore, we had a breadth of data available that is arguably not used to its full extent, including hyperspectral imagery (Suomalainen et al., 2014). We have explicitly mentioned that we seek to model disease in potato plants by morphological features alone (exclusively computed from UAV-based VHR RGB imagery) in an effort to mimic the HVS (which is also only sensitive to visible light), as we subscribe to its exclusive potential to provide highly accurate and regularly updated geo-spatial information required for crop management (section 1.1.1). We could have however employed the hyperspectral imagery in one particular part, without endangering these principles, namely the estimation of healthy plant locations. As only locations of diseased plants were known on two dates, we exploited this fact (relying on the competence of the plant experts) by only looking in areas at a set distance from the diseased plants that were present in both dates (as disease presence is expected to persist). We would however like to emphasize here that most plants said to exhibit disease by plant experts, barely show clear signs (empirically speaking) of infection which did pose some initial concern whether our morphological features would be able to capture this.

Continuing, we could also have exploited the hyperspectral imagery to further delineate appropriate locations by, for instance, employing information in the red-edge region (indicative of healthy vegetation, section 1.1.1). Additionally, the class polygons were created from sampled locations that reflect the variance found in the research area (e.g. illumination, plant type, date), of which only 7 were employed as part of the data associated with the first observation (our LSMSS-based approach was only performed on data depicting 19-06-17). This was mostly done out of time constraint, but also because the dates for RTK measurements and the VHR imagery for the second observation vary by 8 days, making any direct correlation difficult. If one were to perform our modelling approach on several data, we would be able to capture their morphological change over time (as required for accurate crop management, section 1.1.1).

The segmentation and classification preluding to the *class objects* were validated using the manually segmented *class polygons* (see section 3.2.1), which was also used to create the associative classifier. We would argue that this approach remains valid as only a subset of the *class data* is used, therefore allowing the models to be evaluated on data they have not seen yet, and would relate it to a necessary evil due to the costs of manual segmentation.

By sampling and evaluating in the same locations, we could have introduced considerable bias that would not be picked up during model evaluation. This does not appear to be the case as empirical evaluation of the *class polygons* showed rather accurate segmentation in line with our evaluation (less so with *flower* which was most frequently misclassified, often necessitated object filtering, also in line with our evaluation, as seen in figure 21 and table 5).

One last point of concern is our method of estimating the spectral threshold from given class distributions. As mentioned, we were not able to fully prevent class overlap (section 4.2) which either means that our class definitions are too broad (and thus exhibit *genuine* overlap) or that our sampling scheme is flawed (which introduces *false* overlap). The pixel set related to this overlap exhibits equal probability of them belonging to two or more classes, which is why we remove them from the equation (relying on the distinctiveness of the resulting class distributions). We have chosen to remove the pixels sets in the smallest class (with regards to sample size) as that gives us less data to evaluate (see section 3.2.5), saving computational time. By doing this, we are effectively discarding already scarce class information (as stated by Strahlers theorem of small objects requiring more semantic knowledge, making sampling costly, and exhaustive sample sizes rare, see Strahler, 1986) which in turn could introduce a new bias for the other classes and result in over-segmentation of the smallest class, as the threshold is set too high. This is not the case, going by our results in figure 20 and table 3.

5.3 Method modifications and comparison

Sampling did indeed prove to be particularly costly for classes of higher semantics or smaller stature (e.g. *flower,* see section 4.1*)*. This was one of the reasons why we employed kernel density estimation for the evaluation of a models separability and generalizability, instead of histograms like the original authors. Although they are both appropriate non-parametric methods that approximate any complex shape resulting from the (often non-linear) transformations between color spaces, histograms require an exorbitant amount of samples in higher dimensions to ensure that enough bins are filled to make reliable classifications (i.e. *curse of dimensionality*; mentioned in the respective paper). We expect that this is the reason why channel mixing is not allowed and channel sets are limited to 3. KDE however essentially interpolates (and extrapolates) between given samples, always along for *filled bins* and probability estimates, even outside the given data. Although we subscribe to the benefits of such approximation, it does means that any particular distribution will just expand to incorporate new data, necessitating careful sampling and discrete class definitions. As detailed in section 3.2.2, we employ Scotts rule of thumb to select the bandwidth size, and mention its importance to the estimate made, but we do not evaluate the approximation of distributions.

Direct comparison of our results to those found in García-Mateos et al. (2015) is not really appropriate as they are evaluating a binary classifier on its ability to segment vegetation from soil (very distinctive classes), whereas we are evaluating a non-binary classifiers ability to distinguish between *flower, veg,* and *soil* (with varying distinctiveness), along with the simple fact that our method of evaluation differ slightly as to incorporate the KDE function. Separability for example is evaluated as the area ratio of one particular class distributions against the collective distributions, which would make our results seem considerably lower if compared to results associated with a smaller collective distribution. Despite this, we have computed the same dimension sets shown in the respective paper (see appendix A and B) to

enable some comparison. Surprisingly, separability results for both approaches are fairly similar for channel 1, channel 3, and all channels (albeit considerably more overlap for the KDE-based approach for aforementioned reasons). The channel hue (HSV_0) performs consistently high across all implementations and channel sets. Generalizability also exhibits similar results (if only *veg* is evaluated) as their distribution of high accuracy overlap. Both approaches were successful in severely reducing class distribution overlap by employing various color transformations, implying that both class definitions exhibit some distinctiveness within each other. Both resulting classifiers however are naïve and just expands any distribution if given new data, meaning that users themselves have to ensure distinct class definitions and non-faulty sampling.

5.4 Modelling and classification

Our LSMSS-based approach incorporates semantic knowledge by characterizing plant objects (grouped instances of *veg* and *flower*) by their morphological features (table 6), but we do not explicitly acknowledge scale or object hierarchy. All class segments are namely built up simultaneously according to **one** set criterion of similarity (i.e. homogeneity), potentially limiting the segment size of one class by the spectral boundaries of another. Class hierarchy is only exploited once with the filtering of the class objects (section 3.2.6), with the removal of *flower* objects outside *veg* objects. Both *texture* and *structure* feature sets were computed using GLCM-based texture analysis, employing 1 and 5 pixel distance strata, as to capture subtle omnidirectional patterns. Although figure 24 does not allow valid comparison (individual feature importance is simply too low), a pixel distance of 1 is arguably too small given the inherent heterogeneity of VHR imagery, potentially oversaturating metrics.

The collective of these features, representing both healthy and diseased plants, is used to model said disease by employing random forest. Although we were initially simply following the paper by Feng et al. (2015) which proposed its use, CART (classification and regression trees) models are increasingly being used in conjunction with object-oriented methods as shown by Blaschke (2010), thus validating its use. Some notable examples of OBIA + CART use cases as mentioned by Blaschke include: extraction of forest inventory information (Chubey et al., 2006), and land cover mapping (Laliberte et al., 2007).

Going by the model results alone, we were indeed able to model disease in potato plants strictly by their morphological features as they performed above expectation (if normalized, averages 0.8 across all metrics). Variable importance however exhibited considerably low values across all features (figure 24), meaning that although their collective was proven to hold information (table 7), there are no individual features that were particularly expressive. We would argue that this is a direct result of the sheer subtlety of pathogen traits which varied greatly in scale, severity and visibility not only per pathogen but also per host (Sankaran etal., 2010). This was preluded to as most plants said to be diseased by plant experts barely exhibited clear traits, at least not during empirically evaluation of the VHR imagery, which is why we cannot guarantee the status of plants that were marked as healthy.

6. Conclusions and recommendations

The most important conclusion here is that the proposed method of disease classification, employing only morphological features computed from object-based approximations of potato plants given only UAV-based VHR RGB imagery, was proven to be possible, and potentially viable for crop management if improved in a few key areas. We were able to come to this conclusion by answering the following research questions, and by suggesting the following recommendations critical to any potential follow-up research:

1. What are the morphological signs of Erwinia and PVY^{NTN} infection in potato plants?

Summarizing our findings, the majority of the pathogen traits express themselves as discoloration on the leaf level, which is ultimately a function of *structure* and *texture* (at different scales). This is because necrosis, mosaics, and stipple-streak all affect the appearance, and consistency of the leaf surface (and collectively of the canopy) by introducing new patterns or colors (or by deteriorating existing ones). Mottling and desiccation also affect the feel (or structure) of the surface. Arguably the most palpable morphological sign however is growth stunting, which can be exhibited in a myriad of plant components, is most commonly induced by *blackleg* courtesy of the Erwinia pathogen, and can primarily be expressed as functions of *size* and *shape.*

2. *What feature space transformations of UAV-based orthophotos (RGB) can improve the separability of classes found in potato crop site scenes?*

Employment of higher dimensions overall have shown to improve the separability of the classes under consideration, but the most separable dimensions set is [DSM_0, EXB_0, $I23_2$, HSV_1, RGB_2] representing height, blue-ness, green-ness, and saturation respectively. We must agree however with the statement of Garcia (2015) which emphasized that separable models do not necessarily equate to generalizable models (i.e. perform well on new data), as the results would imply a bias towards *veg* and flower. Our most general, *and* separable was ultimately determined to be [DSM_0, HLS_2, HSV_0, LUV_1, LUV_2] representing height, saturation, hue, large/medium wavelengths, and luminance respectively, as it averaged around 0.5 weighted MCC across all classes.

3. What morphological features, derived from object-based approximations of potato plant, best detect the presence and severity of Erwinia or PVYNTN infection?

We would argue that the spatial resolution employed by our VHR imagery is sufficient to capture most traits, and we also subscribe to the statement made by Lang (2010), which suggested that segmentation is an efficient means of aggregating high levels of detail, as empirical and validated evaluation of our class objects confirms accurate segmentation. It is also worth emphasizing that no distinction was made in the model by plant type, disease type or severity of said disease. All of this leads us to conclude that our model indeed captures potato plant traits indicative of disease (as our collective feature sets appears to exhibit deterministic properties, as it scores 0.47 and 0.73 in MCC and F1 respectively) but also that individual feature importance varies considerably not only due to the sheer subtlety of the traits, but also due to the varying properties of the given potato plants themselves (that vary in plant type, disease type, and sheer uncertainty of the status as healthy plant).

As such, although the 7 most expressive features were clearly all associated with the DSM making *size* and *structure* the the most important morphological features, the low variable importance overall increases the uncertainty of this statement.

We would argue that the ambiguity in both the explanatory and response variables could be minimized by incorporating more semantic knowledge that characterizes the healthy and diseased potato plants, which can be done with hyperspectral imagery. We would also suggest an object-oriented approach that employs texture and shape in every iteration, as to better exploit empirical knowledge (would require robust sampling, class modelling, knowledge organization systems). Lastly, ambiguous class boundaries and other geometric problems prevent automatically updated geo-spatial information following this approach. We would suggest further research on fuzzy classification and the estimation of plant location by which one could approximate said boundaries (see Prokop & Reeves, 1992, and Sebari & He, 2013)

Bibliography

Atli Benediktsson, J., Chaunussot, J., & M. Moon, W. (2013). Advances in Very-High-Resolution Remote Sensing. *Proceedings of the IEEE*, *101*(3), 566–569. http://doi.org/10.1109/JPROC.2012.2237076

Barbedo, J. G. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. *SpringerPlus*, *2*(1), 660–671. http://doi.org/10.1186/2193-1801-2-660

Blancard, Dominique; Laterrot, Henri; Marchoux, Georges; Candresse, T. (2012). Diagnosis of Parasitic and Nonparasitic Diseases. In *Tomato Diseases* (pp. 35–411). http://doi.org/10.1016/B978-0-12-387737-6.50002-9

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16. http://doi.org/10.1016/j.isprsjprs.2009.06.004

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., … Tiede, D. (2014). Geographic Object-Based Image Analysis - Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, *87*, 180–191. http://doi.org/10.1016/j.isprsjprs.2013.09.014

Burton, W. G. (1989). *The Potato*. Longman Singapore Publishers (Pte) Ltd.

Canadian Food Inspection Agency. (2015). The Biology of Solanum tuberosum (L.) (Potatoes). Retrieved November 6, 2018, from http://www.inspection.gc.ca/plants/plants-with-novel-traits/applicants/directive-94-08/biology-documents/solanum-tuberosum-l-/eng/1330982063974/1330982145930

Celis-Gamboa, C. (2002). *The life cycle of the potato (Solanum tuberosum L.): from crop physiology to genetics*. *Plant Breeding*.

Chubey, M. S., Franklin, S. E., & Wulder, M. A. (2006). Object-based Analysis of Ikonos-2 Imagery for Extraction of Forest Inventory Parameters. *Photogrammetric Engineering & Remote Sensing*, *72*(4), 383–394. http://doi.org/10.14358/PERS.72.4.383

Clewley, D., Bunting, P., Shepherd, J., Gillingham, S., Flood, N., Dymond, J., … Moghaddam, M. (2014). A python-based open source system for Geographic Object-Based Image Analysis (GEOBIA) utilizing raster attribute tables. *Remote Sensing*, *6*(7), 6111–6135. http://doi.org/10.3390/rs6076111

D. Opitz, S. B. (2010). Object recognition and image segmentation: the Feature Analyst® approach. In *Object-Based Image Analysis*.

Datiles, M. J. &, & Acevedo-Rodríguez, P. (2014). Solanum tuberosum (potato). Retrieved from https://www.cabi.org/isc/datasheet/50561

des Essarts, Y. R., Cigna, J., Quêtu-Laurent, A., Caron, A., Munier, E., Beury-Cirou, A., … Faure, D. (2016). Biocontrol of the potato blackleg and soft rot diseases caused by Dickeya dianthicola. *Applied and Environmental Microbiology*, *82*(1), 268–278.

http://doi.org/10.1128/AEM.02525-15

E. Weinke, S. Lang, M. P. (2010). Strategies for semi-automated habitat delineation and spatial change assessment in an Alpine environment. In *Object-Based Image Analysis*.

Feng, Q., Liu, J., & Gong, J. (2015). UAV Remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sensing*, *7*(1), 1074–1094. http://doi.org/10.3390/rs70101074

Food and Agriculture, & Organization of the United Nations; Earthscan. (2011). *The state of the world's land and water resources for food and agriculture (SOLAW) – Managing systems at risk*. *The State of the World's land and water resources for Food and Agriculture. Managing systems at risk.* Rome; London.

Free Software Foundation. (2014). Various Licenses and Comments about Them. Retrieved September 5, 2018, from https://www.gnu.org/licenses/license-list.html

García-Mateos, G., Hernández-Hernández, J. L., Escarabajal-Henarejos, D., Jaén-Terrones, S., & Molina-Martínez, J. M. (2015). Study and comparison of color models for automatic image analysis in irrigation management applications. *Agricultural Water Management*, *151*, 158–166. http://doi.org/10.1016/j.agwat.2014.08.010

Gartner, G., Meng, L., & Peterson, M. P. (2010a). *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*. *Technology*. http://doi.org/10.1007/978-3-540-88183-4

Gartner, G., Meng, L., & Peterson, M. P. (2010b). *Object-based Image Analysis*. *Technology*. http://doi.org/10.1007/978-3-540-88183-4

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

Haralick, R. M., & Shapiro, L. G. (1985). Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, *29*(1), 100–132. http://doi.org/10.1016/S0734-189X(85)90153-7

Haralick, R., Shanmugan, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*. http://doi.org/10.1109/TSMC.1973.4309314

Kamp, J., Blok, P., Polder, G., van der Wolf, J., & Jalink, H. (2016). Smart Ziekzoeker 2015 : detectie van virus- en bacteriezieke pootaardappelen met behulp van vision- en sensortechnologie. Retrieved from http://edepot.wur.nl/385490

Knoth, C., & Nüst, D. (2016). Enabling Reproducible Obia With Open-Source Software in Docker Containers. *GEOBIA 2016 : Solutions and Synergies*.

Koehrsen, W. (2018). Hyperparameter Tuning the Random Forest in Python. Retrieved June 6, 2018, from https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74

Laliberte, A. S., Fredrickson, E. L., & Rango, A. (2007). Combining Decision Trees with Hierarchical Object-oriented Image Analysis for Mapping Arid Rangelands. *Photogrammetric Engineering & Remote Sensing*, *73*(2), 197–207. http://doi.org/10.14358/PERS.73.2.197

Landgrebe, D. A. (1999). Some fundamentals and methods for hyperspectral image data analysis . *BiOS'99 International Biomedical Optics Symposium*, 104–113.

Lang, S. (2010). Object-based image analysis for remote sensing applications: modeling reality – dealing with complexity. In *Object-Based Image Analysis*.

Li, L., Zhang, Q., & Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors (Switzerland)*, *14*(11), 20078–20111. http://doi.org/10.3390/s141120078

Liu, P. (2017). *The future of food and agriculture: Trends and challenges. Fao*. http://doi.org/ISBN 978-92-5-109551-5

M. Ehlers, D. T. (2010). On segment based image fusion. In *Object-Based Image Analysis*.

Mahlein, A.-K. (2016). Present and Future Trends in Plant Disease Detection. *Plant Disease*, *100*(2), 1–11. http://doi.org/10.1007/s13398-014-0173-7.2

Maillot, N., Thonnat, M., & Boucher, A. (2004). Towards ontology-based cognitive vision. *Machine Vision and Applications*, *16*(1), 33–40. http://doi.org/10.1007/s00138-004-0142-9

Michel, J., Youssefi, D., & Grizonnet, M. (2015). Stable mean-shift algorithm and its application to the segmentation of arbitrarily large remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(2), 952–964. http://doi.org/10.1109/TGRS.2014.2330857

Nolte, P., Alvarez, J. M., & Whitworth, J. L. (2002). Potato Virus Y Management for the Seed Potato Producer. *University of Idaho Extension*, (CIS 1165), 8.

OpenCV. (2018). Shape descriptors. Retrieved June 6, 2018, from https://docs.opencv.org/3.4/d1/d32/tutorial_py_contour_properties.html

Pascale, D. (2003). A review of RGB color spaces... from xyY to R'G'B'. *Babel Color*, 1–35. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Review+of+RGB+color+spaces...+from+xyY+to+R'G'B#0

Pérombelon, M. C. M. (2002). Potato diseases caused by soft rot Erwinia - an overview of pathogenesis. *Plant Pathology*, *51*(1), 1–12. http://doi.org/10.1046/J.0032-0862.2001.SHORTTITLE.DOC.X

Prokop, R. J., & Reeves, A. P. (1992). A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP: Graphical Models and Image Processing*, *54*(5), 438–460. http://doi.org/10.1016/1049-9652(92)90027-U

Rich, A. E. (1983). *Potato diseases* (United Kin). London: ACADEMIC PRESS INC. (LONDON) LTD.

Sandwell, D. (2004). *Use of color in remote sensing*.

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. http://doi.org/10.1016/j.compag.2010.02.007

Savary, S., Ficke, A., Aubertot, J. N., & Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. *Food Security*, *4*(4), 519–537. http://doi.org/10.1007/s12571-012-0200-5

Scipy. (2013). Reference Guide: scipy.stats.gaussian_kde. Retrieved November 6, 2018, from https://docs.scipy.org/doc/scipy-0.13.0/reference/generated/scipy.stats.gaussian_kde.html

Scott, D. W. (2010). Scott's rule. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 497–502. http://doi.org/10.1002/wics.103

Scott, D. W. . H.-U. B. C. for A. S. and E. (CASE). (2004). Multivariate Density Estimation and Visualization, (16).

Sebari, I., & He, D. C. (2013). Automatic fuzzy object-based analysis of VHSR images for urban objects extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, *79*, 171–184. http://doi.org/10.1016/j.isprsjprs.2013.02.006

Shih, P., & Liu, C. (2005). Comparative assessment of content-based face image retrieval in different color spaces. *International Journal of Pattern Recognition and Artificial Intelligence*, *19*(7), 873–893.

Singh, R. P., Valkonen, J. P. T., Gray, S. M., Boonham, N., Jones, R. A. C., Kerlan, C., & Schubert, J. (2008). Discussion paper: The naming of Potato virus Y strains infecting potato. *Archives of Virology*, *153*(1), 1–13. http://doi.org/10.1007/s00705-007-1059-1

Su, W., Li, J., Chen, Y., Liu, Z., Zhang, J., Low, T. M., … Hashim, S. A. M. (2008). Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery. *International Journal of Remote Sensing*, *29*(11), 3105–3117. http://doi.org/10.1080/01431160701469016

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. http://doi.org/10.3390/rs61111013

Toth, I. K., Bell, K. S., Holeva, M. C., & Birch, P. R. J. (2003). Soft rot erwiniae: From genes to genomes. *Molecular Plant Pathology*, *4*(1), 17–30. http://doi.org/10.1046/j.1364-3703.2003.00149.x

Toth, I. K., van der Wolf, J. M., Saddler, G., Lojkowska, E., Hélias, V., Pirhonen, M., … Elphinstone, J. G. (2011). Dickeya species: An emerging problem for potato production in Europe. *Plant Pathology*, *60*(3), 385–399. http://doi.org/10.1111/j.1365-3059.2011.02427.x

Weih, R. C., & Riggan, N. D. (2010). Object-based classification vs. pixel-based classification: Comparitive importance of multi-resolution imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(4), 4–7. Retrieved from http://www.isprs.org/proceedings/xxxviii/4-c7/pdf/Weih_81.pdf

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, *2017*. http://doi.org/10.1155/2017/1353691

Zheng, L., Zhang, J., & Wang, Q. (2009). Mean-shift-based color segmentation of images containing green vegetation. *Computers and Electronics in Agriculture*, *65*(1), 93–98. http://doi.org/10.1016/j.compag.2008.08.002

Appendices

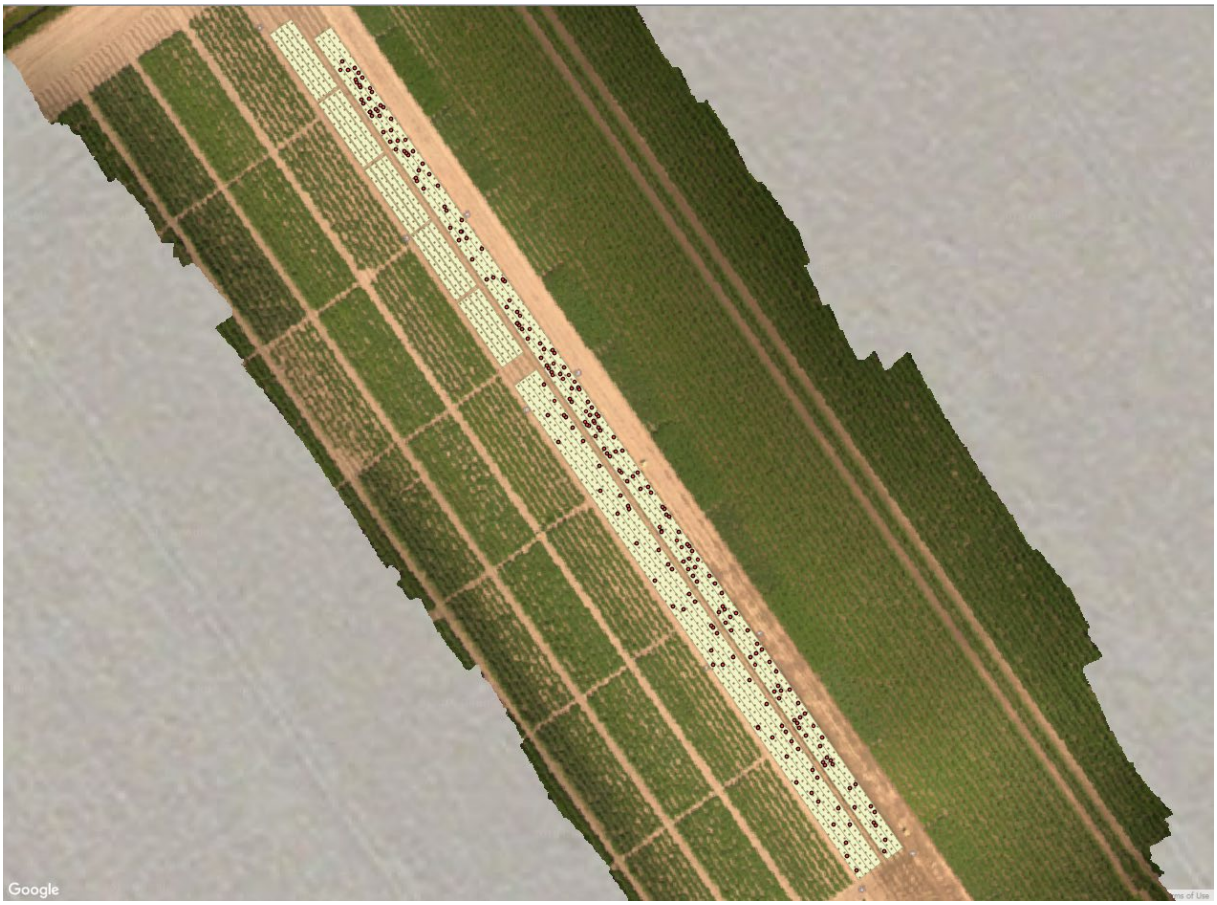
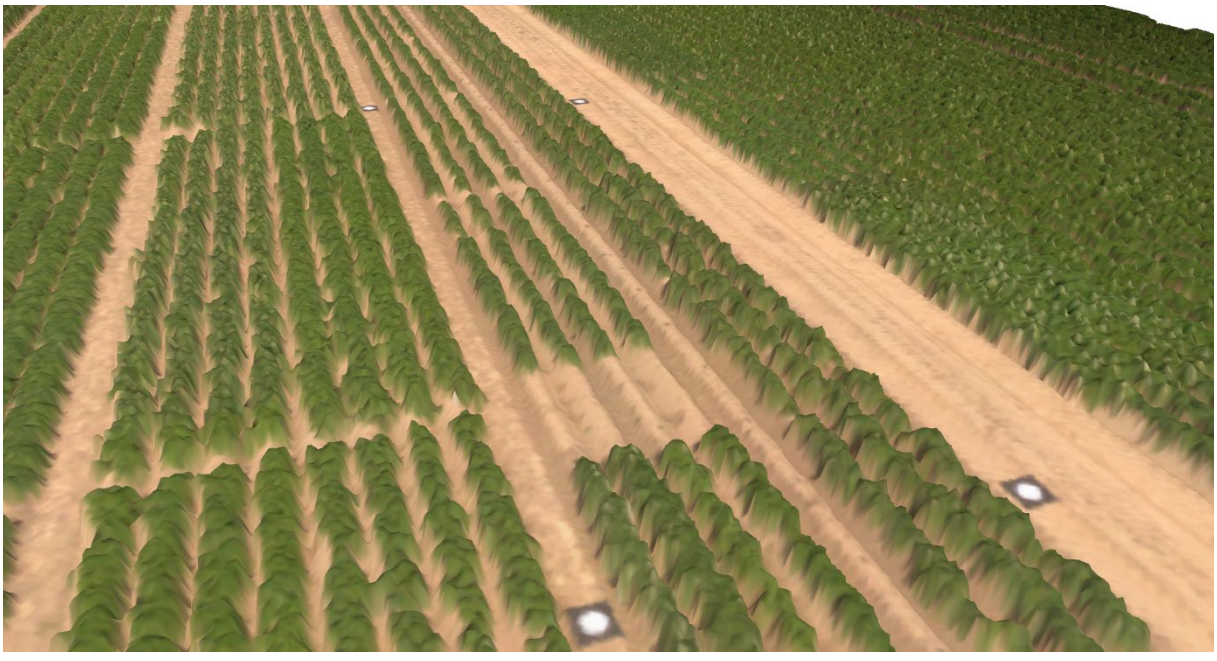
A. Separability comparison

Model	Chan.#1	Chan.#2	Chan.#3	Chan.#1-#2	Chan.#1-#3	Chan.#2-#3	All channels	
LUV	0.356044	0.348013	0.311733	0.373353	0.318529	**0.28199**	0.294739	**Veg overlap**
HSV	**0.2583**	**0.308416**	0.340745	**0.178716**	**0.30047**	0.290433	**0.206298**	
LAB	0.354765	0.321276	0.331924	0.341498	0.335619	0.302635	0.318021	
HLS	0.260605	0.361779	0.437493	0.313156	0.414402	0.437621	0.385857	
XYZ	0.341385	0.354616	0.363617	0.350655	0.360874	0.331767	0.319732	
RGB	0.340106	0.371957	0.362438	0.377744	0.374098	0.319668	0.307408	
YUV	0.353062	0.332922	0.351676	0.335647	0.385541	0.307897	0.324721	
I23	0.349156	0.345894	**0.278775**	0.378145	0.333828	0.31594	0.317945	
LUV	0.105869	0.06792	0.223198	0.031055	0.076169	0.045588	0.022617	**Minimum overlap**
HSV	0.048169	0.165481	0.08293	0.038839	**0.02515**	0.069983	**0.019211**	
LAB	0.104643	**0.052779**	0.209874	0.029187	0.084511	0.046017	0.022529	
HLS	**0.047979**	0.082424	0.142267	**0.027225**	0.042335	0.070771	0.027745	
XYZ	0.080523	0.10639	0.100783	0.028592	0.068168	0.082899	0.02256	
RGB	0.073782	0.132572	0.100283	0.027817	0.062545	0.075424	0.022621	
YUV	0.105645	0.269993	**0.064516**	0.081667	0.029117	0.04148	0.022804	
I23	0.089726	0.146361	0.06518	0.067855	0.032783	**0.041291**	0.022068	
LUV	0.88	**0.02**	0.32	0.011	0.1	0.0063	0.0037	**Results (García)**
HSV	**0.021**	0.41	0.91	0.014	0.01	0.27	**0.0031**	
LAB	0.88	0.0095	0.28	0.0048	0.18	0.0058	0.0035	
HLS	**0.021**	0.88	0.4	0.012	0.01	0.27	0.0032	
XYZ	0.91	0.9	0.79	**0.0051**	0.48	0.18	0.0036	
RGB	0.75	0.79	0.73	0.009	0.63	0.077	0.0031	
YUV	0.93	0.4	0.02	0.25	0.016	0.0046	0.0032	
I23	0.9	0.8	**0.017**	0.71	**0.007**	**0.0044**	0.0032	

B. Accuracy comparison

Model	Chan.#1	Chan.#2	Chan.#3	Chan.#1-#2	Chan.#1-#3	Chan.#2-#3	All channels	
LUV	51.85	97.61	84.89	97.86	92.44	98.73	98.59	**Results (García)**
HSV	**96.18**	80.68	49.56	96.84	97.88	82.73	98.86	
LAB	52.11	**99.2**	85.72	**99.08**	88.28	**99**	**98.89**	
HLS	96.18	56.2	79.88	97.23	97.83	82.72	98.86	
XYZ	53.37	50.9	66.84	98.94	76.45	88.39	98.27	
RGB	60.57	60.09	70.11	98.22	69.05	93.52	98.48	
YUV	50.75	81	96.05	85.35	96.51	98.93	98.49	
I23	57.22	53.57	**98.71**	57.2	**98.86**	98.85	98.67	
LUV	55.84	57.79	47.04	61.2	57.68	58.58	61.78	**Accuracy (mean)**
HSV	57.62	46.26	57.5	59.68	**61.61**	58.86	**61.98**	
LAB	55.84	56.54	44.98	61.47	56.72	58.29	61.78	
HLS	57.62	**58.04**	56.01	**61.66**	60.47	**59.31**	62.38	
XYZ	57.81	55.84	57.22	61.47	58.85	56.72	61.78	
RGB	**58.24**	51.48	57.39	61.5	59.8	58.48	61.78	
YUV	55.84	37.1	**57.83**	56.83	61.3	57.76	61.78	
I23	57.13	55.61	55.94	58.92	61.4	57.76	61.78	
LUV	80.38	**86.68**	61.87	89.98	82.18	**88.86**	91.03	**Accuracy (veg)**
HSV	84.92	57.91	**84.79**	88.27	**90.88**	84.57	91.11	
LAB	80.38	82.77	56.73	**90.92**	80.01	85.81	91.03	
HLS	84.92	84.62	78.51	90.81	89.62	84.92	**91.56**	
XYZ	84.96	80.38	81.42	**90.92**	84.62	80.01	91.03	
RGB	**86.27**	70.34	81.75	90.76	86.74	84.03	91.03	
YUV	80.38	41.67	86.42	80.17	90.2	85.98	91.03	
I23	82.8	78.35	80.79	84.41	90.7	85.98	91.03	

C. Python scripts

https://github.com/jaspersiebring/disease_detection_OBIA

D. Various imagery