University of Zurich[UZH]

Assessing Machine Learning Methods to Detect Bois Noir in Vineyards - Towards a systematic monitoring approach with spectroradiometry and RPAS data to reduce farmers' manual detection time

GEO 511 Master's Thesis

Author
Jeremiah Huggel
14-612-949

Supervised by
Prof. Dr. Meredith Schuman

Faculty representative
Prof. Dr. Meredith Schuman

30.08.2024
Department of Geography, University of Zurich

University of Zurich[UZH]

GEO 511
Master Thesis
30. August 2024

Assessing Machine Learning Methods to Detect Bois Noir in Vineyards

Towards a systematic monitoring approach with spectroradiometry and RPAS data to reduce farmers' manual detection time

Jeremiah HUGGEL
14-612-949

Supervised by:
Prof. Dr. Meredith Christine SCHUMAN

Faculty representative:
Prof. Dr. Meredith Christine SCHUMAN

Remote Sensing Laboratories
Department of Geography
University of Zurich

Abstract

Remotely piloted aircraft systems (RPAS), also known as unmanned aerial vehicles (UAVs) are used more and more in every part of life, from movie cameras, to package delivery systems, and to monitor plant health. Their use allows for more consistent and reproducible results in all these sectors. Bois noir (BN) and flavescence dorée (FD) are two grapevine yellowing diseases which reduce yields for vine growers and increase management time for their plots by requiring manual monitoring of each plant and removal of infected plants. Using RPAS to monitor BN and FD would reduce labour and allow for more targeted intervention when a disease is detected. The aim of this thesis is to examine the relationship between phenotypic, imageable symptoms and the infection of BN to inform future monitoring systems of BN and FD. To identify the most relevant wavelengths in detection of BN, field spectroscopy measurements were taken which span from the visible (VIS) to the near- and shortwave-infrared wavelengths (NIR/SWIR). Additionally, RPAS images covering the VIS and NIR regions were taken from the air to assess feasibility of transferring results of the spectroscopy to the RPAS. From the spectroscopy measurements, the most relevant wavelengths were calculated using a recursive feature elimination (RFE) technique based on a support vector machine (SVM) model. Other models were used in tandem to evaluate the consistency of the results. To evaluate the effectiveness of the feature selection and to see how accurately infected plants can be identified, machine learning (ML) models were employed. This thesis is part of the project SmartGrape and acts as a pilot study to inform later stages of the project. There was only limited data available from the first year of collection, so comparisons to literature were made to verify the feasibility of the results. Wavelengths relating to pigments (520 and 650 nm), photosynthetic activity like fluorescence (710 nm), and non-pigment dry matter and water content of the leaf (1500-1600 and 2000+ nm) were identified to be the main differentiators when trying to classify infected and non-infected leaves. Affordable, commercially available RPAS-based sensors can be used to monitor the VIS wavelengths (520, 650, 710 nm), but not those in the SWIR. This means that classification is possible, but with less accuracy than if the higher wavelengths were included. In the future, a sensor covering higher wavelengths could help improve the *in-situ* detection of BN and FD.

Contents

List of Figures

List of Tables

1 Introduction

Phytoplasma yellowing diseases like bois noir (BN) and flavescence dorée (FD) are a threat to viticulture, resulting in loss of yield and possibly whole plants (Agroscope, 2024; Bertaccini et al., 2014). They are most easily spotted in plants towards the end of the growing season. Both are spread by cicada vectors; BN by *Hylasthes obsoletus* (Kehrli et al., 2010; Kessler et al., 2011; Maixner et al., 2014) and FD by *Scaphoideus titanus* (Chuche & Thiéry, 2014; Galetto et al., 2005). Flavescence dorée is known to be spread amongst vineyards in the west of Switzerland and will, probably in the near future, spread towards the east (Kehrli et al., 2009; Rizzoli et al., 2023). FD follows similar spreading as BN, which also started in the west of Switzerland a few decades ago and in recent years, was found around lake Zurich and in 2023, was newly confirmed in Grisons (Agroscope, 2024).

FD is a disease which can spread between plants, which means that if it is found within a vineyard, the plants have to be quarantined and eliminated. BN does not spread between plants and thus the measures against it are not as drastic (Dubuis et al., 2023). Therefore, it is possible to study BN in the field, while it is much harder to do so with FD. Early detection of BN and FD are in the interest of vine growers, as plants that are infected with FD need to be quarantined and removed, while plants that are infected with BN have less yield. Since the diseases are phenotypically similar (Agroscope, 2024; Tessitori et al., 2018), it should be possible to take conclusions drawn from BN to apply to FD (similar to Bendel et al., 2020). Ideally, they could be differentiated to monitor the spread of FD and protect as well as possible from it in a timely manner.

At the time of writing, there is no standardized way to monitor BN or FD within Switzerland. Farmers assess their plants manually and report diseases when they detect them (Agroscope, 2024). This is labour intensive and makes the assessment subjective. They also have many other diseases to worry about, like esca or mildew (Dubuis et al., 2023) and might not be sure how to distinguish them. With a more structured approach, the spread of the disease could be monitored and farmers in risk areas could be informed more directly and timely as to when they need the information.

Monitoring these diseases are ideally done with remotely piloted aircraft systems (RPAS, also known as unmanned aerial vehicles (UAVs) and drones), as to not disturb the plants and minimise human effort needed. This is an approach that is and will be used more and more often in different agricultural fields (Albetis et al., 2017; Cavender-Bares et al., 2020; Dehkordi et al., 2020; Petibon et al., 2021; Yang et al., 2017). Using a high-resolution spectrometer (also known as hyperspectral spectrometer) as an alternative to the RPAS for the analysis will help identify wavelengths needed for detection (Albetis et al., 2017; Backhaus et al., 2011; Bendel et al., 2020; Hueni et al., 2017),

since RPAS-mounted sensors have more limited spectral resolution. The high resolution promises to highlight features in the spectral response of the vines which can be used to differentiate between infected and non-infected vines (Al-Saddik et al., 2017). To differentiate the collected spectra between infected and non-infected plants, there are multiple methods that can be used (Kuhn & Johnson, 2013). Approaches range from simpler linear models and 'classical statistical' models (Chambers, 1992) to machine learning (ML) (Alloghani et al., 2020) or deep learning (DL) models (Audebert et al., 2019). Linear models use linear combinations of independent numeric or categorical variables to predict a numerical response variable (Chambers, 1992). They are usually used in simple modelling tasks where the input variables are not very complex. For binary classification tasks (i.e. infected vs. non-infected), a special case is used, the logistic regression model. In remote sensing it is mostly used for feature selection and classification (Cheng et al., 2006).

Machine learning and especially deep learning are currently used all around the world with "Artificial Intelligences" (AI) such as ChatGPT. Their potential to unearth patterns in high dimensional space which are not easily recognisable by humans make them powerful tools for classification tasks. But in order to use DL methods, a great amount of data is needed, generally referred to as big data (Najafabadi et al., 2015). This encompasses millions of datapoints and requires a great amount of them to be annotated (i.e. referenced as 'ground-truth') in order to develop a reliable model (P. Zhu et al., 2020). There is also the drawback, that most currently available DL models hide their structure and cannot be interpreted by humans anymore. There is a shift towards DL methods for research in remote sensing, but it is not yet a common approach, due to data availability, computational requirements and not knowing what decisions the model makes, i.e. black-box behaviour (X. X. Zhu et al., 2017).

Therefore, machine learning is still the current state of the art in remote sensing. The term machine learning is not new and was already well established in the 80s (Carbonell et al., 1983) and encompasses the notion of feeding training data to a statistical model which can then 'learn' from it and predict unseen data with the same variables. It has found wide use in the world of remote sensing (Al-Saddik et al., 2017; Bendel et al., 2020; Colkesen & Kavzoglu, 2016, to name a few) especially support vector machines (SVM) (Huang et al., 2002; Mountrakis et al., 2011; Rumpf et al., 2010) which work well with high dimensionality data, which high-resolution spectroscopy data is. Other ML models like random forest (RF), k-nearest neighbours (kNN) and radial basis function (RBF) are also often seen but generally work better with fewer dimensions (Pal, 2005). Fewer dimensions are generally preferred as high dimensional data takes a long time to process and contains a lot of redundancy due to correlations between wavelengths (H. Li et al., 2022). To this effect there are many methods that can be used and, in this thesis, recursive feature elimination (RFE) (Colkesen & Kavzoglu, 2016; Ramezan, 2022) based on support vector machines is used as a main method. For comparison, a random forest RFE (RF-RFE) (Demarchi et al., 2020), a genetic algorithm (GA) (Al-Saddik et al., 2017) and a random forest univariate filter (Albetis et al., 2017) are run as well. To support the findings, a completely new method to segment spectra based on their similarity, a hierarchical spectral clustering with parallel analysis (HSC-PA) (C. Li et al., 2023a) and a simple principal component analysis (PCA) (Wold et al., 1987) are run as well. Other papers have explored the space of dimensionality reduction for yellowing diseases,

and even though they use different methods, they all have one commonality: they discuss the fact that all grape varieties are different and each application may require differing wavelengths (Albetis et al., 2017; Al-Saddik et al., 2019b; Becker et al., 2021; Bendel et al., 2020). Most of the literature also uses laboratory setups for detection, so doing a field-based study brings its own chances. Therefore, exploring new methods and discussing them in the light of existing literature will bring new insights into the complex world of plants affected by yellowing diseases.

This thesis is part of the project SmartGrape led by the University of Zürich (UZH) Spatial Genetics group together with the ETH Environmental Robotics group, Agroscope, and the Weinbauzentrum Wädenswil. It is a three-year project to find suitable methods to survey bois noir and flavescence dorée in vineyards. One part of this project is detecting bois noir with RPAS or leaf spectroscopy and is where this thesis comes into play. The data used is from the first year of data collection and should help inform the procedure of future field seasons. The aim of this thesis is to help formulate a framework where farmers are supported in identifying and monitoring BN.

To this end, the following research questions are asked:

1. Can machine learning models be used to reliably identify bois noir in vineyards with Pinot noir and Zweigelt?
2. What wavelengths are ideal to identify bois noir in vineyards with compact remote sensing equipment?
3. Can a commercially available, multispectral RPAS like the Mavic 3M be used to detect bois noir-infected plants in an otherwise healthy field?
4. Are there early symptoms that can be detected before they present visually towards the end of the year?

2 Material and Methods

Parts of the text in this section is adapted from a co-authored publication in Acta Horticulturae ((Bertschinger et al., 2024); Experimental Setting, Bois Noir and Flavesence Dorée, RPAS, High resolution spectroradiometry).

2.1 Experimental setting

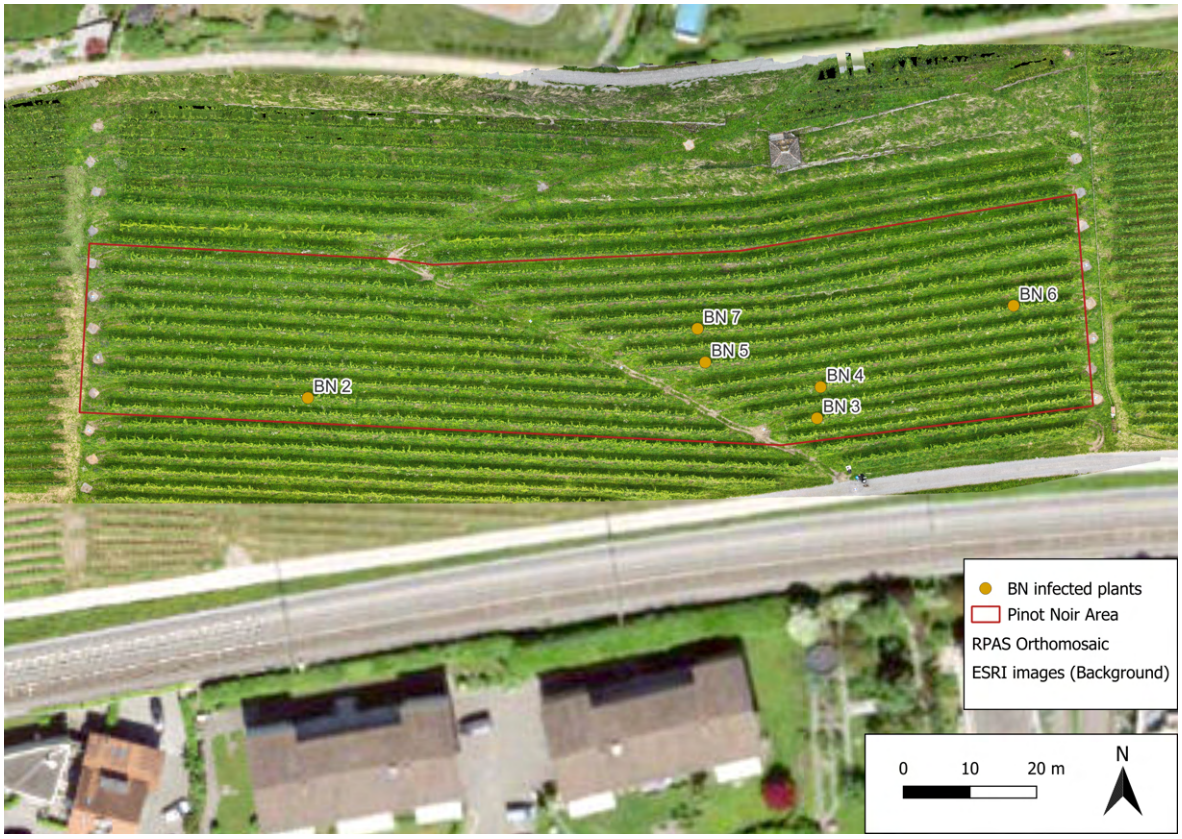
Two vineyards showing bois noir symptoms were selected for field experiments, located around Stäfa in the canton of Zürich, Switzerland. The vine growers kindly allowed to mark the selected, infected plants with metal tags and black-yellow tape at the stems so that they can be recognized throughout multiple years. In both vineyards data were collected by an ASD hand-held field spectrometer (FieldSpec4) and a DJI Mavic 3 multispectral (M3M), as well as a DJI Matrice 300, with a MicaSense RedEdge Dual sensor, both being remotely piloted aircraft systems (RPAS).

Those vineyards were: **Weber (WEB)** (Figure 2.1a); grape variety is Pinot noir, planted in 2006; 1634 vines; plant distance: 2.5 x 0.9 m; gross surface area: 4635 m^2, 7 infected vines. **Rütihof (RUT)** (Figure 2.1b); grape variety is Zweigelt, planted in 2009, cut back due to BN infection and regrafting with Zweigelt in 2019; 468 vines; plant distance: 2.6 x 1 m; gross surface: 2000 m^2, estimated 35% BN incidence.

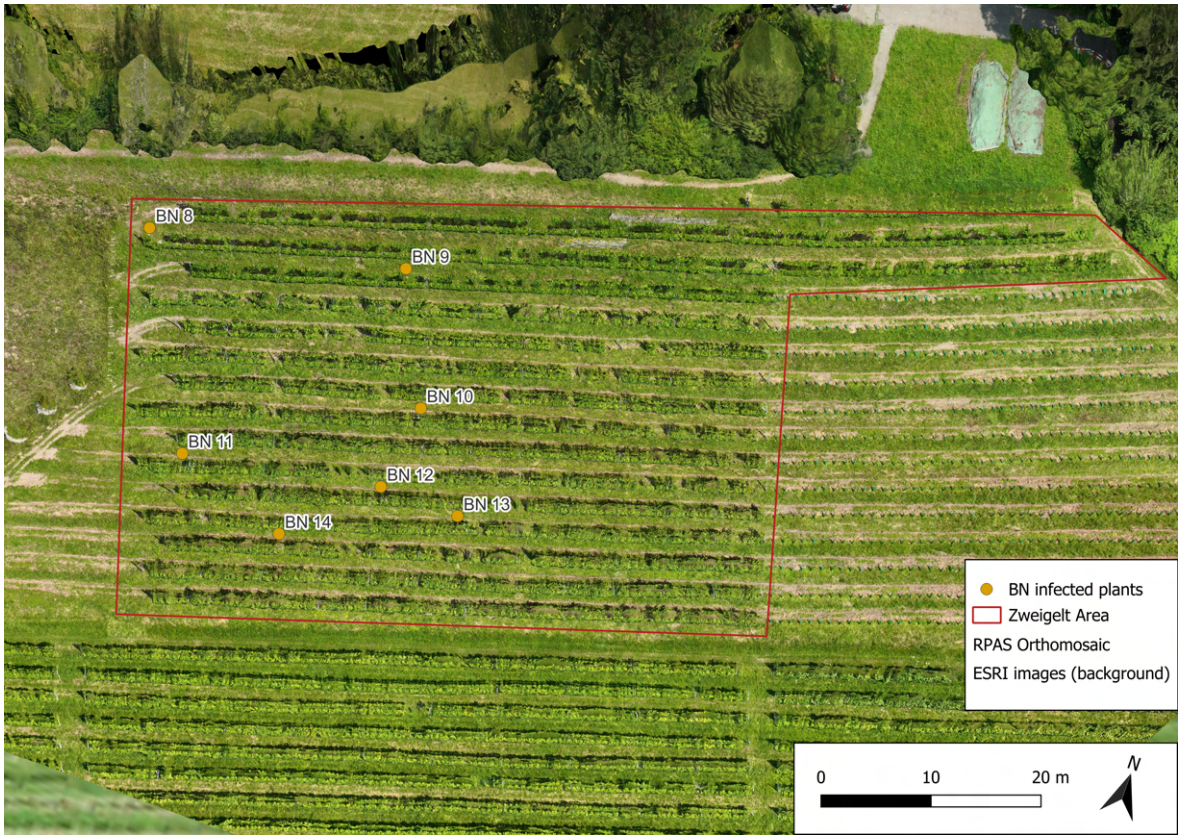
Dates	M3M	Matrice 300	FieldSpec WEB	FieldSpec RUT	qPCR
06.07.23					X
11.07.23	X	X	X		
27.07.23	X		X	X (only BN 8-10)	
08.08.23	X		X		
30.08.23			X	X	
01.09.23	X				
31.08.23					X
27.09.23	X	X			

TABLE 2.1: Overview of the measurement dates and when which data was collected. The first FieldSpec measurement day at Rütihof (RUT) was not completed due to technical difficulties.

The low incidence rate in Weber allowed to sample all the infected plants, with a control plant right next to it. Rütihof posed a bigger challenge, and it was decided to measure the same amount of plants as in Weber and picking plants that had an uninfected neighbour. The infectivity of the plants was tested using a qPCR-test (Pelletier et al., 2009), once at the beginning (July 6th, 2023) and once towards the end of the season (August 31st, 2023) (also see Table 2.1). Plants were pre-



(A) BN infected plants at Weber; grape variety Pinot noir; Orthomosaic from 08.08.2023



(B) Selected and labeled BN infected plants at Rütihof; grape variety Zweigelt; Orthomosaic from 27.06.2024

FIGURE 2.1: Overview of selected BN infected plants at both vineyards.

selected visually by an expert in the disease and leaf-tissue samples were taken for diagnosis by qPCR (Bertschinger et al., 2024). The presumed positive plants were individually tested, while the controls were bulk tested per vineyard. At Weber, one positive plant (BN 1) was tested negative by the qPCR method and subsequently excluded from the analysis and measurements.

2.2 Bois noir and flavesence dorée

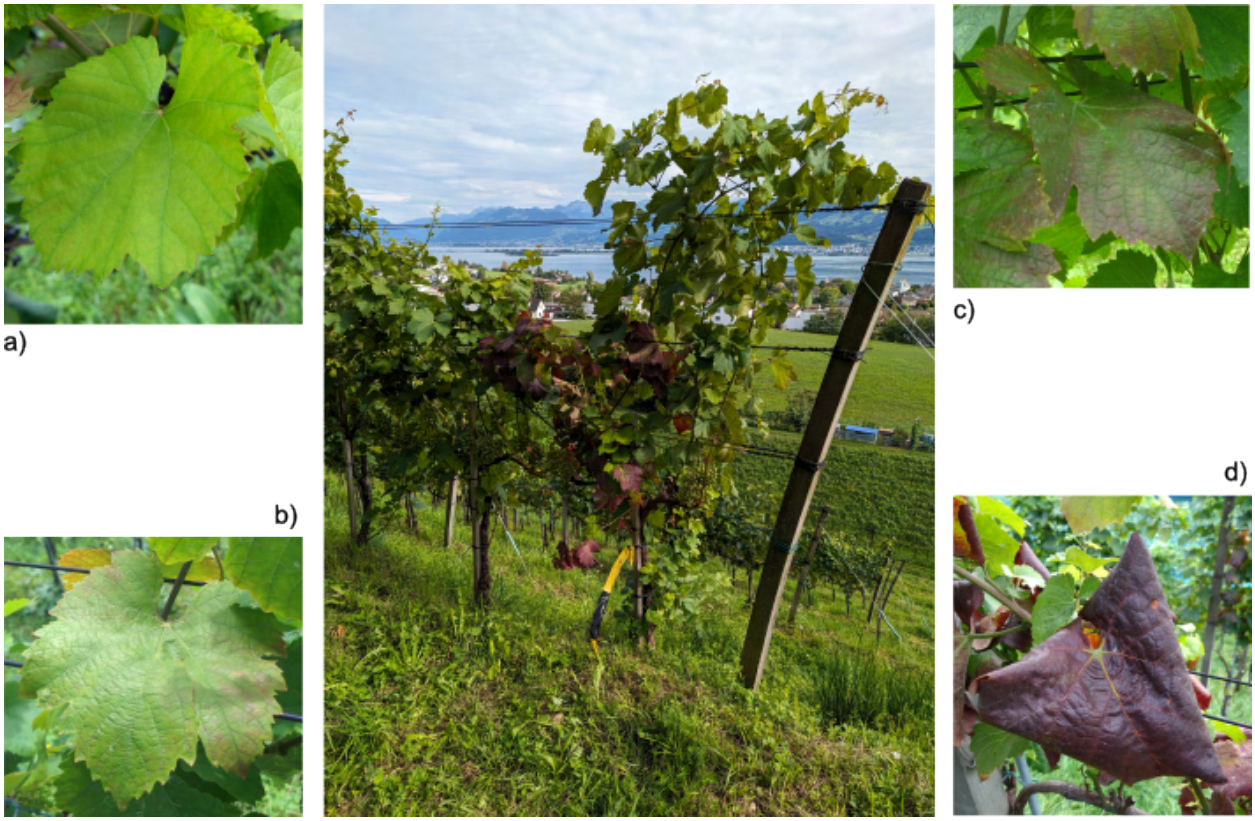


FIGURE 2.2: In the middle, a representative BN infected plant *in situ* at Rütihof (date: 08-08-2023, photographer: Jeremiah Huggel); a) a new, asymptomatic leaf; b) a mildly symptomatic leaf; c) a 'normal' symptomatic leaf; d) a heavily symptomatic leaf.

Bois noir (BN) is a yellowing disease of vines, which gets transferred by so-called vectors (an organism that transfers diseases) from its host plant. It is caused by *Candidatus Phytoplasma solani* (16SrXII-A), transferred from a non-grape host, usually stinging nettle (*Urtica dioica*) or bindweed (*Convolvulus arvensis*), to grapevine (*Vitis vinifera*) by *Hyalesthes obsoletus* Signoret (Hemiptera: Cixiidae) (Galetto et al., 2005; Kessler et al., 2011). Transfer occurs when an infected *H. obsoletus* feeds on grapevines because its favoured host plants have become unavailable. The current recommended, but not required, control measures for BN are to remove host plants prior to planthopper emergence from soil, and to remove symptomatic grapevines (Dubuis et al., 2021). There are no systematic measurements of the current spread and distribution of the insect vector, its host plants, or BN, because specific testing for the disease is limited to spot tests. Yet from visual assessment of symptoms, BN is considered a common problem in Swiss viticulture, and *H. obsoletus* and the BN phytoplasma are present in all viticultural regions of Switzerland (Dubuis

et al., 2021; Kehrli et al., 2009).

BN is visually identified by three symptoms: i) leaf discoloration and curling, ii) green rubbery wood on one or more canes, and iii) wilt and premature loss of grapes. However, there are alternative causes for these symptoms, which occur at a late stage of infection at the end of the season (August to October), and presence of BN must be confirmed by molecular analyses of symptomatic plants (Dubuis et al., 2021; Galetto et al., 2005). Diseased grapevines are thus difficult to identify, and their status can change from one year to the next. Varieties also differ in presentation of symptoms, e.g., Chardonnay and Pinot noir often show more distinct symptoms than Chasselas or Merlot (Rizzoli et al., 2023).

The planthopper vector of BN is difficult to detect and control. Controlling host plants of the pathogen and its vector, bindweed and stinging nettle, seems a more effective way to address the disease, but both plants are common in vineyards, also harbor beneficial insects, and are aggressively invasive (Kehrli et al., 2009).

BN is related to, and is often confused with, the quarantine disease flavescence dorée (FD), caused by *Candidatus Phytoplasma vitis*. FD spreads between grapevines via another cicada vector, the American grapevine leafhopper *Scaphoideus titanus* Ball (Hemiptera: Cicadellidae) (Chuche & Thiéry, 2014; Debonneville et al., 2022). *S. titanus* is associated primarily with grapevines, not necessarily with other host plants. FD is present in the cantons Ticino, Vaud, Wallis, Geneva and Grisons, with restricted distributions in the latter two. As a quarantine organism, FD must be reported to authorities, and controlled by the mandatory removal of affected grapevines upon discovery (Agroscope, 2024; 'SR 916.201 - Verordnung des WBF und des UVEK vom 14. November 2019 zur Pflanzengesundheitsverordnung (PGesV-WBF-UVEK)', n.d.). The visual symptoms of FD are the same as for BN and so BN and FD can only be differentiated by molecular testing.

There is no cure for either disease. Both are systemic diseases, i.e., once the plant is infected, it is not sufficient to cut back the symptomatic portion. Both diseases thus have long-lasting consequences and result in (possibly extensive) vine removal. It is of strategic importance to avoid their introduction into vineyards, to monitor systematically, recognize the disease early, and prevent spread. In regions where FD is not yet present, BN monitoring and prevention can help avoid costly activities necessitated by quarantine regulations and support early detection of FD spread to new areas. These aims would be supported by the establishment of automated monitoring platforms coupled to early detection methods.

2.3 Multispectral imaging

2.3.1 RPAS

For mapping purposes and multispectral data, the vineyards were imaged with two remotely piloted aircraft systems (RPAS, also known as unmanned arial vehicles (UAVs) or drones). One was a DJI Mavic 3 Multispectral (M3M) with images taken on July 11th, July 27th, August 8th, September 1st and September 27th 2023 (see Figure 2.1). For a larger selection of spectral bands

and a higher spatial resolution, a DJI Matrice 300, equipped with a MicaSense RedEdge-MX Dual multispectral camera was used to image on July 11th and September 27th. Both systems have downwelling light sensors. The M3M gimbal can be tilted and produces images with Real Time Kinematic (RTK) corrected geolocation information and is smaller and lighter (ca. 895 g versus 3.6 kg for the M300).

Sensor:	**Mavic 3 Multispectral**	**RedEdge-MX**	**RedEdge-MX blue**
Number of bands	4	5	5
Spectral wavelengths [nm]	Green (560 ± 16) Red (650 ± 16) Red Edge (730 ± 16) Near Infrared (860 ± 26)	Blue (475 ± 32) Green (560 ± 27) Red (668 ± 14) Red Edge (717 ± 12) Near Infrared (842 ± 57)	Coastal Blue (444 ± 28) Green (531 ± 14) Red (650 ± 16) Red Edge (705 ± 10) Red Edge (740 ± 18)
Image size [px]	2592 x 1944	1280 x 960	1280 x 960
Focal length [mm]	4.34	5.4	5.4
Field of view vertical	48.1°	35.4°	35.4°
Field of view horizontal	61.2°	47.2°	47.2°
Adaptive camera angle	Yes	No	No

TABLE 2.2: Wavelengths and other specifications of the three RPAS sensors; the later two are part of the MicaSense dual sensor setup on the DJI Matrice 300. A graphical representation can be found in Appendix Figure A.1.

For the project, it was initially planned to use the Matrice 300 for all field days and test whether the M3M performed similarly to detect bois noir, as the M3M would be much more user-friendly if the method were to be optimised for farmers. However due to availability constraints, an early estimation was made that the sensor differences between the two platforms should not influence the classification too much, simply since they cover roughly the same wavelengths. But as this could prove to be wrong, the bigger Matrice 300 was used twice, once for early- and once for late-stage symptom capture to be able to do a comparison if necessary.

2.3.2 RTK

In order to more accurately track the position of the RPAS, real time kinematic positioning (RTK) was used on both RPAS with the Swipos NTRIP service ('Swiss Positioning Service (Swipos)', 2024). RTK corrects common errors in global navigation satellite system (GNSS)-based geolocation by using a base-station which computes the error of its position calculated with the GNSS method through the phase of the carrier signal from the satellite and calculates the correction of this error. This correction is then transmitted to the rover (e.g. a RPAS) which uses it to improve it's own positional accuracy (van Diggelen, 1997). Swipos uses their own base-stations and interpolates the correction for the users position. A personal ground-based RTK base-station (Emlid Reach M2 with Swipos corrections) was also tested and used on the first field day, but the setup was found to be too time consuming to set up and did not increase the accuracy enough.

2.3.3 GCP

Ground control points (GCPs) were also laid out at the beginning of each field day, which allows to more accurately geolocate images e.g. when creating orthomosaics. They also help in stitching

the images together in post-processing as they are fixed points present in multiple images, helping the software know which photos overlap at which pixels. They were geolocated using a JAVAD Triumph LS+ system (JAVAD, San Jose, USA) for high positional accuracy. It was not possible to leave the GCPs between field days, as all slopes and vines in the vineyard were to be mowed and manipulated in some way by the vine growers during the field campaign.

2.3.4 Calibration

For radiometric calibration, a MicaSense calibration panel (number RP06-2229048-0B) was imaged before and after every flight. These images are used to calibrate the other images in processing later on and normalises the illumination condition such that they are more comparable between flights.

2.3.5 Flight planning software

To ensure that the flights are reproducible, a flight planning software was used. At first this was Drone Harmony and later UgCS (from August 8th). Drone Harmony is highly automated and does not permit changes to the route except by changing the survey area. The flight lines were sometimes uneven and the aircraft jerked around to accommodate the route, which was undesirable especially for the Matrice 300 which did not have a gimbal system for the camera. Being able to control factors like the camera tilt, the orientation of the aircraft and the orientation of flight lines ensures similar conditions between sites and therefore more comparable results, which was permitted by UgCS and why it was chosen.

The settings for the flights were: 65° camera angle from horizontal, camera/RPAS angled perpendicular to the rows of vines (i.e. yaw: 355 N (WEB) / 347 N (RUT)), 20 m flight height (terrain following, 2 m tolerance), 2 s interval between images, 3 m/s flight speed => 85% overlap of images on both axis. The camera angle was chosen to provide a side-view of the vines, since the vine rows are very narrow and images from a top-down view include more ground than vines. With a sideways-looking camera, the area of vines is much higher per image and allows to not only survey the top of the vines (which are often dominated by new leaves), but also the older and often more symptomatic leaves.

2.3.6 Data

Initial processing was performed using Pix4Dmapper software (Pix4D SA, Switzerland; educational license), in which the data was taken from the RPAS, read into Pix4Dmapper alongside calibration images and processed with the 'Ag Multispectral' template and standard settings except to output a Raster DSM as GeoTIFF with inverse distance weighting and an orthomosaic as GeoTIFF. This resulted in an output of a 3D voxel cloud, an orthomosaic and several indices which can also be calculated after the processing. An initial visual estimate of NDVI , GRVI and ARI indexes (see paragraph below and Results 3.5) did not depict any differences between infected plants and their neighbours, even with images taken very late in the season (27.09.2023). This was the driver to focus on determining relevant wavelengths of BN and therefore the potential of

the RPAS imagery. The orthomosaics resulting from Pix4D were used in combination with GNSS measurements of the infected vine stems to produce the maps in section 2.1.

The vegetation indices used in Pix4D are the normalised vegetation index (NDVI, Formula 2.1) (Tucker, 1979), one of the most used indices in the remote sensing community and often used in commercial products for RPAS in agriculture. The green red vegetation index (GRVI, Formula 2.2) (Motohka et al., 2010) has shown high potential to differentiate FD-symptomatic from asymptomatic vines in Albetis et al., 2017 and grapevine trunk disease in Albetis et al., 2019. Lastly, the anthocyanin reflectance index (ARI, Formula 2.3) (Gitelson et al., 2001) is used in Al-Saddik et al., 2017, 2019b as a good indicator to changing leaf structure in infected plants. Those are only a selection of vegetation indices. They were chosen as they stood out in the BN and grapevine yellowing disease literature as indices that work well for classification tasks, however they are often used in combination with other indices. This combination was not attempted in this thesis, as the focus was put on the spectroscopy data analysis. A proposed workflow for the RPAS imagery can be found in the outlook section 4.5.1.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{2.1}$$

$$GRVI = \frac{Green - Red}{Green + Red} \tag{2.2}$$

$$ARI = \frac{1}{Green} - \frac{1}{Red\text{-}edge} \tag{2.3}$$

2.4 Leaf spectroscopy

Specification	Value
Wavelength Range	350 nm - 2500 nm
Number of Sensors	3
Resolution VNIR	3 nm
Resolution SWIR	10 nm
Resampled spectral resolution	1 nm
Channels	2151
Field of view	25°

TABLE 2.3: Technical data of the FieldSpec 4 (adapted from Malvern Panalytical, 2023).

To identify spectral features distinguishing infected leaves, a field spectroradiometer (ASD FieldSpec 4 Standard-Res, S/N 18140, Bonsai Advanced Technologies, Carry Le Rouet, France; Table 2.3) with an ASD Plant Probe attachment (Bonsai Advanced Technologies) having a halogen light source was used. The measurements were taken against a black background (Labsphere, North Sutton, NH, USA), with five measurements per leaf. A white Leaf Clip sticker (ASD PN 145382, Bonsai Advanced Technologies) was used for optimization (gain and integration) and white reference (WR) measurements. The following settings were used for the FieldSpec: fiber optic, bare

fiber; number of samples per measurement: spectrum 25, dark current 100, white reference 25; integration time, 8.5 ms (workflow similar to (Al-Saddik et al., 2019a)).

For the data collection, a paired design was used, such that for every BN-positive vine that was measured, a BN-negative vine was measured immediately afterwards. Two to three leaves per infected plant, as well as of its neighbouring uninfected plants were measured. For the infected plants, leaves were classified as asymptomatic, mildly symptomatic, or heavily symptomatic, based on visible discoloration typical of BN, and leaves were measured in one to three of these categories for each infected plant depending on overall symptom development (i.e. which categories were present). For healthy neighbours, two similar leaves at a similar position on the vine as those measured on infected plants were measured. Leaves were measured with the plant probe facing the top (adaxial) side of the leaf, on the lamina between major veins. A white reference (WR) was measured alongside the internal calibration before the leaves were measured. The FieldSpec also performs a black reference measurement (black current) during the internal calibration.

2.4.1 Pre-processing and jump correction

Fieldspec data were curated, visualised and analysed using the R Statistical Software (v4.3.2; (R Core Team, 2023). All code can be found in the GitHub repository https://github.com/JereHu/Master_Thesis. For a full write-up of the packages and their versions used in this thesis, see "SessionInfo_R.txt" on the GitHub page.

The data were read from the raw ASD files with the package spectrolab (Meireles & Schweiger, 2023) as reflectance measurements. The reflectance is calculated as:

$$Reflectance = \frac{tar\_reflect(leaf)}{ref\_reflect(WR)} \tag{2.4}$$

There are other ways to calculate reflectance (e.g. C. Li et al., 2023b; Petibon et al., 2021), but there were no measurements conducted with a white background, as the aim is not to model leaf optical properties, but rather to do a standardised screening, for which the applied method should suffice.

Additionally, all wavelengths <400 nm were removed, as they are prone to have a bad signal to noise ratio (Cavender-Bares et al., 2016). There were also measurements in the data that were not used in this thesis (e.g. references or measurements using experimental settings), as well as invalid measurements (infinite reflectance values) and one symptomatic outlier, which did have a very different shape in the VIS region to the other spectra.

The FieldSpec uses three sensors to measure the whole spectra, which might lead to differences between the sensors, visible as breaks or jumps at 1000 nm and 1800 nm (e.g. spectra in Appendix A.2). These jumps are there because as the sensors warm up (even after an hour of pre-warming) they react differently to the change in temperature (see Hueni & Bialek, 2017) and become what we call miss-matched (see Figure 2.3). If the re-matching is done with physical models, as they are in the method proposed by Hueni and Bialek, 2017, then the sensor matching should not introduce

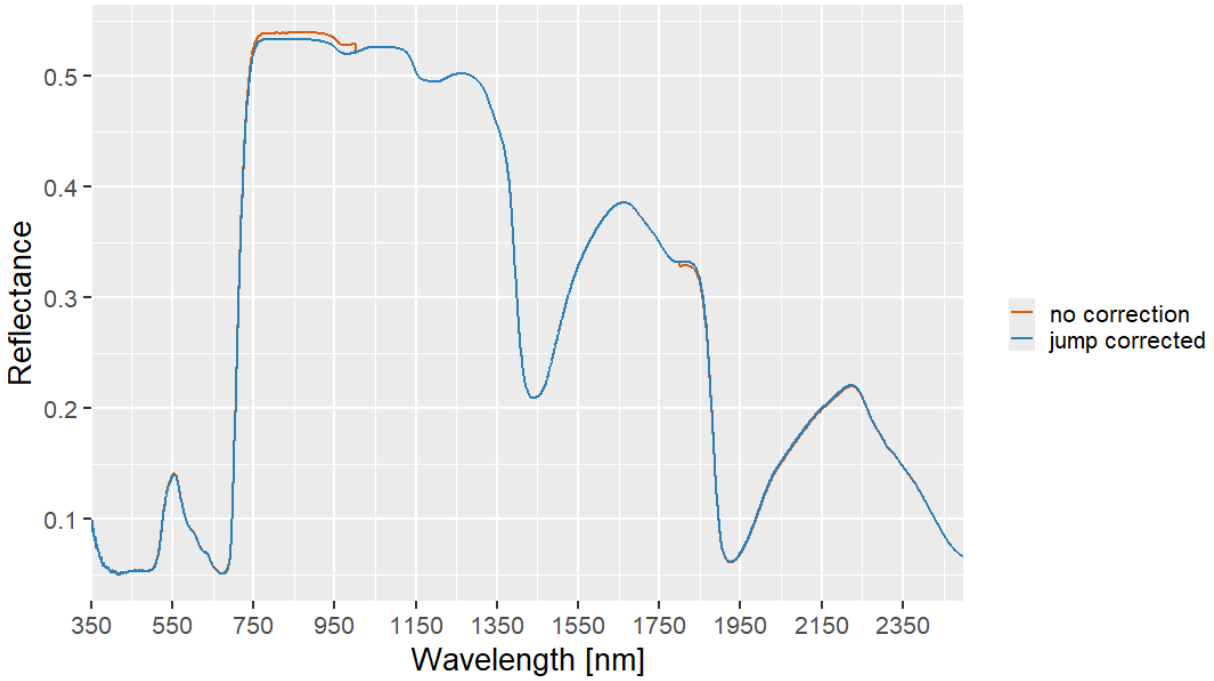


FIGURE 2.3: Visual demonstration of what jump correction from spectrolab does demonstrated on an example spectrum. The orange line is the original spectrum and the blue line is the corrected spectrum line.

artificial artefacts, but we refrain from implementing this method ourselves here.

There used to be a jump correction done with spectrolab, but due to missing understanding of how the method was implemented, it was left out. It is believed that they use a statistical method, which might introduce artefacts. Since we use a paired measurement setup, i.e. always measure an infected plant and then a control plant, any differences in the sensors (jumps) will be present in both measurements. They do vary over time, but with the paired setup, they vary about equally in both measurements. For the next year of measurements, it is made sure to systematically perform internal white-reference measurements, which mitigates these jumps by re-calibrating (setting the white reference) often.

2.4.2 Asymptomatic leaves

Asymptomatic leaves were those that belonged to infected plants but were not visibly symptomatic at the time of measurement. This meant that they looked like control plants. To determine whether there was a difference between control and asymptomatic leaves, a difference spectra was created to assess the differences in the mean on Weber data (see Figure 3.15 in section 3.6).

To be on the safe side it was decided to exclude the asymptomatic data points from the non-infected data. Another reason was that asymptomatic leaves might play an important role in early detection of the disease, as asymptomatic leaves might turn symptomatic later on and show a trend over time. Since measurements were not necessarily conducted on the same leaves over time, but just the same plant, this difference was not used in this thesis.

2.4.3 Data

After pre-processing, the data were averaged over the five samples collected per measurement and over the data per plant per day of measurement. This resulted in 26 control, 19 asymptomatic and 31 symptomatic data points available for the Weber vineyard (see Figure 2.4 and Appendix A.2) and 8 control, 9 asymptomatic and 17 symptomatic points for Rütihof. The symptomatic measurements can be differentiated into mildly symptomatic (n = 5 (WEB), n = 0 (RUT)), normal symptomatic (n = 18 (WEB), n = 10 (RUT)) and heavily symptomatic (n = 8 (WEB), n = 7 (RUT)). For most applications in this thesis, they were pooled together as symptomatic, as the target is to see if any symptomatic plants can be differentiated and it is assumed that their signals differ mostly in amplitude.

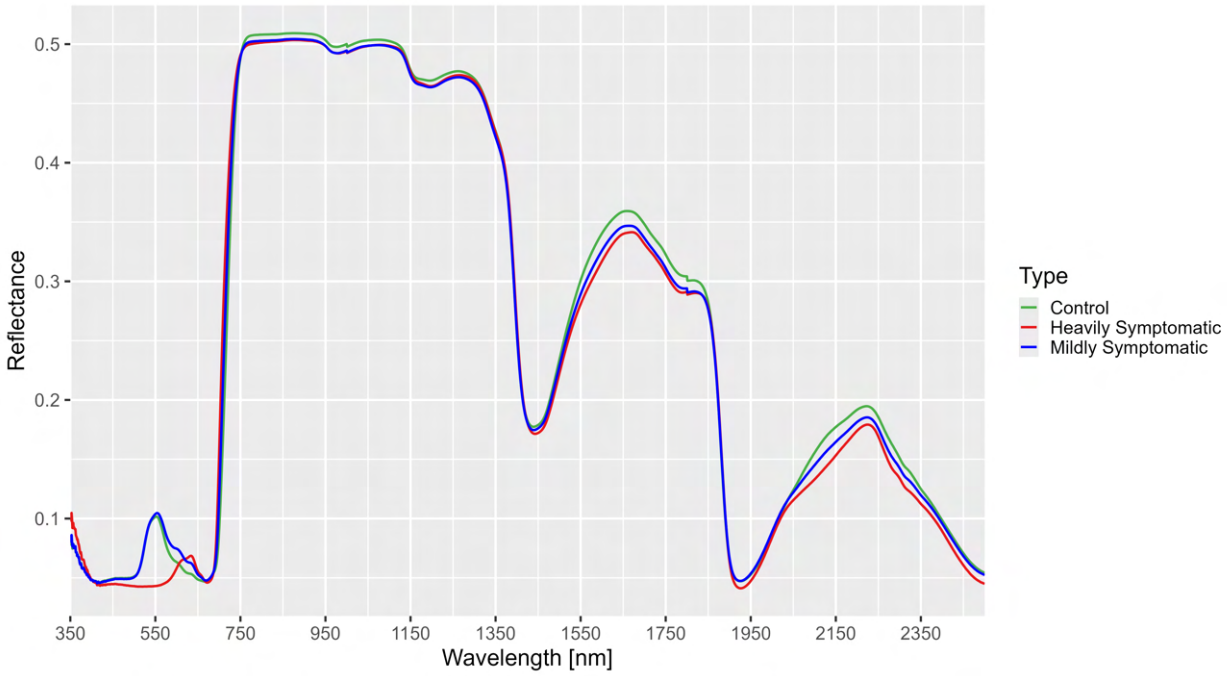


FIGURE 2.4: Mean spectra of control, mildly symptomatic and heavily symptomatic plants from the Weber vineyard.

2.5 Models

2.5.1 K-fold CV

The models below all went through the same process: First, the model parameters were defined using a k-fold cross validation (CV) method (Kohavi, 1995) with k = 10. This means that the data was partitioned into ten roughly equally distributed chunks, such that a similar number of infected vs non-infected datapoints were in each chunk. Then the data is trained on nine of these chunks and tested/validated on the last chunk. This is repeated ten times until each chunk was used once as test data and nine times as training data. From this, the best model is created with parameters tuned on the results of the 10-fold CV. Before each tune, the same seed was set for the randomness, such that every fold is the same every run and the results become comparable

between models.

Ideally, the k-fold CV would have been used to evaluate the model with a limited dataset, but the consensus in the programming world (e.g. Berk, 2017; Kuhn, 2008) is that it should only be used to calculate the parameters of an ideal model and then train that model on the full dataset to predict new data. This is also reflected in the functionality in caret, one of the most used R packages for ML, as a model that is trained using k-fold CV does not output the metrics that are usually output from a 'normally trained' model. The issue is that all values would be averaged and the community thinks that this does not give a meaningful representation of the evaluation metrics.

A final model was then trained on a predefined training set from Weber, which consists of a randomly selected 80% of the full dataset, and the model is then tested on the other 20%. This split was created with the R package 'caTools' (Tuszynski, 2021) and it distributes the data based on the BN-infected attribute such that there is an even distribution of infected and non-infected data-points in each split. Since we also had an independent dataset from Rütihof, the model was also tested on that dataset to get an impression on how the model would perform with truly unseen data.

2.5.2 Machine learning

In this thesis, three machine learning algorithms, as well as one linear model are assessed for their suitability to be used in modelling remote sensing data of infected vines. Each problem in this thesis can be described as a binary classification problem of infected vs. non-infected. All three ML algorithms are supervised classification algorithms, meaning that they use a training set with annotated labels (infected vs non-infected) to build a model and use the labels/ground truth to assess their accuracy.

2.5.2.1 SVM

Support vector machines (SVM) are used quite commonly in remote sensing (Albetis et al., 2019; Huang et al., 2002; Mountrakis et al., 2011) and data science (Alloghani et al., 2020). A spatial analysis lecturer told us that SVMs are generally used when the number of dimensions/variables overweigh the number of datapoints, which is often the case when using high resolution data. While this might not be strictly true, I found it worked out more often than not during my studies.

An SVM is a supervised machine learning algorithm which maximizes a boundary, i.e. a line or hyperplane in such a way that the error of classification is minimized (as defined in Vapnik and Golowich, 1995). In a way it is similar to a regression line, except that it is not straight and in much higher dimensions, so that it is a (hyper)plane, not a line anymore. The support vectors are the points closest to the boundary.

The SVM makes use of kernels for the calculation of the decision boundary. When using non-linear kernels (like a polynomial or a radial basis function), then the boundary becomes non-linear and more flexible to fit the points as well as possible to their respective classes. As a kernel, the

radial basis function is used, inspired by literature (Al-Saddik et al., 2017; Bendel et al., 2020) and the CV results. The implementation used here is mostly from the package e1071 (Meyer et al., 2023) and kernlab (Karatzoglou et al., 2004).

2.5.2.2 RF

Random forest classifiers (RF) are also often used successfully in remote sensing (e.g. Pal, 2005). They are based on decision trees, which use a certain feature of the dataspace and weigh it depending on a pre-defined metric (a decision). Many such decisions lead to a tree-like structure, therefore the name decision tree. For the forest, as the name 'random forest' implies, the method uses randomly selected features in a certain number of trees and tries to minimize errors while classifying the dataset (Ho, 1995). The best combination of trees is then used as a final model. The implementation in this thesis is done by the package randomForest (Liaw & Wiener, 2002) with 200 trees.

2.5.2.3 kNN

The k-nearest neighbour classifier is one of the simplest supervised machine learning algorithms. It uses a distance (often Euclidian) per point to its neighbours as a way to determine which group/class the point belongs to (Cover & Hart, 1967). It is also used in remote sensing (e.g. Meng et al., 2007) to classify high resolution imagery. A low k means that not many neighbours are considered in the classification, which can lead to low accuracy in the predictions. A high k can mean that while the predictions are stable, they might not accurately reflect the data, as the model is not complex enough to fit the training data. Therefore, choosing the correct k is a challenge. In this thesis the choice is made with the k-fold cross validation described above. The implementation in this thesis originates from the package caret (Kuhn, 2008).

2.5.3 Linear models

To not only use machine learning algorithms and having an easy-to-understand way to predict/model the dataspace, a linear model was chosen to run alongside the machine learning methods. Not only was it used for modelling, but more importantly, its results gave a reference to how many wavelengths a dataset can have to be modeled successfully with a linear model. The models are constructed with formulas of form: 'bn~predictors', where bn is the infected/non-infected attribute and all wavelengths of the respective dataspace are used as predictors.

As we are using a binary classification problem, the logistic regression (logit) was chosen to model the logarithmic odds of infection based on the linear combination of independent variables (Hosmer & Lemeshow, 2000). The implementation used here comes from the base stats package (R Core Team, 2023).

2.5.4 Model evaluation

The metrics used in here to evaluate the models are based on the confusion matrix calculated for the results of the classification (created by the package caret (Kuhn, 2008)). A confusion matrix (Table 2.4) uses the classification results from a model and compares them to the measured results/ground truth. It simply counts how many plants are predicted as healthy and how many are actually healthy (true negative, TN), how many are falsely predicted as negative (false negative, FN), how many are predicted as infected, but are healthy (false positive, FP) and how many are predicted to be infected and are actually infected (true positive, TP).

Predicted \ Actual	No / 0	Yes / 1
No / 0	**TN**	FN
Yes / 1	FP	**TP**

TABLE 2.4: Confusion matrix modified from the 'caret' R package (Kuhn, 2008); Positive class = 1

2.5.4.1 Accuracy

Accuracy is the metric of how many measurements were classified in the correct class. It is calculated on a confusion matrix (Rohit, 2022).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.5}$$

Interpreting accuracy can be thought of as: "If it is less than 50%, we could be flipping a coin and on average get better results".

2.5.4.2 F1 / precision / recall

Precision is how often the model predicts a positive (here non-infected) sample correctly, from all positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{2.6}$$

Recall is how often the model predicts positive samples correctly from all actual positive samples (which includes false negatives).

$$Recall = \frac{TP}{TP + FN} \tag{2.7}$$

The F1 score is an amalgamated score of the above explained precision and recall. It uses a harmonic mean of the two scores, such that if they deviate from another, the score worsens.

$$F1 = \frac{2 * precision * recall}{precision + recall} \quad \textbf{OR} \quad \frac{TP}{TP * \frac{1}{2}(FP + FN)} \tag{2.8}$$

2.5.4.3 Kappa

Cohens kappa (Cohen, 1960) is in essence a similar metric to the F1 score, but it considers the imbalance between the datasets. It calculates the agreement between the predicted values and the ground-truth. The Weber dataset does generally not have an issue with imbalance, as there are very similar amounts of infected and non-infected datapoints. In the case of the Rütihof validation dataset however, there is an imbalance. Therefore, this evaluation metric was added to give an additional estimation because the F1 score may be misleading with the Rütihof dataset.

$$Kappa = \frac{2 * (TP * TN - FN * FP)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)} \tag{2.9}$$

For interpretation of the values, the table from Landis and Koch, 1977 can be used. It states that >0.8 is almost perfect, >0.6 is substantial, >0.4 is moderate, >0.2 is fair, 0-0.2 is slight and <0 is poor. This scale is common for statistical interpretation and can also be used for the F1 score. Fleiss et al., 2003 argues that a kappa with >0.75 is excellent for most purposes and 0.40 – 0.75 is acceptable. Mary L, 2012 says that with a kappa score of 0.8 – 0.9, 64-81% of the data is reliable (calculated with the squared kappa, the equivalent to the squared correlation coefficient). In this thesis the guidelines from Fleiss et al., 2003 will be generally used for interpretation.

2.5.5 Relevant wavelengths

Multiple approaches were used to find the most relevant wavelengths for detecting BN-infected vines in this dataset. The main SVM-RFE approach was inspired by Bendel et al., 2020 and also Backhaus et al., 2011 and Becker et al., 2021, but the method was adjusted due to programmatical issues with the implementation of the relevance radial basis function (rRBF). The main issue was with the implementation in R, which is done over the package RSNNS. When used within the caret wrapper, this crashed the R session every time, no matter what changes were implemented. Without the caret wrapper, the method was more complicated and would have needed a great deal of restructuring of the code and data, which led to a change of strategy with different relevancy calculations to support the findings of this thesis. The main method (section 2.5.5.1 below) also uses an RBF as a kernel for the SVM, which makes these methods similar.

In all calculations, the wavelengths around 1000 and 1800 nm ($\pm$ 50 nm) were removed if they were detected by a feature selection method. This is due to the missing jump correction, which might introduce features at the boundaries of the sensors, which are artefacts of the sensor. It was judged preferable to simply omit these artefacts rather than to apply a smoothing function that might introduce new artefacts, as stated above under the spectroscopy methods 2.4.1. The boundaries between sensors in the FieldSpec furthermore have poor signal to noise ratios and so exclusion seemed sensible.

2.5.5.1 SVM-RFE

The main method for finding relevant wavelengths is the support vector machine – recursive feature elimination (SVM-RFE) method. It is used in remote sensing for e.g. land cover classification (Colkesen & Kavzoglu, 2016; Ramezan, 2022). Backhaus et al., 2011 also used a similar method, but with a linear kernel. The way an RFE works, is that it trains an classification algorithm (here SVM) and estimates which features are more or less important. The least important feature is left out and the training is repeated. This is recursively done until all features have been evaluated (Kuhn, 2008). Since this method is applied independently to all wavelengths, any variation of neighbouring wavelengths is a feature of the data, not the method.

From the resulting importance of wavelengths, the selection of 'best' wavelength was done via a local maxima detector with a moving window size of 100 nm (using Zeileis & Grothendieck, 2005). This assures that locally best solutions are considered and not only high peaks are chosen (as can be seen in the results section 3.1.1). This also assures that areas within the whole of the spectrum are used and not only some concentrated sections. This gives a more accurate representation of the plant over the whole spectrum, rather than just some certain areas. This selection method is based on the authors' rationale, rather than any literature, but is assessed for its usefulness later in the thesis (see Section 3.3).

2.5.5.2 RF-RFE

The 'cousin' of the SVM-RFE is the random forest RFE (RF-RFE). It is the same procedure as the SVM-RFE but uses a random forest classifier to model the dataspace. This has also been used in remote sensing, e.g. to classify grasslands with high-resolution data (Demarchi et al., 2020). As this method was mostly chosen to compare the results of feature importance to those of the SVM-RFE, there was no feature selection method applied. This is also the case for the following methods and the reasoning can be taken from the results (Section 3) and discussion (Section 4).

2.5.5.3 RF-GA

As a complement for the RF-RFE and inspired by Al-Saddik et al., 2017 a random forest genetic algorithm was tested for feature selection. The idea is based on the darwinian theory of survival of the fittest (Kuhn & Johnson, 2013; Mitchell, 1996), where the populations with the fittest individuals will outperform populations with worse individuals. The individuals in feature selection are subsets of predictors and the performance/fitness metric is the classification accuracy (Scrucca, 2013). The individuals with the best fitness are randomly selected to reproduce offspring that is then tested in the next generation. The sampling method for the initial individuals used here is 5-fold cross-validation, meaning that the whole algorithm is conducted 5 times. The main issue with this approach is that it tends to fit aggressively and overfits to the predictors (Kuhn, 2008). This is the reason why an external function is used to check the classification accuracy from the selected predictors with the held-out data set from the cross validation.

2.5.5.4 RF univariate filter

This random forest method pre-screens predictors using simple univariate statistical methods and only uses the predictors that pass some criterion (Kuhn, 2008). This method is also used in bioinformatics (Menze et al., 2009) and even in research on FD (Albetis et al., 2017) and should form a complement to the rest of the random forest feature selection methods.

2.5.5.5 HSC-PA

The hierarchical spectral clustering with parallel analysis (HSC-PA) is a new method proposed by C. Li et al., 2023a to segment spectral data into clusters based on similarity. This similarity measure is based on a PCA and uses parallel analysis (Franklin et al., 1995) to determine the number of segments and number of principal components (PCs) retained per segment. This method was evaluated to determine whether it is possible to use this method to train a classifier and if it was not possible, the segmentation results could at least be used to discuss the other methodology.

2.5.5.6 PCA

A principal component analysis (PCA) (Wold et al., 1987) is often suggested when trying to reduce a high-dimensional dataset but has been in critique in the remote sensing community for a while (Cheriyadat & Bruce, 2003; Prasad & Bruce, 2008). It has been used in this thesis to see if there are any discernible differences between BN positive and uninfected vines. It is only used to discuss the results later on and to show which differences are picked up by the principal components.

2.5.6 Resampling with a probability density function

Sensors like those of the Mavic 3M and the field spec are often not perfectly situated on a certain wavelength and are therefore represented with a certain margin (See Table 2.2). To output a value of a certain wavelength, a probability density function (pdf) is used to calculate how much of the light that hits the sensor is of the wanted wavelength. To simulate this behaviour, when resampling from the full spectra to only a few wavelengths, e.g. those calculated by the SVM-RFE, a gaussian distribution is used which takes values from the full spectra based on the probability density function around the desired wavelength. To simulate a sensor similar to the M3M, a standard deviation of 8 nm is chosen that leads to approximately a spectral sensor width of 16 nm. The gaussian shape is commonly used, but not the most accurate to describe how sensors take in light (Trim et al., 2021), but as this study is a proof of concept and the data is not greatly influenced by the choice of distribution, it was chosen here.

3 Results

3.1 Wavelength selection

3.1.1 SVM-RFE

3.1.1.1 Full year

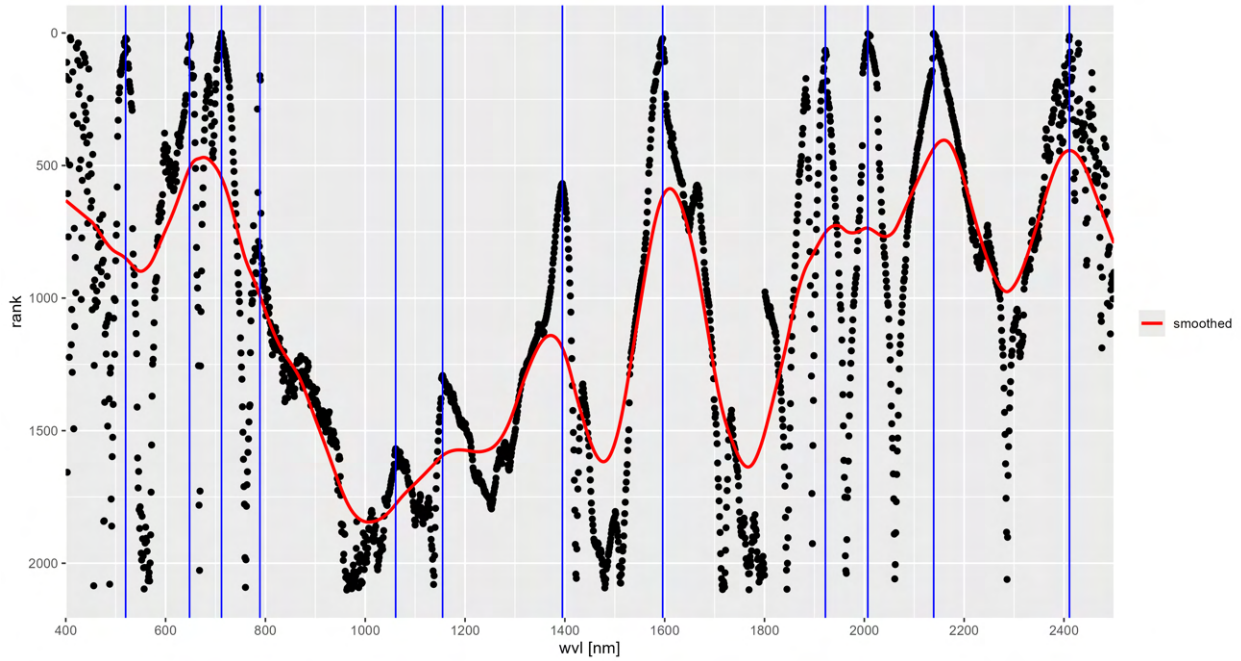


FIGURE 3.1: Result of the SVM-RFE of the full Weber dataset. Black dots denote rank; lower rank means higher importance, therefore the scale is inverted to ease visual interpretation. The blue lines indicate local maxima of the result, i.e. the wavelengths that are selected for further analysis. The red line is a smoothed version of the black dots for visual aid.

For the full dataspace, the SVM-RFE method with a local maxima finder resulted in the selected wavelength features at 520, 648, 712, 789, 1061, 1155, 1395, 1596, 1922, 2007, 2139 and 2411 nm (see Figure 3.1). Wavelengths around 1000 and 1800 nm ($\pm$50 nm) were excluded due to the sensor transition (see section 2.4.1). In the visible part of the spectrum (VIS; <~780 nm) there are 4 features that are quite similar to the wavebands used by the M3M. Towards the near-infrared region (NIR, ~780 nm - ~1100 nm), there are no notable features.

In the range of the middle sensor of the FieldSpec, i.e. the lower short wave infrared (SWIR, 1100 nm - 1800 nm) there are two features that are not highly important and one at 1600nm which is as important as the remaining SWIR features. The high SWIR region (>1800 nm) harbours

four highly important features, which is also indicated by the ranking of the most important wavelengths in the Appendix Table A.1. These SWIR features are all outside of the range that can be monitored by the RPAS' used in this thesis and also most other commercially available RPAS-mounted sensors.

The above results are based on the numbers and the dots of the data, but when looking at the smoothed line, which is mostly for ease of visual interpretation, we see that in the VIS, the 648 and 712 nm features stand out much more than the others. The NIR and low SWIR are not very important, while the high SWIR is still very important.

3.1.1.2 July

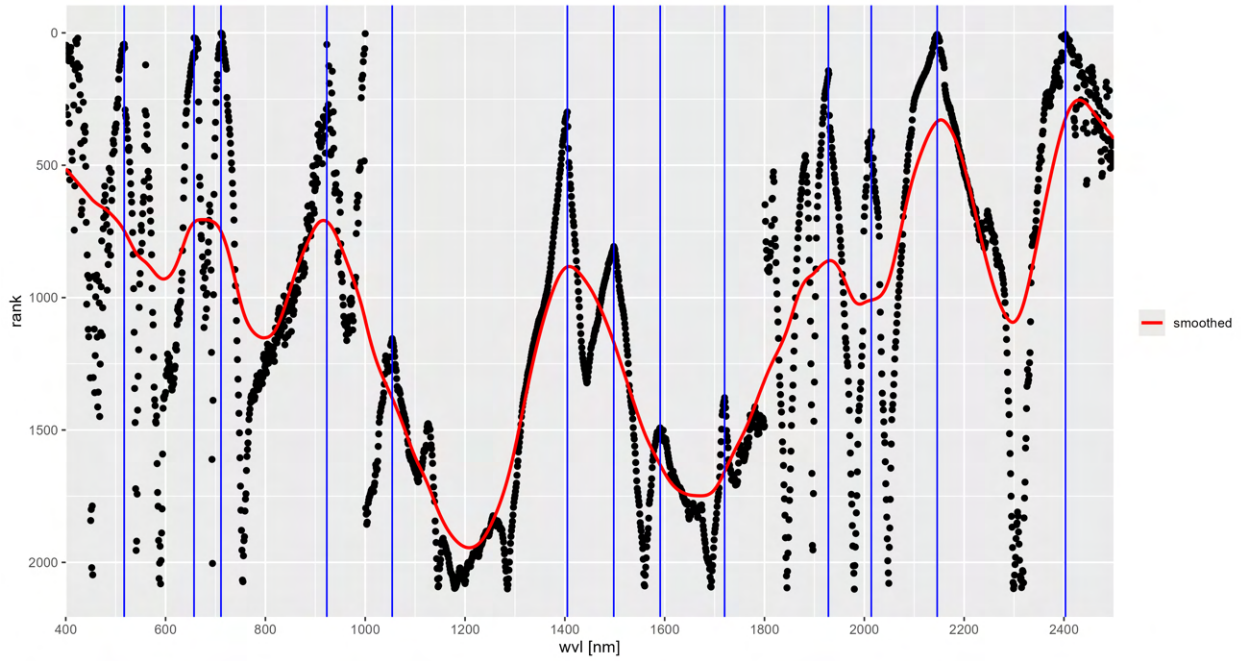


FIGURE 3.2: Result of the SVM-RFE of the July part of the Weber dataset. Black dots denote rank; lower rank means higher importance, therefore the scale is inverted to ease visual interpretation. The blue lines indicate local maxima of the result, i.e. the wavelengths that are selected for further analysis. The red line is a smoothed version of the black dots for visual aid.

The resulting wavelength features for July are: 517, 657, 711, 923, 1054, 1405, 1498, 1591, 1720, 1928, 2014, 2146, 2403 (see Figure 3.2). When applying the same SVM-RFE only on data from July (12 infected, 14 control) which represents less severe symptoms than the full dataset, we see similar features in the VIS with the ~780 nm feature missing, one more feature at 923 nm and at 1054 nm, which are far less important for the full dataset. The 1400 nm feature is more prominent, while the 1600nm diminished. In the last part of the SWIR spectrum, the selected wavelengths are nearly identical to those from the full year dataset, with less importance around 2000 nm.

3.1.1.3 August

The resulting wavelength features for August are: 547, 722, 793, 1093, 1218, 1371, 1477, 1612, 1919, 2014, 2142, 2246, 2364 (see Figure 3.3). When applying the same SVM-RFE only on data from July (19 infected, 12 control) which represents more severe symptoms than the full dataset, we again see similar features in the VIS but missing the ~650 nm feature found in the full dataset. Compared to July, the ~900 nm feature is missing, with a much more prominent feature at ~1093 nm, also compared to the full year. The features at ~1400 nm and ~1500 nm are even more prominent compared to July, while the 1600 nm feature is highly prominent. The ~2000 nm feature dropped in importance compared to the full year and July. Everything past 2000 nm seems to be less important in differentiating late-stage plants from controls.

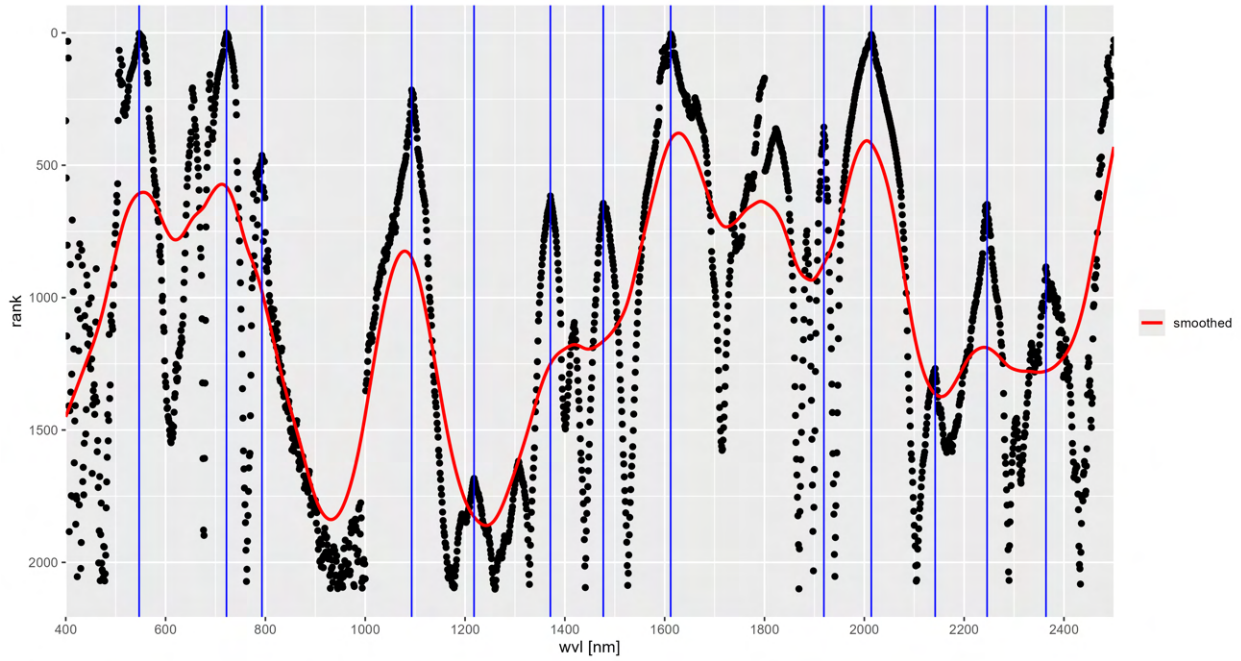


FIGURE 3.3: Result of the SVM-RFE of the August part of the Weber dataset. Black dots denote rank; lower rank means higher importance, therefore the scale is inverted to ease visual interpretation. The blue lines indicate local maxima of the result, i.e. the wavelengths that are selected for further analysis. The red line is a smoothed version of the black dots for visual aid.

3.1.2 RF-RFE

The random forest recursive feature elimination produces a very centralized feature distribution, unlike the SVM-RFE. We don't see importance values for the whole range of wavelengths, but there are only a few points around a few key wavelengths. This is not due to a post-processing step, but the output from the applied method. The most important areas are determined to be around 540, 706 and 1508 nm (see Figure 3.4). The overall importance provided here and the following methods is a relative measure of importance. The number itself is meaningless, and only gains meaning through the comparison to the points within the same method/plot. So when looking at a plot, the wavelengths with high overall importance are the ones to be focused on.

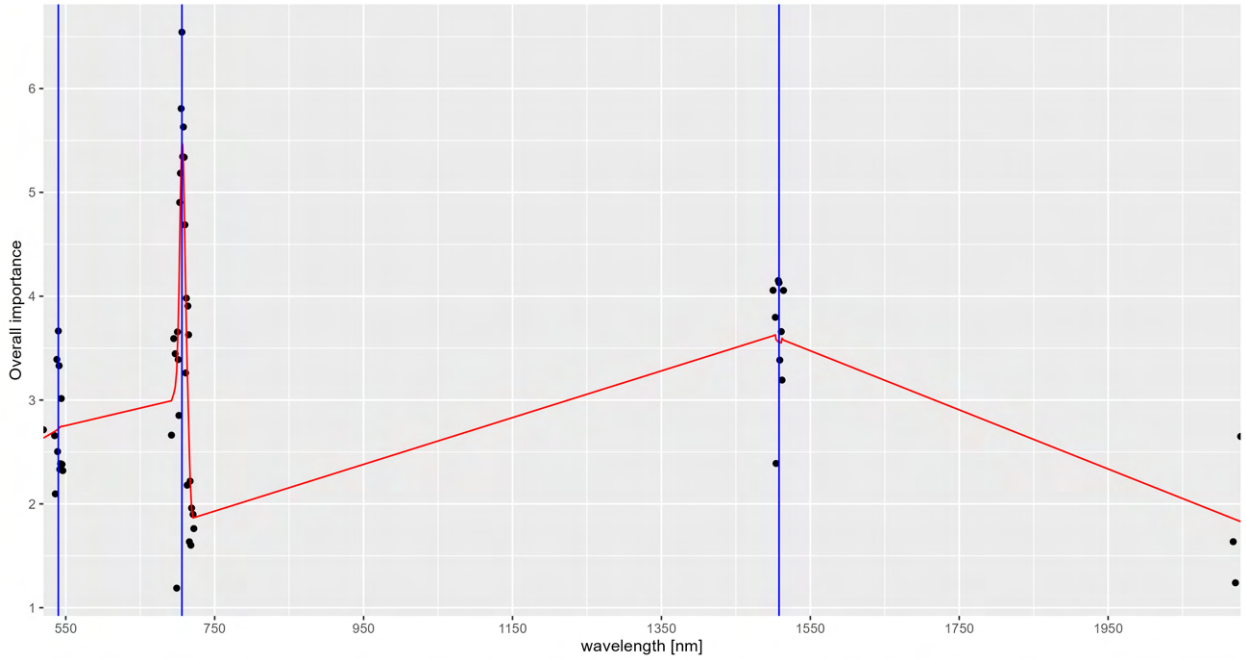


FIGURE 3.4: Result of the RF-RFE importance method on full-year data. The important wavelengths are clustered around 540, 706 and 1508 nm wavelengths. The red line is a smoothed version of the black dots for visual aid.

The fact that it groups around some wavelengths probably goes back to one of the main issues with random forest when using little data. It has a tendency to overfit on training data and focusing on certain features more than others. We can also see a similar result when only using the variable importance feature built in the random forest package (See section 3.2). The wavelengths around 700 nm seem to be very important for random forest classifiers.

3.1.3 RF-GA

Genetic Algorithms have the advantage that they use lots of iterations which reduces the chance that there is an outlier, but they can start optimizing on certain features that are not optimal but lead to better results (Kuhn, 2008). This is probably the case here as well. We can see in the mean error plot below (Figure 3.6) that the internal RMSE goes down over the generations, but the external mean actually worsens slightly, but stays pretty much the same all around. A divergence between the internal and external function usually means that the reduction with the chosen features don't improve the algorithm. The number of predictors used are 238 at iteration 77, which was deemed to be the best iteration by the algorithm. The improvements of the mean RMSE of the internal validation function are small, within 0.04. As a comparison: the examples of the caret package have an improvement of 0.5, with a higher RMSE (max 3.5). This means that here we have a similar percentage of improvement (10% for us compared to 14% of the example), but the divergence makes this result not great anyways.

We can see in the resulting importance plot (Figure 3.5) that much like the other random forest feature selectors, there is a heavy weight on a few wavelengths around 710 nm, and one feature at around 550 nm. There is also a small group around 2130 nm.

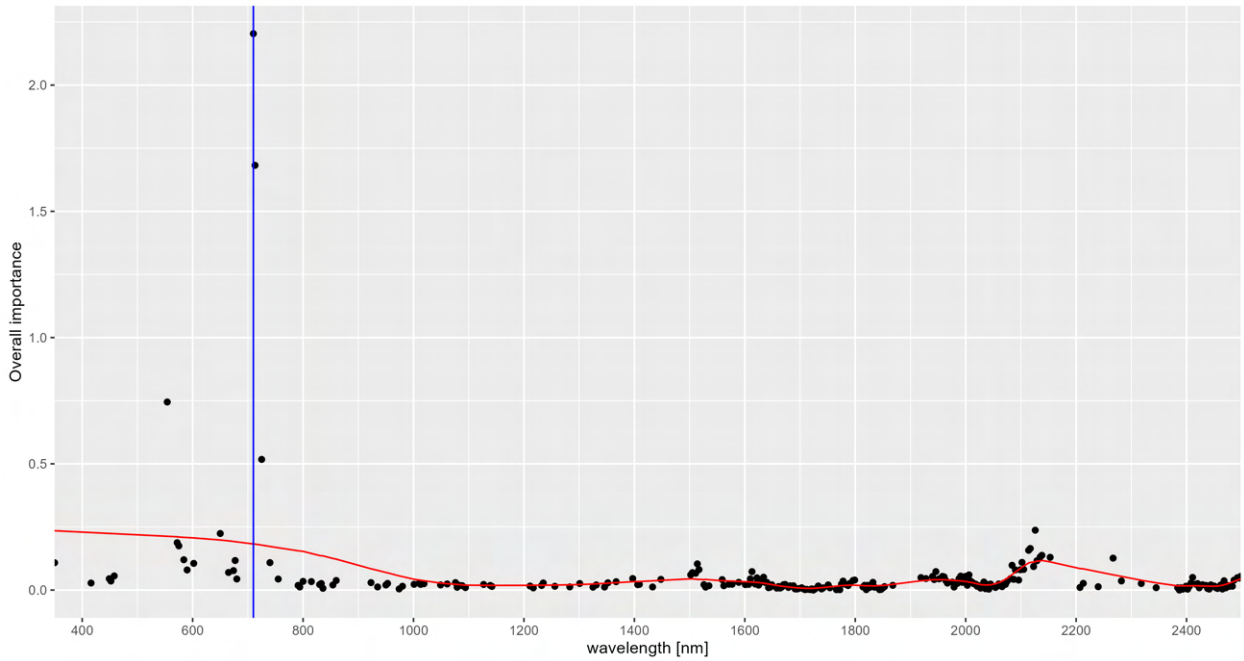


FIGURE 3.5: Result of the RF-GA importance method on full-year data at the best iteration. There is one major spike of importance around 710 nm, a smaller around 550 nm and a group around 2130 nm. The red line is a smoothed version of the black dots for visual aid.

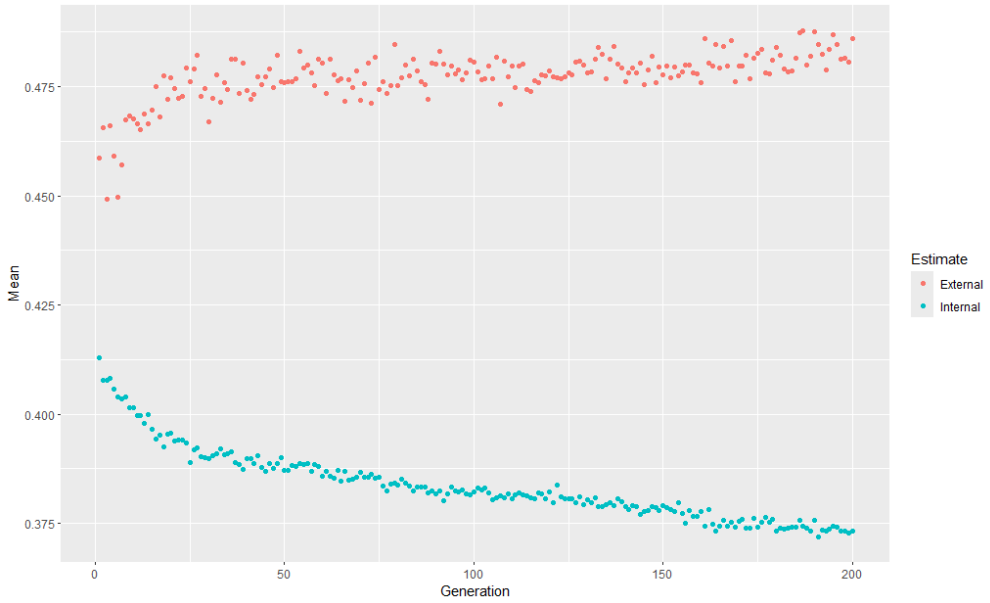


FIGURE 3.6: Mean RMSE of the internal and external evaluation function. Internal determines the best features per generation on part of the dataset, while external validates on a held-back sample of the dataset.

3.1.4 RF univariate filter

As with the other Random Forest approaches, we see a clustering of important wavelengths. Namely around 706 nm, exactly the same as in RF-RFE and around 2100 nm. There are also clusters around ~500 nm and ~2400 nm.

As can be seen in Figure 3.7, the random forest feature selection methods identify only a few

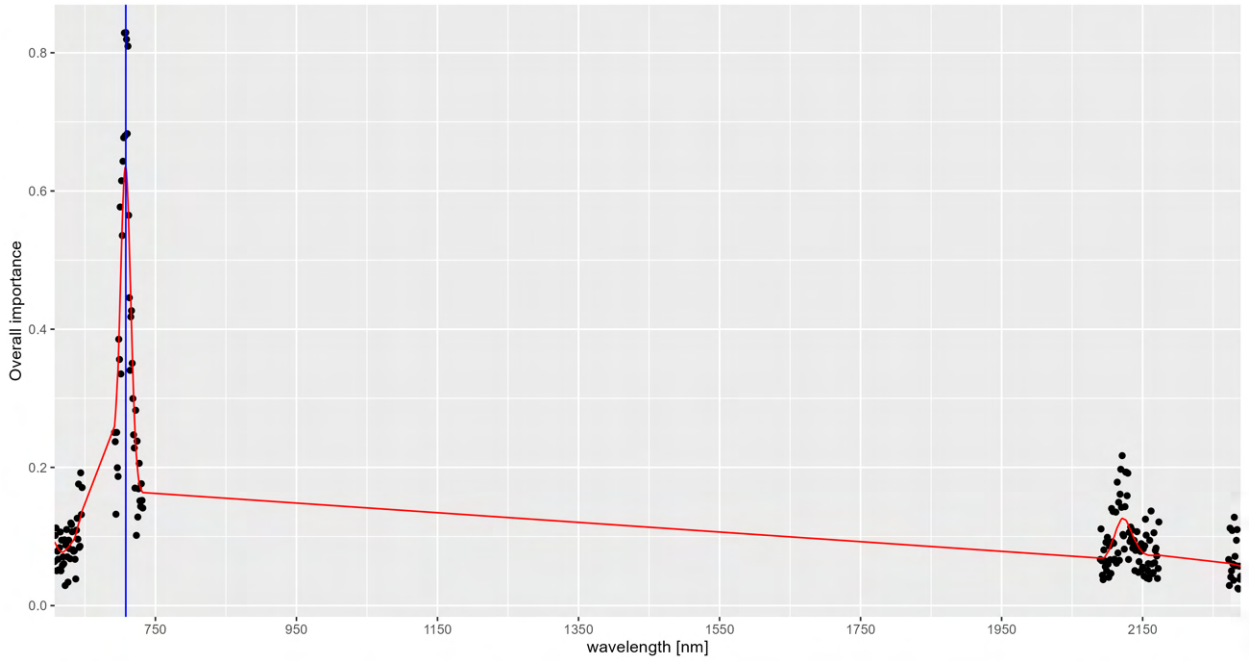


FIGURE 3.7: Results of the RF univariate filter method. There is a major spike of importance around 706 nm. The red line is a smoothed version of the black dots for visual aid.

features, with the main focus on around 710 nm, with mostly only a few points of very high relative importance. Due to their clear bias, their results were not used to inform any models, as they would be quite one-sided (also see Results 3.2). The SVM-RFE methods provide a more continuous evaluation of relative importance, making them more useful to evaluate the dataspace.

3.1.5 Other selections

As a comparison to the local maxima wavelength selection method, different wavelengths were selected. In particular, the best 10/20/50 ranked wavelengths from the SVM-RFE were selected, i.e. the top-ranked in Figure 3.1. This comparison should also provide a better understanding whether it is enough to reduce the feature space to any top wavelengths, or if the local maxima selection approach is helpful. To check whether the M3M wavelengths from the RPAS is helpful and can be used in the field, those center wavelengths were also selected. The reflectance values for all wavelengths were obtained by using the resampling approach with the pdf (see section 2.5.6).

The results from Table 3.1 are summarised here. The best 10 wavelengths are mostly (80%) far SWIR Wavelengths (>2000 nm), with one wavelength each at 712 nm and 648 nm. The best 20 are more spread, with 60% of the values in the far SWIR, 20% around 700 nm, 10% at 650 nm and one each at 409 nm and 520 nm. The Top 50 has 46% in the far SWIR, 10% around ~1600 nm, 18% around ~710 nm, 10% around 650 nm and 16% between 400 nm and 522 nm.

For a ranking of the top 50 most important wavelenghts, see Appendix Table A.1. There we see similar to the RF-RFE results, that 712 nm is the most important, followed by 2000+ nm wavelengths as the important features.

Selector	Wavelengths [nm]
Best 10	648, 712, 2007, 2008, 2010, 2011, 2138, 2139, 2140, 2141
Best 20	409, 520, 648, 649, 710, 711, 712, 713, 2007, 2008, 2010, 2011, 2012, 2138, 2139, 2140, 2141, 2142, 2143, 2411
Best 50	409, 410, 425, 517, 518, 520, 521, 522, 646, 647, 648, 649, 650, 708, 709, 710, 711, 712, 713, 714, 715, 716, 1591, 1593, 1594, 1595, 1596, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2410, 2411, 2429
M3M	560, 650, 730, 860

TABLE 3.1: Selected wavelengths based on SVM-RFE top results of the full year data and Mavic 3M center wavelengths.

3.1.6 HSC-PA

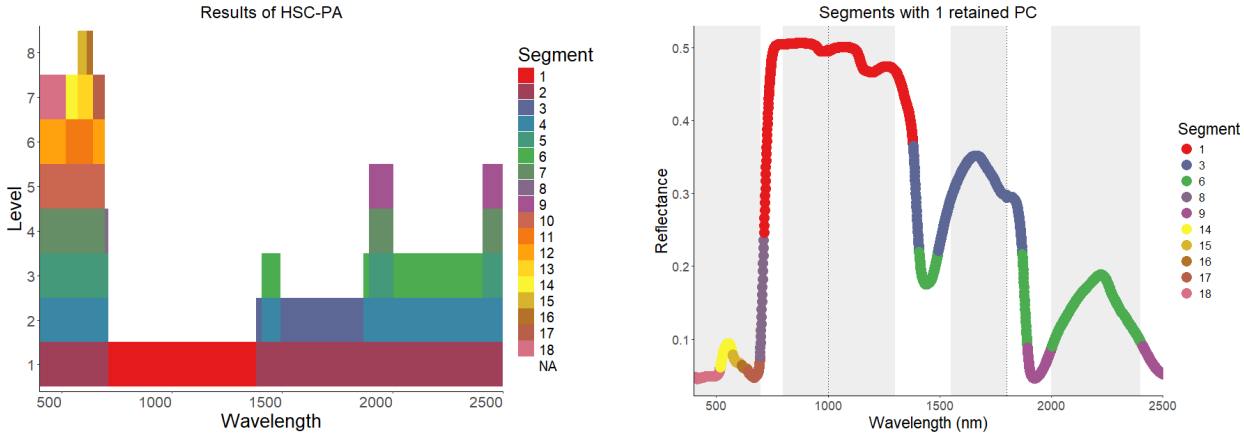


FIGURE 3.8: Results of the Hierarchical Spectral Clustering with Parallel Analysis (HSC-PA) (left) and Presentation of the segments on an example spectra when one Principal Component remains (right).

The segmentation approach from C. Li et al., 2023a yields multiple segments in the VNIR and a few in the SWIR, as well as one single, long segment from the red-edge around 750 nm until about 1350 nm in the SWIR. These segments hint towards a very differentiated VIS region with only minor variability in the SWIR region. This is also shown by the result of the HSC-PA (Figure 3.8, which indicates much more variation in the VIS by there being 8 segmentation steps (levels) versus 1 in the NIR/SWIR and up to 5 in the SWIR.

There was an attempt at using these segments, or rather the first PC per segment to train a classifier, but it did not classify well (<40% accuracy). This approach was therefore not developed further, although the different levels of segmentation are interesting as descriptors of the variation within the dataset.

3.1.7 PCA

Two simple principal component analyses were conducted to potentially reduce the feature space (see Figure 3.9). The first was on the full dataset and did not yield any clearly differentiated and interpretable results, as there are too many variables to chose from. The second uses the SVM-RFE local maxima deduced wavelengths. It also does not show a clear differentiation, but when

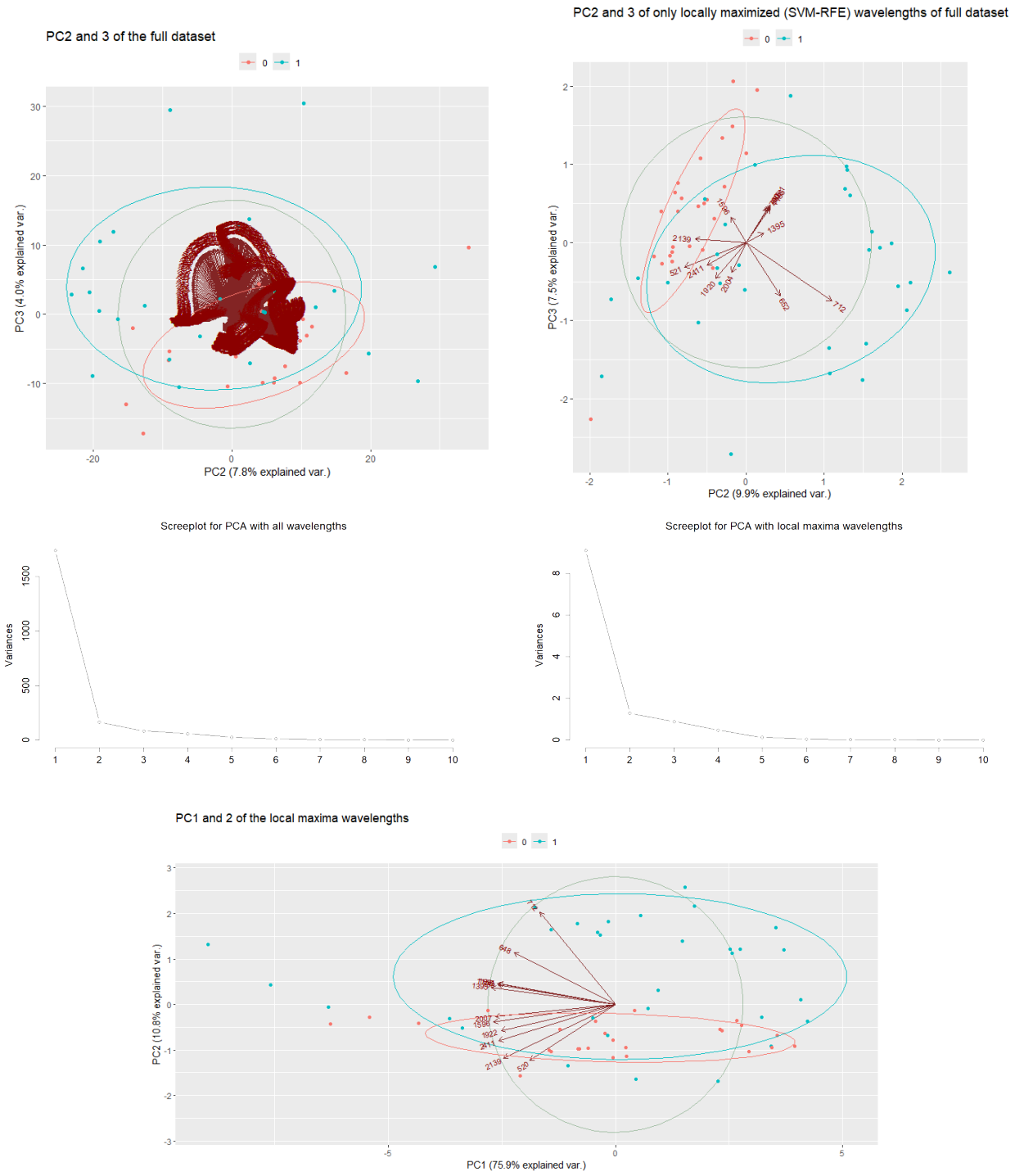


FIGURE 3.9: Results of the PCA on full wavelengths (left) and SVM-RFE wavelengths (right). Top row: PC2 and PC3 to show differences in color. Second row: The screeplots show the respective distribution of explained variance. Third row: PC1 and 2 of the local maxima wavelength, to show and example that there is not much separation between the two PCs (Looks even less separated for the full dataset).

looking at PC2 and PC3, we see that the infected plants (1 in the figure) differ along PC2, which is the direction of the 712 nm feature. We can also see that 2139 nm, 1596 nm and 521 nm point in the direction of the control plants. This separation can also be seen in the PC1 and 2 plot at the

bottom, but less differentiated. For legibility and space reasons, the PC1 and 2 of the full dataset was not shown, but the two elipses are nearly fully intersect and show much less variation. The screeplots show that most of the variation is explained in PC1 in both datasets, but there is no clear indication which wavelengths cause it.

3.2 Models

The results of the models are separated into three groups, depending on the selection of input wavelengths used. The first set of results shows the models trained and tested on the full dataspace, so all wavelengths from 400 nm to 2500 nm are used in the training and testing of the model. The second set shows the dataset resampled with a pdf to the ideal wavelengths determined by the SVM-RFE method (see results 3.1.1). The third set shows the data resampled to the wavelengths that are covered by the M3M RPAS.

The results are presented grouped by the model type, split in three rows; train, test and validation with the second vineyard Rütihof (RUT val). The train value is calculated by using the trained model (on the training set) and predicting the trained dataset. This can show how well the model fits the data. This does not need to be 1 as a perfect fit, because if a model is fit too well it might be an indication of over-fitting. The test value shows what is usually expected of a model and predicts the unseen portion of the original data (test data) with the trained model. The RUT val value shows the validation with the Rütihof data. This is completely new data, from a different vineyard and red variety, and should give an indication of how well the model is transferable to a different dataset.

3.2.1 Full spectra

In Table 3.2 we see that the SVM train and test values are around 0.8, which means they have a 20% misclassification rate. The F1 score is also rather high with both. Precision is perfect for all three values, meaning that all positives that are predicted are actually positive, so no false positives. We can see that the main issue with classification comes from the recall, with a value of around 65%, or 35% misclassification rate, meaning that not all infected samples are correctly identified. The Rütihof validation set has much higher values, even the recall with roughly 95% correctly classified infected plants. One can also see the kappa score, which is a better indicator than F1 when it comes to unbalanced datasets (as Rütihof is), which is also very high with 0.91 and indicates agreement between the measured and modelled output.

For RF, the training accuracy (and all other scores) is perfect, but the test set and the Rütihof validation set are much lower. This is a good indication of an overfitted model, as it works perfectly for the training data that was used, but not well for unseen data. The test data still has a decent classification with 80% accuracy, but for the Rütihof data, we see that it has a perfect recall, but a lower precision (60%), meaning that from the actual positive plants it misclassifies a few.

For the random forest, we can print out a variable importance plot (see Figure 3.10) of the classification, which shows which wavelengths were used in the decision trees most often and

Model	Subset	Accuracy	F1	Precision	Recall	Kappa
SVM	train	0.783	0.750	1.000	0.600	0.578
	test	0.818	0.800	1.000	0.667	0.645
	RUT val	0.960	0.970	1.000	0.941	0.911
RF	train	1.000	1.000	1.000	1.000	1.000
	test	0.818	0.800	1.000	0.667	0.645
	RUT val	0.800	0.872	0.773	1.000	0.449
kNN	train	0.630	0.622	0.700	0.560	0.269
	test	0.818	0.800	1.000	0.667	0.645
	RUT val	0.920	0.938	1.000	0.882	0.828
logit	train	1.000	1.000	1.000	1.000	1.000
	test	0.455	0.400	0.500	0.333	-0.065
	RUT val	0.400	0.211	1.000	0.118	0.079

TABLE 3.2: Summary of ML results of the full spectra-dataset.

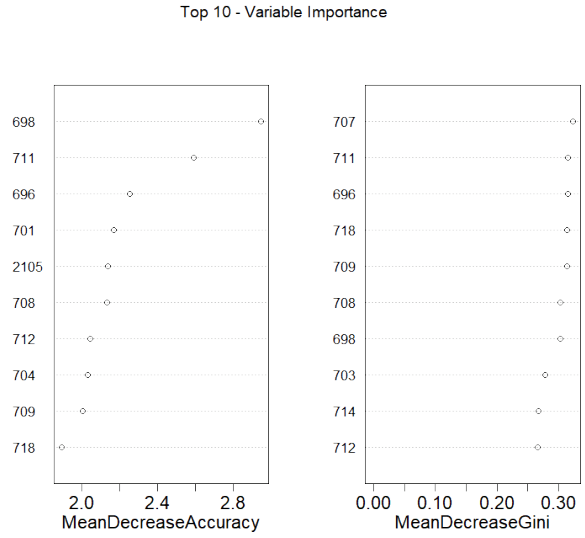


FIGURE 3.10: Variable importance of RF model trained on full dataset.

their impact on accuracy, if left out, and the Gini coefficient. The higher the Gini coefficient, the more it contributes to the homogeneity of the nodes and leaves (Martinez-Taboada & Redondo, 2020), meaning it helps balance the tree more. Here there is high importance on the wavelengths around 700 nm, with only one around 2100 nm.

With the kNN model, we see a very low training accuracy (60%), but a 20% higher test accuracy. This could just be that the few test samples fitted the model nicely. But we can also see that the Rütihof validation set is classified pretty accurately, with the most errors in the false negative space (i.e. recall).

The logit model shows a perfect training mean, so it fits perfectly to the training data, but is extremely unreliable when it comes to new data, seen in all values, especially the kappa score of nearly 0. This is expected, as we are using 2000+ variables in a linear model. This will never result in a good output. In the R code, one can also see that only about 50 of the variables were used in

the fitting process, as it was not able to converge afterwards. We can also see a PR(>|z|) value of 1 for all variables, which means that the model had too many variables and was not able to come to a good conclusion. The logit model here is used to show how metrics look like for an inappropriate model. It will only show useful results with less variables.

3.2.2 Local-maxima-derived SVM-RFE wavelengths

Model	Subset	Accuracy	F1	Precision	Recall	Kappa
SVM	train	0.804	0.780	1.000	0.640	0.619
	test	0.818	0.800	1.000	0.667	0.645
	RUT val	1.000	1.000	1.000	1.000	1.000
RF	train	1.000	1.000	1.000	1.000	1.000
	test	0.818	0.800	1.000	0.667	0.645
	RUT val	0.800	0.872	0.773	1.000	0.449
kNN	train	0.783	0.750	1.000	0.600	0.578
	test	0.818	0.800	1.000	0.667	0.645
	RUT val	0.960	0.970	1.000	0.941	0.911
logit	train	1.000	1.000	1.000	1.000	1.000
	test	1.000	1.000	1.000	1.000	1.000
	RUT val	0.440	0.462	0.667	0.353	-0.017

TABLE 3.3: Summary of ML results on local maxima SVM-RFE wavelengths.

With the optimized wavelengths (Table 3.3), we see a similar picture as for the full dataset. The SVM has a 2% higher training accuracy, with the improvement coming from a higher recall. The test set is the same, but we see a perfect classification from the Rütihof validation data.

The random forest shows exactly the same values as the full dataset, still with a perfect training accuracy. The variable importance plot of the random forest structure (Figure 3.11) shows that 712 nm is still very important, followed by 1600 nm and 2100 nm.

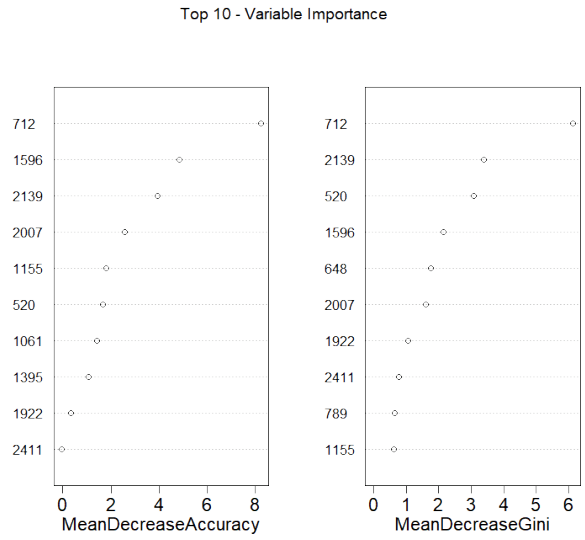


FIGURE 3.11: Variable importance of RF model trained on local maxima wavelengths.

The kNN model also has an improvement, with a higher training and validation accuracy. The training accuracy improved by about 15%, with most of the improvement coming from the precision, so there are no false positives anymore. The kappa score also jumped from 0.25 to 0.65. However, many positive samples are still not detected.

The logit model shows a perfect training performance as before, but contrary to before the test set is also perfectly classified. As with the full feature set, the validation with Rütihof data is very poor, worse than flipping a coin, as again can be seen by a kappa score of around 0.

3.2.3 Mavic wavelengths

With the simulated Mavic wavelengths (Table 3.4), we generally see a decrease in performance for most models. SVM has similar train and test score as with the other wavelengths, but a lot lower scores for the Rütihof validation (e.g. 0.6 accuracy vs 0.9 and 1 before). Here we see a very low kappa score, which indicates that there is a disagreement between the actual and predicted values and that the prediction is not reliable.

Model	Subset	Accuracy	F1	Precision	Recall	Kappa
SVM	train	0.783	0.750	1.000	0.600	0.578
	test	0.818	0.800	1.000	0.667	0.645
	RUT val	0.680	0.800	0.696	0.941	0.083
RF	train	1.000	1.000	1.000	1.000	1.000
	test	0.727	0.769	0.714	0.833	0.441
	RUT val	0.760	0.850	0.739	1.000	0.312
kNN	train	0.761	0.732	0.938	0.600	0.534
	test	0.727	0.667	1.000	0.500	0.476
	RUT val	0.800	0.872	0.773	1.000	0.449
logit	train	0.913	0.917	0.957	0.880	0.826
	test	0.909	0.923	0.857	1.000	0.814
	RUT val	0.720	0.829	0.708	1.000	0.163

TABLE 3.4: Summary of ML results on Mavic wavelengths.

Random forest still has perfect train values, with acceptable test and validation accuracy, but the kappa score indicates that the data is barely reliable. The variable importance plot (Figure 3.12) shows that the most important wavelength is 560 nm, followed by 730 nm. The least important is 650 nm, which for the first time shows up in the RF importance plots at all.

The kNN also shows similar train values compared to the 'ideal' wavelengths, test and validation set are much lower. Also, kappa is considerably lower in test and validation.

The logit model is the outlier in these results, meaning that for the first time, it produced a meaningful output (see Figure 3.13). This is due to the low number of predictors used. It has a high train and test value and reached significance for wavelengths 730 nm and 860 nm and slight significance for 560 nm. The transferability of the model is not guaranteed, with excellent accuracy values, but very low kappa score once again.

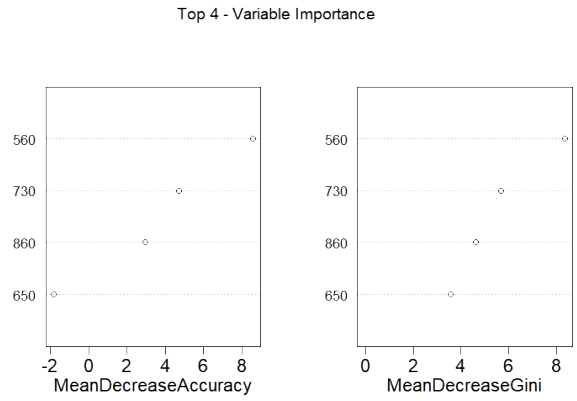


FIGURE 3.12: Variable importance of RF model trained on Mavic wavelengths.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.130      6.695  -0.318   0.7503
`560`       -393.646    167.519  -2.350   0.0188 *
`650`        225.434    183.640   1.228   0.2196
`730`        429.171    159.430   2.692   0.0071 **
`860`       -264.658     98.481  -2.687   0.0072 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

FIGURE 3.13: Output of linear model using the Mavic wavelengths.

3.3 Selected wavelengths compared

These models are all trained with 10-fold CV on the full Weber dataset and each model is evaluated with the parameters they provide. The implementation of SVM is different here than above, as it uses the package caret for k-fold CV with ‘svmRadial’, which is implemented in the package kernlab (Karatzoglou et al., 2004). So, the results may differ compared to the e1071 implementation (Meyer et al., 2023).

RF is not shown here, as we have seen that it mostly overfits and does not give reliable classification. There was a similar issue with logit, with the fact that it did not produce reliable results with too many wavelengths, but as a rough comparison, it can be used.

In the logit plot (Figure 3.14) we see the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) which are used to compare the same model trained on different data and give information about how well the model fits the data. The AIC (Akaike, 1974) estimates the likelihood of a model to predict/estimate future values, while the BIC (Wright & Stone, 1979) measures the trade-off between model fit and model complexity (E. A. Mohammed et al., 2015). Generally, the model with the lowest BIC and AIC indicates the best fit for the data. As we can see, the lowest value in both BIC and AIC corresponds to the RFE top 10 wavelengths, with only a slight difference to the local maxima and Mavic wavelengths. These results have to be taken with a grain of salt, as the logit models throw warnings when executed, since the data has too little variance within the variables to create a well-fitting model.

SVM and kNN do not have this restriction of too many variables. kNN has a clear decrease

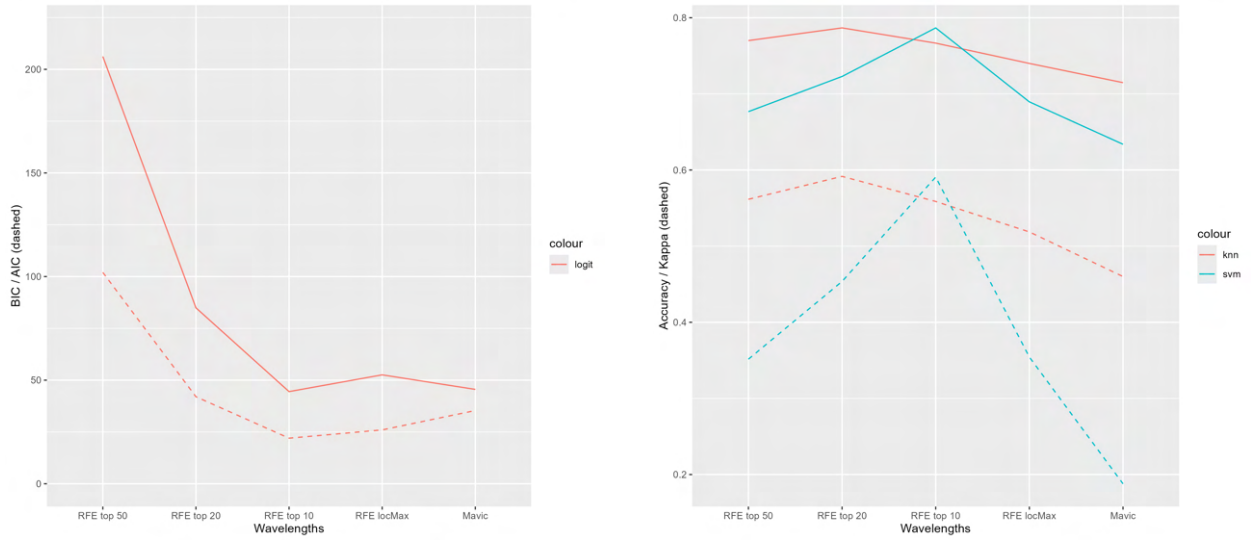


FIGURE 3.14: Results of the models compared over different wavelength selection. Left: logistic regression classifier (full line: BIC, dashed line: AIC); Right: kNN and SVM classifier (full line: Accuracy, dashed line: Kappa score)

after the top 20 variables, while SVM has the highest accuracy and kappa value with the top 10 dataset.

3.4 Early detection

Name	Accuracy	F1	Precision	Recall	Kappa
Day 1 SVM	0.846	0.800	1.000	0.667	0.683
Day 2 SVM	0.692	0.500	1.000	0.333	0.350

TABLE 3.5: Summary of ML results of early symptom modelling.

As a preliminary test, data from the first two days (11-07-2023, 27-07-2023) was used to train an SVM to see whether it could distinguish between infected and healthy vines before they were fully symptomatic (inspired by Rumpf et al., 2010). The data for these days is very limited (7 controls, 6 infected each) and therefore it did not make sense to split the set in training and test sets. There was also no equivalent data from Rütihof (as they presented much more severe when measured). The presented results (see Table 3.5) show that on the first day, the samples can be classified with an 85% accuracy and an acceptable kappa value, however on day two, it seems less efficient with 70%. We can also see that on both days, the issue lay with capturing all symptomatic plants, as some were misclassified as healthy, with 33% and 67% misclassification respectively.

3.5 RPAS indices

To show how RPAS detection might look like when using Pix4D, a set of indices (NDVI, GRVI, ARI) were calculated on the orthomasaic (Figure 3.16). Vines are mostly in the red and green mixed zone, below each green row. The green rows are actually the shadowy grass between the

rows of vines. As can be seen, the rows of vines don't stand out, mostly because it is already very late in the season and the growing has stopped. The top of the vines seems to still be active, as there are mixed responses from the rows. All indicators seem to give a similar response, with the NDVI showing the 'greenest' within the rows of vines and the GRVI the 'reddest' for the top of vines. With the default settings of Pix4D (with which all images here were generated), the GRVI values below 0.25 are left out. This makes it easy to see that there are very low values from the stems and dirt areas in the vineyard. The groups could be manually adjusted for the other maps, but then the equal area criterion is not met anymore, which makes the maps less comparable.

3.6 Asymptomatic leaves

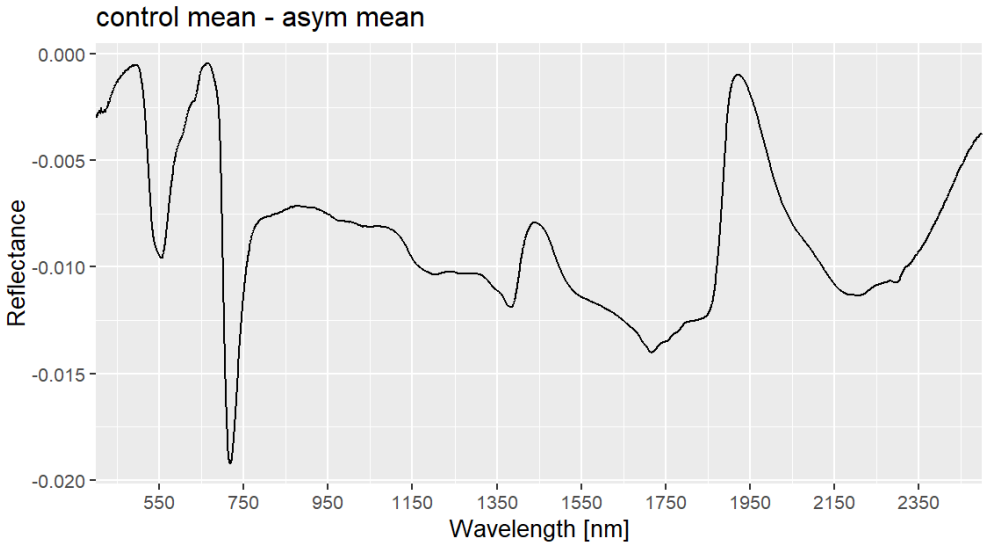


FIGURE 3.15: Difference between the mean of Weber control plants and asymptomatic leaves.

To assess the differences between the control plants and the asymptomatic plants, a difference was calculated between mean spectra (see Section 2.4.2). The difference only shows a maximum variation of reflectance of 0.02 in magnitude (Figure 3.15). Also, an SVM was trained to check if there was a discernible difference, but the accuracy was 40%.

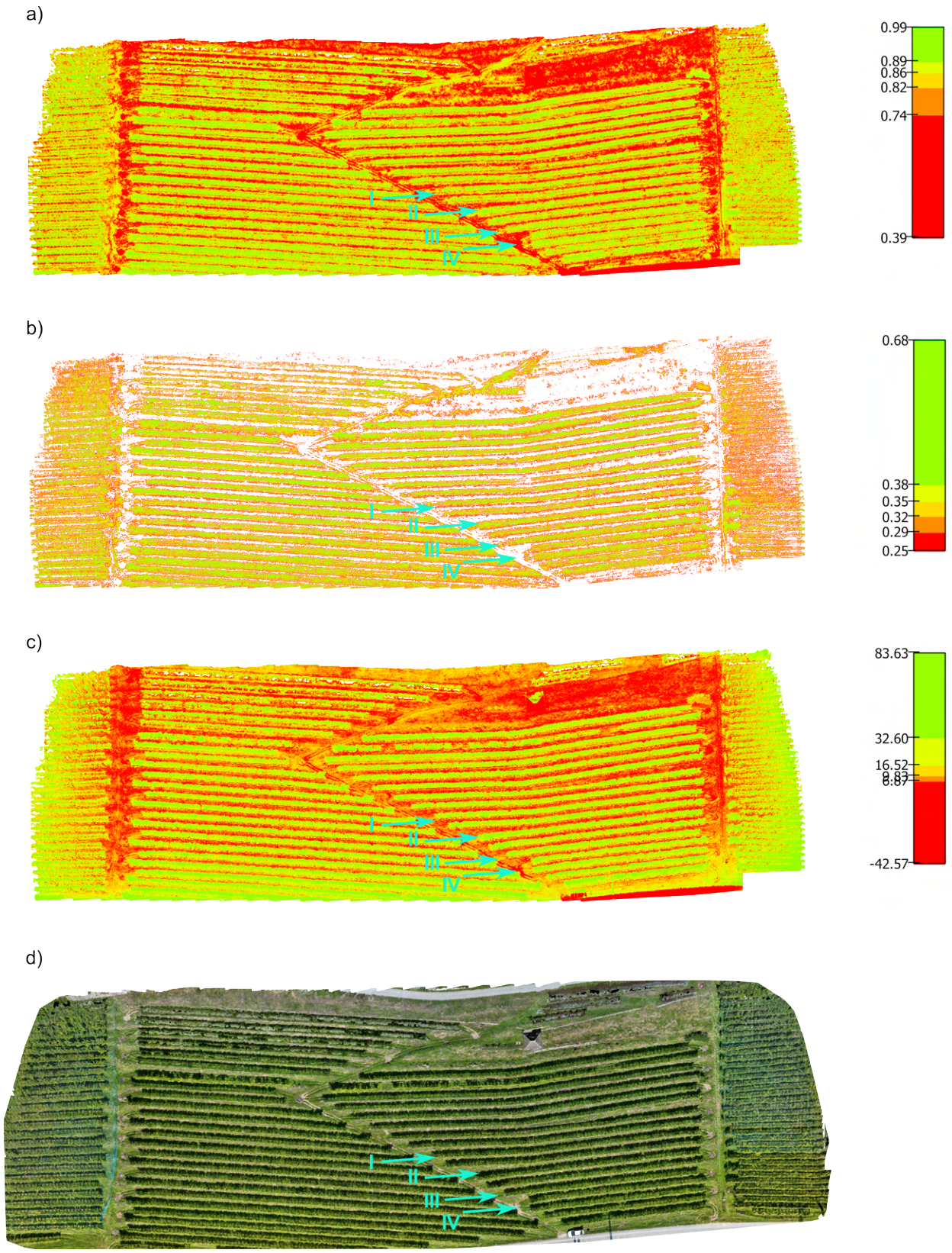


FIGURE 3.16: Orthomosaic of RPAS images 29.09.2023, exported from Pix4D, with respective color scale next to it. The scales were generated by Pix4D and set to have equal area of pixels per group. a) NDVI - bormalised vegetation index, b) GRVI - green red vegetation index, c) ARI - anthocyanin reflectance index, d) RGB orthomosaic for reference. Arrows indicate example areas of interest: I) vine row, II) grass shadowed by the vine row, III) grass, IV) dirt.

4 Discussion

4.1 ML methods

To find a suitable machine learning algorithm (ML) is not an easy task, as there are hundreds of options when looking at literature and programming environments. In this thesis, three of the more common algorithms in remote sensing were tested to differentiate bois noir (BN)-infected vines from healthy vines. This approach of differentiating with machine learning was used by other authors (Al-Saddik et al., 2017, 2018; Bendel et al., 2020; Rumpf et al., 2010), most of which concluded that the methods have to be adapted to the dataset, mainly due to variation in grape varieties and the differences in how diseases present themselves. Yellowing grape vine diseases were found to be differentiable with support vector machines (SVM) with a radial basis function kernel (Al-Saddik et al., 2019b; Bendel et al., 2020). This thesis also found an 80% accurate classification with SVM on unseen test data and 96% accuracy and kappa of 0.91 with a separate validation dataset. A k-nearest neighbour classification was close second, yielding 80% accuracy on test data and 92% accuracy and kappa of 0.83 on the validation dataset. The validation dataset is from a different location and grape variety than the training and test set (Zweigelt vs Pinot noir), which could mean that that Zweigelt has a more distinct spectral response when infected. Another reason could be that most datapoints of the validation dataset were recorded late into the season (Table 2.1). This would mean that the symptoms are more pronounced and therefore the spectral response more intense. It could also mean that the important variables determined by the SVM are transferable to some degree.

The main limitation in the analysis presented here is the limited number of data points. Twenty-six control and thirty-one symptomatic spectra were used in the analysis, which is not sufficient to reliably train machine learning models, which generally require hundreds of datapoints. However, the results are comparable to those from literature (e.g. Al-Saddik et al., 2019b) which claim over 90% classification accuracy on grape vine yellowing diseases with SVM. The findings also agree with literature that the model is accurate in precision (no false positives) but that it is not great with false negatives, i.e. it does not find all infected plants, but does not mark non-infected plants.

4.2 Important wavelengths

To use this approach more widely for bois noir (BN) or flavescence dorée (FD) monitoring, the idea to use a field-spectrometer to measure every plant is not feasible. One alternative would be to use a high-resolution spectral camera on a RPAS, but this is very costly and would not fit the

idea of wide-spread use. Ideally, light-weight and commercially available RPAS like the Mavic 3 multispectral (M3M) would be used for monitoring. Therefore, it is important to know which wavelengths of the spectra are most important for the classification of BN. There are many ways to approach this, and a recursive feature elimination was chosen as the main method (Colkesen & Kavzoglu, 2016; Ramezan, 2022). The choice was made since SVM is often used successfully to classify spectral data and since feature elimination techniques have been used successfully not only in remote sensing, but biomedical research (e.g. Guyon et al., 2002). For comparison of the resulting wavelengths, some other methods were explored, namely: RF-RFE (Demarchi et al., 2020), RF-GA (Al-Saddik et al., 2017; Mitchell, 1996), RF univariate filtering (Kuhn, 2008), HSC-PA (C. Li et al., 2023a) and a simple PCA (Wold et al., 1987). These methods produced various results, with the main SVM-RFE method identifying wavelengths at 520, 650, 712, (790,) 1596, 2007, 2139 and 2411 nm as important. The random forest-based analyses focuses on three to four main regions: around 540, 710, 1500 and 2400 nm. The PCA on SVM-RFE selected wavelengths showed a bias to 712 nm for infected and a bias of 520 nm and 2140 nm for uninfected. The HSC-PA was not used to get numerical results or for prediction, but showed high variation in the VIS regions (500 - 700 nm) and some variation in the SWIR regions, mostly around 2000 nm and 2500 nm. This is in accordance with the other findings.

Taken together, the different methods mostly identified similar wavelengths as important. The easiest to interpret and also the ones most often resulting from the methods are the ~520 nm and 710 nm features. Those correspond to green (~520 nm) and red (710 nm) in the visible spectrum. This is due to the fact that BN-infected leaves turn red when they turn symptomatic (see Figure 2.2 and Figure 2.4) and that the pigment content influences the light reflectance of the leaf (Jacquemoud & Ustin, 2019). This effect is however not only limited to the late stage of the symptoms, but also the case when we focus on July data (see results 3.2), where the leaves have not turned fully red yet. There, the SVM-RFE method identified these wavelengths as important. It is possible that this region is important even for detecting early symptom development as well.

The 520-550 nm range is influenced heavily by chlorophyl, carotenoid and anthocyanin content (Jacquemoud & Ustin, 2019, p. 182). Chlorophyl and carotenoids are photosynthetically active and therefore present during the growing season, while anthocyanins are produced during autumn. High chlorophyll content means that the 550 nm peak is more present (as the leaf is greener and absorbs more blue and green light) and high anthocyanin content means a shift towards the red peak at 650 nm. A high carotenoid content, which produces yellow, orange to brown leaf discolorations, leads to a 'stair' effect (an increase at around 600 and 630 nm), which can be seen in the mean of the mildly-symptomatic leaves (Figure 2.4). The absorption features of chlorophyll a+b and carotenoids are also well established (Blackburn, 1998, as discussed in Al-Saddik et al., 2019b) at 470 nm (carotenoids), 635 nm (chlorophyll b) and 680 nm (chlorophyll a). The content of these pigments is not a definitive answer to why the models select these wavelengths but helps to imagine why some of these features stand out.

The 710 nm feature, which is present in all results, as well as a large focus of the random forest analyses, is in the so-called red-edge area, which spans from ~680 nm to ~750 nm and denotes the

wavelengths which are on the edge of the visible spectrum. This region is photosynthetically active and reflects a lot more of the incoming light than the rest of the visible spectrum. Horler et al., 1983 showed that the red-edge position is tightly linked to chlorophyll content across all species. Wang et al., 2020 also notes that these wavelenghts are coupled to chlorophyll and carotenoids. From the PCA results, we can assume that a higher value in at 710nm is linked to being infected, a loss in productivity as anthocyanin levels increase, or both.

The region of 685-690 and 730-740 nm is also known to be heavily influenced by sun-induced chlorophyll fluorescence (G. H. Mohammed et al., 2019). While the exact wavelengths do not fully coincide with the results presented here, they might still influence the measurements, as these boundaries are usually not very sharp due to differences in plants, measurement equipment and so on. Fluorescence is energy that is emitted by the plant when it does not need all the energy received by the sunlight. So, when a plant is stressed and perhaps not able to use all light for photosynthesis, maybe also due to a decrease in chlorophyll, it often is more fluorescent. Therefore, an increase in fluorescence might be an indicator of an infected plant. This is also supported by Al-Saddik et al., 2019b who discusses a chlorophyll decrease from infected leaves.

The SVM-RFE method produced a feature at 790 nm, which is present in the overall and August results, but not at all in the July results. This wavelength denotes the beginning of the Near Infrared (NIR) region and is generally attributed to leaf structure, with very little absorbance and influence of pigments (Jacquemoud & Ustin, 2019). With it being lightly present in August, one could interpret that this area becomes more differentiated when the leaf starts changing its shape (internal and external), such that the light returned is influenced. This 'ledge' is sometimes visible in the spectra, just above the red-edge in infected leaves (visually assessed only, see Figure in Appendix A.2).

Water is the next major influencer of leaf spectra, with major absorptance features at 1450, 1940 and 2500 nm. These are visible in all spectra, with a huge dip in the reflectance (see Figure 2.4 and Appendix A.2). The random forest-based analyses focus on regions at 1500 and 2400, but not necessarily on the dips directly. The SVM-approach also does not focus on these features themselves, but on adjacent values, which are still influenced by water absorption. Not only water influences the absorption in the SWIR region, but according to Jacquemoud and Ustin, 2019 these regions also depend on abundant cellular organic compounds of the leaf cell. These compounds are partially or entirely structural. 1500-1600 nm and 2000 nm are largely influenced by cellulose, sugar, and starch, while 2150 nm and 2400 nm are influenced by protein and lignin. The spectra seem to be influenced by a combination of the above-named factors, especially in wavelengths higher than 1800 nm. Determining exactly what the influence per wavelength is, is not entirely explored, but attempts have been made, by e.g. Wang et al., 2020. We don't have the data in this dataset to assess the relative influences in the SWIR. These influences may be from water content, contents of non-pigment organic compounds, or a combination of the two. For the application presented in this thesis, a next step based on both cost-benefit analysis and known disease mechanisms would be to assess what portion of these differences can be attributed to differences in leaf water content, as determining the other factors is a lot more resource intensive.

As a summary, the wavelengths that are determined by the feature selection methods are influenced by a) pigments in the VIS and b) water content and cellular organic compounds in the leaf. All feature selection methods identified features not only in the VIS, but also the SWIR spectral region. Additionally comparing the success of models using all wavelengths versus the "important" selection versus the Mavic 3M wavelengths indicates that including the wavelengths in the SWIR in a model improves detection.

4.3 Wavelengths compared to literature

As the relevancy method is largely based on the idea of Bendel et al., 2020, we compare to their results first. For BN field measurements on red varieties, the commonalities lie at around 520 nm, 650 nm, 712 nm, 1590 nm, 2140 nm and 2410 nm. Their results deviate $\pm$20 nm from mine (with the closest being 712 vs 718 nm), but they are very close to each other. When looking at their relevancy curves, we can also see that these values are not necessarily to be seen as exact values, as they are center-wavelengths. Therefore there is some leeway in interpretation whether the exact numbers are that meaningful. I believe that for all wavelengths there is a certain margin of error, due to measurement differences/inaccuracies, variation in calculation methods used and fuzziness of biological traits. Compared to Bendel et al., 2020, the main differences are that they identify strong importance in the 620-650 nm range and highest importance at 1590 nm, which in my analysis shows up as a relevant, but weaker signal, included in the top 50 selection. This puts the region in roughly the 3rd important spot, behind the VIS and far SWIR regions. I personally don't think that the exact relevancy of center wavelengths can be determined, but that regions of interest, perhaps based on biological traits could be ranked based on the relevancy results. E.g. influence of cellulose, sugar and starch at 1500-1600 and 2000 nm (as mentioned above) ranked by the importance of 1590 nm and 2000 nm features. Serbin et al., 2014 also discusses the varied importance of leaf structural components per wavelength and notes that the peaks might shift due to compositions of the leaf. Wang et al., 2020 provides a detailed placement of leaf traits and their respective important wavelengths for models. While this placement is very helpful to gauge the area of the spectra where certain features are to be expected, I am unsure if the high resolution provided by them is to be trusted in imaging applications. Their results are also on large-scale forests, observed with airborne sensors which might influence the exact wavelengths.

Compared to Al-Saddik et al., 2019b and Al-Saddik et al., 2017, the results might not be directly transferrable, as they did their analyses on FD, but similarities can also be found. These similarities are especially important for the SmartGrape project, which this thesis is a part of, as it aims to differentiate BN and FD. Their results for red varieties and both severities of infection suggest that the most important features lie at 1485, 1725 and 2226 nm (Al-Saddik et al., 2017). They only focused on three wavelengths for their spectral disease index but vary them for different grape varieties. They also discuss the importance of the carotenoid band at 550 nm and a various wavelengths between 550 and 700 nm as influenced by pigments. They found that the ARI index was one of the best indices to determine FD in red varieties. They also point out that reflectance at around 700 nm is an essential feature of green vegetation. Due to changes in chloro-

phyll production in diseased plants, this reflectance could be an "unspecific indicator to detect diseased crops" (Al-Saddik et al., 2017). They also mention sugars, starch and proteins being a driver of changing reflectance in infected plants, but they situate it between 800 and 1400 nm, which is complementary to my results and also supported by Jacquemoud and Ustin, 2019.

4.4 Modeling fewer wavelenghts

The next step is to explore the influence of the reduction of features in the model generation. To this end, the same models as used for the full dataset were trained and tested on the SVM-RFE-defined 'best' wavelengths, with a local maxima finder (see results for local maxima 3.2.2). There we see that the SVM performed similarly in training and testing on the full dataset, but perfectly classified the validation set. As already discussed, this could just be due to lack of data availability but puts SVM as the best classifier for this data. kNN is close behind again, with an elevated accuracy in all metrics compared to the full dataset. The random forest classifier performed similarly to the full dataset and did not benefit from the feature space reduction. This is also reflected in the variable importance plot which has nine out of ten wavelengths around 700 nm in the ten most important wavelengths. These results show that decreasing the feature space leads to an improvement in performance for models that use the data well. Similar findings also come from Al-Saddik et al., 2017 and Albetis et al., 2019 and is generally viewed as good practice, to minimize the impact of irrelevant variables (Becker et al., 2021; Kuhn, 2008; Mitchell, 1996). With a higher percentage of the variables containing meaningful information, the model complexity decreases (e.g. less decision trees) which decreases computation time and decreases the chance of misclassification.

When comparing the feature reduction to just the wavelengths that represent the DJI Mavic 3M, the results become mostly worse, with the only increase in performance for the logit model, which for the first time has significance in the coefficients (see Figure 3.13). While the logit model performs well in this case on train and test datasets, it fails on external validation, showing that it might be useful when classifying within a field for e.g. a next season, but beyond that should not be used for modelling spectra due to their complexity. The decrease in accuracy with the Mavic wavelengths might be remedied if more data were available and should be explored in the future of the SmartGrape project.

For further comparison, a CV approach was used to train models on the full dataset of Weber with varying selected wavelengths to see whether the feature reduction in itself is important, or whether the localized maxima perform better. It should be mentioned again that the implementation of the algorithms might differ to the rest of the analysis for SVM and kNN, as the CV approach uses the caret inbuilt algorithms and therefore the accuracies might vary. We see a clear indication that ten wavelengths are a good number even for SVM, which is known to handle high dimensional data very well. The kNN and SVM show that the localized maxima perform worse than simply using the top 10 wavelengths determined by the SVM-RFE. This might stem from the importance curve, which puts less emphasis on wavelengths 800 - 1800 nm, of which none are present in the top 10, but due to local maxima included in the other selection. It is not surpris-

ing that an SVM performs better on the best features selected by an SVM approach and that this difference is less pronounced in the kNN.

What this thesis is not suggesting, is that only the top 10 wavelengths should be used, but that the corresponding regions should be considered. Every approach points to a high importance at 650 nm, 712 nm and wavelengths in the SWIR (2000+ nm). This is complementary to the prior discussion, which showed that some of the features included in the local-maxima wavelengths include regions that are valued by other papers, like 1400 and 1600 nm, which are not included when using the top results. A weighting of the wavelengths based on the importance curve could improve the performance of the local-maxima-selected wavelengths, but it could not be implemented in this thesis. The used models could only weight the different training plants by importance, not the wavelengths. This might be attempted again later in the project, as to represent all regions of the spectra as well as possible and not leave out a range of 1000 nm.

4.5 Outlook

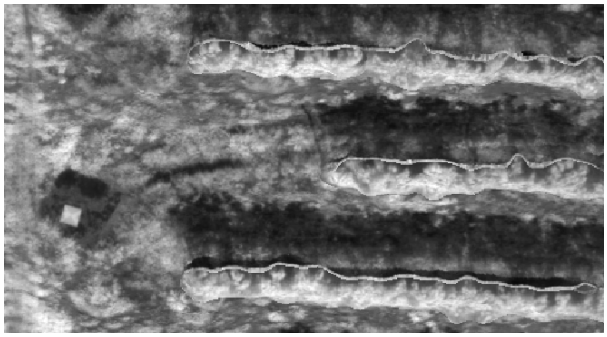
4.5.1 Proposed workflow for RPAS imagery

The following is a proposed workflow for the RPAS imagery, which was originally planned as part of this thesis. However, using high-resolution spectroscopy data to determine relevant wavelengths took first priority.

There are two main ways in which processing might be done. One is to use the images directly, without any orthorectification or stitching done by Pix4D. This approach is similar to Albetis et al., 2017 and has the advantage that less processing is done that might introduce errors. The other is to use the Pix4D voxel cloud and the generated orthomosaics. The issue with the orthomosaics in Pix4D is that the 3D pixel cloud is calculated from the images and the images are then re-projected onto that model to create a planar view. This introduces artefacts, which might not be radiometrically correct, where the rows of vines stand high above the rows of grass in-between (see Figure 4.1). These artefacts might also exist because of the side-looking orientation of the camera, which means that sideways images are projected from a top-down view, which Pix4D might not correct for. The backside of the vines is not visible to the camera at any point, so it might project a top-of-plant pixel on the backside and therefore the seams don't look natural, but stand out as border lines. At least that would be an explanation for the 'upper' border.

In either case, the first step is to distinguish between grass/floor pixels and vine pixels (Albetis et al., 2017; Polder et al., 2014). For the single images, this would have to be done with each image and the relevant pixels are to be extracted. In the case of the orthoimage, this could produce a mask which can be repurposed for all dates to easily extract the relevant vegetation pixels. The second step is to create a buffer around the GNSS coordinates of the infected vines, to mostly include the correct vine. This is a major error source, as the vines don't grow separately from each other, but intertwine in an unpredictable way. So much in fact, that for the field spectroscopy measurements it often took a while to make sure that the leaves measured belonged to the correct plant. Also, not all leaves of an infected plant show symptoms, so every infected buffer area would include

(A) RGB channel of Mavic 3M, orthomosaic processed by Pix4D.



(B) NIR channel of Mavic 3M, orthomosaic processed by Pix4D.

FIGURE 4.1: Orthomosaics in Pix4d showing one exemplary part of Weber vineyard, where line-like artefacts around the vine-rows can be seen due to projection distortion.

many non-symptomatic pixels anyways.

This step is also much more complicated for the single images and would probably have to be changed since the images are taken with a side-view and determining exact positions in a 3D space from a 2D image is not trivial. For the single images, the approach would have to be done on a pixel-by pixel basis and manually identified infected pixels would have to be used as a ground-truth. Albetis et al., 2017 has found a solution for this, however they took images from a top-down view and their vines showed very clear symptoms over the whole plant (70%+), which made it easier to create a clear boundary.

With the infected pixels classified and extracted, a classifier would have to be trained in order to detect infected pixels from uninfected ones in new data. This would follow a similar approach as was presented is this thesis with the spectroscopy data.

4.5.2 BN detection with RPAS

The early detection explored in this thesis shows that it is possible to classify data based on early symptoms with an SVM and full spectra, however this should be tested again with more data in the next years of the SmartGrape project. It seems also plausible with the rest of the data, that early detection is possible with a reduced feature space and therefore might be done with RPAS. The main issue that we see here is that there seems to be a great importance on some wavelengths in the SWIR region of the spectra, which are not provided with the commercially available M3M or most other RPAS-compatible sensors. We also see that the 1600 nm feature is less important in July and gains more towards August, while 2000+ nm features are more important in July and less in August. This points to an importance of those SWIR wavelengths for early detection.

The M3M wavelengths roughly correlate with what has been found in this thesis, even though the center wavelengths might be better situated at slightly different wavelengths than the M3M provides. The Matrice 300 with the red-edge dual sensor setup (see Table 2.2) might provide a better coverage of the wavelengths explored in this thesis, with two sensors at around 712 nm, one at 530 and one at 560 nm. These improvements would have to be explored in more detail

and they might not be worth pursuing considering that the Mavic 300 is bulkier, pricier and less available.

Counterintuitively to what has been presented in this thesis (except maybe the results of comparing the wavelength selection 3.3), it might be a good approach to create more 'channels' from RPAS data. This would mean for example, that each channel is recombined to a new value in such a way that more information about each pixel is generated. This is not a new concept and is the basis of vegetation indices (VIs) which are widely used in remote sensing. This has been briefly explored here by creating NDVI, GRVI and ARI maps (see results 3.5), but on singular index maps there was not even a clear differentiation between vines and surrounding vegetation, mostly shadowy grass stands out. However, different indices show different levels of differentiation (e.g. the GRVI map has somewhat clear red rows which indicate vines) and a combination of them could lead to better results. This could also transfer to a per-leaf-resolution, such that infected plants and their neighbours could be differentiated. E.g. Albetis et al., 2017 showed that it is possible to classify FD pixels with a combination of vegetation indices on vines that show nearly full symptoms, so it should be possible to create an approach similar to theirs. Their plants showed very clear FD symptoms, as mentioned above, which makes annotating a training and test set and therefore validation easier. With high resolution/close-up BN-infected images, it would be possible to have multiple pixels per BN-infected leaf. Therefore, fewer annotated leaves would be needed for the same training effect, which decreases the amount of images needed. A trained model could then potentially be scaled to a lower resolution (images taken from further away), to increase detection efficiency.

Al-Saddik et al., 2017 also explored an approach to identify FD using spectral indices and even created their own, with the caveat that their research is leaf-based, not on a full scale for a field of vines. Al-Saddik et al., 2019b concluded that a mixture of scatter correction methods and spectral derivatives are best at detecting FD on a leaf basis. They even propose to build a multispectral camera to leverage their findings, but as far as I could tell, this was not yet reported. The studies available for comparison with the work in this thesis have generally been performed in lab conditions, with a regulated background or on vines that are fully symptomatic. As was already mentioned in section 2.3.6, the less clear presentation of symptomatic leafs per plant in the used vineyards will make it more difficult to identify diseased plants. Based on the spectroscopy results, the ideal scenario would be to combine a conventional RPAS with the VNIR bands and measurements from some sensor that includes far SWIR wavelengths. This is a hypothetical system as of now but could lead to more robust results. The main issue with such a system is that cameras that are able to detect SWIR wavelengths usually require a cooling system and are therefore bulkier, heavier and more expensive.

Some recent studies report promising results with custom sensors (Barjaktarovic et al., 2024; Barjaktarović et al., 2023) in the VIS to NIR range. They classified FD with 'hyperspectral' data using linear discriminant analysis and went on to produce their own multispectral camera with a thermal sensor from cheap and easily available parts. Their choice of a thermal camera (thermal infrared being of wavelengths 3'000 - 14'000 nm) might be a complement to the SWIR sensor

mentioned in this thesis, or indicates that adding a SWIR sensor might be possible, but this would have to be assessed separately. Their system seems to be a field-focused version and one should follow their research in the near future to see whether they can develop a robust version that could be mounted on a RPAS. In Barjaktarovic et al., 2024 they discuss open source methods to co-register images from the sensor (align them pixel by pixel), which DJI handles by measuring the offset of the camera centers between each other and shifting the images respectively. For self-built systems, where sensors might not be in an exact plain or the focal lengths of the cameras are not the same, having a co-registration step is important, so that in the final analysis the pixels match between different bands. This study shows that it would be possible to construct a camera with VIS and thermal detection with tailored bands for increased accuracy. Whether this system could be mounted to a RPAS is still under examination.

5 Conclusion

This thesis aimed to determine whether machine learning can be used to identify BN in Pinot noir from spectroscopy data and transfer the models to Zweigelt, which it achieved by means of a SVM model and compared to other literature. The transferability showed great promise, which could have been due to late-stage symptoms in the Zweigelt plot compared to the Pinot noir plot. This showed however, that the symptoms show up in similar regions of the spectra, changing mostly in amplitude. The resulting accuracy of 90+% is comparable to other approaches, with the caveat that the results reported here are from a pilot dataset with a small number of data points. The evaluation criteria used, namely accuracy in combination with F1 score, precision and recall was useful to evaluate how the different models performed for detecting infected plants. The Cohens' kappa score was additionally used to identify mismatches with classifications that performed well based on the other indicators. It helped pinpoint mismatches between actual and predicted values and pointed to unreliable results from SVM trained with few features.

The selection methods used to determine the most important wavelengths pointed to important features at 520 nm, 650 nm, 710 nm (very important for RF classifiers) and wavelengths >2000 nm. There were also some other importances at 1400 nm and 1600 nm, with 1600 nm being valued by some literature. These values are not to be seen as exact, but mostly indicate regions of importance. These wavelengths correspond to changes in pigments (VIS) and structure, water, and non-pigment components of the leaf (NIR and SWIR) as the season progresses and the symptoms grow more pronounced. To achieve early detection, a preliminary test on a small dataset indicates that it is possible with the full spectrum and has to be further evaluated whether it works with fewer wavelengths (i.e. with Mavic 3M wavelengths).

The last question posed by this thesis is whether (early) detection can be done with a Mavic 3M RPAS, or other commercially available RPAS systems. The initial estimate based on different indices showed no clear distinction between infected and non-infected plants, or any plants for that matter, but was only carried out on single indicators. Other authors have shown that a combination of indices can lead to better results, even if only the wavelengths from an RPAS are used, effectively increasing the number of features / channels by recombining the 4 main ones. This approach will have to be evaluated on the data for this project when more data are available. With more data, which are collected this year, the SmartGrape project will revisit the approaches presented here, test the conclusions of this thesis and develop the approaches further. The ultimate goal would be to use commercially available RPAS or ground-based systems to reduce the manual workload of farmers in the field.

To conclude, it would be beneficial to create a system specifically tailored to wavelengths 520

nm, 650 nm, 712 nm, and at least one SWIR wavelength (e.g. 1600 or 2140 nm). This system would be able to detect the pigment changes as well as the physiological and water content changes that an infected leaf goes through and would allow for a more nuanced classification. RPAS systems that are on the market can lead to promising results already, if a combination of vegetation indices (VI) is used. It is still to be determined if they work well on field data, as most current research is lab based. This thesis was not yet able to detect meaningful differences in Mavic 3M data, but a combination of VIs was not tested. To minimise manual annotation for future projects, close-up images of infected and non-infected leaves should be annotated, so that an identified leaf could provide many annotated pixels. A model would then be trained and tested on the clearly symptomatic and non-symptomatic pixels. The model could then be used on images taken from further away, where an infected leaf might only be captured on a few pixels. This could remedy some of the single vine detection issues presented above. This approach is best when using images that are taken from an angle, so that as many symptomatic leaves are visible by the sensor as possible.

Bibliography

Agroscope. (2024). Goldgelbe Vergilbung. http://ggv.agroscope.ch/

Akaike, H. (1974). A New Look at the Statistical Model Identification. In E. Parzen, K. Tanabe & G. Kitagawa (Eds.), *Selected papers of hirotugu akaike* (pp. 215–222). Springer New York. https://doi.org/10.1007/978-1-4612-1694-0{_}16

Albetis, J., Duthoit, S., Guttler, F., Jacquin, A., Goulard, M., Poilvé, H., Féret, J. B., & Dedieu, G. (2017). Detection of Flavescence dorée grapevine disease using Unmanned Aerial Vehicle (UAV) multispectral imagery. *Remote Sensing*, *9*(4). https://doi.org/10.3390/rs9040308

Albetis, J., Jacquin, A., Goulard, M., Poilvé, H., Rousseau, J., Clenet, H., Dedieu, G., & Duthoit, S. (2019). On the potentiality of UAV multispectral imagery to detect Flavescence dorée and Grapevine Trunk Diseases. *Remote Sensing*, *11*(1). https://doi.org/10.3390/rs11010023

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. Springer, Cham. https://doi.org/10.1007/978-3-030-22475-2{_}1

Al-Saddik, H., Simon, J. C., & Cointault, F. (2017). Development of spectral disease indices for flavescence dorée grapevine disease identification. *Sensors (Switzerland)*, *17*(12). https://doi.org/10.3390/s17122772

Al-Saddik, H., Laybros, A., Billiot, B., & Cointault, F. (2018). Using image texture and spectral reflectance analysis to detect Yellowness and Esca in grapevines at leaf-level. *Remote Sensing*, *10*(4). https://doi.org/10.3390/rs10040618

Al-Saddik, H., Laybros, A., Simon, J. C., & Cointault, F. (2019a). Protocol for the Definition of a Multi-Spectral Sensor for Specific Foliar Disease Detection: Case of Flavescence Dorée. In M. Rita & L. Pagliari (Eds.), *Phytoplasmas: Methods and protocols* (pp. 213–238). Springer New York. https://doi.org/10.1007/978-1-4939-8837-2{_}17

Al-Saddik, H., Simon, J. C., & Cointault, F. (2019b). Assessment of the optimal spectral bands for designing a sensor for vineyard disease detection: the case of Flavescence dorée. *Precision Agriculture*, *20*(2), 398–422. https://doi.org/10.1007/s11119-018-9594-1

Audebert, N., Le Saux, B., & Lefevre, S. (2019). Deep learning for classification of hyperspectral data: A comparative review. *IEEE Geoscience and Remote Sensing Magazine*, *7*(2), 159–173. https://doi.org/10.1109/MGRS.2019.2912563

Backhaus, A., Bollenbeck, F., & Seiffert, U. (2011). Robust classification of the nutrition state in crop plants by hyperspectral imaging and artificial neural networks. *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*. https://doi.org/10.1109/WHISPERS.2011.6080898

Barjaktarovic, M., Santoni, M., & Bruzzone, L. (2024). Design and Verification of a Low-Cost Multispectral Camera for Precision Agriculture Application. *IEEE Journal of Selected Top-*

ics in Applied Earth Observations and Remote Sensing, *17*, 6945–6957. https://doi.org/10.1109/JSTARS.2024.3377104

Barjaktarović, M., Santoni, M., Faralli, M., Bertamini, M., & Bruzzone, L. (2023). Data acquisition for testing potential detection of Flavescence dorée with a designed, affordable multispectral camera. *Telfor Journal*, *15*(1), 2–7. https://doi.org/10.5937/telfor2301002B

Becker, F., Backhaus, A., Johrden, F., & Flitter, M. (2021). Optimal multispectral sensor configurations through machine learning for cognitive agriculture. *At-Automatisierungstechnik*, *69*(4), 336–344. https://doi.org/10.1515/auto-2020-0069

Bendel, N., Backhaus, A., Kicherer, A., Köckerling, J., Maixner, M., Jarausch, B., Biancu, S., Klück, H. C., Seiffert, U., Voegele, R. T., & Töpfer, R. (2020). Detection of two different grapevine yellows in Vitis vinifera using hyperspectral imaging. *Remote Sensing*, *12*(24), 1–22. https://doi.org/10.3390/rs12244151

Berk, U. (2017). How to choose a predictive model after k-fold cross-validation? - Cross Validated. https://stats.stackexchange.com/questions/52274/how-to-choose-a-predictive-model-after-k-fold-cross-validation/650945#650945

Bertaccini, A., Duduk, B., Paltrinieri, S., & Contaldo, N. (2014). Phytoplasmas and Phytoplasma Diseases: A Severe Threat to Agriculture. *American Journal of Plant Sciences*, *05*(12), 1763–1788. https://doi.org/10.4236/ajps.2014.512191

Bertschinger, L., Schuman, M. C., Linus Reichert, Huggel, J., Debonneville, C., Geckeler, C., Kehrli, P., Mackie-Haas, K., Mintchev, S., Ramos, S., & Schneider, S. (2024). SmartGrape - toward early detection of invasive cicada-borne grapevine diseases through in-field spectroscopy and plant volatile sensing. *Procceedings of the European Horticulture Congress*.

Blackburn, G. A. (1998). Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *International Journal of Remote Sensing*, *19*(4), 657–675. https://doi.org/10.1080/014311698215919

Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983, January). AN OVERVIEW OF MACHINE LEARNING. In *Machine learning* (pp. 3–23). Elsevier. https://doi.org/10.1016/b978-0-08-051054-5.50005-4

Cavender-Bares, J., Meireles, J., Couture, J., Kaproth, M., Kingdon, C., Singh, A., Serbin, S., Center, A., Zuniga, E., Pilz, G., & Townsend, P. (2016). Associations of Leaf Spectra with Genetic and Phylogenetic Variation in Oaks: Prospects for Remote Detection of Biodiversity. *Remote Sensing*, *8*(3), 221. https://doi.org/10.3390/rs8030221

Cavender-Bares, J., Gamon, J. A., & Townsend, P. A. (Eds.). (2020). *Remote Sensing of Plant Biodiversity*.

Chambers, J. M. (1992, October). Linear Models. In T. Hastie (Ed.), *Statistical models in s* (1st Edition, pp. 95–144). Routledge. https://doi.org/10.1201/9780203738535-4

Cheng, Q., Varshney, P. K., & Arora, M. K. (2006). Logistic regression for feature selection and soft classification of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, *3*(4), 491–494. https://doi.org/10.1109/LGRS.2006.877949

Cheriyadat, A., & Bruce, L. M. (2003). Why Principal Component Analysis is not an Appropriate Feature Extraction Method for Hyperspectral Data. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *6*, 3420–3421. https://doi.org/10.1109/igarss.2003.1294808

Chuche, J., & Thiéry, D. (2014). Biology and ecology of the Flavescence dorée vector Scaphoideus titanus: A review. *Agronomy for Sustainable Development*, *34*(2), 381–403. https://doi.org/10.1007/s13593-014-0208-7

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Colkesen, I., & Kavzoglu, T. (2016). Performance evaluation of rotation forest for svm-based recursive feature elimination using hyperspectral imagery. *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*, *0*. https://doi.org/10.1109/WHISPERS.2016.8071792

Cover, T. M., & Hart, P. E. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Debonneville, C., Mandelli, L., Brodard, J., Groux, R., Roquis, D., & Schumpp, O. (2022). The Complete Genome of the Flavescence Dorée Phytoplasma Reveals Characteristics of Low Genome Plasticity. *Biology*, *11*(7), 953. https://doi.org/10.3390/biology11070953

Dehkordi, R. H., Burgeon, V., Fouche, J., Gomez, E. P., Cornelis, J. T., Nguyen, F., Denis, A., & Meersmans, J. (2020). Using UAV collected RGB and multispectral images to evaluate winter wheat performance across a site characterized by century-old biochar patches in Belgium. *Remote Sensing*, *12*(15). https://doi.org/10.3390/RS12152504

Demarchi, L., Kania, A., Ciezkowski, W., Piórkowski, H., Oświecimska-Piasko, Z., & Chormański, J. (2020). Recursive feature elimination and random forest classification of natura 2000 grasslands in lowland river valleys of poland based on airborne hyperspectral and LiDAR data fusion. *Remote Sensing*, *12*(11), 1842. https://doi.org/10.3390/rs12111842

Dubuis, P.-H., Gfeller, A., Künzler, L., Kehli, P., Linder, C., Reynard, J.-S., Debonneville, C., Spring, J.-L., Zufferey, V., & Mackie-Haas, K. (2021). *Pflanzenschutzempfehlungen für den Rebbau 2021/2022* (tech. rep.). Agroscope, Schweizerische Eidgenossenschaft. https://doi.org/10.34776/at370g

Dubuis, P.-H., Gfeller, A., Egli-Künzler, L., Kehrli, P., Linder, C., Reynard, J.-S., Debonneville, C., Spring, J.-L., Zufferey, V., Mackie-Haas, K., Blouin, A., & Verdenal, T. (2023). *Pflanzenschutzempfehlungen für den Rebbau 2023/2024* (tech. rep.). Agroscope. www.pflanzenschutz-rebbau.agroscope.ch

Fleiss, J. L., Levin, B., Paik, M. C., et al. (2003). The measurement of interrater agreement. In *Statistical methods for rates and proportions* (pp. 22–23, Vol. 2). Citeseer.

Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T., & Fralish, J. S. (1995). Parallel Analysis: a method for determining significant principal components. *Journal of Vegetation Science*, *6*(1), 99–106. https://doi.org/10.2307/3236261

Galetto, L., Bosco, D., & Marzachì, C. (2005). Universal and group-specific real-time PCR diagnosis of flavescence dorée (16Sr-V), bois noir (16Sr-XII) and apple proliferation (16Sr-X) phytoplasmas from field-collected plant hosts and insect vectors. *Annals of Applied Biology*, *147*(2), 191–201. https://doi.org/10.1111/j.1744-7348.2005.00030.x

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochemistry and Photobiology*, *74*(1), 38. https://doi.org/10.1562/0031-8655(2001)074<0038:opaneo>2.0.co;2

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1-3), 389–422. https://doi.org/10.1023/A:1012487302797

Ho, T. K. (1995). Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, *1*, 278–282. https://doi.org/10.1109/ICDAR.1995.598994

Horler, D. N., Dockray, M., Barber, J., & Barringer, A. R. (1983). Red edge measurements for remotely sensing plant chlorophyll content. *Advances in Space Research*, *3*(2), 273–277. https://doi.org/10.1016/0273-1177(83)90130-8

Hosmer, D. W., & Lemeshow, S. (2000, September). *Applied Logistic Regression*. Wiley. https://doi.org/10.1002/0471722146

Huang, C., Davis, L. S., & Townshend, J. R. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*(4), 725–749. https://doi.org/10.1080/01431160110040323

Hueni, A., Damm, A., Kneubuehler, M., Schlapfer, D., & Schaepman, M. E. (2017). Field and Airborne Spectroscopy Cross Validation -Some Considerations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(3), 1117–1135. https://doi.org/10.1109/JSTARS.2016.2593984

Hueni, A., & Bialek, A. (2017). Cause, Effect, and Correction of Field Spectroradiometer Interchannel Radiometric Steps. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(4), 1542–1551. https://doi.org/10.1109/JSTARS.2016.2625043

Jacquemoud, S., & Ustin, S. (2019, September). *Leaf Optical Properties*. Cambridge University Press. https://doi.org/10.1017/9781108686457

Karatzoglou, A., Hornik, K., Smola, A., & Zeileis, A. (2004). kernlab - An S4 package for kernel methods in R. *Journal of Statistical Software*, *11*, 1–20. https://doi.org/10.18637/jss.v011.i09

Kehrli, P., Kessler, S., Schaerer, S., & Delabays, N. (2009). Die Verbreitung der Schwarzholzkrankheit und ihres Überträgers in der Schweiz. *Schweizerische Zeitschrift für Obst- und Weinbau*, *14*.

Kehrli, P., Schaerer, S., & Delabays, N. (2010). *Hyalesthes obsoletus, vecteur du bois noir: répartition et biologie* (tech. rep.). Agroscope. https://www.researchgate.net/publication/265336526

Kessler, S., Schaerer, S., Delabays, N., Turlings, T. C., Trivellone, V., & Kehrli, P. (2011). Host plant preferences of Hyalesthes obsoletus, the vector of the grapevine yellows disease 'bois noir', in Switzerland. *Entomologia Experimentalis et Applicata*, *139*(1), 60–67. https://doi.org/10.1111/j.1570-7458.2011.01107.x

Kohavi, R. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection* (tech. rep.). http://robotics.stanford.edu/~ronnyk

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M., & Johnson, K. (2013, January). *Applied predictive modeling*. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Li, C., Czyż, E. A., Ray, R., Halitschke, R., Baldwin, I. T., Schaepman, M. E., & Schuman, M. C. (2023a). Association of leaf spectral variation with functional genetic variants Leaf reflectance-

Field spectroscopy-Multiparent Advanced Generation Inter-Cross (MAGIC)-Genome-wide association studies (GWAS)-Hierar-chical Spectral Clustering with Parallel Analysis-Nicotiana attenuata. https://doi.org/10.1101/2023.10.03.560760

Li, C., Czyż, E. A., Halitschke, R., Baldwin, I. T., Schaepman, M. E., & Schuman, M. C. (2023b). Evaluating potential of leaf reflectance spectra to monitor plant genetic variation. *Plant Methods*, *19*(1). https://doi.org/10.1186/s13007-023-01089-9

Li, H., Cui, J., Zhang, X., Han, Y., & Cao, L. (2022). Dimensionality Reduction and Classification of Hyperspectral Remote Sensing Image Feature Extraction. *Remote Sensing*, *14*(18), 4579. https://doi.org/10.3390/rs14184579

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. https://CRAN.R-project.org/doc/Rnews/

Maixner, M., Albert, A., & Johannesen, J. (2014). Survival relative to new and ancestral host plants, phytoplasma infection, and genetic constitution in host races of a polyphagous insect disease vector. *Ecology and Evolution*, *4*(15), 3082–3092. https://doi.org/10.1002/ece3.1158

Malvern Panalytical. (2023). ASD FieldSpec 4 - High Resolution Spectroradiometer | Malvern Panalytical. https://www.malvernpanalytical.com/en/products/product-range/asd-range/fieldspec-range/fieldspec4-hi-res-high-resolution-spectroradiometer

Martinez-Taboada, F., & Redondo, J. I. (2020). Variable importance plot (mean decrease accuracy and mean decrease Gini). https://doi.org/10.1371/journal.pone.0230799.g002

Mary L, M. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Meireles, J. E., & Schweiger, A. (2023, September). meireles/spectrolab: CRAN-v0.0.18. https://doi.org/10.5281/zenodo.5524855

Meng, Q., Cieszewski, C. J., Madden, M., & Borders, B. E. (2007). K nearest neighbor method for forest inventory using remote sensing data. *GIScience and Remote Sensing*, *44*(2), 149–165. https://doi.org/10.2747/1548-1603.44.2.149

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*, *10*(1), 1–16. https://doi.org/10.1186/1471-2105-10-213

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2023). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. https://CRAN.R-project.org/package=e1071

MicaSense RedEdge-MX Dual - Professional Drones | DJI Enterprise. (2018). https://www.aerialsolutions.be/en/product/micasense-rededge-mx-dual/

Mitchell, M. (1996). *An introduction to genetic algorithms*. MIT Press.

Mohammed, E. A., Naugler, C., & Far, B. H. (2015, August). Emerging Business Intelligence Framework for a Clinical Laboratory Through Big Data Analytics. In *Emerging trends in computational biology, bioinformatics, and systems biology: Algorithms and software tools* (pp. 577–602). Elsevier Inc. https://doi.org/10.1016/B978-0-12-802508-6.00032-6

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J. P., Miller, J. R., Guanter, L., Moreno, J., Moya, I., … Zarco-

Tejada, P. J. (2019). Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sensing of Environment*, *231*, 111177. https://doi.org/10.1016/j.rse.2019.04.030

Motohka, T., Nasahara, K. N., Oguma, H., & Tsuchida, S. (2010). Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology. *Remote Sensing*, *2*(10), 2369–2387. https://doi.org/10.3390/rs2102369

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1–21. https://doi.org/10.1186/s40537-014-0007-7

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

Pelletier, C., Salar, P., Gillet, J., Cloquemin, G., Very, P., Foissac, X., & Malembic-Maher, S. (2009). *Triplex real-time PCR assay for sensitive and simultaneous detection of grapevine phytoplasmas of the 16SrV and 16SrXII-A groups with an endogenous analytical control* (tech. rep. No. 2).

Petibon, F., Czyż, E. A., Ghielmetti, G., Hueni, A., Kneubühler, M., Schaepman, M. E., & Schuman, M. C. (2021). Uncertainties in measurements of leaf optical properties are small compared to the biological variation within and between individuals of European beech. *Remote Sensing of Environment*, *264*. https://doi.org/10.1016/j.rse.2021.112601

Polder, G., van der Heijden, G. W., van Doorn, J., & Baltissen, T. A. (2014). Automatic detection of tulip breaking virus (TBV) intulip fields using machine vision. *Biosystems Engineering*, *117*(100), 35–42. https://doi.org/10.1016/j.biosystemseng.2013.05.010

Prasad, S., & Bruce, L. M. (2008). Limitations of principal components analysis for hyperspectral target recognition. *IEEE Geoscience and Remote Sensing Letters*, *5*(4), 625–629. https://doi.org/10.1109/LGRS.2008.2001282

R Core Team. (2023). R: A Language and Environment for Statistical Computing. https://www.R-project.org/

Ramezan, C. A. (2022). Transferability of Recursive Feature Elimination (RFE)-Derived Feature Sets for Support Vector Machine Land Cover Classification. *Remote Sensing*, *14*(24), 6218. https://doi.org/10.3390/rs14246218

Rizzoli, A., Debonneville, C., Linder, C., & Kehrli, P. (2023). *Vergilbungskrankheiten der Rebe: Goldgelbe Vergilbung und Schwarzholzkrankheit-Beschreibung, Biologie und Bekämpfung* (tech. rep.). Agroscope. http://ggv.agroscope.ch

Rohit, K. (2022, December). F1 Score in Machine Learning: Intro & Calculation. https://www.v7labs.com/blog/f1-score-guide

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

Scrucca, L. (2013). GA: A package for genetic algorithms in R. *Journal of Statistical Software*, *53*(4), 1–37. https://doi.org/10.18637/jss.v053.i04

Serbin, S. P., Singh, A., McNeil, B. E., Kingdon, C. C., & Townsend, P. A. (2014). Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecological Applications*, *24*(7), 1651–1669. https://doi.org/10.1890/13-2110.1

SR 916.201 - Verordnung des WBF und des UVEK vom 14. November 2019 zur Pflanzengesundheitsverordnung (PGesV-WBF-UVEK). (n.d.). https://www.fedlex.admin.ch/eli/cc/2019/787/de

Swiss Positioning Service (Swipos). (2024). https://www.swisstopo.admin.ch/de/swipos-der-swiss-positioning-service

Tessitori, M., La Rosa, R., & Marzachì, C. (2018). Flavescence dorée and bois noir diseases of grapevine are evolving pathosystems. *Plant Health Progress*, *19*(2), 136–138. https://doi.org/10.1094/PHP-10-17-0057-MR

Trim, S. A., Mason, K., & Hueni, A. (2021). Spectroradiometer spectral calibration, ISRF shapes, and related uncertainties. *Applied Optics*, *60*(18), 5405. https://doi.org/10.1364/ao.425676

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Tuszynski, J. (2021). caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc. https://CRAN.R-project.org/package=caTools

van Diggelen, F. (1997). GPS and GPS+GLONASS RTK. *Proceedings of the 10th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1997)*, 139–144. https://www.ion.org/publications/abstract.cfm?articleID=2744

Vapnik, V., & Golowich, S. E. (1995). *Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing* (tech. rep.).

Wang, Z., Chlus, A., Geygan, R., Ye, Z., Zheng, T., Singh, A., Couture, J. J., Cavender-Bares, J., Kruger, E. L., & Townsend, P. A. (2020). Foliar functional traits from imaging spectroscopy across biomes in eastern North America. *New Phytologist*, *228*(2), 494–511. https://doi.org/10.1111/nph.16711

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1-3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Wright, B., & Stone, M. (1979, January). *Best test design*. MESA Press. https://research.acer.edu.au/measurement/1

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z., Li, H., & Yang, H. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01111

Zeileis, A., & Grothendieck, G. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, *14*(6), 1–27. https://doi.org/10.18637/jss.v014.i06

Zhu, P., Tan, Y., Zhang, L., Wang, Y., Mei, J., Liu, H., & Wu, M. (2020). Deep Learning for Multilabel Remote Sensing Image Annotation with Dual-Level Semantic Concepts. *IEEE Transactions*

on Geoscience and Remote Sensing, *58*(6), 4047–4060. https://doi.org/10.1109/TGRS.2019.2960466

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017, December). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. https://doi.org/10.1109/MGRS.2017.2762307

A Appendix - Additional Figures and Tables

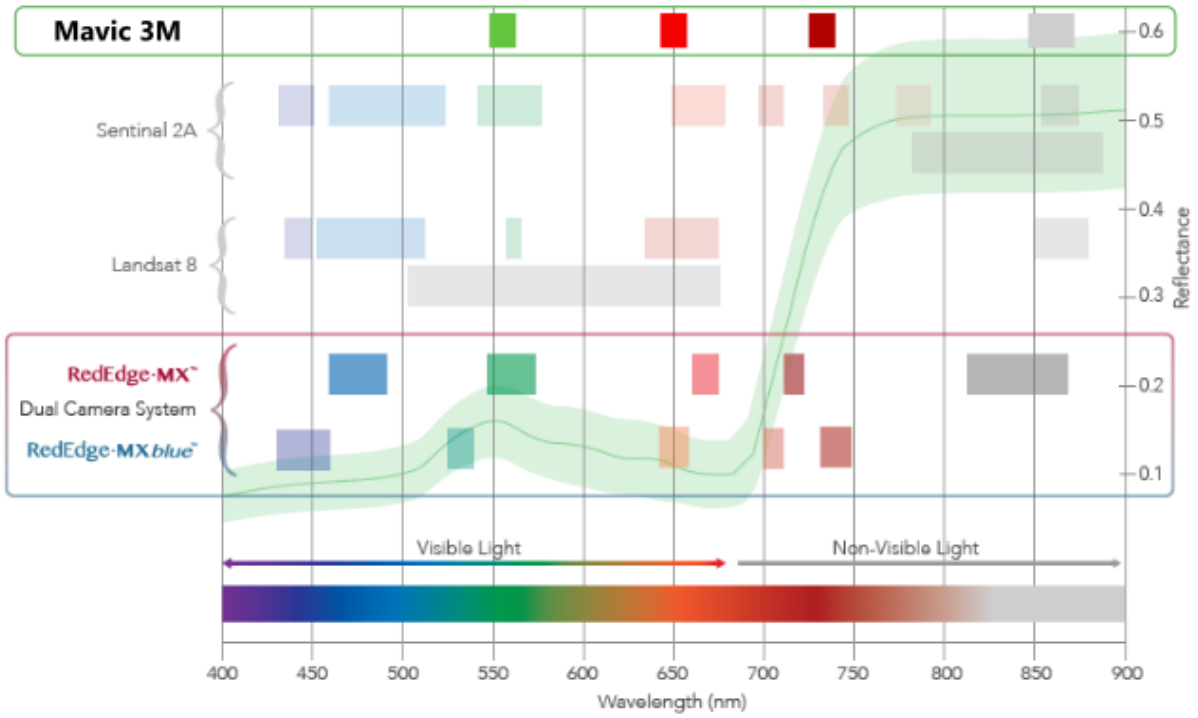
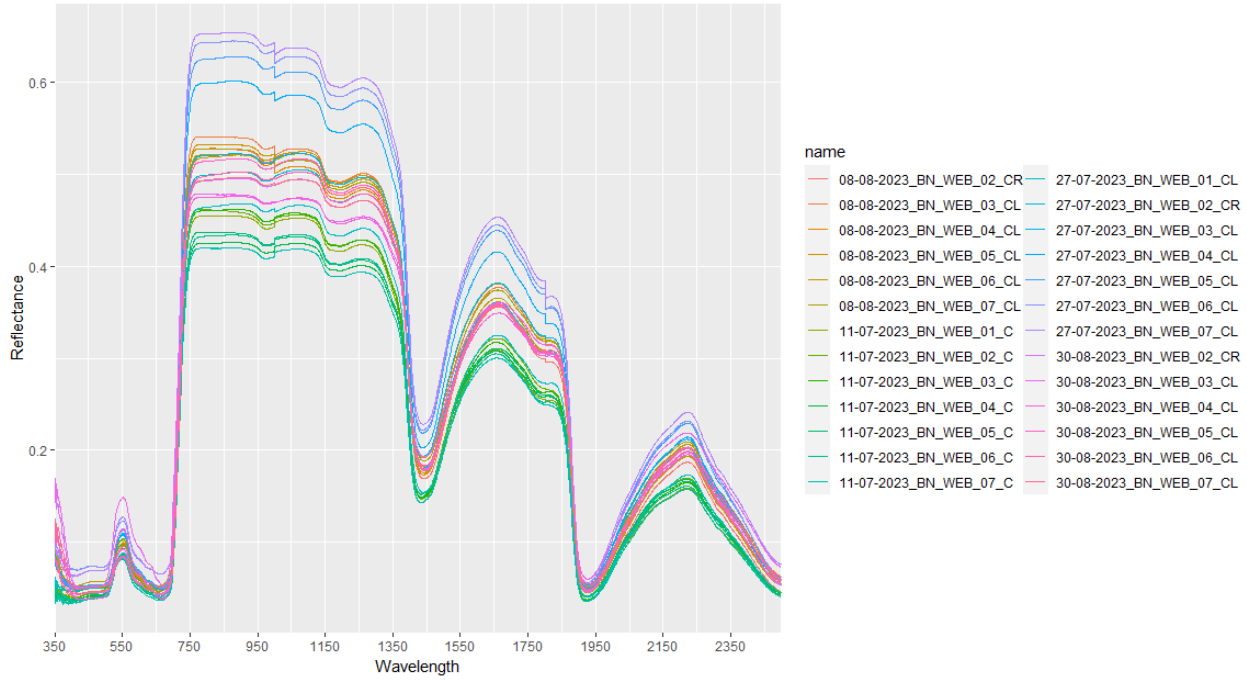


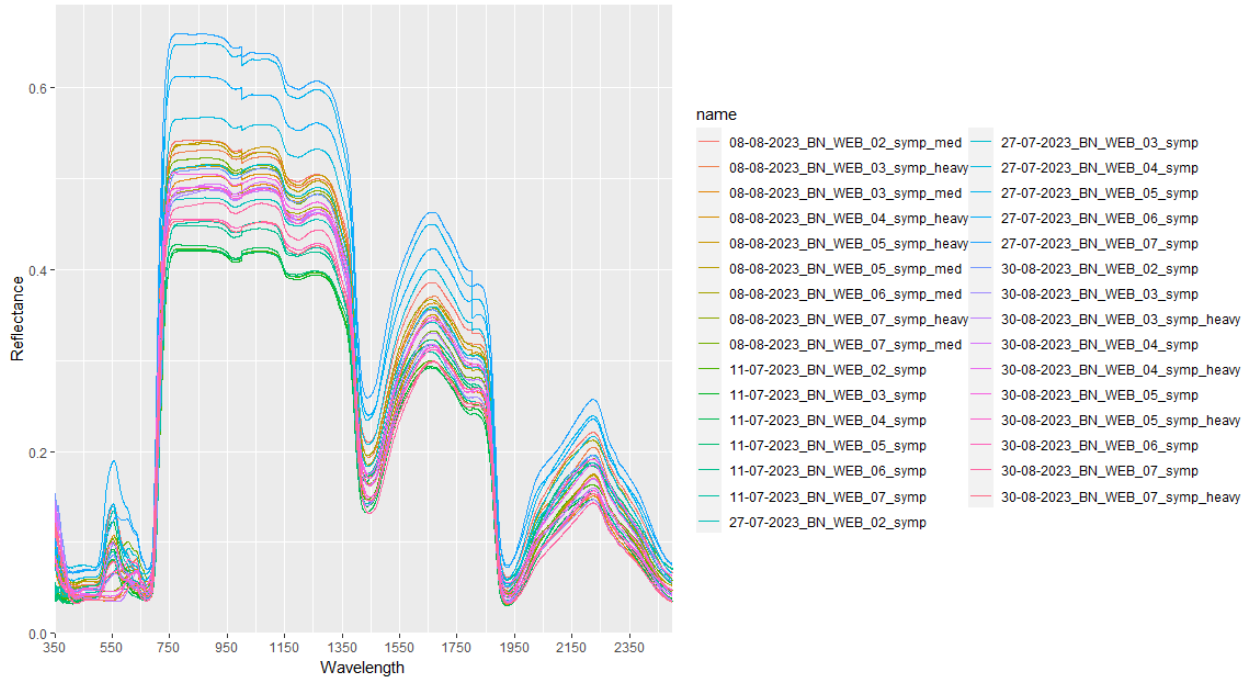
FIGURE A.1: Visual representation of RPAS sensor wavelength coverage, modified from 'MicaSense RedEdge-MX Dual - Professional Drones | DJI Enterprise', 2018.

Rank	Wavelength [nm]	Rank	Wavelength [nm]
1	712	26	410
2	2139	27	2410
3	2007	28	715
4	2138	29	1594
5	2008	30	2144
6	2140	31	2006
7	2010	32	650
8	2141	33	2005
9	2011	34	2145
10	648	35	709
11	2012	36	517
12	2142	37	1593
13	2411	38	425
14	2143	39	2429
15	711	40	708
16	649	41	2146
17	409	42	2004
18	710	43	647
19	520	44	2009
20	713	45	716
21	1596	46	2147
22	521	47	518
23	714	48	1591
24	1595	49	2003
25	522	50	646

TABLE A.1: Top 50 SVM-RFE wavelengths of full year dataset, ranked from most to least important.



(A) All Control spectra from the Weber vineyard.



(B) All symptomatic spectra from the Weber vineyard.

FIGURE A.2: Plots to show all used spectra of the Weber vineyard.

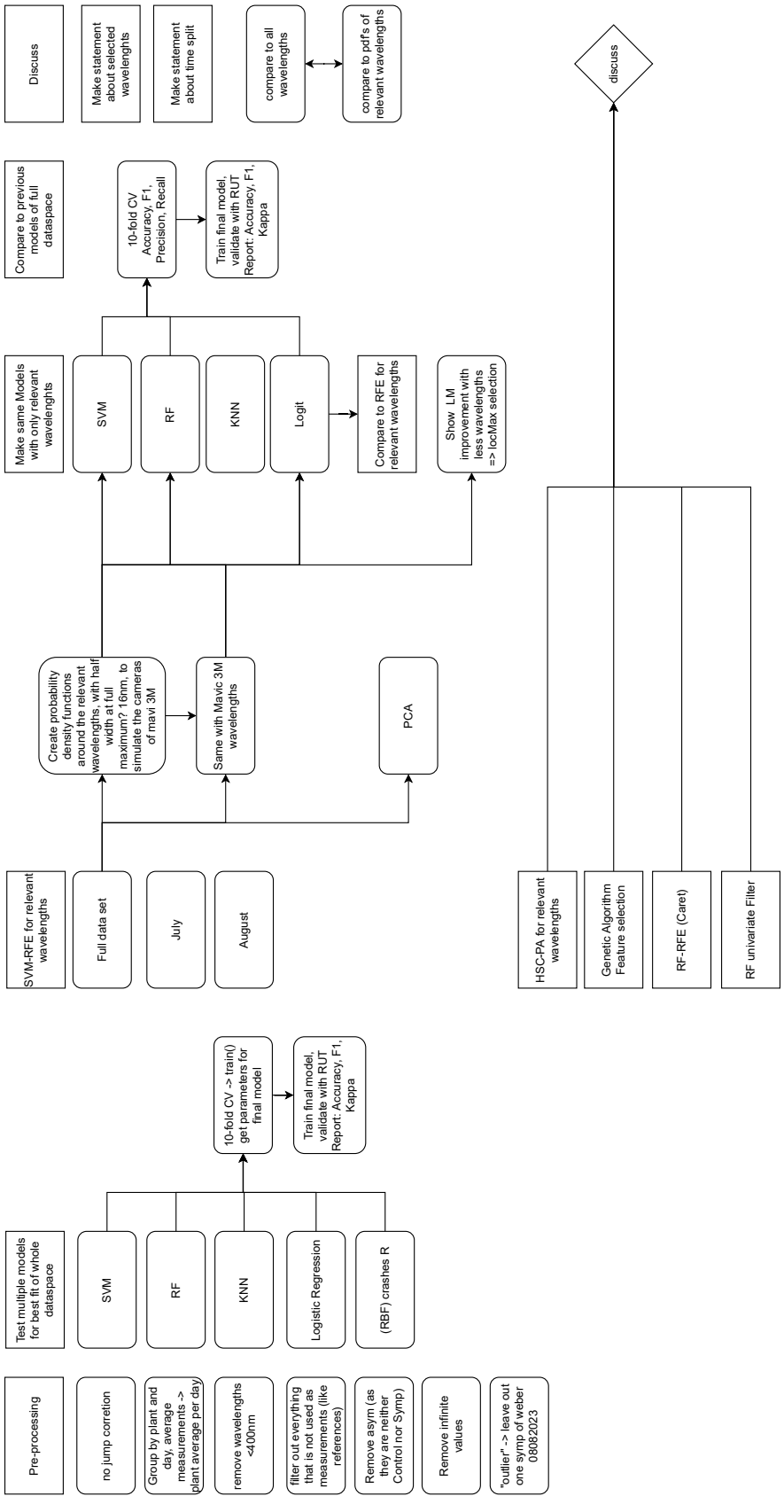


FIGURE A.3: Overview of the programming workflow of this thesis (draft).

Acknowledgements

It was a pleasure working with so many people during the process of my thesis and being able to learn the applied side of remote sensing in the field. The biggest thank you goes to my supervisor Meredith C. Schuman, who took me onboard for a master thesis on very short notice, even though she is extremely busy. Thanks for showing me how to do fieldwork and taking the time to explain everything, from RPAS and RTK to field spectrometry and everything around it. Also, a huge thanks to all participating members of the Smartgrape project: without them it would have not been possible to do all measurements; and thanks to lake Zurich for cooling me down after many long field days. Thanks to the Spatial Genetics group who were always open for questions and helped me wherever possible, especially Cheng Li who provided her code for the HSC-PA with an explanation so I could implement it. A penultimate thanks to my friends Jonathan and Anna who read the thesis and gave me constructive feedback and to Isabelle and Jasmin who helped me with the overleaf template. And finally, a huge thanks to my girlfriend Alina, without whom I might have struggled a lot more with staying focused on writing and who helped me keep on track when I started to veer off.

Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Date:
August 29, 2024

Signature
Jeremiah Huggel