



Grapevine red blotch virus detection in the vineyard: Leveraging machine learning with VIS/NIR hyperspectral images for asymptomatic and symptomatic vines[☆]



E. Laroche-Pinel^a, K. Singh^a, M. Flasco^b, M.L. Cooper^c, M. Fuchs^d, L. Brillante^{a,*}

^a Department of Viticulture & Enology, California State University Fresno, Fresno, CA, USA

^b Department of Plant Pathology & Crop Physiology, Louisiana State University Agricultural Center, Baton Rouge, LA, USA

^c University of California, Agriculture & Natural Resources 1710 Soscol Ave, Suite 4, Napa, CA, USA

^d Plant Pathology and Plant-Microbe Biology, Cornell University, Geneva, NY, USA

ARTICLE INFO

Keywords:

Precision viticulture
Non-destructive disease detection
Phenomics
Spectral feature selection
Machine vision

ABSTRACT

A decade after the discovery of grapevine red blotch virus (GRBV), there is ample evidence of its detrimental impacts on grapevine physiology, grape composition, and wine production. To mitigate the spread of GRBV in vineyards, roguing is recommended as a disease management response. The imperative to identify and remove diseased vines justifies the development of autonomous scouting. In this study, nearly 700 ground-based hyperspectral images, encompassing both symptomatic and asymptomatic vine canopies, were collected in a Cabernet Franc vineyard during two growing seasons, capturing pre- and post-veraison vine development stages. Spanning 230 bands from visible (VIS) to near-infrared (NIR) domains (510 to 900 nm with 1.7 nm width), canopy spectral signals were isolated from the background through semantic segmentation using U-Net. Simultaneously, the GRBV status of each vine was established in the laboratory through polymerase chain reaction. These two intertwined datasets were used for training various machine learning algorithms and their ensembles. In addition, strategies to reduce dataset size through spectral binning and testing three different feature selection methods (Recursive Feature Elimination, Univariate Feature Selection, and taking into consideration autocorrelation) were explored. Our findings revealed that hyperspectral imagery identified GRBV-infected vines with an accuracy of 75.7 % around harvest, coinciding with the peak of disease symptom expression, utilizing only 19 bands with a 16 nm bin width. Prior to veraison when most vines are asymptomatic, an accuracy of 74.2 % was achieved, employing 5 bands with a 16 nm bin width. This study substantiates the utility of hyperspectral images in the identification of GRBV-infected vines, offering a robust foundation for the development of a streamlined sensing system that holds great promise for the grape and wine industry in effectively scouting vineyards for GRBV.

1. Introduction

Vineyards are facing unprecedented challenges due to an increased occurrence of diseases in relation to climate change, globalization, and the introduction of new pathogens into different regions, changing viticultural practices, pesticide resistance, or restrictions in pesticide usage (Pugliese et al., 2011; Wilcox et al., 2015). Red blotch disease caused by grapevine red blotch virus (GRBV) affects vineyards by delaying fruit ripening and decreasing the accumulation of sugars and phenolic compounds (Ricketts et al., 2017; Rumbaugh et al. 2021).

Infected black berried grapevine cultivars exhibit characteristic red blotches on leaves, while chlorosis and necrosis are observed on leaves of infected white berried grapevine cultivars. Leaves show red blotches starting in the middle of the blade or at edges mostly around veraison (August) but sometimes at pre-veraison (late June to July), and blotches coalesce as the season progresses. These foliar disease symptoms are prominent in September–November in North America (Rumbaugh et al. 2021; Sudarshana et al. 2015). Disease symptoms are largely restricted to the basal leaves early on, with leaves in the middle of the shoots becoming symptomatic as the season progresses. The disease was first

[☆] This article is part of a special issue entitled: 'Specialty Crop Automation' published in Computers and Electronics in Agriculture.

* Corresponding author.

E-mail address: lucabrilante@csufresno.edu (L. Brillante).

identified in 2008 in California vineyards and has since been found in grape-growing regions worldwide (Sudarshana et al., 2015). If no action is taken to reduce the spread of GRBV, the economic impact of red blotch disease can reach up to \$68,548 per hectare over the 25-year lifespan of a vineyard (Ricketts et al., 2017).

GRBV is transmitted by the three-cornered alfalfa hopper (Flasco et al., 2021; Flasco et al. 2023). Controlling the spread of the disease by removing infected vines, and replacing them with clean vines, a strategy known as “roguing”, is recommended as a disease management response (Ricketts et al. 2017). Relying solely on visual symptoms to diagnose infected plants is time-consuming and not always reliable. This approach is also impractical because symptoms often only appear later in the vegetative season when growers are preoccupied with harvesting, leaving little time for additional tasks such as scouting for diseased vines. Accurately identifying infected plants requires expertise to avoid misdiagnosis, as red blotch disease can easily be mistaken for other issues, such as nutritional deficiencies, other pathogens, and abiotic stresses (Sudarshana et al., 2015). In addition, symptom expression can be limited to only a few spurs, shoots, or leaves of an infected vine, while the rest of the canopy remains asymptomatic due to the uneven distribution of the virus in an infected grapevine (DeShields and KC 2023; Setiono et al. 2018). Molecular diagnostic techniques are considered the most reliable to evaluate GRBV infections in grapevines (DeShields and KC, 2023). The downside is that these assays are expensive and time-consuming, and their accuracy depends on solid sampling protocols and minimized human errors. Therefore, a more automated and efficient approach for quickly detecting and diagnosing red blotch disease would be desirable for vineyard managers.

Diseases can alter plant optical signatures due to changes in their biochemical and biophysical properties (Carter and Knapp, 2001). Infected vines can, therefore, be identified using spectroscopy tools such as remote sensing imaging, which can be used at a large scale to screen entire vineyards. Hyperspectral imagery provides individual wavelength information across a wide range of the electromagnetic spectrum and can be particularly effective in monitoring grapevine vegetation. As a result, there has been a surge in research in this field (Terentev et al., 2022) with many studies using hyperspectral images to detect diseases in vineyards (Vanegas et al., 2018; Junges et al., 2018; Bendel et al., 2020a). To our knowledge, only two studies have so far been conducted using hyperspectral imagery to detect GRBV on leaf images. The first differentiated infected and non-infected leaves to perform a binary classification using a Support Vector Machine (SVM) classifier (Mehrubeoglu et al., 2016). The second documented up to 87 % overall binary accuracy using machine learning models with hyperspectral images taken in the visible domain under laboratory conditions (Sawyer et al., 2023). These two studies were promising but only used images of leaves rather than of the entire canopy under vineyard conditions.

Hyperspectral technologies are extremely effective for grapevine vegetation analysis, including disease identification (Thomas et al., 2018). They usually perform better than multispectral data (Marshall and Thenkabail 2015; Sluiter and Pebesma 2010; Sun et al., 2017). However, due to higher costs, larger storage requirements, and increased technical complexity in terms of time and computational resources for analysis, it may be challenging for grape growers to utilize these technologies. Identifying the critical spectral bands to achieve accurate GRBV identification, both in terms of location in the light spectra and wavelength width in nanometers, would simplify the system, opening the path to simplified and cheaper sensors with a reduced number of bands that could be used directly by grape growers.

The first objective of the study was to investigate the impact of GRBV infection on the hyperspectral signature of vine canopies in vineyards. We summarize results acquired on images with 230 bands in the 510 nm (visible) to 900 nm (near infrared) range from August to October in 2020 and 2021. Nearly 700 vines were imaged and sampled for GRBV diagnostic PCR in the laboratory. The second objective was to explore different algorithms to predict infected vines. Hyperspectral images

were processed and used as predictor of the infection status using three different machine-learning methods and their ensembles. The last objective was to reduce the number of wavelengths used by the models for the development of simpler solutions that could pave the road to commercial applications. The best models were interpreted for understanding the relative contribution of the most relevant wavelengths in different regions of the light spectrum. To better understand hyperspectral sensing capabilities, results were also compared with expert-based assessments of the infection status in the vineyard or on RGB images.

2. Materials & methods

2.1. Summary of experimental approach

The experimental approach is summarized in Fig. 1. A single vine on one given date constituted the basic experimental unit of the trial. Vines were individually sampled for molecular analysis and imaged with a hyperspectral camera. We selected a random and variable number of vines on five different dates, across two growing seasons and from pre-veraison (no-symptomatic expression of the virus) to harvest (maximum symptomatic expression of the virus). A total of 671 vines were imaged and tested for GRBV by PCR using petioles to distinguish between healthy and infected samples. The next step consisted of imaging the grapevine canopy with a hyperspectral camera mounted on a tripod in vineyard conditions. Reflectance standards (75 %, 44 %, and 2 %) were acquired every hour during the imaging time. Images were then pre-processed to segment the canopy from the background, and data were transformed into reflectance values using the reflectance standards. Different machine-learning models were applied to the segmented images to classify the infection status of the select vines and compare their diagnostic predictions with the PCR results. Experts were also involved in this study to predict the infection status of the vines by visually assessing both RGB images and the grapevine canopy in the vineyard in October 2020 and 2021. Expert predictions were also compared to PCR results. Both computer and expert predictions were evaluated using accuracy metrics and confusion matrices (Fig. 1).

2.2. Study vineyard site

The study was conducted in a ~3.5 ha commercial and mature vineyard in Napa Valley, California. This vineyard was planted in 2008 with Cabernet Franc grafted onto the rootstock 101–14 Millardet et de Grasset. Vines were planted 1.2 m apart, trained in a vertical shoot positioning system, and spur pruned in a bilateral cordon.

2.3. GRBV infection assessment and disease expression

GRBV symptoms were assessed in October of 2020 and 2021. Four petioles were collected per vine in September and October 2020, and in August, September, and October 2021. Petioles sampling and image acquisition were conducted simultaneously within a couple of days, which for GRBV does not influence the spectral signature or the infection status of the sampled plants. Petioles were selected close to the trunk, two per cordon. Viral infection was assessed by end-point multiplex PCR (Krenz et al., 2014) using the four petioles that were sliced into small pieces with sterile razor blades for nucleic acid isolation using the MagMAX 96 AI/ND Isolation kit (Thermo Fisher Scientific) on a King Fisher instrument, as previously described (Flasco et al., 2023; Sawyer et al., 2023).

2.4. Hyperspectral images

2.4.1. Image acquisition

A hyperspectral camera (Senop HSC-2, Senop Oy Optronics, Finland) was mounted on a tripod at a height of 1 m above the ground and a

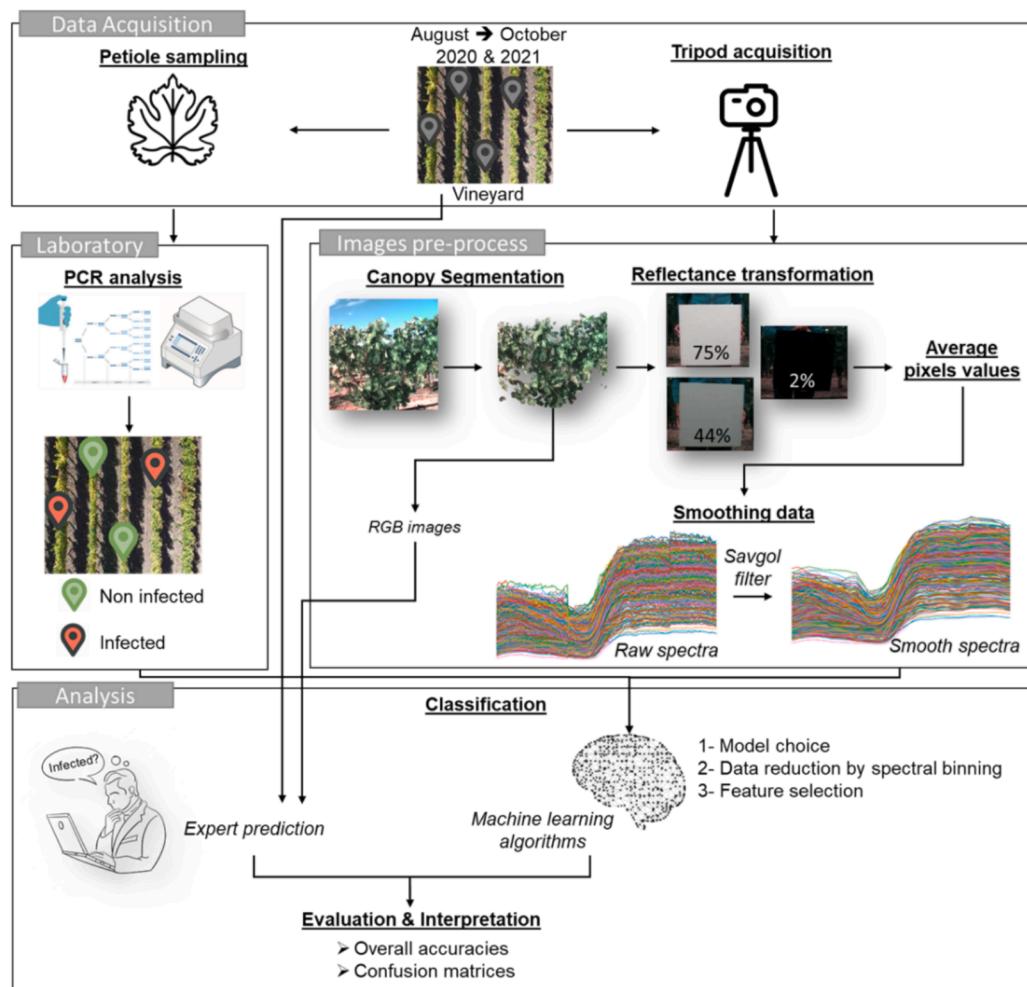


Fig. 1. Workflow of the analyses carried out in this study. The data acquisition was done in 2020 and 2021 from early August (pre-veraison) to October (post-veraison). Leaves and petioles were sampled from each vine for PCR testing of GRBV. The hyperspectral acquisition was done with the camera on a tripod in the vineyard. Images were pre-processed to segment the canopy signal. The classification was done using machine learning models to predict PCR results. Experts predicted the GRBV infectious status using RGB images and observing the vine canopy for visual symptoms in the vineyard. Models and expert predictions were evaluated using accuracies and confusion matrices.

distance of ~ 1 m from each vine. The camera was programmed to capture 1024×1024 -pixel images from 510 nm to 900 nm in 230 bands with a spatial resolution of ~ 7 mm per pixel. One vine was captured at a time from each hyperspectral image. The images were taken around solar noon, the same day as the leaves were sampled. Three standards (75 %, 44 %, and 2 % reflectance) were imaged every hour during the field measurements to convert the radiance values into reflectance. Sampled and imaged vines were distributed throughout the vineyard block, and no more than ten vines sampled within the same row. Vines were selected randomly based on previous knowledge of the vineyard (infection maps made in previous years), striving to maintain a balanced dataset between infected and non-infected vines. Each sampled vine was identified, tagged, and precisely located with a centimeter-level GNSS. A GNSS was also installed on the tripod to link each image with the corresponding PCR result.

2.4.2. Image pre-processing

The first and the last spectral bands were removed from the dataset because of noise in the signal. The final dataset was composed of 228 spectral bands. The first step of pre-processing the images was to extract the canopy signal and remove the background (soil, sky, shadow). Two hundred fifty images were segmented by manual labeling using PerClass Mira software (PerClass BV, Delft, Netherland). Then a U-Net model was trained on this subset, and the segmentation was applied to the whole

dataset with an accuracy of ~ 90 %. Imagery acquired in radiance was converted into reflectance through linear regressions using the standards whose images were the closest in time (with a maximum time difference of ~ 30 min). The conversion was done using equation 1, where a (slope) and b (intercept) were retrieved from the linear regression between the standard's radiance and their corresponding reflectance ($x_a = 0.012$, $\sigma_a = 0.014$ $x_b = 0.014$, $\sigma_b = 0.047$, $x_{R2} = 0.965$, $\sigma_{R2} = 0.012$, where x is the average and σ the standard deviation for all calibrations). Then the reflectance values were averaged per image, and a Savitzky-Golay filter was applied to smooth the spectra and reduce the noise (Savitzky & Golay, 1964). Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) pre-processing were also tested but did not improve accuracy and were not reported in this article.

$$\text{Reflectance} = a \cdot \text{Radiance} + b \quad (1)$$

2.5. Infection classification

2.5.1. Datasets for analysis

The data acquired were re-organized for 1) assessing the ability of hyperspectral imaging to identify GRBV-infected vine at different stages of disease expression, and 2) achieving dimensionality reduction by removing redundancy and irrelevant predictors, identifying the most

relevant spectral regions to discriminate infection status, and increasing the operational efficiency of the sensing method. To assess the performance of hyperspectral images at pre-veraison (no visible disease symptoms) and post-veraison (visible disease symptoms), the imagery dataset was divided into six sub-datasets (Table 1). The first sub-dataset considered all the data. The second sub-dataset included only images at pre-veraison when disease symptoms are not visible. The third sub-dataset considered all the post-veraison images when symptoms start to become visible. Finally, the fourth and fifth sub-datasets divided the imagery acquired post-veraison, when disease symptoms are visible. A binary classification was performed for each dataset to identify diseased or non-diseased vines in images.

A three-category classification was also performed with the 2021 August and October dataset to differentiate 1) non-infected, 2) infected vines with no visible symptoms (images acquired at pre-veraison), and 3) infected vines with the highest chance of visible symptoms (images acquired monthly from post-veraison to harvest). The number of samples for the October and the non-infected categories was reduced to balance each category and match an even 1/3 distribution rule.

Dimensionality reduction was achieved through multi-model stacking, feature selection (as described in section 2.5.3), and feature engineering on the spectral signal. Feature engineering held three different datasets, where the number of spectral bands was aggregated by averaging the reflectance values every 5, 10, or 20 spectral bands, or 8, 16, and 32 nm (called binned datasets herein) (Table 2).

2.5.2. Expert classification

The disease status of individual vines was determined by analysis of RGB segmented images by four experts without previous knowledge of the origin of the samples and virus status (Fig. 2). Experts also assessed the disease status in the vineyard by visual monitoring of typical red botch symptoms in October 2020 and 2021. Symptom data were used to produce a prediction map. Visual assessments were compared with PCR results of the October 2020 and 2021 petiole samples.

2.5.3. Machine learning algorithms

Three machine learning algorithms and their combinations were tested for binary classification (infected vs. non-infected) on the phenological datasets. The first one was Random Forest (RF), a common method for classification based on decision trees (Breiman, 2001; Pal, 2005; Belgiu and Drăguț 2016). The second algorithm was Partial Least Square Discriminant Analysis (PLS-DA) which is commonly used for spectral data (Barker and Rayens, 2003). The last algorithm was Support Vector Machine (SVM) that is well suited for binary classification (Boser et al., 1992).

A grid search was used to determine the optimized parameters to use for the models. For Random Forest, the optimized maximum level of each tree was 5 (2, 5, 7, 10 tested), the optimized number of trees was 100 (100, 200, 300, 400 tested) and the optimized minimum number of samples was 2 (2, 5, 10 tested). The maximum depth values cover shallow trees (2) to moderately deep trees (10), balancing underfitting and overfitting risks. The number of trees tested starts at 100 and increases to 400, observing validation results. The minimum number of samples test different pruning levels from minimal (2) to more conservative (10). For the PLS-DA, the optimized number of components was 6

Table 2

Bands width in nanometer (nm) for each dataset used in this study.

	Raw	Bin 5	Bin 10	Bin 20
Bands width (nm)	1.6	8	16	32
Total # of bands (predictor)/image	228	46	23	12

to 8 depending on the dataset (2, 5, 10, 15, 20 tested). For the SVM, the optimized kernel was polynomial (linear, RBF, and polynomial tested) with a degree of 4 (1, 2, 3, and 4 tested).

Multi-model ensembles through simple averaging (simple arithmetic mean of the predictions) were tested for the best models. Condorcet 1785, demonstrates that when each voter's probability of being correct exceeds 0.5 and voters act independently, adding more voters increases the likelihood of a correct majority vote. Although initially formulated in political science without any knowledge of machine learning, this principle mirrors the effectiveness of ensemble models. Ensemble learning's success can be attributed to various factors, including statistical, computational, and representation learning, bias-variance decomposition, and strength-correlation (Ganaie et al., 2022; Dietterich, 2000).

Multi-model stacking by using the PLS-DA scores as predictors for RF or SVM were also evaluated. PLS-DA effectively reduces dimensionality while preserving relevant information, providing informative features for subsequent models. By leveraging diverse learning algorithms like RF or SVM, the ensemble can capture complementary aspects of the data, enhancing generalization and robustness. This approach not only mitigates the curse of dimensionality but also boosts predictive performance by aggregating predictions from multiple models (Pavlyshenko, 2018; Džeroski & Ženko, 2004).

The predictions were assessed using repeated-stratified 10-fold with 10 repetitions cross-validations to include a balanced number of samples from all the sampling dates. We used 10x repeated 10-fold independently from dataset size, as most of the models in this study are binary classifications. This cross-validation procedure offers more granular performance metrics for each fold, more stable and realistic performance estimates, and more reliable model selection in this scenario when compared to other approaches, such as leave-one-out-cross-validation, which suffers from high variance due to its binary (correct/incorrect) outcomes for each one-sample fold. For the PLS-DA and SVM models, raw predicted values >0.5 were considered infected and raw values ≤ 0.5 were considered non-infected. In multi-model ensembles, all models were assessed contemporarily on the same subsets to avoid data leakage, as none of the models had access to the whole population prior to assessment (thus k-fold split was performed before PLS-DA application, the first step in multi-model). The model performance was assessed using overall test accuracy, which is the ratio of the correctly predicted images to the whole pool of test images, standard error (SE) was also computed for the test set. Confusion matrices were computed to evaluate the accuracy of each infection category and examine the false positive and false negative rates. The best models with higher accuracies were investigated in terms of feature importance to highlight the most significant bands used by the algorithms. They were also used to classify the different binned datasets.

Table 1

Details of the six datasets used in this study.

	DATASET	SEP 2020	OCT 2020	AUG 2021	SEPT 2021	OCT 2021	# data
BINARY Classification	All data	X	X	X	X	X	671
	Pre-veraison – 2021			X			76
	Post-veraison – 2020 & 2021	X	X		X	X	595
	Post-veraison – 2020	X	X				387
	Post-veraison – 2021				X	X	208
MULTICLASS Classification	Pre- & post-veraison – 2021			X		X	226
# data		154	233	76	58	150	

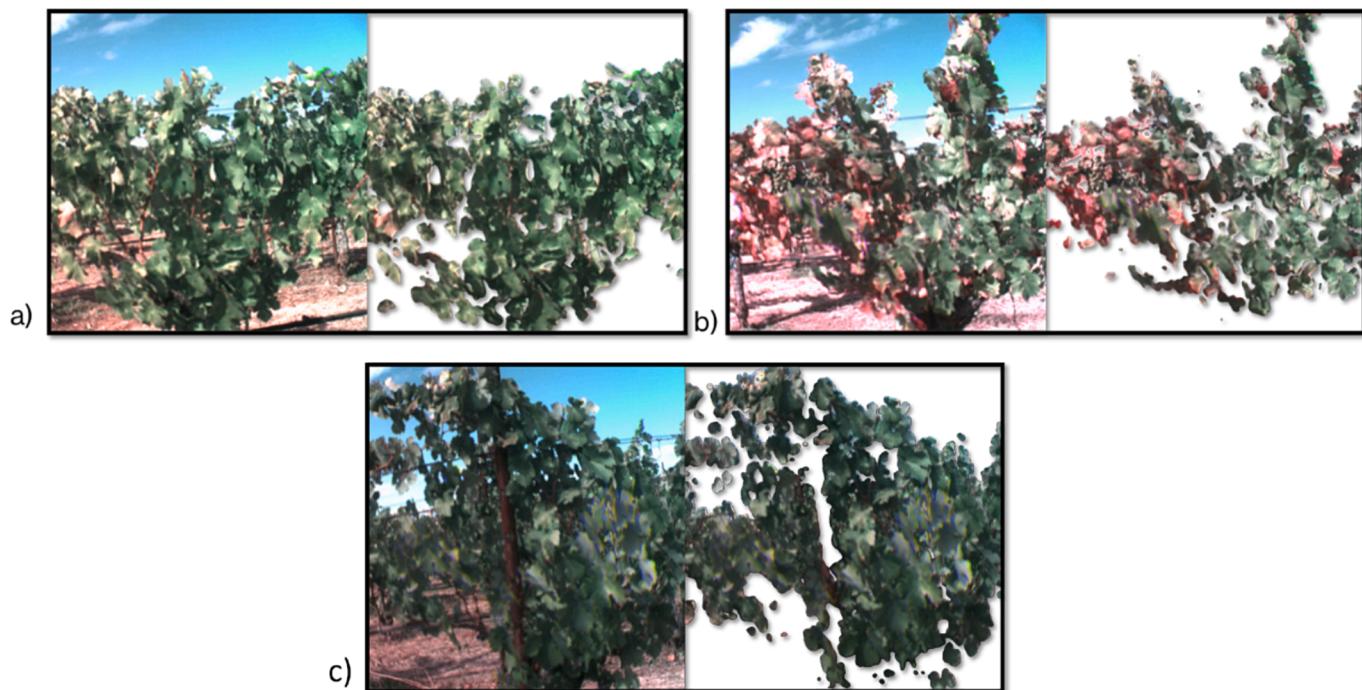


Fig. 2. Illustration of the segmentation process. A) Segmentation of a non-infected vine image, b) Segmentation of a GRBV-infected vine image with visible symptoms, c) Segmentation of an GRBV-infected vine image without visible symptoms. Images displayed in RGB colors using 525.3 nm, 555.7 nm, and 601.3 nm.

2.5.4. Feature selection methods

Feature selection is crucial when working with hyperspectral data due to their high dimensionality which can lead to computational inefficiency and overfitting when using classification models. By applying feature selection methods, it is possible to identify the bands that contribute the most to classification accuracy and only select those in model prediction to reduce computational complexity, improve model performances, and enhance interpretability (Jović et al., 2015). For this study, three feature selection methods were applied to both raw and binned spectra using the SVM model.

The first method to be tested was a Recursive Feature Elimination, RFE, (Guyon and Elisseeff, 2003). RFE is a technique to select the most relevant features from a dataset by recursively removing less important features. The RFE algorithm works by training a model on the entire feature set and then ranking the importance of each feature based on their contribution to the model's performance. The least important features are eliminated from the dataset, and the model is retrained on the remaining features. By selecting the most important features, RFE can reduce overfitting, improve model performance, and simplify the model by removing unnecessary predictors.

The second approach employed Univariate Feature Selection (UFS); a technique aimed at identifying the most significant features within a dataset. This method independently assesses the relationship between each feature and the target variable, selecting those with the strongest correlations. Iteratively, each feature undergoes evaluation, with the top-performing ones determined based on their highest correlation scores computed through Analysis of Variance (ANOVA). Classification is then computed using the top 5 % to 100 % of the best features with 5 % increments.

In UFS, the focus is on individual features and their contribution to the target variable, rather than in considering the relationships between features. This method is simple and straightforward, but it does not consider any interactions or dependencies between features. Therefore, for the third method, which takes into consideration autocorrelation (AC-SVM), a correlation matrix was computed to evaluate the correlation between each band, and features with a Pearson correlation coefficient (R) greater than 0.90 were removed from the dataset before

performing the classification.

3. Results

3.1. Data collection and PCR results

A total of 671 vine samples was tested for GRBV in PCR and imaged in this study. Of these samples, 56 % (376 of 671) were non-infected with GRBV, and 44 % (295 of 671) were GRBV-infected, as shown by PCR (Table 3). A well-balanced dataset was maintained on all sampling dates. A total of 387 and 208 samples was tested at post-veraison when disease symptoms were visible in 2020 and 2021, respectively. A total of 76 samples were tested at pre-veraison when disease symptoms were not visible in 2021.

3.2. Extraction of the canopy signal

All images were segmented using 525.3 nm, 555.7 nm, and 601.3 nm and displayed in RGB colors (Fig. 2).

3.3. Impact of the infection on the grapevine hyperspectral signatures

The spectral signatures for the whole dataset (Fig. 3a), the pre-veraison 2021 dataset (Fig. 3b), the entire post-veraison 2020–2021 dataset (Fig. 3c), the post-veraison datasets in September 2020 (Fig. 3d), October 2020 (Fig. 3e), September 2021 (Fig. 3f) and October 2021

Table 3

Percentage of non-infected and GRBV-infected vines according to PCR results at five petiole collection months in 2020 and 2021.

Sampling Date	Non- GRBV- infected	GRBV- Infected
Sep 20	56.5 %	43.5 %
Oct 20	55 %	45 %
Aug 21	52 %	48 %
Sep 21	38 %	62 %
Oct 21	51 %	49 %
Total	56 %	44 %

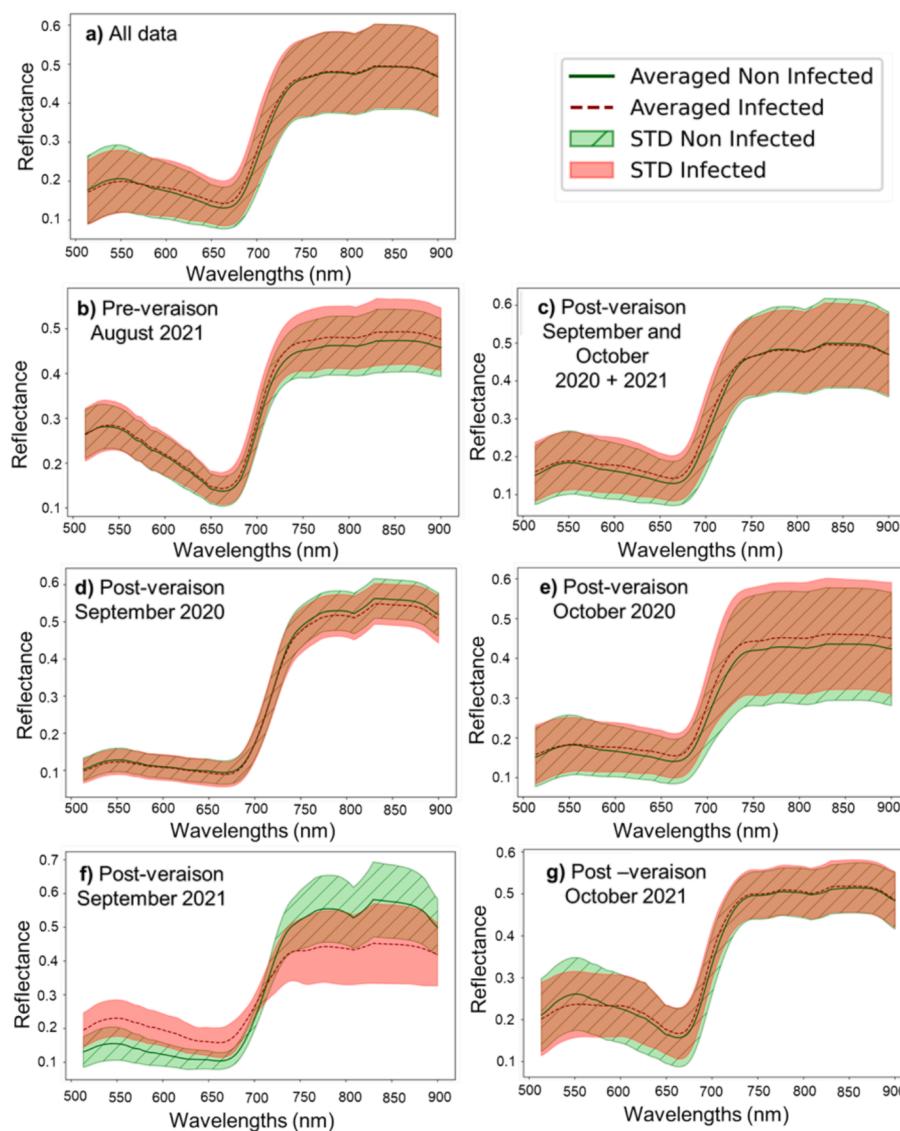


Fig. 3. Spectral signatures of the non-infected (green line) or the GRBV-infected (red dotted line) vines for the whole dataset (a), pre-veraison (b), post veraison for 2020 and 2021 (c), post-veraison for September 2020 (d) and October 2020 (e), and post-veraison for September 2021 (f) and October 2021 (g). The standard deviation (STD) is also represented in red for the GRBV-infected and in green with lines for the non-infected category.

(Fig. 3g) were analyzed. The standard deviation showed a large variability within each category (Fig. 3). The spectral signatures did not show any evident separation, except for September 2021, however, average trends demonstrated signal differences. A lower reflectance in the NIR for GRBV-infected vines at the post-veraison stage was observed in September 2020 (Fig. 3d) and 2021 (Fig. 3f). Additionally, a peak in the green domain (around 550 nm) for the non-infected category and a peak in the red domain (around 650 nm) for the GRBV-infected category was obtained for the whole dataset (Fig. 3a). A trough is observed around 800 nm especially for post-veraison datasets in September 2020 (Fig. 3d) and September 2021 (Fig. 3f), possibly linked to a change in water status, as previously described (Beget & Di Bella, 2007, Kutser et al., 2016). Anecdotally, 2021 was a drier season than 2020 in the study area (Williams et al., 2022), but assessment of plant water status was out of purpose in this research. As the signal varied from plant to plant, and the standard deviation was high, it was not possible to use a simple, heuristic method to classify infection.

3.4. Infection prediction by experts

3.4.1. Prediction using RGB images

Four experts analyzed the RGB segmented images and classified them as GRBV-infected or non-infected. Analysis was done to compare their predictions with the results of the molecular testing (PCR). The overall accuracy using the whole dataset was 75.8 %, including 53.9 % with the pre-veraison dataset and 78.6 % with the post-veraison dataset.

Confusion matrices were built for the three phenological datasets to determine the accuracy per infection category (Fig. 4). The non-infected category was always well predicted with an accuracy ranging from 97 % (all data, Fig. 4a) to 100 % (pre-veraison, Fig. 4b). However, for the infected category, the accuracy reached only 52 % for the whole dataset (Fig. 4a), 59 % for the post-veraison dataset (Fig. 4c) and 3 % for the pre-veraison dataset (Fig. 4b). Almost half of the images of infected vines (48 %) were predicted as non-infected (false negatives) using the whole dataset (Fig. 4a). This percentage decreased to 41 % using the post-veraison dataset (Fig. 4c). For the pre-veraison dataset (Fig. 4b), 97 % of infected vines were classified as non-infected.

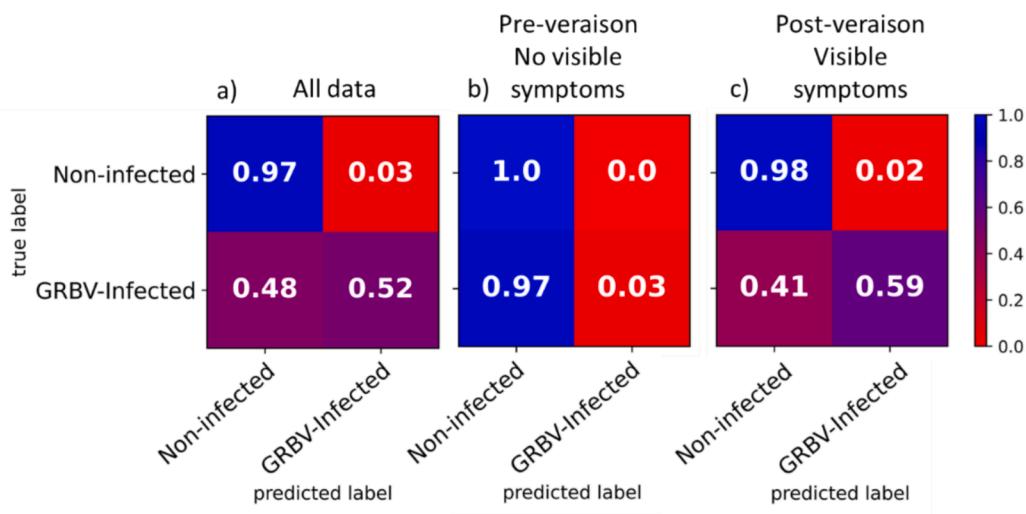


Fig. 4. Confusion matrices of expert prediction using RGB images for the whole dataset (a), the dataset with no visible symptoms (b) and the dataset with visible symptoms (c). The top left corner shows the percentage of non-infected vines that were well predicted as non-infected. The top right corner shows the percentage of non-infected vines that were wrongly predicted as infected (false positive). The bottom left corner shows the percentage of infected vines that were wrongly predicted as non-infected (false negative). The bottom right corner shows the percentage of infected vines that were well predicted as infected. Color is mapped to percentage points going from 0 (red) to 1 (blue) and does not necessarily indicate more accurate results.

3.4.2. Prediction based on disease symptoms in the vineyard at post-veraison

Diseased vines were mapped throughout the Cabernet Franc vineyard in October 2020 and 2021 at the peak of symptom expression. Expert predictions were compared with the results of the PCR testing of petioles sampled at the same time. An overall accuracy of 75.3 % was obtained in 2020 and 83.6 % in 2021. For 2020 the two experts were able to identify non-infected vines at 71 %, and GRBV-infected vines at 83 % (Fig. 5a). The false positive rate was 29 % and false negative rate was 17 % (Fig. 5a). In 2021 two experts reach an accuracy of 92 % for the non-infected vines but a lower accuracy of 76 % for the infected ones (Fig. 5b).

3.5. Infection prediction by machine learning models

3.5.1. Investigation of different machine learning models for binary prediction

The overall accuracy of the phenological datasets using the models by themselves, multi-model ensemble (simple arithmetic average of predictions), or multi-model stacking (using PLSDA scores as predictors for the second model) was calculated (Table 4). Multi-model ensembles are described in section 2.5.3. The best single model was the SVM model, with an overall accuracy of 71.4 % for the whole dataset and 75.2 % for the post-veraison dataset in 2020 and 2021. For the other dataset, PLS-DA performed slightly better than SVM. A slight improvement in accuracy was reached using the multi-model ensemble averaging PLS-DA and SVM predictions. The accuracy increased by 0.2 points for the whole dataset (69.5 %), while the accuracy of the post-veraison dataset

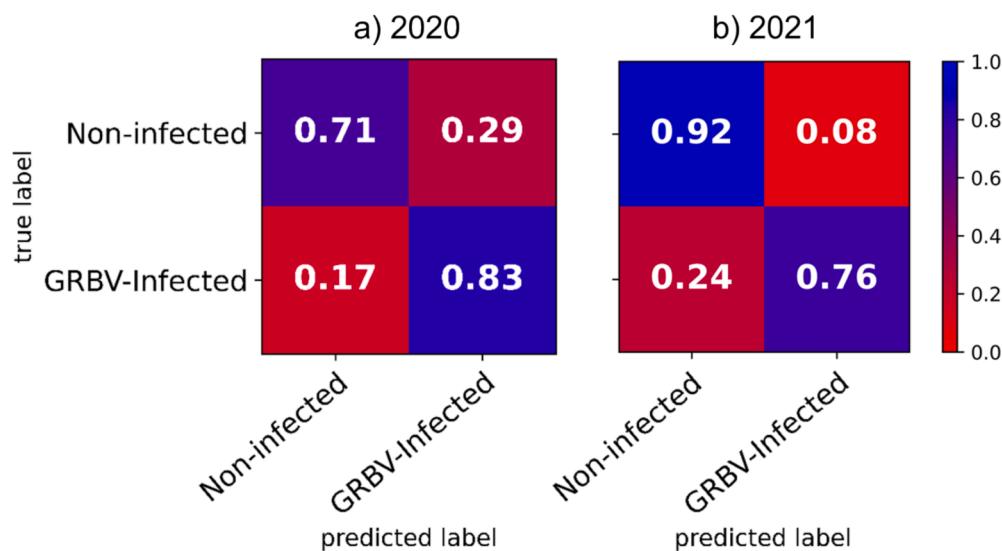


Fig. 5. Confusion matrix of expert predictions made in the field in October 2020 and 2021 compared to the PCR results obtained for the same date. The top left corner shows the percentage of non-infected vines that were well predicted as non-infected. The top right corner shows the percentage of non-infected vines that were wrongly predicted as infected (false positive). The bottom left corner shows the percentage of infected vines that were wrongly predicted as non-infected (false negative). The bottom right corner shows the percentage of infected vines that were well predicted as infected. Color is mapped to percentage points going from 0 (red) to 1 (blue) and does not necessarily indicate more accurate results.

Table 4

Accuracies of the different machine learning models tested in this study and their multi-model ensemble and stack for each dataset (standard error SE are given in parenthesis).

	All data	Pre-veraison No visible symptoms 2021	Post-veraison Visible symptoms 2020 + 2021	Post-veraison Visible symptoms 2020	Post-veraison Visible symptoms 2021
Model					
RF	62.5 % (1.4)	65.9 % (2.6)	62.7 % (1.6)	61.5 % (1.9)	67.0 % (2.5)
PLS-DA	68.2 % (1.1)	68.5 % (20)	71.8 % (1.2)	75.2 % (2.0)	75.8 % (2.4)
SVM	71.4 % (1.4)	57.9 % (4.0)	75.2 % (1.5)	72.6 % (2.3)	75.4 % (2.5)
Multi-Model ensemble					
<i>PLS-DA_SVM</i>	71.6 % (1.9)	67.3 % (2.9)	75.2 % (1.9)	73.6 % (2.1)	77.7 % (2.0)
<i>PLS-DA_SVM_RF</i>	70.5 % (2.5)	65.1 % (3.4)	73.9 % (1.3)	71.2 % (1.8)	76.1 % (2.5)
Multi-Model stacking					
<i>PLS-DA_RF</i>	54.8 % (2.7)	64.5 % (2.4)	56.9 % (2.1)	60.5 % (2.1)	50.2 % (2.4)
<i>PLS-DA_SVM</i>	52.3 % (2.9)	63.4 % (3.6)	54.2 % (2.7)	60.6 % (2.1)	49.3 % (3.0)

remained identical (Table 4). Using the PLS-DA in a multi-model stacking approach decreased the accuracies. In 2021, the accuracies for the post-veraison dataset consistently outperformed those in 2020, with the highest overall accuracy achieving 77.7 % by combining the PLS-DA and SVM methods (Table 4).

The SVM permutation importance (Fig. 6a) and the PLS-DA coefficients attribute (Fig. 6b) to each spectral band used by the multi-model ensemble SVM-PLS-DA were mapped on the post-veraison dataset. The most significant spectral bands were different according to the model, except for the region around 600 nm (orange domain), which was important for both models. For the SVM model, the other important spectral bands were located around 525 nm (green), 540 nm (green), 710 nm (red-edge), and 850 nm (NIR). For the PLS-DA model, the spectral bands are located around 550 nm (green), 575 nm (orange), 740 nm (red-edge), 780 nm (NIR), and 810 nm (NIR).

According to these results, not all bands were important in predicting the infection status. Therefore, reducing the redundancy of information could be useful to reduce the noise and improve accuracy.

3.5.2. Reducing the number of features with spectral binning

The PLS-DA and SVM algorithms and their highest accuracy ensemble (PLS-DA SVM ensemble) applied to the different binning datasets were analyzed (Table 5). The best overall accuracy for the whole dataset (71.6 %) was obtained using the binning by 5 (46 bands) with the SVM model. Considering the pre-veraison dataset, the best overall accuracy was slightly lower, with 68.5 % obtained using the whole bands with the PLS-DA model. The best overall accuracy (75.6 %)

was obtained with the binning by 10 (23 bands) and the PLS-DA and SVM ensemble on the post-veraison dataset.

3.5.3. Identifying the best bands using different feature selection methods

The overall classification accuracy using SVM_RFE according to the number of features selected out of the 46 (bin by 5) for the whole dataset across the 10-cross validation splits was mapped as an example (Fig. 7). The best previous accuracy found using 228 bands and the SVM model was 71.4 %. This accuracy was reached using 16 bands and even exceeded (73.5 %) when 24 features were selected. This number of bands was considered the best.

Using RFE to reduce the redundancy of information while also improving the signal-to-noise ratio through spectral binning improved the accuracy of predictions with SVM models (Table 6). Large improvements in accuracy through applying RFE were generally observed when starting with all the bands (no binning), resulting in an increase of 2.5 points when using all data and 2.2 points when using the pre-veraison dataset. For the post-veraison dataset, the largest improvement in accuracy was observed when combining RFE with spectral binning, with an increase of 1.9 points. The combination of spectral binning and RFE yielded better models with improved accuracy of 73.5 % when using all data (8 nm binning using 24 of the 46 available bands), and 75.7 % (16 nm binning using 19 of the 23 available bands) when using the post-veraison dataset. This was not validated for the pre-veraison dataset where PLS-DA combined with spectral binning outperformed SVM (Table 5).

The best accuracy for the pre-veraison dataset is achieved with UFS-

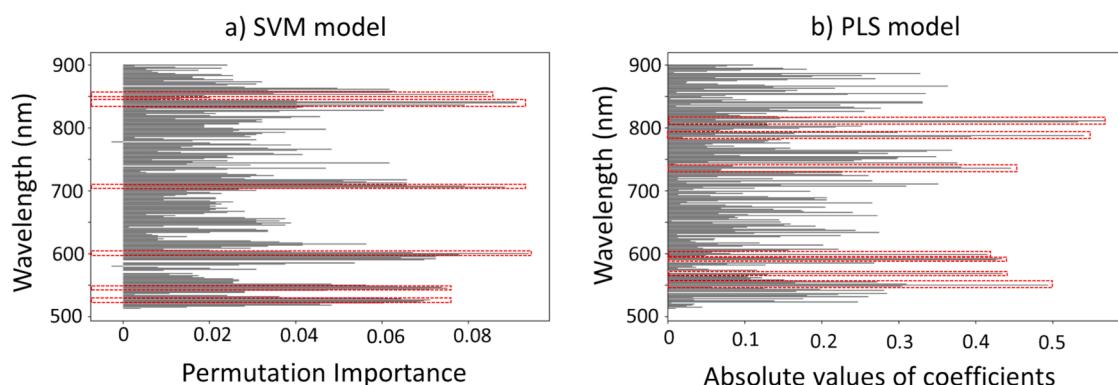


Fig. 6. SVM permutation importance (a) and PLS-DA coefficients (b) attribute to each band using post-veraison dataset (visible symptoms) to predict vine infection in VIS/NIR hyperspectral images. The most significant bands are highlighted in red dashed lines.

Table 5

Overall accuracies of the best models for the different spectral binning dataset (standard error SE in percentage point for the cross-validations are given in parenthesis).

Dataset	Band #	Bands width (nm)	All data	Pre-veraison No visible symptoms 2021	Post-veraison Visible symptoms 2020 + 2021
SVM model					
RAW	228	1.6	71.4 % (1.4)	57.9 % (4.0)	75.2 % (1.5)
BIN 5	46	8	71.8 % (1.6)	56.1 % (5.1)	74.4 % (1.8)
BIN 10	23	16	69.8 % (1.7)	62.9 % (3.9)	73.8 % (1.6)
BIN 20	12	32	70.9 % (1.6)	64.0 % (5.2)	71.2 % (1.7)
PLS-DA model					
RAW	228	1.6	68.2 % (1.1)	68.5 % (2.0)	71.8 % (1.2)
BIN 5	46	8	67.9 % (1.2)	57.1 % (3.1)	71.4 % (1.2)
BIN 10	23	16	66.3 % (1.5)	61.6 % (2.6)	70.5 % (1.3)
BIN 20	12	32	65.7 % (1.6)	65.6 % (2.0)	68.8 % (2.0)
PLS-DA SVM ensemble					
RAW	228	1.6	71.6 % (1.9)	67.3 % (2.9)	75.2 % (1.5)
BIN 5	46	8	60.8 % (1.2)	61.5 % (4.2)	61.2 % (1.6)
BIN 10	23	16	64.2 % (2.4)	62.2 % (2.8)	75.6 % (1.2)
BIN 20	12	32	64.5 % (2.8)	64.2 % (3.1)	70.7 % (1.8)

SVM attaining 74.2 % using only 5 bands with a 16 nm width (Table 6). Overall, this method increased accuracy by 5–10 percent points on the pre-veraison dataset; no improvement was observed with the other

datasets.

When removing autocorrelated bands (AC-SVM), accuracies only improved for the no-visible symptom dataset when using 228 bands, 46 or 23 bands. Specifically, utilizing only 10 bands out of the 228 yielded to a 58.9 % accuracy, surpassing the 57.9 % accuracy achieved with the original SVM employing all 228 bands. Similarly, employing 12 bands out of the 46 resulted in a 57.4 % accuracy, outperforming the 56.1 % accuracy with the original SVM utilizing all 46 bands (Table 6).

The best overall accuracy (75.7 %) was obtained using SVM-RFE with the post-veraison dataset with the 19 bands selected in the 16 nm binning dataset, including a category-based accuracy of 85 % for the non-infected vines, and 62 % for the GRBV-infected ones (Fig. 8).

The selected spectral bands for the pre-veraison dataset were in the green (from 530 nm to 570 nm), the red (from 600 nm to 650 nm), the red-edge (670 to 715 nm), and in two NIR regions (from 750 nm to 780 nm and around 875 nm). For the post-veraison dataset, the selected bands were mostly in the green (520–570 nm), red (600–650 nm) and red-edge domains (650–730 nm). These results revealed that asymptomatic and symptomatic GRBV-infected vines can be distinguished as the relevant spectral domains are different.

3.5.4. Evaluating a multiclass classification according to symptom expression

A three-class classification was evaluated by splitting the infected category into an asymptomatic sub-category and a symptomatic sub-category. The corresponding confusion matrix of the SVM model revealed relative accuracies for the three categories (Fig. 9). The overall accuracy for this model was 70.5 %, a value higher than the accuracy obtained for the SVM model in binary classification (section 3.4.1.). Both infected categories were well predicted at 73 % (symptomatic) and 77 % (asymptomatic), respectively, both levels being higher than those obtained with binary models. The non-infected category accuracy was lower at 63 %.

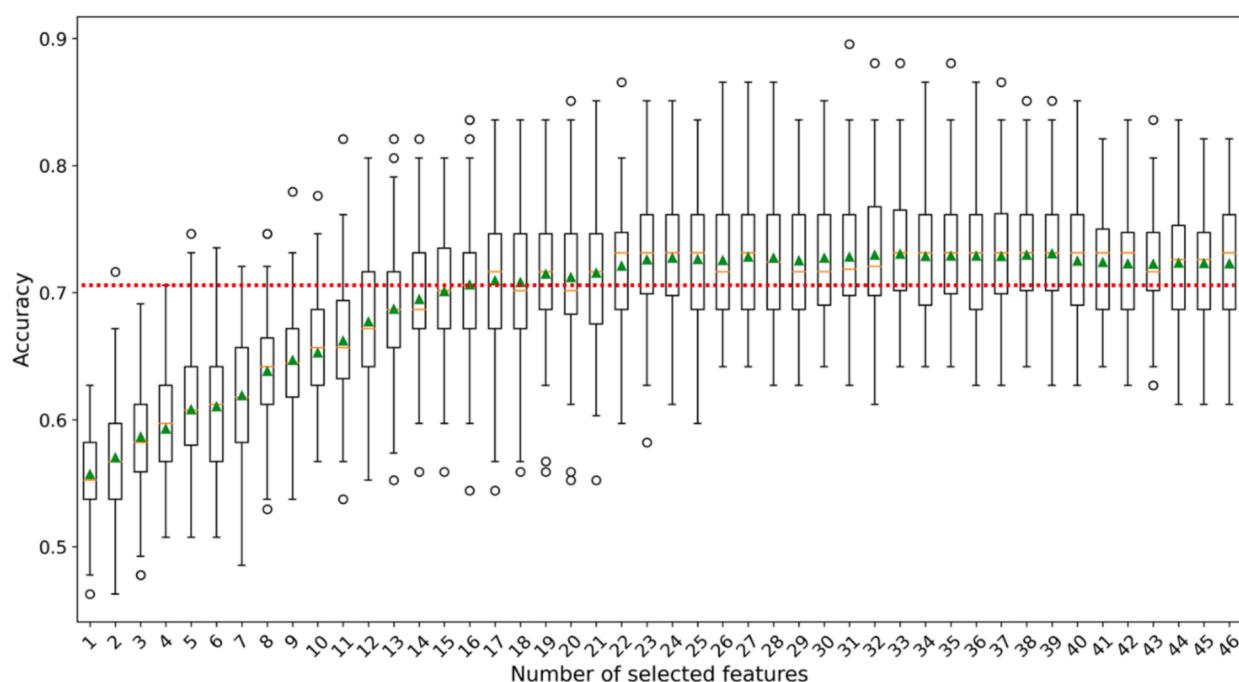


Fig. 7. Classification accuracy with RFE according to the number of selected features using the spectral binning by 5 and the whole dataset with SVM model. The red dot line represents the accuracy 71.4 % (the best founded previously using all the 228 bands with the SVM model). Green triangle represents the mean and orange line the median.

Table 6

Overall accuracies (in %) of the SVM model for the different dataset with RFE, UFS and auto-correlation selections compared to the original SVM model accuracies without feature selection. The best results are highlighted in bold for each dataset (standard error in percentage point for the cross-validations are given in parenthesis). The number of features selected is given for each model and each dataset.

	Raw-228 bands Accuracy (SE)	# Bands selected	Bin 5–46 bands Accuracy (SE)	# Bands selected	Bin 10–23 bands Accuracy (SE)	# Bands selected	Bin 20–12 bands Accuracy (SE)	# Bands selected
All dataset								
<i>SVM model</i>	71.4 % (1.4)	–	71.8 % (1.6)	–	69.8 % (1.7)	–	70.9 % (1.6)	–
<i>RFE-SVM</i>	73.9 % (1.5)	84	73.5 % (1.9)	24	70.1 % (2.5)	16	71.5 % (1.8)	10
<i>UFS-SVM</i>	62.4 % (1.1)	148	65.2 % (2.5)	32	56.8 % (3.2)	23	56.4 % (4.2)	9
<i>AC-SVM</i>	65.4 % (1.8)	35	58.7 % (3.0)	10	58.4 % (2.8)	6	56.1 % (2.4)	4
Pre veraison dataset								
<i>SVM model</i>	57.9 % (4.0)	–	56.1 % (5.1)	–	62.9 % (3.9)	–	64.0 % (5.2)	–
<i>RFE-SVM</i>	60.1 % (2.3)	102	58.1 % (3.4)	41	70.9 % (5.8)	6	64.7 % (7.3)	10
<i>UFS-SVM</i>	67.2 % (2.1)	135	70.2 % (1.3)	38	74.2 % (2.6)	5	71.8 % (5.6)	8
<i>AC-SVM</i>	58.9 % (2.9)	10	57.4 % (2.5)	12	65.1 % (3.5)	4	61.3 % (4.2)	4
Post veraison dataset								
<i>SVM model</i>	75.2 % (1.5)	–	74.4 % (1.8)	–	73.8 % (1.6)	–	71.2 % (1.7)	–
<i>RFE-SVM</i>	75.2 % (2.6)	56	75.6 % (3.2)	36	75.7 % (2.1)	19	72.2 % (1.4)	10
<i>UFS-SVM</i>	67.5 % (2.8)	182	64.7 % (1.7)	46	65.4 % (2.2)	20	65.5 % (1.5)	11
<i>AC-SVM</i>	59.3 % (1.7)	25	62.1 % (1.8)	10	59.1 % (1.4)	8	60.5 % (1.1)	4

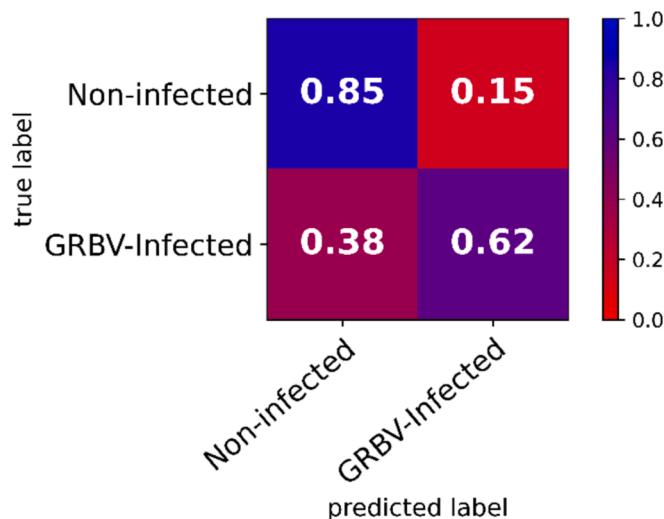


Fig. 8. Confusion matrix for the late season dataset using the binning by 10 with the 19 selected bands. The top left corner shows the percentage of non-infected vines that were well predicted as non-infected. The top right corner shows the percentage of non-infected vines that were wrongly predicted as infected (false positive). The bottom left corner shows the percentage of infected vines that were wrongly predicted as non-infected (false negative). The bottom right corner shows the percentage of infected vines that were well predicted as infected. Color is mapped to percentage points going from 0 (red) to 1 (blue) and does not necessarily indicate more accurate results.

3.6. Comparison of PLS-DA and SVM models predictions with no visible symptoms

The prediction values for the SVM and PLS-DA models at pre-veraison using the dataset with 23, 16 nm wide bands revealed good predictions for the non-infected category with values less or equal to 0.50 in 60 % of the cases for the PLS-DA (Fig. 10a) and 85 % for the SVM model (Fig. 10c). Approximately 40 % of the well-predicted images had values ranging from 0.43 to 0.50 for the SVM model, which is close to the classification threshold. For the prediction of infected vine images, 48 % are well predicted as infected with a prediction value superior to 0.50 with the PLS-DA model (Fig. 10b) and only 30 % with the SVM model (Fig. 10d). A total of 58 % of infected-vine images have predicted values ranging from 0.35 to 0.50 (classified as non-infected) with the SVM model.

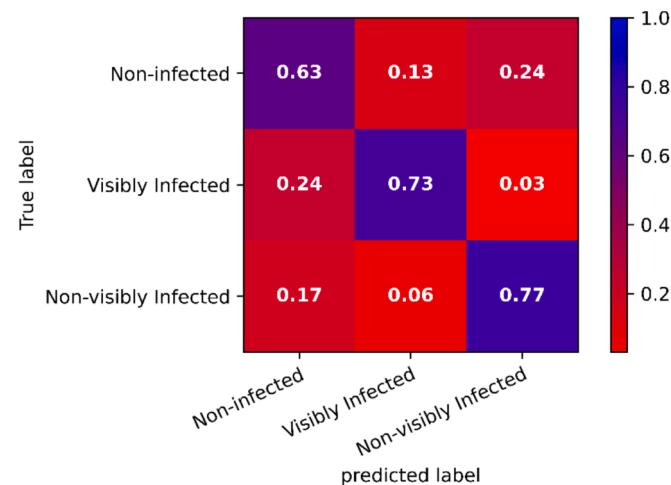


Fig. 9. Confusion matrix of the predictions for the three classes classification using data of 2021 (August for the early season and October for the late season) with a SVM model using the raw 228 spectral bands. The overall accuracy of this model is 70.5 %. The top left corner shows the percentage of non-infected vines that were well predicted as non-infected. The top right corner shows the percentage of non-infected vines that were wrongly predicted as non-visibly infected. The bottom left corner shows the percentage of non-visibly infected vines that were wrongly predicted as non-infected. The bottom right corner shows the percentage of non-visibly infected vines that were well predicted as non-visibly infected. Color is mapped to percentage points going from 0 (red) to 1 (blue) and does not necessarily indicate more accurate results.

4. Discussion

4.1. A first study with promising results

This study confirmed the potential of hyperspectral images in detecting plant diseases as already demonstrated for leaf stripe (Junges et al., 2018), leafroll (Bendel et al., 2020; MacDonald et al., 2016), Esca (Al-Saddik et al., 2018), or grapevine vein clearing virus (Nguyen et al., 2021). For vein clearing, the authors obtained better results with SVM than random forest. A preliminary study exploring the use of hyperspectral imaging in VIS-NIR domains for the identification of red blotch disease in images of detached leaves used a small dataset and did not detail the accuracy of the detection methods (Mehrubeoglu et al., 2016). A more detailed and comprehensive study using the VIS range in a mixed dataset with red blotch and leafroll diseases in detached leaves under controlled conditions obtained an accuracy of 57 % using the CNN

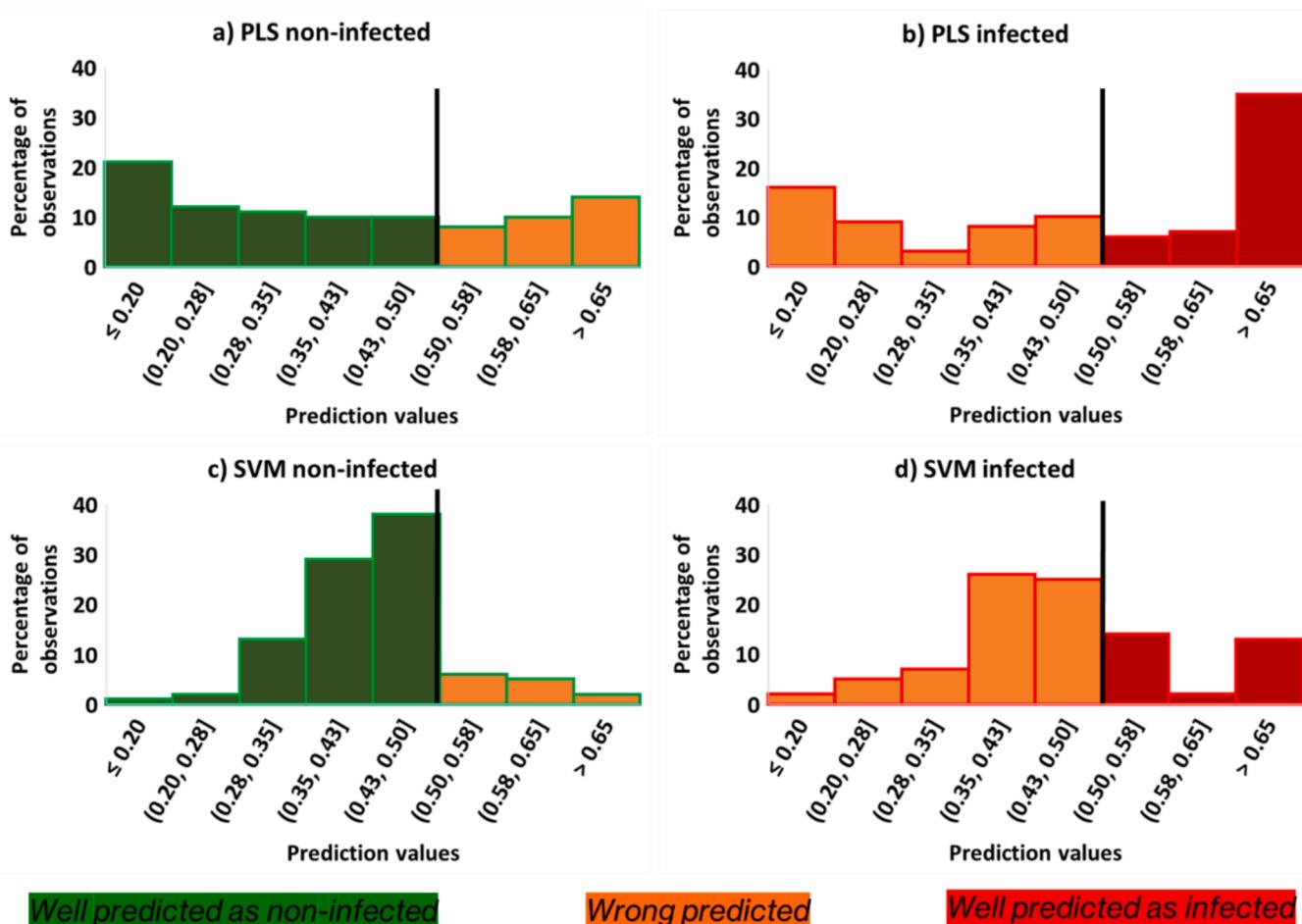


Fig. 10. Detail of the prediction values for both SVM and PLS-DA model for the non-infected and the infected vines images using the binning by 10 datasets with 23 bands for no visible symptoms stage of the infection.

model at post-veraison (Sawyer et al., 2023). The small number of samples could have contributed to the overall lower accuracy than in the present study. The current study is the first to use hyperspectral VIS-NIR images of the entire canopy to identify red blotch-infected vines. An overall accuracy of 73.5 % was achieved for the whole dataset with the accuracy exceeding 75 % when considering only the data acquired at post-veraison when disease symptoms are most visible (Table 6). The accuracy of a SVM model using 19 bands of 16 nm width (75.7 %) was slightly lower than the expert prediction using RGB images (78.6 %) (section 3.4.1.) or in the field (averaged 79.45 % for both years). These results demonstrated that an expert-like level of accuracy can be obtained by employing a field camera equipped with a limited number of bands. This approach could offer several advantages, including reduced costs (Abd-Elrahman et al., 2011), storage space requirements, and time and power consumption during analysis (Silva et al., 2022).

4.2. Assessment of symptoms variability

4.2.1. Visible and no visible symptoms

Symptoms of red blotch in infected vines evolve during the season and are fully visible post-veraison in California (around harvest) (Sudarshana et al., 2015). From the pre-veraison dataset with no visible symptoms (August), the only significant difference in terms of the spectral signal was in the NIR domain, while for the post-veraison dataset (September and October), the separation was in the visible domain (mainly the red domain) (Fig. 3). In fact, by looking at the expert predictions using the RGB images, the pre-veraison dataset was much more difficult to predict, with an accuracy of less than 54 % (slightly

higher than chance, 50 %) against 78.6 % for the late season.

The virus is unevenly distributed within the vine, meaning even leaves without visible symptoms can still be infected. Molecular diagnostic techniques, like PCR, confirm that the virus may be present in asymptomatic leaves (Flasco et al., 2024). This irregular distribution in the canopy poses challenges for detection, as it complicates both visual assessments and the use of hyperspectral imaging. While hyperspectral imaging offers a powerful tool for detecting infections even in asymptomatic stages, the uneven distribution of the virus adds complexity to its application and analysis.

Regarding the machine learning model, the best accuracy for the pre-veraison dataset was 74.2 % against 75.7 % for the post-veraison dataset (Table 6). With the pre-veraison dataset, it is more challenging to predict the infection for a machine learning model, as already shown by previous studies using leaf images to detect red blotch (Sawyer et al., 2023) or Esca (Pérez-Roncal et al., 2022). The low number of samples for this category (76 samples) (Table 3) could also have impacted the accuracy of the model, especially when comparing to the post-veraison dataset (over 500 samples). This was particularly evident when assessing the performance on the three-category classification (Fig. 9). This model predicted infected vines in images taken with canopies without visible symptoms with good accuracy for both infected categories (more than 70 %). Curiously, the accuracy for the post-veraison dataset (73 %) was lower than the accuracy for the pre-veraison dataset (77 %). This result can be explained by the fact that the distinction in the NIR domain for the pre-veraison dataset (Fig. 3b) is higher than the distinction in the visible domain for the post-veraison dataset (Fig. 3c). This may be due to the symptom variability (the post-veraison dataset includes multiple

acquisition dates), but also in the method used to analyze the data.

4.2.2. Variability in symptom expression

Disease symptoms in vineyards may vary in intensity from year to year, and the accuracy of the predictive model will also differ, as shown by Gao et al. (2020). These authors detected grapevine leafroll disease using hyperspectral images of leaves, with infected category accuracies ranging from 72 % to 93 % according to the year of the study. Here, using canopy images, better accuracies were found for 2021 (77.7 %) compared with 2020 (75.2 %) (Table 4), and more differences were observed in the spectral signatures of 2021 compared with 2020, especially in September (Fig. 3f and 3d). The expert's accuracy also increased from 75.3 % in 2020 to 83.6 % in 2021 (section 3.4.2). Both years (2020 and 2021) were dry, but 2021 was drier than 2020 (Williams et al., 2022). It is known that red blotch symptoms are exacerbated in conditions of water stress, as reported by Levin and KC (2020). The phenological phase of acquisition may impact the results, and for maximum performances, it may be important to image vineyards later in the season at the time of maximum symptom expression. This is well known and applied also in sampling protocols for virus diagnostics using molecular techniques in the laboratory.

Another source of disease symptom variability besides the phenological time of acquisition and seasonal conditions is the cultivar. Each cultivar can show differences in symptom expression and have a different spectral baseline. This is most evident in white cultivars (Krenz et al., 2014). It is not possible to directly apply a model developed for red cultivars to white cultivars, as reported by Albetis et al. (2017). An interesting expansion of this research would be the investigation of the methods described here in multiple red and white cultivars. Improving the detection of GRBV in vineyards of white cultivars is critical because monitoring of symptoms is less reliable than for red cultivars.

4.2.3. Influence of the method used

The method used in this study (using the average value of the canopy pixels of each image) can impact the results. As the overall canopy signal is averaged, when symptoms are variable or little expressed, and the virus is unevenly distributed in the infected vine, the information is easily diluted and can impact classifier performances. Nonetheless, annotating images and using them in their entirety would have only been achievable through human vision, therefore limiting the study to the visible domain. In addition, such an approach would not have allowed the study of images of asymptomatic vines. In other words, because pre-symptomatic disease expression is non-visible to the human eye, there is a limitation to the machine-learning techniques that need to be leveraged, particularly, the spectral signal must be aggregated. Since it is not possible to annotate imagery, limiting the effectiveness of techniques such as convolutional neural networks, studying methods that could leverage the richness of the hyperspectral signal is worth investigating.

4.3. Determination of spectral bands importance

Both visible (510 nm to 700 nm) and NIR domains (700 nm to 900 nm) were analyzed in this study with wavelengths ranging from 510 nm to 900 nm. The first domains most used by the models tested in this study were the green and the red domains (Fig. 6). Hodáňová (1985) demonstrated that the visible domain is impacted by pigments concentration, such as chlorophyll (green domain) or anthocyanins (red domain) which are affected by the presence of diseases. In fact, diseases decrease the production of chlorophyll and increase the production of anthocyanins (Carter and Knapp, 2001). This resulted in a peak in the green domain for the non-infected vines and a peak in the red domain for the GRBV-infected vines (Fig. 3). The second domain most used by the models in this study was the "red-edge" (Fig. 6). The absorption of pigments experienced a steep decrease beyond 700 nm, where the absorption of the photosystem I reaction center (P700) is maximal, leading

to a spectral region of rapid transition known as the "red-edge". This significant rise in reflectance, corresponding to the decline in absorption, is the most distinct attribute of vegetation spectra (Hallik et al., 2019). Yet the photosystem I can be highly affected by the presence of a pathogen (Asada 2006; Yang and Luo 2021). This was evidenced in our study by a distinct signal of GRBV-infected and non-infected vines (Fig. 3). Finally, the last domain which was important to identify red blotch infection is the NIR (Fig. 6). NIR domain is impacted by macro components such as water for example (Walsh et al., 2020) or by changes in the leaf mesophyll structure (Zahir et al., 2022), among others. Diseases can alter these parameters (reducing the water content of the plant or changing the leaf structure). This impact was stronger for vines at pre-veraison (Fig. 3b) than post-veraison (Fig. 3c). This may be due to the fact that the disease can be the dominant factor affecting the leaf spectral signal at an early stage; while at a late stage in the season, other factors such as water stress or leaf senescence can influence the spectral signal.

4.4. Model performance comparison

4.4.1. Impact of algorithm, feature reduction and feature selection

The PLS-DA and SVM models do not use the same wavelengths for predicting GRBV infection (Fig. 6). PLS-DA performed better at pre-veraison while SVM performed better at post-veraison, and the combination of their predictions gave better results only at post-veraison (Table 5). This result can be explained by the fact that the contribution to the reduction in the error rate was different for similar wavelengths between the two models, with their combined outputs proving beneficial. At pre-veraison, SVM better identified the non-infected category with an accuracy of 85 % (Fig. 10c) while PLS-DA yielded better results for the prediction of the infected category with an accuracy of 48 % (Fig. 10b). PLS-DA was used as a reference in this study as it is a standard practice in the treatment of spectroscopic data, but SVM has raised in popularity because of its ability with high dimensional input vectors and approach of nonlinear problems. The low performance of RF (Table 4) could be related to the fact that it uses a relatively high percentage of non-informative bands using a random selection of predictors at each tree split.

Using spectral binning to reduce the noise in the signal yielded a higher accuracy in most cases (Table 5). Spectral data often contain noise or irrelevant fluctuations that can obscure meaningful information. Binning averages the spectral values within each bin, effectively reducing the impact of noise and improving the signal-to-noise ratio. This noise reduction can enhance the accuracy of subsequent data analysis. Binning also reduces the dimensionality of the data by collapsing multiple spectral data points into a smaller number of bins. This simplification of features can prevent overfitting in machine learning models and improve generalization performances. Binning can also increase the robustness of spectral data analysis to variations in experimental conditions especially when working at different stages of the disease infection during two seasons. By aggregating spectral values within each bin, the analysis becomes less sensitive to small fluctuations or shifts in the data, resulting in more consistent and reliable results (Yoon and Park, 2015). In this study, the best results were achieved with binning by 10, which reduced the dataset to 23 bands. This likely provided the optimal balance between removing noise from the data and retaining important spectral information. Binning by 10 allowed us to simplify the data without losing significant details that contribute to accurate model predictions, whereas other binning methods either retained too much noise (binning by 5) or discarded too much information (binning by 20).

Reduction of redundant information through binning and applying RFE improved even more the overall accuracy by an average of 3 percent points (Table 6), leading to an accuracy of 75.7 % at post-veraison using only 19 bands with a width of 16 nm. At pre-veraison, the best model (74.2 %) is obtained using UFS-SVM with 5 bands (16 nm width)

(Table 6). In general, in this study, RFE gave better results for the dataset with visible symptoms while UFS gave better results for the dataset with no visible symptoms. This could be explained by the different nature of each method. UFS is simpler and may work well with fewer samples and less variability. The dataset with no visible symptoms is smaller and includes only data from one year. While RFE iterative approach is better suited for datasets with more samples and more complex relationships between features and the target variable which may be the case for the dataset with visible symptoms with two different years of study.

Removing auto-correlated bands using ANOVA led to lower accuracies in most of the cases except for the no visible symptoms dataset with 228 and 46 bands. In those two cases with fewer samples and more features, removing auto-correlated features can have led to better accuracies by reducing dimensionality (in datasets with fewer samples and more features, the curse of dimensionality becomes more pronounced), mitigating multicollinearity (with fewer samples, multicollinearity can be more detrimental because it exacerbates the instability of parameter estimates), and reducing noise (auto-correlated features may contain noise or irrelevant information that could adversely affect model performance, especially in datasets with limited sample).

These findings are important in advancing the methodologies toward a commercially viable operational service, as a reduced number of bands simplifies the sensor and data processing and makes them more cost-effective.

4.4.2. Machine learning and expert predictions

The purpose of this study was not to directly compare automatic identification with expert predictions. However, data from experts were included as a reference to better understand model performances and evaluate the potential for practical applications. The best model (SVM binary model with 19 bands) obtained comparable overall accuracy with expert predictions on the RGB images for the visible symptoms stage (75.7 % against 78.6 %). At pre-veraison, the overall accuracy was 74.2 % with 5 bands, while it was less than 55 % for the expert predictions. This is due to the fact that, if symptoms are not visible to human eyes yet, they can be observed by the camera in the NIR. The overall accuracy of the expert predictions in the vineyard is a few points lower than the other methods, although this vineyard assessment was carried out around harvest, the time of maximum symptom expression. The intensity of symptom expression is critical for virus assessment both for expert and machine-learning predictions from hyperspectral images. The best performances were obtained close to the harvest date, at the time of maximum symptom expression.

Expert predictions did not reach 100 % accuracy because even at post-veraison, the time of maximum symptomatic expression, some vines that tested positive for GRBV in PCR did not manifest visible symptoms. It was recently shown that it takes at least 15 months for a vine that is inoculated by viruliferous three-cornered alfalfa hoppers to exhibit red blotch disease symptoms (Flasco et al. 2023). This situation resulted in 41 % false negatives for the expert's prediction using the RGB images (Fig. 4c) and 17 % to 24 % false negatives for the vineyard predictions in 2020 and 2021, respectively (Fig. 5). One important consideration is that expert predictions can appear as a quick and easy method to collect large datasets to train machine-learning models, as in recent work on grapevine leafroll identification (Romero Galvan et al., 2023). However, it is limited by an expert's ability to observe visible symptoms, and therefore, prone to inaccuracy in assessing false negatives. The testing of each vine in the training set, as done in this study, is therefore paramount to obtain an accurate assessment of model performances and avoid disastrous consequences in practical applications.

Experts can predict infected vines with decent accuracy, as previously documented (Bell et al., 2017). In our study, the hyperspectral assessment performed similarly to experts at the post-veraison stage and outperformed the experts at the pre-veraison stage. The hyperspectral system could be further improved through additional work evaluating the best time of image acquisition, using multiple images of the same

vines, spectral change detection with time, and different aggregation methods of the spectral signal. The clear advantage of the hyperspectral assessment is autonomy: cameras can be embedded in mobile systems (field vehicles, unmanned ground vehicles) and allow a scalable, rapid vine-by-vine infection assessment over large surfaces, which is not operationally possible by expert scouting.

4.5. Potential for future research

In considering the potential for future research, it is important to acknowledge the limitations of current approaches in hyperspectral imaging for vineyard disease detection. Promising results have been achieved, such as a 76.6 % detection accuracy for post-veraison red blotch virus identification. However, challenges persist in scaling these technologies and applying them practically in agriculture, as acknowledged by previous studies (Liu et al., 2024; Ang and Seng, 2021; Ram et al., 2024). Key areas for future investigation include the development of more cost-effective and user-friendly sensors, as well as optimizing image acquisition timing across different phenological stages. Further exploration of hyperspectral detection in white grape cultivars, which exhibit different disease symptom types, will be essential to enhance generalizability. These advancements, alongside efforts to address variability in disease symptoms and environmental conditions, will be pivotal in improving the efficiency and accuracy of vineyard disease management.

5. Conclusion

GRBV has a negative impact on vineyard performance and quality of wines. There is no known cure for the virus in the vineyard. The only effective way to limit the virus spread is to remove infected plants and replace them with healthy ones. This process, known as roguing, is both time-consuming and expensive due to the need to accurately identify infected vines in the vineyard. Infected vineyards are commonly scouted based on visual monitoring of disease symptoms, a challenging approach due to several confounding factors. To address this limitation, nearly 700 hyperspectral images of vines canopies were acquired in the vineyard and tested by machine-learning identification of infected vines for which the presence of GRBV was determined by PCR. This research evaluated methods of dimensionality reduction to simplify image acquisition and improve model performances. The ultimate objective is to lay a solid foundation for autonomous scouting of vineyards to help growers manage virus spread while saving time and money.

Models demonstrated an overall accuracy of 74.2 % at pre-veraison, when disease symptoms are not visible (obtained with SVM using only 5 bands with a 16 nm width), and an overall accuracy of 75.7 % at post-veraison, when disease symptoms are visible (obtained with SVM using 19 bands with a width of 16 nm). These accuracies are comparable to expert assessment for the visible symptom stage but are much higher for the non-visible symptom stage.

This work is the first to investigate the use of hyperspectral images of vine canopies acquired in the vineyard to identify red blotch infection. It is also one of the few studies attempting dimensionality reduction through spectral binning and recursive feature elimination to move toward simpler sensors for the operational identification of grapevine viruses by growers in the field. It is possible to improve hyperspectral assessment even further by optimizing the phenological stage for image acquisition. In future research, it would be interesting to automatize image acquisition in the field and expand the investigation to white cultivars and to more complex datasets including confusing leaf symptoms cause by different biotic and abiotic factors or cultivars.

CRediT authorship contribution statement

E. Laroche-Pinel: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation,

Conceptualization. K. Singh: Writing – review & editing, Investigation. **M. Flasco:** Writing – review & editing, Investigation. **M.L. Cooper:** Writing – review & editing, Validation, Supervision, Resources, Investigation, Funding acquisition, Data curation, Conceptualization. **M. Fuchs:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **L. Brillante:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Luca Brillante reports financial support was provided by California Department of Food and Agriculture. Luca Brillante reports financial support was provided by California State University Agricultural Research Institute. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper].

Acknowledgements

Funding support for this project originated from the CDFA-SCGB #19-0001-039-SF, and from the CSU-ARI System Grant #21-01-123. Authors wish to thank Bronco Wine Co. for continued support of the lab activities, the National Grape Research Alliance for early support of the project, St. Supery Estate Vineyards and Winery for access to vineyards and co-operation. Takuy Frnzian, Guadalupe Partida and Selena Vengco for help with sample collection, image and data acquisition, and Benjamin Corrales for his work on the semantic segmentation.

Data availability

Data will be made available on request.

References

- Abd-Elrahman, A., Pande-Chhetri, R., Vallad, G., 2011. Design and development of a multi-purpose low-cost hyperspectral imaging system. *Remote Sens. (Basel)* 3 (3), 570–586. <https://doi.org/10.3390/rs3030570>.
- Albetis, J., Duthoit, S., Guttler, F., Jacquin, A., Goulard, M., Poilv , H., F ret, J.B., Dedieu, G., 2017. Detection of Flavescence dor e grapevine disease using unmanned aerial vehicle (UAV) multispectral imagery. *Remote Sens. (Basel)* 9 (4), 308.
- 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 Sens. (Basel)* 10 (618), 1–18. <https://doi.org/10.3390/rs10040618>.
- Ang, K.L.M., Seng, J.K.P., 2021. Big data and machine learning with hyperspectral information in agriculture. *IEEE Access* 9, 36699–36718.
- Asada, K., 2006. Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiol.* 141 (2), 391–396.
- Barker, M., Rayens, W., 2003. Partial least squares for discrimination. *J. Chemom.* 17 (3), 166–173. <https://doi.org/10.1002/cem.785>.
- Beget, M.E., Di Bella, C.M., 2007. Flooding: The effect of water depth on the spectral response of grass canopies. *J. Hydrol.* 335 (3–4), 285–294.
- Belgiu, M., Dr  gut, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31.
- Bell, V.A., Blouin, A.G., Cohen, D., Hedderley, D.I., Oosthuizen, T., Spreeth, N., Lester, P.J., Pietersen, G., 2017. Visual symptom identification of grapevine leafroll-associated virus 3 in red berry cultivars supports virus management by roguing. *J. Plant Pathol.* 99 (2), 477–482. <http://www.jstor.org/stable/44686794>.
- N. Bendel, A. Kicherer, A. Backhaus, J. K ckerling, M. Maixner, E. Bleser, et al., 2020a. Detection of grapevine leafroll-associated virus 1 and 3 in white and red grapevine cultivars using hyperspectral imaging. *Remote Sens.* 12(10). doi: [10.3390/rs12101693](https://doi.org/10.3390/rs12101693).
- Bendel, N., Kicherer, A., Backhaus, A., et al., 2020. Evaluating the suitability of hyper- and multispectral imaging to detect foliar symptoms of the grapevine trunk disease Esca in vineyards. *Plant Methods* 16, 142. <https://doi.org/10.1186/s13007-020-00685-3>.
- B.E. Boser, I.M. Guyon, V.N. Vapnik, 1992. A training algorithm for optimal margin classifiers, Fifth Annual Workshop on Computational Learning Theory, ACM, 144–152, June 1992.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32.
- Carter, G.A., Knapp, A.K., 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *Am. J. Bot.* 88 (4), 677–684. <https://doi.org/10.2307/2657068>.
- Condorcet, M.D., 1785. *Essay on the application of analysis to the probability of majority decisions*. Imprimerie Royale, Paris.
- DeShields, J.B., Kc, A.N., 2023. Comparative diagnosis of grapevine red blotch disease by endpoint PCR, qPCR, LAMP, and visual symptoms. *Am. J. Enol. Vitic.* 74 (1).
- T.G. Dietterich, 2000. Ensemble Methods in Machine Learning. In: *Multiple Classifier Systems*. MCS 2000. Lecture Notes in Computer Science, vol 1857. Springer, Berlin, Heidelberg. doi:[10.1007/3-540-45014-9_1](https://doi.org/10.1007/3-540-45014-9_1).
- D  zrovska, S., Zenko, B., 2004. Is combining classifiers with stacking better than selecting the best one? *Mach. Learn.* 54, 255–273. <https://doi.org/10.1023/B:MACH.0000015881.36452.6e>.
- M. Flasco, V. Hoyle, E.J. Cieniewicz, B.G. Roy, H.L. McLane, K.L. Perry, G. Loeb, B. Nault, M. Heck, M. Fuchs, 2021. Grapevine red blotch virus is transmitted by the three-cornered alfalfa hopper in a circulative, nonpropagative transmission mode with unique attributes. *Phytopathology*, 111: 1851-1861, doi:[10.1094/phyto-02-21-0061-r](https://doi.org/10.1094/phyto-02-21-0061-r).
- Flasco, M., Hoyle, V., Cieniewicz, E.J., Loeb, G., McLane, H., Perry, K., Fuchs, M., 2023. The three-cornered alfalfa hopper, *Spissistilus festinus*, is a vector of grapevine red blotch virus in vineyards. *Viruses* 15, 927. <https://doi.org/10.3390/v15040927>.
- Flasco, M., Cieniewicz, E.J., Cooper, M.L., McLane, H., Fuchs, M., 2024. Investigating the latency period of grapevine red blotch virus in a diseased Cabernet franc vineyard experiencing secondary spread. *Am. J. Enol. Vitic.* 75 (1).
- Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., Suganthan, P.N., 2022. Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* 115, 105151.
- Gao, Z., Khot, L.R., Naidu, R.A., Zhang, Q., 2020. Early detection of grapevine leafroll disease in a red-berried wine grape cultivar using hyperspectral imaging. *Comput. Electron. Agric.* 179, 105807.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3 (Mar), 1157–1182.
- Hallik, L., Kuusk, A., Lang, M., Kuusk, J., 2019. “reflectance Properties of Hemiboreal Mixed Forest Canopies with Focus on Red Edge and near Infrared Spectral Regions” *Remote Sensing* 11 (14), 1717. <https://doi.org/10.3390/rs11141717>.
- D. Hod  nov , 1985. “Leaf optical properties,” in *Photosynthesis during leaf development*. Tasks for vegetation science, vol. 11. Ed. Zdenek, Sestak (Dordrecht: Springer), 107–127. doi:[10.1007/978-94-009-5530-1_5](https://doi.org/10.1007/978-94-009-5530-1_5).
- Jovi , A., Brki , K., Bogunovi , N., 2015. A review of feature selection methods with applications. 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). Opatija, Croatia 2015, 1200–1205. <https://doi.org/10.1109/MIPRO.2015.7160458>.
- Junge, A.H., Ducati, J.R., Scalvi Lampugnani, C., Alman a, M.A.K., 2018. Detection of grapevine leaf stripe disease symptoms by hyperspectral sensor. *Phytopathol. Mediterr.* 57 (3), 399–406. https://doi.org/10.14601/Phytopathol_Mediterr-22862.
- Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., Casal, G., 2016. Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. *Remote Sens. (Basel)* 8 (6), 497. <https://doi.org/10.3390/rs8060497>.
- Krenz, B., Thompson, J., McLane, H., Fuchs, M., Perry, K.L., 2014. Grapevine red blotch-associated virus is widespread in the United States. *Phytopathology* 104, 1232–1240.
- Levin, A.D., Kc, A.N., 2020. Water deficits do not improve fruit quality in grapevine red blotch virus-infected grapevines (*Vitis vinifera* L.). *Front. Plant Sci.* 11, 1292.
- Liu, Y., Feng, H., Yue, J., Jin, X., Fan, Y., Chen, R., Bian, M., Ma, Y., Li, J., Xu, B., Yang, G., 2024. Improving potato AGB estimation to mitigate phenological stage impacts through depth features from hyperspectral data. *Comput. Electron. Agric.* 219, p108808.
- MacDonald, S.L., Staid, M., Staid, M., Cooper, M.L., 2016. Remote hyperspectral imaging of grapevine leafroll-associated virus 3 in cabernet sauvignon vineyards. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2016.10.003>.
- Marshall, M., Thenkabail, P., 2015. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS GeoEye-1 WorldView-2 Landsat ETM+ and MODIS vegetation indices in crop biomass estimation. *ISPRS J. Photogrammetry Remote Sens.*, 108, 205–218.
- M. Mehrubeoglu, K. Orlebeck, M.J. Zemlan, W. Autran, 2016. Detecting red blotch disease in grape leaves using hyperspectral imaging. In Velez-Reyes, M. and Messinger, D. W., editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXII*, volume 9840 of *Society of PhotoOptical Instrumentation Engineers (SPIE) Conference Series*, 98400D, doi:[10.1117/12.2223814](https://doi.org/10.1117/12.2223814).
- Nguyen, C., Sagan, V., Maimaitiyiming, M., Maimaitijiang, M., Bhadra, S., Kwasniewski, M.T., 2021. Early detection of plant viral disease using hyperspectral imaging and deep learning. *Sensors* 21 (3), 742. <https://doi.org/10.3390/s21030742>.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* 26 (1), 217–222.
- Pavlyshenko, B., 2018. Using stacking approaches for machine learning models. In 2018 IEEE second international conference on data stream mining & processing (DSMP). 255–258. IEEE.
- P  rez-Roncal, C., Arazuri, S., Lopez-Molina, C., Jar n, C., Santesteban, L.G., L  pez-Maestresalas, A., 2022. Exploring the potential of hyperspectral imaging to detect Esca disease complex in asymptomatic grapevine leaves. *Comput. Electron. Agric.* 196, 106863.
- M. Pugliese, M.L. Gullino, A. Garibaldi, 2011. Effect of climate change on infection of grapevine by downy and powdery mildew under controlled environment. *Commun Agric Appl Biol Sci.* 2011;76(4):579–82. PMID: 22702176.

- Ram, B.G., Oduor, P., Igathinathane, C., Howatt, K., Sun, X., 2024. A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Comput. Electron. Agric.* 222, 109037.
- Ricketts, K.D., Gómez, M.I., Fuchs, M.F., Martinson, T.E., Smith, R.J., Cooper, M.L., Moyer, M.M., Wise, A., 2017. Mitigating the economic impact of grapevine red blotch: Optimizing disease management strategies in U.S. Vineyards. *Am. J. Enol. Vitic.* 68, 127–135. <https://doi.org/10.5344/ajev.2016.16009>.
- Romero Galvan, F.E., Pavlick, R., Trolley, G., Aggarwal, S., Sousa, D., Starr, C., Forrestel, E., Bolton, S., del Mar Alsina, M., Dokoozlian, N., Gold, K.M., 2023. Scalable early detection of grapevine viral infection with airborne imaging spectroscopy. *Phytopathology* 113 (8), 1369–1614.
- Rumbaugh, A.C., Sudarshana, M.R., Oberholster, A., 2021. Grapevine red blotch disease etiology and its impact on grapevine physiology and berry and wine composition. *Horticulturae* 7, 552. <https://doi.org/10.3390/horticulturae7120552>.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36 (8), 1627–1639. <https://doi.org/10.1021/ac00214a047>.
- Sawyer, E., Laroche-Pinel, E., Flasco, M., Cooper, M.L., Corrales, B., Fuchs, M., Brillante, L., 2023. Phenotyping grapevine red blotch virus and grapevine leafroll-associated viruses before and after symptom expression through machine-learning analysis of hyperspectral images. *Front. Plant Sci.* 14. <https://doi.org/10.3389/fpls.2023.1117869>.
- Setiono, F.J., Chatterjee, D., Fuchs, M., Perry, K.L., Thompson, J.R., 2018. The distribution and detection of Grapevine red blotch virus in its host depend on time of sampling and tissue type. *Plant Dis.* 102 (11), 2187–2193.
- Silva, D.M., Bernardin, T., Fanton, K., Nepaul, R., Pádua, L., Sousa, J.J., Cunha, A., 2022. Automatic detection of Flavescens Dorée grapevine disease in hyperspectral images using machine learning. *Procedia Comput. Sci.* 196, 125–132. <https://doi.org/10.1094/PHYTO-12-14-0369-FI>.
- Sun, J., Yang, J., Shi, S., Chen, B., Du, L., Gong, W., Song, S., 2017. Estimating rice leaf nitrogen concentration: influence of regression algorithms based on passive and active leaf reflectance. *Remote Sens.* 9, 951. <https://doi.org/10.3390/rs9090951>.
- Terentev, A., Dolzenko, V., Fedotov, A., Eremenko, D., 2022. Current state of hyperspectral remote sensing for early plant disease detection: A review. *Sensors* 22 (3). <https://doi.org/10.3390/s22030757>.
- Thomas, S., Kuska, M.T., Bohnenkamp, D., et al., 2018. Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. *J. Plant Dis. Prot.* 125, 5–20. <https://doi.org/10.1007/s41348-017-0124-6>.
- Vanegas, F., Bratanov, D., Powell, K., Weiss, J., Gonzalez, F., 2018. A novel methodology for improving plant pest surveillance in vineyards and crops using UAV-based hyperspectral and spatial data. *Sensors* 18 (1), 260. <https://doi.org/10.3390/s18010260>.
- Walsh, K.B., Blasco, J., Zude-Sasse, M., Sun, X., 2020. Visible-NIR ‘point’ spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. *Postharvest Biol. Technol.* 168, 111246.
- Wilcox, W.F., Gubler, W.D., Uyemoto, J.K., 2015. Compendium of grape diseases, disorders, and pests. APS Press, The American Phytopathological Society, Tomball, TX, pp. 39–45.
- Williams, A.P., Cook, B.I., Smerdon, J.E., 2022. Rapid intensification of the emerging southwestern North American megadrought in 2020–2021. *Nat. Clim. Chang.* 12, 232–234. <https://doi.org/10.1038/s41558-022-01290-z>.
- Yang, H., Luo, P., 2021. Changes in photosynthesis could provide important insight into the interaction between wheat and fungal pathogens. *Int. J. Mol. Sci.* 22 (16), 8865. <https://doi.org/10.3390/ijms22168865>.
- S.C. Yoon, B. Park, 2015. Hyperspectral image processing methods. *Hyperspectral Imaging Technology in Food and Agriculture*, 81–101, doi: [10.1007/978-1-4939-2836-1_4](https://doi.org/10.1007/978-1-4939-2836-1_4).
- Zahir, S.A.D.M., Omar, A.F., Jamlos, M.F., Azmi, M.A.M., Muncan, J., 2022. A review of visible and near-infrared (Vis-NIR) spectroscopy application in plant stress detection. *Physical, Sensors and Actuators A*.