19. Analysis of Flavescence dorèe leaf symptoms using hyperspectral imaging and machine learning

C. Nuzzi[1*], E. Saldi[2], I. Negri[2] and S. Pasinetti[1]
[1] *Department of Mechanical and Industrial Engineering (DIMI), University of Brescia, Via Branze 38, Brescia, 25123, Italy; * cristina.nuzzi@unibs.it*
[2] *Department of Sustainable Crop Productions (DI.PRO.VES.), Catholic University of the Sacred Heart, Via E. Parmense 84, 29122, Piacenza, Italy*

Abstract

This work presents a novel approach for the detection and classification of Flavescence dorèe disease based on hyperspectral imaging. Three machine learning models were trained to classify leaf samples of Pinot Noir into healthy, asymptomatic, and diseased classes according to a combination of vegetation indices calculated on the per-pixel spectrum of the sample. The dataset used included 201 hypercubes collected from the same field in 2023 and 2024, equally divided into the three classes. Results highlighted good performance (accuracy) for the model trained on the population of features extracted from the vegetation indices for each leaf and normalized using z-score.

Keywords: Flavescence dorèe, hyperspectral imaging, machine learning, spectroscopy

Introduction

Flavescence dorèe (FD) is a quarantine disease affecting European vines. Known symptoms are evident in several parts of the plant, including wood, fruits, and leaves (European Food and Safety Authority *et al*., 2020). According to the grape variety, leaf symptoms are different (leaves become red in the case of dark grapes, yellow for white grapes). The variety also influences the plant's resistance to the disease; however, even if it survives the infection, it cannot fully recover, leading to large cost for winemakers. The phytoplasma is transmitted to the plants by a vector insect, *Scaphoideus titanus*, and most control strategies focus primarily on the elimination of the insect in the field to prevent or limit the disease transmission. Several southern European countries are affected by FD, including Italy, France, Spain, Portugal, Croatia, and Slovenia. To control the spread of the disease and reduce potential contagion to nearby fields, local regulations force winegrowers to destroy the whole vineyard if at least 20% of it is affected by FD (Boulent *et al*., 2020). Since a whole growing season may pass between infection and the appearance of symptoms, it is fundamental to develop strategies for early detection to mitigate the impact of FD. Currently, the only reliable method to certify the presence of FD (which may also be confused with other diseases with similar symptoms) is to conduct a polymerase chain reaction (PCR) analysis on leaf samples; however, this method is effective only if the leaves exhibit clear symptoms, making early detection challenging (Pelletier *et al*., 2009). The research community is trying to come up with cost and time effective solutions, especially involving contactless sensors. The research described by Al-Saddik *et al*. (2017) focused on the analysis of spectral bands that highlight the presence of FD by using a portable spectro-radiometer to analyse the leaves of several affected plants of different varieties. Similar experiments were conducted by Junges *et al*. (2020), enforcing the conclusion that most physiological changes can be detected in the visible spectral range (400–900 nm) and in the range 1600–2200 nm, where water absorption and phenolic compounds accumulation effects can be observed. In addition to contact spectrometers, hyperspectral imaging (HSI) is rapidly gaining momentum as well. The main advantage is that the spectral information is acquired for each pixel of the leaf sample at once, producing a higher amount of data compared to contact spectrometers thanks to higher spatial

resolution. For example, Silva *et al.* (2022) adopted machine learning (ML) techniques to analyse hyperspectral images, obtaining promising preliminary results.
The objectives of the study were to deepen the approach presented by Nuzzi *et al.* (2024), in which only 100 leaf samples of Pinot Noir were used and the ML approach proposed focused on a single approach based on vegetation indices (VIs) computed from the average leaf spectra. One of the key points of this work is the addition of the "asymptomatic" class into the study, representing samples that should be positive to FD but are not yet showing symptoms (hence, they are undetectable by PCR methods). In this novel work, the dataset is expanded to 201 samples (67 per class) and three different feature-based ML approaches are proposed.

Materials

Instrumentation
The hyperspectral camera adopted was HERA VIS-NIR (Nireos, Milan, Italy), which has a spectral range of 400–1000 nm with spectral resolution of 5 nm (bands). The acquisition software is based on LabVIEW (National Instruments, Austin, Texas, USA). The camera overall dimensions are 205.0 × 150.0 × 83.5 mm and weighs about 2 kg. The focal length of the equipped lens is equal to 25 mm, and the image resolution is equal to 10254×1080 pixels. The camera was managed by a laptop computer equipped with Windows 11 Pro, CPU Intel i7 2.80 GHz and 32 Gb RAM, connected to the camera by two USB cables.

Data collection procedure
The experimental field of Pinot Noir considered for this work is located in Pometo, Colli Verdi (Pavia, Italy) and was suspected to be affected by FD based on the assessment of an expert agronomist. Leaf samples were collected in two campaigns: one in 2023 and another in 2024.
In 2023, leaf samples were collected (i) from specific plants identified by the expert agronomist as likely positive for FD, and (ii) from other plants in the field that were suspected to be free from FD. Healthy samples were collected from Pinot Noir plants cultivated in a protected environment. The samples were collected in three different days according to the field's availability: (i) on 14 July, (ii) on 6 September, and (iii) on 18 September (right before harvest). In 2024, a specific row of the vineyard was selected for the experiment that did not contain plants that were affected by FD during 2023. Data was collected in 4 days selected according to field availability: (i) on 31 May, (ii) on 1 July, (iii) on 6 August, and (iv) on 20 September. For both campaigns, the choice of acquisition days was motivated by the field availability, weather conditions and growth stage of the plants. Vines start sprouting new leaves for the season in May and start losing leaves at end of the season in late September; hence, the acquisitions were conducted when a suitable number of leaves could be collected from the plants, also according to the development of the symptoms (e.g., samples collected in May do not show symptoms yet, so they serve as the baseline to label ground-truth asymptomatic samples).
After taking the leaf samples from the vineyard, the acquisition of the corresponding hypercubes was conducted indoors in a nearby facility that provided the necessary power to run the equipment and the necessary light conditions. The time elapsed between sample collection and acquisition was approximately 15 min, ensuring that the degradation of the samples over time was avoided. The acquisitions were conducted nearby a window during summer mornings (from 10 AM to 12 AM), providing the camera with only natural light coming from the outside and ensuring a diffused light general illumination of the sample with no overexposed areas on their surface. This allowed simulation of a natural environment as much as possible. The leaves were placed on a white paper sheet positioned on the ground. The camera-object distance was set to 1 m, ensuring that the leaf was always in focus and providing a field of view (FoV) of 271×217 mm (that is enough to acquire a single leaf). The hypercubes were saved on disk in binary format and analysed by the proprietary

software afterward to compute the normalized reflectance spectrum of each pixel in the hypercube. The hypercube of the reference object (a Teflon sample) used for the normalization process was acquired before the acquisition of the samples. The normalized hypercubes were then converted in MATLAB format and each of them required around 1.4 Gb of space. After the hypercubes acquisition, all the leaf samples were analysed by an external laboratory to detect the presence or absence of FD following Pellettier *et al.* (2009) grapevine extraction protocol for Phytoplasma detection.

Composition of the datasets

The leaves were always collected from the sample plants selected for the experiment during the corresponding campaigns. Therefore, samples that were negative to FD at the start of the season but positive at the end were labelled as asymptomatic samples. For example, considering three samples taken from the same plant, the ones collected during the first and second acquisition dates were labelled as asymptomatic, while the one collected in the third date was labelled as diseased. As for healthy samples, in 2023 a total of 10 samples were collected from another field of Pinot Noir cultivated in the same conditions as the testing one. This choice was motivated by the fact that the whole experimental field was suspected to be affected by FD, so certified healthy samples were needed to validate the approach. For the data collected in 2024, healthy samples correspond to leaf samples taken from plants of the same field that always were negative to FD during the acquisitions. The data collected in 2023 is divided into 10 healthy samples, 45 asymptomatic samples, and 45 diseased samples (Nuzzi *et al.*, 2024). The data collected in 2024 contained only 22 diseased and 22 asymptomatic samples, while the other 124 samples were either healthy or affected by other pathologies (e.g., Peronospera). For the purposes of this work, the data collected in both campaigns was merged into a bigger dataset. The three classes were balanced using the same number of samples for each of them (67) as follows: for the healthy class, 10 samples from the 2023 campaign and 57 from the 2024 one, for the asymptomatic and diseased classes, 45 samples were taken from the 2023 campaign and 22 from the 2024 one.

For the ML model training and testing, the dataset was divided into two sub-datasets: the training dataset (D_{train}), containing 70% of the data (47 samples per class), and test dataset (D_{train}), containing the remaining 30% of the data (20 samples per class).

Methodology

Vegetation indices

A plethora of VIs can be adopted to facilitate the detection of stress in plants. VIs are designed to summarize the leaf reflectance spectrum according to specific areas of interest, e.g. the green, red, or NIR spectral regions. A total of 10 VIs were adopted: (i) the normalized vegetation Index (NDVI), (ii) the red-green index (RGI), (iii) the clorophyll index (CI), (iv) the modified chlorophyll absorption in reflectance index (mCARI), (v) the red-edge position (REP), (vi) the curvature of the spectral reflectance curve in the red-edge region (CUR), (vii) the modified anthocyanin reflectance index (mARI) (viii) a modified version of mARI proposed in (Nuzzi *et al.*, 2024) (mARI2), (ix) the anthocyanin content index (ACI), and (x) the modified anthocyanin content index (mACI). See Albetis *et al* (2017), Steele *et al* (2009) and Main *et al* (2011) for mathematical definitions. NDVI and RGI are commonly used as general indicators of plants' health, thus they were chosen as a mean of comparison with existing literature. This was also the case for red-edge indices REP and CUR, also used by Al-Saddik *et al.* (2017) in their work based on contact spectrometers. Clorophill-based indices (CI, mCARI) were selected because the presence of FD was suspected to reduce chlorophill production. Considering the selected vine variety (Pinot Noir), symptoms were expected to produce red pigments in the leaves. As a result, anthocyanin-based indices were chosen to detect them (mARI, mARI2, ACI, mACI).

Pre-processing

To prepare the data for the consequent analysis, a modified version of the pre-processing procedure detailed in Nuzzi *et al.* (2024) was applied (see Figure 1):

1. RGB projection. The corresponding colour image was reconstructed from the original hypercube of levels by considering the mean of the color channels. This was done to take into account light variations and reduce color alterations. The wavelengths used to average the channels were: from 581 to 667 nm for red, from 510 to 576 nm for green, from 481 to 496 nm for blue.
2. Computation of binary masks. To remove the background pixels and only keep pixels belonging to the leaf, the easiest way was to apply a binary mask on the hypercube and quickly delete unwanted pixels. The masks were computed using the segmentation open-source platform Roboflow (Dwyer *et al.*, 2024) that allows creation of polygonal labels with simple clicks thanks to powerful AI models. However, if the leaf contains holes, these were included in the contour and should be filtered out in a second step. The polygonal contour was afterwards converted to a binary mask M in MATLAB.
3. Leaf pixels extraction. By looping on the $i = 1 \ldots 120$ channels of the original hypercube, the resulting 2D matrix H_i and the mask M are both rolled, obtaining a vector each. The rolling procedure was used to convert multi-dimensional matrices into vectors by moving each row at the end of the previous one. Then, the 1D mask M was applied on the 1D hypercube channel H_i, obtaining a matrix containing only pixels belonging to the leaf, I_{leaf}, of size N_{px}×120, where N_{px}, was equal to the number of pixels belonging to the leaf.
4. Outliers removal. Since, for some leaves, holes could be present and retained in the mask, it was necessary to filter them out afterwards. To do so, the NDVI index was computed for each pixel of the leaf I_{leaf}. Pixels for which the NDVI value was less than 0.3 were considered background and removed from I_{leaf}. NDVI was chosen since it is the most effective for separating biological from inert material.
5. Mean spectra computation. After the outliers removal, the mean spectra of the sample was calculated as the mean for each column of I_{leaf}. This step produced an averaged spectrum S_{leaf} for each sample.
6. Vegetation indices computation. For each pixel of the filtered leaf hypercube I_{leaf}, the 10 VIs were calculated. As a result, each VI corresponded to a population of N_{px} values.
7. Extraction of features. A total of 10 features were extracted from each VI population, corresponding to the three quartile values (VI_{Q25}, VI_{Q50}, VI_{Q75}), the lower and upper outlier values (VI_{Lout}, VI_{Uout}), the population mean VI_{mean}, the position of the two whiskers of the population (VI_{Lwhisk}, VI_{Uwhisk}), the population kurtosis and skewness (VI_{ku}, VI_{sk}).

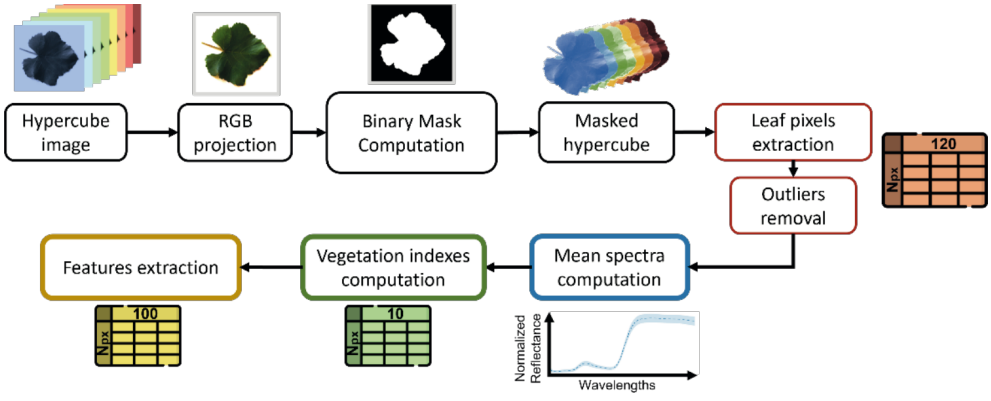

Figure 1. Scheme of the proposed pre-processing procedure.

Machine learning classification

MATLAB 2024b (The MathWorks, Natick, MA, USA) was adopted to conduct the analysis, specifically the Statistics and Machine Learning Toolbox. Three different approaches were tested: (i) classification using the full mean spectrum S_{leaf} of each leaf sample, for which the 120 values of the spectrum were the predictors, (ii) classification using only the mean values of each VI, VI_{mean}, for a total of 10 predictors, and (iii) classification using all the features extracted per VI as detailed in the "Features extraction" step of the pre-processing procedure, for a total of 100 predictors. Compared to the work in Nuzzi *et al.* (2024), the proposed approaches and their comparison are an important advancement in the topic, never proposed before, paving the way for more sophisticated solutions involving per-pixel analyses. For each classifier category available, the values of the corresponding approach calculated for the data in D_{train} were used for the training phase, and the results were evaluated using the values of the corresponding approach for the data in D_{test}. For all the approaches tested, cross-validation was conducted using 5 folds after training. It is worth noting that, for each approach, all the available ML models in MATLAB toolbox were tested, but for the sake of brevity only the best-performing ones are described in the following Section.

Results and discussion

For the first approach that involved the adoption of the mean leaf spectrum S_{leaf} (120 predictors) for each leaf, the best performing model was the support vector machine (SVM) kernel, which resulted in an accuracy after validation equal to 85% and, after test, equal to 75%. Figure 2a shows the confusion matrix obtained on D_{test}. The confusion matrix highlights that several occurrences of asymptomatic and diseased class were classified as healthy, which is a concerning result considering that diseased samples appear visibly different from healthy samples (e.g., red color, leaf edges rolled, different surface texture). This result could be due to the fact that the average spectrum of a diseased leaf could contain also green areas, especially in cases where the leaves have not yet completely turned red.

The second approach utilized 10 predictors only, corresponding to the 10 mean VI values calculated from the spectra population of each leaf sample. The best performing ML model was the linear discriminant adopting a full covariance structure, achieving 77% accuracy during validation and 80% during test. Figure 2b shows the confusion matrix obtained on D_{test}. In this case, only 10% of the diseased occurrences were classified as healthy or asymptomatic (1 occurrence each), while healthy and asymptomatic classes are confused between each other more often (4 asymptomatic occurrences classified as healthy, 3 healthy occurrences classified as asymptomatic). Despite being sub-optimal, this approach is still better compared to the previous one.

The last approach adopts a total of 100 predictors, 10 per VI. In this case, for each leaf sample the VIs were computed for each pixel belonging to the leaf, to which corresponds a full spectrum. Hence, by obtaining a population of VIs, it is possible to extract parameters representing their distribution that potentially include information about the presence of specific interesting spots in the leaf. The best performing model was a wide neural network with 1 fully connected layer, a first layer size of 100 neurons emplying ReLU activation function, iteration limit set to 1000 and a regularization strength set to 0. Data was standardized by the model before starting training. The model achieved a validation accuracy of 82% and a test accuracy of 88%. By looking at the confusion matrix in Figure 2c, it is evident how the model outperforms the other two, since all the diseased samples were correctly classified. Misclassifications occurred for the healthy class (3 samples classified as asymptomatic) and for the asymptomatic class (3 samples classified as diseased, 1 sample classified as healthy). It is worth noting that since asymptomatic samples should have FD but are not showing the symptoms, classifying them as diseased is better than classifying them as healthy, despite having more visual similarities with healthy samples. The results highlight how adopting all the VI's population features boost the recognition performance. Another test was conducted for this approach, this time normalizing the predictors using z-score normalization. The results were improved: the

best performing model was the efficient linear SVM with relative coefficient tolerance 0.001 and multiclass coding set to one-versus-one, achieving a validation accuracy of 82% and a test accuracy of 92%. The confusion matrix is shown in Figure 2d: asymptomatic and diseased classes were both recognized perfectly, while for the healthy class 5 occurrences were misclassified as asymptomatic. Considering that for FD it is best to have a false positive detection than a false negative one, this is an impressive result that highlights that feature normalization is of utmost importance to obtain accurate results, especially for VIs that have values very different from each other (e.g., NDVI values range from −1 to 1, while REP values were over 700). All the confusion matrices in Figure 2 also show row-normalized and column-normalized results, also known as the true positive rate (TPR), the false negative rate (FNR), the positive predicted value (PPV), and the false discovery rate (FDR).

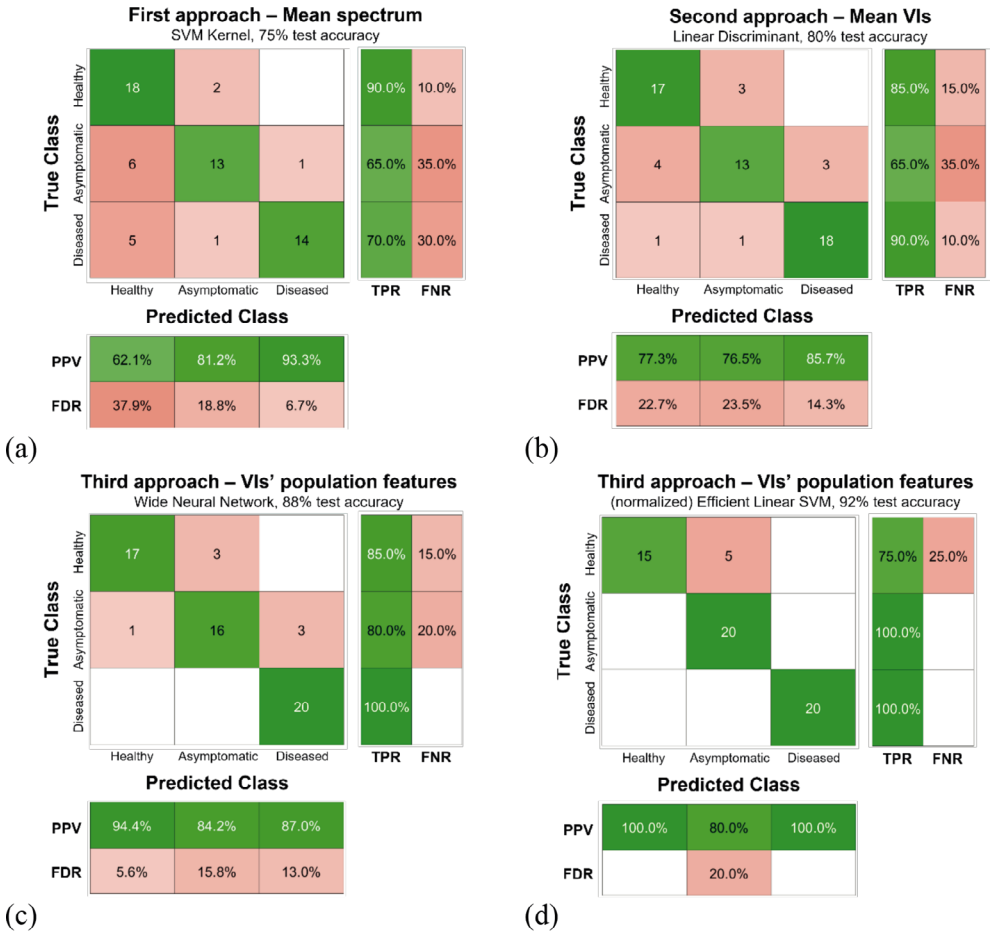

Figure 2. Confusion matrices obtained on the test dataset for each approach tested. (a) First approach (mean spectrum), (b) second approach (mean VIs), (c) third approach (VI population features), (d) third approach with normalized data (VI population features).

Conclusions

The work presented addresses the detection of FD in leaf samples using HSI combined with ML. The results highlight that the proposed methodology is robust and that all 10 vegetation indices are necessary for the ML model to reach high accuracy levels. In particular, the best results were obtained for the third approach that leverages a total of 100 features, 10 per VI, calculated to take into account the spectrum of each pixel belonging to the leaf samples. By normalizing those features using z-score normalization it is possible to further improve the classification accuracy, since the VI adopted have very different ranges. Although this method does not include image processing techniques that preserve the spatial information of red spots in the leaf, it represents a significant advancement in FD detection through HSI. Future developments will be focused on the development of ad-hoc image processing techniques. Moreover, to obtain reliable results the dataset dimension should increase as well; hence, new campaigns will be conducted to obtain more samples. In addition, data related to other varieties and diseases with similar symptoms (e.g., Bois Noir) will also be included in the dataset to improve robustness against false positive and false negative detections.

Acknowledgements

This work was supported by the European Union, FSE-REACT-EU, PON "Research and Innovation 2014-2020", D.M. 1062/2021, contract number 46-G-13219-3.

References

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), 308.

Al-Saddik, H., Simon, J.C., & Cointault, F. (2017). Development of spectral disease indices for 'Flavescence dorée' grapevine disease identification. Sensors, 17(12), 2772.

Boulent, J., St-Charles, P.-L., Foucher, S., & Théau, J. (2020). Automatic detection of Flavescence dorée symptoms across white grapevine varieties using deep learning. Frontiers in Artificial Intelligence, 3, 564878.

Dwyer, B., Nelson, J., & Hansen, T. (2024). Roboflow (Version 1.0). Available online at https://roboflow.com (accessed 29 January 2024).

European Food and Safety Authority, E. F., Tramontini, S., Delbianco, A., & Vos, S. (2020). Pest survey card on Flavescence dorée phytoplasma and its vector Scaphoideus titanus. EFSA Supporting Publication, 17(8), 1909E.

Junges, A., Almança, M., Fajardo, T., & Ducati, J. (2020). Leaf hyperspectral reflectance as a potential tool to detect diseases associated with vineyard decline. Tropical Plant Pathology, 45(5), 522–533.

Main, R., Cho, M., Mathieu, R., O'Kennedy, M., Ramoelo, A., & Koch, S. (2011). An investigation into robust spectral indices for leaf chlorophyll estimation. ISPRS Journal of Photogrammetry and Remote Sensing, 66(6), 751–761.

Nuzzi, C., Micheli, M., Papa, G., Negri, I., Saldi, E., & Pasinetti, S. (2024). Hyperspectral imaging combined with machine learning to classify Flavescence dorée symptoms. In IEEE Xplore 2024 IEEE International Workshop on Metrology for Agriculture and Forestry. pp 29–31.

Pelletier, C., Pascal, S., Gillet, J., Cloquemin, G., Very, P., Foissac, X., et al. (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. Vitis: Journal of Grapevine Research, 48(2), 87–95.

Silva, D., Bernardin, T., Fanton, K., Nepaul, R., Pádua, L., Sousa, J., & Cunha, A.. (2022). Automatic detection of Flavescense Dorée grapevine disease in hyperspectral images using machine learning. Procedia Computer Science, 196, 125–132.

Steele, M., Gitelson, A., Rundquist, D., & Merzlyak, M. (2009). Nondestructive estimation of anthocyanin content in grapevine leaves. American Journal of Enology and Viticulture, 60(1), 87–92.