MACHINE LEARNING METHODS FOR SUGAR QUANTIFICATION IN GRAPES BASED ON NEAR-INFRARED HYPERSPECTRAL IMAGING

Julien BOYER[1], Jordane POULAIN[1], Sylvie ROUSSEL[1], Nicolas SAURIN[2], Maxime RYCKWAERT[3], Ryad BENDOULA[3], Carole FEILHES[4] and Eric SERRANO[4]

[1]Ondalys, 4 rue Georges Besse, 34830 Clapiers, France

[2]ITAP, Univ Montpellier, INRAE, Institut Agro, Montpellier, France

[3]UE Pech Rouge, Univ Montpellier, INRAE, Gruissan, France

[4]IFV Sud-Ouest, V'innopôle, 1920, route de Lisle/Tarn, 81310 Peyrole, France

Keywords: Hyperspectral Imaging (HSI), Machine Learning, quantification, sugar, grape bunch

Hyperspectral imaging [1] has applications in many fields, including agriculture, environment, medicine and industry. It can be used to classify objects according to their composition, or to quantify compounds present on the surface and show their spatial distribution. This last case is often challenging as the reference is not known for each pixel. Image processing is even more important in this case to extract the Region of Interest (ROI) to build the model but also to be able to apply it on new images.

This scientific study presents the methodology for analyzing hyperspectral images for grape bunch maturity prediction directly in vineyards. In this example, the parameter to be predicted is the sugar content (in °Brix) on each whole bunch of grapes, providing an average measurement of sugar content per image.

The database includes 131 hyperspectral images of 2 red grape varieties (Fer Servadou N. and Syrah N.) acquired with the SPECIM IQ[1] (SPECIM - Konica Minolta) portable visible – near-infrared hyperspectral camera, with a spectral range of 400 to 1000 nm. Grape bunches are measured in the vineyard with the natural illumination of sunlight and without a uniform background. As a result, the image bank is quite heterogeneous. Extracting the object of interest (single bunch) is therefore a complex matter and a specific methodology and algorithm had to be created.

A methodology to automatically process the images has been developed. It consisted of different steps: 1) spectral preprocessing with row normalization, 2) extraction of the Region of Interest (ROI) by a Soft Independent Modeling of Class Analogy (SIMCA) [2] approach, combined with classical thresholding, and 3) average of the pixels.

The results of the quantitative models built on the average spectra shown that a Support Vector Machines (SVM) [3] regression model is significantly outperforming Partial Least Squares (PLS) regression [4].

The automatic tool for image processing has been successfully applied to predict the sugar content of the whole image ROI to visualize the pixel spatial distribution of the sugar in the bunch.

[1] https://www.specim.com/iq/

Julien BOYER, Jordane POULAIN, Sylvie ROUSSEL, Nicolas SAURIN, Maxime RYCKWAERT, Ryad BENDOULA, Carole FEILHES and Eric SERRANO. Machine learning methods for sugar quantification in grapes based on near-infrared hyperspectral imaging, ICNIRS 2025, June 2025, Rome, Italia.

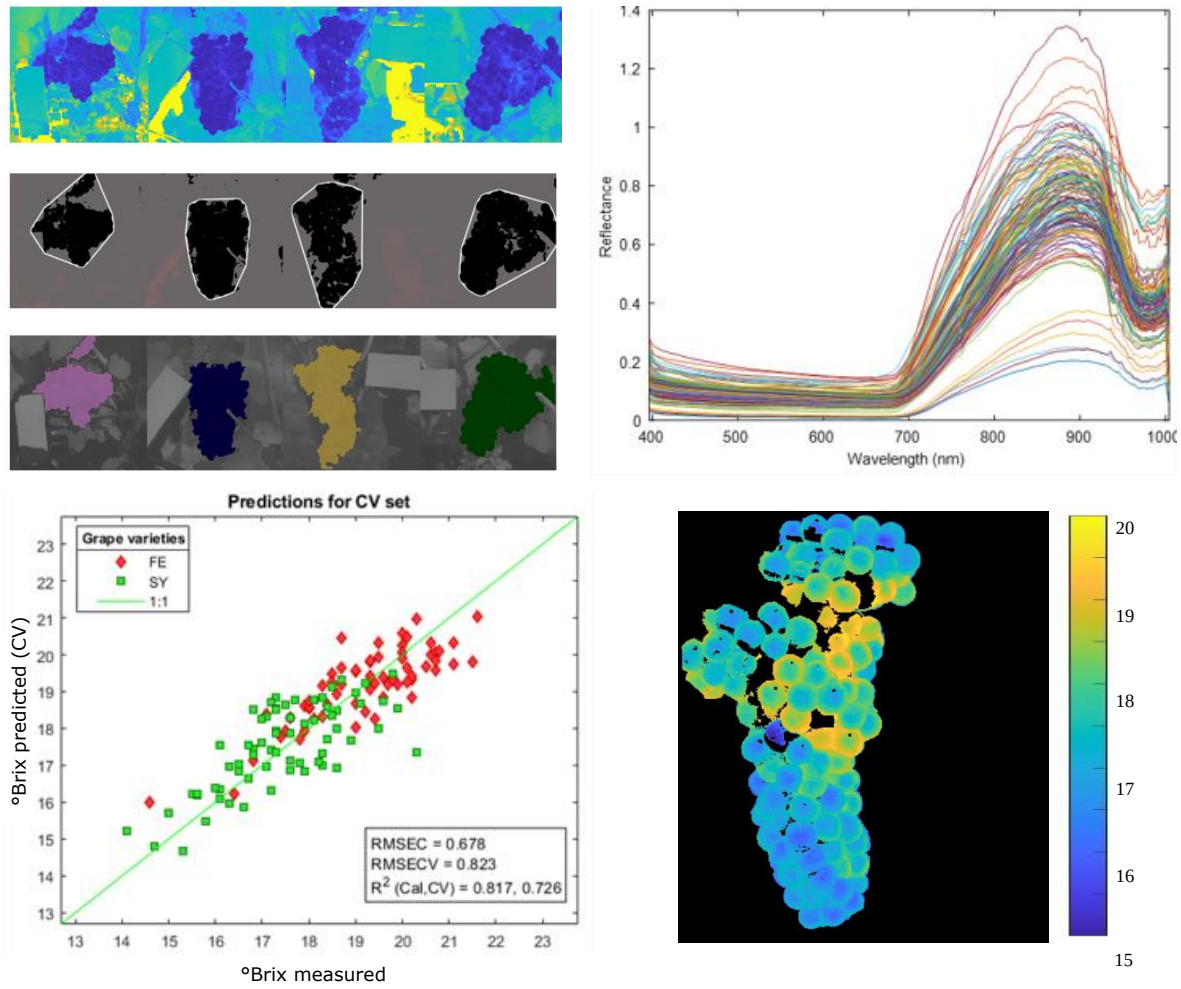

Figure 1: Example of extraction of the ROI on 4 images (top-left), resulting raw averaged spectra for all the images of the whole database (top-right), predictions in cross-validation of the average sugar content for each image of the SVM model (bottom-left) and predictions for each pixel for one of the images (bottom-right).

ACKNOWLEDGMENTS

This HIS data analysis study was carried out by Ondalys in collaboration with INRAE (UMR ITAP, Team COMIC and UE Pech Rouge) and IFV (IFV South West) as part of the VINIoT project, an Interreg SUDOE project funded by the European Regional Council.

[1] Amigo, José Manuel, éd. Hyperspectral imaging. Data handling in science and technology 32. Amsterdam, Netherlands; Cambridge, MA, United States: Elsevier, 2020.

[2] S. Wold, M. Sjostrom, SIMCA: A method for analyzing chemical data in terms of similarity and analogy. In Kowalski, B.R., ed., Chemometrics Theory and Application, American Chemical Society Symposium Series 52, 1977, p. 243-282.

[3] C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. 20, 273–297 (1995).

[4] H. Wold. Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiaah (Ed.). Multivariate Analysis. 1966. pp.391-420.