proceedings

MDPI

Proceedings

Non-destructive estimation of grape ripeness in Syrah variety via VNIR–SWIR spectroscopy

Nikiforos Samarinas [1], Nikolaos Tsakiridis [2], Nikolaos Tziolas [1], Eleni Kalopesa [1], Konstantinos Karyotis [3], Dionysis Bochtis [4], George Zalidis [1,5] and Zoe Doulgeri [6]

[1] School of Agriculture, Faculty of Agriculture, Forestry, and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki, 54123, Greece; {smnikiforos@auth.gr, ntziolas@auth.gr, kalopesa@auth.gr, zalidis@auth.gr}

[2] Department of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, 54123, Greece; tsakirin@auth.gr

[3] School of Science and Technology, International Hellenic University, 14th km Thessaloniki-N, Moudania, 57001 Thermi, Greece; kkaryotis@ihu.edu.gr

[4] Institute for Bio-Economy and Agri-Technology (IBO), Centre of Research and Technology-Hellas (CERTH), 6th km Charilaou-Thermi Rd, Thessaloniki, 57001, Greece; d.bochtis@certh.gr

[5] Interbalkan Environment Center, 18 Loutron Str., Lagadas, Greece; zalidis@auth.gr

[6] Automation Robotics Lab, Department of Electrical Computer Engineering, Aristotle University of Thessaloniki, Greece; doulgeri@ece.auth.gr

* Correspondence: smnikiforos@auth.gr;

† Presented at the title, place, and date.

Received: date; Accepted: date; Published: date

Abstract: Spectroscopy is a widespread technique used in many scientific fields such as in the food production. The use of hyperspectral data and specifically in the visible and near infrared (VNIR) and in the short-wave infrared (SWIR) regions in grape production is of great interest. Due to its fine spectral resolution, hyperspectral analysis can contribute to both fruit monitoring and quality control at all stages of maturity with a simple and inexpensive way. This work presents an application of a contact probe spectrometer that covers the VNIR–SWIR spectrum (350–2500 nm) for the quantitative estimation of the wine grapes' ripeness. A total of 110 samples of grape vine Syrah (*Vitis vinifera Syrah*) variety were collected over the 2020 harvest and pre-harvest seasons from Ktima Gerovassiliou located in Northern Greece. Their total soluble solids content (°Brix) was measured in-situ using a refractometer. Two different machine learning algorithms, namely partial least square regression (PLS) and random forest (RF) were applied along with several spectral pre-processing methods in order to predict the °Brix content from the VNIR–SWIR hyperspectral data. Additionally, the most important features of the spectrum were identified, as indicated by the most accurate models. The performance of the different models was examined in terms of the following metrics: coefficient of the determination (R^2), root mean square error (RMSE) and ratio of performance to interquartile distance (RPIQ). The values of R^2 = 0.90, RMSE =1.51 and RPIQ = 4.41 for PLS and 0.92, 1.34, 4.96 for RF respectively, indicate that by using a portable VNIR–SWIR spectrometer it is possible to estimate the wine grape maturity in-situ.

Keywords: vis–NIR; SWIR; feature importance; wine grapes; explainable artificial intelligence

1. Introduction

The common practice today used by the wine producers to assess the maturity of the grapes is to use a refractometer to measure in-situ the °Brix value which is the total soluble solids content in the grape. On the other hand, spectroscopy is a widespread technique used in many scientific fields such as in the food production [1] and soil science [2]. The regions of the visible and near-infrared

(VNIR) and the short-wave infrared (SWIR) in grape production have piqued the interest of researchers in the past [3]. The main goal of this research was to estimate the wine grape ripeness in a simple, inexpensive, and non-destructive way, taking advantage of the synergy of the aforementioned spectral regions with machine learning algorithms.

2. Materials and Methods

2.1. Main methodology

Figure 1a compares the two different techniques (destructive and non-destructive) in order to estimate the grape ripeness. Furthermore, for convenience, our methodology is presented in the form of a flowchart (Figure 1b).

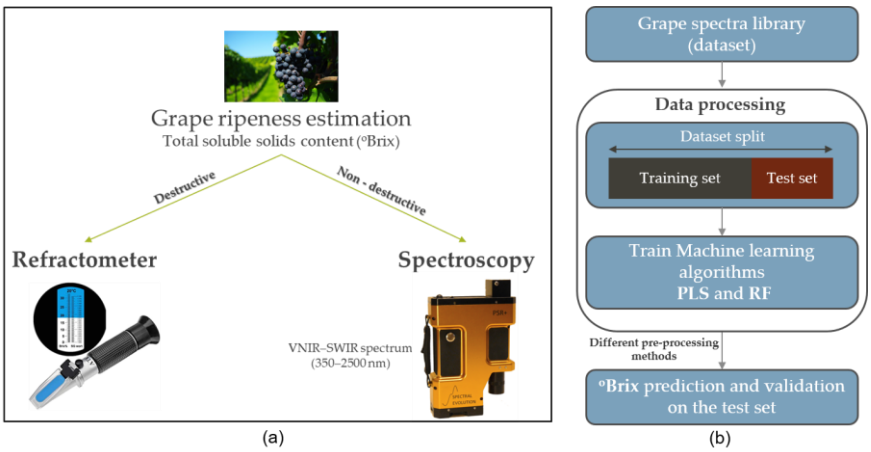

Figure 1. (a) Different techniques for grape ripeness estimation (b) Methodology flowchart

2.1. Pilot area measurements – Grape spectra library

Our pilot area is the vineyard of Ktima Gerovassiliou (https://www.gerovassiliou.gr/), located in Northern Greece, 25 km southeast of the city of Thessaloniki (40° 27' 04'' N, 22° 55' 23'' E). Ktima Gerovassiliou is one of the largest wine producers in the country cultivating several varieties of both white and red wine.

The first stage of the methodology was the creation of the grape spectral library. All measurements were performed in-situ using a portable contact probe spectrometer namely PSR +3500 (Spectral Evolution Inc., Lawrence, Massachusetts, USA) that covers the VNIR–SWIR spectrum (350–2500 nm). The final dataset created consists of 110 samples of the Syrah variety during the 2020 harvest and pre-harvest season from heterogeneous points in the vineyard. In addition, the grapes measured with the spectrometer were also used to measure their °Brix content with the help of a portable RHB-32ATC refractometer (Laxco Inc., Bothell, Washington, USA) but with a destructive effect of the fruit.

2.2. Data processing

2.1.1. Dataset split

The dataset is usually divided into training and test set in order to train and find the model hyperparameters (model selection) and estimate the model prediction error or accuracy, respectively. In our case the conditioned Latin hypercube method (cLHS) [4] was used and the grape spectra dataset was split with 67% as a training and with 33% as an independent test set.

2.1.2. Pre-processing methods

Data pre-processing is an important step in chemometrics analysis and having often a significant impact on the generalization performance of a supervised machine learning algorithm. In

this work several pre-processing methods were used (see Table 1). Initially the reflection data recorded were converted to pseudo-absorbance, which may present a more linear relationship between absorption and concentration of the grape properties according to the Beer–Lambert law. Furthermore, the method of the first derivative is a method that removes the baseline from the spectra and at the same time emphasizes the absorption characteristics. According to Ertlen et.al. [5] important information can be extracted from VNIR spectra if spectral derivatives are obtained. A Savitzky–Golay 1st order filter (SG1) is usually used to calculate the first derivative [6]. Furthermore, standard normal variate (SNV) and de-trending (DT) are common mathematical transformations used in spectroscopy.

2.1.3. Machine learning prediction algorithms

The following two algorithms were employed in the present study: (i) the Partial Least Squares regression algorithm (PLS) transforms the input space by selecting a number of orthogonal factors (named latent variables) that maximize the covariance between predictors and response variable, and then builds a linear model on this projection, (ii) the random forest algorithm (RF) which constructs a number of decision tree models and uses bagging to sequentially improve the ensemble decision. To find the optimal hyperparameters in both models (namely the number of latent variables for PLS and the number of trees in RF) an internal five-fold cross-validation was employed in the training set.

3. Results and Discussion

3.1. Spectra signatures and pre-processing methods

Figure 2a illustrates the reflectance spectra of the grape spectra library. The continuous line is the mean spectrum, while the shaded area indicates the standard deviation. In addition, the histogram shown in Figure 2b demonstrates the distribution of the recorded °Brix values in our dataset for all the grape samples, with higher values indicating that the grape is ripe.

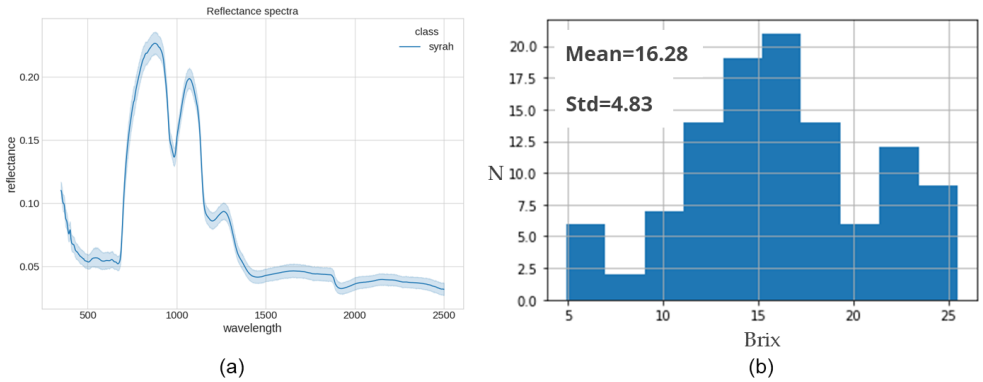

Figure 2. (a) Dataset reflectance spectra, (b) Distribution of the recorded Brix values for all samples

3.2. Comparison and evaluation of results

Table 1 below presents the results of the two machine learning algorithms based on three different metrics for the best pre-processing method according to each model. The metrics are coefficient of the determination (R^2), root mean square error (RMSE) and ratio of performance to interquartile distance (RPIQ). We can easily observe high accuracy for all methods with R^2 values being around 0.9, RMSE values to be at low levels approximately 1.5 degrees of °Brix and RPIQ values to be at high levels with values greater than 2; indicating that there is a good performance. The best result was achieved by the reflectance SG1 pre-process method with the RF model having the lowest prediction errors in the independent test set. In addition, based on the results of the best PLS model and using the Variable Importance in Projection (VIP), the region with the most information emerged in the range from 700 to 1300 nanometers.

Table 1. Accuracy results of the two machine learning models (PLS and RF) using various pre-processing methods in the independent test set

Pre-processing method	PLS			RF		
	R^2	RMSE	RPIQ	R^2	RMSE	RPIQ
Absorbances	0.83	1.98	3.38	0.70	2.59	2.57
Absorbances SG1	0.85	1.85	3.62	0.90	1.49	4.49
Absorbances SG1 SNV	**0.90**	**1.51**	**4.41**	0.89	**1.52**	4.38
Absorbances SNV DT	0.83	1.96	3.40	0.88	1.63	4.10
Absorbances SNV	0.88	1.64	4.07	0.86	1.79	3.74
Reflectances SG1	0.89	1.58	4.23	**0.92**	1.34	**4.96**
Reflectances SNV	0.86	1.78	3.75	0.85	1.83	3.66

4. Conclusions

Overall, by using a portable VNIR–SWIR spectrometer it is possible to estimate the wine grape maturity in-situ. The proposed methodology can support the wine producers in estimating the ripening grape parameters and making decision on harvest time. Also, some future steps are to examine more grape varieties (including white varieties) as well as the use of hyperspectral cameras.

Author Contributions: "Conceptualization, NS and NTS; methodology, NTS, NS, and NTZ; writing—original draft preparation, NS and NTS; writing—review and editing, NS, NTS, and NTZ; visualization, KK; supervision, EK; project administration, ZD, GZ and DB;"

Funding: The research leading to these results has received funding from the European Community's Framework Programme Horizon 2020 under grant agreement No 871704, project BACCHUS.

Acknowledgments: The authors would like to thank Ktima Gerovassiliou for providing the grape samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Peng, J.; Xie W.; Jiang J.; Zhao Z.; Zhou, F.; Liu, F., Fast Quantification of Honey Adulteration with Laser-Induced Breakdown Spectroscopy and Chemometric Methods, *Foods*. **2020**, 9, 341. https://doi.org/10.3390/foods9030341.
2. Karyotis, K.; Angelopoulou, T.; Tziolas, N.; Palaiologou, E.; Samarinas, N.; Zalidis, G. Evaluation of a Micro-Electro Mechanical Systems Spectral Sensor for Soil Properties Estimation, *Land*. **2021,** 10, 63. https://doi.org/10.3390/land10010063.
3. Giovenzana, V.; Beghi, R.; Malegori, C.; Civelli, R.; Guidetti, R. Wavelength Selection with a View to a Simplified Handheld Optical System to Estimate Grape Ripeness, *Am. J. Enol. Vitic*. **2014,** 65, 117–123. https://doi.org/10.5344/ajev.2013.13024.
4. Minasny, B.; McBratney, A.B. A conditioned Latin hypercube method for sampling in the presence of ancillary information, *Comput. Geosci.* **2006,** 32, 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009.
5. Ertlen, D.; Schwartz, D.; Trautmann, M.; Webster, R.; Brunet, D. Discriminating between organic matter in soil from grass and forest by near-infrared spectroscopy, *Eur. J. Soil Sci.* **2010,** 61, 207–216. https://doi.org/10.1111/j.1365-2389.2009.01219.x.
6. Vasques, G.M.; Grunwald, S.; Sickman, J.O. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra, *Geoderma*. **2008,** 146, 14–25. https://doi.org/10.1016/j.geoderma.2008.04.007.