Computers and Electronics in Agriculture 140 (2017) 244–254

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Comparison of different approaches for the prediction of sugar content in new vintages of whole Port wine grape berries using hyperspectral imaging

Véronique M. Gomes [a,*], Armando M. Fernandes [b], Arlete Faia [d,e], Pedro Melo-Pinto [a,c]

[a] CITAB-Centre for the Research and Technology of Agro-Environmental and Biological Sciences, Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
[b] INOV – INESC Inovação, Rua Alves Redol, 9, 1000-029 Lisboa, Portugal
[c] Departamento de Engenharias, Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
[d] BioISI – Biosystems & Integrative Sciences Institute, University of Lisboa, Faculty of Sciences, Campo Grande, Lisboa, Portugal
[e] Departamento de Genética e Biotecnologia, Escola das Ciências da Vida e Ambiente, Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal

ARTICLE INFO

Article history:
Received 1 August 2016
Received in revised form 13 April 2017
Accepted 10 June 2017

Keywords:
Prediction
Hyperspectral imaging
PLSR
Neural networks
Grapes berries

ABSTRACT

Two different approaches, PLS regression and neural networks, were compared for monitoring the quality of grapes using sugar content predictions based on hyperspectral imaging. The present work expands the result analysis and updates the state-of-the-art published in a conference article of the authors which concern the prediction of sugar content for vintages not used in model creation when the measured samples are composed of a small number of whole berries. This is highly innovative. The prediction models were established upon training under each approach and the generalization ability of both methodologies was determined through using n-fold-Cross-Validation and test sets. Sugar content was estimated using a model trained with spectra from samples of 2012. The test sets were composed of samples with six whole berries of 2012 or 2013 vintages.

The results for PLS regression and Neural Networks for a test set with 2012 samples, were 0.94 °Brix and 0.96 °Brix for the root mean square error (RMSE), and 0.93 and 0.92 for squared correlation coefficients (R^2), respectively, for each approach. When using test data containing 2013 samples, the RMSE values were 1.34 °Brix and 1.35 °Brix, and the R^2 values were 0.95 and 0.92. These errors are competitive with those of works from other authors executed under less demanding conditions. The results obtained suggest that when combining hyperspectral imaging with appropriate chemometric techniques or machine learning algorithms, it is possible to have a satisfactory generalization for vintages not employed in model creation.

1. Introduction

Over the last years, Portugal has become very competitive in producing wines. One of the most famous Portuguese wines with renowned quality is Port wine which is a unique fortified wine produced from grapes that grow in Douro appellation. In this work, we will focus on the development and comparison of two chemometric and machine learning algorithms models, namely, PLSR and neural networks, for sugar content prediction in samples composed by only six whole Port wine grape berries, using hyperspectral imaging data collected in reflectance mode. The major novelty is that the grapes from one of the vintages used for test were not employed in the model creation. This evaluation, rare in scientific literature, is extremely relevant since testing a model with the same vintage employed in the training model does not ensure that the model will be adequate when tested with future vintages. In addition, the determination of sugar content using a small number of berries per sample is also uncommon and is more difficult than using a large number of berries.

The relevance of the method developed for sugar content measurement comes from the need to maintain prominence in current markets by ensuring the high quality of the wines produced through the continuous improvement of the winemaking process. By measuring sugar content of grapes, which allows to evaluate the degree of ripening and is directly related with the alcoholic strength of the wine produced, it is possible to harvest wine grapes at the optimal point of maturity and to select them according to their quality making it possible to improve the quality of the grapes and resulting wines. Due to large differences in terroir,

* Corresponding author.
E-mail address: veroniquegomes@gmail.com (V.M. Gomes).

http://dx.doi.org/10.1016/j.compag.2017.06.009
0168-1699/

characterized by conditions such as sun exposition, water availability, soil quality and altitude, inside a vineyard, the grapes can be in various ripening stages and present different values of the sugar content. However, traditional analytical methods used in enological measurements of grapes, which are time consuming, costly and invasive cannot provide sugar content measurement at vineyard scale during the whole ripening process. Consequently, the use of cost-effective, non-destructive, and fast methods, such as hyperspectral imaging, which allow a large sampling to assess grape conditions throughout the vineyard, can become an important added value for the grape and wine industry.

Hyperspectral imaging integrates spectroscopy and digital imaging techniques (de Juan et al., 2009; Prats-Montalbán et al., 2011) to collect both spatial and spectral information that allows the development of models that estimate the enological parameters of grapes. This imaging method, in reflectance mode, collects information about the intensity of light reflected by objects as a function of their wavelength, measuring, simultaneously, thousands of points over a sample, without requiring contact between the spectrometer/camera and the grapes. Transmittance mode spectroscopy, in which light must pass through the sample, has also been used to measure enological parameters in grape (Larrain et al., 2008). It competes directly with reflectance mode spectroscopy but has the disadvantage of requiring contact with measured sample to avoid mixing the transmittance signal with the significantly more intense reflectance signal.

Given the large amount and complexity of the information collected by hyperspectral imaging, powerful data analysis tools are required. These tools can be chemometric (González-Caballero et al., 2011; Hernández-Hierro et al., 2013; Le Moigne et al., 2008; Nogales-Bueno et al., 2014) or machine learning methods (Dębska and Guzowska-Świder, 2011; Fernandes et al., 2011, 2015; Gomes et al., 2017; Vijayaraghavan et al., 2014) which have the ability of learning models from training samples composed of the measured spectra and the enological parameters measured. Once the training is finished they are able to generalize, i.e., they are able to estimate the enological parameters for new sets of samples. Several calibration models based on spectroscopy information have been employed in the last few years. Partial least squares (Arana et al., 2005; Diago et al., 2016), least squares support vector machines (Cao et al., 2010) and committee machines of neural networks (Fernandes et al., 2011, 2015; Gomes et al., 2017) were used to create models from spectra collected in reflectance mode for samples composed of a small number of whole grape berries. The comparison of PLSR and neural networks, in the present work, is relevant because both methods use different learning algorithms which result in different models, therefore it is impossible to know in advance which method will provide the best results. Two test sets from different harvesting year were employed to assess the created models generalization ability. So, in the present work, we propose to do an evaluation and comparison of the two developed methodologies in order to check if they are able to generalize and give correct predictions for samples from vintages that are and that are not included in the training model.

Using a small quantity of berries is important because some wineries want to select the best berries for producing high quality wines (Noguerol-Pato et al., 2012a,b,c; Tarter and Keuter, 2005). This would be possible with the technology described in the present article using already available destemmers that extract berries one-by-one from bunches (Goldfarb, 2008). After being extracted the berries would be placed over a conveyor belt that would pass underneath the hyperspectral camera (Fernandes et al., 2015). However, samples containing few berries will provide, for the enological parameter in question, a worse estimate of the average value of the distribution of all berries that can be collected than a sample containing a large number of berries. Consequently, there will be a larger variability of the enological values between samples when these contain a small number of berries which, in turn, makes it more difficult to create accurate prediction models (Fernandes et al., 2015; Gomes et al., 2017). In fact, works estimating sugar content in samples containing a small number of whole berries are still uncommon (Arana et al., 2005; Cao et al., 2010; Fernandes et al., 2015). More works can be found in literature (Fadock, 2011; Fadock et al., 2016; Nogales-Bueno et al., 2014) for spectroscopic measurements of sugar content in reflectance mode, but for a large number of grapes. Other works exist that use small number of whole berries per sample but that determine other enological parameters different from sugar content (Gomes et al., 2017; Martínez-Sandoval et al., 2016; Nogales-Bueno et al., 2015).

Using for test grapes from a vintage that was not employed in the model creation is rare in scientific literature; the authors have found two works where this is done (Fadock, 2011; Janik et al., 2007) besides two from their authorship (Gomes et al., 2017, 2014a). Janik et al. (2007) determines anthocyanin content in grape homogenates while we determine sugars in whole berries. Gomes et al. (2017) uses samples with six whole berries, as the present work does, but to analyze pH and anthocyanin content. On the other hand, Fadock (2011) analyses sugar content but uses a large number of berries per measured sample while the present work uses only six whole berries per sample. Gomes et al. (2014a) trains, using 2012 data, neural networks for sugar content determination in samples with six whole berries and tests with 2013 data which is a vintage not used for training. However, this test set contains less samples than those employed here. Gomes et al. (2014a) together with Gomes et al. (2014b) are two previous conference works from the authors that are prototypes of the present work. The latter, published as an abstract, compares neural networks and PLS but only for 2012 data.

In conclusion, up to the authors' knowledge, the present work is one of the first to test sugar prediction models with grapes from a vintage not used in the model creation, when the grapes are whole and in small quantity and, the first to check if PLSR provides good results under these conditions.

2. Materials and methods

2.1. Grape sampling

The analysis focuses on grapes of Touriga Franca, an important variety related to Port wine, harvested from the vineyards of Quinta do Bonfim in Pinhão, Portugal, which is property of Symington Family Estates, one of the world's largest producers of Port wine. In order to analyze grapes between the beginning of veraison (end of July) and maturity (end of September), a total of 240 bunches were collected in 2012, twenty-four per day; and a total of 84 grapes bunches were harvested in 2013, twelve per day. All bunches were from vine trees with small, medium and large vigor. For samples harvested in 2012 there were 210 samples available for training and validation and 30 samples for testing the models. All samples collected in 2013 were available for test. Six grape berries, randomly collected inside a single bunch, were used for each sample to measure the hyperspectral spectrum and to determine the analytical enological values. The berries were removed from bunches with their pedicel attached. After imaging, and before conventional enological analysis, all the samples were frozen at −18 °C.

2.2. Experimental setup

Hyperspectral data was collected using a hyperspectral imaging system (Fig. 1) containing the following components: a hyperspectral camera, composed of a JAI Pulnix (JAI, Yokohama, Japan) black