Postharvest Biology and Technology 140 (2018) 60–66

Contents lists available at ScienceDirect

Postharvest Biology and Technology

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

On-site images taken and processed to classify olives according to quality – The foundation of a high-grade olive oil

Enrique S. Pariente[a], John C. Cancilla[b], Kacper Wierzchos[c], José S. Torrecilla[a,*]

[a] Departamento de Ingeniería Química, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040, Madrid, Spain
[b] Scintillon Institute, San Diego, CA, USA
[c] Department of Physics, University of South Florida, Tampa, FL, USA

ARTICLE INFO

Keywords:
CCD Olive images
Image processing
Quality estimation
Neural networks

ABSTRACT

It is well known that the quality of the olives used to produce olive oil is directly related to the quality of this product. In this research, the classification of olives according to their quality grade has been achieved by combining image processing and mathematical modeling. The images of 190 different olives of four varietals (Ascolano, Manzanillo, Mission, and Sevillano), which were categorized into four quality-based groups (optimal, acceptable, borderline, and unacceptable), were employed to design a set of linear (partial least squares) and non-linear (artificial neural networks) mathematical models. Three different binary classifications were attempted: (**a**) optimal/others, (**b**) optimal and acceptable/others, and (**c**) unacceptable/others. The results obtained for (**a**) were a perfect classification rate using either one of the modeling approaches. On the other hand, (**b**) and (**c**) were clearly favored by the use of neural networks, which reveals a more complex task (images are more similar). In these two classifications, the linear models offered 75% and 70% correct classification rates, whereas the non-linear ones provided 93% and 90%, using comparable validation procedures. In addition, an external validation has been done using photographs taken directly from the olive grove. The misclassification percentage is less than 13%. Therefore, proper data extraction from olive images combined with neural networks can lead to accurate tools for the location of olives of a specific quality grade, which can be of great use for olive oil producers that seek a specific quality in their raw materials and, consequently, in their final product.

1. Introduction

In the past, many applications based on image processing have been developed within the food technology field, as well as the techniques to extract complementary information from different types of images for further modeling purposes (Brosnan and Sun, 2004; Du and Sun, 2004). For instance, the quality of lamb meat has been assessed using near infrared hyperspectral imaging (Kamruzzaman et al., 2012), thermal images have been used to locate bruised tomatoes (Van Linden et al., 2003), or a technique that uses images from two perpendicularly positioned cameras has been employed to determine mass and volume of various citrus fruits (Omid et al., 2010).

A different fruit that has been studied through image-based approaches are olives, which have been classified according to external damage (Riquelme et al., 2008) or analyzed to predict their olive oil content (Ram et al., 2010). In the current research, information has been extracted from olive photographs to reach a fast method to separate olives according to their quality. Achieving a successful system in this regard would be of great use for the extra virgin olive oil (EVOO) production sector, as high-quality olives are directly related to the production of high-quality oils.

Olive oil is known for playing a key role in the popular and healthy Mediterranean diet (Visioli et al., 2002), as it is the main source of dietary lipids and can be used for cooking as well as for its raw consumption (Bach-Faig et al., 2011). Research has shown that a regular intake of olive oil is beneficial for various reasons such as leading to a lower incidence of cardiovascular diseases (Estruch et al., 2013) or favoring a healthier cholesterol metabolism (less concentration of low-density lipoproteins) (Kris-Etherton et al., 2002). Due to its complex composition, rich in antioxidants, EVOO possesses a high nutritional value which is directly correlated with the quality of the oil and its beneficial effects towards human health (Bach-Faig et al., 2011). For these reasons, the implementation of a fast system that would help producers reach higher quality EVOOs would have a relevant impact on the sector. As already mentioned, this is precisely the goal of the present research, as data has been gathered from multiple photographs of olives which possess different qualities to design and train mathematical classifiers to distinguish olives according to their quality, enabling the

* Corresponding author.
E-mail address: jstorre@ucm.es (J.S. Torrecilla).

https://doi.org/10.1016/j.postharvbio.2018.02.012
Received 29 December 2017; Received in revised form 22 February 2018; Accepted 23 February 2018
0925-5214/

selection of only the olives that meet the producers' requirements.

In essence, images such as photographs or video frames can be broken down into numbers, and, in turn, mathematically processed. Therefore, if adequately interpreted and processed, the data contained in images can be used as variables in classifying mathematical models to distinguish characteristic groups, which, in this case, are olives with different qualities. The main modeling and classifying tools employed were non-linear artificial neural networks, which have shown their value in other image processing applications (Mohebbi et al., 2009; Sujana et al., 1996). The statistical performance of these intelligent algorithms was compared with the results offered by classic partial least squares discriminant analysis algorithms (Aroca-Santos et al., 2015). Both types of models were fed with data that is easy and fast to gather from the olive images, leading to systems that can potentially be used by the producers at the olive mill for real-time selection of the desired olives.

2. Materials and methods

In this section, the olive images processed and analyzed, as well as the mathematical tools and procedure followed to reach the olive classification will be described.

2.1. Images and data

The images that have been used during the present research were gathered from a book (Sibbett and Ferguson, 2005) where 190 photographs of four different olive varietals were found (Ascolano (40), Manzanillo (50), Mission (60), and Sevillano (40)). They appeared catalogued according to their quality into four groups which were optimal, acceptable, borderline, and unacceptable (some examples of the images can be seen in Fig. 1). These qualities will be used during the modeling phase of this work as dependent variables or, in other words, as labels to classify the olives according to their quality, regardless of their varietal. In addition, another set of photographs of olives, which were taken by a Nikon Coolpix L11 6MP Digital Camera directly from an olive grove in the southern part of the Community of Madrid (Spain), between the months of October and November (2017), were used to externally validate one of the models developed (Fig. 2). The external validation database is fully composed by images taken before and right after the harvest of the olives. In particular, the external validation dataset is formed by 45 photos of olives (12 of them were taken directly from a tree branch). These olives mainly belong to the Cornicabra and Castellana varietals in different stages of maturation. The olive varietals used in the external validation are commonly employed for extra virgin olive oil production and they present a high olive oil rate (18–20%) in comparison with other Spanish varietals.

The photographs from the book and the olive grove were digitalized or converted into mathematical matrices, where each pixel of the images became a set of numbers. This conversion enabled the mathematical analysis of the olive images. Each image was therefore turned into three matrices which represent color channels (red, green, and blue; RGB), each possessing values which range from 0 to 255, which in the end will be used as independent variables by the designed classifying models. Nevertheless, prior to modeling or extracting these variables, the images have to be properly treated.

2.2. Image processing

In all cases, and in order to extract information from an object (olive) present in an image, it is required to previously identify it and separate or distinguish it from the rest of the image, which may be a simple background or even other interfering images. To do this, different tools present in the Image Processing Toolbox of the MatLab 2014b software package were employed. This led to the isolation of the olives within each image, enabling the accurate and reliable extraction of the above mentioned RGB variables, which solely contained information regarding the actual olives. Following this, the mean values of each color channel were calculated leading to three variables that characterized each image (mean R, mean G, and mean B). These were used during the design and optimization of the classifying models.

2.3. Classifying models

The true goal of the present research is to reach a mathematical model that is rapidly capable of classifying olives according to their

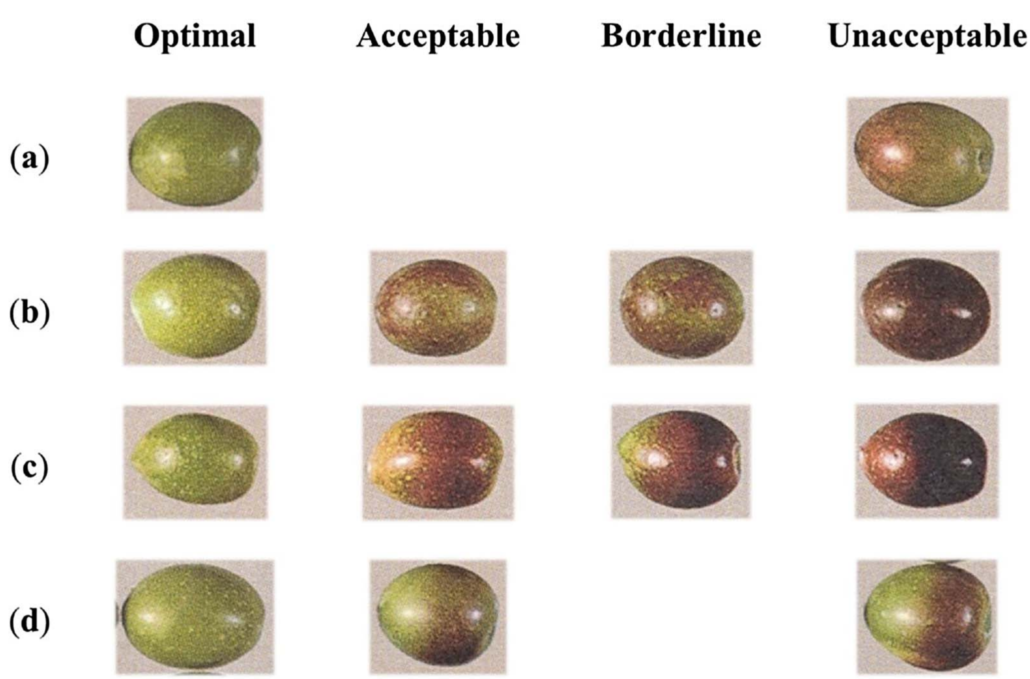

Fig. 1. Photographs of some of the olives that have been used during the current research. A single example of each quality grade and varietal combination that was available is shown. (**a**) Ascolano, (**b**) Manzanillo, (**c**) Mission, and (**d**) Sevillano. The photographs were obtained from a book by Sibbett and Ferguson (2005).