Postharvest Biology and Technology 120 (2016) 204–212

Contents lists available at ScienceDirect

Postharvest Biology and Technology

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

Near-infrared spectroscopy for detection of hailstorm damage on olive fruit

Roberto Moscetti[a], Ron P. Haff[b], Danilo Monarca[c], Massimo Cecchini[c], Riccardo Massantini[a,*]

[a] Department for Innovation in Biological, Agro-food and Forest system, Tuscia University, Via S. Camillo de Lellis snc, 01100 Viterbo, Italy
[b] United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center, 800 Buchanan St., Albany, CA 94710, United States
[c] Department of science and technology for Agriculture, Forest, Nature and Energy, Tuscia University, Via S. Camillo de Lellis snc, 01100 Viterbo, Italy

ARTICLE INFO

Article history:
Received 26 January 2016
Received in revised form 8 June 2016
Accepted 13 June 2016
Available online 2 July 2016

Keywords:
Olea europaea L.
Impact damage
Bruise
Acousto-Optic Tunable Filter-Near Infrared spectroscopy
Discriminant analysis
Wavelengths selection

ABSTRACT

A rapid, robust, and economical method to detect hailstorm-damaged olive fruit (*Olea europaea* L.) would benefit both consumers and producers of olives and olive oil. Here, the feasibility of using Near-Infrared (NIR) spectroscopy for olive fruit sorting (cv. Canino) into hailstorm-damaged and undamaged classes is demonstrated. Features selected from the entire spectra by the genetic algorithm (two to six features per model) were input to Linear Discriminant Analysis, Quadratic Discriminant Analysis and k-Nearest Neighbor routines to develop models to classify olive fruit. Spectral pretreatment and feature selection were optimized through an iterative routine developed in R statistical software. Each model was evaluated based on false positive (α-error), false negative (β-error) and total error rates. The most accurate models yielded total error rates of less than five percent. The optimal features corresponded to *R*[1320 nm], *R*[~1460 nm], *R*[~1650 nm], *R*[~1920 nm], *R*[~2080 nm], *R*[~2200 nm] and *R*[~2220 nm], where *R*[x] represents the reflectance of light from the sample at a wavelength of x nm. The results indicate that single-point NIR spectroscopy is a feasible basis for hailstorm damage detection in olive fruit with the potential to allow on-line implementation on milling production lines.

1. Introduction

Olea europaea (L.) is widely cultivated in the Mediterranean basin and its oil is a key component of the Mediterranean diet. The processed olive is either sold as table olives or further processed into oil. Worldwide production of olive oil increased from 1.5 million metric tons in 1990 to roughly three million metric tons in 2015, with 90% coming from Spain, Italy, or Greece (IOC, 2016). The unique chemical composition of virgin olive oil positively impacts sensory, nutritional and health properties as compared to the oil of other vegetable crops. These beneficial properties have been attributed to a fatty acid composition rich in monounsaturated acids such as oleic acid (Visioli and Galli, 2002), and to certain minor constituents such as phenolic antioxidants, including some derivative secoiridoids. Consumption of extra-virgin olive oil has been identified as a contributing factor to reduced rates of atherosclerosis, oxidative-stress associated diseases, autoimmune diseases, as well as some forms of cancer and diabetes (Katan et al., 1995; Tripoli et al., 2005; Bendini et al., 2007). Since olive fruit quality is directly correlated with the chemical composition and overall quality of produced oil, methods for removing defective fruit would benefit producers and consumers alike.

The quality of olive oil depends mainly on the quality of the fruit from which it is derived. Thus, reducing the incidence of external damage is a priority for the agri-food industry (Jiménez et al., 2016). Susceptibility to external damage such as bruising, scarring and impact damage depends on both internal and external factors of the fruit, such as cultivar, texture and firmness, maturity, dry matter, temperature, size and shape, as well as cell wall status, cell shape, etc. (Van linden et al., 2006; Jimenez-Jimenez et al., 2012). When cellular tissue is ruptured, the release of intracellular fluid causes enzymatic reactions resulting in oxidative and hydrolytic degradation of oil and phenolic fractions (Segovia-Bravo et al., 2009; Jiménez-Jiménez et al., 2013; Moscetti et al., 2015; Ramirez et al., 2015; Rojnić et al., 2015). The subsequent browning is visible within a few hours and increases over time (Ben-Shalom et al., 1978). The chief enzyme that leads to browning, Polyphenol oxidase (PPO), catalyzes the oxidation of both hydroxytyrosol and

* Corresponding author.
E-mail address: massanti@unitus.it (R. Massantini).

http://dx.doi.org/10.1016/j.postharvbio.2016.06.011
0925-5214/

verbascoside to o-quinones, which condense to form brown polymers. Chemical oxidation of hydroxytyrosol may occur simultaneously (Martinez and Whitaker, 1995; Segovia-Bravo et al., 2011). Unless the damaged fruit is detected and removed the result is reduced quality in the produced oil. Thus, human inspection and classification of olive fruit according to quality regulations as defined by the European Commission Regulation No. 299/2013 (European Commission, 2013) is a fundamental tool to ensure the quality and economic value of olive oil. However, human inspection is labor intensive and highly subjective, and thus subject to high classification error rates. Thus, classification based on machine vision has great potential to improve quality (Guzmán et al., 2013), reduce labor expense, and minimize the impact of extreme weather events such as hailstorm. Botzen et al. (2010) estimate hailstorm frequency could increase by 25–50% by 2050 under standard climate change scenarios, with economic implications due to increased fruit damage. Most European Countries already experience severe hailstorms, especially in northern Italy where they are most frequent, although central and southern Italian regions are also affected (Baldi et al., 2014). Thus, as climate change progresses there will be a corresponding need for research to offset the damaging economic impacts on agriculture (Schiermeier, 2015).

Technological advances in visible/near-infrared (vis/NIR) spectroscopy and hyperspectral imaging have provided the means for real-time on-line quality assessment of both olive fruit and oil. A substantial amount of research has been reported addressing the assessment of olive fruit defects, such as bruising, impact damage, insect infestation, etc. (Stella et al., 2015a, 2015b). Although ongoing advances in computing speed and data transfer rates have enhanced the speed and accuracy of spectroscopic devices, data handling continues to be one of the more serious technological challenges to practical real-time applications (Wu and Sun, 2013). This is particularly true for hyperspectral imaging, where the "curse of dimensionality" results in a high computational load (Burger and Gowen, 2011). Currently, the acquisition of both spectral and spatial information is time-consuming and, because hyperspectral imaging is an indirect method of analysis, standardized calibration and model transfer procedures are required. Thus, commercially available hyperspectral imaging systems remain impractical for on-line implementation. However, systems customized for specific applications based on a limited number of wavebands (multi-spectral imaging) in combination with data reduction schemes (ElMasry and Sun, 2010) have the potential to meet the speed requirements for an olive-oil production line. Overviews of chemometric methods used for data handling of vis/NIR spectroscopic datasets are available (Nicolaï et al., 2007; Burger and Gowen, 2011). Thus, conventional vis/NIR spectroscopic systems are currently the most practical spectroscopic technology available for rapid on-line applications.

The objective of this research was to demonstrate the use of single-ray NIR spectroscopy for hailstorm damage detection on olive fruit. The expected impact is a new tool to help the olive industry maintain quality today and in the future with the expected increase in the frequency of hailstorm and other extreme weather events.

2. Materials and methods

2.1. Sample preparation

Approximately 1.2 kg of olives (*O. europaea* L., cv. Canino) was manually harvested on a local farm in Central Italy at the beginning of November 2014. The harvest was conducted 13 days after a severe storm with the following characteristics: Torro code H0-H3, size code 2 and hail diameter between 11 and 15 mm (Baldi et al., 2014). The storm caused significant damage to fruit and vegetation as well as production loss. In order to establish a homogeneous sample set, only mature fruits of the appropriate green color were collected. Detection and identification of hailstorm damage was visually performed by consulting experts from local olive oil mills.

The use of fruit that were naturally damaged by hailstorm avoided the need for lab-based simulation of fruit damage, which would lead to less robust and transferrable classification models. Immediately after harvesting, the samples were taken to the laboratory in appropriate thermal boxes. From the original 1.2 kg, 744 olives were selected that were free from decay. The fruit were kept at $25 \pm 0.5\,^{\circ}C$ for 24 h to allow for temperature and moisture equilibrium prior to NIR spectra acquisitions.

2.2. NIR spectral acquisition

Olive spectra were acquired using a Luminar 5030 Acousto-Optic Tunable Filter-Near Infrared (AOTF-NIR) Miniature 'Hand-held' Analyzer (Brimrose Corp., Baltimore, USA). The instrument was equipped with a reflectance post-dispersive optical configuration, a pre-aligned dual beam lamp assembly and an indium gallium arsenide (InGaAs) array (range 1100–2300 nm, 2 nm resolution) with an integrating time of 60 ms. Each sample was measured in triplicate along the fruit equatorial line, and the average spectrum was used for further analysis. The reference spectrum was automatically measured by the instrument as described by Cayuela and Weiland (2010). Diffuse reflectance spectra were acquired by the SNAP! 2.04 software (Brimrose Corp.). Before the spectral acquisition, olives were visually evaluated for presence or absence of hailstorm damage, thus assigning each spectrum into damaged (Unsound) or not-damaged (Sound) classes.

Each olive was modeled as a 'data vector', where the spectral reflectance values (otherwise called features) were vector components. Fifty percent of the samples were randomly assigned to the calibration set (372 fruits) and 50% were split into three prediction sets (124 fruits each). Features for use in classification were extracted from the whole spectra of the calibration set. Features were extracted following spectral pretreatments including Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), and Savitzky-Golay first and second derivatives (*D1f* and *D2f*, respectively) with a second or third order polynomial fitted over a window of five (S5), nine (S9) or thirteen (S13) variables (Savitzky and Golay, 1964; Boysworth and Booksh, 2008). For each dataset, Mean Centering (MC) was also tested (Fig. 1). Every possible combination of preprocessing was also tested and only the best results, in terms of model performance, were retained.

2.3. Chemometrics

Because of high correlation among spectral data, a Genetic Algorithm (GA) was used to select features for input to classification algorithms, with the goal of selecting a series of wavebands which could describe the correlation between the predictor variables and the response variables (Xing et al., 2008). The GA selects a small subset of spectral bands with biological or biochemical importance, which are representative of the entire spectral dataset. In this study, GA was used to seek n-feature subsets (where n ranged from 2 to 6) which are optimal surrogates for the whole dataset (Cerdeira et al., 2013). A maximum of 6 features was chosen to minimize overfitting.

Sets of features selected by the GA were input into three different classification algorithms: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and k-Nearest Neighbors (kNN). LDA and QDA classification procedures were both based on Bayes' rule, while the kNN algorithm used Euclidean