Chemometrics and Intelligent Laboratory Systems 156 (2016) 241–248

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

VIS/NIR hyperspectral imaging and N-way PLS-DA models for detection of decay lesions in citrus fruits

A. Folch-Fortuny [a], J.M. Prats-Montalbán [a], S. Cubero [b], J. Blasco [b], A. Ferrer [a]

[a] Multivariate Statistical Engineering (GIEM), Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Camino de Vera s/n, Edificio 7A, 46022 Valencia, Spain

[b] Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), Cra. Moncada-Náquera, km 5, 46113 Moncada, Valencia, Spain

ARTICLE INFO

Article history:
Received 16 February 2016
Received in revised form 26 April 2016
Accepted 10 May 2016
Available online 12 May 2016

Keywords:
Hyperspectral imaging
NIR
NPLS-DA
Variable selection
Permutation test

ABSTRACT

In this work an N-way partial least squares regression discriminant analysis (NPLS-DA) methodology is developed to detect symptoms of disease caused by *Penicillium digitatum* in citrus fruits (green mould) using visible/near infrared (VIS/NIR) hyperspectral images. To build the discriminant model a set of oranges and mandarins was infected by the fungus and another set was infiltrated just with water for control purposes. A double cross-validation strategy is used to validate the discriminant models. Finally, permutation testing is used to select a few bands offering the best correct classification rates in the validation set. The discriminant models developed here can be potentially implemented in a fruit packinghouse to detect infected citrus fruits at their arrival from the field with affordable multispectral (3–5 channels) cameras installed in the packinglines.

1. Introduction

Citrus production exceeded 115 million tons in 2011 [1]. They are cultivated in over one hundred countries worldwide, being Spain as one of the most important producer countries and the world leader in fresh citrus exports [1]. Citrus are, indeed, the most widely produced fruits for human consumption, especially oranges (62%) and mandarins (23%). To ensure product quality and reduce production losses, it is mandatory to enhance postharvest handling in citrus packinghouses. Many issues arise in this process due to pathological diseases in fruits. This problem can be potentially harmful, since a small set of rotten and sporulated fruits can contaminate the whole batch, especially during storage or transport. *Penicillium digitatum* (the cause of green mould) and *Penicillium italicum* (the cause of blue mould) are two examples of the most deleterious fungi causing fruit decay, and they affect several cultivars over the world [2,3].

Green mould lesions at early stages cannot be detected with the naked eye because the appearance of the damage is very similar to the appearance of sound fruit. The first symptoms of this disease appear as a slightly discoloured soft, water soaked around a point of injury. The spot expands rapidly to a 30–40 mm diameter. As the infection advances, a white fungal growth appears on the surface of the rot [4]. Before the sporulation, the appearance of the lesions is very similar to the sound skin, being difficult for the workers to detect damaged fruit, especially when they work on an inspection table, examining fruit travelling at high speed. Therefore, the application of visual inspection or computer vision systems based on colour images is limited. Nowadays, novel machine vision technologies are being incorporated in the citrus postharvest to detect this dangerous disease, mostly based on ultraviolet (UV) induced fluorescence. Ogawa et al. [5] presented a system to detect decay lesions in citrus using fluorescence images, and Blanc et al. [6] patented an automatic machine for in-line decay detection and fruit sorting using UV illumination. However, Momin et al. [7] demonstrated that different cultivars of citrus fruits have different excitation wavelengths to produce UV induced fluorescence in the infected areas, which makes it difficult to create a system valid for all cases based only on this technology. Also, this kind of automatic detection can be potentially jeopardized by fluorescence measurements from other non-related defects [8]. Alternatively, this disease can often be observed using other techniques like image backscattering [9,10] or hyperspectral imaging (HSI) [11]. In this sense, different hyperspectral sensors are being investigated to detect non-visible fruit damage [12] like decay lesions in citrus fruits [13].

Using spectral devices, a set of images is obtained at different wavelengths, capturing a huge amount of chemical information. Some works have been focused on reducing the redundant information in this procedure, compressing the high-dimensional original variable space into a low-dimensional one that preserves the main properties of the data. Gomez-Sanchis et al. [14] and Lorente et al. [15] used the features from spectral images of infected fruit as inputs for classification algorithms, in order to improve the discrimination between sound and symptomatic skin. In addition, HSI systems have also been developed to detect other dangerous diseases. Qin et al. [16] used a portable imaging spectrograph to acquire hyperspectral images of red grapefruits affected by canker and other defects. In that work, the spectral images of the different defects were analysed using principal component

http://dx.doi.org/10.1016/j.chemolab.2016.05.005

analysis (PCA) [17] and spectral information divergence as classification method, detecting 97.6% of infected fruits. In Qin et al. [18], the authors exploited the bands selected using PCA and correlation analysis to obtain a system capable of detecting the canker using ratios of two bands. Afterwards, a system to detect canker lesions in-line was developed by Qin et al. [19]. Also, PCA and band ratios were used by Li et al. [20,21] to select relevant bands for the detection of this disease among other common defects.

Multivariate Image Analysis (MIA) uses a wide number of models and approaches to deal with hyperspectral images [22,23]. PCA is probably the most used method within MIA (some examples are shown in the previous paragraph), but other two-way methods are commonly used, as partial least squares regression (PLS) [24] or multivariate curve resolution (MCR) [25]. In some cases, it is convenient or interesting to use three-way models such as N-way PLS (NPLS) [26] or Tucker [27].

The aim of this work is to develop multivariate models based on hyperspectral images able of discriminating between infected and sound citrus fruits while at the same time reducing as much as possible the number of wavelengths used. For this, we will use NPLS discriminant analysis (NPLS-DA) [28,29] to build a latent variable-based regression model using specific features extracted from a pool of images of different orange and mandarin cultivars collected at the Instituto Valenciano de Investigaciones Agrarias (IVIA) (Valencia, Spain). This kind of models has been successfully applied in many research works within fruit industry, *e.g.* for tomato [30], coffee [31], loquats [32] and apple [33] discrimination. In our case, the present study represents an attempt to implementing automatic classification procedures in fruit packinghouses to prevent the storage of infected citrus fruits, which may ultimately rot and sporulate causing contamination of packinghouse facilities and spread of the disease to healthy stored fruit.

The structure of the paper is as follows. Section 2 gives specific details on the data and the image acquisition. In Section 3 the data preprocessing, feature extraction and latent variable modelling are described. Section 4 shows the results of the multivariate discriminant modelling. Finally, some conclusions are drawn on Section 5.

2. Materials

Eight different orange and mandarin varieties are analysed in this paper: Clementine, Navel Lane Late, Mioro, Nadorcott, Nova, Salustiana, Blood orange, and Washington Navel. In each variety, 150 fruits were harvested from the field collection of the Citrus Germplasm Bank at the IVIA (Spain) [34]. After two days of storage with controlled temperature and humidity, 100 fruits of each variety were inoculated with a concentration of 10^6 spores/ml of *P. digitatum* [35]. These citrus fruits represent the fungus group. The remaining 50 fruits were inoculated with water, and they represent the control group to know if the inoculation process influences the results. Both inoculations were produced around 2 days after the fruit collection.

Between 1 and 4 days after inoculation, when the fruit started to show slight external symptons of decay, a camera coupled with a visible/near infrared (VIS/NIR) liquid crystal tunable filter (LCTF) was used at IVIA to obtain a RGB and hyperspectral images from each fruit of each variety. Fig. 1 shows the RGB images of a control and an infected mandarin, in order to illustrate how difficult is to discriminate between both classes with visual inspection. 44 wavelengths were registered from 650 to 1080 nm with a resolution of 10 nm. Each image has 1040 times 1392 pixels per wavelength. Therefore, the hyperspectral images can be represented as $1040 \times 1392 \times 44$ datacubes.

3. Methods

3.1. Data preprocessing

The citrus fruits appear centred in the images (see Fig. 2, first block of images). The spherical shape of the fruits causes some undesirable effects in the fruit images, one of the most important being that the pixels in the borders (pale blue areas around the fruit in Fig. 2) appear darker than those in the centre of the fruit due to the reflexion laws of the light. Therefore, it is convenient to remove the pixels near the border from the analysis, which was done in this experiment by applying a mask. After defining an intensity threshold, pixels exceeding this limit were selected, representing the inner area of the fruit (see Fig. 2, second block of images). The pixel selection is performed at each wavelength of the image. Then, the joint area across all wavelengths is defined as the mask for the whole image. This way, if a pixel is above the threshold for, at least, one wavelength, it is guaranteed that it is included in the fruit mask. This procedure is repeated for all fruits in each variety.

Five different data preprocessings were applied in this study. The first one consisted of analysing the images using the original intensities (no preprocessing), *i*, measured with the VIS/NIR-LCTF system. The second one consisted of transforming the intensities into reflectance values, *r*, using black (*b*) and white (*w*) references taken with the HSI system:

$$r = 100 \times \frac{i - i_b}{i_w - i_b}. \tag{1}$$

The third preprocessing consisted of obtaining the absorbance values, *a*, from the reflectance. That is:

$$a = log_{10}\left(\frac{r}{100}\right). \tag{2}$$

The fourth and fifth preprocessings consisted of applying multiplicative scatter correction (MSC) and standard normal variate (SNV)

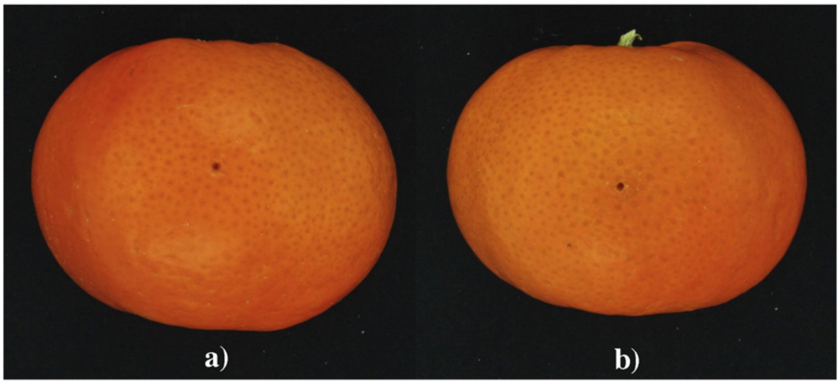

Fig. 1. RGB images of a control (a) and an infected (b) mandarin.