Computers and Electronics in Agriculture 123 (2016) 297–303

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Peach variety identification using near-infrared diffuse reflectance spectroscopy

Wenchuan Guo *, Jingsi Gu, Dayang Liu, Liang Shang

College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China

ARTICLE INFO

Article history:
Received 4 July 2015
Received in revised form 5 March 2016
Accepted 6 March 2016
Available online 16 March 2016

Keywords:
Peach
Near-infrared spectroscopy
Variety identification
PCA
LSSVM
ELM

ABSTRACT

More than 1000 peach varieties with significant differences in qualities are cultivated in China. Distinguishing peach varieties is not only needed by peach sellers, but also demanded by consumers. To offer information on identifying peach varieties, near-infrared (NIR) diffuse reflectance spectra between 833 and 2500 nm were collected for four peach varieties, 100 samples for each variety. Kennard–Stone algorithm method was used to divide all samples into calibration set (320 peaches) and prediction set (80 peaches). Eight principal components (PCs), 1067 and 10 characteristic wavelengths were extracted by principal component analysis (PCA), uninformative variable elimination based on partial least squares (UVE-PLS) and successive projections algorithm (SPA) from full spectra (FS) with 2074 initial wavelengths, respectively. Least squares support vector machine (LSSVM) and extreme learning machine (ELM) were used to establish peach varieties identification models using the FS, selected PCs and characteristic wavelengths as input variables. Experimental results showed that all models based on PCA reached 100% accuracy for identifying the four peach varieties. The accuracy of LSSVM models based on UVE-PLS also reached 100%. This study indicated that peach varieties could be distinguished successfully by using NIR spectroscopy.

1. Introduction

Peaches which originated in China have been cultivated for more than 3000 years (Capitani et al., 2013). Because of its tasteful, sweet, and juicy flesh, peach has become the third most important fruit in the world (Davidovic et al., 2013). China has been the largest peach-planted area and production country since 1993. In 2010, Chinese peach cultivation area and production reached 719.4 km^2 and 1.04 million tons, respectively (Lü et al., 2012). Nowadays, more than 1000 peach varieties with significant difference in quality are cultivated in China (Chen et al., 2011). Most of the peaches belong to mid-ripening varieties. It means that most of the peaches are mature at almost same time, usually from the middle ten days of July to the first ten days of August in North China. Generally, more than one peach variety is planted in a peach orchard, and several varieties are sold by sellers at the same time. Therefore, different peach varieties are easily mixed during harvesting and selling. However, each variety has special taste, which causes different price and being appreciated by different people. Therefore, distinguishing peach varieties is not only needed by peach sellers, but also demanded by consumers.

Near-infrared (NIR) spectroscopy outperformed classical physical and chemical methods with the advantages of fast analytical speed, easy operation, and nondestructive measurements in detecting qualities of agricultural products (Shao et al., 2011). It has been used to determine the soluble solids content (SSC) and firmness of fruits, such as pear (Li et al., 2013), strawberry (Sanchez et al., 2012), kiwifruit (Lee et al., 2012), grape (Gonzalez-Caballero et al., 2011) and melon (Greensill et al., 2001). In addition, NIR spectroscopy has also been used to nondestructively discriminate varieties of agricultural products, such as, red and pink tomatoes using partial least squares (PLS) regression with distinguishability of 96.85% (Sirisomboon et al., 2012), Thai orange varieties using logistic regression with 100% classification accuracy (Suphamitmongkol et al., 2013), Chinese bayberries using BP artificial neural network (BP-ANN) with 95% accuracy (Li et al., 2007), yogurt varieties applying BP-ANN with 100% recognition rate (He et al., 2006), coffees varieties with 100% accuracy (Esteban-Diez et al., 2004), and coffee beans varieties using PLS with misclassification errors in the range of 5–10% (Myles et al., 2006). These studies shows that the varieties could be identified with excellent discrimination rates.

* Corresponding author. Tel.: +86 29 87092476; fax: +86 29 87091737.
E-mail address: guowenchuan69@126.com (W. Guo).

http://dx.doi.org/10.1016/j.compag.2016.03.005
0168-1699/

The internal quality indices of peaches, such as SSC, pH, firmness and flesh color, and woolly peaches, have also been successfully predicted using NIR spectroscopic techniques (Lin and Ying, 2009; Ortiz et al., 2001; Shao et al., 2011; Slaughter et al., 2013; Takano et al., 2007). If NIR spectroscopy could also be used to identify peach varieties, peach variety and its internal qualities could be predicted using one spectrum. However, to our knowledge, there is limited information on peach variety identification. Therefore, the main objective of the work is to explore the feasibility of NIR spectroscopy in peach variety identification. For this purpose, NIR diffuse reflectance spectra were collected for four peach varieties, three characteristic variable selection methods were applied to reduce the data dimension, and two modeling methods were used to establish peach variety identification models. The influence of full spectra and selected characteristic wavelengths on model precision was compared, and the best models were put forward.

2. Materials and methods

2.1. Materials

Three mid-ripening varieties (Shahong, Beijing 8 and Hongmi) and a late-ripening variety (Laishanmi) were chosen as samples in the study. The four varieties are widely planted in Shaanxi Province, even in North China. They resemble each other in rose color of skin and round shape, and have close or same harvesting periods. The flesh color of 'Hongmi' peach is light rose, but the flesh color of other three varieties is milk-white. Ninety percent mature peaches, which are regarded as commercial ripeness stage, were used as samples. The features of 90% mature peaches are: green has faded away, true background color has presented, and the special texture of each variety has exhibited (SB/T, 1992). The peach samples of 'Shahong', 'Beijing 8', 'Hongmi' and 'Laishanmi' were randomly picked from 6- to 8-year-old peach tress in two local orchards, belongs to different owners, in Yangling, Xi'an, Shaanxi Province, P.R. China on July 9, July 18, August 8, and August 20, 2012, respectively. The two orchards are located at 34°19′ north latitude and 108°02′ east longitude. They were irrigated, fertilized and sprayed by their owners guided by peach fruit experts. After the samples were transported to laboratory, the peaches were kept in airtight polyethylene bags and stored at 4 °C in refrigerators. Before experiment, peaches without defects were taken away from refrigerators, washed with tap water to remove foreign materials on surface, wiped dry and kept at room temperature (24 ± 2 °C) to warm for about 6 h. During the next 3 days after the peaches of each variety were picked, measurements were taken on 30–35 peaches everyday. One hundred peaches from each variety were measured. Totally, 400 peaches were used in the work.

2.2. Spectral acquisition

The NIR spectra were collected in diffuse reflectance mode using a Fourier transform NIR spectrometer (Model MP.0331.04, BRUKER, Germany) equipped with a NIR optical fiber probe, an interferometer, a detector, a wide-band quartz halogen light source (50 W), and software OPUS (Version 5.5, BRUKER, Germany). The NIR diffuse reflectance spectra were collected from 833 nm to 2500 nm with an interval of 0.804 nm, which resulted in 2074 variables. Before collecting the spectra, the NIR spectrometer was turned on and allowed to warm up for at least 1 h. Then optical fiber background spectra were collected to avoid any spectral variation due to unknown factors. When collecting the spectra, keep the NIR optical fiber probe contact with the surface of peach closely to avoid surface reflectance and air interference. Two spectra were obtained at two points, which were located at the center parts of blush side and non-blush side, respectively, around the fruit equator by 180° interval for each peach. The average of the two spectra represented the spectrum data of each peach, and was used in further analysis. All spectra were collected at room temperature (24 ± 2 °C) by using OPUS software.

2.3. Measurement on peach mass and quality parameters

The mass of each peach was measured using an electronic balance with the precision of 0.1 g. The firmness of peach pulp was measured with a GY-3 fruit penetrometer (Sundoo Instruments, Zhejiang, China) with an 8-mm-diameter penetrometer tip. Two readings at two points, where the NIR spectra were collected, were obtained for each fruit. The soluble solids content was determined with an Atago Pallete Series Model PR101α digital refractometer (Atago Co. Ltd., Tokyo, Japan). The pulps near the points where firmness were measured were used to press juice for SSC measurement. Two SSC readings at each point were recorded. The means of two readings for firmness and four readings for SSC on each sample were calculated. The mass was measured before, and the firmness and SSC were obtained after NIR spectral acquisition, respectively. All measurements were undertaken at room temperature (22 °C ± 2 °C).

2.4. Spectral pretreatment

Smoothing and baseline correction are fundamental pretreatment methods to the experimental data to reduce noise and to correct baseline variations. In the study, Savitzky–Golay smoothing was used for original spectra to reduce noise and multiplicative scatter correction (MSC), which presented better determination results than other methods, was used to correct baseline variations.

2.5. Sample dividing

Rational division of the sample sets is important to improve the identification accuracy. Random sampling (RS), Kennard–Stone (KS) algorithm, Duplex, and sample set partitioning based on joint *x–y* distances (SPXY) methods are classical sample dividing methods (Zang et al., 2012). SPXY was usually used to divide samples in quantitative analysis. Previous studies have shown that RS and Duplex methods just divided the maximum into the calibration set, while KS algorithm divided both the maximum and the minimum into the calibration sets. Moreover, the data in calibration and prediction sets distributed more evenly when the samples were divided by KS algorithm than that divided by RS and Duplex methods. The main process of the KS algorithm is as follows:

Step 1: Calculate the Euclidean distance between every two samples of all samples for each peach variety. Euclidean distance $d_x(p,q)$ is calculated by Eq. (1).

$$d_x(p,q) = \sqrt{\sum_{j=1}^{J}\left[x_p(j) - x_q(j)\right]^2};\ p,q \in [1,N] \quad (1)$$

where $x_p(j)$ and $x_q(j)$ are the instrumental response at the j-th variable of samples p and q, respectively. J refers to the number of all variables, and it is 2074 in this study. N refers to the number of samples for each variety, and it is 100 here. The samples with the largest Euclidean distance are chosen as the first and second samples in the calibration set.

Step 2: Calculate every remaining sample's Euclidean distance to the selected samples, and the minimum distance was selected. Until every remaining sample's distance is calculated,