Sensors and Actuators B 229 (2016) 534–544

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

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

Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds

Lalit Mohan Kandpal[a], Santosh Lohumi[a], Moon S. Kim[b], Jum-Soon Kang[c], Byoung-Kwan Cho[a,*]

[a] Department of Biosystems Machinery Engineering, College of Agricultural and Life Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 305-764, South Korea
[b] Environmental Microbial and Food Safety Laboratory, Agricultural Research Service, US Department of Agriculture, Powder Mill Rd. Bldg. 303, BARC-East, Beltsville, MD 20705, USA
[c] Department of Horticultural Bioscience, Pusan National University, Miryang 627-706, South Korea

ARTICLE INFO

Article history:
Received 28 October 2015
Received in revised form 15 January 2016
Accepted 4 February 2016
Available online 8 February 2016

Keywords:
Seed viability
Hyperspectral imaging
Partial least-squares discriminant analysis (PLSDA)
Variable selection methods
Germination ability
Muskmelon seeds

ABSTRACT

A near-infrared (NIR) hyperspectral imaging (HSI) system was used to predict viability and vigor (in term of germination periods) in muskmelon seeds. Hyperspectral images of muskmelon seeds were acquired using a NIR push-broom HSI system covering the spectral range of 948–2494 nm. After NIR spectra collection, all seeds underwent a germination test to confirm their viability and vigor. The spectra from seeds with 3 and 5 germination days and nongerminated seeds were further used for development of a classification model of partial least-squares discriminant analysis (PLSDA). Most effective wavelengths were selected using three model-based variable selection methods, i.e., variable important in projection (VIP), selectivity ratio (SR), and significance multivariate correlation (sMC), which selected 23, 18, and 19 optimal variables, respectively, from full set of 208 variables. The selected variables from different waveband selection methods were found genuine and significant for interpreting the prediction results of seed viability and vigor. Subsequently, the PLS-DA model was constructed using individual VIP-, SR-, or sMC-selected variables. The results demonstrated that the PLSDA model developed with the selected optimal variables from the different methods provided comparable results for the calibration set; however, the PLSDA-SR method afforded the highest classification accuracy (94.6%) for a validation set used to determine the viability and vigor of muskmelon seeds. The wavelengths selected by the different methods represents chemical components of the seed and the attribute of germination ability was chosen most often.

1. Introduction

Seed viability and vigor are the two most important parameters directly related to seed germination performance and seedling emergence. During storage, seeds undergo physiological and physiochemical changes, termed aging, including deterioration of the chemical component of the seeds [22,33]. Many physiological and biochemical changes that take place during seed deterioration are or have been used to assess seed quality. However, with prolonged storage, seed viability decreases with the change in lipid peroxidation [45]. The rate at which the seed ages depends on its ability to resist degradation and on its protective mechanisms. Previous studies have defined seed viability as the development of the seed embryo and the ability to produce normal seedlings under favorable conditions (i.e., moisture, humidity, mechanical resistance, temperature), while seed vigor comprises those attributes that determine the potential for rapid, uniform development of normal seedlings under different field conditions [4]. Therefore,

---

Abbreviations: HSI, hyperspectral imaging; TZ, tetrazolium test; FOV, field of view; NIR, near infrared; FT-NIR, Fourier transform-near infrared; PLS, partial least-square; PLSDA, partial least square discriminant analysis; VIP, variable importance in projection; SR, selectivity ratio; sMC, significance multivariate correlation; SNV, standard normal variate; MSC, multiple scatter correction; RMSE, root mean square error; TP, target projection.

* Corresponding author. Fax: +82 42 823 6246.
E-mail addresses: lalitm85@gmail.com, bxc195@gmail.com (L.M. Kandpal), Santosh123@cnu.ac.kr (S. Lohumi), moon.kim@ars.usda.gov (M.S. Kim), kangjs@pusan.ac.kr (J.-S. Kang), chobk@cnu.ac.kr (B.-K. Cho).

http://dx.doi.org/10.1016/j.snb.2016.02.015
0925-4005/

a good quality seed has strong vigor and steady germination and establishes quickly in the field.

In general, seeds start to lose vigor before they lose their ability to germinate; therefore, vigor (germination potential) testing is an important practice in seed production programs. Some of the conventional methods used to estimate seed viability and vigor include the germination test, the tetrazolium test (TZ), and the electric conductivity test. However, these tests are time consuming, based on chemicals, and destroy the seed. To overcome these shortcomings, various spectroscopic techniques, i.e., Fourier transform-near infrared (FT-NIR) and mid infrared (Mid-IR), and Raman spectroscopy, are used to detect seed viability [37,29]. These spectroscopic techniques are nondestructive and provide a large amount of chemical information on the scanned sample. However, they are point-based scanning techniques and examine a relatively small area of a specimen, so they do not provide spatial information that is important for many seed inspection applications.

Hyperspectral imaging (HSI) is a powerful and fast emerging technique that is used for measuring the quality of agricultural and food products [21,25,26,30]. HIS provides spectral and spatial information simultaneously from the scanned samples in the form of a hypercube with two spatial dimensions and one spectral (wavelength) dimension. In addition, the HSI technique is faster than point-based spectroscopic techniques because many samples can be analyzed at same time. HSI can be used to obtain spatial and spectral information about an object in the ultraviolet, visible, and near-infrared regions (100–2500 nm) of the electromagnetic spectrum [15]. Besides a spectral range, HIS sensors have the flexibility to collect hyperspectral data for specimens of different sizes and shapes, and the spatial resolution and field of view (FOV) can be adjusted depending on the application.

HSI is a potential alternative to the point-based spectroscopic techniques used in a wide range of applications in the agro-food sector. It has been used to classify grains, including maize, wheat, barley, oat and groat, soybean, and rice seed [17], and, in particular, to determine the viability of pepper seed [36], native Australian plant seeds [38], and barley, wheat and sorghum seeds [34]. The HSI technique is generally used to identify or classify samples by recognizing their particular complex chemical or physical properties of the sample. However, spectral data consist of several hundreds to thousands of variables, which can be difficult to interpret without the help of the multivariate analytical method of chemometrics. Analysis of complex spectra, such as those containing relatively weak or greatly overlapping spectral bands, poses a challenge. Multivariate analysis is used for large amount of data, but there is the possibility of overfitting because of the presence of unwanted variables in the data. These unwanted variables make model interpretation difficult and affect the classification and prediction efficiency of the model. It is preferable to use fewer wavebands for different HSI applications because it can lead to faster sensor systems, thus reducing total integration time and increasing the number of samples inspected per second. Therefore, selecting the important variables and discarding unwanted variables can improve the performance of the model [53].

The aim of variable selection methods in HSI is to reduce dimensionality while preserving relevant information for later classification. A range of variable selection techniques have been used to select important variables for spectroscopic data [53,2]. In addition, Renzullo [42] used a wavelength selection method in combination with hyperspectral reflectance spectrometry to separate hyperspectral reflectance spectra of grapevine leaves. In addition, a successive projection algorithm in combination with partial least square (PLS) was used with the hyperspectral data of salmon flesh to visualize Enterobacteriaceae contamination. The combination yielded a higher prediction value and smaller root mean square (rms) error of prediction with the selected variables than the model developed with all of the spectral data [16].

To the best of our knowledge, our study is the first to use broader NIR-HSI region (over 1000 nm) in combination with model-based variable selection methods to predict viability and vigor (in term of germination period) of seeds. However, some previous studies used short NIR region (usually less than 700–1000 nm) to determine germination ability of the seeds [36,38]. We artificially aged muskmelon seeds to produce enough nonviable seeds to use in the development of a classification model, and to produce seeds that germinated at different periods to determine seed vigor. We hypothesized that chemical changes that occur in cell membrane during seed aging are responsible for seed viability and vigor and generate changes in reflectance profile in particular spectral region.

In this article, we present a novel approach to model-based variable selection methods that addresses the optimal separation of (seed) classes and the number of spectral bands to be selected. The performances of the variable selection methods in combination with PLS discriminant analysis (PLSDA) are compared with respect to classification accuracy and by interpreting the significance of individually selected variables.

2. Experimental

2.1. Samples preparation

The 288 muskmelon seeds used in this study were kindly provided by the Hyundai seed Co., Ltd. (Yeoju, Gyeonggi-do, Korea) in November 2013. Balesevic-Tubic et al. [7], reported the effects of accelerated and natural aging on soybean seeds and observed a decrease in lipid peroxidation and superoxide dismutase and peroxidase activities that have a negative effect on seed viability and vigor [49]. In this study, we used artificial aging (AA) treatment to yield seeds with different viability and vigor. According to Olouch and Welbaum [39], the viability and vigor of muskmelon seed are greatly affected by their variable moisture content during storage. Thus, AA seeds were immersed in water to get the moisture content to 20% by following the method of Lohumi et al. [29]. Three subsets of 72 seeds were vacuum-packed in plastic bags and stored in 45 °C hot water to age for 2, 4 and 6 days, while another set of 72 seeds did not undergo AA and were kept as a control (0 h). After aging, the seeds were dried in an incubator at 20 °C to bring them back to their original weight. All samples were vacuum-packed and stored at 4 °C until subjected to HSI data collection.

2.2. Hyperspectral data acquisition

A laboratory-based push-broom NIR-HSI system (shown in Fig. 1) was used to acquire the hyperspectral images of muskmelon seeds. The system comprised a line scan image spectrograph (SWIR, Headwall Photonics, Fitchburg MA, USA) that covered the spectral range 948–2494 nm with 8-nm spectral resolution, MCT (Mercury cadmium Telluride) detectors to detect the radiation reflected back from the samples, a high performance 320 × 208 camera (MCT, Headwall Photonics, Fitchburg, MA, USA), a stepper motor to move the conveyer belt, six tungsten halogen lamps (100 W each) with fiber optics to illuminate the samples, data acquisition software, and a display unit. To improve the signal to noise ratio, the samples was set to move at 0.25 mm/scan through the conveyor unit to be able to cover the spatial shape of the samples, the exposure time was set at 6000 μs, and the distance of the samples from the camera was set at 42 cm. Seed samples were placed on a black plate and transferred to the conveyer belt to be scanned line by line. The movement of a sample fixed to the translation stage was controlled by the step interval and the number of steps. Spatial and