Use of advanced chemometrics on visible-near infrared image data to improve discrimination of fungi on seeds

Nisha Shetty, Merete Halkjær Olesen, Birte Boelt and Rene Gislum

Aarhus University, Faculty of Agricultural Sciences, Department of Agroecology, Research Centre Flakkebjerg, DK-4200 Slagelse, Denmark
*Corresponding author: nisha.shetty@agrsci.dk; nisha.stat@gmail.com

Introduction

Seed health is an important aspect of quality in the seed production chain for increased crop productivity and improved food quality. In seed production of spinach (*spinacia oleracea* L.), more and more attention is paid to reducing disease incidence which, until now, has mainly been done by applying fungicides in the field. It is possible to reduce the amount of fungi on seeds[1] and in some countries seeds with fungi like e.g. *Verticillium* are not allowed. The test method used today is the freeze-blotter test in which the seeds are evaluated under microscope. This method for testing seed health is time consuming and the results are highly dependent on the laboratories and their experience.[2] Earlier studies demonstrated the possibility of using multispectral VIS-NIR spectroscopy for detection of infected spinach seeds[3]. However, in earlier study, it was not possible to discriminate between seeds infected with the three most important fungi namely *Cladosporium*, *Stemphylium* and *Verticillium* due to the similarity in mean spectral intensity. These fungi are seed-borne and seed-transmitted pathogens of spinach.

In the present experiment, first order image features along with textural features were used to classify *Cladosporium*, *Stemphylium* and *Verticillium*. A supervised classification method (extended canonical variates analysis, ECVA)[4] was used; ECVA is based on standard canonical variates vnalysis (CVA) [5,6]. CVA is a method for estimation of directions in space that maximise the differences between groups in the data according to a well-defined optimisation criterion. A drawback of CVA is that it cannot deal with highly collinear data such as those found in spectroscopic applications. For this purpose, a modification of the CVA method was developed[4] to cope with collinear high-dimensional data called - ECVA. This method utilises partial least squares (PLS) regression as an engine for solving an eigenvector problem involving singular covariance matrices. The purpose of this study was to evaluate classification of the three fungi (*Cladosporium, Stemphylium* and *Verticillium)* using advanced chemometrics on multispectral VIS-NIR image data.

Materials and Methods

Samples

Seeds (n=400) from four different seed lots of spinach were subjected to a modified[1] freeze-blotter seed health test. A total of 28 spinach seeds that were only infected with one of the three different fungi species *Cladosporium* (13), *Stemphylium* (9) and *Verticillium* (6) were selected for imaging and analysis in the present study.

Imaging

One week after plating the seeds, images were captured using a VideometerLab instrument (Videometer A/S, Denmark). The instrument is equipped with a camera inside an integrating sphere illuminated by light emitting diodes (LEDs) ranging in wavelengths from 395 to 970 nm. The integrating sphere ensures that light is scattered evenly, giving a uniform diffuse lighting. The diodes emit a narrow band of light that ensures that only light of the desired wavelength is present when an image is acquired. Images of 1280x960 pixels were captured within a range of all spectral bands. Prior to image capture, the image was calibrated with respect to colour, geometry and self-illumination, thereby gaining a set of directly comparable images.

Data analysis

First order statistics of the image (such as mean, standard deviation, median, kurtosis, minimum and maximum) were extracted using a grey-level histogram and Haralick texture features (such as angular, entropy, inverse, contrast and correlation) were extracted using a grey level co-occurrence matrix. Texture and greyscale data were extracted from the image using a macro written in the free software ImageJ 1.43 (http://rsbweb.nih.gov/ij/). Principal component analysis (PCA) was used to reduce the multi-dimensional features to simplified features. The important features thus obtained were used to train the classification

Reference paper as:
N. Shetty, M.H. Olesen, B. Boelt and R. Gislum (2012).Use of advanced chemometrics on visible-near infrared image data to improve discrimination of fungi on seeds, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 342-344.

model, which was validated using random cross-validation (10 iterations and 6 segments). ECVA, a supervised classification method was used for discriminating the three fungi groups. Prior to the analysis, data was scaled block-wise. Data analyses were performed using MATLAB version 7.9 (R2009b) (The Mathworks, Inc., Natick, MA, USA) along with the PLS toolbox 5.5.2 (Eigenvector Research, Inc., Manson, WA, USA) and ECVA toolbox 2.01(http://www.models.life.ku.dk).

Results and Discussion

Some of the features extracted allowed discrimination of spinach seeds infected with the three types of fungi *Cladosporium*, *Stemphylium* and *Verticillium*. PCA loading plots of first order statistics showed that wavelength bands of 'maximum' and 'mean' had the highest variation. Median had similar information to mean (Figure 1a). PCA loadings on texture features indicated that wavelength bands of 'contrast' had maximum variation compared to the rest of the features namely angular, entropy, inverse and correlation (Figure 1b). ECVA on mean, maximum, contrast separate and the combination of mean+maximum, mean+contrast, maximum+contrast showed discrimination according to three types of fungi; best separation was obtained for mean (Figure 2a) and maximum+contrast data (Figure 2b). It was found that, although histogram-based features are quite effective, sometimes a histogram cannot provide sufficient information required for processing. In those cases, it is desirable to use higher order image statistics such as texture features extracted using a grey level co-occurrence matrix in combination with histogram-based statistics.

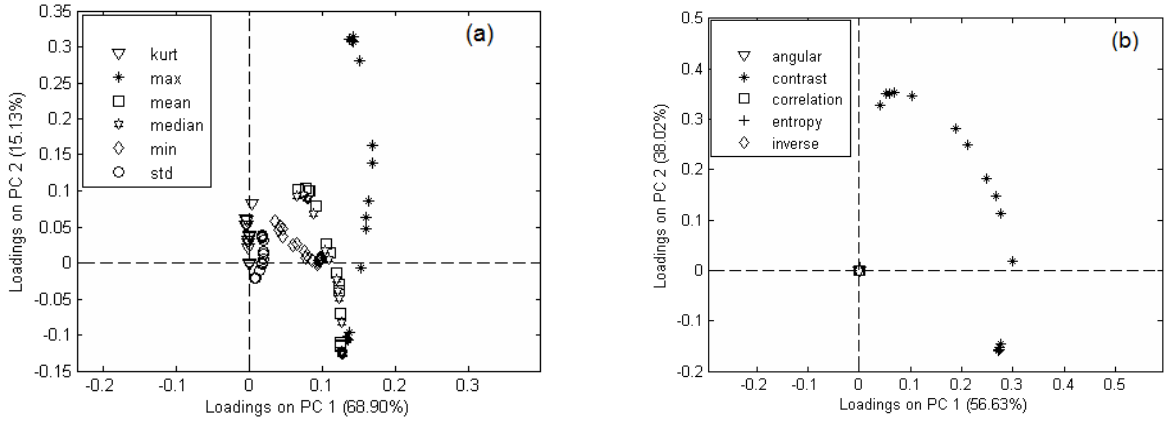

Figure 1. PCA loadings plot of first two components (a): histogram-based features and (b): texture features.

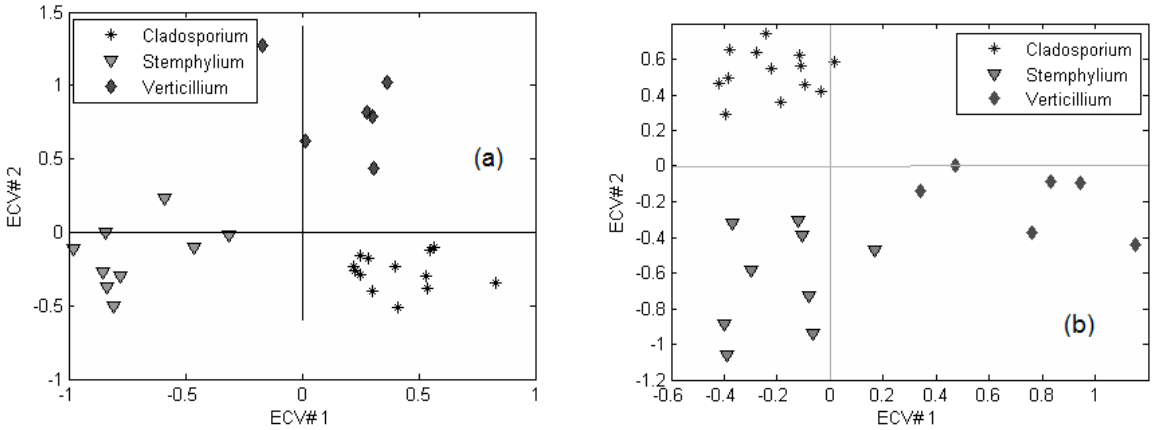

Figure 2. ECVA score plot of first two components (a): using mean, (b): using maximum + contrast.

Conclusion

A multivariate classification model based on VIS-NIR multispectral image data using advanced chemometrics method is demonstrated and it is plausible that spinach seeds infected with the different fungi can be separated using some of the histogram-based and Haralick texture features. ECVA score plots using VIS-NIR multispectral image features, 'mean' and 'maximum+contrast' allowed for the best discrimination of spinach seeds infected with the three fungal species: *Cladosporium*, *Stemphylium* and *Verticillium*. The cross-validated misclassification percentage using 'maximum+contrast' was 3.5% and zero using 'mean' data.

Reference paper as:
N. Shetty, M.H. Olesen, B. Boelt and R. Gislum (2012).Use of advanced chemometrics on visible-near infrared image data to improve discrimination of fungi on seeds, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 342-344.

References

1. L.J. du Toit and P. Hernandez-Perez, *Plant Dis*. **89**, 1305-1312 (2005).
2. M.B. McDonald, A.F. Evans and M.A. Bennett, *Seed Sci. Technol.* **29**, 683-689 (2001).
3. M.H. Olesen, J.M. Carstensen and B. Boelt, *Seed Sci. Technol.* **39**, 140-150 (2011).
4. L. Nørgaard, R. Bro, F. Westad and S.B. Engelsen, *J. Chemom*. **20**, 425-435 (2006).
5. C.R. Rao, *Advanced Statistical Methods in Biometric Research*. John Wiley & Sons, Inc., New York (1952).
6. W.J. Krzanowski, *Principles of Multivariate Analysis: A User's Perspective*, revised edition. Oxford University Press, Oxford (2000).

Reference paper as:
N. Shetty, M.H. Olesen, B. Boelt and R. Gislum (2012).Use of advanced chemometrics on visible-near infrared image data to improve discrimination of fungi on seeds, in: Proceedings of the15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 342-344.