CAFEi2016-133

Comparison of chemometrics methods for classification of sugarcane brix using visible and shortwave near-infrared technology

C.D.M Ishkandar[1], N.M. Nawi[1,a], G. Chen[2], T. Jensen[2] and S.A. Mehdizadeh[3]

[1]Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Selangor, Malaysia; [2]Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia; [3]Department of Agricultural Machinery and Mechanization, Ramin Khuzestan University of Agriculture and Natural Resources, Khuzestan, Iran.

Abstract

The potential of visible and shortwave near-infrared (VSWNIR) diffuse reflectance spectroscopy in the range of 400 to 1000 nm, in combination with three classifier algorithm techniques, was investigated to classify sugarcane quality into three quality classes (high, medium and low). Two hundred and ninety sugarcane internode samples were used to evaluate the ability of this technology. Each internode sample was scanned at four scanning points to obtain the spectral data which was later correlated with its sugar content (Brix value). The Brix values were later classified using three classifier algorithms namely Bayesian discriminant analysis (BDA), artificial neural network (ANN) and support vector machine (SVM). The results shows that the overall classification accuracies achieved by BDA, SVM and ANN were 77.8, 83.1 and 88.7% respectively. The results demonstrated that the VSWNIR spectroscopy together with chemometrics techniques could be a rapid tool to be used for classification of sugarcane quality based on spectral data.

Keywords: VSWNIR spectroscopy, sugarcane, classification, chemometrics techniques

INTRODUCTION

Sugarcane is an important crop in many countries including Brazil, India, United States and Australia. In order to improve the yield and quality of the crop, the application of precision agriculture (PA) technologies in this industry was recommended by Bramley (2009). PA is a valuable management tool to maximise farm profits through efficient application of crop inputs by matching them to potential variations of crop yield and quality in the field (Wendte et al., 2001). One of the requirements for PA is a crop map. Nawi et al. (2014) proposed the application of a spectroscopic technology on a chopper harvester to measure and map sugarcane quality in a field during harvest. The use of spectral data coupled with a powerful classification algorithm techniques could be applied to rapidly map crop quality over the paddock (Menesatti et al., 2010).

Lately, many classification methods have been employed to classify agricultural produces. For examples, Xie et al. (2009) applied least-squares support vector machines (LS-SVM), discriminant analysis (DA), soft independent modelling of class analogy (SIMCA), and discriminant partial least squares (DPLS) for tomatoes classification. Shahin et al. (2002) applied Bayesian discriminant Analysis (BDA) to classify sweet onions based on internal damage. Besides, the artificial neural network (ANN) was applied to discriminate varieties of Chinese bayberry (Li et al., 2007). ANN was also used by Nawi et al. (2013) to classify sugarcane Brix into five quality classes. The average accuracy for classification reported by this study was 83.1%.

---

[a] Email: nazmimat@upm.edu.my

Although various classification methods are available, there has been no systematic comparison among these approaches in their relative ability to classify spectral data for sugarcane samples. Thus, this paper aims to compare the performance of classification methods namely BDA, SVM and ANN to classify sugarcane quality based on their spectral data. The specific objectives of this study were (1) to classify sugarcane Brix into three quality classes using BDA, ANN and SVM (2) to identify the best classifier for classification of sugarcane Brix.

MATERIALS AND METHODS

Sample Preparation

A total of 290 internodes were extracted from 22 sugarcane stalk samples. The stalk samples were collected from the research station of Bureau of Sugar Experimental Station (BSES), Bundaberg, Australia. The stalk samples were first topped and cut into individual internode using a cutter. Each internode sample was scanned at four different scanning points from bottom to the top (Figure 1). Each section was imaginarily labeled as S1, S2, S3 and S4 following the sequence from S1 to S4 (bottom to the top). Detailed explanation on the stalk preparation can be found in Nawi et al. (2013).

Reflectance Measurement

In this study, a handheld visible and shortwave near-infrared (VSWNIR) spectroradiometer (FieldSpec HandHeld and FieldSpec Pro FR, 325 to 1075 nm, Analytical Spectral Devices (ASD), Inc., Boulder, USA) was used to collect reflectance readings from the skin surface over the wavelength range of 400 to 1000 nm. The scanning was carried out using the 25° field of-view (FOV) of the spectroradiometer inside a light-proof measurement box (90 x 60 x 40 cm). Two halogen lamps were used to provide illumination to the sensor. All spectral data were stored in a computer and processed using the RS3 software for Windows (Analytical Spectral Devices, Boulder, USA Pro). The reflectance spectra were collected from four different locations on each internode. Four spectra from each internode (S1 to S4) were then averaged and later used for calibration against Brix values.

Brix Measurement

After the spectral acquisition, each internode was squeezed using a clamp to extract juice samples for the determination of sugar content by measuring the Brix value using a handheld Brix refractometer (Model: RHB-32ATC, China, the Brix range is from 0 to 32% with automatic temperature compensation).

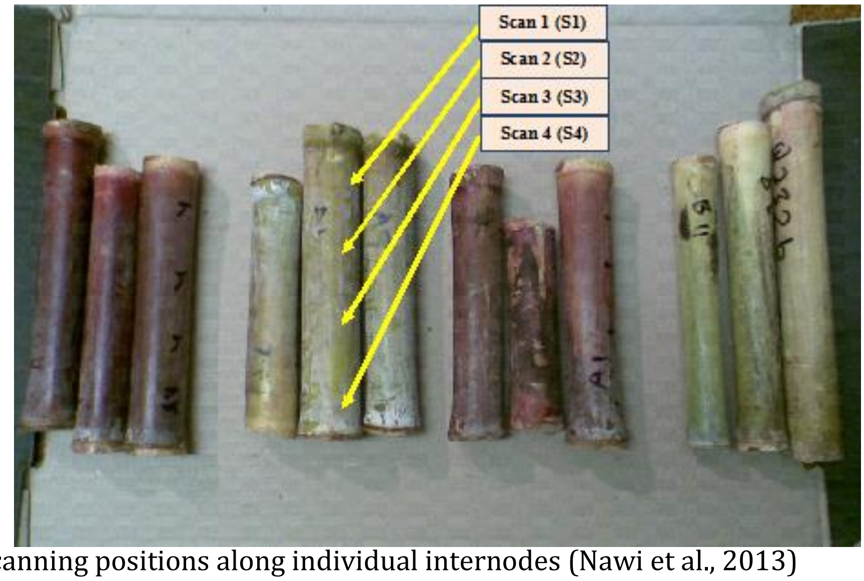

Figure 1. Scanning positions along individual internodes (Nawi et al., 2013)

Principal Component Analysis (PCA)

PCA is a well-known multivariate data reduction method which transforms original data into new orthogonal variables referred to as principal components, or PCs (Purcell et al., 2005). Usually, only a few of the new PCs or latent variables are required to describe most of the data variance with the first PC accounting for the greatest amount of variance. A low number of PCs were desirable in order to avoid inclusion of signal noise in the modeling (Xiaobo et al., 2007). In this paper, PCA for all the classifier algorithms was executed using Matlab (Version 7, The Mathworks Inc).

Description of Classification Models

The classifier algorithms used for Brix classification in this study were Bayesian discriminant analysis (BDA), support vector machine (SVM) and artificial neural network (ANN). Detail explanation of each classifier is described below.

1. Bayesian discriminant analysis (BDA)

BDA is a simple and practical classification method developed based on Bayesian discriminant theory (Anderson et al., 2003). Xiaobo et al. (2010) used this method to classify an apple external quality by combining the size, shape, and color. A Bayesian classifier was also successfully applied for quality classification of sweet onions (Shahin et al., 2002) and oranges, peaches and apples (Blasco et al., 2003). The classification algorithms for BDA was developed and executed in Matlab (Version 7, The Mathworks Inc.).

2. Support Vector Machine (SVM)

Originally intended for binary classification, SVM has been investigated for solving multi-class problems using various discrimination strategies including the one-versus all, the one-versus-one and the directed acyclic graph (DAG) SVM (Platt et al., 2000). The study from Zhao et al. (2006) showed that NIRS combined with SVM could be efficiently utilized for rapid and simple identification of the tea categories. Nashat et al. (2011) applied image processing applications in bakery and food quality assessment and concluded that the SVM classifier offered several advantages because this algorithm is simple, cost-effective, and rapid. The development of classification algorithm adopted in this paper was based on DAG SVM method described by Nashat and Abdullah (2010).

3. Artificial Neural networks (ANN)

ANN is a well-known non-linear method which could provide a robust classification model (Lee et al., 2010). Most applications of ANNs in postharvest technology have been for classification purposes (Hahn et al., 2004). The ANN used in this paper was a *Perceptron* model also known as *Back-propagation Perceptron*. This type of network was selected because it is an excellent pattern classifier (Torrecilla et al., 2004). The ANN selected was a feed-forward network with supervised learning. The ANN model consists of two layers with connections to the outside world (an input layer where data are presented to the network and an output layer which holds the network's response to given inputs), and one hidden layers with ten neurons. The input layer had three neurons which corresponded to the first three PCs as determined by PCA methods. In this study, the sigmoid function was applied as the transfer function (Liu et al., 2007). The ANN algorithm including PCA method for classification purposes in this study was programmed and executed in Matlab (Version 7, The Mathworks Inc.).

Classification of quality data

To assess the performance of different classification methods on classification of spectral data, the sugarcane Brix values were divided into three classes using the equal interval classification method (Table 1). The division into three classes was consistent with a method reported by Bramley et al. (2012) who used three quality categories on their work to map sugarcane quality in a paddock.

Training and testing data sets

To apply all classifier methods, spectral data were divided into training and testing sets. The training set (219 samples) was used to develop the classifier models and while the testing set (71 samples) was used to evaluate the performance of the classifier models.

Table 1. Classification table for sugarcane Brix

Class	Brix range
High (H)	17.5-22.2
Medium (M)	12.6-17.4
Low (L)	7.6-12.5

RESULTS AND DISCUSSIONS

Bayesian discriminant analysis (BDA)

Table 2 shows the results of classification performance by using BDA classifier. In this model, based on three quality classes, the accuracy of BDA classification ranged from 60% to 79.5% with the average accuracy being 77.8%. The accuracy reported in this study is similar to the accuracy reported by Ventura et al. (1998) for scoring of the apples according to sugar content which yielded good classification of 72 and 76% with thresholds of 12 and 13 Brix, respectively. Class high (H) and medium (M) have achieved reasonable prediction performance with 79.5 and 78.2% accuracy, respectively. The most misclassification was observed in class low (L) with only 60% accuracy. This low accuracy was probably due to low number of sample in that particular group.

Table 2. Bayesian classification results for three quality classes

Class	Brix range	No. of test sample in each group	No. of correct classification	Accuracy (%)
High (H)	19.3-22.2	44	35	79.5
Medium (M)	13.5-16.4	22	17	78.2
Low (L)	7.6-10.4	5	3	60
Total		72	56	77.8

Support vector machine (SVM)

In this model, the accuracy of classification using SVM for three quality classes ranged from 60% to 90.9% with the average accuracy being 83.1% (Table 3). This finding is higher than the classification accuracies of 79% reported by Ravikanth et al. (2015) who applied SVM classifier on the raw NIR reflectance spectra to classify all the foreign material types from wheat. However, the finding is slightly lower than a study carried out by Nashat and Abdullah (2010) who reported the correct classification rate of 87.3% when applying SVM to classify biscuits into one of eight distinct groups, corresponding to different degrees of baking. It can be seen that SVM performs better than BDA in order to classify sugarcane quality classes.

Table 3. SVM classification results for three quality classes

Class	Brix range	No. of test sample in each group	No. of correct classification	Accuracy (%)
High (H)	19.3-22.2	44	40	90.9
Medium (M)	13.5-16.4	22	16	72.7
Low (L)	7.6-10.4	5	3	60
Total		71	59	83.1

Artificial Neural Network (ANN)

Table 4 shows the accuracy of classifying sugarcane °Brix using ANN with three PCs. Three classes with their respective threshold values were predefined from the distribution of °Brix as discussed in Table 1. In this model, the accuracy of classification ranged from 40 to

95.5% with the average accuracy being 88.7%. The accuracy ranges obtained in this study were comparable to the work reported by Park et al. (2003) for classifying soluble solids of Gala apples with 45.2 to 93.5% accuracy. However, the result of this study was lower than the result (99.5% accuracy) reported by Mohan et al. (2005) who used ANN to classify grain quality. For sugarcane study, this result is better than the result reported by Nawi et al. (2013) who obtained an average accuracy of 83.1%. In that study, the authors classified quality classes into 5 categories. Since this report was based on the same data used by Nawi et al. (2013), it is concluded that the fewer class number could increase the classification performance.

Table 4. ANN classification results for three quality classes

Class	Brix range	No. of test sample in each group	No. of correct classification	Accuracy (%)
High (H)	19.3-22.2	44	42	95.5
Medium (M)	13.5-16.4	22	19	86.4
Low (L)	7.6-10.4	5	2	40
Total		71	63	88.7

CONCLUSIONS

The following conclusions can be drawn from the study:

– This study has demonstrated that the application of VSWNIR spectroscopy combined with classification algorithm can be applied to classify sugarcane Brix based on spectral data.

– The ANN used to classify °Brix into several quality classes had yielded good classification performance ranging from 40 to 95.5% accuracy with overall accuracy being 88.7%.

– BDA classifier has shown an acceptable classification performance and could classify all the spectra with accuracy up to 77.8%. For SVM, its overall classification accuracy was 83.1%. This study has shown that ANN has outperformed the BDA and SVM as shown by its average accuracy.

– The results show that VSWNIR spectroscopy together with chemometrics techniques could be applied as a rapid tool for classification of sugarcane quality based on spectral data.

Literature cited

Anderson, D. C., Li, W., Payan, D. G., and Noble, W. S. (2003). A New Algorithm for the Evaluation of Shotgun Peptide Sequencing in Proteomics: Support Vector Machine Classification of Peptide MS/MS Spectra and SEQUEST Scores. J. Proteome Res. 137-146.

Blasco, J., Aleixos, N., and Moltó, E. (2003). Machine Vision System for Automatic Quality Grading of Fruit. Biosys. Eng. *85(4),* 415-423.

Bramley, R. G. V. (2009). Lessons from nearly 20 years of Precision Agriculture research, development, and adoption as a guide to its appropriate application. Crop Pasture Sci. *60,* 197-217.

Bramley, R. G. V., Panitz, J. H., Jensen, T., and Baillie, C. (2012). Within block spatial variation in CCS-another potentially important consideration in the application of precision agriculture to sugarcane production. In Proceedings of the Australian Society of Sugar Cane Technologists. *34*, 1-8.

Hahn, F., Lopez, I., Hernandez, G. (2004). Spectral detection and neural network discrimination of Rhizopus stoloniferspores on red tomatoes. Biosys. Eng. *89(1),* 93–99.

Lee, W. S., Alchanatis, V., Yang, C., Hirafuji, M., Moshou, D., and Li, C. (2010). Sensing technologies for precision specialty crop production. Comput Electron Agric. *74(1),* 2-33

Li, X., He, Y., and Fang, H. (2007). Non-destructive discrimination of Chinese bayberry varieties using Vis/NIR spectroscopy. J. Food Eng. *81(2),*357-363.

Liu, H., Darabi, H., Banerjee, P., & Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), *37(6),* 1067-1080.

Menesatti, P., Antonucci, F., Pallottino, F., Roccuzzo, G., Allegra, M., Stagno, F., & Intrigliolo, F. (2010). Estimation of plant nutritional status by Vis–NIR spectrophotometric analysis on orange leaves [Citrus sinensis (L) Osbeck cv Tarocco]. Biosyst. Eng. *105(4),* 448-454.

Mohan, L. A., Karunakaran, C., Jayas, D. S., and White, N. D. G. (2005). Classification of bulk cereals using visible and NIR reflectance characteristics. Canadian Biosystems Engineering, *47(7),* 7-14.

Nashat, S., and Abdullah, M. (2010). Multi-class colour inspection of baked foods featuring support vector machine and Wilk's λ analysis. J. Food Eng, 370-380.

Nashat, S., Abdullah, A., & Abdullah, M. Z. (2011, May). A robust crack detection method for non-uniform distributions of coloured and textured image. In 2011 IEEE International Conference on Imaging Systems and Techniques, 98-103. IEEE.

Nawi, N. M., Chen, G., Jensen, T., and Mehdizadeh, S. A. (2013). Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared. Biosyst. Eng. *115*, 154–161.

Nawi, N. M., Chen, G., and Jensen, T. (2014). In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: A review. Precis. Agric. *15(6),* 684-703.

Park, B., Abbott, J. A., Lee, K. J., Choi, C. H., and Choi, K. H. (2003). Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and gala apples. Trans. ASAE. *46(6),* 1721-1731.

Platt, J. C., Cristianini, N., and Shawe-Taylor, J. (1999, November). Large Margin DAGs for Multiclass Classification. In nips. 12, 547-553.

Purcell, D. E., Leonard, G. J., O'Shea, M. G., and Kokot, S. (2005). A chemometrics investigation of sugarcane plant properties based on the molecular composition of epicuticular wax. Chemometr. Intell. Lab. *76*, 135-147.

Ravikantha, L., Singha, C. B., Jayasa, D. S., and White, N. D. (2015). Classification of contaminants from wheat using near-infrared hyperspectral imaging. Biosyst. Eng. *135*, 73-86.

Shahin, M. A., Tollner, E. W., Gitaitis, R. D., Sumner, D. R., and Maw, B. W. (2002). Classification of sweet onions based on internal defects using image processing and neural network technique. American Society of Agricultural and Biological Engineers, 1613-1618.

Torrecilla, J. S., Otero, L., and Sanz, P. D. (2004). A neural network approach for thermal/pressure food processing. J. Food Eng. *62*, 89-95.

Ventura, J., Libermana, R. P., Greena, M. F., Shanera, A., and Mintza, J. (1998). Training and quality assurance with the structured clinical interview for DSM-IV (SCID-I/P). Psychiatry Res. *79(2),* 163-173.

Wendte, K. W., Skotnikov, A. and Thomas, K. K. (2001). Sugar cane yield monitor, U.S. Patent No. 6,272,819.

Xiaobo, Z., Jiewen, Z., Xingyi, H., and Yanxiao, L. (2007). Use of FTNIR spectrometry in non-invasive measurements of soluble solid contents SSC of "Fuji" apple based on different PLS models. Chemometr. Intell. Lab. Sys. *87*, 43-51.

Xiaobo, Z., Jiewena, Z., Poveyb, M. J., Holmesb, M., and Hanpina, M. (2010). Variables selection methods in near-infrared spectroscopy. Anal. Chim. Acta, *667(1-2),* 14-32.

Xie, L., Ying, Y., and Ying, T. (2009). Classification of tomatoes with different genotypes by visible and short-wave near-infrared spectroscopy with least-squares support vector machines and other chemometrics. J. Food Eng. *94*, 34-39.

Zhao, J. W., Chen, Q. S., Huang, X. Y., and Fang, C. H. (2006). Qualitative identification of tea categories by near infrared spectroscopy and support vector machine. J. Pharm. Biomed. Anal. *41*, 1198–1204.