Advance Journal of Food Science and Technology 7(9): 691-696, 2015
DOI:10.19026/ajfst.7.1629
ISSN: 2042-4868; e-ISSN: 2042-4876

Submitted: July 18, 2014 Accepted: September 13, 2014 Published: March 25, 2015

Research Article

Discrimination of Rice Varieties using LS-SVM Classification Algorithms and Hyperspectral Data

[1]Jin Xiaming, [1, 2]Sun Jun, [2]Mao Hanping, [1]Jiang Shuying, [2]Li Qinglin and [1]Chen Xingxing
[1]School of Electrical and Information Engineering,
[2]Laboratory Venlo of Modern Agricultural Equipment, Jiangsu University, Zhenjiang 212013, P.R. China

Abstract: Fast discrimination of rice varieties plays a key role in the rice processing industry and benefits the management of rice in the supermarket. In order to discriminate rice varieties in a fast and nondestructive way, hyperspectral technology and several classification algorithms were used in this study. The hyperspectral data of 250 rice samples of 5 varieties were obtained using FieldSpec®3 spectrometer. Multiplication Scatter Correction (MSC) was used to preprocess the raw spectra. Principal Component Analysis (PCA) was used to reduce the dimension of raw spectra. To investigate the influence of different linear and non-linear classification algorithms on the discrimination results, K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Least Square Support Vector Machine (LS-SVM) were used to develop the discrimination models respectively. Then the performances of these three multivariate classification methods were compared according to the discrimination accuracy. The number of Principal Components (PCs) and K parameter of KNN, kernel function of SVM or LS-SVM, were optimized by cross-validation in corresponding models. One hundred and twenty five rice samples (25 of each variety) were chosen as calibration set and the remaining 125 rice samples were prediction set. The experiment results showed that, the optimal PCs was 8 and the cross-validation accuracy of KNN (K = 2), SVM, LS-SVM were 94.4, 96.8 and 100%, respectively, while the prediction accuracy of KNN (K = 2), SVM, LS-SVM were 89.6, 93.6 and 100%, respectively. The results indicated that LS-SVM performed the best in the discrimination of rice varieties.

Keywords: Classification algorithm, hyperspectral technology, rice variety

INTRODUCTION

Rice, one of the major eating foods, is the main raw material for daily meal of people in China. The nutritional value and taste of rice in diverse regions and varieties are different. In China, the main producing domains of rice lay in East and South of the Yangtze River area. In order to meet the nutritional needs and purchase demand of customers, it is necessary to classify the rice based on quality and variety reasonably and it is also a trend of marketing management of large-scale food supermarket. At present, the classification of rice varieties in China is still in the stage of manual sorting, which is time-consuming and laborious. There have been a few reports about the application of variety classification or grading in fruit, fish and meat. Sarbu *et al*. (2012) used the UV-Vis spectroscopy to classify the kiwi and pomelo based on the combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). Pholpho *et al*. (2011) used the pattern recognition method to realize the classification of intact longan and bruised longan based on the visible spectrum. Cen used visible/near infrared spectroscopy to classify the orange varieties and compared the classification accuracy of neural network with that of partial least squares (Cen *et al*., 2007).

In addition, Wang *et al*. (2011) used hyperspectral reflectance imaging technique to discriminate insect infestation from other confounding surface features in jujubes. Zhu *et al*. (2013) investigated the potential of visible and near infrared hyperspectral imaging as a rapid and nondestructive technique to determine whether fish has been frozen-thawed and obtained a good classification performance. Barbin *et al*. (2012) developed a hyperspectral imaging technique to achieve fast, accurate and objective determination of pork quality grades. Zhao *et al*. (2010) used NIR and support vector data description to discriminate egg's freshness and achieved good result. In the above literatures, the method of fruit classification was effective and it also showed the feasibility of using the spectral technology to classify the fruit grades. However, there were few reports referring to the classification of rice according to variety and quality.

In recent years, hyperspectral technology has gained wide application in different fields by virtue of its advantages over other analytical technology and it

Corresponding Author: Sun Jun, School of Electrical and Information Engineering of Jiangsu University, Zhenjiang 212013, China, Tel.: +86-13775544650; Fax: +86-051188780088

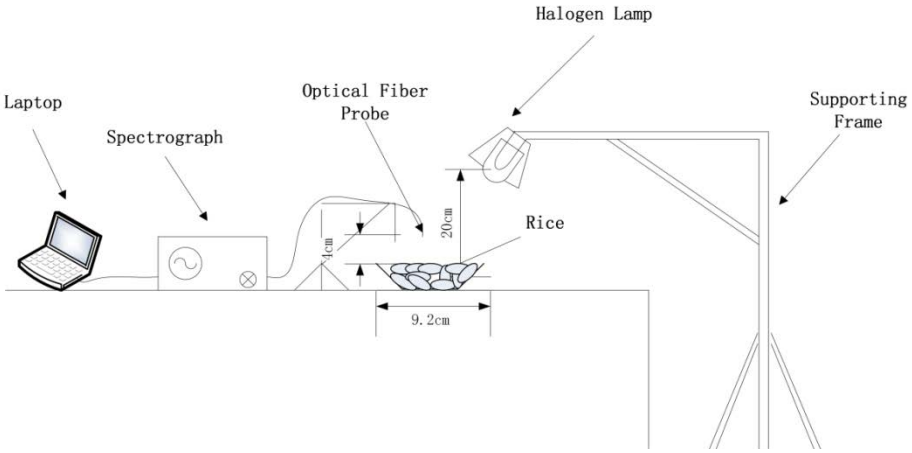

Fig.1: Schematic representation of hyperspectral data acquisition device

has been one of the most dominant in the field of nondestructive detection (Qin *et al*., 2012; Li *et al*., 2012; Watanabe *et al*., 2013). However, hyperspectral data has overlap bands and a large amount of information, so it is difficult to process the hyperspectral data directly. Therefore, spectral preprocessing, feature extraction and classification algorithm will be investigated and an optimal model for classifying rice variety will be found in this study

MATERIALS AND METHODS

Hyperspectral data acquisition device: Hyperspctral data acquisition device is composed of portable spectrometer, the auxiliary light source, notebook computer, round containers and experiment platform. FieldSpec®3 portable spectrum analyzer made by ASD company in American was used to obtain hyperspectral data, with the spectral measurement range of 350-2500 nm. In the spectral range of 350-1000 nm, the sampling interval is 1.4 nm and the spectral resolution is 3 nm, while in the range of 1000-2500 nm, the sampling interval is 2 nm and the spectral resolution is 10 nm. Finally, Hyperspectral data were derived in the form of ASCII and stored in the computer for the following process. The spectral data analysis software was ASD View SpecPro. Hyperspectral data acquisition device is shown as Fig. 1.

The halogen lamp which has a wide range of spectrum and adjustable light was chosen as auxiliary light source and it can meet the need of spectral detection. Spectral information was captured using spectral optical fiber probe and transferred to spectrometer through an optical fiber. Signal was parsed by the spectrometer and transferred to portable computer. Hyperspectral data were read by computer using spectral analysis software and saved as binary file automatically.

Sample preparation and spectral data acquisition: In this study, rice samples were purchased from WAL-MART supermarket laying at DingMao Road No. 198 in Jiangsu Zhenjiang, including a total of 5 species, such as:

- Ruan-Ya-Xiang-Si
- Jiang-Su
- Chang-Li-Xiang
- Zhen-Zhu
- Si-Miao

These samples were manually labeled five tags and then stored in the plastic bag under the room temperature.

During the experiment of spectral data acquisition, the rice sample was firstly placed on the black velvet and then spectral probe was placed 4 cm above table, perpendicular to the circular vessels with diameter of 9.2 cm and height of 2 cm. The angle of auxiliary light and experimental platform was kept as 45^{o} and the vertical distance between auxiliary and experimental platform was 20 cm. The field of view was set as 25^{o}. Before the measurement of rice sample, the standard reflecting plate was measured to eliminate the system error caused by the environmental factors such as light intensity. Finally, all the measurements of each sample were repeated for 3 times and the average value was taken as the final measurement results.

Multiplication Scatter Correction (MSC): Sample in homogeneity would cause the great difference in the sample spectrum and the spectral changes caused by scattering will be greater than that caused by sample components. In MSC method, each spectrum should be linear with the ideal spectrum and the ideal spectrum can be approximated as the average spectrum of calibration set. The reflection absorbance value under arbitrary wavelength of each sample has an approximately linear relationship with the corresponding absorbance spectrum of average spectra. Linear intercept and slope can be regressed by spectra set and be used to calibrate each spectrum. Intercept size reflects the unique reflection action of sample and

the slope size reflects the uniformity of samples (Sirisomboon *et al*., 2012; Zhang *et al*., 2012).

The expression of average spectrum is shown as formula (1), linear regression is shown as formula (2) and correction formula of MSC is shown as formula (3):

$$\overline{X_i} = \sum_{i=1}^{n} \frac{X_i}{n} \quad (1)$$

$$X_i = m_i \overline{X_i} + b_i \quad (2)$$

$$X_{i(MSC)} = \frac{(X_i - b_i)}{m_i} \quad (3)$$

where, X is the spectral matrix of calibration set, X_i is the spectrum of ith sample, m_i, b_i are the slope and intercept respectively of linear regression of the ith spectrum X_i and average spectrum X. Through the adjustment of m_i and b_i, the spectral difference is reduced, at the same time, the original information relevant to chemical composition is tried to be kept. Through the correction, the random variation can be deduced in maximum degree.

Principal Component Analysis (PCA): PCA is an unsupervised pattern recognition method and is used for visualizing data trends in a dimensional space. It has been applied in many fields. Generally, PCA is one of the techniques that commonly used for intending to eliminate the redundant information and reduce the computational burden by using mathematical method. The main objective of PCA is to use fewer variables to explain most of the variation in the raw data and many highly relevant variables are changed into those which are independent or unrelated to each other (Peng *et al*., 2014; Liu and Ngadi, 2013; Serranti *et al*., 2013).

K-Nearest Neighbors (KNN): KNN is a simple linear classification algorithm in the machine learning and it is also one of the most popular classification methods in pattern recognition. In KNN classification, an unknown sample of the prediction set is classified according to the majority of its K-nearest neighbors in the calibration set (Ji-yong *et al*., 2011). In this study, KNN was used, the k samples which are the nearest to test sample in the all N samples will be found. The identification rate of KNN model is influenced by parameter K which can be determined by calibration process.

Support Vector Machine (SVM): SVM is a non-linear supervised learning method for linear or nonlinear classification problems which was developed by Vapnik and his co-workers. At the beginning, SVM only can be used for Binary classification problems and then with the development of its theory, it can also be used for multi-class problem. It works by obtaining the optimal boundary of two groups in vector space independent on the probabilistic arrangements of vectors in the calibration set. When the linear boundary in the low dimension input space is not enough to

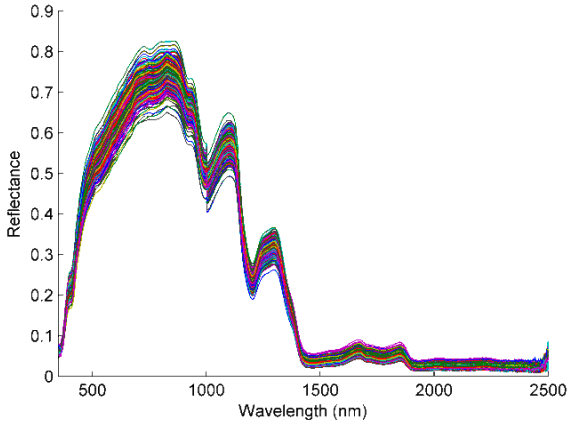

Fig. 2: Raw spectra of all rice samples

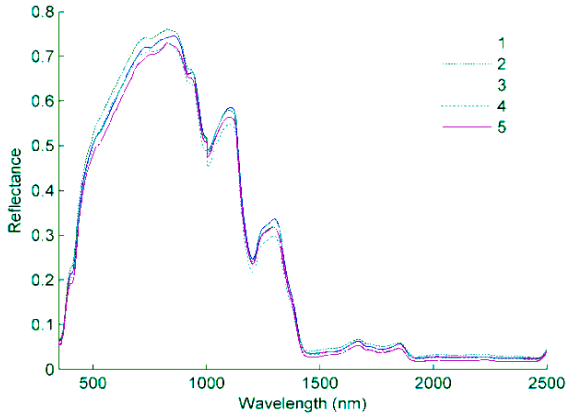

Fig. 3: Average spectra of 5 varieties of rice

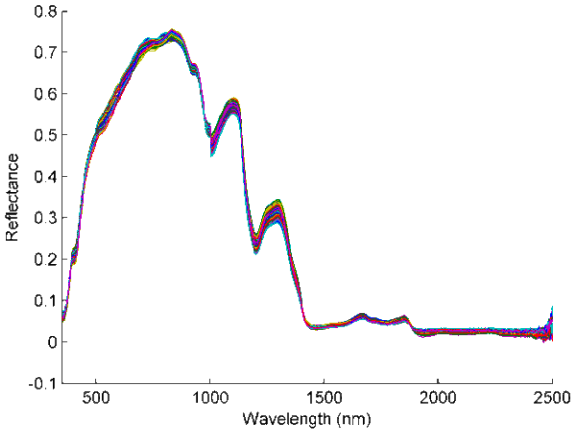

Fig. 4: Spectra of all samples preprocessed by MSC

separate the two classes, SVM can create a hyperplane that allows linear separation in the higher dimension feature space by using kernel function (Teye *et al*., 2013; Li *et al*., 2011).

Least Squares Support Vector Machine (LS-SVM): LS-SVM is an improved version based on the standard support vector machine. It has been successfully applied in many classification problems at present. LS-SVM works well based on the margin-maximization principle of performing structural risk minimization and it trains more easily than SVM (Wu *et al*., 2012). In this study, three crucial problems such as the selection of optimal input subset, appropriate kernel function and optimal kernel parameters would be resolved by using grid search and 10-fold cross validation methods (Gao *et al*., 2013). And the free LS-SVM toolbox (LS-SVM v1.5) with the MATLAB version was used to develop the calibration and prediction models.

The spectra characteristic of each rice variety: The raw spectra of all samples (350-2500 nm) were shown in Fig. 2 and the average spectrum of each rice variety ware shown in Fig. 3. From Fig. 3, it can be seen that, there were obvious differences among the spectra of the 5 rice varieties. Especially the differences among spectral data were greater at the peak of the spectral curve. Therefore, the 5 rice varieties can be classified according to hyperspectral data.

MSC preprocess: As the spectra may be affected by the inevitable noise resulting from the hardware and a great difference among the spectral data would be caused by the inhomogeneity of the rice samples and the scattering from the light. Therefore, it is necessary to use a suitable preprocess method to correct the raw spectra before the development of the models. In this study, MSC preprocess method was used to deal with the raw spectra and the processed spectral curve ware shown as Fig. 4.

RESULTS AND DISCUSSION

Preliminary:

Dimension reduction using PCA: As hyperspectral data provides much more information than general spectral data, the problem of huge data, noisy data and redundant data are more prominent during the procedure of hyperspectral data processing. In order to improve the processing efficiency and meet the online industrial application, a few of dimension reduction methods should be investigated. In this study, PCA was used to reduce the dimension of raw spectra of 5 rice varieties. The three-dimensional map of PC1, PC2 and PC3 of 5 rice varieties was shown in Fig. 5. Where, "1" expresses Ruan-Ya-Xiang-Si rice, "2" expresses Jiang-Su rice, "3" expresses Chang-Li-Xiang rice, "4" expresses Zhen-Zhu rice and "5" expresses Si-Miao rice. From Fig. 5, it can be clearly seen that 5 rice varieties can be well separated from each other in the three-dimensional map of PC1, PC2 and PC3. It indicated that the same result was obtained from PCA and the observation from spectral curve of 5 rice varieties.

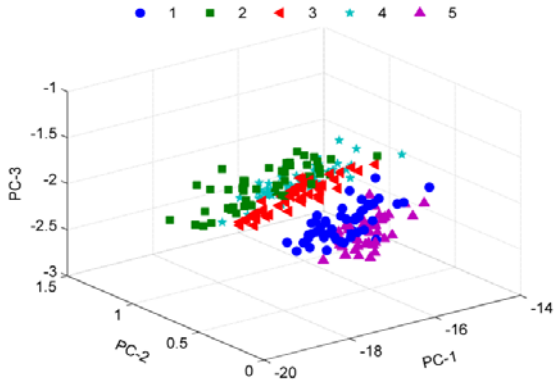

Fig. 5: Three-dimensional map of PCs of 5 varieties of spectra

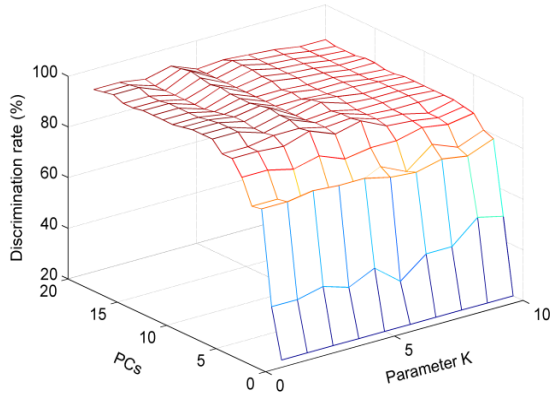

Fig. 6: Result of cross validation recognition rates in KNN model

Determination of PCs and model parameters: The parameters used in the model have great influence on the performance of the final discrimination models. Different parameters may lead to great difference for the same classification algorithm. Therefore, the parameters should be firstly determined in the calibration set. In this study, the 10-fold cross validation method was used to choose the optimal number of PCs and determined the K value of KNN and kernel function of SVM and LS-SVM. PCs from 1 to 20, K value from 1 to 10 in KNN, linear and RBF kernel function in SVM and LS-SVM were investigated respectively. The final parameters were determined according to the maximum cross validation discrimination rate. The cross validation results of KNN, SVM and LS-SVM were shown in Fig. 6 to 8 respectively. It can be seen from Fig. 6 to 8 the cross

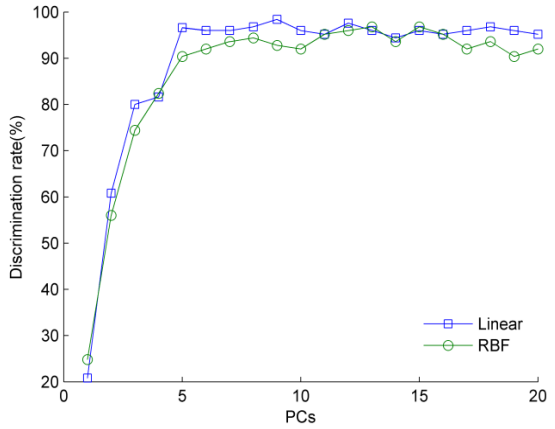

Fig. 7: Result of cross validation recognition rates in SVM model

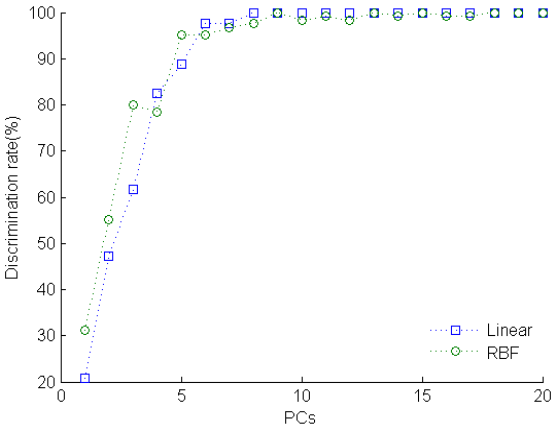

Fig. 8: Result of cross validation recognition rates in LS-SVM model

Table 1: Results of 3 discrimination models

Models	Kernel function	PCs	Cross validation accuracy	Prediction accuracy
KNN (K = 2)	-	8	94.4	89.6
SVM	Linear	8	96.8	93.6
LS-SVM	Linear	8	100	100

validation discrimination rate improved with the increasing of the PCs and when the PCs reached 8, the discrimination rate changed little. From Fig. 6, it can be found that the optimal PCs is 8 and the optimal K of KNN is 2. From Fig. 7, the optimal PCs of SVM is 8 and its kernel function is linear kernel function. Similarly, the optimal PCs of LS-SVM is also 8 and its kernel function is linear kernel function.

Analysis of three classification models in prediction set: All samples were divided into two parts. One hundred and twenty five rice samples (25 of each variety) were randomly chosen as calibration set and the remaining 125 rice samples were prediction set. According to the optimal PCs and parameters, three classification algorithms (KNN, SVM and LS-SVM) were respectively used to establish the discrimination models for rice varieties and the performance of three models were mainly evaluated by cross validation accuracy and prediction accuracy, the final results were shown in Table 1.

For the three models, KNN model is the lowest both in the cross validation accuracy and prediction accuracy, while the other two models are relatively high. Therefore, nonlinear models (SVM and LS-SVM) performed better than the linear model (KNN). Probably because there is some nonlinear relationship among the hyperspectral data, so it is hard to separate all the samples successfully only using linear algorithm. In addition, for the two nonlinear models, LS-SVM model performed better than the SVM model with the cross validation accuracy of 100% and prediction accuracy of 100%. This is because LS-SVM is the improved algorithm for SVM and more suitable for the hyperspectral data.

CONCLUSION

The FieldSpec®3 spectrometer was used to discriminate the rice varieties with the spectral range of 350-2500 nm. Principal Component Analysis (PCA) was used to reduce the dimension and move the noisy data in the hyperspectral data. Then using K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Least Square Support Vector Machine (LS-SVM) to develop three discrimination models for 5 rice varieties. Similarly, cross validation method was employed to determine the optimal number of Principal Component (PCs) and the parameters in the KNN, SVM and LS-SCM models. Based on the results of cross validation accuracy and prediction accuracy, KNN model give the lower performance than the SVM and LS-SVM models. In addition, LS-SVM models achieved the best performance with the accuracy of 100%. These results indicated that nonlinear classification algorithm performed better than the linear classification algorithm in the hyperspectral data of rice variety and LS-SVM can be used to develop the optimal discrimination model for rice variety.

ACKNOWLEDGMENT

This study is supported by National natural science funds projects (31101082), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) and Jiangsu University College Student Scientific Research Project (No: 13A541, No: 13A526).

REFERENCES

Barbin, D., G. Elmasry, D.W. Sun and P. Allen, 2012. Near-infrared hyperspectral imaging for grading and classification of pork. Meat Sci., 90(1): 259-268.

Cen, H.Y., Y. He and M. Huang, 2007. Combination and comparison of multivariate analysis for the identification of orange varieties using visible and near infrared reflectance spectroscopy. Eur. Food Res. Technol., 225(5-6): 699-705.

Gao, J.F., X.L. Li, F.L. Zhu and Y. He, 2013. Application of hyperspectral imaging technology to discriminate different geographical origins of *Jatropha curcas* L. seeds. Comput. Electron. Agr., 99: 186-193.

Ji-yong, S., Z. Xiao-Bo, Z. Jie-wen, M. Han-ping, W. Kai-liang, C. Zheng-wei and H. Xiao-wei, 2011. Diagnostics of nitrogen deficiency in mini-cucumber plant by near infrared reflectance spectroscopy. Afr. J. Biotechnol., 10(85): 19687-19692.

Li, P., G. Du, W. Cai and X. Shao, 2012. Rapid and nondestructive analysis of pharmaceutical products using near-infrared diffuse reflectance spectroscopy. J. Pharmaceut. Biomed., 70: 288-294.

Li, S.J., H. Wu, D.S. Wan and J. Zhu, 2011. An effective feature selection method for hyperspectral imaging classification based on genetic algorithm and support vector machine. Knowl-Based Syst., 24(1): 40-48.

Liu, L. and M.O. Ngadi, 2013. Detecting fertility and early embryo development of chicken eggs using near-infrared hyperspectral imaging. Food Bioprocess. Technol., 6(9): 2503-2513.

Peng, X.L., X. Li, X. Shi and S. Guo, 2014. Evaluation of the aroma quality of Chinese traditional soy paste during storage based on principal component analysis. Food Chem., 151: 532-538.

Pholpho, T., S. Pathaveerat and P. Sirisomboon, 2011. Classification of longan fruit bruising using visible spectroscopy. J. Food Eng., 104(1): 169-172.

Qin, J.W., K.L. Chao, M.S. Kim, 2012. Nondestructive evaluation of internal maturity of tomatoes using spatially offset Raman spectroscopy. Postharvest Biol. Tec., 71: 21-31.

Sarbu, C., R.D. Nascu-Briciu, A. Kot-Wasik, S. Gorinstein, A. Wasik and J. Namiesnik, 2012. Classification and fingerprinting of kiwi and pomelo fruits by multivariate analysis of chromatographic and spectroscopic data. Food Chem., 130(4): 994-1002.

Serranti, S., D. Cesare, F. Marini and G. Bonifazi, 2013. Classification of oat and groat kernels using NIR hyperspectral imaging. Talanta, 103: 276-284.

Sirisomboon, P., M. Tanaka, T. Kojima and P. Williams, 2012. Nondestructive estimation of maturity and textural properties on tomato 'Momotaro' by near infrared spectroscopy. J. Food Eng., 112(3): 218-226.

Teye, E., X.Y. Huang, H., Dai and Q. Chen, 2013. Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. Spectrochim. Acta A, 114: 183-189.

Wang, J., K. Nakano, S. Ohashi, Y. Kubota, K. Takizawa and Y. Sasaki, 2011. Detection of external insect infestations in jujube fruit using hyperspectral reflectance imaging. Biosyst. Eng., 108(4): 345-351.

Watanabe, K., I. Kobayashi, S. Saito, N. Kuroda and S. Noshiro, 2013. Nondestructive evaluation of drying stress level on wood surface using near-infrared spectroscopy. Wood Sci. Technol., 47(2): 299-315.

Wu, D., J. Chen, B. Lu, L. Xiong, Y. He and Y. Zhang, 2012. Application of near infrared spectroscopy for the rapid determination of antioxidant activity of bamboo leaf extract. Food Chem., 135(4): 2147-2156.

Zhang, X.C., J.Z. Wu and Y. Xu, 2012. Near Infrared Spectroscopy Technology and its Application in Modern Agriculture. Publishing House of Electronics Industry, Beijing, pp: 104-105.

Zhao, J.W., H. Lin, Q.S. Chen, X. Huang, Z.B. Sun and F. Zhou, 2010. Identification of egg's freshness using NIR and support vector data description. J. Food Eng., 98(4): 408-414.

Zhu, F.L., D.R. Zhang, Y. He, F. Liu and D.W. Sun, 2013. Application of visible and near infrared hyperspectral imaging to differentiate between fresh and frozen-thawed fish fillets. Food Bioprocess. Technol., 6(10): 2931-2937.