African Journal of Biotechnology Vol. 10(85), pp. 19687-19692, 28 December, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.557
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

Diagnostics of nitrogen deficiency in mini-cucumber plant by near infrared reflectance spectroscopy

Shi Ji-yong[1], Zou Xiao-bo[1]*, Zhao Jie-wen[1], Mao Han-ping[2], Wang Kai-liang[1], Chen Zheng-wei[1] and Huang Xiao-wei[1]

[1]School of Food and Biological Engineering, Jiangsu University, 301 Xuefu Road, 212013 Zhenjiang, Jiangsu, China.
[2]Key Laboratory of Modern Agricultural Equipment and Technology, 301 Xuefu Road, 212013 Zhenjiang, Jiangsu, China.

Accepted 9 June, 2011

In protected agriculture, deficiency of an essential element may drastically affect plant growth, appearance and most importantly yield. Information about nutrient deficiencies in crops grown in controlled environment is essential to optimize food productivity. In this study, near infrared reflectance spectroscopy (NIRS) analysis was used to identify nitrogen (N) deficiency coupled with pattern recognition methods in mini-cucumber plants grown under non-soil conditions. Leaves at the first three nodes of nitrogen deficient plants and control plant were used for NIRS data acquisition. K-nearest neighbors (KNN) and artificial neural network (ANN) were applied to build diagnostics models, respectively. Some parameters of the model were optimized by cross-validation. The performance of the KNN model and the ANN model based on NIRS data was compared. Experiment results showed that the ANN model was better than the KNN model. The optimal ANN model was achieved when principle component factors were equal to 5 and identification rate of the ANN model were 100% in both the training set and the prediction set. This study demonstrated that the NIRS coupled with ANN pattern recognition method can be successfully applied to the diagnostics of nitrogen deficiency in mini-cucumber plant grown under non-soil conditions.

Key words: Deficiency, nitrogen, near infrared reflectance spectroscopy (NIRS), artificial neural network.

INTRODUCTION

Due to the increasing costs of crop production and to the progressing environmental pollution by agroche-micals, mineral fertilizers should be applied more efficiently (Marian W et al., 2009). This concerns primarily N, because the over application of this element leads to low N recovery efficiency and to a risk of nitrate pollution of ground waters. Protected agriculture characterized by a fast growth rate, show high nutrient demand which must be satisfied on a rational basis (Bacci et al., 1998).

Conventional approaches to ascertaining plant nutrient content in plant tissues and quick field tests on plant sap status rely primarily on determining total elemental and fluids (Le Bot et al.,1998; Lemaire et al., 2008; Scarpeci et al., 2007). These methods are time consuming, tedious, expensive and involve destructive sampling of plant parts. A further disadvantage of conventional methods is that the interpretations of total elemental content in tissues do not generally include consideration of the distribution and form of the element within cells. Hence, simple analysis of total elemental content may be of limited value for evaluation of nutrient adequacy within plant tissues in some cases. The diagnostics of disease symptoms in plants, including those resulting from nutrient deficiencies require quick, reliable and precise instrumental techniques enabling to recognize the symptoms of physiological disorders prior to the occurrence of damages to plant growth and crop yield.

*Corresponding author. E-mail: zou_xiaobo@ujs.edu.cn. Tel: +86 511 88780174. Fax: +86 511 88780201.

Abbreviations: NIRS, Near infrared reflectance spectroscopy; KNN, K-nearest neighbors; ANN, artificial neural network; PCA, Principal components analysis.

Numerous studies involving a rapid estimating of the N requirement of corps have been carried out with non-invasive technologies. Netto et al. (2005) studied the color of robusta (*Cofea canephora* Pierce ex Froehner) leaves and noted the occurrence of a significant linear depen-dence between the N content of leaves and readings on a chlorophyll meter. Miguel et al. (2008) presented a low-cost method for analysing barley N-nutrition status by analysing the colour of leaf images captured with a conventional digital camera and also compared the methodology with the conventional use of SPAD-502 measurements. Wiwart et al. (2008) analyzed changes in the color of the first three leaves of faba bean (*Vicia faba* L.), pea (*Pisum sativum* L.) and yellow lupine (*Lupinus luteus* L.) plants under conditions of nitrogen (N), phosphorus (P), potassium (K) and magnesium (Mg) deficiencies by digital color image technology. Mistele et al. (2008) found spectral measurements were useful to describe the NNI in winter wheat, whether or not the canopy was deprived or adequately supplied with N in particular.

The information about the N status of crop stands by using spectral reflectance measurements is useful to support nitrogen fertilizer application within a precision farming, especially because the tractor-mounted measurements allowed N status to be estimated quickly and non-destructively. GarcõÂa-Ciudad (1998) used near infrared reflectance spectroscopy to estimate nitrogen content in grasses from semiarid grasslands and the results showed that the calibration equation is robust enough for an accurate prediction of the nitrogen concentration in samples of grasses collected in different years.

Near-infrared reflectance spectroscopy (NIRS) has been developed and applied to the estimation of chemical composition in plant samples (Petisco et al., 2005, Shi et al., 2010). NIRS overcomes some of the drawbacks of conventional methods used for a mineral analysis in plant samples. The measurement by NIRS is rapid, nondestructive and it is a multi-element technique because one measurement of the sample is sufficient to estimate many parameters. Several works have demonstrated that NIR spectra are correlated with macronutrients such as P, calcium and sulfur (Petisco et al., 2005). NIRS can be used to determine the concentrations of certain cations, owing to their association with organic or hydrated inorganic molecules as well as for feed identification and authenti-cation (Givens et al., 1999). Hence, there are good reasons to use near infrared reflectance spectroscopy analysis in the early diagnostics of physiological symp-toms cause by various stress factors.

In this study, near infrared reflectance spectroscopy was attempted to identify nitrogen deficiency in mini-cucumber plant coupled with pattern recognition methods. In building identification model, K-nearest neighbors (KNN) and artificial neural network (ANN) were attempted comparatively and some parameters of model are optimized by cross-validation.

MATERIALS AND METHODS

Plants

Leaves of cucumber (*Cucumis sativus*, Biyu-3, FuNong seeds Co. Ltd., Shanghai, China) were taken for the experiments. All investigated plants were grown under non-soil conditions (perlite rock) in greenhouses at Jiangsu University in China from May 5 to July 16, 2009 (32.11°N, 119.27°E). Plants were spaced in double rows with a planting density of 20 cm × 20 cm. Following seed sowing and germination, one plant was left per pot. Two treatments were analyzed: no N fertilization (-N) and control (fertilization with all macronutrients). The composition of the complete nutrient solution was: NO_3 11.75 mmol l^{-1}, NH_4 1.0 mmol l^{-1}, H_2PO_4 1.25 mmol l^{-1}, K 6.5 mmol l^{-1}, Ca 2.75 mmol l^{-1}, Mg 1.0 mmol l^{-1}, SO_4 1.0 mmol l^{-1}; Fe 15 µmol l^{-1}, Mn 10 µmol l^{-1}, Zn 5 µmol l^{-1}, B 25 µmol l^{-1}, Cu 0.75 µmol l^{-1}, Mo 0.5 µmol l^{-1}. In the control group (Ctrl), the plants continued to be maintained in complete nutrient solution where they were supplied with all the essential nutrient elements. In the -N group, the nutrient solution was formulated to eliminate NO_3 and NH_4 without changing the concentrations of the other essential elements. The nutrient solutions were supplied at rates of 500 ml/plant/day for 2 weeks after transplanting and of 800 to 1000 ml/plant/day for the rest 2 weeks until harvesting. The nutrient solution was provided to each plant using a trickle nozzle. Drainage water was collected and reused.

Recommended pesticides (carbendazim, sulphur colloidal suspension agent) were used to control insects as needed. One week after the plants blooming, 15 leaves at the first 3 nodes per day were randomly taken for measurements. The measurement process was conducted over a 7 day period. In total, 105 leaves were taken into account for nutrient diagnostics.

Spectra collection

The NIR spectra were collected in the reflectance mode using the Antaris near infrared spectrophotometer (Thermo Electron Co., USA) with an integrating sphere. Each spectrum was the average spectrum of 32 scans. The range of spectra is from 10,000 to 4000 cm^{-1} and the data were measured in 1.928 cm^{-1} intervals, which resulted in 3112 variables.

Each sample was collected three times at three different points of leaf surface. The spectral information was for the upside of the leaves. The average of the three spectra, which were collected from the same leaf sample, was used in further analysis. The temperature was kept around 25°C and the humidity was kept at a steady level in the laboratory.

Spectra preprocessing

Standard normal transformation (SNV), as a mathematical transformation method (Hu et al., 2010; Moghimi et al., 2010), was applied to process NIR spectra. Figure 1 shows NIR spectra of leaves after SNV preprocessing.

Basic principle of KNN

In KNN classification, an unknown sample of the prediction set is classified according to the majority of its K-nearest neighbors in the training set (Candolfi et al., 1998; Chen et al., 2008). Parameter K had a great influence on the identification rate of KNN model and

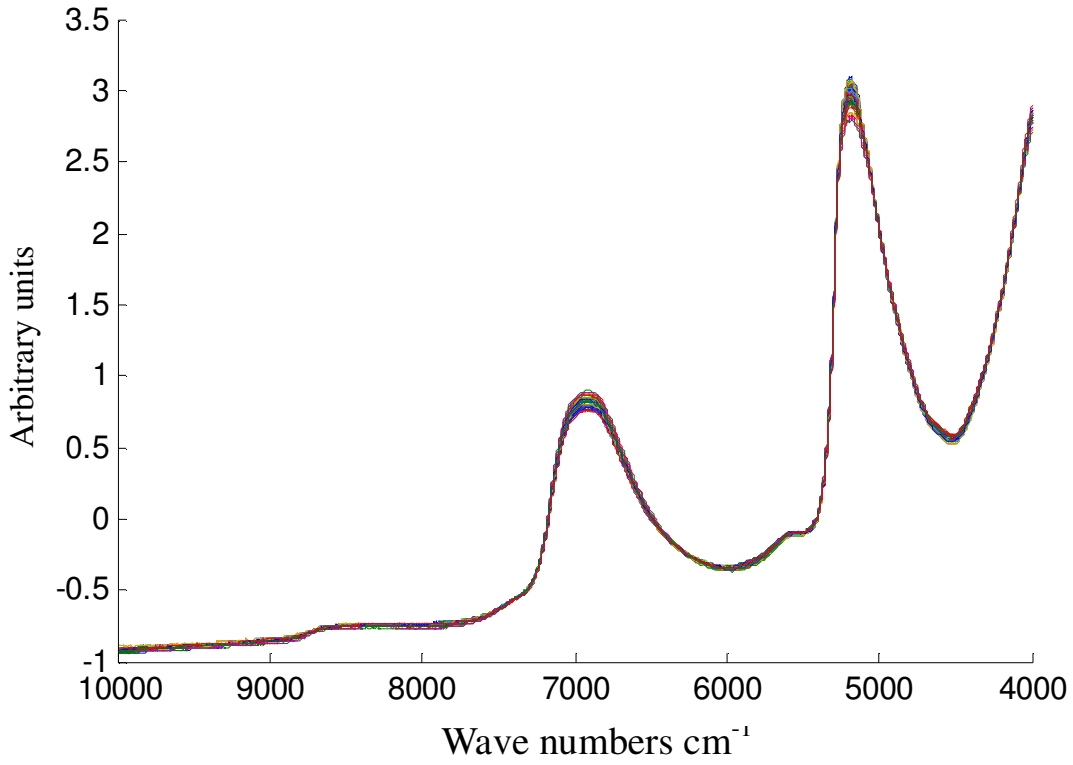

Figure 1. Spectra of leaf samples after SNV preprocessing.

optimal values for these parameters were selected in the calibration process. KNN parameters were optimized with the minimum prediction error estimated by cross-validation in calibration set. The value of K which gives the lowest error rate was selected.

Basic principles of ANN

The algorithm of ANN is able to capture and represent complex relationships between inputs and outputs (Dou et al., 2005). In this study, the ANN was employed as one of the calibration methods for comparison. Back propagation artificial neural network, as a classical feed-forward multilayer networks consisting of neurons arranged in layers was used to build models. The selected PCs as inputs were processed by neural network. The output expressed the resemblance that an object corresponded with a training pattern. After the adjustments of the nodes of hidden layers and other parameters, the ANN models could be developed. Three layers (that is an input layer, a hidden layers and an output layer) of ANN were arranged. The optimal PCs determined were used as the input layer and output layer was the classification labels of -N leaves and control leaves. Hidden nodes of ANN were optimized by 'trial and error' and optimal of hidden nodes was evaluated by the minimal mean square error (MSE) value (Zhao et al., 2006). The learning rate factor and momentum factor were all set to 0.1; the initial weights were 0.3; scale function was the 'tan h' function. The permitted training error was set as 0.002 and the maximal time of training was 20,000.

Software

All algorithms were implemented in Matlab V7.0 (Mathworks, USA) under Windows XP. Result software (Antaris System, Thermo Electron Co., USA) was used in NIR spectral data acquisition.

RESULTS AND DISCUSSION

Principal components analysis (PCA)

All wave numbers were used in PCA; Figure 2 shown a 3-dimensional (3D) space of leaf samples represented by PC1, PC2 and PC3. PC1 intercepted 80.71% variances, PC2 15.35% variances and PC3 2.73% variances. Geometrical exploration based on PCA score plots gave the clusters trend in the 3-dimension space, but cannot be used directly as a tool for discriminating leaves. Through PCA, the accumulated variance contribution rate was up to 99.51% for the top five PCs. In other words, PCs 1 to 5 could load most of the whole spectral information. Therefore, the top five principal components vectors were extracted by PCA. These vectors were inputted into the KNN/ANN classifiers as latent variables. Figure 2 shown that there was a neat separation of two groups in the 3D space represented by PC1, PC2 and PC3. Such classification trend in this 3D space could be explained by the chemical background of mini-cucumber leaves and PCA methods. Leaf can exhibit considerable differences in their own chemical characteristics according to different nitrogen nutrient level. The differences from chemical characteristics of leaf can be reasonably differentiated in the NIR spectroscopy. Therefore, NIR

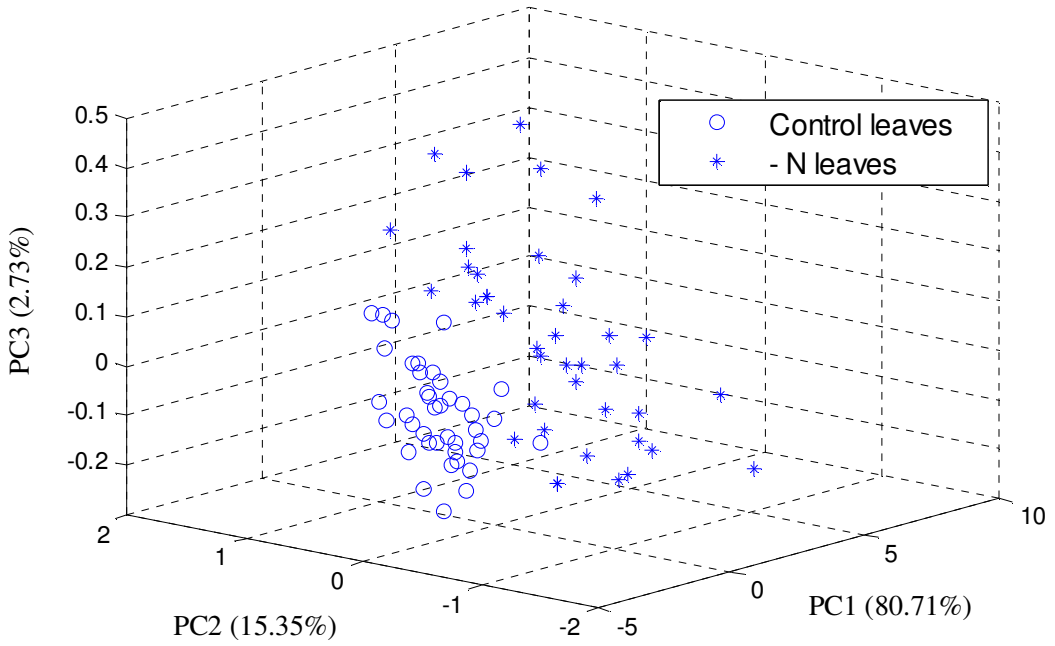

Figure 2. Score cluster plot of the top three principal components (PCs) for all samples.

spectroscopy data can exhibit the cluster trend of leaf samples according to nitrogen level by means of PCA.

Identification results of the model

Nutrient exploration based on PCA score plots only gave the clusters trend in the 3-dimension space but can not be used as a classification tool. The ultimate aim of this study was to classify two nitrogen level of leaf by NIRS and supervised pattern recognition method. Supervised pattern recognition refers to techniques in which a priori knowledge about the category membership of samples was used for classification. The identification model was developed on a training set of samples with categories (Zou et al., 2010; Wu et al., 1996a, b). The model performance was evaluated by the use of a prediction set. In this study, 105 samples were divided into two subsets. One of the subsets is called the training set which was used to build model and the other is called the prediction set which was used to test the robustness of model. The training set contained 80 samples and each grade had 40 samples. The remaining 25 samples constitute the prediction set and nitrogen deficient grade also had 15 samples.

Supervised pattern recognition methods are numerous and the main problem is to choose the most appropriate method. In this study, linear and non-linear supervised pattern recognition methods (KNN, BP-ANN) were applied comparatively. The identification rate (%) is used to evaluate the performance of the identification model as an important criterion. It was calculated according to:

$$R = \frac{N_1}{N_2} \times 100 \quad \% \tag{1}$$

Where, R was the identification rate (%) in the training set or prediction set; N_1 was the number of the correctly identified samples in the training set or prediction set; N_2 was the number of all samples in the training set or prediction set. Some principal components vectors were extracted from NIRS data as the inputs of the identification model. The number of principal components factors (PCs) was crucial to the performance of the identification model. Therefore, the number of PCs should be optimized in building an identification model.

Identification results of KNN model

KNN was a non-parametric method. An unknown sample of the prediction set was classified according to the majority of its K-nearest neighbors in the training set (Candolfi et al., 1998). Parameter K had a high influence on the identification rate of KNN model and the optimum K number was often determined by cross-validation procedure (Chen et al., 2008). The number K that gave the lowest error rate was chosen. At the same time, the number of principal component factors (PCs) also should be optimized. Therefore, in this study, five K values (K = 1, 3, 5, 7 and 9) and 9 PCs (PCs = 1, 2, to 9) were tested simultaneously by a cross-validation. The identification results of cross-validation influenced by the number of PCs and the K value were shown in Figure 3. As seen from

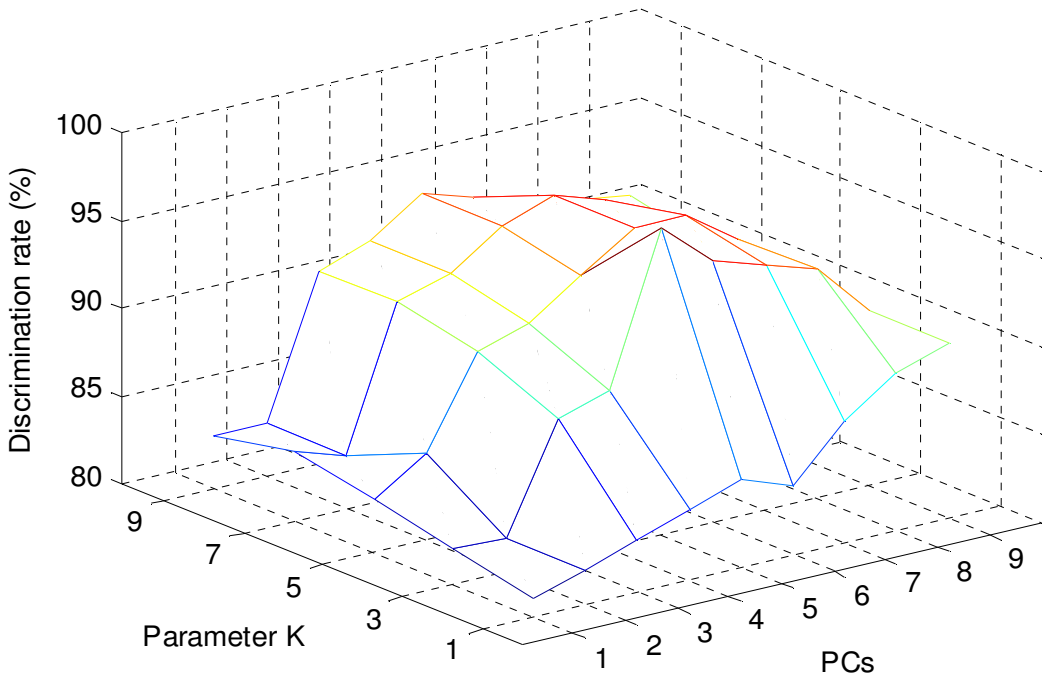

Figure 3. Discrimination rates of KNN model optimized by cross-validation under different parameter K and PCs.

Table 1. Comparison of the identification results from KNN and BP-ANN models.

Model	Optimal PCs	Identification result of model	
		Training set (%)	Prediction set (%)
KNN	5	97.5	96
BP-ANN	5	100	100

Figure 3, the optimal KNN was achieved when K = 3 and PCs = 5. The identification rate by cross-validation was 97.5%. Here, only two nitrogen deficient samples were identified wrongly to the control group and the remaining samples are all identified correctly. In the prediction set, the identification rate was 96% and one nitrogen deficient sample was identified wrongly to the control group.

Identification results of ANN model

Considering that the linear model did not provide a complete solution to the classification problem, non-linear approach such as artificial neural networks (ANN) was used. Some parameters including the number of neurons, scale functions, learning rate factor, momentum factors and initial weights had some influence on the performance of ANN model to some extent (Zhao et al., 2006). These parameters were optimized by an experiment in this study as follows: three layers (that is, an input layer, a hidden layers and an output layer) of BP-ANN were arranged; the number of neurons in the hidden layer was set as 5, the learning rate factor and momentum factor were all set to 0.1; the initial weights were set to 0.3; scale function is set as the 'tan h' function.

The number of PCs was also optimized in this study. The optimal BP-ANN model is obtained when 5 PCs were used. The identification rates of BP-ANN model were all 100% in the training set and prediction set, respectively; in other words, all samples were identified correctly.

Discussion of identification results

Table 1 compared the identification results from KNN and BP-ANN models. Their optimal numbers of PCs were all 5, the identification rates of two models were 97.5 and 100%, respectively, in training set and the identification rates of two models were 95 and 100% in prediction set. BP-ANN model was a little better than KNN model. It indicated that the identification results from the non-linear pattern recognition method were better than the linear pattern recognition method, because the ability of

self-learning and self-adjusting from non-linear method was stronger than linear method.

Conclusion

Near infrared reflectance spectroscopy was used to diagnose nitrogen deficiency correlated with an appropriate pattern recognition methods in this study. K-nearest neighbours (KNN) and artificial neural network (ANN) were used comparatively to build identification models using near infrared reflectance spectroscopy data in the experiment; experimental results shown that the performance of ANN model was a little better than KNN model. The optimal ANN model was achieved when principal components factors equaled to 5. Identification rates of model were all 100% for both training set and prediction set.

It can be concluded that near infrared reflectance spectroscopy technique had high potential to diagnose nitrogen nutrient of mini-cucumber plant. Therefore, a simple, rapid and reliable overall characterization of nitrogen nutrient may be obtained at a low cost. In comparison to time-consuming chemical methods, the results obtained by near infrared reflectance spectroscopy represents a considerable improvement in diagnostic of nitrogen deficiency.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support provided by the foundations of NSFC (Grant no. 6091079), Chinese 863 Program (Grant nos. 2008AA10Z208 and 2011AA1008047), the foundation of Chinese top 100 doctoral dissertation (200968), postdoctoral foundation (0601003C, 20070411024), program sponsored for scientific innovation research of college graduate in Jiangsu province (CX10B_277Z) and the talent foundation of Jiangsu University.

REFERENCES

Bacci L, De Vincenzi M, Rapi B, Arca B, Benincasa F (1998). Two methods for the analysis of colorimetric components applied to plant stress monitoring. Comput. Electron Agr. 19(2): 167-186.

Candolfi A, Wu W, Massart DL, Heuerding S (1998). Comparison of classification approaches applied to NIR-spectra of clinical study lots. J. Pharmaceut. Biomed. 16(8): 1329-1347.

Chen QS, Zhao JW, Vittayapadung S (2008). Identification of the green tea grade level using electronic tongue and pattern recognition. Food Res. Int. 41(5): 500-504.

Dou Y, Sun Y, Ren YQ, Ren YL (2005). Artificial neural network for simultaneous determination of two components of compound paracetamol and diphenhydramine hydrochloride powder on NIR spectroscopy. Anal. Chim. Acta. 528(1): 55-61.

Givens DI, Deaville ER (1999). The current and future role of near infrared reflectance spectroscopy in animal nutrition: a review. Aust. J. Agric. Res. 50 (7):1131-1145.

Hu Y, Erxleben A, Ryder AG, McArdle P (2010). Quantitative analysis of sulfathiazole polymorphs in ternary mixtures by attenuated total reflectance infrared, near-infrared and Raman spectroscopy. J. Pharmaceut. Biomed. 53(3): 412-420.

Le Bot J, Adamowicz S, Robin P (1998). Modelling plant nutrition of horticultural crops: a review. Sci. Hortic-Amsterdam, 74(1-2): 47-82.

Lemaire G, Jeuffroy M, Gastal F (2008). Diagnosis tool for plant and crop N status in vegetative stage: Theory and practices for crop N management. Rur. J. Agron. 28(4): 614-624.

Marian W, Gabriel F, Krystyna Z, Elzbieta S (2009). Early diagnostics of macronutrient deficiencies in three legume species by color image analysis. Comput. Electron. Agric. 65(1): 125-132.

Moghimi A, Aghkhani MH, Sazgarnia A, Sarmad M (2010). Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. Biosyst. Eng.106(3): 295-302.

Petisco C, García-Criado B, de Aldana BRV, Zabalgogeazcoa I, Mediavilla S, García-Ciudad A (2005). use of near-infrared reflectance spectroscopy in predicting nitrogen, phosphorus and calcium contents in heterogeneous woody plant species. Anal. Bioanal. Chem. 382(2): 458-465.

Scarpeci TE, Marro ML, Bortolotti S, Boggio SB, Valle EM (2007). Plant nutritional status modulates glutamine synthetase levels in ripe tomatoes (Solanum lycopersicum cv. Micro-Tom). J. Plant. Physiol. 164(2):137-145.

Shi JY, Zou XB, Yin XP, Chen ZW (2010). A new method of characteristic Wavelength Sub-Range Selection of Near infrared Spectroscopy.Spectrosc spec Anal.30(12): 3199-3202.

Wu W, Walczak B, Massart DL, Heuerding S, Erni F, Last IR, Prebble KA (1996). Artificial neural networks in classification of NIR spectral data: Design of the training set. Chemometr. Intell. Lab. 33(1): 35-46.

Wu W, Walczak B, Penninckx W, Massart DL (1996). Feature reduction by Fourier transform in pattern recognition of NIR data. Anal. Chim. Acta. 331(1-2): 75-83.

Zhao JW, Chen QS, Huang XY, Fang CH (2006). Qualitative identification of tea categories by near infrared spectroscopy and support vector machine. J. Pharmaceut. Biomed. 41(4): 1198-1204.

Zou XB, Zhao JW, Malcolm PJW, Mel H, Mao H (2010). Variables selection methods in near-infrared spectroscopy. Anal Chim Acta.667(1-2):14-32.