http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (19): 4845-4851, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.4845.4851

Detecting Nitrogen Content in Lettuce Leaves Based on Hyperspectral Imaging and Multiple Regression Analysis

[1,2]Sun Jun, [2]Jin Xiaming, [3]Mao Hanping, [2]Wu Xiaohong, [3]Gao Hongyan, [3]Zhu Wenjing and [2]Liu Xiao
[1]Key Laboratory of Agri-informatics, Ministry of Agriculture, Beijing, 100081 P.R.China
[2]School of Electrical and Information Engineering of Jiangsu University,
[3]Laboratory Venlo of Modern Agricultural Equipment, Jiangsu University, Zhenjiang, 212013, P.R. China

Abstract: This study was carried out to detect nitrogen content in lettuce leaves rapidly and non-destructively using visible and near infrared (VIS-NIR) hyperspectral imaging technology. Principal Component Analysis (PCA) was performed on the average spectra to reduce the spectral dimensionality and the principal components (PCs) were extracted as the input vectors of prediction models. Partial Least Square Regression (PLSR), Back Propagation Artificial Neural Network (BP-ANN), Extreme Learning Machine (ELM), Support Vector Machine Regression (SVR) were, respectively applied to relate the nitrogen content to the corresponding PCs to build the prediction models of nitrogen content. R^2_p of the PLSR model for nitrogen content was 0.91 and RMSEP was 0.32. BP model of structure 5-2-1 with R^2_p of 0.92 and RMSEP of 0.21, ELM model of structure 5-10-1 with R^2_p of 0.95 and RMSEP of 0.19 and SVR model for nitrogen with R^2_p of 0.96 and RMSEP of 0.18, all got good prediction performance. Compared with the other three models, SVR model has the better performance for predicting nitrogen content in lettuce leaves. This work demonstrated that the hyperspectral imaging technique coupled with PCA-SVR exhibits a considerable promise for nondestructive detection of nitrogen content in lettuce leaves.

Key words: Hyperspectral imaging, lettuce leaves, nitrogen content

INTRODUCTION

Nitrogen is the most crucial element in different physiological processes of crops. Nitrogen nutrient deficiency would affect crop growth and lead to poor growth quality. Over-application of nitrogen fertilizer would lead to low nitrogen absorption efficiency and environment pollution. Suitable amount of nitrogen fertilizer can promote the high yield and high quality of crops, so the detection of nitrogen status in crop in time is the premise of the suitable nitrogen fertilizer.

Conventional approaches to detecting nutrient content in crop rely primarily on chemical measurement method in laboratory which are time-consuming, tedious, expensive and involve destructive sampling of crop parts. In recent years, Nondestructive Detection Technology (NDT) is applied widely, by which the internal nutrient elements in crop can be detected without destroying crops organization.

In NDTs, spectroscopy technology has been widely used in estimating nitrogen content of crops. Several scholars studied the correlation between crop nitrogen level and spectral reflectance. Shi J.Y. *et al.* used near Infrared Reflectance Spectroscopy (NIRS) analysis to identify nitrogen deficiency coupled with pattern recognition method (Shi *et al.*, 2011£©.Tian Y.C. *et al* made a systematic analysis on the quantitative relationship between leaf nitrogen concentrations and different hyperspectral vegetation indices with multiple field experiments under varied nitrogen rates and varied types in rice (Tian *et al.*, 2011).Zhang Y.S. used the canopy reflectance spectra at multiple wavelengths to predict the nitrogen content of rice and wheat leaf (Zhang *et al.,* 2010).Although the above literatures have all achieved good prediction precision, there is a large disparity from the practical application. Because the spectroscopic method has a great drawback that it acquires the spectral data from a single point or from a small portion of the tested sample it can't show the plentiful characteristic information of crop nutrient status, so the error of estimating result is great inevitably.

In NDTs, imaging technology has also been applied in N detection of crops. Graeff *et al* used multi-spectral images in different wavebands to study the nitrogen deficiency recognition of rice and found that the images of wavebands in 380-390 nm'516-IR'516-780 '430-780 nm' 540-600 nm can be used to identify crop nutrient (Graeff *et al.*, 2001). Yang, W. and Nick, S. applied

Corresponding Author: Sun Jun, School of Electrical and Information Engineering of Jiangsu University, Zhenjiang 212013,China

CCD camera with optical filter method to detect nutrient elements of cucumber leaves based on multi-spectral images and the results showed that this detection method was feasible and there was high linear relationship between leaf nitrogen content and vegetation index, leaf area of multi-spectral images (Yang *et al.,* 2010).It can be seen from the results of these above literatures that, there are considerable error in the prediction results, because the imaging technology is based on several wavebands information and information source is not overall. However, hyperspectral imaging technique combines spectroscopy and imaging technique to acquire both spectral and spatial information from objects and it has been applied in nondestructive detection in food quality widely, including strawberry (ElMasry *et al.*, 2007). pork (Barbin *et al.*, 2012).

In recent years, hyperspectral imaging technology has been developed as a useful tool for determining internal and external attributes of crops. Zhang XL *et al* used hyperspectral imaging technology to provide N, P and K concentration information to show the nutrient distribution in oilseed rape leaves (Zhang *et al.,* 2013). Shi J.Y. *et al* used Chlorophyll concentration distribution map of cucumber leaf given by near infrared hyperspectral imaging to determine nitrogen deficiency (Shi, J.Y. *et al.*, 2012). As a daily food crop, lettuce is popular with the public. However, there were little articles about nitrogen detection in lettuce based on hyperspectral imaging technology reported. In this work, lettuce was chosen as research object. Using hyperspectral imaging technology, combined with multiple regression methods, we attempted to establish regression models of nitrogen content in lettuce leaves.

MATERIALS AND METHODS

Lettuce leaf samples preparation: All investigated lettuces (Italy annual bolting seeds Co.Ltd., Suqian, China) were cultivated under non-soil conditions (perlite rock) in greenhouse at Laboratory Venlo of Modern Agricultural Equipment in Jiangsu University (Zhenjiang of China, 32.11°N,119.27°E). After seeds were sown and bud, one plant was left per pot. The nutrient solution was config\ured in accord with Japan Yamazaki nutrient solution formula configuration. In order to cultivate lettuce samples in different nitrogen levels, from transplanting, the samples were irrigated using the nutrient solution of different nitrogen concentration between 25 and 150% (namely 25, 50, 75, 100, 150%) of the nitrogen concentration in standard nutrients solution respectively, simultaneously other nutrient element in the normal content.

At the rosette stage, one spotless lettuce leaf, in same leaf position with full leaf mesophyll, was picked randomly from each plant. There were 12 pieces of leaves from a group, so a total of 60 samples were acquired from the 60 plants in this experiment. After being picked off, the lettuce leaf samples were immediately sealed in plastic bags and carried to laboratory for image collection.

Hyperspectral imaging system: It consists of an imaging spectrograph with spectral resolution of 2.8 nm (ImspectorV10, Spectral Imaging Ltd., Finland), a CCD video camera with image resolution of 672*512 and wavelength range from 350 to 1050 nm, a mobile platform used for moving lettuce leaf samples, a computer supported with Spectral-cube data acquisition software (Spectral Imaging Ltd., Finland) to control image acquisition, an illumination unit consisting of two 150w fiber optical halogen lamps (2900 version, Illumination Technologies, USA).The hyperspectral imaging system can record a whole line of an image rather than a single pixel at a time.

Extraction of spectra features: PCA is a dimension reduction method which has been commonly used in many fields. It can make the data space map to a low dimensional subspace by orthogonal transform, let less new variables replace original variables and at the same time make the new variables retain as much as possible the original information. So PCA was used to reduce dimensions of mean spectra for excluding the redundant information in this paper.

Measurement of nitrogen content: When the hyperspectral images of lettuce leaf samples were captured, the samples were sent to laboratory quickly to be weighed, steamed, drought, for chemical analysis. The total nitrogen content of lettuce leaves sample were determined by Kjeldahl method.

Partial least squares regression (PLSR): It is a regression method available for multivariate calibration method. PLSR technique is particularly useful when it is necessary to predict a set of dependent variables from a large set of independent variables. It will set up linear relationship between spectra (X) and parameters under investigation (Y). Both X and Y matrices are transformed into new spaces and the obtained data called X scores and Y scores are then carefully selected and correlated in an attempt to maximize the interpretation of Y scores by X scores. The predicted Y scores are subsequently employed to produce the prediction of Y. The following simple function can be used to feature the best linear relationship between X and Y (Feng and Sun, 2013):

$$Y = x\beta+\varepsilon$$

where, β is the regression coefficients and ε is the prediction error.

Support vector regression (SVR): The regression of support vector machine is a good regression method. Ingenious application of kernel function is the essence of SVR algorithm and proper kernel function can project the complex practical problems to high-dimensional feature space through nonlinear transformation. A linear decision function was constructed in feature space to solve the complex nonlinear problems. At present, the common kernel functions are linear kernel function, polynomial kernel function, radial basis function, Sigmoid kernel function. SVR will eventually translate original problem into a two programming problems and in theory, the global optimal solution can be obtained (Ding *et al.*, 2011).

Back propagation (BP) algorithm: Back Propagation artificial neural network (BP-Net) consists of input layer, hidden layer and output layer. S-type transfer function was used by hidden layer neurons and hidden nodes were evaluated by the Minimal Mean Square Error (MSE) value. Linear transfer function was used by output layer neurons. It is proved in theory that, when the number of hidden layer neurons is enough large, a three-layer BP network can approximate any nonlinear function with finite point of discontinuity.

ELM(Extreme Learning Machine) algorithm: It is a new algorithm of feed forward neural network with single hidden layer. The connection weights between input layer and hidden layer and the threshold value of hidden layer neurons were generated randomly and they needn't be adjusted in the training process. Simply set up the number of hidden layer neurons, we can get the unique optimal solution. Compared with the traditional training methods, ELM algorithm has the advantages of fast learning speed and good generalization performance.

RESULTS AND DISCUSSION

Spectral extraction: The spectral data directly affects the performance of model. The Environment for visualizing images software (ENVI) was used to select 3 regions of interest (ROIs) with 50*50 in which every pixel corresponds to the spectral reflectance of 390-1050nm wavelength value, avoiding the vein of lettuce leaves. In order to avoid the influence of random noise, the average spectrum of all pixels in ROI was calculated as the spectral reflectance value of each leaf sample. The flow chart of main steps for extracting spectral data is shown as Fig. 1.

The above steps were operated for all lettuce leaf samples one by one and the spectral value of all 60 lettuce leaves were calculated eventually, their spectral curves were shown as in Fig. 2.

Principal component analysis: As shown in Fig. 2, the original spectral range is from 350 to 1050 nm with a total of 512 wavelengths. Because original spectra have too large amount data and high redundancy, the Principal Component Analysis (PCA) was used to deal with original spectra to reduce the dimensionality of original spectra. The results of PCA were shown in Table 1.

Modeling test: There were a total of 60 lettuce leaf samples in sample library, 40 samples of which were selected as calibration set to build a multiple regression model and the remaining 20 samples were used as the test set to evaluate the performance of modeling.

In order to build a prediction model of nitrogen content in lettuce leaves with a strong stability and high prediction precision, PLSR, SVR, BP-ANN, ELM were

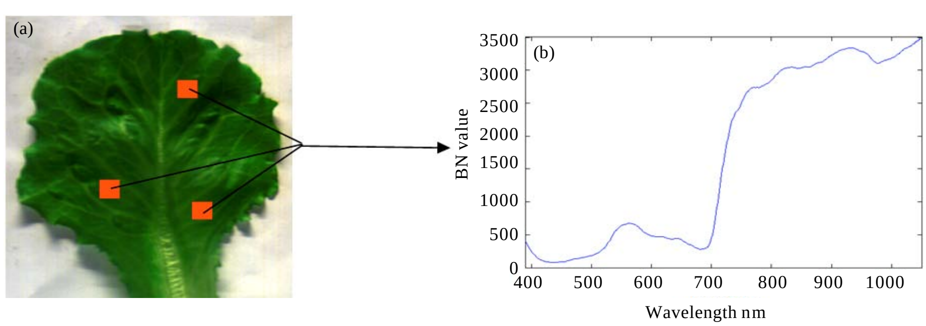

Fig. 1: Spectral extraction from the ROI of lettuce leaf sample

Table 1: Results of PCA

Principal components	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Contribution Ratio (%)	47.13	40.01	10.22	1.05	0.93	0.28	0.13	0.12	0.05	0.03
Cumulative Contribution Ratio (%)	47.13	87.14	97.36	98.41	99.33	99.61	99.74	99.86	99.91	

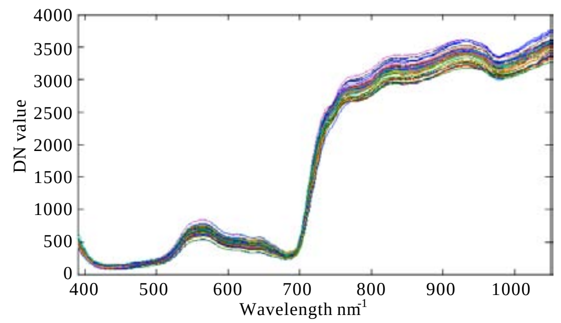

Fig. 2: Spectral curve of all lettuce leaves

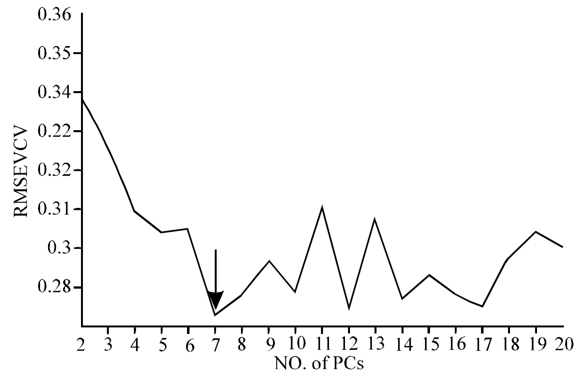

Fig. 3: Performances of PLSR models using different number of PCs

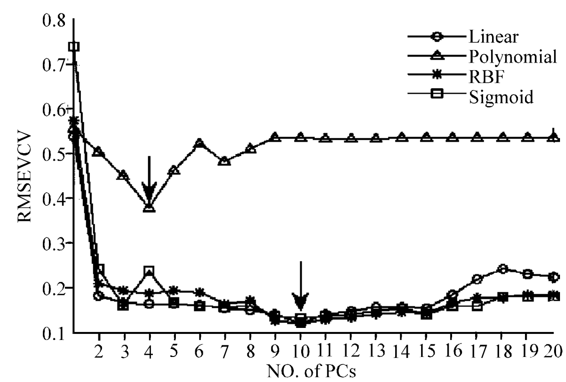

Fig. 4: Performances of SVR models using different kernel functions and different number of PCs

used to establish the relationship between hyperspectral image information of lettuce leaves and their nitrogen contents. The determination coefficients and root mean square errors of these 4 models were compared with each other, in order to find the optimal model for predicting nitrogen content of lettuce leaves.

Determination of the number of PCs and parameters: The stability and accuracy of multivariate regression models are mainly affected by the input variables and model parameters. In this work, the principal component analysis was used to reduce the dimensionality of the original data. So it is essential to judge the optimal dimensionality (the number of PCs) and effective parameters.

The number of principal component was set as from 1 to 20 one by one, based on which the calibration models were established by using PLSR, SVR, BP-ANN, ELM algorithms respectively. At the same time, 10-fold-Cross-Validation was used to determine the optimal PCs and parameters. And the parameters of each model are set as below.

In PLSR models, Latent factors are the most important to the performance of PLSR model, In this study, the optimal number of latent variable is determined corresponding to the minimum of PRESS value. Latent factors will vary with the number of PCs. The performances of PLSR models with different number of PCs are shown in Fig. 3.

The performance of SVR models is mainly affected by the kernel function, penalty factor C and gamma parameter. In this work, the common four kernel functions such as linear kernel function, polynomial kernel function, RBF kernel function, Sigmoid kernel function were respectively used to build models and the grid searching and cross validation were used to search the optimal parameter C and kernel function. The performances of SVR model with different number of PCs based on four kernel functions are shown in Fig. 4.

In BP-ANN model, the transfer function of hidden layer is set as tansig function, the transfer function of output layer as purelin function, the training times as 500, expected error as 0.001, the learing rate as 0.1.

ELM model was built only with determined network activation function and the number of hidden layer nodes. In this work, Sigmoidal function was used as excitation function of ELM model.

The performances of BP-ANN model and ELM model were mainly affected by the number of hidden layer nodes. Therefore, with the number of hidden layer nodes arising from 1 to 20, the performances of corresponding models were discussed, respectively. The predictive results of BP-ANN and ELM models were not fixed each time and their randomness is relatively large. So the results from Fig. 5 and 6 were the mean results of 10 tests.

From Fig. 3, 4, 5 and 6 it can be seen that, when the number of PCs is small, the root mean square errors estimated by 10-fold-Cross-Validation (RMSECV) of the four models are all large and all the prediction performances of every model are poor. Because the less principal components cannot contain larger information of original data. With the number of PCs increasing, the performances of model are constantly changing with a trend of better first and worse later. It is because that redundancy of information increased gradually, affecting the prediction performance of models.

Finally, the optimal PCs in PLSR model was 7. The optimal PCs of SVR models based on linear kernel function, ploynomial kernel function, RBF kernel function, Sigmoid kernel function respectively were 10, 4, 10, 10.

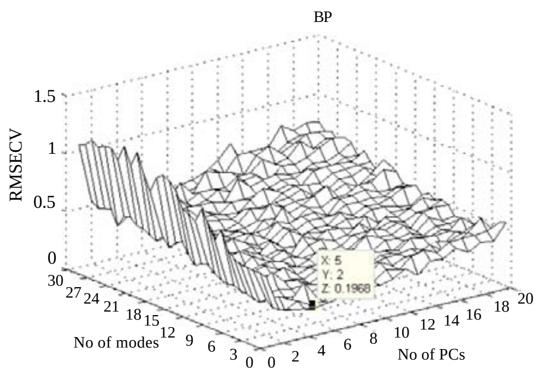

Fig. 5: Performance of BP-ANN models using different number of PCs

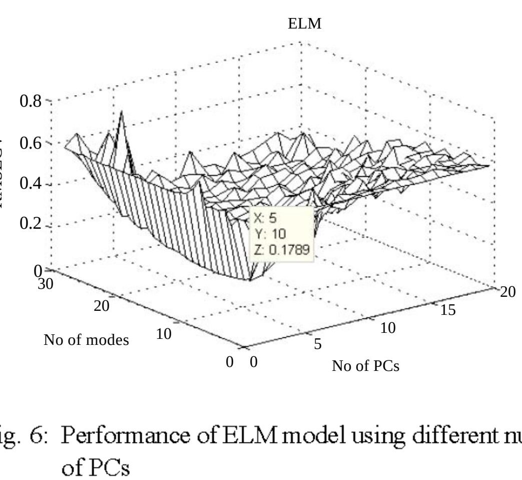

Fig. 6: Performance of ELM model using different number of PCs

The optimal number of PCs in BP-ANN model was 5 and the optimal number of hidden layer nodes was 2, with the corresponding topological structure in 5-2-1. The optimal number of PCs in ELM model was 5 and the optimal number of hidden layer nodes was 10, with the corresponding structure in 5-10-1.

Prediction of nitrogen using four modeling methods: According to the discussion above, the optimal number of PCs and the optimal parameters were set to build 4 models in different algorithms. The relationship between the hyperspectral image information of lettuce leaves and nitrogen content was established finally. The performances of 4 models were shown in Table 2. And the prediction results of PLSR model, SVR model, BP model and ELM model respectively were shown in Fig. 7, 8, 9 and 10.

From Table 2 it can be seen that, SVR model based on RBF kernel function has the best prediction performance, with determination coefficient of Prediction (R^2_p) of 0.96, the Root Mean Square Error of Prediction (RMSEP) of 0.18. ELM model also has a good prediction performance,

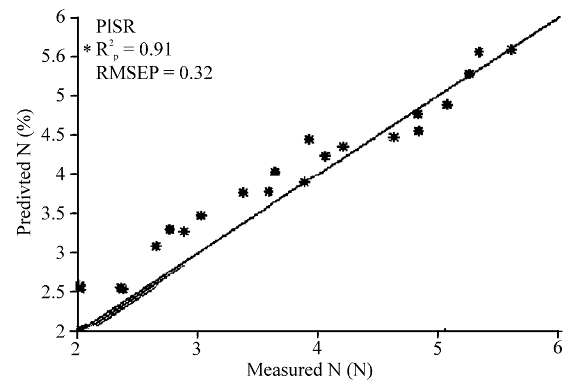

Fig. 7: Prediction result of PLSR model

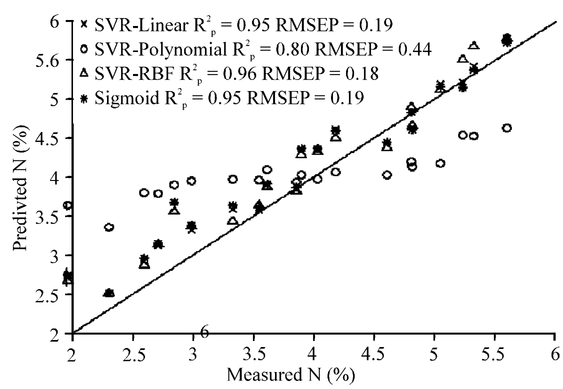

Fig. 8: Prediction result of SVR model

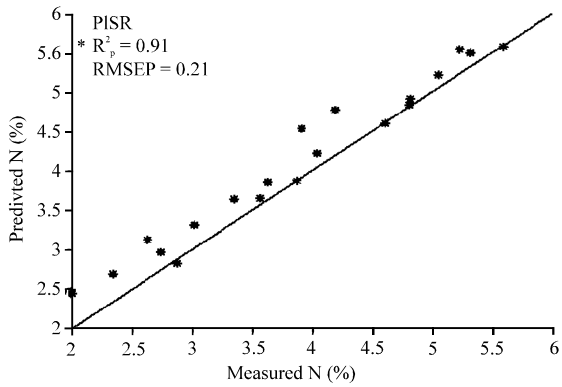

Fig. 9: Prediction result of BP-ANN model

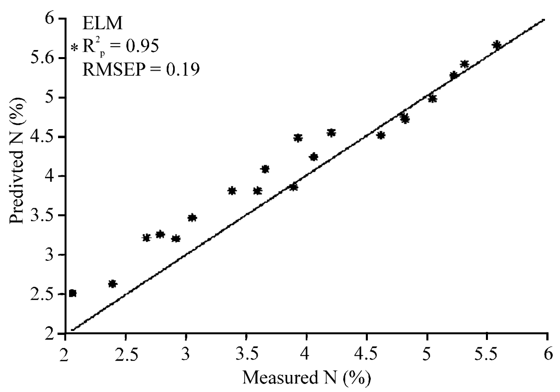

Fig. 10: Prediction result of ELM model

Table 2: Performances of 4 models

Model	No. of PCs	R^2	RMSEP
PLS	7	0.91	0.32
SVR			
Linear	10	0.95	0.19
Polynomial	4	0.80	0.44
RBF	10	0.96	0.18
Sigmoid	10	0.95	0.19
BP			
Minimum	5	0.85	0.18
Maximum		0.97	0.24
Mean		0.92	0.21
ELM			
Minimum	5	0.92	0.17
Maximum		0.97	0.21
Mean		0.95	0.19

with R^2_p of 0.95 and RMSEP of 0.19. In contrast, SVR model based on polynomial kernel function has the worst performance with R^2_p of 0.80 and RMSEP of 0.44.

CONCLUSION

Partial Least Square Regression (PLSR), Back Propagation Artificial Neural Networ k(BP-ANN), ELM, SVR were applied to relate the nitrogen content to the corresponding PCs to build the prediction model of nitrogen content and reasonable estimation results were obtained. The determination coefficient for prediction (R^2_p) of the PLSR model for nitrogen content was 0.91 and Root Mean Square Error for Prediction (RMSEP) was 0.32. BP model of structure 5-2-1 for nitrogen with R^2_p of 0.92 and RMSEP of 0.21, ELM model of structure 5-10-1 with R^2 of 0.95 and RMSEP of 0.19 and SVR model for nitrogen with R^2_p of 0.96 and RMSEP of 0.18, all got good prediction performance. Compared with other three models, SVR model based on RBF kernel function has the best prediction performance for determination of the nitrogen content in lettuce leaves. This work demonstrated that the hyperspectral imaging technique coupled with PCA-SVR(RBF kernel function) exhibits considerable promise for nondestructive diagnostic of nitrogen deficiency in lettuce leaves.

ACKNOWLEDGMENT

This work is supported by the Open Fund Project of KeyLaboratory of Agri-informatics, Ministry of Agriculture(2013007), National natural science funds projects (31101082, 61075036), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions(PAPD) and the National College Students' innovation and entrepreneurship training program (201310299011), the college students' innovation and entrepreneurship training program of higher education of Jiangsu province (201310299011Z).

REFERENCE

Barbin, D.F., G. ElMasry, D.W. Sun and P. Allen, 2012. Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging. Analytica Chimica Acta, 719: 30-42.

Ding, L., T.Q. Liao and L. Tao, 2011. The method of sensors data fusion based on SVR. Chin. J. Sensors Actuators, 24: 710-713.

ElMasry, G., N. Wang, A. ElSayed and M. Ngadi, 2007. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. J. Food Eng., 81: 98-107.

Feng, Y.Z. and D.W. Sun, 2013. Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of *Pseudomonas* loads in chicken fillets. Talanta, 109: 74-83.

Graeff, S., D. Steffens and S. Schubert, 2001. Use of reflectance measurements for the early detection of N, P, Mg and Fe deficiencies in *Zea mays* L. J. Plant Nutr. Soil Sci., 164: 445-450.

Shi, J.Y., X.B. Zou, J.W. Zhao, H.P. Mao, K.L. Wang, Z.W. Chen and X.W. Huang, 2011. Diagnostics of nitrogen deficiency in mini-cucumber plant by near infrared reflectance spectroscopy. Afr. J. Biotechnol., 10: 19687-19692.

Shi, J.Y., X.B. Zou, J.W. Zhao, K.L. Wang and Z.W. Chen *et al.*, 2012. Nondestructive diagnostics of nitrogen deficiency by cucumber leaf chlorophyll distribution map based on near infrared hyperspectral imaging. Scientia Horticulturae, 138: 190-197.

Tian, Y.C., X. Yao, J. Yang, W.X. Cao, D.B. Hannaway and Y. Zhu, 2011. Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground-and space-based hyperspectral reflectance. Field Crops Res., 120: 299-310.

Yang, W., S. Nick and M.Z. Li, 2010. Nitrogen content testing and diagnosing of cucumber leaves based on multispectral imagines. Spectrosc. Spectral Anal., 30: 210-213.

Zhang, X., F. Liu, Y. He and X. Gong, 2013. Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. Biosyst. Eng., 115: 56-65.

Zhang, Y.S., X. Yao, Y.C. Tian, W.X. Cao and Y. Zhu, 2010. Estimating leaf nitrogen content with near infrared reflectance spectroscopy in rice. Chin. J. Plant Ecol., 34: 704-712.