Sensors & Transducers, Vol. 169, Issue 4, April 2014, pp. 55-60

Sensors & Transducers

Detection of Isoflavones Content in Soybean Based on Hyperspectral Imaging Technology

Tan Kezhu, Chai Yuhua, Song Weixian, Cao Xiaoda
College of Electrical and Information, Northeast Agricultural University,
59 Mucai Street, Harbin, 150030, China
Tel.: +86 045155191383, fax: +86 045155191383
E-mail: tankezhu@yeah.net

Received: 10 February 2014 /Accepted: 7 April 2014 /Published: 30 April 2014

Abstract: Because of many important biological activities, Soybean isoflavones which has great potential for exploitation is significant to practical applications. Due to the conventional methods for determination of soybean isoflavones having long detection period, used too many reagents, couldn't be detected on-line, and other issues, we propose hyperspectral imaging technology to detect the contents of soybean isoflavones. Based on the 40 varieties of soybeans produced in Heilongjiang province, we get the spectral reflection datum of soybean samples varied from the soybean's hyperspectral images which are collected by the hyperspectral imaging system, and apply high performance liquid chromatography (HPLC) method to determine the true value of the selected samples of isoflavones. The feature wavelengths for isoflavones content prediction (1516, 1572, 1691, 1716 and 1760 nm) were selected based on correlation analysis. The prediction model was established by using the method of BP neural network in order to realize the prediction of soybean isoflavones content analysis. The experimental results show that, the ANN model could predict isoflavones content of soybean samples with of 0.9679, the average relative error is 0.8032 %, and the mean square error (MSE) is 0.110328, which indicates the effectiveness of the proposed method and provides a theoretical basis for the applications of hyerspectral imaging in non-destructive detection for interior quality of soybean.

Keywords: Soybean isoflacones, Hyperspectral images, Nondestructive examination, BP neural network.

1. Introduction

People mainly rely on soybean to obtain isoflavone. However the isoflavone content of soybean is various obviously according to different soybean seed varieties and the isoflavone content is also relation to appearance quality of soybean seed (shape, color, disease et al.) [1]. Therefore we can apply nondestructive detection technique to select genetic resources of soybean with high isoflavone content, besides, isoflavone content rate of raw material affects production cost seriously for the same production technology. Nevertheless, traditional analytical techniques are time consuming, destructive and sometimes, can require sophisticated equipments (i.e. the analyses of phenolic compounds by HPLC, volatile compounds by GC–MS, and mechanical features by texture analyzer). Moreover, destructive analyses can be performed only on a limited number of samples and thus, their statistical relevance could be limited. In recent years, researches have been focused on the development of non-destructive techniques suitable to increase the number of samples that can be analyzed, to repeat more times the same analysis on the same sample at a given time or during its physiological evolution, and

to achieve real-time information [2] .The study of nondestructive detection technique for isoflavone can supply high-quality soybeans for oil and fatty enterprise. However, fewer researches focus on isoflavone applying nondestructive detection technique. Several non-destructive techniques such as Visible/Near Infra Red spectroscopy (vis/NIR) can be applied for the assessment of soybean quality traits. Some researchers analyzed soybean fatty acid [3], soybean protein [4], 18 amino acids [5], and phosphorus content using Visible/Near Infra Red spectroscopy [6, 8]. Jin Hwan Lee identified transgenic soybeans applying NIR. However, NIR method can express only partial information of samples, which affects detection accuracy seriously because of distribution inconsistency of components in space [9].

Hyperspectral imaging, known also as chemical or spectroscopic imaging, is an emerging technique that integrates conventional imaging and spectroscopy to attain both spatial and spectral information from an object [10]. A hyperspectral image consists of a series of sub-images, each one representing the intensity distribution at a certain spectral band. When a sample is exposed to light, the reflected radiation can be measured and recoded as a reflectance spectrum. This spectrum is related to chemical composition of the sample, and spectra collected from sample at different quality levels can therefore be quite different [9]. Hyperspectral imaging is advantageous with respect to spectroscopic techniques, which acquire the spectral data from a single point or from an integrating of a small region on the tested fruit, and it is advantageous respect to the conventional RGB techniques, which can be poor identifiers of surface features and chemical composition of the sample sensitive to wavebands other than RGB. Thus, hyperspectral imaging can be a useful tool for quality investigation of agricultural products. In research applications, hyperspectral imaging was used to measure sugar content and surface damage of some fruits such as apple, pear, and orange [10, 14], leaf diseases and pesticide residue of tomato and cucumber [15], freshness and PH value of pork and chicken [16], maturity of some fruits such as oil palm, table grapes and tomatoes [17], firmness of strawberries, soluble solid content, bitter pit, bruise and surface defects and contaminations in apple fruit [18, 21].These researches have illustrated the potential of using hyperspectral imaging in agricultural products detection.

In this study, hyperspectral data of soybean in Heilongjiang province was acquired using hyperspectral sensor system (1000 nm-2500 nm). Five bands were selected from total 256 bands after computing correlationship. ANN algorithm was investigated in predicting soybean isoflavones content.

2. Materials and Methods

2.1. Soybean Samples

In this study, 40 Heilongjiang province common soybean seed varieties were acquired from Soybean Research Center of Northeast Agricultural University. The soybean samples were stored in dry and temperate condition. The soybean samples for the experiment are shown in Table 1.

Table 1. The soybean samples for the experiment.

ID	Sample name	Number	ID	Sample name	Number
1	Beijiao08-266	50	21	Heijiao1582	50
2	Donghai09-51083	50	22	Heinong44	50
3	Dongjiao03-2176	50	23	Heinong48	50
4	Dongnong4400	50	24	Kangxian5	50
5	Dongnong82216	50	25	Kedou28	50
6	Dongyin	50	26	Ken07-4373	50
7	Ha06-3869	50	27	Ken07-5203	50
8	Hefeng44	50	28	Kendou26	50
9	Hefeng 50	50	29	Kendou31	50
10	Hefeng 52	50	30	Kenfeng20	50
11	Hefu03-775	50	31	Kennong30	50
12	Hehang-05-450	50	32	Ken07-5054	50
13	Hejiao06-1148	50	33	Kenfeng16	50
14	Hejiao 07-410	50	34	Longsheng1	50
15	Hejiao 07-482	50	35	Mengdou31	50
16	Hejiao 07-482-1	50	36	Nengfeng16	50
17	Hejiao 08-1148	50	37	Sui08-5331	50
18	Hejiao 60	50	38	Suinong25	50
19	Hejiao 61	50	39	Suinong35	50
20	Hefeng 25	50	40	Beifeng9	50

2.2. Hyperspectral Imaging System

Hyperspectral images of the soybean samples were acquired using a hyperspectral imaging system. The system is consisted of a charge-coupled device (CCD) camera equipped with an imaging spectrograph (IM-Spector V25E, Spectral Imaging Ltd., Finland). The optical sensor system allowed to study the soybean quality in the spectral range of 1000-2500 nm. The light source is consisted of a 150 W halogen lamp (EKE 21 V 150 W, Japan) and an optic fiber that transfer the radiation to a linear light diffuser. The camera spectrograph assembly was supplied with a stepper motor to move the unit through the field of the camera's view and scan the samples line by line. The spectral images were collected in a dark room where only the halogen light source was used. The value of the camera gain was fixed at 0 to minimize the background noise. Each collected spectral image was stored as a three-dimensional image (x, y, λ). The spatial components (x, y) included 320*256 pixels, and the spectral component (λ) included 256 bands within 1000-2500nm range. The hyperspectral imaging system was controlled by software called Spectral Scanner for images acquisition. The soybean samples were placed in a small container with the size of 10 cm*10 cm. Hyperspectral images of each soybean varieties were collected by using the hyperspectral imaging system with about 60 cm distance between the sample and the camera. Before collecting hyperspectral data, we should compute exposure time exactly to acquire clear images. The speed of stepper motor was adjusted to get appropriate spatial resolution. The exposure time was 10 ms, and the stepper motor speed was 1.25 mm/s. The schematic diagram of the hyperspectral imaging system used is shown in Fig. 1.

The software of ENVI4.7 (ITT, USA), Matlab7.1.2 (The Math Works, Natick, USA) and IBMSPSS Statistics 21 were used for hyperspectral data analysis.

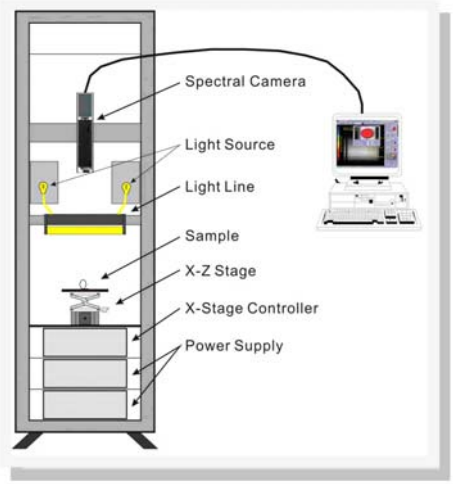

Fig. 1. Hyperspectral imaging system.

2.3. Calibration of Hyperspectral Images

During operation, light intensity will change for various spectral bands and the image noise will be caused by dark current. The hyperspectral images were firstly corrected with a white and a dark reference. The dark reference was used to remove the effect of dark current of the CCD detectors, which are thermally sensitive. The corrected image (R) is estimated by using the following equation:

$$R = \frac{I_{raw} - I_{black}}{I_{white} - I_{black}}, \quad (1)$$

where I_{raw} is the recorded hyperspectral image, I_{black} is the dark image (with 0 % reflectance) recorded by turning off the lighting source with the lens of the camera completely closed, and I_{white} is the white reference image (The white board with 99 % reflectance). The corrected images will be the basis for the subsequent image analysis to extract the spectral response of each sample. As a result of length reason, only corrected image of soybean seeds for 1666 nm wavelength is shown in Fig. 2.

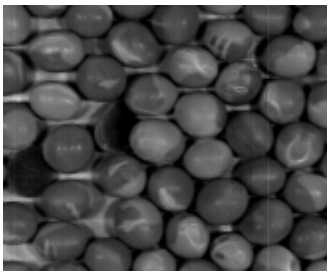

Fig. 2. Corrected image of soybean seeds for 1666 nm wavelength.

The average spectra reflectance value of interesting region was selected from calibration hyperspectral images. The spectrum was in range of 1000 nm-2500 nm. The original reflectance spectrum of soybean samples is shown in Fig. 3.

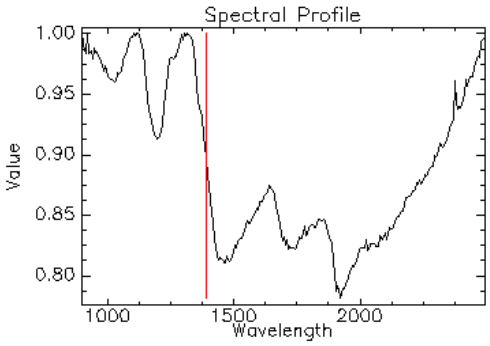

Fig. 3. Original reflectance spectra of soybean samples.

3. Results and Discussion

3.1. Spectrum Curve Smooth

The median filter is a nonlinear smoothing technique, the gray value of each pixel is set for all pixel gray value of the point value of a neighborhood window. The median filter is a kind of effective noise suppression of nonlinear signal processing technology based on the order statistical theory, the basic principle of median filter is the median of each sample value instead of a neighborhood of the point of a digital image or sequence of numbers, to get around the pixel values close to the true value, thereby eliminating the isolated noise points. The smoothing results were acquired by the following equation.

$$g(x,y) = med\{f(x-k,y-l),(k,l \in W)\}, \quad (2)$$

where $f(x,y)$ is the original reflectance spectra, $g(x,y)$ is the spectrum after smoothing, W is the 2D template.

In this experiment, 3×3 sliding window was selected to smooth the original reflectance spectra of soybean. The smoothing result is shown in Fig. 4. We can see that the spectral curve is smoother, the noise is lower and the S/N (signal to noise ratio) is more higher compared to the original reflectance spectra.

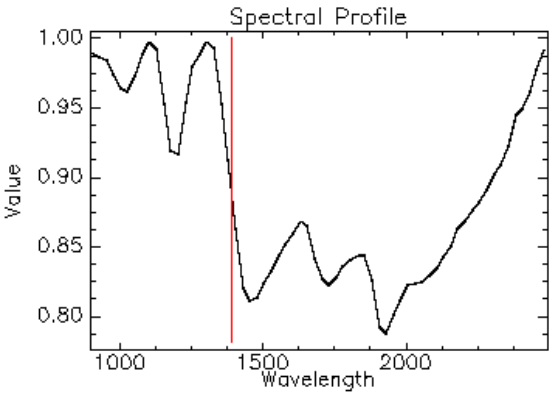

Fig. 4. Reflectance spectra after median filtering algorithm.

3.2. Regression Analysis of Isoflavone Content and Spectral Reflectance Values

For reasons of space, correlation analysis the of isoflavone content of 34 soybean samples and wavelengths (1002 nm and 1716 nm) are shown in Table 2.

The regression curve of isoflavone content and spectral reflectance values is shown in Fig. 5. As can be seen from the figure, correlation coefficient of isoflavone content and 5-7 bands are about 0.4, therefore, 5 bands were selected for characteristic wavelength in forecasting isoflavone content. The characteristic wavelengths are shown in Table 3.

Table 2. Correlation analysis between isoflavones contents and feature wavelength.

		Isoflavones content	Band 1	Band 2
isoflavones content	Pearson Correlation	1	.404*	.262
	Significance (bilateral)		.018	.134
	N	34	34	34
Band 1	Pearson Correlation	.404	1	.958**
	significance (bilateral)	.018		.000
	N	34	34	34
	Pearson Correlation	.262	.958	1
	significance (bilateral)	.134		
Band 2			.000	
	N	34	34	34

*. Significant at 0.05 level
**. Significant at 0.01 level

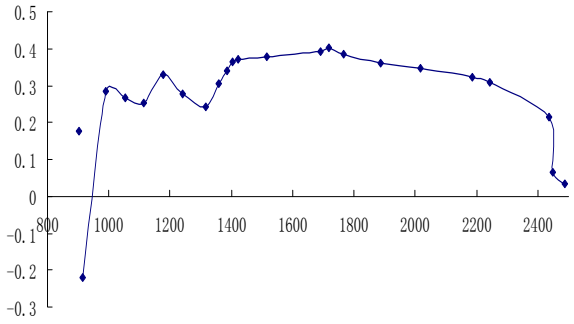

Fig. 5. Correlation coefficient curve for isoflavones content.

Table 3. Feature wavelength for isoflavones content.

Quality factor	Characteristic wavelength/nm				
Isoflavone	1516	1572	1691	1716	1760

3.3. Prediction Models of BP Neural Network

There are many statistical and expert system approaches to handle complex mathematical transformations. In these approaches, artificial neural network (ANN) is relatively easy to implement, reliable. Here the attractiveness of using a neural network model to predict soflavone is demonstrated.

In this study, multilayer feed-forward network with back propagation was used. The back-propagation (BP) algorithm and its variants are supervised by training rules that involve an iterative procedure to move the weights along the negative gradient of the error function, so as to minimize the difference between the networks' output and the desired target. In each step, errors are used as inputs to feedback connections from which adjustments are made to the synaptic weights layer by layer in a backward direction. The weight associated with each connection is adjusted by an amount proportional to the strength of the signal on the connection and the total measure of the error. All ANN neurons used were configured on the model by Haykin [22]. The hidden and output neurons can be mathematically characterized with the following equations:

$$g(x,y) = med\{f(x-k, y-l), (k,l \in W)\} \quad (3)$$

$$y_k = \varphi(v_k), \quad (4)$$

where x_j is the input signal; w_{kj} is the synaptic weight of neuron k; v_k is the linear combiner or summing junction; b_k is the bias; y_k is the output of the neuron and φ is the transfer function. This study considered two of the most common transfer functions: the hyperbolic tangent sigmoid (tansig) and the logarithmic sigmoid (logsig) functions. If the output layer of the network has sigmoid neurons, then the output values are limited to a small range. So, linear output neurons were used.

The Levenberg-Marquardt algorithm was used to train the networks as it is suitable to deal with ill-conditioned minimization problems. So, an approximation to the Hessian matrix is used in the Newton-like update:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e_k, \quad (5)$$

where μ governs the step size and I is the unit matrix; J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is the vector of network errors.

The neural network designed with an input layer, an output layer and one hidden layer in this work. 40 soybean verities were used in the experiment; five characteristic bands were selected for each sample. 34 soybean verities were selected randomly for training, the rest for predicting. There are 5 neurons at input layer which represents spectral reflectance values based on 5 bands. After several tests, the number of neurons in hidden layer is 10. The artificial neural network was trained by using MATLAB software. After several tests, learning rate is determined to 0.1, and momentum factor is determined to 0.2. We can see from Fig. 6, the coefficient of determination (R^2) is 0.9679. It is evident from the result that it is feasible using ANN to predict isoflavone content. The measured value and predicted value of isoflavone is shown in Fig. 6.

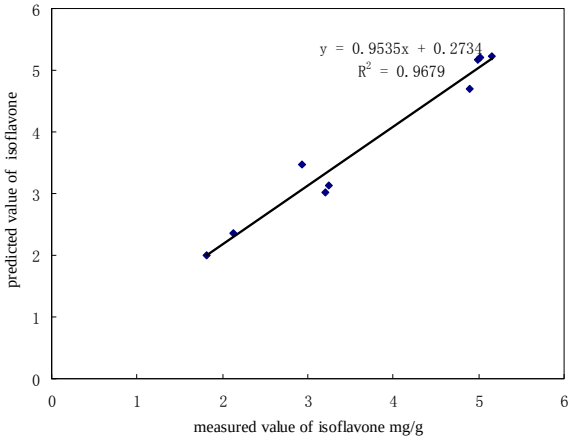

Fig. 6. Measured versus estimated isoflavones content using ANN method.

4. Conclusions

The following conclusions were drawn from this study:

1) The hyperspectral images (100 nm-2500 nm) of different soybean verities were acquired after testing the hyperspectral imaging system, and the reflectance spectra was acquired.

2) Correlation of isoflavone content and spectra values in 256 bands was analyzed, and correlation coefficient curve for isoflavones content was acquired. According to the result, the wavelengths (1516 nm, 1572 nm, 1691 nm, 1716 nm and 1760 nm) were selected as characteristic wavelength.

3) Prediction model for isoflavone was constructed using ANN. The coefficient of determination (R^2) is 0.9679. The average relative error is 0.8032 %, MSE is 0.110328. Thus, the hyperspectral imaging technique proposed in this paper can potentially be useful for predicting soybean isoflavone.

Acknowledgements

This study was financially supported by Heilongjiang Provincial Natural Science Foundation (ZD201303) and Youth Scientific Research Fund of Northeast agricultural University.

References

[1]. Wang Jifeng, Zhao Xiaoming, Song Yufa, et al, Effects of appearance quality on the chemical quality of soybean, *Heilongjiang Agricultural Sciences*, Issue 5, 2006, pp. 23-24.

[2]. Li Jiangbo, Rao Xiuqin, Ying Yibin, Advances in application of hyperspectral imaging to nondestructive detection of external quality of agricultural products, *Spectroscopy and Spectral Analysis*, Vol. 31, Issue 8, 2011, pp. 2021-2026.
[3]. Chai Yuhua, Tan Kezhu, Measurement of soybean fatty acid content by near infrared spectroscopy, *Transactions of the Chinese Society of Agricultural Engineering*, Issue 1, 2007, pp. 238-241.
[4]. Yan Long, Jiang Chun-Zhi, et al, Development and Reliability of Near Infrared Spectroscopy Models of Protein and Oil Content in Soybean, *Soybean Scienc*e, Vol. 27, Issue 5, 2008, pp. 32-38.
[5]. Li Nan, et al, The Prediction of Nondestructive Measurement of Amino Acids Composition, *Journal of Plant Genetic Resources*, Vol. 13, Issue 6, 2012, pp. 121-127.
[6]. Delwiche S. R., Protein content of single kernds of wheat by nearinfrared reflectance spectroscopy, *Journal of Cereal Science*, Vol. 30, Issue 6, 1995, pp. 241-254.
[7]. R. P. Cogdill, C. R. Hurburgh, G. R. Rippke, Single-kernel maize analysis by near-infrared hyperspectral imaging, *Transactions of the ASAE*, Vol. 47, Issue 1, 2004, pp. 311-320.
[8]. C. B. Singh, D. S. Jayas, J. Paliwa, et al, Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging, *Cereal Chemistry*, Vol. 86, Issue 3, 2009, pp. 256-260.
[9]. G. Elmasry, N. Wang, A. Elsayed, et al., Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry, *Journal of Food Engineering*, Vol. 81, Issue 1, 2007, pp. 98-107.
[10] R. F. Lu, Y. K. Peng, Hyperspectral scattering for assessing peach fruit firmness, *Biosystems Engineering*, Vol. 93, Issue 2, 2006, pp. 161-171.
[11]. Hong Tiansheng, Qiao Jun, Wang Ning, et al, Nondestructive inspection of Chinese pear quality based on hyperspectral imaging technique, *Transaction of the Chinese Society for Agricultural Engineering*, Vol. 23, Issue 2, pp. 151-155.
[12]. Zhao Juan, Peng Yankun, Zhao Songwei, Detection of apple surface defects based on hyperspectral technology, *Journal of Food Safety and Quality*, Vol. 3, Issue 6, 2012, pp. 681-684.
[13]. Guo Enyou, Liu Muhua, Zhai Jiewen, et al, Nondestructive testing technology of hyperspectral image sugar content of navel orange, *Journal of Agricultural Machinery*, Issue 5, 2008, pp. 91-93.
[14]. Shan Jiajia, Peng Yankun, Wang Wei, Simultaneous detection of external and internal quality parameters of apples using hyperspectral technology, *Journal of Agricultural Machinery*, Vol. 42, Issue 3, 2011, pp. 140-144.
[15]. Chai Ali, Liao Ningfang, et al, Identification of Cucumber Disease Using Hyperspectral Imaging and Discriminate Analysis, *Spectroscopy and Spectral Analysis*, Vol. 30, Issue 5, 2010, pp. 151-156.
[16]. Zhang Leilei, Peng Yankun, et al, Determination of pork freshness attributes by hyperspectral imaging technique, *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 28, Issue 7, 2012, pp. 254-259.
[17]. Osama Mohammed Ben Saeed, Sindhuja Sankaran, et al, Classification of oil palm fresh fruit bunches based on their maturity using portable four-band sensor system, *Computers and Electronics in Agriculture*, Issue 82, 2012, pp. 55-60.
[18]. Min Huanga, Renfu Lu, Apple mealiness detection using hyperspectral scattering technique, *Postharvest Biology and Technology*, Vol. 58, Issue 3, 2010, pp. 168-175.
[19]. Muhammad A. Shahin, Stephen J. Symons, Detection of fusarium damage in Canadian wheat using visible/near-infrared hyperspectral imaging, *Food Measure*, Issue 6, 2012, pp. 3-11.
[20]. Gowen A. Burger, Data handling in hyperspectral image analysis, *Chemometrics and Intelligent Laboratory Systems*, Vol. 108, Issue 1, 2011, pp. 13-22.
[21]. Ma Benxue, Ying Yibin, Rao Xiuqin, et al., Advance in nondestructive detection of fruit internal quality based on hyperspectral imaging, *Spectroscopy and Spectral Analysis*, Vol. 29, Issue 6, 2009, pp. 1611-1615.
[22]. Haykin S., Machinery, Neural network and machine learning, *Machinery Industry Press*, 2009.