Advance Journal of Food Science and Technology 11(4): 337-343, 2016
DOI:10.19026/ajfst.11.2420
ISSN: 2042-4868; e-ISSN: 2042-4876

Submitted: August 21, 2015 Accepted: September 11, 2015 Published: June 05, 2016

Research Article

Identification of Black Tea from Four Countries by Using Near-infrared Spectroscopy and Support Vector Data Description Pattern Recognition

Jingming Ning, Jingjing Sun, Xiaoyuan Zhu, Xuyu and Zhengzhu Zhang
State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, P.R. China

Abstract: In this study, black teas from four different countries were successfully identified using near-infrared (NIR) spectroscopy combined with the Support Vector Data Description (SVDD) algorithm. The original spectra of tea ranged in wavelength from 12500 to 4000 cm^{-1}. We used SVDD to optimize the parameters and calibrate the discrimination model. As a comparison, the K-Nearest Neighbor algorithm (KNN) and Partial Least Square (PLS) were also used in this study. Compared with the KNN and PLS classifications, the SVDD model was better able to deal with imbalance training samples and outperformed the other models in the prediction set. The optimal SVDD model was achieved with principal components (*PC*) = 5. Identification rates were 96.25% in the training set and 92.50% in the prediction set. These results indicate that NIR spectroscopy combined with SVDD is a useful tool in building a one-class calibration model for discrimination of black tea from different countries.

Keywords: Black tea, different origins, FTIR spectroscopy, identification, SVDD

INTRODUCTION

Tea (Camellia sinensis (L.) O. Kuntze) is a common beverage consumed by more than two billion people in more than 120 countries (Yumei, 2012). Tea has many health benefits, including antioxidant activity to prevent cancer, cardiovascular diseases and diabetes (Sharangi, 2009).

Additionally, tea can improve digestion, aids in fat transformation and reduction and improve thinking (Basu and Lucas, 2007; Serban *et al*., 2015). Black tea, which is fermented, is the most popular tea in the world; it accounts for about 75% of the world's tea trade volume. To produce black tea, fresh leaves are withered, rolled (cut), fermented and dried (Ullah *et al*., 1984). During fermentation, tea polyphenols are enzymatically oxidized, which reduces the polyphenol content by 50% and results in the appearance of theaflavins, thearubigins and other new compounds. This significantly changes the chemical composition of the leaves (Hodgson *et al*., 2012; Bahorun *et al*., 2012) and improves the quality of the tea, thereby giving it a specific brightness and taste (Roberts and Smith, 1961; Owour, 1990; Owuor and Othieno, 1989; Millin and Swaine, 1981).

Teas are processed in over 50 countries in the world. China, India, Indonesia and Kenya are the main producers. Although black tea looks similar regardless of the country of origin, there are large differences in the chemical composition of black teas from different countries owing to differences in the geographical environment and climate.

Traditionally, chemical methods and sensory evaluation are used to discriminate between black teas from different countries (Chen *et al*., 2008a). While the chemical analysis methods used to identify tea are precise (Sereshti *et al*., 2013), they are complex, time-consuming, labour-intensive, costly and require large amounts of organic solvent. Using sensory evaluation to identify tea is imprecise, as it can be easily influenced by other factors, including the environment and the mood of the evaluator. Recently, electronic noses (Sharma *et al*., 2013) and electronic tongues (Khaydukova *et al*., 2015) have been used to evaluate tea quality; however, these methods cannot distinguish teas from different countries. Therefore, a rapid and accurate analytical method is required to discriminate the geographical origin of black teas.

Fourier Transform Near-Infrared (FT-NIR) spectroscopy is a powerful analytical tool because it is fast, accurate and non-destructive. NIR spectroscopy has been successfully used to analyse some active compounds found in tea. For example, Yan (2005) and Liu *et al*. (2010) used NIR to analyse the amino acids, caffeine, water and other components in tea. However, NIR spectroscopy is seldom used to discriminate between black teas from different countries.

In this study, we developed a robust method using NIR spectroscopy to discriminate between black teas from four countries (China, India, Kenya and

Corresponding Author: Zhengzhu Zhang, State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, P.R. China

Indonesia). In addition, we optimized the Support Vector Data Description (SVDD) algorithm to build a calibration model. We evaluated the performance of the final model based on the identification rates in a prediction set of samples.

MATERIALS AND METHODS

Technical route: The methodology of the study is described in Fig. 1. We obtained the original spectra of black teas from China, India, Indonesia and Kenya by using a FT-NIR spectrometer and we used a discriminated model. To identify an unknown sample, the original spectrum is obtained and the model is applied, simultaneously. The discriminated result is obtained quickly.

Sample preparation: Black tea samples of the same species (C. sinensis (L.) O. Kuntze) were collected from India, Indonesia, Kenya and China. In May 2011, young sprouts with one bud and two leaves were picked from tea shrubs. We collected 30 samples from each country, which resulted in 120 samples in total. Then, the samples were divided into two subsets: a calibration set to build the model and a prediction set to test the efficiency of the model. Table 1 details the testing samples.

To prepare the samples, all samples were dried in a forced draught oven (Shanghai Yi-Heng Machine Co., Shanghai, China) at 50 °C for about 2 h. The samples were crushed into powder by a cyclone mill. Black tea powder of the appropriate particle size was obtained by using a 60-mesh griddle. The teas were stored in airtight jars until further analysis.

Spectrum acquisition: NIR spectral curves of the black tea samples were obtained by using a FT-NIR spectrometer with an integrating sphere and a standard

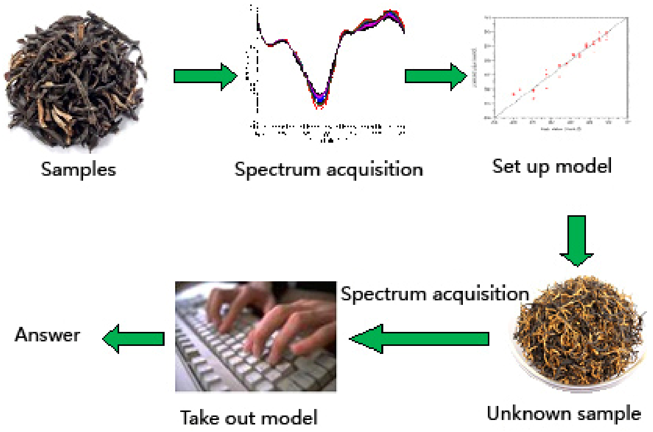
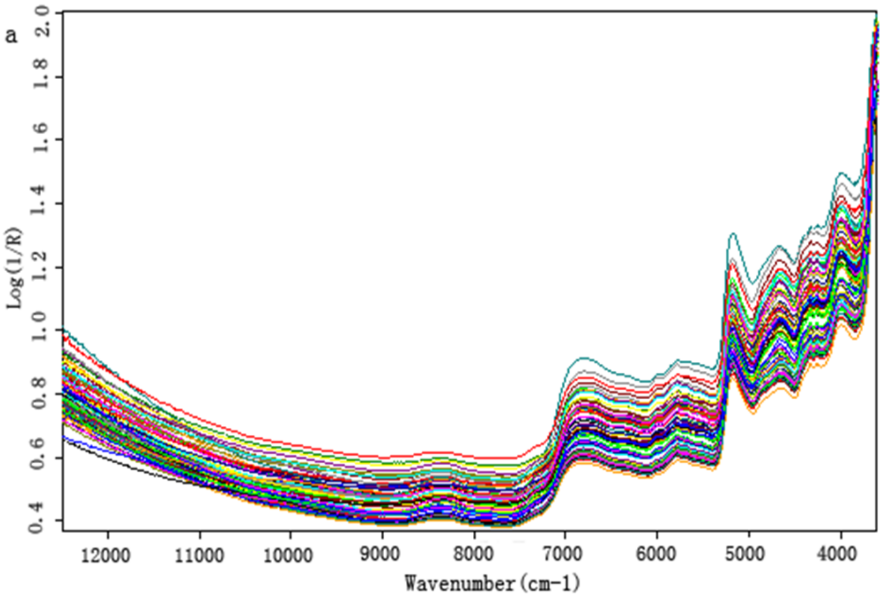


Fig. 1: The methodology of this study



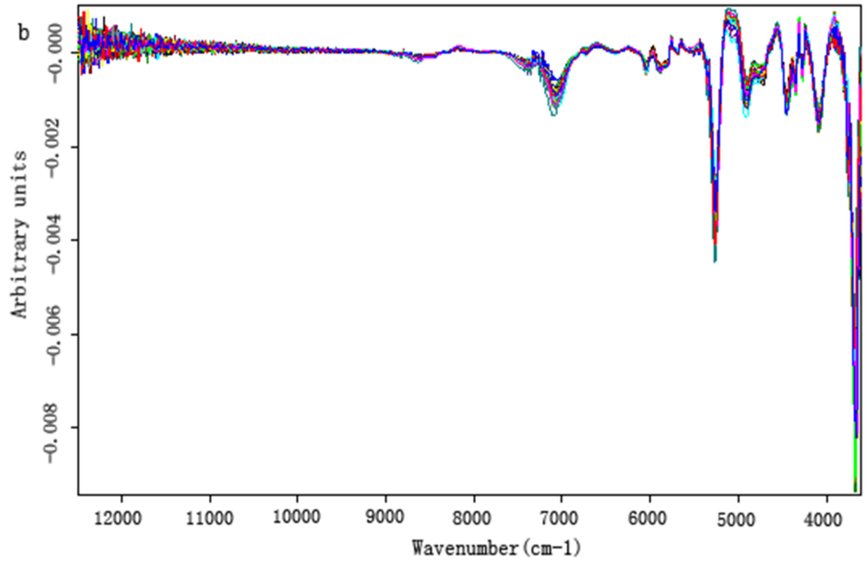


Fig. 2: Raw spectra of black tea samples; (a): before and; (b): after first derivative +9-point smooth preprocessing

Table 1: Origin and number of samples

Sample	Origin	Training set	Prediction set
India black tea	India	20	10
Indonesia black tea	Indonesia	20	10
Kenya black tea	Kenya	20	10
China black tea	China	20	10
Total		80	40

sample accessory (a cup) (Bruker Co., Karlsruhe, Germany) to hold the samples. The sample spectra were collected within the spectral range of 12,500-1-4,000 cm-1. The data was measured with an interval of 8 cm-1, which contained 2307 variables. During measurements, the temperature was 25±2°C and the humidity remained steady.

Each spectrum was an average of 32 (about 15 s) scanning spectra. For each black tea sample, 5 g black tea powder was placed into a sample cup. In order to get a stable database, each sample was scanned three times and averaged. The average value was used in the following analysis. Figure 2a shows the original NIR spectra of black tea samples from different geographical regions. All spectra were recorded as log (1/R), where R was the relative reflectance.

Basic theory of SVDD: Support Vector Data Description (SVDD) is custom-tailored for one-class classification. This algorithm is inspired by the theory of a two-class Support Vector Machine (SVM). SVDD defines boundaries around the different black tea samples by using the smallest sample volume possible (Chen *et al*., 2008b). The SVDD has been used to solve the multi-class classification problem and the methodology is as follows: Given the training data (x, y), …, (x, y), x∈R, y∈{1,…,K}, the class of all training samples with the minimal super group data is established. The radius R is defined by using the centre of each hyper sphere. The K super ball is obtained by solving a quadratic programming problem under K two:

$$\text{minimize}\, R + C\sum \xi, m = 1,\cdots,K \tag{1}$$

The constraint conditions are:

$$\|\phi(x) - a\| \le R + \xi, \forall i, y = m, m = 1,\cdots,K \tag{2}$$
$$\xi \ge 0, \forall i$$

where, ξ is the relaxation factor, C regulates the hyper sphere and controls the error. Similar to SVDD, the Lagrange operator is used to solve two quadratic programming problems and the dual form can be written as follows:

$$\text{Maximize} \sum \alpha K(x,x) - \sum \alpha\alpha K(x,x), m = 1,\cdots,K \tag{3}$$

$$\text{Constraint conditions} \sum \alpha = 1 \tag{4}$$

and $y = m$, $0 \le \alpha \le C$ $m = 1, \ldots, K$

When K a hyper sphere defines test point and K a hyper sphere S represents decision function, then

$$\text{Class of } x = \arg\max \text{sim}\,(x, S), m = 1, \ldots, K \tag{5}$$

Software: MATLAB V7.0 (Math Works, USA) under Windows XP was used for data analysis. The spectral curves were acquired through NIR spectral Software (Bruker Co., Karlsruhe, Germany).

RESULTS AND DISCUSSION

In this study, 30 samples were collected from each of the four countries, China, India, Indonesia and

Kenya; in total, 120 tea samples were collected. The ultimate aim of this study was to identify the origin of a tea sample as one of the four countries by supervised pattern recognition. Supervised pattern recognition refers to the techniques in which a classification model is developed by using a training set of samples. The model's performance is then evaluated using the prediction set of samples.

Spectral preprocessing investigation: Figure 2a presents the raw spectral curves of black tea samples from different countries. The raw NIR spectral curves are complex since there are hundreds of compounds in black tea. The most intensive band in the spectra belongs to the vibration of the overtone of the carbonyl group, followed by the C-H stretch and C-H deformation vibration, the-CH2 and the-CH3 overtone from the mid-infrared. The vibrations of the carbonyl group, -C-H and-CH2 are attributed to several specific ingredients of tea, including catechins, caffeine and amino acids

All the black tea samples cannot possess the same physical properties such as particle size and compactness. Therefore, the diffuse reflection of light in solid particles is inevitably affected. Consequently, the original spectral curves of the black tea samples have to be preprocessed. Different preprocessing methods have different impacts on the established model. In this study, three methods of spectrum preprocessing, the Standard Normal Variate (SNV), the first derivative and 9-point smooth (1stDer+9-point smooth) and the second derivative and 9-point smooth (2stDer+9-point smooth), were used to conduct a comparison study on the classification of the black tea samples from the four countries. The 1stDer+9-point smooth spectrum preprocessing method was the best in eliminating the effects of an uneven sample size and compactness on the spectral curves. In all three methods, the correlation coefficient (R) was 0.9826 for the calibration set, the root mean square error of cross validation (RMSECV) of interaction validation was 0.1021 and the Root Mean Square Prediction Error (RMSEP) of external validation was low (Table 2). We chose the 1stDer+9-point smooth method to preprocess the original NIR spectral curves because it best controlled for differences in sample preparation. The recognition rates with different spectral preprocessing methods are shown in Fig. 3. The original spectrum and preprocessed 1stDer+9-point smooth spectral curves of the black tea are shown in Fig. 2b.

Principal Component Analysis (PCA) results: Black tea contains a large number of organic ingredients, including catechins, amino acids and soluble carbohydrates. The hydrogen-containing groups in these organic ingredients can produce absorptions of multiplication and sum frequencies in the near-infrared area. Therefore, there is interference among the NIR spectral curves of the black tea samples. When a

Table 2: Model statistical parameters for different spectral preprocessing methods

Spectral preprocessing	Calibration set		Prediction set	
	RMSECV	R	RMSEP	R
Standard normal variate	0.4216	0.8768	0.5561	0.8471
1^{st}Der+9-point smooth	0.1021	0.9826	0.2135	0.9456
2^{st}Der+9-point smooth	0.2158	0.9352	0.2726	0.9025

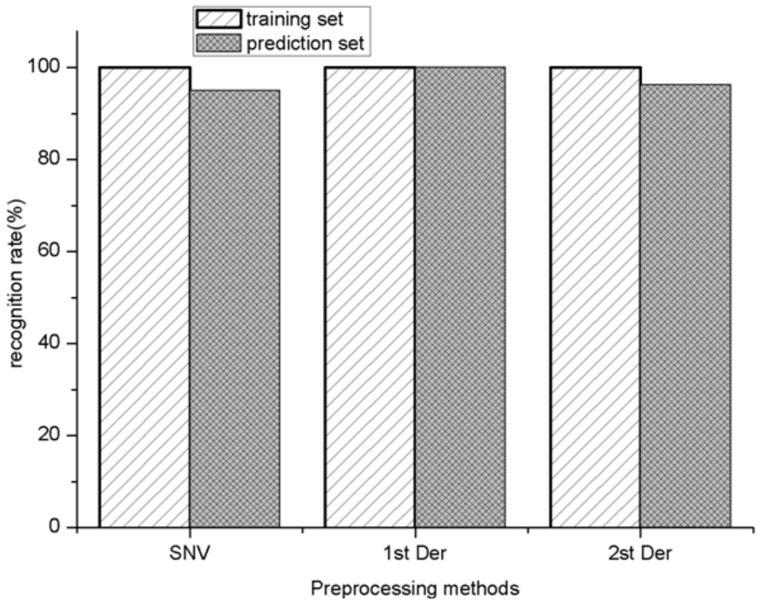


Fig. 3: The recognition rates with different preprocessing methods

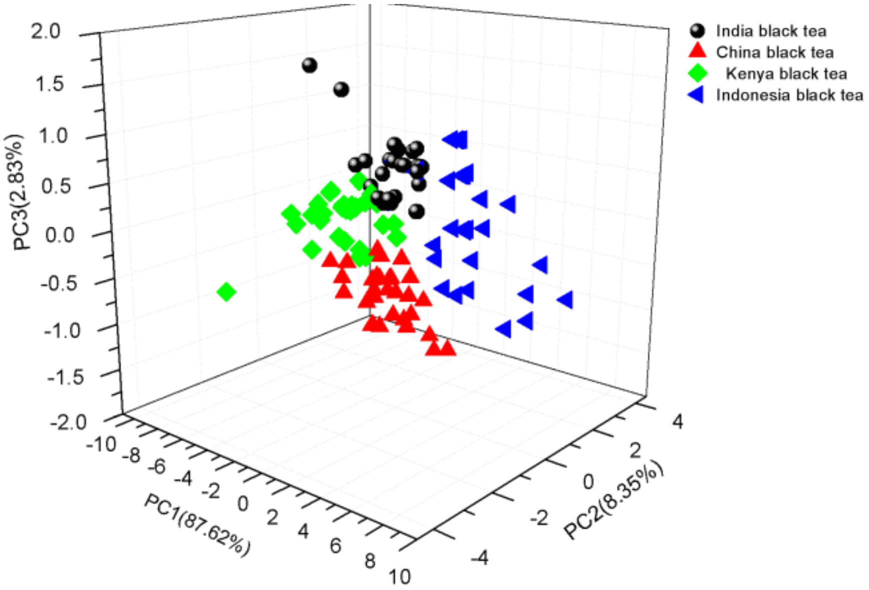


Fig. 4: Three-dimensional PCA scatter plot of black tea samples from four different countries

Table 3: Comparison of recognition results between SVDD, KNN and PLS models

	Recognition results	
Models	Training set (%)	Prediction set (%)
SVDD	95	90
KNN	85	70
PLS	95	85

classification model is established, this information redundancy decreases with model forecasting. Principal Component Analysis (PCA) is a statistical method to transform multiple indicators into several representative aggregative indicators. Redundancy information is reduced from a high-dimensional space to a low-dimensional space by using PCA. The vectors obtained from each principal component are orthogonal. A three-dimensional PCA scatter plot with the principal components PC1, PC2 and PC3 is shown in Fig. 4.

From the PCA scatter plot, we used the factor score vectors from the first three principal components. PC1 accounts for 87.62% of the variance, while PC2 and PC3 account for 8.35 and 2.83% of the variance, respectively. Therefore, the first three components of the black tea samples represent 98.80% of the information of the black tea samples. Tea samples from four countries cluster in the PCA scatter plot is differently. During the fermentation process, most organic compounds in tea change with respect to chemical structure and content, resulting in different absorptions of multiplication and sum frequencies in the near-infrared wavelength range. Fermenting for different amounts of time results in different changes in the organic components of tea. Since these black tea samples were all fermented for the same amount of time and thus have similar composition, they appeared close to each other in the three-dimensional scatter. The score construction in the three-dimensional scatter plot shows the similarities and differences in the structure and content of the black tea samples.

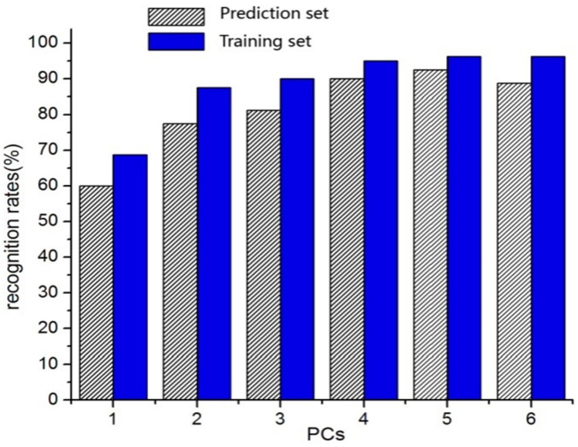


Fig. 5: Discriminating rates of emendation and forecast, using different factors in the SVDD model

Selection model: To highlight the good performance of the SVDD algorithm, we attempted to compare the SVDD algorithm with the K-Nearest Neighbor algorithm (KNN) and Partial Least Square (PLS) approaches. Table 3 shows the recognition results obtained by the SVDD, KNN and PLS approaches in the training and prediction sets. The ability to identify the origin of a tea sample in the training set is equal in PLS (95%) and SVDD (95%) and both are better than KNN (85%). However, in the prediction set, SVDD (90%) performed better than both KNN (75%) and PLS (85%). Therefore, SVDD is the best for the identification of the origin of a tea sample.

Table 4: Discrimination of the origin of black tea samples by using the SVDD model

			Result of recognition				
Sample set	Origin	Sample number	India	Indonesia	Kenya	China	Overall recognition rate (%)
Training set	India	20	19	0	1	0	96.25
	Indonesia	20	0	19	1	0	
	Kenya	20	1	0	19	0	
	China	20	0	0	0	20	
Prediction set	India	10	9	0	1	0	92.50
	Indonesia	10	1	9	0	0	
	Kenya	10	1	0	9	0	
	China	10	0	0	0	10	

SVDD Result: Different numbers of principal component factors were used to establish the SVDD discriminative model. The recognition rates in the training and prediction sets were used as indicators to evaluate the model stability and robustness. The influence of the principal component numbers on the recognition rate is shown in Fig. 5.

The recognition rate of the model increases with increase in the number of principal component factors included. With five principal components, the accumulated variance contribution rate is almost 97% and the recognition rate for the prediction set is at its highest. All nine principal components can reflect the features of the spectral curves of black tea samples. The model recognition rate is 96.25% in the training set and 92.50% in the prediction set of the black tea samples (Table 4).

CONCLUSION

We used SVDD as a pattern recognition tool to develop an identification model. We show that NIR spectra coupled with SVDD pattern recognition can identify the country of origin of black tea samples. The SVDD algorithm outperforms the KNN and PLS approaches in identifying the geographical origin of the tea samples. Therefore, NIR spectra analysis with SVDD pattern recognition could be used to identify the origin of other agricultural products.

ACKNOWLEDGMENT

This study has been financially supported by the Natural and Science Program of Education Department (Grant No. KJ2013A109) and the Natural Science Foundation in Anhui Province (Grant No. 1408085MC61).

REFERENCES

Bahorun, T., A. Luximon-Ramma, V.S. Neergheen-Bhujun, T.K. Gunness, K. Googoolye *et al*., 2012. The effect of black tea on risk factors of cardiovascular disease in a normal population. Prev. Med., 54: 98-102.

Basu, A. and E.A. Lucas, 2007. Mechanisms and effects of green tea on cardiovascular health. Nutr. Rev., 65: 361-375.

Chen, Q., J. Zhao, J. Cai *et al*., 2008b. Inspection of tea quality by using multi-sensor information fusion based on NIR spectroscopy and machine vision. Trans. Chinese Soc. Agric. Eng., 3: 5-10.

Chen, Q.S., Z.M. Guo and J.W. Zhao, 2008a. Identification of green tea's (*Camellia sinensis* (L.)) quality level according to measurement of main catechins and caffeine contents by HPLC and support vector classification pattern recognition. J. Pharmaceut. Biomed., 48: 1321-1325.

Hodgson, J.M., I.B. Puddey, R.J. Woodman, T.P.J. Mulder, D. Fuchs *et al*., 2012. Effects of black tea on blood pressure: A randomized controlled trial. Arch. Intern. Med., 172(2): 186-188.

Khaydukova, M., X. Cetó, D. Kirsanov, M. del Valle and A. Legin, 2015. A tool for general quality assessment of black tea-retail price prediction by an electronic tongue. Food Anal. Method., 5: 1088-1092.

Liu, T., B. Chun-fang and R. Yu-lin, 2010. Determination of quality properties of soy sauce by support vector regression coupled with SW-NIR spectroscopy. Chem. Res. Chin. Univ., 3: 385-391.

Millin, D.J. and D. Swaine, 1981. Fermentation of tea in aqueous suspension. J. Sci. Food Agr., 32: 905-919.

Owour, P.O., 1990. Variations of the chemical composition of clonal black tea due to delayed withering. J. Sci. Food Agr., 1: 56-61.

Owuor, P.O. and C.O. Othieno, 1989. Effects of maceration method on the chemical composition and quality of clonal black teas. J. Sci. Food Agr., 49: 84-94.

Roberts, E.A.H. and R.F. Smith, 1961. Spectrophotometric measurements of theaflavins and thearubigins in black tea liquors in assessment of quality of teas. Analyst, 86: 94-98.

Serban, C., A. Sahebkar and S. Ursoniu, F. Andrica and M. Banach, 2015. Effect of sour tea (*Hibiscus sabdariffa* L.) on arterial hypertension: A systematic review and meta-analysis of randomized controlled trials. J. Hypertens., 33: 1119-1127.

Sereshti, H., S. Samadi and M. Jalali-Heravi, 2013. Determination of volatile components of green, black, oolong and white tea by optimized ultrasound-assisted extraction-dispersive liquid-liquid microextraction coupled with gas chromatography. J. Chromatogr. A, 1280: 1-8.

Sharangi, A.B., 2009. Medicinal and therapeutic potentialities of tea (*Camellia sinensis* L.)-a review. Food Res. Int., 42: 529-535.

Sharma, M., D. Ghosh and N. Bhattacharya, 2013. Electronic nose - A new way for predicting the optimum point of fermentation of black tea. Int. J. Eng. Sci. Invent., 2: 56-60.

Ullah, M.R., N. Gogoi and D. Baurah, 1984. Effect of withering on fermentation of tea leaf and development of liquor characters of black tea. J. Sci. Food Agr., 35: 1142-147.

Yan, S.H., 2005. Evaluation of the composition and sensory properties of tea using near infrared spectroscopy and principal component analysis. J. Near Infrared Spec., 6: 313-325.

Yumei, L., 2012. Research on the dynamic comparative advantage of china tea foreign trade. Ph.D. Thesis, Southwest University, China.