Identification of Blueberry Beverage Using Vis/NIR Spectroscopy

Dongze Li, Guifang Wu*, Hai Ma, Zhao Liu, Guiquan Liu, and Junhua Hou

College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot, China

Abstract. Four kinds of blueberry beverage from different varieties, a total of 140 samples were acquired and analyzed by applying of spectrum technology. Using Savitzky-Golay spectral smoothing and multiplicative scatter correction (MSC) on the sample data pretreatment, four varieties of blueberry beverage were cluster analyzed by using principal component analysis method (PCA), a three-dimensional score view was achieved by the first 3 principal components of all samples (PC1, PC2 and PC3), which shows an obvious classification effect on the blueberry beverage. The first three principal components of the load diagram analysis, the characteristic bands related with the blueberry beverage varieties were 420-430nm, 490-500nm, 570-580nm and 1350-1365nm. According to the cumulative contribution rate (99.20%) of the first 6 principal components, the first 6 principal components was choosed as the input of multilayer perceptron (MLP) neural network, 100 samples in all the blueberry beverage samples were selected as a training set, and the remaining 40 samples were used as the prediction set. Training set were trained and prediction set were predicted by applying the multilayer perceptron neural network, and the correct rate of prediction were 100%. Research shows, using principal component analysis combined with multilayer perceptron neural network to identify the varieties of blueberry beverage is feasible.

Key words: Blueberry beverage, Vis/NIR spectroscopy, PCA, MLP neural network

1 Introduction

Blueberry, also known as bilberry, is called the King of Berries because of their special texture, sweet-tart taste, and a nutritious abundance in VC^[1-2]. The earliest cultivation of blueberries took place in North America, but that happened less than a hundred years ago^[3]. In China, the cultivation area of blueberries is mainly in Greater Khingan Mountains, Lesser Khingan Mountains, which are in northeast of China, and Shandong Peninsula. In Zhejiang, Hubei and Sichuan Basin and other areas also have a small amount of cultivation. With an abundance in vitamin, Superoxide Dismutase (SOD), arbutin and flavonoids compounds, blueberries are antioxidant food. It not only works in erythropoiesis synthesis, vision improvement, immune enhancement and cardiac strengthen^[4], but also helpful in medical efforts against aging, ulceration, inflammation and cardiovascular diseases^[5]. Wild blueberries contain many great substances such as folic acid and ursolic acid that are good for hypertension treatment, and anthocyanin that is capable of preventing and curing inflammation^[6-7]. Blueberries are also natural anti-cancer fruit because of the existence of many active constituents that inhibit the activity of cancer cell growth and even accelerate their apoptosis^[8]. For Healthy adults, a moderate drinking of blueberry beverage everyday could not only reduce the memory of the recession, but also can improve the body's ability to resist oxidative stress, while reducing the loss of lymphocyte DNA^[9].

Blueberry beverage is mainly made from blueberries, which gains much popularity among people. However, blueberry beverage in market is various and the identifications of blueberry beverage varieties are also complex and diverse^[10]. To better develop blueberry beverage, it's vital and necessary to find a fast and effective way among these complex and diverse identifications

Visible/near infrared spectroscopy is a spectroscopic method that bases on different results of the absorption of electromagnetic waves against different substances, which is typically applied in the field of chemistry, medical, agriculture and detection of agricultural products^[11-12]. And if NIR Spectroscopy is applied in the identification of blueberry beverage, there will be many advantages such as fast speed, high efficiency, no damage, high stability and accuracy^[13-14].

Principal Component Analysis (PCA), can effectively find the most important elements and structure among a large number of data. It can also remove the noise and redundancy to deal with the original complex data for dimension reduction processing so that the simple structure inside the complex data can be extracted. That is, without affecting the main spectral information, many original variables can be replaced with few variables. As one of neural network^[15-16]. Multilayer Perceptron (MLP) neural network, which is widely used in pattern recognition

* Corresponding author: wgfarsa@126.com

with good distribution ability, can solve the complex classification problem of pattern distribution^[17-18].

In this study visible/ near infrared spectroscopy (NIRS) and principal Component Analysis (PCA) with Multilayer Perceptron (MLP), as a new way for blueberry beverage identification, are used to analyze different kinds of blueberry beverage.

2 Materials and Methods Network

2.1 Instrument and Equipment

American ASD (Analytical Spectral Device) company Quality Spec spectrometer, its probe field of view angle is 25 degrees, the resolution is 3.5nm, spectral range of 350 ~ 1830nm, which sampling interval are 1.4nm @ 350 ~ 1000nm, 2nm @ 1000 ~ 1830nm, spectral resolution are 3nm @ 700nm, 10nm @ 1400nm. Analysis software is ASD View Spec Pro 6.0, Unscrambler 9.7, SPSS19.0.

2.2 Sample Source and Spectral Scan

A total of 35 blueberry beverages of four varieties of Genhe blueberry juice, Ye Laoda, Ye Shampo and Wild blueberry were collected randomly, samples were collected before the blueberry beverages were shaken well and 140 samples were prepared. The spectral scanning adopts transmission method with an optical path of 2 mm, each sample was scanned 30 times and three spectral curves were preserved, the average spectrum of the three spectral curves was used as the final transmission spectrum.

2.3 Preprocessing of spectral data

In order to reduce or eliminate the effects of noise, baseline drift and sample nonuniformity during spectral scanning, the original spectrum needs to be mathematically transformed to improve the accuracy and stability of the estimation model.

2.3.1 Smooth method

Savitzky-Golay convolution smoothing is an improvement on moving smoothing by multiplying the measured value by the smoothing factor to reduce the effect of smoothing on useful information and then fitting of the least squares method. The concrete formula is:

$$x_{k,s} = x_k = \frac{1}{H} \sum_{i=-w}^{+w} x_{k+i} h_i \quad (1)$$

In formula (1), h_i represent the smoothing factor, H represent the normalization factor.

2.3.2 Multiplicative scatter correction

When calculating the multiplicative scatter correction (MSC) for the acquired spectrum, the average spectrum is used instead of the "ideal" spectrum to the least squares method fit of each spectrum and the average spectrum so that the average spectrum can as much as possible becomes as linear relationship. The concrete formula is:

2.3.2.1 Calculate the average (ideal) spectrum of the sample

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (2)$$

2.3.2.2 Perform linear regression and the regression coefficients a_i and regression constants b_i were obtained by least squares method

$$X_i = a_i \bar{X} + b_i \quad (3)$$

2.3.2.3 Correct each spectrum

$$X_{i,M} = \frac{X_i - b_i}{a_i} \quad (4)$$

In formula(2)(3)(4), X_i represent the sample spectrum, n represent the number of samples, $\bar{X}$ represent the average spectrum, a_i represent the regression coefficients, b_i represent the regression constants, $X_{i,M}$ represent the corrected spectrum.

2.3.3 Preprocessed spectra

In the Unscrambler 9.7 software, Savitzky-Golay smoothing is used, the smoothing point is 9, and the multiple scattering correction (MSC) is used to preprocess the sample. The processing effect is the best. In order to remove the noise effects of the first and the end of the spectral curve, use the spectrum of the band 400 ~ 1800 nm^[19-21]. The spectral curves of the four blueberry beverages were obtained after pretreatment. As shown in Fig. 1, the spectral lines of one sample were arbitrarily selected in each blueberry beverage sample. Taking the abscissa as the wavelength of the spectrum and the ordinate as the absorbance of each sample.

Fig. 1. Spectral curve four varieties of blueberry beverage.

2.4 Principal component analysis

A total of 140 samples were cluster analyzed by using principal component analysis method of four blueberry beverages, blueberry juice, Ye Laoda, Ye Shanpo and Wild blueberry. As shown in Fig.2, in which the X, Y and Z axes respectively represent the scores of the first principal component (PC1), the second principal component (PC2) and the third principal component (PC3). The four kinds of blueberry beverages in Fig.2 were divided into four categories, indicating that PC1, PC2 and PC3 have very good clustering effect on four kinds of blueberry beverages also can qualitatively characterize the characteristics of the four blueberry beverages. But the edge sample distinction of Genhe blueberry juice, Ye Laoda and Ye Shanpo are not obvious. In order to improve the prediction accuracy, using principal component analysis combined with multilayer perceptron neural network to establish detection and analysis model of four kinds of blueberry beverage.

Fig. 2. Blueberry beverage principal component cluster diagram.

3.1 Characteristic bands selected

According to the principal component 1, 2, 3 in the entire wavelength range of the load value shown in Fig. 3. Taking abscissa as the wavelength of 400 ~ 1800nm, the ordinate as the load of different principal components in each wavelength variable values. The correlation between the principal component and the absorbance of the blueberry beverage at each wavelength. From Figure 3 can be obtained that range in 420 ~ 430nm, 490 ~ 500nm, 570 ~ 580nm and 1350 ~ 1365nm have maximum correlation with principal component. The main component 4-20 has maximum correlation with the absorbance of these four blueberry beverage's bands. Therefore, the four wavelength bands were selected as the characteristic bands of near infrared spectrum of the blueberry beverage over the entire wavelength range.

Fig. 3. The load of the principal component values.

3.2 Principal component analysis results

The cumulative contribution rate of the first six principal components has reached 99.204% by the principal component analysis method. As shown in Table 1, the spectral data of 140 samples were replaced by the first six principal components. All the samples were analyzed by the first six principal components.

Table 1. The cumulative contribution rate of the principal component.

Principal Component	PC1	PC2	PC3	PC4	PC5	PC6
Cumulative Contribution Rate	53.65 %	91.16 %	95.67 %	97.83 %	98.64 %	99.20 %

3 Test Results and Analysis

3.3 Identification of blueberry beverage based on multilayer perceptron neural network

The first six principal components obtained by principal component analysis are used as inputs to the multilayer perceptron (MLP) neural network. The three-layer network structure is chosen as the best training structure, and its structure is 6 nodes, the hidden layer is 5 nodes and the output layer is 4 nodes, with hyperbolic tangent as the activation function of the hidden layer, softmax as the activation function of the output layer. A total of 100 blueberry beverage samples were used as training set for training. The remaining 40 blueberry beverage samples were used as test set. The results show that, as shown in Table 2 ("1" is Genhe blueberry juice, "2" is Ye Laoda, "3" is Ye Shampo, "4" is Wild blueberry) accuracy of four kinds of blueberry beverage classification has reached 100%.

Table 2. Blueberry beverage classification prediction accuracy.

Blueberry Beverage Varieties	1	2	3	4	Accuracy Rate
1	10	0	0	0	100%
2	0	10	0	0	100%
3	0	0	10	0	100%
4	0	0	0	10	100%

4 Conclusion

Based on the combination of principal component analysis and multilayer perceptron(MLP) neural network, the model of blueberry beverage brand identification was established, and the accuracy of classification prediction was 100%.This fully illustrates the use of near infrared spectroscopy technology can quickly and accurately identify blueberry beverage varieties. This method of principal component analysis combined with multilayer perceptron (MLP) neural network is use the principal component as the input of multilayer perceptron(MLP) neural network, which can reduce the calculation of neural network, accelerate the training of the sample and eliminate the spectral signal interference, greatly improving the accuracy of the forecast. At the same time, the characteristic bands of the blueberry beverage brand are 420 ~ 430nm, 490 ~ 500nm, 570 ~ 580nm and 1350 ~ 1365nm., which are closely related to the beverage brands of Blueberry. Therefore, it is feasible to use the principal component analysis and the multi-layer perceptron neural network to identify the varieties of blueberry beverages. It also opens up a good prospect for the future development of blueberry beverage testing equipments.

Acknowledgment

This work was supported by National Natural Science Foundation of China (6126511); Higher School Doctoral Discipline Special Research Fund (20111515120004); Inner Mongolia Natural Science Foundation of China (2102MS0915); Inner Mongolia Agricultural University doctoral research start fund (BJ09-18).

References

1. A. Han Zhang. *Wild blueberry health care value*. Heilongjiang STI. **13**, 87 (2013)
2. K. Kim Mina, Han Sub Kwak. *Influence of functional information on consumer liking and consumer perception for blueberry functional beverages*. IJFST. **50**, 70-76 (2013)
3. Shiow Y. Wang, Hangjun Chen, J CAMP MARY, et al. *Genotype and heavy season influence blueberry antioxidant capacity and other quality attributes*. IJFST. **47**, 1540- 1549 (2012)
4. J. Barba Francisco, J. Esteve Maria, Frigola Ana. *Physicochemica and nutritional characteristics of blueberry juice after high pressure processing*. FRI. **50**, 545-549 (2013)
5. Ying Wang, Shu Wang, Chunguang Ren, etc. *Blueberry wine fermentation process and quality of the relationship between the study*. Guizhou Science. **31**, 75-78 (2013)
6. Pengxiang Han, Bei Zhang, Xuqiao Feng, etc. *Blueberry nutrition and health functions and its development and utilization*. FIT. **36**, 370-375,379 (2015)
7. Yangcheng Song, Yujuan Chen, Hao Li, etc. *Blueberries in the rapid detection of anthocyanins*. Food Science. **31**, 334-336 (2010)
8. Lihua Pan, Jianfei Wang, Xinggan Ye, etc. *Blueberry anthocyanin extraction process and its immunomodulatory activity*. Food Science. **35**, 81-86 (2014)
9. Haiyan Gao, Long Xu, Hangjun Chen, etc. *Blueberry postharvest quality control and anti-oxidation research progress*. Chinese JFS. **6**, 1-8 (2013)
10. Haige Li, Yanjun Wu, Yange Yang, etc. *Advances in technology research to identify the authenticity of berry juice*. Food Science. **37**, 243-250 (2016)
11. Guifang Wu, Hai Ma, Xin Pan. *Identification of varieties of natural textile fiber based on vis / NIR infrared spectroscopy*. IAEAC.Institute of Electrical and Electronics Engineers Inc, 585-589 (2015)
12. R. BEGHI, V. GIOVENZANA, A.SPINARDI, etc. *Derivation of a blueberry ripeness index with a view to a low-cost, handheld optical sensing device for supporting harvest decisions*. Transaction of the ASABE. **56**, 1551 -1559 (2013)
13. Guifang Wu, Wei Jiang, Chunguang Wang, etc. *Quality inspection of tomato shelf life based on TPA and Vis / NIR*. MFST. **31**, 290-294 (2015)
14. Guifang Wu, Chunguang Wang. *Investigating the effects of simulated transport vibration on tomato tissue damage based on Vis / NIR spectroscopy*. PBT. **98**, 41-47 (2014)
15. Shirong Ai, Ruimei Wu, Yan Wu. *Study on the Identification of Tea Beverage by Near Infrared Spectroscopy Based on BP Neural Network*. Anhui JAS. **14**, 7658-7659, 7662 (2010)

16. Daniel Cozzolino, Wies Cynkar, Nevil Shah, etc. *Varietal differentiation of grape juice based on the analysis of near- and mid-infrared spectral data*. Analytical Methods. **5**, 381- 387 (2012)
17. Royston Goodacre, David Hammond, B. Kell Douglas. *Quantitative analysis of the adulteration of orange juice with sucrose using pyrolysis mass spectrometry and chemometrics*. JAAP. **40-41**, 135-158 (1997)
18. I. Suarez Gracirla, A.Ortiz Oscar, M. Aballay Pablo, etc. *Adaptive neural model predictive control for the grape juice concentration process*. // International Conference on Industrial Technology. Universidad Tecnica Federico Santa Maria. Institute of Electrical and Electronics Engineers Inc. 57-62 (2010)
19. Guifang Wu, Yixiang Jiang, Yanyan Wang, etc. *Identification of dry red wine varieties based on independent principal component and BP neural network*. SSA. **29**, 1268-1271 (2009)
20. Jianhu Wu, Juntao Lei, Qi Yang. *Identification of Dry Date by Using Visible / Near Infrared Spectroscopy*. JFSQI. **7**, 1870-1875 (2016)
21. Yulu Wang, Hui Guo, Changjie Han. *Study on identification method of different mutton based on visible / near infrared spectroscopy*. FST. **40**, 298-302 (2015)