Rapid Classification of Honey Varieties by Surface Enhanced Raman Scattering Combining with Deep Learning

Zheng Fang, Wen Wang, Aixia Lu, Ying Wu, Ying Liu, Changchun Yan, Caiqin Han*
Jiangsu Key Laboratory of Advanced Laser Materials and Devices
School of Physics and Electronic Engineering, Jiangsu Normal University Xuzhou, China
*hancq@jsnu.edu.cn

***Abstract*—A new method based on surface enhanced Raman spectroscopy and deep learning is developed to classify the varieties of honey. This method used the difference of honey SERS spectrum to classify. Spectra of honey were accessed through preprocess. The classification method is found to be an effective approach to analyze substances, especially similar substances, so as to improving the accuracy of classification. Experimental results demonstrate that this approach is suitable for classification of similar substances.**

Keywords—classification; SERS; deep learning; honey

I. INTRODUCTION

Honey is a sweet, viscous substance produced from flowers by bees and some other insects. Every 100-gram honey provides 304 kilocalories. Composed of 17% water and 82% carbohydrates, honey has low content of fat, dietary fiber, and protein. There are over 300 types of honey so honey adulteration could be a big problem [1]. Therefore, it is important to identify the different varieties of honey. Some methods had been adopted to classification such gas chromatography-mass spectrometer (GC-MS), high performance liquid chromatography (HPLC) and Fourier transform mid-infrared spectroscopy. Nevertheless, many sample pretreatments are needed while using those methods. Hence, finding an efficient and simple method is required.

Raman spectroscopy is a powerful spectroscopic technique for qualitative or quantitative analysis of a wide range of substances. However, the intensity of Raman spectroscopy is not strong enough. Thus, Raman spectroscopy is not well known until the occurrence of Laser and the discovery of Surface Enhance Raman Scattering (SERS) [2]. SERS is a surface-sensitive technique that enhances Raman scattering by molecules adsorbed on rough metal surfaces such as Ag, Au and Cu. In recent years, SERS was used to identify many substances and has been used in honey analysis [3]. However, there is no research using SERS to classify different varieties of honey.

Many different classification methods have been proposed to deal with SERS spectrum. The traditional classification algorithms include k-nearest-neighbors (KNN), maximum-likelihood, support vector machine (SVM) and logistic regression. The majority of those above algorithms suffer from the "curse of dimensionality". To deal with the high dimensionality, some dimensionality reduction-based methods were proposed such as principal component analysis (PCA) and sparse representation. However, it will cause the feature loss after reducing dimensionality. Deep learning is a part of machine learning methods [4]. The idea of deep learning also known as feature learning. That means the feature can be learned automatically from input data without dimensionality reduction.

In this paper, we develop a method to classify the varieties of honey rapidly by SERS combining with deep learning. This method can rapidly classify honey without any sample pretreatment. So this method can be used in portable instrument. Using this method, spectroscopy data doesn't need dimensionality reduction. It can reduce the loss of features and improve the accuracy of classification.

II. MATERIALS AND METHODS

A. Sample preparation

The experimental material in the study consisted of six varieties of honey from different kinds of flowers including acacia honey, wolfberry honey, codonopsis honey, jujube honey, jinghua honey and eucalyptus honey. The tested honey samples were produced by the same manufacturer.

B. Silver sol preparation [5]

In this study, we used following method to product silver sol:

1) *$AgNO_2$ (90 mg) was dissolved in 500 mL of H_2O.*
2) *Brought the solution to boiling.*
3) *Adding 1% sodium citrate (10mL) to solution.*
4) *Stirring for 1h.*

After stirring for 1h, we can achieve yellowish green silver sol solution.

C. Processing of Raman spectroscopy

The strong fluorescence would reduce signal-to-noise ratio of Raman spectrum and disturb the identification of honey.

Also, the baseline will disturb the classification of honey. Hence, baseline emendation is important in Raman spectroscopy analysis. Some algorithms such as adaptive iteratively reweighted penalized least squares (airPLS)[6] were adopted to emendate baseline. The following algorithm provides the main steps of the proposed approach called airPLS. The steps of the algorithm are as the following:

1) Input Raman spectroscopy vector x and Parameter λ *which control the smoother of resulting background.*

2) assigning $w^0=1$.

3) fit y^0 *using* $w^0=1$ *and penalized least squares.*

4) If not reached the terminative criterion, reweight w^t *using* y^{t-1}.

5) fit y^t *using* $w^t=1$ *and penalized least squares.*

6) Iteration will stop either with the maximal iteration times or when the terminative criterion is reached.

7) finally fitted baseline y.

The airPLS algorithm provides a simple but effective method for estimating baselines for Raman spectroscopy. Fig.1 shows the result of the airPLS algorithm.

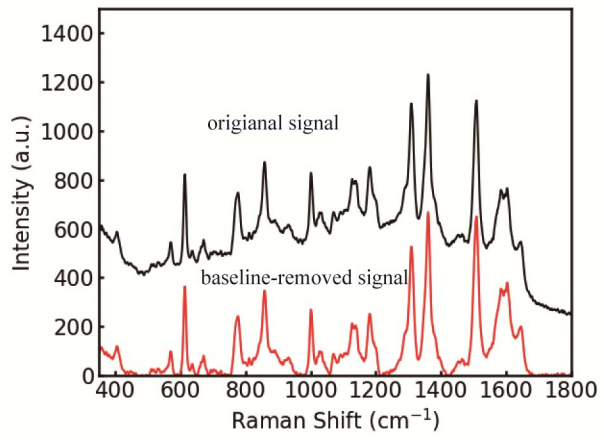

Fig.1. Result of baseline emendation

Since Raman spectrum intensities differ between two measurements due to different experimental conditions, a normalization procedure is required. The following formula shows the procedure of normalization:

$$y = \frac{x - \text{minValue}}{\text{maxValue} - \text{minvalue}} \tag{1}$$

Here x is a Raman spectroscopy vector, minValue and maxValue are minimum and maximum value of x, respectively.

D. Deep learning

A typical deep learning model is a deep neural network (DNN) with multiple hidden layers [7], which is shown in Fig.2.

The neural network layers of DNNs can be divided into three types according to the position of different layers: input layer, hidden layer and output layer. The first layer is the input layer, and the last layer is the output layer, and the middle layers are hidden layers.

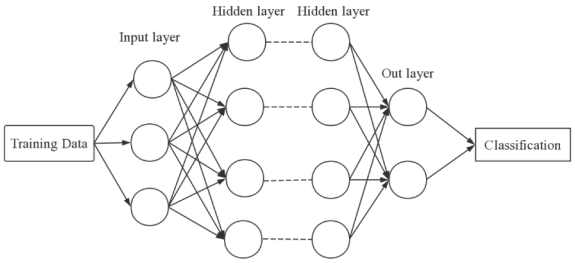

Fig.2. Model of deep neural network

The layers are fully connected, that means any one of the neurons in the ith layer must be connected to any one of the neurons in the $(i+1)$th layer. Although the DNN looks complicated but it is the same as the perceptron: a linear relationship couple with an activation function $\sigma(x)$. The following formula shows the linear relationship:

$$y = \sum w_i x_i + b \tag{2}$$

Where w is the linear coefficient and b is the deviation. All DNNs need to be trained to make them have more generalization ability by using a dataset to train DNNs.

III. Result and Discussion

Raman spectra were collected by ProRaman-L analyzer from Enwave Optronics with a near infrared diode laser and excitation line 785nm. The laser power was set at 100 mW and the integral-time was set at 10 seconds to measure spectrum of honey. The Raman shift range 350 cm^{-1} to 1800 cm^{-1}. SERS spectra of honey are shown in Fig.3.

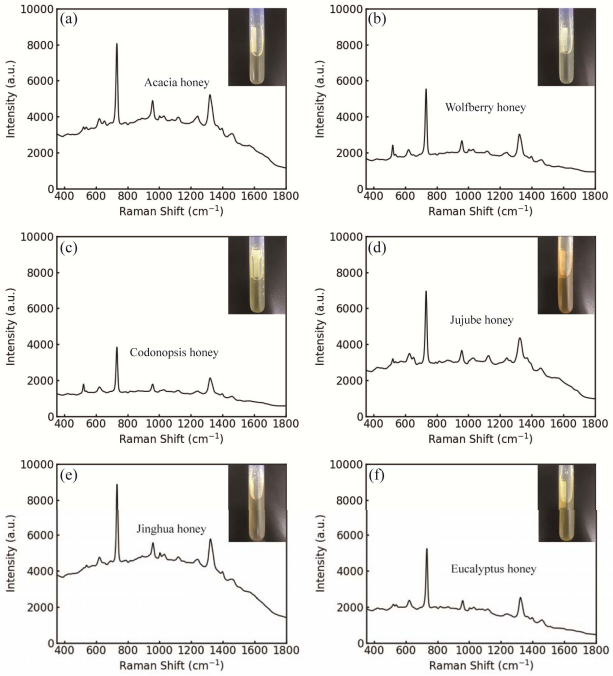

Fig.3. SERS spectra of these six varieties of honey. (a) acacia honey; (b) wolfberry honey; (c) codonopsis honey; (d) jujube honey; (e) Jinghua honey; (f) eucalyptus honey.

Due to the fluorescence, the features of spectrum cannot be found easily. With those spectra, the rate of identification will decrease. The process of removing fluorescence and baseline is essential. Fig.4 shows the SERS spectra of honey after baseline emendation.

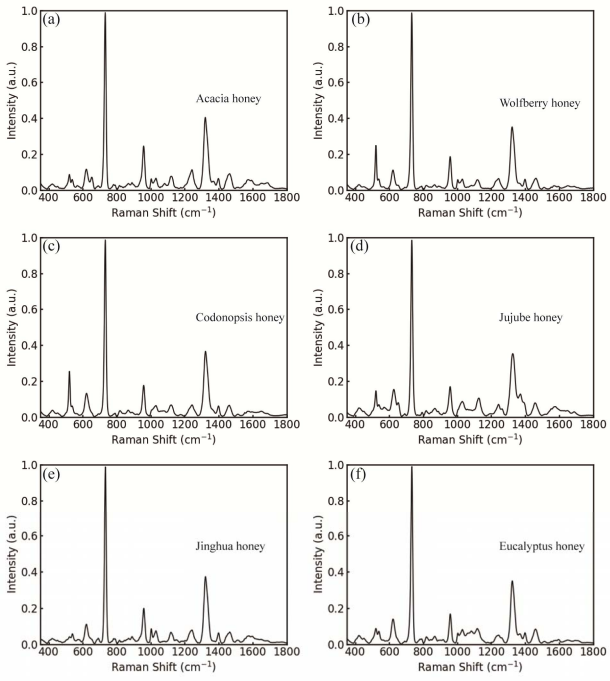

Fig.4. Processed SERS spectra of these six varieties of honey. (a) acacia honey; (b) wolfberry honey; (c) codonopsis honey; (d) jujube honey; (e) Jinghua honey; (f) eucalyptus honey.

The principal component of honey are saccharides and water. Hence, the SERS spectra of honey are similar. However, the SERS spectra are different at 520cm^{-1}, 656cm^{-1}, 1003cm^{-1} and 1371cm^{-1}. Fig.5 shows the comparison among SERS spectrum of these six varieties of honey.

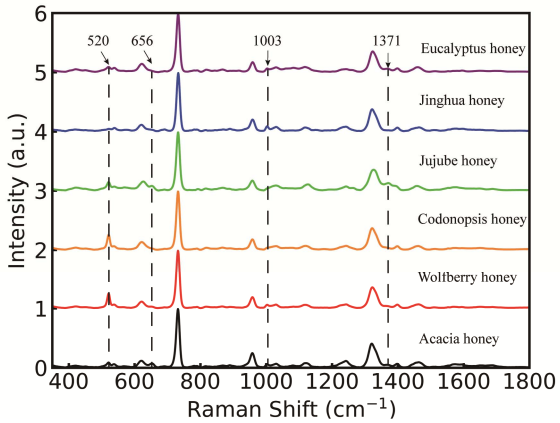

Fig.5. Comparison among SERS spectrum of these six varieties of honey

The data through preprocessing was divided into two sets training set and test set. In this study, training set consist of 540 honey SERS spectra and test set consist of 180 honey SERS spectra. The DNN which has ten hidden layers was established. The result of classification was shown in Table.1. The aggregate accuracy is 96.7% and the accuracy can meet the requirements of actual detection.

TABLE I RESULT OF CLASSIFICATION

Sample name	Total sample	Accuracy sample	Accuracy
Acacia honey	30	29	96.7%
Wolfberry honey	30	30	100%
Codonopsis honey	30	30	100%
Jujube honey	30	28	93.3%
Jinghua honey	30	29	96.7%
Eucalyptus honey	30	28	93.3%
Aggregate	180	174	96.7%

IV. CONCLUSION

In this work, we have presented an efficient and novel method for classifying honey varieties rapidly. This approach is suitable for classification, and its efficiency is better than other method in some aspects. Our method can apply to portable devices for rapid detection. The classification accuracy can also meet the requirements of actual detection. Also, the method can be expanded to classification of other similar substances.

ACKNOWLEDGMENTS

This research is funded by the National Natural Science Foundation of China (Grant No. 61575087), the Natural Science Foundation of Jiangsu Province (Grant No. BK20151164). The authors would like to thank a Project Funded the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

REFERENCES

[1] A. J. Siddiqui, S. G. Musharraf, M. I. Choudhary, and A. U. Rahman, "Application of analytical methods in authentication and adulteration of honey," *Food Chemistry,* vol. 217, pp. 687-698, 2017.

[2] M. Moskovits, "Surface-enhanced spectroscopy," *Review of Modern Physics,* vol. 57, pp. 783-826, 1985.

[3] M. F. Raduan, W. Mansor, K. Y. Lee, A. R. M. Radzol, M. F. Raduan, W. Mansor, *et al.*, "Analysis of honey using surface enhanced Raman spectroscopy," in *IEEE Symposium on Industrial Electronics & Applications*, 2017, pp. 29-31.

[4] G. E. Hinton, S. Osindero, and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation,* vol. 18, pp. 1527-1554, 2014.

[5] P. C. Lee and D. Meisel, "Adsorption and surface-enhanced Raman of dyes on silver and gold sols," *Journal of Physical Chemistry,* vol. 86, pp. 3391-3395, 1982.

[6] Z. M. Zhang, S. Chen, and Y. Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst,* vol. 135, p. 1138, 2010.

[7] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, p. 436, 2015.