

Classification of Blood Species Using Fluorescence Spectroscopy Combined with Deep Learning Method

Jianhong Gan^{1,2}, Linhua Zhou^{1,2*}, Jian Cui^{1,2}, Boqi Man^{1,2}, Xiaoning Jia^{1,2}, Sanzhi Shi^{1,2}, Linna Liu³

¹School of Science, Changchun University of Science and Technology, Changchun, China

²Provincial Demonstration Center for Experimental Mathematics Education (Changchun University of Science and Technology), Changchun, China

³Chinese Academy of Agricultural Sciences, Changchun Veterinary Research Institute, Changchun, China

Email: *zhoulh@cust.edu.cn

How to cite this paper: Gan, J.H., Zhou, L.H., Cui, J., Man, B.Q., Jia, X.N., Shi, S.Z. and Liu, L.N. (2019) Classification of Blood Species Using Fluorescence Spectroscopy Combined with Deep Learning Method. *Journal of Applied Mathematics and Physics*, 7, 2324-2332.

<https://doi.org/10.4236/jamp.2019.710158>

Received: September 5, 2019

Accepted: October 12, 2019

Published: October 15, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this work, a deep belief neural network model (DBN) was developed to classify doves, chickens, mice and sheep blood samples, which have many similarities in composition causing their spectra to look almost identical by visual comparison alone. The DBN model was formulated for the feature extraction from the pretreated fluorescence spectroscopy. Then, cross-validation results showed that the application of deep learning method made it possible to classify the blood fluorescence spectroscopy in a more precise way than previous methods. Especially, the classification accuracy of whole blood with 1% of concentration was up to 97.5%.

Keywords

Neural Network Model, Deep Learning, Classification, Blood Species, Fluorescence Spectroscopy

1. Introduction

In recent decades, with the rapid development of spectroscopy technology, it has been applied to many fields, especially to life sciences. For example, it is widely used in diagnosing disease of biological tissue and also has been used to determine the species of origin of the blood sample [1]-[6]. The identification of blood spectroscopy is an important aspect of forensic investigation, wildlife investigation and import/export detection. The current methods used to identify the blood spectroscopy mainly include HPLC [7] (High Performance Liquid Chromatography) and LC-MS [8] (Liquid Chromatography-mass Spectrometry).

try), etc. These methods are featured with complex operation procedures and low recognition efficiency and easy to cause contamination of blood samples.

In 2008, De Wael *et al.* [9] tested the Raman spectrum and Near-infrared spectrum of human, canine, and feline. They found it is very difficult to visually distinguish the difference among the blood samples of different species. However, in 2009, based on advanced statistical method, Virkler *et al.* [10] developed a three-dimensional score plot that shows separation among Raman spectrum of human, canine, and feline blood samples. In 2017, Lu *et al.* [11] performed blood species classification using fluorescence spectroscopy and support vector machine (SVM), and achieved a remarkable result. In 2018, Dong *et al.* [12] classified the blood samples of human, dogs and rabbits by method of machine learning, with over 90% accuracy.

Although, in the past decade, great progress has been made in the study of classification of blood spectroscopy, to develop a quick and high accuracy analytical technique to classify the blood spectroscopy which is still an important task now. In recent years, the artificial intelligence technology of deep learning has evolved as one of the powerful tools in many fields [13]. It's worth mentioning that deep learning has given conspicuous performance with respect to the automatic feature extraction [14], which endows it with broad and efficient application prospects in analysis of spectral data.

In this work, the combination of fluorescence spectroscopy and deep learning method was used to classify doves, chickens, mice and sheep blood samples, which have many similarities in composition causing their spectra to look almost identical by visual comparison alone. A deep belief neural network model (DBN) is developed for feature extraction from the pretreated data, and the cross-validation results show that the application of deep learning method makes it possible to classify the blood species in a more precise way than previous methods.

The paper is organized as follows. In Section 2, the fluorescence spectrums of whole blood and red blood cell solutions with different concentrations from doves, chickens, mice and sheep are presented and pretreated. In Section 3, a Deep Belief Network (*i.e.*, DBN) [15] [16] [17] model based on a stacking of Restricted Boltzmann Machines (*i.e.*, RBMs) is established to extract the deep learning feature of blood spectrums. Besides, the label-layer of DBN is established to adjust the parameters, so as to obtain a classifier after training, which makes it possible to classify the blood species in a more precise way than traditional methods. In Section 4, the paper ends with some discussions about the methods and future work.

2. Experiment and Data Pretreatment

Fluorescence spectrum is capable of detecting the chemical structures of molecules in the organism and the interaction between different molecules and surroundings, which allows researchers to achieve the cellular composition and

metabolism of organisms via the study of fluorescence spectrum [18]. In this work, the whole blood and red blood cell samples of doves, chickens, mice and sheep are collected. Then, 0.85% NaCl normal saline solution is used to dilute the blood samples into the solution with 1% and 3% of concentration. Finally, Cary Eclipse fluorescence spectrophotometer (See **Table 1** for parameter setting) is used to scan the stand-by blood samples at ambient temperature and collect the spectrum data (See **Table 2**, 50 groups of data are collected from the whole blood and red blood cell solutions (with two different concentrations) of four animals).

In view of that, the fluorescence spectrum data collected includes a mass of noise, the “db3” wavelet mother function, “Heursure” threshold selection rule and “sln” threshold adjustment approach are selected to perform de-noising processing of the original spectral data. The de-noised and normalized spectral curves are shown in **Figure 1**.

3. Methodology and Classification Results

3.1. Methodology

Methodologically, after raw data of fluorescence spectrum is pretreated by wavelet de-noising and normalization, RBM is used to extract the abstract features in the training dataset, and the tuning classifier with label information is used to accomplish the overall training of DBN. Finally, the trained DBN model is used to perform the recognition and classification in the testing dataset (see **Figure 2**).

3.2. DBN Model Formulation and Hyper-Parameter Adjustment

As a generative model proposed by Geoffrey Hinton in 2006, a Deep Belief Network (DBN) is developed by a stack of RBMs (Restricted Boltzmann Machines).

RBM is an energy-based neural network model with a two-layer structure (the visible layer and hidden layer). The neurons between visible layer and hidden layer are completely connected and there is no any connection within the layers. DBN is developed by a stack of several RBMs.

The neurons of RBM belongs to Boolean ones whose values are 0 or 1 (0 = Activated; 1 = Non-activated). The energy function of RBM is defined as:

$$E_{\theta}(v, h) = -\sum_{i=1}^{n_v} a_i v_i - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} h_j \omega_{j,i} v_i \quad (1)$$

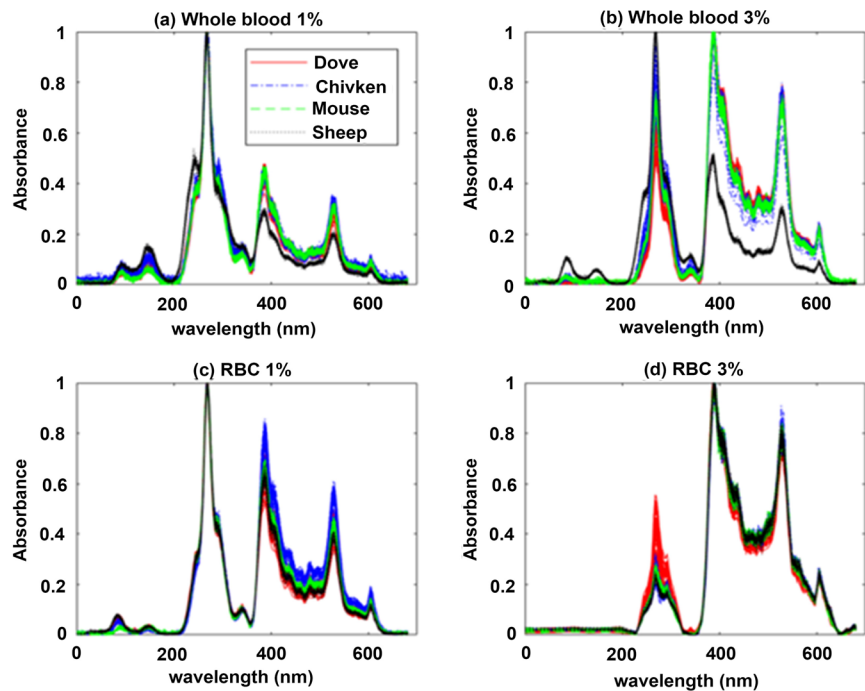
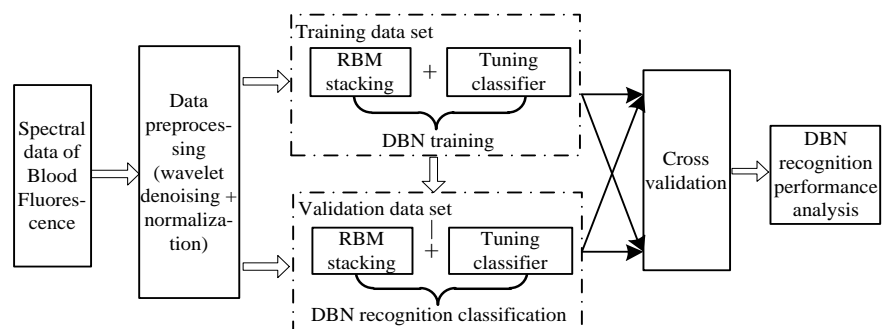
In which, v refers to the state of visible layer, while n_v is the number of neurons in visible layer; h refers to the state of hidden layer, while n_h is the number

Table 1. Parameters of Cary Eclipse fluorescence spectrophotometer.

Laser wavelength	200 nm	Scanning speed	600 nm/min
Laser slit	5 nm	Data interval	1 nm
Emission slit	10 nm	Average time	0.1 s

Table 2. Data description (Note: See the original data in appendix file).

Blood sample	Total Number of samples	Number of Train samples	Number of Validation samples
Dove	50	40	10
Chicken	50	40	10
Mouse	50	40	10
Sheep	50	40	10

**Figure 1.** De-noised and normalized spectral curve. (a) 1% concentration whole blood solution; (b) 3% concentration whole blood solution; (c) 1% concentration red blood cell solution; (d) 3% concentration red blood cell solution.**Figure 2.** Flow chart of the classification method.

of neurons in hidden layer; a refers to the offset of neurons in visible layer; b refers to the offset of neurons in hidden layer; $w_{j,i}$ represents the weight of connection between neurons in visible layer and hidden layer. Suppose that $\theta = \{W_{ij}, \alpha_i, b_j\}$ is a parameter to be learned by RBM. After the confirmation of

parameter, the joint probability distribution of (v, h) based on the energy function is:

$$p_{\theta}(v, h) = \frac{1}{z_{\theta}} e^{-E_{\theta}(v, h)} \quad (2)$$

In which, $z_{\theta} = \sum_{v, h} e^{-E_{\theta}(v, h)}$ is the normalization factor. The boundary distribution of joint probability distribution about v (also known as the likelihood function) can be expressed as:

$$p_{\theta}(v) = \sum_h p_{\theta}(v, h) = \frac{1}{z_{\theta}} \sum_h e^{-E_{\theta}(v, h)} \quad (3)$$

During the learning process of RBM, suppose that there are N sample data, and the value of parameter θ can be confirmed by maximizing the log-likelihood function:

$$\theta^* = \arg \max_{\theta} L(\theta) = \arg \max_{\theta} \sum_{n=1}^N \log P_{\theta}(v^{(n)}) \quad (4)$$

The training process of DBN is divided into two steps: 1) Firstly, the unsupervised training of RBM network is conducted layer by layer so as to ensure that more feature information can be retained when the feature vector is mapped to different feature spaces; 2) The tuning classifier with labels is set to slightly adjust whole DBN network based on back-propagation algorithm.

Structure of deep neural network and hyper-parameter adjustment are two most prominent aspects of DBN construction. A DBN stacked by 3 RBMs is designed in the paper, and the number of nodes at each layer is 681-50-50-200 (see **Figure 3**). According to the results of experimental comparison, the improvement of DBN's depth (the number of layers stacked by RBM) and width (the number of nodes at each layer of RBM) is found to haven't produced higher recognition accuracy rate, but significantly reduces the time efficiency of recognition.

Besides, DBN model in **Figure 3** is subject to hyper-parameter adjustment. As for the learning rate and momentum factor, multiple groups of cross-validation experiments and adjustments on the 1% whole blood training set are carried out, with the learning rate of 0.1 and the momentum factor of 0.5 (see **Table 3**).

3.3. Classification Results and Analysis

In this paper, half of the spectral data of the same blood sample (whole blood or red blood cell), at the same concentration that is collected from the same animal, is randomly selected as the training data. While the rest half data are used as the test data. Secondly, the formulated DBN model and the label-layer of DBN (*i.e.*, the classifier) are trained by the training data. Then, the test data of doves, chickens, mice and sheep is recognized by the trained classifier. Finally, the training set and test set are interchanged to perform cross validation and obtain the average classification accuracy (See **Table 3**).

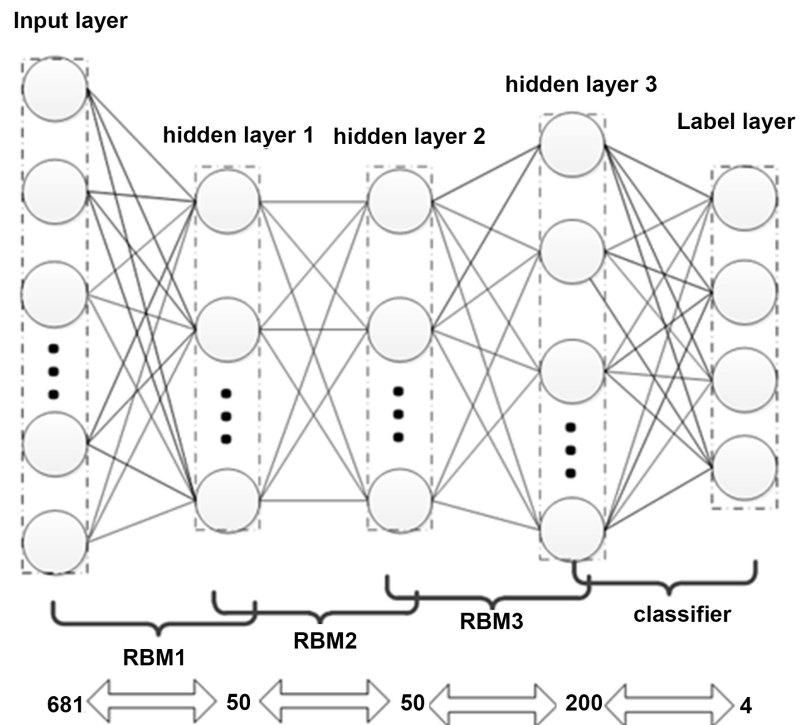


Figure 3. Structure of the deep belief network model.

Table 3. Hyper-parametric adjustment.

Learning rate	Hidden layer	Momentum factor	Classification accuracy	Standard deviation	Selection or not
0.1	2	0.5	96.50%	0.95	No
		0.9	95.5%	0.81	No
0.01	2	0.5	96.62%	0.34	No
		0.9	96.5%	0.71	No
0.1	3	0.5	98.38%	0.56	Yes
		0.9	95.75%	1.03	No
0.01	3	0.5	96.37%	0.52	No
		0.9	97.36%	0.86	No

The classification accuracy of 1% whole blood is achieved 97.5%, with a standard deviation of 2.5, while the classification accuracy of 3% whole blood is 96.5%, with a standard deviation of 2.09. The classification accuracy of 1% and 3% red blood cells are 85% and 82.5%, respectively, with the standard deviations of 8.67 and 2.85 accordingly. According to **Table 3**, the classification accuracy of whole blood solutions is higher than those of red blood cell solutions, which prove that the whole blood solution shall be used to identify the blood species of animals in reality.

All samples of mice and sheep can be recognized accurately under the circumstance of different blood types and concentrations. Among the samples that

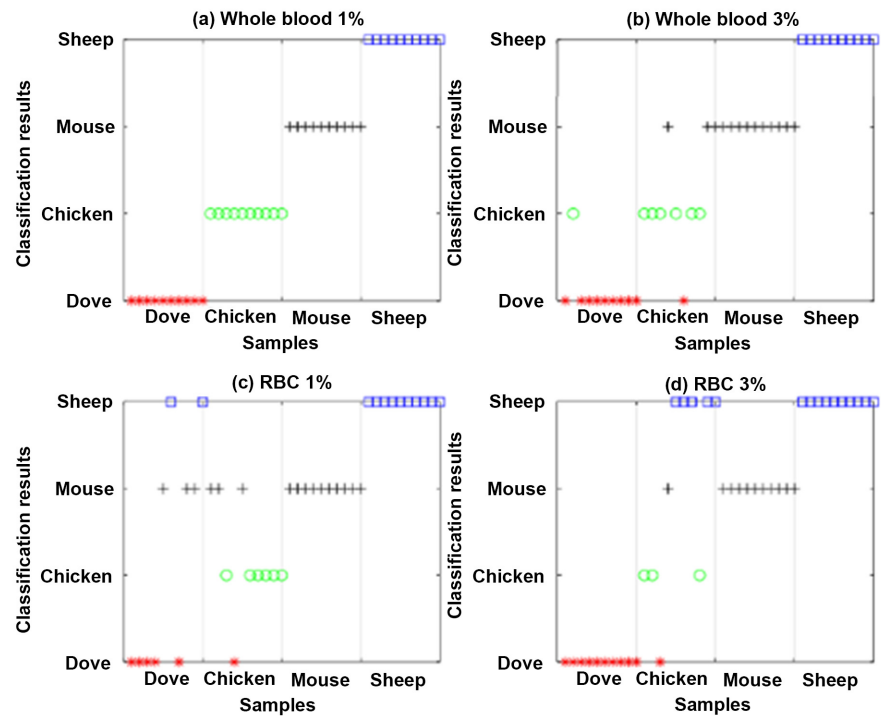


Figure 4. Classification results based on the deep belief network. (a) 1% concentration whole blood solution; (b) 3% concentration whole blood solution; (c) 1% concentration red blood cell solution; (d) 3% concentration red blood cell solution.

Table 4. Cross validation results of the deep belief network.

Classification	Cross validation accuracy	Standard deviation	Classification	Cross validation accuracy	Standard deviation
Whole blood 1%	97.5%	2.5	RBC 1%	85%	8.67
Whole blood 3%	96%	2.09	RBC 3%	79%	2.85

identified the wrong classification, doves were the most likely to misjudge mice, followed by sheep. Whereas the blood spectrum of chickens is quite similar to that of mice and sheep, small part of chicken samples is mistaken for mice and sheep. But even so, most samples of doves, chickens, mice and sheep can be recognized accurately in the case of 1% whole blood (see **Figure 4** and **Table 4**).

4. Conclusions

In recent years, the artificial intelligence technology of deep learning has evolved as one of the powerful tools for many fields. With the rapid development of spectroscopy technology, there is no doubt that spectral data analysis using advanced machine learning method will become a new direction in spectral analysis. In this work, based on fluorescence spectroscopy combined with artificial intelligence technology of deep learning method, a novel analytical approach for classification of blood species was developed to classify doves, chickens, mice and sheep blood samples. A deep belief neural network model stacked by three layers of RBMs is used for the feature extraction from the fluorescence spectral

data. Finally, by the trained DBN classifier, the cross-validation results show that the application of deep learning method makes it possible to classify the blood species in a more precise way than previous methods. Especially, the classification accuracy of whole blood with 1% of concentration is up to 97.5%. Compared with other models of blood spectrum classification, DBN can extract effective feature information preferably so that better classification (*i.e.* higher classification accuracy) is achieved. Therefore, the approach proposed in this paper has great potential in classification of blood species in forensic investigation, wildlife investigation and import/export detection.

However, in the current work, we just classify four different kinds of animals with the fluorescence spectroscopy combined with DBN model. In the future work, we shall collect spectral data of more kinds of animals with different blood concentrations and try to formulate a more useful machine learning model, so as to provide more quick and effective identification and classification methods for forensic investigation, wildlife investigation and import and export detection, etc.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of P. R. China (No. 11401092, 11426045), Scientific and Technological Planning Project of Jilin Province of P. R. China (No. 20180101229JC), Foundation of Jilin Educational Committee (No. JJKH20181100KJ).

Appendix File

The original fluorescence spectroscopy of four kinds of animals with different concentrations is attached.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Frank, C.J., McCreery, R.L. and Redd, D.C.B. (1995) Raman Spectroscopy of Normal and Diseased Human Breast Tissues. *Analytical Chemistry*, **67**, 777-783. <https://doi.org/10.1021/ac00101a001>
- [2] Sun, Y., Liu, R., Elson, D.S., *et al.* (2008) Simultaneous Time- and Wavelength-Resolved Fluorescence Spectroscopy for Near Real-Time Tissue Diagnosis. *Optics Letters*, **33**, 630-632. <https://doi.org/10.1364/OL.33.000630>
- [3] Sikirzhyski, V., Virkler, K. and Lednev, I.K. (2010) Discriminant Analysis of Raman Spectra for Body Fluid Identification for Forensic Purposes. *Sensors*, **10**, 2869-2884. <https://doi.org/10.3390/s100402869>
- [4] Sikirzhyskaya, A., Sikirzhyski, V. and Lednev, I.K. (2017) Determining Gender by Raman Spectroscopy of a Bloodstain. *Analytical Chemistry*, **89**, 1486-1492. <https://doi.org/10.1021/acs.analchem.6b02986>

- [5] McLaughlin, G., Doty, K.C. and Lednev, I.K. (2014) Raman Spectroscopy of Blood for Species Identification. *Analytical Chemistry*, **86**, 11628-11633. <https://doi.org/10.1021/ac5026368>
- [6] Zhang, L., Zhou, M., Li, X., Li, G. and Lin, L. (2014) Discrimination of Human and Nonhuman Blood Using Visible Diffuse Reflectance Spectroscopy. *Analytical Methods*, **6**, 9419-9423. <https://doi.org/10.1039/C4AY01673D>
- [7] Andrasko, J. and Rosā, N.B. (1994) Sensitive Identification of Hemoglobin in Bloodstains from Different Species by High Performance Liquid Chromatography with Combined UV and Fluorescence Detection. *Journal of Forensic Sciences*, **39**, 1018-1025. <https://doi.org/10.1520/JFS13680J>
- [8] Van Steendam, K., De Ceuleneer, M., Dhaenens, M., Van Hoofstat, D. and Deforce, D. (2013) Mass Spectrometry-Based Proteomics as a Tool to Identify Biological Matrices in Forensic Science. *International Journal of Legal Medicine*, **127**, 287-298. <https://doi.org/10.1007/s00414-012-0747-x>
- [9] De, W.K., Lepot, L., Gason, F. and Gilbert, B. (2008) In Search of Blood-Detection of Minute Particles Using Spectroscopic Methods. *Forensic Science International*, **180**, 37-42. <https://doi.org/10.1016/j.forsciint.2008.06.013>
- [10] Virkler, K. and Lednev, I.K. (2009) Blood Species Identification for Forensic Purposes Using Raman Spectroscopy Combined with Advanced Statistical Analysis. *Analytical Chemistry*, **81**, 7773-7777. <https://doi.org/10.1021/ac901350a>
- [11] Deng, L. and Yu, D. (2014) Deep Learning: Methods and Applications. *Foundations and Trends in Signal Processing*, **7**, 197-387. <https://doi.org/10.1561/20000000039>
- [12] Gong, M., Yang, H. and Zhang, P. (2017) Feature Learning and Change Feature Classification Based on Deep Learning for Ternary Change Detection in SAR Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, **129**, 212-225. <https://doi.org/10.1016/j.isprsjprs.2017.05.001>
- [13] Lu, P., Fan, Y., Zhou, L., *et al.* (2017) Feature Extraction and Classification of Animal Blood Spectra with Support Vector Machine. *Spectroscopy and Spectral Analysis*, **37**, 3828-3832.
- [14] Dong, J., Hong, M., Zheng, X., *et al.* (2018) Discrimination of Human, Dog and Rabbit Blood Using Raman Spectroscopy. *Spectroscopy and Spectral Analysis*, **38**, 459-466.
- [15] Zhang, Y. (2016) Research of Deep Learning Method Based on Restricted Boltzmann Machine. Beijing University of Chemical Technology, Beijing.
- [16] Ma, S., Shen, T., Wang, R., *et al.* (2015) Terahertz Spectroscopic Identification with Deep Belief Network. *Spectroscopy and Spectral Analysis*, **35**, 3325-3329.
- [17] Lu, W., Guo, Y., Dai, J., *et al.* (2018) Rice Germination Rate Detection Based on Fluorescent Spectrometry and Deep Belief Network. *Spectroscopy and Spectral Analysis*, **38**, 1303-1312.
- [18] Wang, S., Xu, X., Liu, Q., *et al.* (2001) The Application of Fluorescence Spectroscopy in the Study on Protein Conformation. *Progress in Chemistry*, **13**, 28-31.