Automatic Identification of Nanfeng Mandarin by Using a Stacked Autoencoder-Based Deep Learning Algorithm with Smartphone

Zhiyong Xiao and Ya-Ru Liang

December 6, 2018

Automatic Identification of Nanfeng Mandarin by Using a Stacked Autoencoder-Based Deep Learning Algorithm with Smartphone

Zhi-Yong Xiao
School of Software, Jiangxi Agricultural University
Jiangxi Agricultural University,
Nanchang City, 330045, Jiangxi P.R.China
+86-791-83828084
zhyxiao@jxau.edu.cn

Ya-Ru Liang
School of Engineering, Jiangxi Agricultural University
Jiangxi Agricultural University,
Nanchang City, 330045, Jiangxi P.R.China
+86-791-83828104
liangyaru@jxau.edu.cn

ABSTRACT
With the increasing deep learning applications in agriculture, identification of agriculture product of geographical indication by images had received remarkable attention from both academic and engineering fields. In order to facilitate the identification of Nanfeng mandarin, an identification method based on smartphone image and deep learning was proposed. In this paper, the research team proposed a classification scheme by using a stacked autoencoder based deep learning algorithm with three views of Nanfeng mandarin. As smartphone photography becomes more sophisticated, and the speed of a smartphone's internet connection fully supports real-time transmission of images, the image data can be easily recorded and quickly uploaded by the smart phone than other professional image equipment. A stacked autoencoder based deep learning algorithm was employed here for mandarin image classification so as to precisely recognize four type mandarins that were Nanfeng mandarin, Shaowu mandarin, Liucheng mandarin and Guangchang mandarin respectively. Experimental results indicated that the method based smartphone image and deep learning algorithm achieved a high accuracy of 89.45% for Nanfeng mandarin recognition than traditional identification methods. The method proposed in this paper may provide a convenient, fast and relatively accurate way for people to automatic identify Nanfeng mandarin.

CCS Concepts (ACM'S COMPUTING CLASSIFICATION SYSTEM (CCS))
• CCS → Computing methodologies→ Machine learning→ Machine learning algorithms → Feature selection

Keywords
Automatic identification; Nanfeng mandarin; Stacked autoencoder; Image; Smartphone

1. INTRODUCTION
Nanfeng mandarin is a kind of geographical indication product in Nanfeng country, Jiangxi province, and its research was a single method, or a comprehensive method combination of spectroscopy and chromatography in the last years. These methods had been applied in the identification of Nanfeng mandarin. Although these methods achieve higher accuracy of identification, these methods require not only professional devices but also time-consuming. Therefore, ordinary people were unable to use these methods to identify whether a mandarin was really a geographical indication product. The proposed method of identification based on smart phone and stacked autoencoder (SAE) was relatively practical. People used smart phones to take the three views of the orange, and uploaded the image data to the background network. The background server handled and analyzed the image, got the type of classification, and then sent the results to people's smartphones in time.

The detection technology based on near infrared reflectance spectroscopy (NIR) of the agricultural product quality at home and abroad has been mature. In China, researchers have applied NIR or atomic absorption spectroscopy to study the quality of agricultural products, mainly including fruits [1, 2], meat [3, 4] and dairy products [5, 6]. A study [7] conducted on the NIR technique for the identification of Yangmei varieties. A study [8] used NIR technology to test the soluble solids content (SSC) of Nanfeng mandarin. A study [9] used NMR spectroscopy and NIR to detect the origin of oolong tea. With the application and popularization of artificial neural network (ANN) theory, many scholars began to integrate ANN and spectral detection technology into the research of agricultural product identification. According to the principle of metal tracking analysis, a study [10] used statistics and neural network classifier to identify potato geographical sources. A study [11] used NIR technology to identify honey by geographical indications. A study [12] used principal component analysis method and sequence clustering analysis to determine the characteristic elements and mechanism of olive oil of geographical indication product. With the development and integration of computer technology, image processing technology and artificial intelligence theory, machine vision has developed rapidly. In the field of agriculture, machine vision [13] can accurately and rapidly identify different kinds of objects that human eyes cannot quickly or accurately identify. Many scholars have conducted a series of studies in this field and achieved certain results. For example, a study [14] proposed an algorithm for the automatic recognition of Fuji apples on the tree. A work [15] conducted grading studies on potatoes. A study [16] identified crops and weeds. In recent years, deep learning (DL) [17], a new development direction of machine learning, has received unprecedented attention. DL has shown the magic charm of intelligent prediction and classification in the fields of speech recognition, natural language processing and image processing [18]. With the development of DL, some scholars [8-10] conducted researches on source identification and identification of agricultural products based on machine learning. Above scholars have made certain progress in the research and the classification accuracy was as high as 88.3%. DL also shows the huge advantage in agricultural products classification, but these researches are still in the experimental stage and are still a long way from practical application.

Most of mandarin farmers in Nanfeng country can accurately identify the Nanfeng mandarin from mixed oranges based on size, color, texture and shape, and they can make a correct judgment with the naked eye. However, for the customers, most people cannot judge the real and fake Nanfeng mandarin from the appearance factor. According to the comparative analysis of the appearance of Nanfeng mandarin and other kinds of mandarin, Nanfeng mandarin has the appearance features of flat round concave on the top of fruit, small concave on the base of fruit, smooth and non-sandy fruit surface and orange peel. These appearance features are not found in other kinds of mandarin and have high identification. Based on the comprehensive analysis of the above research reports and literature, the research team proposed a research scheme. First, a smartphone is used to obtain the original three views of Nanfeng mandarin. These images are uploaded to the background through the network. Then, the image feature is formed into test data, input the deep neural network, and get the result of identification. In order to obtain the classification model with high accuracy, the research team had collected a large number of four kinds of mandarin from four different regions as training data. After training and adjusting the parameters of the deep learning model, the classification accuracy was achieved, and finally, the geographical indications product identification model of Nanfeng mandarin was determined.

In the study, a simple identification of Nanfeng mandarin was demonstrated through the DL algorithm and three views, and so experimental results showed that this method had high identification accuracy and feasibility.

2. Materials and methods

2.1 Experimental materials

The experimental samples of mandarin in Nanfeng county, Jiangxi province were from the orchards of three different towns (Baishe, Sangtian and Qiawan) in Nanfeng county, Jiangxi province. The team collected 1,100 superior mandarin from different fruit trees in each orchard and 3,300 from three orchards. The team shipped these samples to the laboratory as required, and finally a total of 2,720 prime fruits were selected as formal samples. A total of 3,300 samples were collected from an orchard in Liucheng county, Guangxi province. The team shipped these samples to the laboratory as required, and finally a total of 2,300 prime fruits were selected as formal samples. A total of 3,300 samples were collected from an orchard in Shaowu county, Fujian province. The team shipped these samples to the laboratory as required. Due to accidents, picking samples caused more damage during transportation and finally a total of 1,180 prime fruits were selected as formal samples. A total of 3,300 samples were collected from an orchard in Guangchang county, Jiangxi province. The team shipped these samples to the laboratory as required. Due to accidents, picking samples causes more damage during transportation and finally a total of 2,280 prime fruits were selected as formal samples. The number of all samples reached 8480.Therefore, the number of samples met the requirements of this study's DL framework. Demographic data for these samples were shown in Table 1.

Table 1. Demographic data of the four different mandarins

Type	Number	Mean size[SD](mm)	Max size(mm)	Mini size(mm)	Mean weight[SD](g)
Nanfeng mandarin	2720	38.2[5.0]	42	36	35.8[6.5]
Liucheng mandarin	2300	41.3[4.9]	44	37	40.1[7.3]
Shaowu mandarin	1180	42.2[5.1]	45	38	40.6[7.2]
Guangchang mandarin	2280	42.5[5.2]	45	36	40.3[7.3]

2.2 Proposed method

In general, a sparse autoencoder (AE) [19] was adopted for features subtraction and dimensionality reduction. A softmax classifier was adopted in the phase of classification. The sparse AE is a framework of unsupervised learning, and a stacked autoencoder (SAE) [20] is a neural network consisting of multiple layers of sparse AE. The softmax classifier is one type of supervised learning. SAE and softmax classifier can be well united in the proposed system, as shown in Figure 1. The outputs of each layer in SAE are wired to the inputs of the successive layer. The sparse AE is a key substructure of the SAE, and it has a symmetrical architecture, as shown in Figure 2.Given a training set with m samples $\{(x^{(1)},y^{(1)}),\ldots,(x^{(m)},y^{(m)})\}$, where $y^{(i)}$ is the category label of i th sample, $y^{(i)}\in\{1,2,\ldots,k\}$, for the first layer, the activation units are the input data. The mapping from input layer into hidden layer can be regarded as encoding, while the mapping from hidden layer into output layer can be regarded as decoding. SAE architecture is constructed by stacking the input and hidden layers of sparse AE, where a subsequent layer is trained through the output of its previous layer. In the proposed SAE, there are four layers, i.e., an input layer, two hidden layers and an output layer. The structure of the proposed SAE is shown in Figure 3.

The *i*-th activation unit $a_i^{(l)}$ of the *l*-th layer can be represented by the units of (*l*-1)-th layer through the weight parameter *W*, a bias term *b* and an activation function *f* :

$$z_i^{(l)} = \sum_{j=1}^{n} W_{ij}^{(l-1)} a_j^{(l-1)} + b_i^{(l-1)} \quad (1)$$

$$a_i^{(l)} = f\left(z_i^{(l)}\right) \quad (2)$$

where *f* is chosen as the sigmoid function in this study:

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

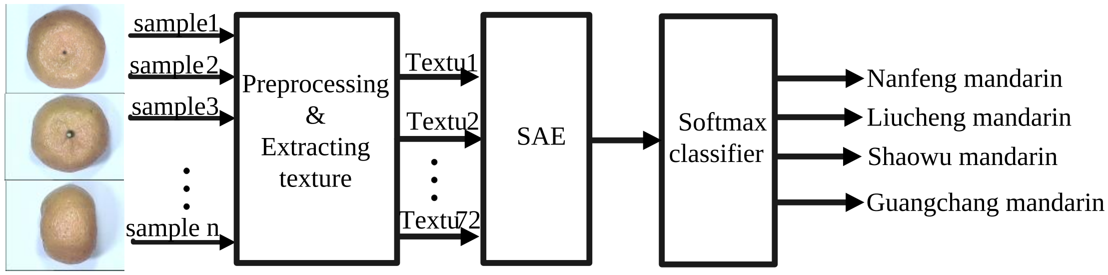

Figure 1. The proposed system based on SAE and softmax classifier

So the SAE is considered as one framework of deep learning. It is remarkable that deep learning has been a research focus in latest years. This novel machine learning method has been successfully applied in the areas of speech recognition [21], human face identification [22] and computer vision [23]. SAE enjoys all the benefits of deep learning frames of greater expressive power. After each layer of the SAE is set up, the model of network is considered as a whole used to fine-tune all pre-trained parameters with a softmax regression classifier. The softmax regression classifier is a generalized model which can be employed to resolve a multi-task classification. The goal of SAE is to minimize the distance between the l-th layer and the (l-1)-th layer to learn W and b. The cost function can be computed as:

$$J(W,b)=\left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|h_{W,b}\left(x^{(i)}\right)-y^{(i)}\right\|^{2}\right)\right]+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(W_{ji}^{(l)}\right)^{2} \quad (4)$$

where m is the number of training samples, n_l is the number of layers, and s_l is the number of units in the l-th layer. In equation (4), the first term is the mean square error between inputs and outputs, representing the quality of learning, while the second term is a regularization term (also called a weight decay term) that tends to decrease the magnitude of the weights and helps prevent the learning from over-fitting, λ is the weight decay parameter. To ensure that the features of hidden layer are desirably sparse, one sparsity constraint must be introduced in the cost function to control the learning process. This sparsity constraint parameter controls the average activation of the hidden units. The average activation of the j-th hidden unit is defined as:

$$\hat{\rho}_j=\frac{1}{m}\sum_{i=1}^{m}\left(a_j^{(2)}x^{(i)}\right) \quad (5)$$

where a_j denotes the activation of hidden unit j in the SAE. If $\hat{\rho}_j$ is small enough (close to 0.05), most units of hidden layer will be inactive. To force $\hat{\rho}_j$ be equal to a very small value ρ, the sparsity penalty term is designed based on the concept of Kullback - Leibler (KL) divergence:

$$D_{KL}\left(\rho\|\hat{\rho}_j\right)=\rho\log\frac{\rho}{\hat{\rho}_j}+(1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j} \quad (6)$$

Then, the sparsity penalty term is incorporated into the new cost function:

$$J_{sparse}(W,b)=J(W,b)+\beta\sum_{j=1}^{s_2}D_{KL}\left(\rho\|\hat{\rho}_j\right) \quad (7)$$

where β is a weight of the sparsity penalty term. During the process of learning, the new cost function $J_{sparse}(W,b)$ is ultimately minimized by updating W and b. It is necessary to calculate the contributions of all units in the hidden layer and output layer to the cost function before each iteration. As shown in the study, limited-memory BFGS (L-BFGS) [24] is a suitable optimization algorithm for updating W and b in the process of back propagation. In this study, L-BFGS is adopted. Concretely, the Softmax regression function $h_\theta\left(m^{(i)}\right)$ takes the form:

$$h_\theta\left(m^{(i)}\right)=\frac{1}{\sum_{j=1}^{k}e^{\theta_j^T m^{(i)}}}\begin{bmatrix}e^{\theta_1^T m^{(i)}}\\ e^{\theta_2^T m^{(i)}}\\ \vdots\\ e^{\theta_k^T m^{(i)}}\end{bmatrix} \quad (8)$$

Where $m^{(i)}$ is the key feature from the output layer of SAE. Through fine tuning softmax weight decay, it will take care of the numerical problems associated with over-parameterized representation. This model generalizes logistic regression to classification problems where the class label can take on more than two possible values.

2.3 Data preprocessing

Each mandarin image set included three views (i.e. the front view, the top view and the bottom view). For example, three views of a mandarin were shown in Figure 4-6. In the actual living environment, due to the influence of external factors such as human and smart phone brand, the three views obtained by smart phones were usually less stable and clear than those obtained in the laboratory environment. In order to improve the stability and reliability of the classification system, the research team carried out a series of preprocessing on the mandarin images, including converting the images of the three channels into single channel images. Secondly, image filtering can remove the interference and influence of image noise and improve image quality. The research team performed the same preprocessing on all mandarin image data, then extracted the 72-dimensional texture features, normalized them, and added labels. Finally these texture features were converted into specific training data. After 0-1 normalization, these features were extracted and input to the SAE.

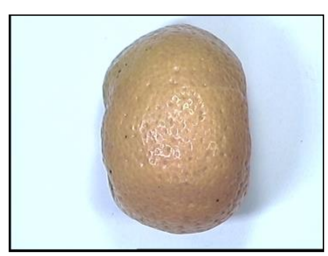

Figure 4. The front view of a mandarin

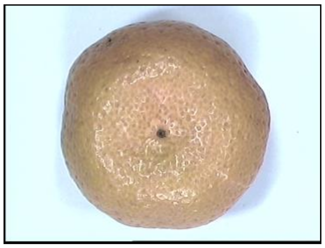

Figure 5. The top view of a mandarin

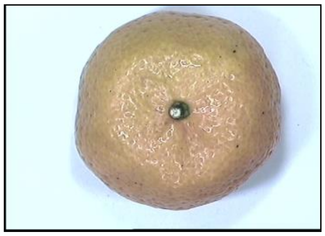

Figure 6. The bottom view of a mandarin

3. Parameters of SAE-based classification algorithm

After data preprocessing, these texture features selected from all mandarins were input into the SAE, and then the outputs from the SAE was input into the subsequent softmax classifier. In the procedure, the input layer's size of SAE was 72, and the first and second hidden layer's size was 200; sparsity parameter of the SAE was 0.05. The weight decay parameter term of the SAE was 1e-8. The weight of sparsity penalty term of the SAE was 3. The Softmax weight decay was 1e-6. Finally the softmax classifier was used to discriminate mandarin type.

4. Experimental results and analyses

In the study, the classification accuracy of Nanfeng mandarin was computed by using cross validation (CV). For example, in the 4240-fold CV, all the samples are randomly partitioned into 4240 groups, with each group including two subjects. Of the 4240 groups, 4239 groups (8478 samples) were chosen as the training set, and remaining one group was used as the test data. The validation was repeated 4240 times such that each group was used exactly once as test data. The final estimation was produced by averaging the 4240 results during the validation. The classifier predicted the labels (Nanfeng mandarin = 1, Shaowu mandarin = 2, Liucheng mandarin = 3 and Guangchang mandarin = 4) of the two subjects who were left out, and the accuracy of each classification was assessed. Finally, total accuracy was computed for each category. The prediction parameters included true positive (TP), false negative (FN), true negative (TN), and false positive (FP) classifications. The accuracy was computed using the following formula.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) * 100\% \qquad (9)$$

The parameter of accuracy was reported in the study, and the results of some CVs were shown in Table 2.

Table 2. The comparison results of different CV

CV	**Accuracy**
20 fold	82.97%
53 fold	85.51%
106 fold	87.31%
212 fold	87%
424 fold	89.45%
848 fold	88.81%
4240 fold	88.97%
8480 fold	88.81%
Average	87.35%

The results of various CVs were satisfactory, with a minimum accuracy of 82.97%, a maximum accuracy of 89.45% and an average accuracy of 87.35%.Therefore, the experimental results showed that the proposed classification method was feasible. To further analyze the details of the classification, the team conducted a detailed study of the 4240-fold CV. The classification results of the study were shown in Table 3.

Table 3 The results of mandarin type in 4240-flod CV

Type label	**1**	**2**	**3**	**4**
Total number	2720	1180	2300	2280
Number of prediction type 1	2470	18	267	0
Number of prediction type 2	58	954	5	169
Number of prediction type 3	190	3	2025	16
Number of prediction type 4	2	205	3	2095
Number of wrong prediction	250	226	275	185

Table 3 showed that Nanfeng mandarin was predicted to be Liucheng mandarin 190 times, and Liucheng mandarin was predicted to be Nanfeng mandarin 267 times. This phenomenon may have something to do with the introduction of a large number of mandarin seedlings from Nanfeng country to Liucheng country since 1981. Secondly, Shaowu mandarin was predicted to be Guangchang mandarin 205 times, and Guangchang mandarin was predicted to be Shaowu mandarin 169 times. Finally, there were few misclassifications among other species. For example, Nanfeng mandarin was predicted to be Shaowu mandarin 58 times, and to be Guangchang mandarin 2 times. Other CV results showed the same classification trend as the 4240-fold CV. In general, different CV had produced consistent results for the classification of different mandarin.

In the study of 2720 Nanfeng mandarins, 2300 Liucheng mandarins, 1180 Shaowu mandarins and 2280 Guangchang mandarin, texture feature array extracted from three views was classified as Nanfeng mandarin or other mandarins with an accuracy of approximately 90% using SAE algorithm. In a previous study [25], three types of mandarins (Nanfeng mandarin, Liucheng mandarin and Shaowu mandarin) had been studied by NIR and principal component analysis, and the literature demonstrated that they could be distinguished completely. The literature did not give a quantitative description of the experimental samples and also used professional equipment. Therefore, this method had certain limitations and could not be popularized. In another study [26] using chemical method，the researchers measured the volatile matter in the skins of three different types of mandarin, then analyzed their 17 main components, and finally divided them into three types. The chemical method had accurate result, but was still not widely available. Another study [27] combined spectroscopy and chromatography with chemometrics. The comprehensive method had been successfully applied in the identification of geographical indication product-Nanfeng mandarin. This comprehensive method was relatively accurate, but it was still time-consuming, labor-intensive and difficult to implement. All of the above studies had achieved certain achievements, but they had adopted professional methods, so they were not efficient and not widely spread.

Using a novel machine learning method and smart phone, the accuracy in the study was approximately below 10% compared with the best result of the reference [27] cited in this article. But the method may be easily applied in the identification of geographical indication product. It was suggested that the framework of SAE and softmax classifier should be a better choice to discriminate Nanfeng mandarin from other three kinds of mandarin through image.

5. Conclusions

In conclusion, the method blending three views and using deep learning successfully classified Nanfeng mandarin and other three different mandarins with an average accuracy of 87.35%. This method is the most convenient and economic method among the state-of-the-art results. So this proposed method in the study may provide a good and easy way to indicate Nanfeng mandarin.

6. ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (Nos. 61662047) and by Science and Technology Research Foundation of Jiangxi education department (GJJ170272).

7. REFERENCES

[1] Fu, X.P., et al., Detection of pear firmness using near infrared diffuse reflectance spectroscopy. Guang Pu Xue Yu Guang Pu Fen XI, 2006. 26(6), 1038-1041.

[2] Ma, G., et al.,Nondestructive sugar content determination of peaches by using near infrared spectroscopy technique. Guang Pu Xue Yu Guang Pu Fen XI, 2007. 27(5), 907-910.

[3] Zhao, J.W., et al., The determination of beef tenderness using near-infrared spectroscopy. Spectroscopy & Spectral Analysis, 2006. 26(4), 640.

[4] Zhao, L.L., et al., Quality analysis of Chinese bacon with near infrared spectroscopy. Spectroscopy & Spectral Analysis, 2007. 27(1), 46-49.

[5] Chang, M., P.J. Chu, and K.X. Xu, Study on noninvasive detection using NIR diffuse reflectance spectrum for monitoring protein content in milk powder. Spectroscopy & Spectral Analysis, 2007. 27(1), 43.

[6] Han, D.H., et al., Identification of reconstructed milk in raw milk using near infrared spectroscopy. Journal of Agricultural Mechanization Research, 2008. 27(3), 465-468.

[7] Yong, H.E. and L.I. Xiao-Li, Discriminating varieties of waxberry using near infrared spectra. Journal of Infrared & Millimeter Waves, 2006. 25(3), 192-194+212.

[8] Sun, T., et al., Determination of soluble solids content in Nanfeng Mandarin by Vis/NIR spectroscopy and UVE-ICA-LS-SVM. Spectroscopy & Spectral Analysis, 2013. 33(12), 3235.

[9] Meng, W., et al., Geographical Origin Discrimination of Oolong Tea (TieGuanYin, Camellia sinensis (L.) O. Kuntze) Using Proton Nuclear Magnetic Resonance Spectroscopy and Near-Infrared Spectroscopy. Food Analytical Methods, 2017. 10(11), 3508-3522.

[10] Anderson, K.A., et al., Determining the geographic origin of potatoes with trace metal analysis using statistical and neural network classifiers. Journal of Agricultural & Food Chemistry, 1999. 47(4), 1568-75.

[11] Herrero, L.C., et al., A fast chemometric procedure based on NIR data for authentication of honey with protected geographical indication. Food Chemistry, 2013. 141(4), 3559-65.

[12] Gumus, Z.P., et al., Determination of trace elements and stable carbon isotope ratios in virgin olive oils from Western Turkey to authenticate geographical origin with a chemometric approach. European Food Research and Technology, 2017. 243(10), 1719-1727.

[13] Ying, Y. and Y. Jiang, Advance on Application of Machine Vision Technique to Automatic Quality Identification of Agricultural Products. Transactions of the Chinese Society of Agricultural Engineering, 2000. 16(3), 4-8.

[14] Bulanon, D.M., et al., A segmentation algorithm for the automatic recognition of Fuji apples at harvest. Biosystems Engineering, 2002. 83(4), 405-412.

[15] Tao, Y., et al., Fourier-Based Separation Technique for Shape Grading of Potatoes Using Machine Vision. Transactions of the Asae, 1995. 38(3), 949-957.

[16] Tillett, N.D., et al., Mechanical within-row weed control for transplanted crops using computer vision. Biosystems Engineering, 2008. 99(2), 171-178.

[17] Schmidhuber, J., Deep learning in neural networks: An overview. Neural Networks, 2015. 61, 85-117.

[18] Najafabadi, M.M., et al., Deep learning applications and challenges in big data analytics. Journal of Big Data, 2015. 2(1), 1.

[19] Xu, J., et al., Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. IEEE Trans Med Imaging, 2016. 35(1), 119-130.

[20] Xiao Z, Wang C, Jia N, et al. SAE-based classification of school-aged children with autism spectrum disorders using functional magnetic resonance imaging[J]. Multimedia Tools & Applications, 2018(1):1-12.

[21] G. E. Hinton, L.D., D. Yu, et al., Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,. IEEE Signal Process. Mag., 2012. 29(6), 82-97.

[22] Jiang, H., A Study of Deep Learning and Human Face Recognition Methods. Technology Innovation & Application, 2018.

[23] M. Zeiler, G.T.a.R.F., Adaptive deconvolutional networks for mid and high level feature learning. Proc. IEEE Int. Conf. Comput. Vision, 2011, 2018-2025.

[24] Liu, D.C. and J. Nocedal, On the limited memory BFGS method for large scale optimization. Mathematical Programming, 1989. 45(1-3), 503-528.

[25] Wei yuanlong, et al., Near infrared spectroscopy combined with principal component analysis to identify nanfeng honeberry from different habitats, 2013. 33(11), 3024-3027.

[26] Huang jie, et al. Fisher bee combined hs-gc/MS to identify the geographical indication of Nanfeng mandarin. National chromatographic symposium and instrument exhibition. 2015.

[27] Yin changhai. Chromatography and near-infrared spectroscopy analysis and evaluation of nanfeng mandarin. 2013,