

Deep belief network-based drug identification using near infrared spectroscopy

Huihua Yang^{*,†,§}, Baichao Hu^{*,¶}, Xipeng Pan[†], Shengke Yan^{*}, Yanchun Feng[‡],
Xuebo Zhang[‡], Lihui Yin[‡] and Changqin Hu[‡]

**College of Electronic Engineering and Automation
Guilin University of Electronic Technology
1 Jinji Road, Guilin 541004, P. R. China*

*†Automation School
Beijing University of Posts & Telecommunications
10 Xitucheng Road, Beijing 100876, P. R. China*

*‡National Institutes for Food and Drug Control
10 Tiantanxili Road, Beijing 100050, P. R. China*

§yhh@bupt.edu.cn, 406611592@qq.com

¶hubaichao8185@163.com

Received 13 March 2016

Accepted 4 May 2016

Published 28 June 2016

Near infrared spectroscopy (NIRS) analysis technology, combined with chemometrics, can be effectively used in quick and nondestructive analysis of quality and category. In this paper, an effective drug identification method by using deep belief network (DBN) with dropout mechanism (dropout-DBN) to model NIRS is introduced, in which dropout is employed to overcome the overfitting problem coming from the small sample. This paper tests proposed method under datasets of different sizes with the example of near infrared diffuse reflectance spectroscopy of erythromycin ethylsuccinate drugs and other drugs, aluminum and nonaluminum packaged. Meanwhile, it gives experiments to compare the proposed method's performance with back propagation (BP) neural network, support vector machines (SVMs) and sparse denoising auto-encoder (SDAE). The results show that for both binary classification and multi-classification, dropout mechanism can improve the classification accuracy, and dropout-DBN can achieve best classification accuracy in almost all cases. SDAE is similar to dropout-DBN in the aspects of classification accuracy and algorithm stability, which are higher than that of BP neural network and SVM methods. In terms of training time, dropout-DBN model is superior to SDAE model, but inferior to BP neural network and SVM methods. Therefore, dropout-DBN can be used as a

[§]Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

modeling tool with effective binary and multi-class classification performance on a spectrum sample set of small size.

Keywords: Deep belief networks; near infrared spectroscopy; drug classification; dropout.

1. Introduction

Counterfeit and substandard drugs, cause great harms to people and require identification under pharmaceutical supervision. Moreover, due to the differences in production process, raw materials, packaging and other forms, even the same drug produced by different manufacturers or different specifications of the same drug may have different features. So, it is of great value to identify these differences in pharmaceutical supervision. Near infrared spectroscopy (NIRS) analysis technology can be used for quick and nondestructive detection of the drugs and has been widely used in the pharmaceutical industry,¹⁻³ for it has abundant frequency doubling and frequency vibration information in molecular groups.

NIRS combined with pattern recognition technology has been widely used in the rapid and non-destructive quantitative and qualitative analysis of drugs. In terms of constructing classification model, Deconinck *et al.*⁴ used decision tree to identify the spectra of Viagra and Cialis drugs, which obtained good classification accuracy; Storme *et al.*⁵ adopted support vector machines (SVM) to do the drug identification of NIRS, the result of which was proved more accurate than that of decision tree. Yu Ke *et al.*⁶ used least squares SVM (LS-SVM) to distinguish the genuine and fake traditional Chinese medicine (Danshen Salvia Miltiorrhiza powder samples), which achieved a higher accuracy than the SVM. However, there are few researches on the spectroscopy multi-classification, for the multi-classification is more complex than binary classification, and the shallow structure of SVM and the two-layer back propagation (BP) neural network make the ability of its function mapping weak, the ability of which still remains to be investigated in the identification of multi-classification drugs.

Deep learning, a machine learning method,⁷ has been successfully used in the fields of natural language processing, text classification and image recognition.⁸⁻¹⁰ Meanwhile, it has deep network structure and nonlinear activation function, which makes all kinds of deep learning models be

appropriate for big data model, especially for the ones with higher dimensions and which are non-linear. However, the number of samples is relatively small in spectral analysis, and direct application of deep learning model may result in overfitting problem. There were few papers about deep learning in the spectral classification. Only in Ref. 11, we have learned that Luo used the technology of NIRS combined with sparse denoising auto-encoder (SDAE) algorithm to conduct the identification of genuine and fake drugs.

This paper has proposed the near infrared spectral drug identification method based on dropout deep belief network (dropout-DBN). The DBN model can ensure the strong nonlinear modeling capabilities, and the dropout mechanism can overcome the problems of the small sample. To verify the validity of the algorithm, we have made experiments of binary and multi-class classification by using the diffuse reflection spectrum data coming from some drugs such as erythromycin ethylsuccinate; verified the function of dropout-DBN in terms of its accuracy, stability and training time; and given detailed comparison among the BP neural networks, SVM and SDAE. All of these show that the proposed method was effective.

2. Methods

2.1. DBN model

The DBN is a semi-supervised learning method, which combines the advantages of unsupervised learning and supervised learning method, and has strong ability of classification and prediction for high dimensional feature vector. Mimicing the multi-layer structure of the human brain, DBN has made a feature extraction of input data from the low layer to the top layer. As shown in Fig. 1, the DBN is comprised of multi-layer unsupervised restricted Boltzmann machine (RBM) and a layer of a supervised BP network.¹²

The training process of the DBN model mainly includes two stages: pre-training stage and fine-tuning stage.¹³ In the pre-training stage, we first

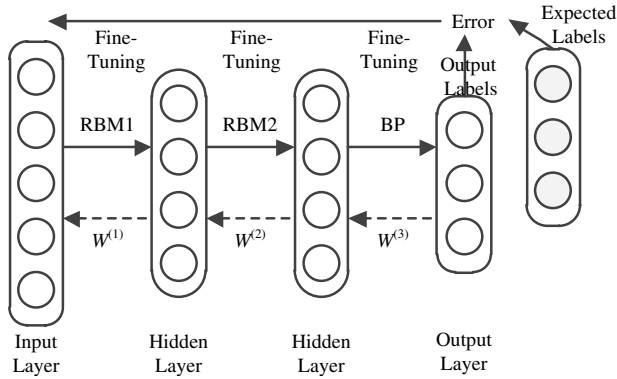


Fig. 1. The structure of DBN model.

initialized the parameters of DBNs, and then used the layer by layer network training method. We used the unsupervised learning method to train one layer of RBM each time, and the output result of which was used as the input layer to the next RBM. The layers were trained one by one like that, until all the layers of RBM were completed, and the weight and bias value of each layer were retained accordingly. In order to make the network performance better, the gradient descent algorithm was used to fine-tune the parameters of the whole network. This stage is similar to the BP neural network algorithm. The weight of each layer has been pre-trained by the RBM before fine-tuning, which is not in random initialization like the BP neural networks do, thus avoiding the situation of local convergence.

2.2. RBM

The RBM is an energy-based model,¹⁴ which consists of a visible layer and a hidden layer. There is no neuronal connections between the same inner layer, and inter layer neurons are fully connected. The main function is to determine the probability distribution of the output layer, as shown in Fig. 2.

In a binary RBM, both the hidden and visible units are binary and stochastic. \mathbf{v} represents the visible layer vector and \mathbf{h} stands for the hidden layer vector. Then, for a group of (\mathbf{v}, \mathbf{h}) , the energy function of RBM can be defined as follow:

$$E(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta}) = -\sum_{i=1}^m a_i b_i - \sum_{j=1}^n b_j h_j - \sum_{i=1}^m \sum_{j=1}^n v_i w_{ij} h_j, \quad (1)$$

where $\boldsymbol{\theta} = (w_{ij}, a_i, b_j)$ is the parameter of RBM, w_{ij} is the connecting weight between visible unit i and

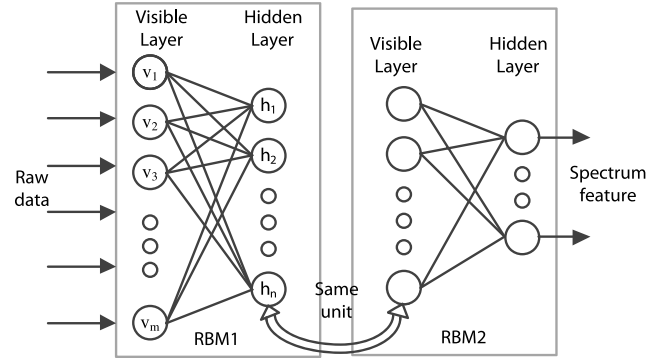


Fig. 2. The structure of RBM model.

hidden unit j , while a_i and b_j are bias terms of visible layer and hidden layer, respectively. m and n are the numbers of visible and hidden units. The goal of the training algorithm is to learn corresponding parameter $\boldsymbol{\theta}$. Based on the energy function, the joint distribution over the visible and hidden units can be defined as follows:

$$P(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta}) = \frac{e^{-E(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta})}}{Z(\boldsymbol{\theta})}, \quad Z(\boldsymbol{\theta}) = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta})}, \quad (2)$$

where $Z(\boldsymbol{\theta})$ is a normalization factor, and then joint distribution of the marginal distribution can be defined as follow:

$$P(\mathbf{v}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta})}. \quad (3)$$

This marginal distribution is the probability of visible unit in RBM. Since there are no hidden–hidden connections, the conditional distribution $P(\mathbf{h}|\mathbf{v})$ is factorial and given by

$$P(h_j = 1|\mathbf{v}) = \sigma\left(\sum_i v_i w_{ij} + b_j\right), \quad (4)$$

where $\sigma(x) = 1/[1 + \exp(-x)]$. Similarly, since there are no visible–visible connections, the conditional distribution $P(\mathbf{v}|\mathbf{h})$ is factorial and is given by

$$P(v_i = 1|\mathbf{h}) = \sigma\left(\sum_j w_{ij} h_j + a_i\right), \quad (5)$$

where $\boldsymbol{\theta}$ can be obtained by calculating the maximum likelihood estimation of the training set (the number of samples is T):

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta}} \sum_{t=1}^T \ln p(\mathbf{v}^{(t)}|\boldsymbol{\theta}). \quad (6)$$

But maximum-likelihood learning is infeasible in a large RBM because it is exponentially expensive to compute the derivative of the log probability of the training data. Nevertheless, RBMs have an efficient approximate training procedure called “contrastive divergence”¹⁵ which makes them suitable as building blocks for learning DBN. We repeatedly update each parameter θ using the difference between two measured, pairwise correlations:

$$\begin{cases} w_{ij} = w_{ij} + \varepsilon(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}}), \\ a_i = a_i + \varepsilon(\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{recon}}), \\ b_j = b_j + \varepsilon(\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{recon}}), \end{cases} \quad (7)$$

where ε is a learning rate, $\langle \cdot \rangle_{\text{data}}$ is the expectation on the distribution defined by the training data, $\langle \cdot \rangle_{\text{recon}}$ is the expectation on the distribution defined by the reconstructed model.¹⁶

2.3. Fine-tuning stage

BP algorithm, a kind of supervised training method, is used to adjust and optimize weight parameters extracted in the pre-training stage. The classification results were obtained by logistic or softmax regression algorithm. The training process is divided into two steps:

- (1) Forward propagation: Set the parameter value to each layer, then input the feature vectors into the BP neural network, according to the preset network structure training and output the training value.
- (2) BP: Compared to the actual results of BP algorithm output with correct results of expected output, the resultant error value resulting from which was put back into the input layer from the output layer, so as to adjust and optimize the parameters of the network.

Through the multiple iterative forward and BP, the weights between neurons were modified. When the error between the actual value and the output value meets the requirement, the training will stop.

2.4. Dropout-DBN model

Deep learning network model is mainly used for massive data network training. If the number

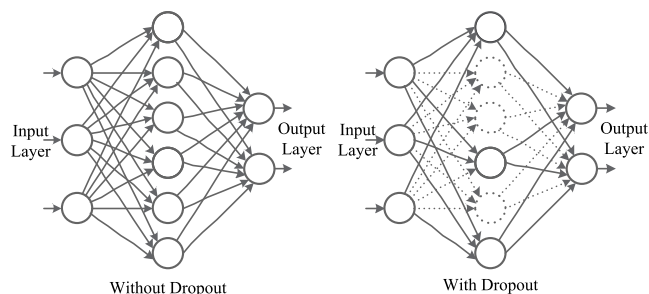


Fig. 3. The role of dropout.

of training sets or the feature dimension of the data is small, the system may produce local convergence and lead to overfitting. The core idea of dropout-DBN is to randomly ignore hidden layer neurons in the training process, so as to prevent overfitting.

Dropout was proposed by Hinton and his team.¹⁷ The basic idea of dropout is that it randomly ignores neurons of the hidden layer. That means, some hidden layer neurons did not participate in this training network and the weight will be reserved. These neurons may be involved in training in the next iteration. In the case of small training samples, due to the excessive number of iterations of the network, the system was likely to have mutual dependence between nodes. During each iteration process, dropout made some of the random section of nodes in the hidden layer out of work, so that it could effectively prevent the interdependence among features, as shown in Fig. 3.

3. Experiments

3.1. Dataset

Experimental data were collected by National Institutes for Food and Drug Control, including erythromycinethylsuccinate from Xi’an Li Jun pharmaceutical Ltd. and other factories (Wutai Group, Shijiazhuang Pharmaceutical Ltd. *et al.*), with aluminum and nonaluminum packaging, and a few other similar drugs (erythromycin, clarithromycin, melemycinum *et al.*). The experiments could get the spectrum curves by measuring absorbance values of each sample in different wavelengths. The wavelength of each spectral data ranged from 4196 nm to 9002 nm, with an interval of 4 nm. The detailed information of the NIRS sample is shown in Table 1.

Table 1. The profile of the pharmaceutical samples.

Category	Species	Number of aluminum	Number of nonaluminum	Total
Positive sample	Non-Xi'an Li Jun erythromycin ethylsuccinate	25	24	49
	Non-Xi'an Li Jun Non-erythromycin ethylsuccinate	9	20	29
	Total	34	44	78
Negative sample	Xi'an Li Jun erythromycin ethylsuccinate	65	106	171

3.2. Pre-processing

Firstly, all drugs spectroscopy samples were derivated (13 smooth pixels) to eliminate baseline and other background interference, as shown in Fig. 4. Secondly, all the input data were converted into the number between 0 and 1 by vector normalization, which were used to eliminate the orders of magnitude difference among the data of different dimensions, and avoid great errors of network prediction.

3.3. Dropout parameters selection

Dropout could inhibit the overfitting problem of small samples, and improve the prediction accuracy of algorithm. Figure 5 shows the relationship between dropout factor and prediction accuracy under the binary classification where the size of training set is 140. The experimental results indicate that dropout factor has comparatively stable classification accuracy in a wide range.

3.4. Classification model

Deep learn toolbox was selected for experiment and the software platform is MATLAB R2014a.

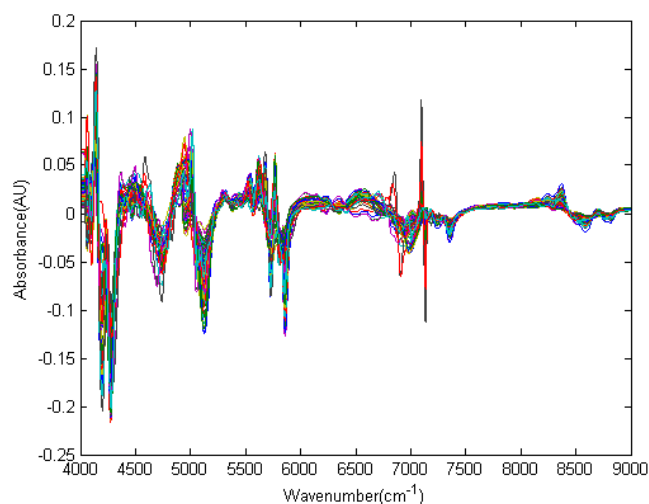


Fig. 4. The NIR spectrum of pharmaceutical samples.

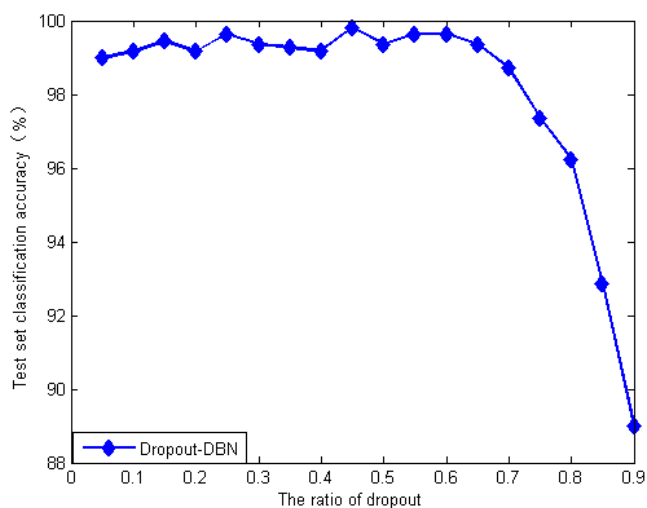


Fig. 5. The classification accuracies of different dropout ratios.

A four-layered network structure was adopted, namely, 1247-200-100-2/4. Binary classification and multi-classification are the same in pre-processing and pre-training stage, and the difference between the two is only in the fine-tuning stage. The output layer of binary classification uses logistic classifier, while the multi-classification uses Softmax classifier. The detailed process is as follows:

- (1) Pre-processing: The spectroscopy data of erythromycin ethylsuccinate was pre-processed and normalized by max-min method.
- (2) Pre-training: Two RBM network structures were used in hidden layer, and the network structure was set as 100-200. In the stage of RBM training, iteration of each RBM was 200; the learning rates of two layers were set as 0.05, and sigmoid function was used as activation function.
- (3) Fine-tuning: After the completion of the RBM training, the values of weights and bias obtained from the pre-training were used to initialize the parameters of the first two layers in BP neural network. Then optimize the parameters by BP

algorithm. The number of iteration of fine-tuning was 200, the learning rate was set as 0.05, and sigmoid function was used as activation function.

- (4) Testing: Firstly, carry out the feed-forward network calculation for test sets to compute the probability of each samples belonging to each classification, and then determine the classification of the samples according to the maximum principle, and extract category labels. Then compare predict labels with expect labels to get the accuracy.
- (5) BP neural network, SVM and SDAE were chosen as contrast methods. BP neural network was available in the toolbox of MATLAB software, and the network structure was the same as the DBNs. LibSVM toolbox of National Taiwan University was used as the experiment of SVM models, and linear kernel function and Gaussian kernel function were used as a comparison. From the cross validation of the grid searching function in the toolbox, SVM Gaussian kernel parameters, $c = 1$, $\gamma = 0.32$ were obtained.

4. Results and Discussion

To verify the identification effects of dropout-DBN model to NIRS drugs, this paper will use dropout-DBN model to, respectively, verify binary classification performance (the ability to identify genuine and fake drugs) and multi-classification performance (the ability to identify multiple drugs) of drugs, and then compare with other methods and evaluate it.

4.1. Binary classification experiment

Firstly, we took genuine and fake drugs identification as samples to test the prediction ability of dropout-DBN model. As shown in Table 1, the experiment collected 249 kinds of drugs spectrum samples. 171 erythromycin ethylsuccinate spectrum samples of different batches of Xi'an Li Jun pharmaceutical Ltd. were considered as the negative samples. Other erythromycin ethylsuccinate spectrum samples and species of spectrum samples in total 78 samples served as the positive samples. In order to verify the prediction performance of the algorithm in different datasets, we did experiments

Table 2. The size configuration of training sample set for binary-class identification.

Number of training sample in total	Number of positive sample	Number of negative sample
60	29	31
80	38	42
100	48	52
120	58	62
140	66	74
160	70	90
180	74	106

with the datasets in Table 2. In the configuration, each dataset in Table 2 by random sampling built independent datasets, 10 times on the basis of evaluating the average performance. Meanwhile, the paper adopted SDAE, BP neural network and SVM to the comparative experiments with dropout-DBN. We evaluated dropout-DBN performance in drug identification applications with classification accuracy, training time and classification stability.

In terms of classification accuracy, according to the datasets configuration of Table 2, the average prediction accuracy on datasets which were chosen at random 10 times independently of two-layer BP neural network, linear kernel and RBF kernel SVM, SDAE, DBN and Dropout-DBN algorithms are as shown in Table 3. The results showed that the performance with the dropout of DBN was better than that of without dropout processing, especially in the small sample sets, dropout-DBN had the best performance. This indicated that dropout could effectively improve the classification accuracy of the model. As Table 3 shows, SDAE, DBN and SVM (RBF kernel) had similar high accuracy, indicating that their complex nonlinear modeling capability could be applied to the binary classification problem. However, the BP neural network was easy to fall into local convergence and learning depth was not enough. At the same time, linear kernel SVM did not have a nonlinear modeling capability, so they had low prediction accuracy.

In terms of algorithm stability, according to the datasets configuration in Table 2, the average prediction accuracy on datasets was chosen at random 10 times independently of each algorithm whose standard deviation (STD) is as shown in Fig. 6. The stability of DBN algorithm was similar with SDAE network and much better than BP neural network. When the training datasets were small, the depth of

Table 3. The binary-classification accuracy on different ratios of training samples (unit: %).

Training/Test set	BP (2 layers)	SVM (linear)	SVM (RBF)	SDAE	DBN	DBN (dropout)
60/189	92.86	95.19	98.31	97.62	97.30	98.41
80/169	93.14	96.15	99.05	98.82	98.46	99.13
100/149	94.23	95.91	98.52	98.76	98.59	99.21
120/129	95.35	94.88	99.22	99.36	99.15	99.57
140/109	96.33	94.95	99.17	99.35	98.90	99.52
160/89	96.85	95.96	99.59	99.78	99.32	99.72
180/69	96.38	95.65	99.40	99.71	99.57	99.69

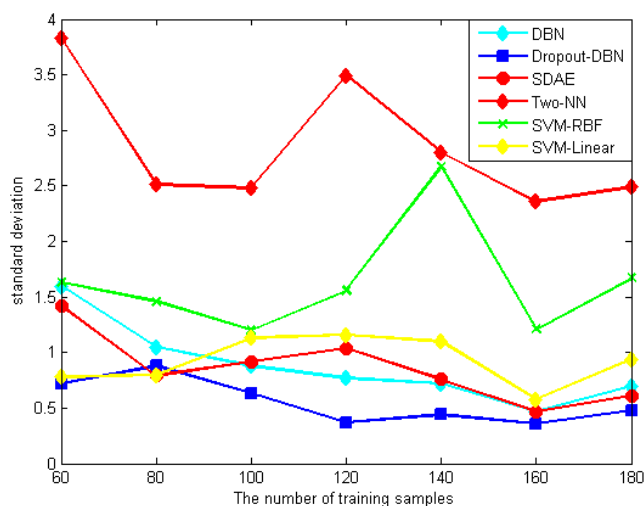


Fig. 6. The STD of accuracy of different binary-classification models.

belief network still had good stability. With the training dataset increasing, the stability of the DBNs generally got higher than SVM algorithm.

In terms of training time, according to the datasets configuration of Table 2, the average training time on datasets which was chosen at random 10 times of each algorithm is as shown in Table 4. SDAE network needs to construct a

Table 4. Training time of different binary classifier on different ratios of training samples (unit: s).

Training/Test set	BP (2 layers)	SVM (linear)	SVM (RBF)	SDAE (time)	DBN (dropout)
60/189	3.25	0.01	0.01	123.16	65.15
80/169	3.99	0.01	0.01	141.30	83.65
100/149	4.31	0.01	0.02	159.33	102.57
120/129	4.76	0.01	0.03	177.66	121.94
149/100	6.59	0.01	0.03	195.59	141.22
169/89	7.28	0.01	0.04	214.71	160.64
180/69	7.88	0.01	0.05	232.98	179.66

three-layer network for data dimensionality reduction and feature extraction, but dropout-DBN network only needs visible-hidden layers with two layers on the network. So, in terms of training time, dropout-DBN has great advantages. However, because BP and SVM algorithm did not need to extract the characteristic variables by pre-training, BP and SVM algorithm cost less time.

4.2. Multi-classification experiment

Secondly, we took multi-classification drug identification as examples to test the prediction ability of dropout-DBN model. The spectrum samples were divided into the following four groups in Table 1. The first kind were 34 spectrum samples with the aluminum packaging and not produced by Xi'an Li Jun Pharmaceutical Ltd. The second kind were 44 spectrum samples without aluminum packaging and not produced by Xi'an Li Jun Pharmaceutical Ltd. The third kind were 65 spectrum samples with the aluminum packaging and produced by Xi'an Li Jun Pharmaceutical Ltd. The fourth kind were 106 spectrum samples without aluminum packaging and produced by Xi'an Li Jun Pharmaceutical Ltd.

Similar to the experiment of binary classification, the samples were randomly selected according to the sizes of datasets and the amount of data in different categories (see Table 5), and the training sets and test sets were built accordingly. By the random sampling, each dataset in Table 5 was configured and datasets were built for 10 times independently to evaluate their average performance.

In terms of accuracy rate, we configure the dataset in Table 5, the average accuracy on datasets of each algorithm in 10 times independently random drawing is shown in Table 6. The results show that the datasets algorithm by using dropout-DBN was almost the highest in multi-class classification.

Table 5. The size configuration of training samples for multi-class identification.

Training sample	Class 1 (34)	Class 2 (44)	Class 3 (65)	Class 4 (106)
60	15	15	15	15
80	19	20	20	21
100	24	25	25	26
120	27	30	31	32
140	28	35	37	40
160	30	37	44	49
180	31	38	49	62

Table 6. The multi-classification accuracy of different classifiers on different ratios of training samples (unit: %).

Training/ Test set	BP (2 layers)	SVM (linear)	SVM (RBF)	SDAE	DBN	DBN (dropout)
60/189	89.84	90.16	96.83	94.86	92.86	96.67
80/169	88.46	91.24	96.69	95.68	94.85	97.46
100/149	90.07	93.36	97.19	95.97	95.84	98.52
120/129	93.02	95.12	97.68	97.67	96.59	98.91
140/109	94.13	95.78	97.50	98.07	97.71	98.82
160/89	96.07	98.20	98.10	97.64	97.75	99.21
180/69	96.52	97.54	97.99	98.41	97.83	99.13

The performance of dropout-DBN was significantly higher than that of DBN without dropout, and the smaller the sample set was, the more obvious advantages it would get. Meanwhile, by further comparing the average accuracy rate of dropout-DBN and DBN on multiple classification problems (Table 6) and binary classification problem (Table 3), the advantages of dropout-DBN in multiple classifications were more significant than that of the DBN. The main reason was that there were fewer samples of each category in the multi-class classification. In contrast, the smaller samples need more dropout mechanisms in order to prevent over-fitting problem. Compared to the drugs identification of binary classification, the accuracy of each algorithm in the drugs identification of multiple classification decreased, but each algorithm still showed previous rule, and SVM (RBF), DBN and SDAE had similar accuracy, which were all higher than BP neural network and SVM (linear kernel). However, when the training samples were extremely rare, SVM had certain advantages.

In terms of algorithm stability, according to the datasets configuration in Table 5, the STD of average prediction accuracy rate in 10 times random

Table 7. Training time of different multi-class classifiers on different ratios of training samples (unit: s).

Training/ Test set	BP (2 layers)	SVM (linear)	SVM (RBF)	SDAE	DBN (dropout)
60/189	6.90	0.01	0.01	165.57	92.00
80/169	9.16	0.01	0.02	190.69	117.63
100/149	11.24	0.01	0.03	214.32	143.66
120/129	11.60	0.01	0.04	238.86	169.25
149/100	14.00	0.01	0.05	264.38	196.26
169/89	14.47	0.01	0.05	289.45	221.33
180/69	15.36	0.02	0.06	314.86	250.50

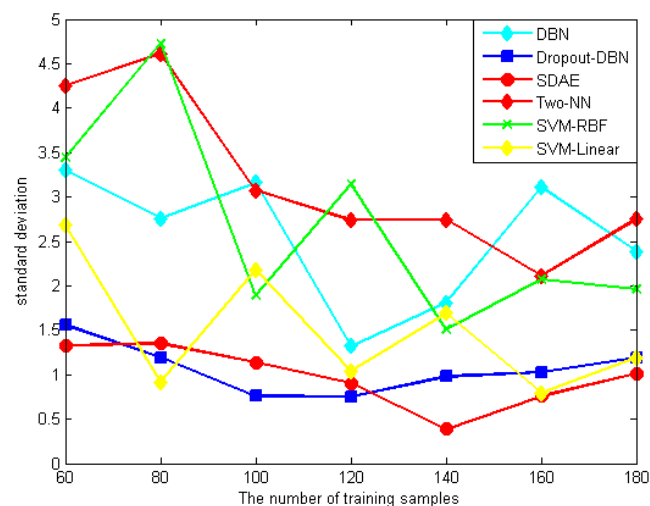


Fig. 7. The STD of accuracy of different multi-class classification models.

drawing datasets for each algorithm is shown in Fig. 7. The DBN network and SDAE network could still maintain high stability, which was generally higher than BP neural network and SVM algorithm.

In terms of training time, according to the datasets configuration in Table 5, the average training time of each algorithm in 10 times random drawing datasets is shown in Table 7. Compared to the binary classification of drugs identification, the training time of each algorithm was extended in the drugs identification of multiple classifications, but the training time of the traditional BP neural network and SVM was still faster.

5. Conclusion

In the drug regulatory applications, there are not only binary classification problems such as genuine

and false drug identification, but also multi-class problems such as identification of pharmaceutical varieties. Due to the difficulties of sample collection, wet chemical analysis and validation, etc., the number of samples is often seldom available. Deep learning network has strong capacity for complex nonlinear modeling, which is especially suitable for complex, high-dimensional and nonlinear analysis of massive data. When the deep learning model is applied to the stoichiometric analysis of NIRS, it will lead to overfitting because of fewness of the samples. The spectral classification method based on dropout-DBNs proposed by the paper can effectively prevent overfitting, through randomly removing network nodes during training. According to the study of binary and multi-class classification on near infrared diffuse reflectance spectroscopy of drugs, the results show that, in terms of small dataset, dropout-DBN had better accuracy and stability than DBN. Simultaneously, the performance of nonlinear classifier is excellent than linear classifiers. Compared to BP neural network, SVM and SDAE methods, dropout-DBN has better prediction ability and stability. However, in terms of training time, DBN algorithm is faster than SDAE, but much slower than BP and SVM. Therefore, dropout-DBN is an effective tool for spectroscopy analysis, and is particularly suitable for small sample model.

Acknowledgments

We are grateful for the financial support of the National Natural Science Foundation of China (Grant Nos. 21365008 and 61562013), Natural Science Foundation of Guangxi (Grant No. 2013GXNSFBFA019279) and Innovation Project of GUET Graduate Education (Grant Nos. GDYCSZ201474 and GDYCSZ201478). We also thank Xi'an Li Jun Pharmaceutical Factory for useful datasets of NIRS samples.

References

1. X. L. Chu, W. Z. Lu, "Research and application progress of near infrared spectroscopy analytical technology in China in the past five years," *Spectrosc. Spect. Anal.* **34**(10), 2595–2605 (2014).
2. K. Järvinen, W. Hoehe, M. Järvinen, S. Poutiainen, "In-line monitoring of the drug content of powder mixtures and tablets by near-infrared spectroscopy during the continuous direct compression tableting process," *Eur. J. Pharm. Sci.* **48**(4–5), 680–688 (2013).
3. T. Miyano, M. Kano, H. Tanabe, "Spectral fluctuation dividing for efficient wavenumber selection: Application to estimation of water and drug content in granules using near infrared spectroscopy," *Int. J. Pharm.* **475**(1–2), 504–513 (2014).
4. E. Deconinck, P. Y. Sacré, D. Coomans, "Classification trees based on infrared spectroscopic data to discriminate between genuine and counterfeit medicines," *J. Pharma. Biomed. Anal.* **57**(1), 68–75 (2012).
5. P. I. Storme, H. Rebiere, M. Matoga, "Challenging near infrared spectroscopy discriminating ability for counterfeit pharmaceuticals detection," *Anal. Chim. Acta* **658**(2), 163–174 (2010).
6. K. Yu, Y. Cheng, "Discriminating the genuineness of Chinese medicines with least squares support vector machines," *Chin. J. Anal. Chem.* **34**(4), 561–564 (2006).
7. G. E. Hinton, R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science* **313**(5786), 504–507 (2006).
8. M. Långkvist, L. Karlsson, A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.* **42**(1), 11–24 (2014).
9. N. Srivastava, R. Salakhutdinov, "Multimodal learning with deep Boltzmann machines," *H. Mach. Learn. Res.* **15**(8), 1967–2006 (2014).
10. E. Variiani, X. Lei, E. Mcdermott, "Deep neural networks for small footprint text-dependent speaker verification," *IEEE Int. Conf. Acoustics Speech and Signal Processing*, 4–9 May 2014, Florence, pp. 4052–4056.
11. Z. C. Luo, "Research on auto-encoder model and model transfer in near infrared spectroscopy," *IEEE Trans. Neural Netw.* (2015).
12. G. E. Hinton, S. Osindero, T. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.* **18**(7), 1527–1554 (2006).
13. A. R. Mohamed, D. E. Dahl, G. E. Hinton, "Acoustic modeling using deep belief networks," *IEEE Trans. Audio, Speech, Lang. Process.* **20**(1), 14–22 (2012).
14. R. Sarikaya, G. E. Hinton, A. Deoras, "Application of deep belief nets for natural language understanding," *IEEE Trans. Audio, Speech, Lang. Process.* **22**(4), 778–784 (2014).
15. G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.* **14**(8), 1711–1800 (2002).

16. T. Kuremoto, S. Kimura, K. Kobayashi, M. Obayashi, "Time series forecasting using a deep belief network with restricted Boltzmann machines," *Neurocomputing* **137**(15), 47–56 (2014).
17. N. Srivastava, G. E. Hinton, A. Krezhevsky, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014).