Journal of Innovative Optical Health Sciences
Vol. 7, No. 4 (2014) 1450020 (10 pages)

DOI: 10.1142/S1793545814500205

World Scientific
www.worldscientific.com

Drug discrimination of Near Infrared spectroscopy based on the scaled convex hull classifier

Zhenbing Liu*, Shujie Jiang† and Huihua Yang‡
School of Electronic Engineering and Automation
Guilin University of Electronic Technology
No. 1 Jinji Road, Guilin, P. R. China
**zbliu@guet.edu.cn*
†*462832784@qq.com*
‡*yang98@guet.edu.cn*

Received 25 July 2013
Accepted 8 October 2013
Published 20 February 2014

Near Infrared spectroscopy (NIRS) has been widely used in the discrimination (classification) of pharmaceutical drugs. In real applications, however, the class imbalance of the drug samples, i.e., the number of one drug sample may be much larger than the number of the other drugs, deceases drastically the discrimination performance of the classification models. To address this class imbalance problem, a new computational method — the scaled convex hull (SCH)-based maximum margin classifier is proposed in this paper. By a suitable selection of the reduction factor of the SCHs generated by the two classes of drug samples, respectively, the maximal margin classifier between SCHs can be constructed which can obtain good classification performance. With an optimization of the parameters involved in the modeling by Cuckoo Search, a satisfied model is achieved for the classification of the drug. The experiments on spectra samples produced by a pharmaceutical company show that the proposed method is more effective and robust than the existing ones.

Keywords: Drug classification; Near Infrared spectroscopy; class imbalance; scaled convex hulls.

1. Introduction

Near Infrared spectroscopy (NIRS) is applied in diverse fields due to its characteristics such as rapidity, simplicity and nondestructive measurements.[1] It can provide highly sensitive, low-cost and nondestructive analysis of various samples.[6] However, the presence of the relatively weak and highly overlapping spectral bands[2] in the NIR spectra poses a challenge for extracting information from the samples. It is essential to apply computational methods to preprocess the measured spectra and to build the analytical model in NIRS.

Existing computational methods can be classified into two categories: quantitative and qualitative. Multiple linear regression (MLR),[3] principal component regression (PCR),[4] partial least squares (PLS),[5–7] artificial neural network (ANN),[8] etc., are typical quantitative methods. Classification trees,[9] K nearest neighbor (KNN),[10] support vector machine (SVM)[11] and PLS-DA[12] are commonly used in qualitative analysis of NIR spectra. In this paper, we focus on the quantitative analysis (classification) of the pharmaceutical drug.

Although many computational methods have been proposed for the discrimination of the pharmaceutical drug,[2,13] there is a common drawback that the class imbalance of the drug samples was not considered. In real applications, different drug classes may contain very different number of samples. Classification algorithms that do not consider class-imbalance tend to be overwhelmed by the majority class and ignore the minority one. However, in many cases, the cost of misclassifying a minor class example is usually more expensive than that of misclassifying a major one. Thus, it is necessary to design a classification algorithm that is biased toward the minority class.

To address the class imbalance problem and obtain the high classification accuracy in the drug discrimination, the scaled convex hull (SCH)-based maximum margin classifier is proposed in this paper. With an optimization of the parameters involved in the SCH modeling by Cuckoo Search, we can produce a model that is biased toward the minority class and has a good performance to detect the genuine and counterfeit drug. Comparing with PLS and SVM classification models, the experiments on NIR samples of erythromycin ethylsuccinate have shown that the SCH classification model has better performance than the others.

2. Theory

2.1. *SCH*

Given a finite training data set $T = (x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$, where x_i is the feature or input vector, $y_i = +1$ or -1 is the class label or output which indicates the membership of x_i. The task of computational methods is to learn a decision function $f(x)$ from the training set that can predict the membership of the unseen sample x.[14] Generally, x is given the label $y = +1$ if $f(x) > 0$, and $y = -1$ otherwise.

For the SVM training algorithms finding the maximum margin between the two sets is equivalent to finding the closest points between the convex hulls that contain each class.[16] $X = \{x_i | x_i \in R^d, I = 1, 2 \ldots, K\}$ is defined as[14]

$$\mathrm{conv}X = \left\{ \omega | \omega = \sum_{i=1}^{k} a_i x_i, \sum_{i=1}^{k} a_i = 1, a_i \geq 0, x_i \in X \right\}, \tag{1}$$

where a_i is the coefficient. It can be seen that the convex hull is the linear combination of the samples with some constraint on the coefficient.

The SCH generated by the sample set $X = \{x_i | x_i \in R^d, I = 1, 2 \ldots, K\}$ is defined as

$$S(X, \lambda) = \left\{ \omega | \omega = \lambda \sum_{i=1}^{k} a_i x_i + (1 - \lambda) m, \sum_{i=1}^{k} a_i = 1, 0 \leq a_i \leq \lambda, x_i \in X \right\}. \tag{2}$$

It can also be rewritten as

$$S(X, \lambda) = \left\{ \omega | \omega = \sum_{i=1}^{k} a_i (\lambda x_i + (1 - \lambda) m), \sum_{i=1}^{k} a_i = 1, 0 \leq a_i \leq \lambda, x_i \in X \right\}, \tag{3}$$

where $m = \frac{1}{k} \sum_{i=1}^{k} x_i$ is the mean value of all original points. The reduction factor λ can be set to different size with the constraint that $\lambda < 1$.[15] It can be seen that the smaller the λ, the smaller is the size of SCH ($S(X, \lambda)$). It is proved that no matter how the λ changes, the SCH has the same geometric shape as the original convex hull, which is the reason why we call it as Scaled Convex Hulls.[14] More details about the SCH can be seen in Ref. 14.

Denote $X^+ = \{x_i | y_i = +1\}$ and $X^- = \{x_i | y_i = -1\}$ as the positive and negative sample set, respectively. The SCH of the positive set can be written as

$$S(X^+, \lambda) = \left\{ \omega | \omega = \sum_{y_i = +1} a_i (\lambda x_i + (1 - \lambda) m^+), \sum_{y_i = +1} a_i = 1, 0 \leq a_i \leq 1, x_i \in X^+ \right\}, \tag{4}$$

where $m^+ = \frac{1}{n^+} \sum_{y_i = +1} x_i (n^+ = |X^+|)$ is the mean value of the positive points.

Similarly, the SCH of the negative set can be written as

$$S(X^-,\lambda)=\left\{\omega|\omega=\sum_{y_i=-1}a_i(\lambda x_i+(1-\lambda)m^-),\ \sum_{y_i=-1}a_i=1,0\le a_i\le 1,x_i\in X^-\right\},\quad(5)$$

where $m^-=\frac{1}{n^+}\sum_{y_i=-1}x_i(n^-=|X^-|)$ is the mean value of the negative points.

For convenience, we denote the "reduced" point $\lambda x_i+(1-\lambda)m$ as x_i', then the SCHs of the positive and negative sample set, respectively can be rewritten as

$$S(X^+,\lambda)=\left\{\omega|\omega=\sum_{y_i=+1}a_ix_i',\ \sum_{y_i=+1}a_i=1,0\le a_i\le 1,x_i'\in X'^+\right\},\quad(6)$$

$$S(X^-,\lambda)=\left\{\omega|\omega=\sum_{y_i=-1}a_ix_i',\ \sum_{y_i=-1}a_i=1,0\le a_i\le 1,x_i'\in X'^-\right\}.\quad(7)$$

Furthermore, the reduction factor λ can be set different for each class, reflecting the importance of each class.

2.2. *SCH classifier*

For nonlinear classification problems, the original convex hull of the two classes, samples are overlapping as seen in Fig. 2(a). But by a suitable selection of λ, the SCHs become smaller (see Fig. 1), and the initially overlapping convex hulls can be reduced to become separable. Once separable, we

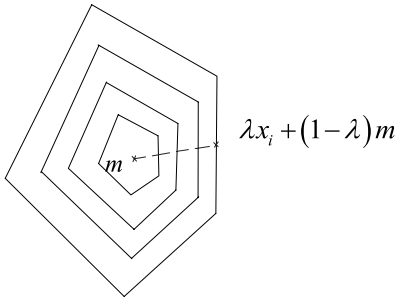
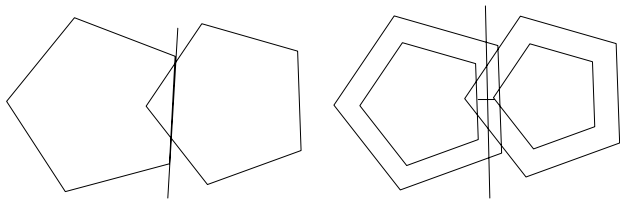


Fig. 1. The SCH of the positive sample set.[15]



(a) (b)

Fig. 2. The two SCHs become separable by a suitable reduction factor.[16]

can find the maximum margin classifiers between the two SCHs through the use of the nearest point pair algorithms.[17] Specifically, the maximum margin classifier is the hyperplane that bisects and is perpendicular to the segment of the nearest point pair.

The finding of the nearest point pair between the two SCHs can be described as the following optimization problem

$$\begin{aligned}\min_{\alpha}\ &\frac{1}{2}\left\|\sum_{y_i=+1}a_ix_i'-\sum_{y_i=-1}a_ix_i'\right\|^2\\ \text{s.t.}\ &\sum_{y_i=+1}a_i=1,\ \sum_{y_i=-1}a_i=1,\\ &0\le a_i\le 1,\quad i=1,2,\ldots,k\end{aligned}\quad(8)$$

where $\sum_{i\in I^+}a_ix_i'$ and $\sum_{i\in I^-}a_ix_i'$ are the SCHs of the positive and negative sample set and (8) presents the shortest distance of the pair of nearest points between two SCHs.

It can be shown that (8) is a convex optimization whose global optimal solution can be solved efficiently. Suppose the optimal values are $a^*=(a_1^*,a_2^*,\ldots,a_k^*)^T$, and the pair of nearest points are simplified as $c=\sum_{y_i=+1}a_i^*x_i'$ and $d=\sum_{y_i=-1}a_i^*x_i'$. The decision function in the feature space will accordingly be of the form

$$f(x)=(\omega^*\cdot x)+b^*,\quad(9)$$

where $w^*=c-d$ and $b^*=-\frac{1}{2}((c-d)\cdot(c+d))$.

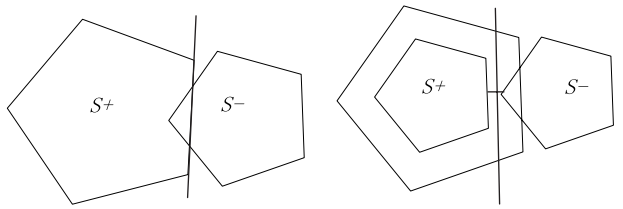
The decision function obtains the maximum margin between the two SCHs, so we called it the SCH based maximum margin classifier. This method can be easily nonlinearized through the kernel trick.

2.3. *SCH classifier for class imbalance problem*

In real applications, different drugs may contain very different number of samples. Classification algorithms

that do not consider class imbalance tend to be overwhelmed by the majority class and ignore the minority one. But in many cases, the minority class is more important to us. Thus, it is necessary to design a classification algorithm that is biased toward the minority class.

As said in Sec. 2.1, the reduction factor λ can be set different for each class to reflect the importance of each class. This can be utilized to solve class imbalance problems. For example, for the majority class, we set a smaller λ to make the corresponding SCH smaller; for the minority class, we set a larger λ to make the corresponding SCH larger. By doing so, the final decision function will move toward the majority class, avoiding the minority class being ignored, as shown in Fig. 3. Fig. 3(a) is the result without considering class imbalance, and Fig. 3(b) addresses the class imbalance by our SCH. Actually, Fig. 3(b) illustrates an extreme case where the minority class is not reduced, and it can be seen that the resulting classifier is biased toward the minority class, compared to that in Fig. 3(a).



(a) (b)

Fig. 3. The SCHs for class imbalance problem.

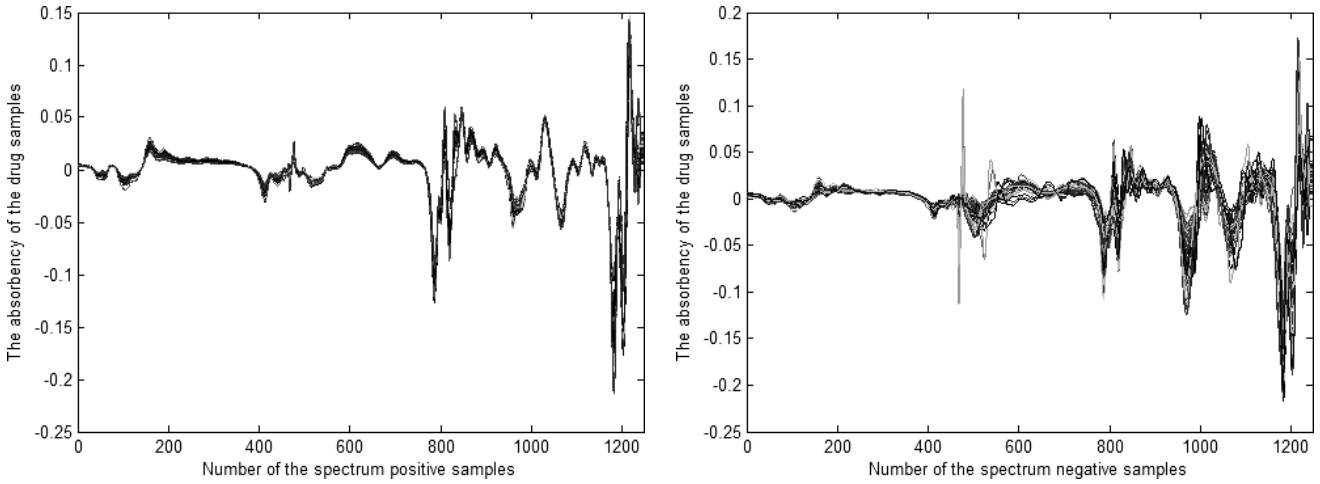
3. Measurement of the Classification Performance

3.1. *Dataset*

All the drug samples of NIR are produced by the Xi'an-Janssen pharmaceutical factores and other factories. To obtain a successive performance of the spectrum samples, the vector normalization and first derivative (SNV-1st D)[17] was applied to preprocess the spectrum data of the drugs, which can eliminate the offset and drift caused by the spectrum deviation.[5] After the processing, the spectrum data were obtained as shown in Fig. 4. Then, the PLS is applied to each sample spectrum to produce a feature vector x_i.

A sample x_i is labeled positively $y_i = +1$ if the spectrum data is the genuine drug [see Fig. 4(a)] or negatively $y_i = -1$ if the spectrum data are counterfeit drug [see Fig. 4(b)]. Finally, the spectrum dataset contains 171 positive and 78 negative drugs.

In order to build a true independent test, the experiments are divided into three categories. The first category, with the gradual decreasing of training samples, does an independent test for the classification performance of the SCH. It is used to study the impact of the number of training data on the classifiers' performance (see Secs. 4.1–4.3). The second category which comprises the class imbalance samples acts as the training samples (see Sec. 4.4). The third category adopts the published pharmaceutical tablet datasets to demonstrate the classification performances of the SCH model (see Sec. 4.5).



(a) (b)

Fig. 4. The samples of NIRS.

3.2. *Experimental setting*

In this paper, data processing and modeling were performed using MATLAB 2012a (The Mathworks, Natick, MA, USA). We experimentally determined the parameters' best values for SVM models by grid optimization. We focus on the RBF kernel for the SCH, which, in practice, is the most widely used for nonlinear SVMs.[18] During the process of optimizing the parameters of the SCH, Cuckoo Search[19,20] is also utilized for a suitable selection of the reduction factor λ. After obtaining the optimal parameters, the SCH performs the classification comparison with the PLS model and SVM model.

3.3. *Measurement of classifier's performance*

Besides the classification accuracy, the F-measure is used as the performance evaluation criteria of all the models[17] for the class imbalance problems, which is amore stable measure, especially for datasets with huge class imbalance. The higher the F-measure, the "better" the quality of the evidence provided by overlaps from the class imbalance sets.[22] For the class imbalance in this paper, F-measure and the classification accuracy are used to assess the prediction performance.[17]

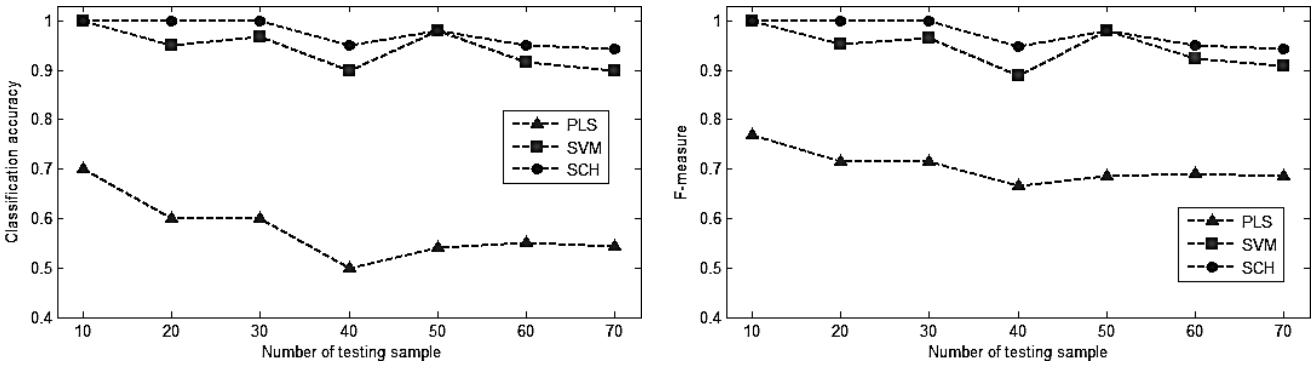
4. Results and Discussion

4.1. *Experiments of discriminating the samples from different manufacturers*

In order to check the classification performances of the SCH model for a kind of drug discrimination from different manufacturers, models are built for recognizing two other types of erythromycin ethylsuccinate tablet. One type is from the Xi'an-Janssen pharmaceutical factory as the positive sample; other type is from other manufacturers as the negative sample. One type is from the Xi'an-Janssen pharmaceutical factory; other type is from other manufacturers. It can be seen in Fig. 5 and Table 1 that the SCH models obtain the obvious performance improvement in comparison with the SVM and the PLS models. As shown in Fig. 5(a), when the training sample has been large enough, the SVM has high classification accuracy in the drug discrimination, and adjusting the reduction factor dose not have any more influence on the classification performances of the SCH. Owing to the property of the PLS algorithm, the fitting curve always exists the accumulating deviation between the actual output and prediction values as shown in Fig. 5. At the same time, the PLS classification accuracy will be declined with decreasing number of training samples.

4.2. *Experiments of discriminating different drugs*

Performance comparison is based on different tablets for discriminating erythromycin ethylsuccinate tablets by the SCH model in comparison with other models. The results in Fig. 6 and Table 2 show that the SCH models reach obvious performance improvement in the drug discrimination by making use of the constantly increasing types of other drugs (see Fig. 6). It can be seen that the SCH model perhaps has the potential to handle more problems with small samples. And it has better classification performance to discriminate the different tablets than other models.

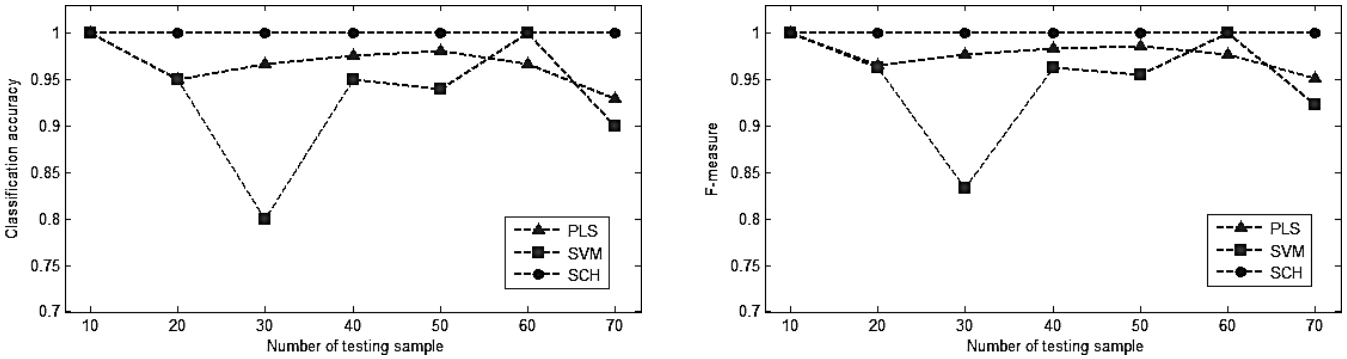


(a) (b)

Fig. 5. The classification accuracy for the different manufacturers samples.

Table 1. Classification accuracy table for the different manufactures samples.

Number of testing samples	10	20	30	40	50	60	70
Classification Accuracy							
PLS	0.7000	0.6000	0.6000	0.5000	0.5400	0.5500	0.5429
SVM	1	0.9500	0.9667	0.9000	0.9800	0.9167	0.9000
SCH	1	1	1	0.9500	0.9800	0.9500	0.9429
F_measure							
PLS	0.7692	0.7143	0.7143	0.6667	0.6849	0.6897	0.6863
SVM	1	0.9524	0.9655	0.8889	0.9796	0.9231	0.9091
SCH	1	1	1	0.9474	0.9796	0.9492	0.9429



(a) (b)

Fig. 6. The classification accuracy for different drug samples.

4.3. *Experiments for the small samples*

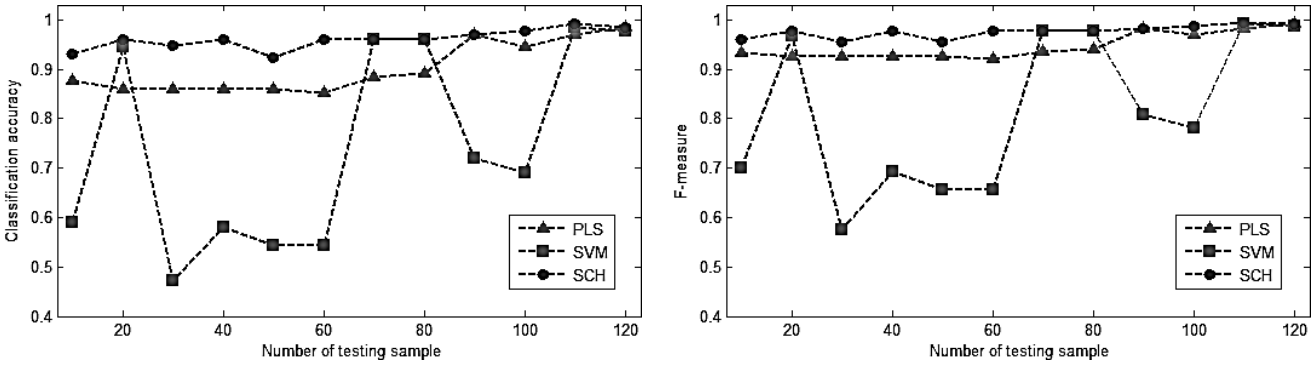
For checking the classification performances of the SCH model on the small training samples, we experiment with a gradual increase of training samples set. The results show that the SCH model is superior than the PLS and SVM models in Table 3 and Fig. 7. When compared with the PLS and SVM models, the results of the SCH model show that the classification performance is improved combined with Cuckoo Search research for selecting the reduction factor. The SCH model is an efficient algorithm to handle more problems with small samples as shown in Fig. 7. For the invariant training samples, the PLS classification accuracy has slightly increased (see Fig. 7).

Table 2. Classification accuracy table for different drugs samples.

Number of test samples	10	20	30	40	50	60	70
Classification Accuracy							
PLS	1	0.9500	0.9667	0.9750	0.9800	0.9667	0.9286
SVM	1	0.9500	0.8000	0.9500	0.9400	1	0.9000
SCH	1	1	1	1	1	1	1
F_measure							
PLS	1	0.9655	0.9767	0.9825	0.9859	0.9767	0.9515
SVM	1	0.9630	0.8333	0.9630	0.9552	1	0.9231
SCH	1	1	1	1	1	1	1

Table 3. Classification accuracy table for the small samples.

Number of training samples	10	20	30	40	50	60	70	80	90	100	110	120
Classification accuracy												
PLS	0.8760	0.8605	0.8605	0.8605	0.8605	0.8527	0.8837	0.8915	0.9690	0.9457	0.9690	0.9845
SVM	0.5891	0.9457	0.4729	0.5814	0.5426	0.5426	0.9612	0.9612	0.7209	0.6899	0.9845	0.9767
SCH	0.9302	0.9612	0.9467	0.9612	0.9225	0.9612	0.9612	0.9612	0.9690	0.9767	0.9922	0.9845
F_measure												
PLS	0.9328	0.9250	0.9250	0.9250	0.9250	0.9205	0.9356	0.9397	0.9823	0.9689	0.9820	0.9911
SVM	0.7006	0.9686	0.5750	0.6932	0.6550	0.6550	0.9780	0.9780	0.8085	0.7802	0.9909	0.9863
SCH	0.9610	0.9778	0.9545	0.9778	0.9545	0.9778	0.9780	0.9780	0.9818	0.9867	0.9955	0.9909

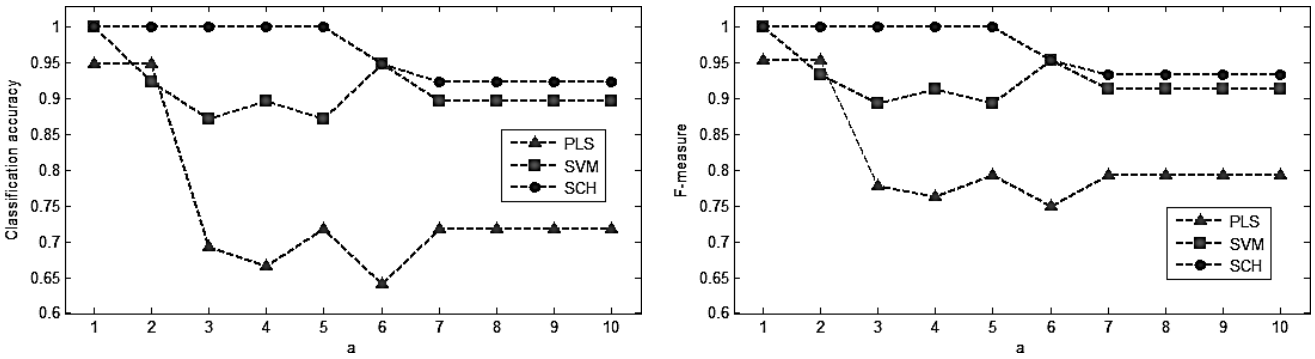


(a) (b)

Fig. 7. The classification accuracy for the small samples.

4.4. *Experiments for the class imbalance samples*

The performance of the three models for the class imbalance in the drug discrimination is shown in Fig. 8 and Table 4, where the parameter a is expressed as the imbalance level of the training sample, i.e., the ratio of the number of minority class samples and reducing the number of majority class. As shown in Fig. 8, the SVM model is built by using only the training examples on second category dataset. The results indicate that the PLS method gives poor performance to decline the classification accuracy to 64.1% (see Fig. 8). A similar situation also happens in the prediction using the PLS method (see Fig. 8). The main reason for poor performances reported by the PLS and SVM models is that all examples are focused on as positive class due to the small size of negative samples and a huge



(a) (b)

Fig. 8. The classification accuracy for the imbalanced samples.

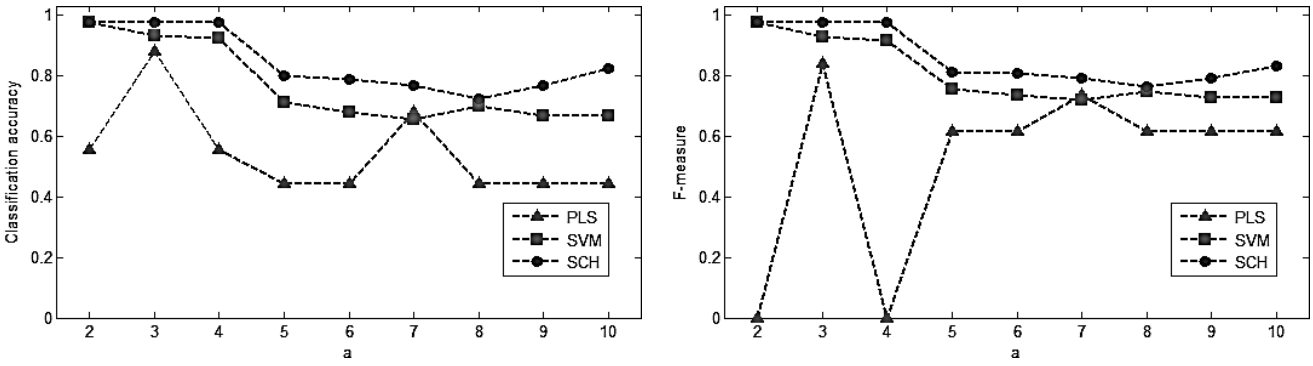
Table 4. Classification accuracy table for the imbalanced samples.

Positive (negative) sample	50 (50)	100 (50)	120 (40)	120 (30)	150 (30)	120 (20)	140 (20)	160 (20)	135 (15)	150 (15)
Imbalance level	01:01	02:01	03:01	04:01	05:01	06:01	07:01	08:01	09:01	10:01
Classification accuracy										
PLS	0.9487	0.9487	0.6923	0.6667	0.7179	0.6410	0.7179	0.7179	0.7179	0.7179
SVM	1	0.9231	0.8718	0.8974	0.8718	0.9487	0.8974	0.8974	0.8974	0.8974
SCH	1	1	1	1	1	0.9487	0.9231	0.9231	0.9231	0.9231
F_measure										
PLS	0.9545	0.9545	0.7778	0.7636	0.7925	0.7500	0.7925	0.7925	0.7925	0.7925
SVM	1	0.9333	0.8936	0.9130	0.8936	0.9545	0.9130	0.9130	0.9130	0.9130
SCH	1	1	1	1	1	0.9545	0.9333	0.9333	0.9333	0.9333

imbalance of positive and negative examples in training classifiers.[21] Then, the SCH classification accuracy is above 92.31% in Fig. 8, and it was proved that the SCH model has higher classification accuracy than other models. Thus, it can be inferred that the SCH model is helpful to address the class imbalance problem in the drug discrimination.

4.5. *Experiments for the active substance on pharmaceutical tablets dataset*

To demonstrate the SCH method, we experiment for the active substance on published pharmaceutical tablets dataset (http://www.models.kvl.dk/Tablets). These tablets exist in four dosages with only two different concentrations of active



(a) (b)

Fig. 9. The classification accuracy for the active substance on pharmaceutical tablets dataset.

Table 5. Classification accuracy table for the active substance on pharmaceutical tablets dataset.

Positive (negative) sample	40 (20)	60 (20)	80 (20)	100 (20)	120 (20)	140 (20)	160 (20)	180 (20)
Imbalance level	02:01	03:01	04:01	05:01	06:01	07:01	08:01	09:01
Classification accuracy								
SVM	0.5556	0.8778	0.5556	0.4444	0.4444	0.6778	0.4444	0.4444
PLS	0.9778	0.9333	0.9222	0.7111	0.6778	0.6556	0.7000	0.6667
SCH	0.9778	0.9778	0.9778	0.8000	0.7889	0.7667	0.7222	0.7667
F_measure								
PLS	0.0000	0.8406	0.0000	0.6154	0.6154	0.7339	0.6154	0.6154
SVM	0.9750	0.9286	0.9176	0.7547	0.7339	0.7207	0.7477	0.7273
SCH	0.9750	0.9750	0.9750	0.8125	0.8081	0.7921	0.7619	0.7921

substance (5 mg (5.6%w/w), and 10, 15 and 20 mg (8.0%w/w) active substance per tablet).[23] We construct artificial sets of patterns in the experiment to discriminate the two different concentrations for pharmaceutical tablets spectrum, where the parameter a is expressed as the imbalance level of the training sample. Finally, we experiment to demonstrate the best classification performances of the SCH model. The results demonstrate that the SCH method is the most effective for the class imbalance (see Fig. 9 and Table 5).

5. Conclusion

To address the problem of imbalance of the drug samples, i.e., a huge imbalance of positive versus negative examples in training classifiers,[21] the SCH-based maximum margin classifier is proposed. The SCH not only can address the imbalance, but also transforms the nonlinear discrimination problems to linear ones. With an optimization of the parameters involved in the modeling by Cuckoo Search, a satisfied model is achieved for the classification of the drug. From comparison with other machine learning methods, the results demonstrate that the SCH method is the most effective one for the class imbalance in the drug discrimination.

Acknowledgments

Our study was funded by the National Natural Science Foundation of China (Grant Nos. 61105004, 61071136 and 21365008), Natural Science Foundation of Guangxi (Grant No. 2013GXNSFBA01 9279) and Innovation Project of GUET Graduate Education (No. ZYC0725). We thank Xi'an-Janssen Pharmaceutical Factory for useful datasets of Near Infrared spectra samples. We would like to express our appreciation to all supporters above mentioned for their strongly financial support.

References

1. Y. Li, X. Shao, W. Cai, "A Consensus least squares support vector regression (LS-SVR) for analysis of near-infrared spectra of plant samples," *Talanta* **72**(1), 217–222 (2006).
2. M. Cocchi, M. Corbellini, G. Foca, M. Lucisano, M. A. Pagani, L. Tassi, A. Ulrici, "Classification of bread wheat flours in different quality categories by a wavelet-based feature selection/classification algorithm on NIR spectra," *Anal. Chim. Acta* **544**(1), 100–107 (2005).
3. H.-Y. Lu, S.-S. Wang, R. Cai, Y. Meng, X. Xie, W. Zhao, "Rapid discrimination and quantification of alkaloids in corydalis tuber by near-infrared spectroscopy," *J. Pharm. Biomedi. Anal.* **59**(5), 44–49 (2012).
4. N. Qu, M. Zhu, H. Mi, Y. Dou, Y. Ren, "Non-destructive determination of compound amoxicillin powder by NIR spectroscopy with The aid of chemometrics," *Spectrochim. Acta A* **70**(5), 1146–1151 (2008).
5. C. Tan, J. Wang, T. Wu, X. Qin, M. Li, "Determination of nicotine in tobacco samples by near-infrared spectroscopy and boosting partial least squares," *Vib. Spectrosc.* **54**(1), 35–41 (2010).
6. S. A. Haughey, S. F. Graham, E. Cancouët, C. T. Elliott, "The application of near-infrared reflectance spectroscopy (NIRS) to detect melamine adulteration of soya bean meal," *Food Chem.* **136**(3), 1557–1561 (2013).
7. Y. Yang, J. Tu, W. Cai, X. Shao, "Feasibility for quantitative determination of deoxyribonucleic acid by using near-infrared diffuse reflectance spectroscopy," *Talanta* **99**, 871–874 (2012).
8. S. Serranti, D. Cesare, F. Marini, G. Bonifazi, "Classification of oat and groat kernels using NIR hyperspectral imaging," *Talanta* **103**(15), 276–284 (2013).
9. E. Deconinck, P. Y. Sacré, D. Coomans, J. D. B. Raman, "Classification trees based on infrared spectroscopic data to discriminate between genuine and counterfeit medicines," *J. Pharm. Biomed. Anal.* **57**(5), 68–75 (2012).
10. W. Luo, S. Huan, H. Fu, G. Wen, H. Cheng, J. Zhou, R. Yu, "Preliminary study on the application of near infrared spectroscopy and pattern recognition methods to classify different types of apple samples," *Food Chem.* **128**(2), 555–561 (2011).
11. A. Candolfi, R. De Maesschalck, D. L. Massart, P. A. Hailey, A. C. Harrington "Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA," *J. Pharm. Biomed. Anal.* **19**(6), 923–935 (1999).
12. P.-Y. Sacré, E. Deconinck, T. De Beer, P. Courselle, R. Vancauwenberghe, P. Chiap, J. Crommen, J. O. De Beer, "Comparison and combination of spectroscopic techniques for the detection of counterfeit medicines," *J. Pharm. Biomed. Anal.* **53**(3), 445–453 (2010).
13. Y. Lai, Y. Ni, S. Kokot, "Discrimination of rhizoma corydalis from two sources by near-infrared spectroscopy supported by the wavelet transform and least-squares support vector machine methods," *Vib. Spectrosc.* **56**(2), 154–160 (2011).

14. L. Zhenbing, L. Jianguo, P. Chao *et al.*, "A novel geometric approach to binary classification based on scaled convex hulls," *IEEE Trans. Neural Netw.* **20**(7), 1215–1220 (2009).
15. M. E. Mavroforakis, S. Theodoridis, "A geometric approach to support vector machine (Svm) classification," *IEEE Trans. Neural Netw.* **17**(3), 671–682 (2006).
16. K. P. Bennett, E. J. Bredensteiner, "Duality and geometry in SVM classifiers," *ICML*, 57–64 (2000).
17. B. Yu, H. Li, M. Lin, H. Bo, L. Zhang, "Support vector machine based on the sample spatial distance for imbalanced sample distribution," *J. Compu. Inf. Sys.* **8**(21), 8753–8761 (2012).
18. Z. Fu, A. Robles-Kelly, J. Zhou, "Mixing linear SVMs for nonlinear classification," *IEEE Trans. Neural Netw.* **21**(12), 1963–1975 (2010).
19. X.-S. Yang. "Cuckoo search via Lévy flights," *Proc. of World Congress on Nature & Biologically Inspired Computing*, pp. 210–214, IEEE Publications (2009).
20. X.-S. Yang, S. Deb, M. Karamanoglu, X. He, "Cuckoo search for business optimization applications," *IEEE National Conf. Computing and Communication Systems*, pp. 1–5 (2012).
21. J.-S. Wu, Z.-H. Zhou, "Sequence-based prediction of microRNA-binding residues in proteins using cost-sensitive laplacian support vector machines," *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **10**(3), 752–759 (2013).
22. S. Banerjee, T. Pedersen, "Extended gloss overlaps as a measure of semantic relatedness," *IJCAI'03 Proc. 18th Int. Joint Conf. Artificial Intelligence*, **3**, 805–810 (2003).
23. M. Dyrby, S. B. Engelsen, L. Nørgaard, M. Bruhn, L. Lundsberg-Nielsen, "Chemometric Quantitation of The Active Substance (Containing C≡N) In A Pharmaceutical Tablet Using Near-Infrared (NIR) Transmittance And NIR Ft-Raman Spectra," *Appl. Spectrosc.* **56**(5), 579–585 (2002).