

Research Paper

Prediction of Milk/Plasma Drug Concentration (M/P) Ratio Using Support Vector Machine (SVM) Method

Chunyan Zhao,¹ Haixia Zhang,^{1,4} Xiaoyun Zhang,¹ Ruisheng Zhang,^{1,2} Feng Luan,¹ Mancang Liu,¹ Zhide Hu,¹ and Botao Fan³

Received April 14, 2005; accepted September 19, 2005

Purpose. Development of reliable computational models to predict/classify milk-to-plasma (M/P) drug concentration ratio remains a challenging object. Support vector machine (SVM) method, as a new algorithm, was constructed to distinguish the potential risk of drugs to nursing infants.

Methods. Each drug was represented by a large pool of descriptors, of which five were found to be most important for constructing the predictive models. Next, two classification models, linear discriminant analysis (LDA) and SVM, were developed with bootstrapping validation based on the selected molecular descriptors.

Results and Conclusions. The classification accuracy of training set and test set for SVM was 90.63 and 90.00%, respectively. The total accuracy for SVM was 90.48%, which was higher than that of LDA (77.78%). Comparison of the two methods shows that the performance of SVM was better than that of LDA, which implies that the SVM method is an effective tool in evaluating the risk of drugs when experimental M/P ratios have not been investigated.

KEY WORDS: bootstrapping validation; linear discriminant analysis (LDA); milk/plasma drug concentration (M/P) ratio; molecular descriptors; support vector machine (SVM).

INTRODUCTION

The issue of which drugs are safe to take during lactation is quite important and complicated. When nursing mothers take drugs, nearly all drugs are excreted into breast milk and are bioavailable to the infants (1). Generally, most drugs do not pose a significant problem to nursing infants, because medication ingested by infants seem to be proportionally small, usually less than 1% of the maternal dose. However, considering the size of infants and differences in metabolism between infants and adults, physicians need to obtain more accurate information to

decide which drugs can be safely used and which drugs should be used with caution during lactation (2–4).

A direct assessment of drugs' risk to infants entails the measurement of drug concentrations in breast milk and extrapolation to other doses/patients by using the milk/plasma (M/P) drug concentration ratio, which is equivalent to drug concentration in the breast milk divided by that in the maternal serum. However, it is difficult to enroll nursing women in a trial solely to assess the pharmacokinetics of a compound in milk and plasma (5). Accurate determination of the M/P ratio also requires careful attention to experimental details, and collection of serial milk and plasma samples. Given the difficulties, combinations of experiments *in vitro/in vivo* and modeling approaches should be sufficient to determine whether a particular drug will show high M/P ratio in human milk, and thus pose a potential risk to nursing infants.

Linear regression methods have been developed relating M/P ratio to the drug's physical–chemical properties. Meskin and Lien (6) were the first to include the physicochemical properties of a drug in the prediction of M/P ratios, and Agatonovic-Kustrin *et al.* (5,7) extended this method by using artificial neural networks (ANNs) to allow the prediction of ratios. It should be noted that in most cases, for a certain drug, the M/P ratios measured under different conditions varied significantly. For instance, the maximum M/P ratio of acyclovir was 4.1, whereas the minimum ratio was only 0.6—representing a sevenfold difference between the two values. It seemed that the predictive models based on these

¹ Department of Chemistry, Lanzhou University, Lanzhou 730000, China.

² Department of Computer Science, Lanzhou University, Lanzhou 730000, China.

³ Université Paris 7-Denis Diderot, ITODYS 1, Rue Guy de la Brosse, 75005 Paris, France.

⁴ To whom correspondence should be addressed. (e-mail: zhanghx@lzu.edu.cn)

ABBREVIATIONS: ANN, artificial neural networks; *C*, capacity parameter; CART, classification and regression tree; $CI_{\min}(CC)$, min coulombic interaction for a C–C bond; ERM, empirical risk minimization; KNN, K nearest neighbors classification; LDA, linear discriminant analysis; $\log P$, octanol water coefficient; M/P, milk/plasma drug concentration; NMC, nearest mean classifier; $PC_{\max}(C)$, max partial charge for a C atom; QP, quadratic programming; $R_{e-e}(CC)$, min e–e repulsion for a C–C bond; SRM, structural risk minimization; SVM, support vector machine; χ^2 , hydrophobicity. Randic index (order 2); γ , gamma.

Table I. Compounds and Their Corresponding Classification

No.	Compounds	Predicted M/P			Experimental classification	Predicted classification	
		Average	Minimum	Maximum		LDA	SVM
Training set ^a							
1	Cefoxitin	0	0	0	–	–	–
2	Ibuprofen	0	0	0	–	–	–
3	Estradiol	0	0	0	–	–	–
4	Warfarin	0	0	0	–	+++	–
5	Tolmetin	0.005	0.005	0.005	–	–	–
6	Cephalexin	0.012	0.01	0.014	–	–	–
7	Suprofen	0.014	0.014	0.014	–	–	–
8	Carbenicillin	0.02	0.02	0.02	–	–	–
9	Amoxycillin	0.028	0.013	0.043	–	–	–
10	Roxithromycin	0.035	0.03	0.04	–	–	–
11	Ceftriaxone	0.045	0.03	0.06	–	–	–
12	Valproic acid	0.053	0.01	0.096	–	–	+++
13	Oxazepam	0.1	0.1	0.33	–	–	–
14	Sulfamethoxazole	0.1	0.1	0.1	–	+++	–
15	Quinapril	0.12	0.12	0.12	–	+++	–
16	Rosaramicin	0.12	0.12	0.12	–	–	–
17	Temazepam	0.14	0.14	0.14	–	–	–
18	Mefloquine	0.145	0.13	0.16	–	–	–
19	Cefotaxime	0.16	0.16	0.16	–	–	–
20	Indomethacin	0.19	0.01	0.37	–	–	–
21	Lorazepam	0.205	0.15	0.26	–	–	–
22	Methyl dopa	0.265	0.265	0.265	–	–	–
23	Noscapine	0.27	0.27	0.27	–	–	–
24	Ampicillin	0.29	0.29	0.29	–	–	–
25	Penicillin G	0.295	0.01	0.58	–	–	–
26	Doxycycline	0.34	0.32	0.36	–	–	–
27	Nitredipin	0.35	0.35	0.35	–	–	–
28	Phenytoin	0.363	0.142	0.584	–	+++	+++
29	Penicillin V	0.37	0.37	0.37	–	–	+++
30	Clemastine	0.375	0.25	0.5	–	–	–
31	Propranolol	0.403	0.107	0.699	–	–	–
32	Lamotrigine	0.425	0.4	0.45	–	+++	–
33	Gentamicin	0.44	0.44	0.44	–	+++	–
34	Methadone	0.44	0.24	0.64	–	–	–
35	Tiapamil	0.44	0.43	0.45	–	+++	–
36	Erythromycin	0.455	0.41	0.5	–	–	–
37	Carbamazepine	0.465	0.24	0.69	–	+++	–
38	Phenobarbitone	0.5	0.4	0.6	–	–	–
39	Triprolidine	0.53	0.5	0.56	–	–	–
40	Flunitrazepam	0.54	0.54	0.54	–	–	–
41	Norfluoxetine	0.56	0.35	0.77	–	–	–
42	Verapamil	0.6	0.6	0.6	+	+	+
43	Chloramphenicol	0.655	0.655	0.655	+	+	+
44	Phenacetine	0.67	0.67	0.67	+	+	+
45	Fluoxetine	0.68	0.52	0.84	+	–**	+
46	Diazepam	0.7	0.1	1.3	+	–**	+
47	Theophylline	0.7	0.7	0.7	+	+	+
48	Caffeine	0.711	0.61	0.812	+	+	+
49	Medroxyprogesterone	0.72	0.72	0.72	+	–**	–**
50	Moclobemide	0.72	0.69	0.75	+	+	+
51	Paroxetine	0.75	0.39	1.11	+	–**	+
52	Imipramine	0.76	0.76	0.76	+	+	+
53	Minoxidil	0.76	0.76	0.76	+	–**	–**
54	Carbamazepine	0.79	0.79	0.79	+	+	+
55	Decarboethoxyloxadine	0.8	0.8	0.8	+	+	+
56	Timolol	0.8	0.8	0.8	+	+	+
57	Nordothiepin	0.85	0.69	1.01	+	–**	+
58	Disopyramide	0.9	0.9	0.9	+	–**	+
59	Zonisamide	0.93	0.84	1.02	+	+	+
60	Metronidazole	0.95	0.9	1	+	+	+
61	Tetracycline	0.95	0.6	1.3	+	+	+

Table I. Continued

No.	Compounds	Predicted M/P			Experimental classification	Predicted classification	
		Average	Minimum	Maximum		LDA	SVM
62	Diltiazem	0.98	0.98	0.98	+	+	+
63	Tinidazole	1.005	1.005	1.005	+	+	+
64	Clomipramine	1.03	0.84	1.22	+	+	+
65	Dothiepsulfoxide	1.18	0.89	1.47	+	+	+
66	Nortriptyline	1.18	0.5	1.62	+	+	+
67	Loratadine	1.2	1.2	1.2	+	-**	+
68	Sertraline	1.275	0.62	1.93	+	+	+
69	Doxepin	1.37	0.4	1.65	+	+	-**
70	Ciprofloxacin	1.495	0.85	2.14	+	+	-**
71	Amitriptyline	1.53	0.5	1.93	+	+	-**
72	Dothiepin	1.59	1.27	1.91	+	+	+
73	N-Desmethylsertraline	1.64	1.64	1.64	+	+	+
74	Cimetidine	1.7	1.7	1.7	+	+	+
75	Labetalol	1.7	0.8	2.6	+	+	+
76	Demethylcitalopram	1.75	1	2.5	+	+	+
77	Nordothiepsulfoxide	1.86	1.57	2.15	+	+	+
78	Atenolol	2.1	1.1	3.1	+	+	+
79	Citalopram	2.1	1.2	3	+	+	+
80	Codeine	2.16	2.16	2.16	+	+	+
81	Mianserin	2.2	2.2	2.2	+	+	+
82	Nicotine	2.25	1.5	3	+	+	+
83	Nitrofurantoin	2.25	2.25	2.25	+	+	+
84	Acyclovir	2.35	0.6	4.1	+	-**	+
85	Morphine	2.46	2.46	2.46	+	+	+
86	Metoprol	2.55	2	3.1	+	-**	-**
87	Clozapine	3.555	2.79	4.32	+	+	+
88	Venlafaxine	3.8	2.8	4.8	+	+	+
89	Quazepam	4.13	4.13	4.13	+	+	+
90	Cannabis	4.24	0.08	8.4	+	+	+
91	Astemizole	4.4	4.4	4.4	+	-**	+
92	Nadolol	4.6	4.6	4.6	+	-**	+
93	Sumatriptan	4.9	4.1	5.7	+	+	+
94	Amphetamine	5.15	2.8	7.5	+	+	+
95	Satalol	5.4	5.4	5.4	+	+	+
96	Bupropion	5.545	5.545	5.545	+	+	+
Test set ^b							
1	Alprazolam	0.001	0.001	0.001	-	-	-
2	Methotrexate	0.04	0.04	0.04	-	-	-
3	Prednisolone	0.13	0.13	0.13	-	-	-
4	Zolpidem	0.13	0.13	0.13	-	-	-
5	Norethindron	0.19	0.19	0.19	-	-	-
6	Nitrazepam	0.27	0.27	0.27	-	+	-
7	Clonazepam	0.33	0.33	0.37	-	-	-
8	Bupivacaine	0.34	0.1	0.58	-	-	-
9	Oxprenolol	0.37	0.37	0.37	-	+	-
10	Zopiclone	0.555	0.41	0.7	-	+	+
11	Norfluoxetine	0.56	0.56	0.56	-	-	-
12	Holoperidol	0.64	0.59	0.69	+	+	+
13	Cotinine	0.78	0.78	0.78	+	+	+
14	Ethosuximide	0.8	0.8	0.8	+	+	+
15	Theobromine	0.82	0.82	0.82	+	-**	+
16	Paracetamol	0.88	0.76	1	+	+	+
17	Ethanol	0.9	0.9	0.9	+	+	+
18	Perfenazine	0.9	0.7	1.1	+	+	+
19	Desipramine	0.915	0.63	1.2	+	+	+
20	Vigabatrin	1	1	1	+	+	+
21	Lidocaine	1.07	0.25	1.89	+	-**	+
22	Nefopam	1.2	1.2	1.2	+	-**	+
23	Desmethyldoxepin	1.275	1.02	1.53	+	-**	+
24	Mexiletine	1.34	0.79	1.89	+	+	-**
25	Clofazim	1.35	1.35	1.35	+	+	-**
26	Chlorprothixene	1.48	0.38	2.58	+	+	+
27	Aspirin	1.63	0.06	3.2	+	+	+

Table I. Continued

No.	Compounds	Predicted M/P			Experimental classification	Predicted classification	
		Average	Minimum	Maximum		LDA	SVM
28	Mepindolol	2.6	1	4.2	+	+	+
29	Procainamide	3.2	3.2	3.2	+	+	+
30	O-Desmethylvenlafaxine	3.3	2.8	3.8	+	−**	+
External test set ^c							
1	Tylenol					−	−
2	Temptra					−	−
3	Acetaminophen					−	−
4	Prozac					+	−
5	Macrobid					−	−
6	Tetracycline					+	+
7	Prednisone					+	−
8	Propylthiouracil					−	−
9	Paxil					−	−

^aThe predictions for the training set are leave-one-out predictions.

^bThe predictions for the test set and the external test set are based on the final model using the training set.

**Misclassified samples.

data could be misleading and were not applicable for predicting new data. Therefore, the regulation M/P values can be applied to indicate the class of drugs. Comparing with regression models, the classification can offer some advantages to allow better management of data that are often noisy. Thus, in this paper, classification models were constructed to distinguish the potential risk of a drug to nursing infants. Many algorithms can be used for classification, e.g., nearest mean classifier (NMC) (8), linear discriminant analysis (LDA) (9), K-nearest neighbors classification (KNN) (10), and classification and regression tree (CART) (11). Recently, support vector machine (SVM), as a new algorithm, was developed by the machine learning community (12,13). Because of its remarkable generalization performance, SVM has attracted attention and gained extensive applications (14–19), etc.

In this paper, SVM models were developed to distinguish the potential risk of drugs to nursing infants. Each compound was encoded with 400 molecular structure descriptors, which can be directly calculated from the drug's structure. The constructed models would evaluate the risk of drugs when experimental M/P ratios have not been investigated. Furthermore, it would provide some insight into the grade of drugs that can transfer into milk.

MATERIALS AND METHODS

Data Set

The compounds under study consisted of 126 commonly used drugs, whose experimental M/P values were taken from

the literature (5–7,20–25). For these drugs, if there existed more than one M/P value, the average one was calculated as the M/P ratio. As shown in Table I, the maximum ratio of the data set was recorded as 5.545, whereas the minimum was 0. Next, the M/P ratios were scaled in the internal [0, 1] and two classes were defined, including Class 1 indicated as “−” (49 drugs, $0 \geq M/P \geq 0.1$) and Class 2 indicated as “+” (77 drugs, $1 \geq M/P > 0.1$).

Model Validation

The data set was randomly divided into a training set (96 compounds) for model development/calibration and an independent test set (30 compounds) for prediction. In addition, to test the classified ability of the predictive model for new drugs, an external data set of nine drugs whose M/P values were currently unavailable was applied for LDA and SVM models. Initially, bootstrapping validation, which implies quantitative assessment of model robustness and its predictive power, was applied to training set (26). Boosting is a related technique that attempts to drive the test set error rapidly to zero. It produces a series of predictors. The training set used for each member of the series is based on the performance of the preceding predictor(s). The method creates new training sets by choosing patterns for which the predictions of the previous predictors were bad more frequently than those for which the predictions were good. Thus, boosting attempts to produce new predictors for its ensemble that are able to make better predictions for patterns for which the current ensemble performance is poor.

Table II. Five Descriptors and the Coefficients for LDA^a

Des	Chemical meaning	Coefficients	F to remove
Constant		1.740	
$\log P$	<i>n</i> -Octanol–water partition coefficient	−6.092	46.881
χ^2	Randic index (order 2)	6.016	17.402
$PC_{\max}(C)$	Max. partial charge for a C atom	−2.149	12.905
$R_{ee}(CC)$	Min. e–e repulsion for a C–C bond	1.250	9.036
$CI_{\min}(CC)$	Min. coulombic interaction for a C–C bond	0.696	5.152

^a Discriminant function $M/P = 1.740 - 6.092 \log P + 6.016\chi^2 - 2.149 PC_{\max}(C) + 1.250 R_{ee}(CC) + 0.696 CI_{\min}(CC)$.

Table III. Correlation Matrix of the Five Selected Descriptors

	$\log P$	χ^2	$PC_{\max}(C)$	$R_{ee}(CC)$	$CI_{\min}(CC)$
$\log P$	1.000				
χ^2	0.590	1.000			
$PC_{\max}(C)$	0.628	0.822	1.000		
$R_{ee}(CC)$	-0.108	-0.465	-0.480	1.000	
$CI_{\min}(CC)$	-0.176	-0.189	-0.274	0.531	1.000

The resampled training set is assembled using probabilistic selection, with the exception that the probability assigned to each sample depends on the prediction error for that sample by the existing ensemble. According to this method, each training pattern is selected with a probability $1/N$ each time, where N is the total number of patterns in the original sample. Many of the original examples may be repeated in the resampled training set, whereas others may be left out. The method constructs a proxy universe equal in size to the original sample and is, in effect, equivalent to sampling without replacement from an infinitely large replicated universe. Typically, 200 bootstrap replications are sufficient to estimate confidence intervals (27). For this study, 1000 replications are used. Goodness-of-fit of the model is tested by the accuracy of classification and the number of misclassified cases. In addition, to test the classified ability of the predictive model for new drugs, an external data set of nine drugs whose M/P values were currently unavailable, was applied for LDA and SVM models.

Descriptor Generation and Selection

The structures of these compounds were drawn with the ISIS DRAW 2.3 program (28). The final geometries were obtained with the semiempirical PM3 method in the HYPERCHEM 6.03 program (29). All calculations were carried out at restricted Hartree-Fock level with no configuration interaction. The molecular structures were optimized using the Polak-Ribiere algorithm until the root-mean-square gradient was 0.001. Then the resulted geometry was transferred into the CODESSA software, developed by

Katritzky *et al.* (30,31), which can calculate constitutional, topological, geometrical, electrostatic, and quantum chemical descriptors. For this study, 400 descriptors were calculated for each compound, including 38 constitutional descriptors, 38 topological descriptors, 12 geometrical descriptors, 71 electrostatic descriptors, and 241 quantum chemical descriptors. Constitutional descriptors are related to the number of atoms and bonds in each molecule, e.g., number of C atoms, number of bonds, and number of rings. Topological descriptors describe the atomic connectivity in the molecule including valence and nonvalence molecular connectivity indices calculated from the hydrogen-suppressed formula of the molecule, encoding information about the size, composition, and the degree of branching of a molecule, e.g., Wiener index, Randic index, and structural information content. The geometrical descriptors describe the size of the molecule and require 3D coordinates of the atoms in the given molecule, e.g., XY Shadow, molecular volume, and molecular surface area. Electrostatic descriptors reflect the characteristics of the charge distribution of the molecule, e.g., max/min partial charge, count of H acceptor sites, and topographic electronic index. The quantum chemical descriptors provide information about binding and formation energies, partial atom charge, dipole moment, and molecular orbital energy levels, e.g., HOMO/LUMO energy, max/min e-e repulsion, and max/min e-n attraction.

Next, the stepwise variable selection method was applied to choose proper features for the large set of descriptors. F value was used as the criterion to decide which descriptor should be removed/entered into the discriminant function. The entry F value is 3.84 and the removal F value is 2.71. Thus, the obtained five descriptors, together with their coefficients in function, and F values are all listed in Table II. The correlation matrix of the five selected descriptors is shown in Table III. It can be seen that the linear correlation coefficient between any of two descriptors is less than 0.85, which means that the descriptors are independent.

LDA Method

Discriminant analysis is useful for situations where a user wants to build a predictive model of group membership based on the observed characteristics of each case. The procedure generates a discriminant function (or, for more than two groups, a set of discriminant functions) based on linear combinations of the predictor variables that provide the best discrimination between the groups. The functions are generated from a sample of cases for which group membership is known; the functions can then be applied to new cases with measurements for the predictor variables but unknown group membership.

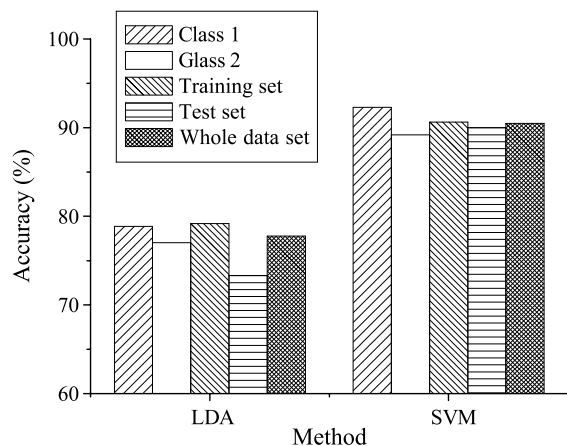


Fig. 1. Comparison of the results obtained from the LDA and SVM method.

Table IV. Optimal γ , C , Number of Support Vectors, and Accuracy of Training Set

γ	C	Number of support vectors	Accuracy (%)
0.215	50	60	90.63
0.220	50	63	90.63
0.785	60	73	90.63

The basic theory of linear discriminant analysis (LDA) is to classify the dependent term by dividing an n -dimensional descriptor space into two regions separated by a hyperplane defined by a linear discriminant function (32,33) as follows:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

where Y is the discriminant score, that is, the dependent variable; $X_1 - X_n$ represents the specific descriptors; and b corresponds to weights associated with the respective descriptors. The linear classification was performed in a stepwise manner: at each step the variable adding the most to the separation of groups is entered into (or the variable adding the least is removed from) the discriminant function. The criteria for the selection were as follows: comparison of the tabulated F , determination of the percentage of molecules correctly classified, and prediction of the classification of molecules not included in the training process. As soon as we have obtained the optimal discrimination conditions for classifying a particular compound, the next step is to obtain new compounds. For this purpose, the final LDA equation was used to select new agents, and the molecules were classified by applying them to the discriminant function.

SVM Method

The following is a brief description of the SVM algorithm. A more detailed description can be found in Vapnik's (12) and Burges's articles (13).

For a binary classification problem, assume that we have a set of samples, i.e., a series of input vectors $\bar{x}_i \in \mathbf{R}^d$ ($i = 1, 2, \dots, N$), with corresponding labels $y_i \in \{-1, +1\}$ ($i = 1, 2, \dots, N$). Here, $+1$ and -1 indicate the two classes. The goal here is to construct one binary classifier or to derive one decision function from the available samples, which has a small probability of misclassifying a future sample. Both the basic linear separable case and the most useful linear nonseparable case (for most real life problems) are considered here.

For a linear separable case, there exists a separating hyperplane whose function is $\bar{\mathbf{w}} \cdot \bar{\mathbf{x}} + b = 0$ which implies the following:

$$y_i(\bar{\mathbf{w}} \cdot \bar{\mathbf{x}}_i + b) \geq 1 \quad i = 1, 2, \dots, N$$

where y_i is the class index, $\bar{\mathbf{w}}$ is a vector normal to the hyperplane, $|b|/||\bar{\mathbf{w}}||$ is the perpendicular distance from the hyperplane to the origin, and $||\bar{\mathbf{w}}||^2$ is the Euclidean norm of $\bar{\mathbf{w}}$.

By minimizing $\frac{1}{2}||\bar{\mathbf{w}}||^2$ subject to this constraint, the SVM approach tries to find a unique separating hyperplane, which maximizes the distance between the hyperplane and the nearest data points of each class. The classifier is called the largest margin classifier.

By introducing Lagrange multipliers a_i , the SVM training procedure amounts to solving a convex quadratic programming (QP) problem. The solution is a unique globally optimized result, and can be shown as:

$$\bar{\mathbf{w}} = \sum_{i=1}^N y_i a_i \bar{x}_i$$

Only if the corresponding $a_i > 0$, are these x_i called support vectors. When an SVM is trained, the decision function can be written as:

$$f(\bar{\mathbf{x}}) = \text{sign} \left(\sum_{i=1}^N y_i a_i (\bar{\mathbf{x}} \cdot \bar{x}_i) + b \right)$$

where sign is the sign of a function. For a linear nonseparable case, allowing for training errors can be done by introducing positive slack variables ξ_i ($i = 1, 2, \dots, N$) in the constraints, which then become:

$$y_i(\bar{\mathbf{w}} \cdot \bar{x}_i + b) \geq 1 - \xi_i \quad \xi_i \geq 0, i = 1, 2, \dots, N$$

We want to simultaneously maximize the margin and minimize the number of misclassifications. This can be achieved by changing the objective function from $\frac{1}{2}||\bar{\mathbf{w}}||^2$ to $\frac{1}{2}||\bar{\mathbf{w}}||^2 + C \sum_{i=1}^N \xi_i^k$.

Minimize $\frac{1}{2}||\bar{\mathbf{w}}||^2 + C \sum_{i=1}^N \xi_i^k$ subject to $y_i(\bar{\mathbf{w}} \cdot \bar{x}_i + b) - 1 + \xi_i \geq 0, i = 1, 2, \dots, N, \xi_i \geq 0, i = 1, 2, \dots, N$.

Error weight C is a regularization parameter to be chosen by the user. It controls the size of penalties assigned to errors. The optimization problem is convex for any positive integer k . For $k = 1$ and $k = 2$, it is also a quadratic programming problem.

For a binary nonlinear classification problem, SVM performs a nonlinear mapping $\Phi(\cdot)$ of the input vector x_i from the input space \mathbf{R}^d into a higher-dimensional Hilbert space \mathbf{H} , and constructs an optimal separating hyperplane. In the linear separable case, we know that the algorithm only depends on inner products between training examples and test examples. Thus we can generalize to nonlinear case. The inner products are substituted by the kernel function $k(\bar{x}_i, \bar{x}_j) = \Phi(\bar{x}_i) \cdot \Phi(\bar{x}_j)$ in the input space. Then, the decision function implemented by SVM can be written as:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^N y_i a_i k(\bar{\mathbf{x}}, \bar{x}_i) + b \right)$$

Two typical kernel functions are listed below;

$$\begin{aligned} \text{Polynomial function} \quad k(\bar{x}_i, \bar{x}_j) &= (\bar{x}_i \cdot \bar{x}_j + 1)^d \\ \text{Gaussian radial basis function} \quad k(\bar{x}_i, \bar{x}_j) &= \exp(-\gamma ||\bar{x}_i - \bar{x}_j||^2) \end{aligned}$$

RESULTS

Results of LDA

Results of the LDA model using bootstrapping validation are listed in Table I, together with misclassified samples marked by double asterisks. This gave an accuracy of 78.85%

for Class 1, 77.03% for Class 2, 79.17% for training set, 73.33% for test set, and 77.78% for the whole data set (see Fig. 1). In addition, the predictive results of the external nine drugs are also listed in Table I.

Results of SVM

SVM was used to develop a nonlinear model. As with other multivariate statistical models, the performance of SVM for classification depends on the combination of several parameters.

First, the kernel functions should be established, because this determines the sample distribution in the mapping space. There are a number of kernel functions, including linear, polynomial, spline, and radial basis function. For classification tasks, a commonly variant is the Gaussian radial basis function because of its good general performance (34). Next is the capacity parameter C , a regularization parameter controlling the tradeoff between maximizing the margin and minimizing the training error. If C is too small, then insufficient stress will be placed on fitting the training data. If C is too large, then the algorithm will overfit the training data. To make the learning process stable, a large value should be set up for C (35). In this study, the initial value was 100. The third parameter is γ . This parameter greatly affects the number of support vectors (SV), which have a close relation with training time. Too many support vectors can produce overfitting and make the training time longer. Parameter γ also controls the amplitude of the Gaussian function and, therefore, controls the generalization ability of SVM. Based on our experience, these two parameters exhibit strong interactions. Therefore, grid search (GS), which has been used either formally or informally for SVM parameter selection, was performed in this study (36). We ranged parameter γ and C from 0.005 to 0.9 with 0.005 as the increment, and from 10 to 1000 with 10 as the increment, respectively. Parameters γ , C , number of support vectors, and accuracy of the training set are shown in Table IV. It can be seen from the table that the three groups of γ and C led to the same accuracy of training set. The lowest number of support vectors prompted the selection of $\gamma = 0.215$ and $C = 50$ as the optimal values. Thus the optimal model was obtained and the predictive results of training set, test set, and external set are all listed in Table I. It gave an accuracy of 92.30% for Class 1, 89.19% for Class 2, 90.63% for training set, 90.00% for the test set, and 90.48% for the whole data set (see Fig. 1).

DISCUSSION

Figure 1 compares the predictive results obtained from LDA and SVM models. Performance-wise, SVM gave better results than LDA, which implies that, using the same descriptors, the SVM method is capable of recognizing nonlinear relationships; in contrast, LDA approaches can only capture linear relationships between molecular properties and descriptors. Thus, SVM method can be used as an effective tool in distinguishing the potential risk of drugs to nursing infants.

In addition, it is necessary to gain some insight into factors that are likely to relate to the transfer process of the drug compounds by interpreting the descriptors used in the study. $\log P$ has been widely used as a measure of hydrophobicity or lipophilicity, which is the ratio of a

chemical's concentration in the *n*-octanol phase to its concentration in the aqueous phase of a two-phase system at equilibrium (37). $\log P$ represents the hydrophobicity of molecules and reflects the ability of molecules to penetrate the biomembrane and reach the interacting sites. In most cases, it seemed evident that $\log P$ terms were used to access biological properties relevant to drug action, cellular uptake, metabolism, bioavailability, and toxicity. For the process of drug's transfer, lipophilicity is approximately correlated to passive transport across cell membranes and the ability of a compound to partition through a membrane. Drugs are expected to partition into milk in accordance with their lipid characteristics. High lipid solubility favors protein binding, reducing the amount of drug available for diffusion into milk. Therefore, as $\log P$ increases, $\log M/P$ decreases. In summary, $\log P$ played an important role and it was evident that very little real progress could be made in understanding the drug transfer process without considering hydrophobicity. Randic index (order 2) (χ^2), a topological descriptor, encodes the size, shape, and degree of branching of the compound, and also relates to the dispersion interaction among molecules (38). It is always used to predict if molecular cavities can be filled up with a candidate molecule. With the application of molecular graphics, the fit docking or intercalation of molecules into cavities in macromolecular simulations became an important consideration for a drug's protein binding. The maximum partial charge for a C atom [$PC_{\max}(C)$] and the minimum coulombic interaction for a C-C bond [$CI_{\min}(CC)$] belong to electrostatic descriptors, which reflect the characteristics of charge distribution in the molecule. The empirical partial charges in the molecule are calculated using the approach proposed by Zefirov (39). This method is based on the Sanderson electronegativity scale and uses the concept that represents molecular electronegativity as a geometric mean of atomic electronegativities. Min e-e repulsion for a C-C bond [$R_{ee}(CC)$] is one of the quantum chemical descriptors used to establish conformational stability, chemical reactivity, and intermolecular interactions. The descriptor characterizes the nuclear repulsion-driven processes in the molecule and may be related to the conformational (rotational, inversional) changes or atomic reactivity in the molecule. Energy was calculated for an optimized conformation with the most stable geometry, or minimum energy structure using molecular or quantum mechanics to determine bond strengths, atomic hybridizations, partial charges, and orbitals from the positions of the atoms and the net charge.

From the discussions above, it can be seen that steric and electric factors are likely two major components in the transfer process for drugs, and all the descriptors involved in the model, which have explicit physical meaning, may account for the structural features responsible for the M/P ratio of drug compounds.

CONCLUSION

In this work, linear discriminant analysis (LDA) and support vectors machine (SVM) were used for classifying the milk/plasma (M/P) concentration ratio of a set of 126 drug compounds using descriptors calculated from the molecular structure alone. The LDA model could provide some insight

into what structural features can best describe the process of drug transfer. The SVM method proved to be a highly effective classification tool because of its structural risk minimization principle, which minimizes an upper bound of the generalization error rather than the training error. This eventually leads to better generalization than neural networks, which implement the empirical risk minimization (ERM) principle and do not converge to global solutions.

Although the classification for M/P ratios is only one component of the complex process for a drug transfer to human milk, it probably forms the most selective filter. We would expect computational methods to have a similar effect on the search for other drugs whose M/P ratios are not currently available.

ACKNOWLEDGMENT

The authors thank the National Natural Science Foundation of China (NSFC) Fund (No. 20305008) for financial support.

REFERENCES

1. S. Kacew. Adverse effects of drugs and chemicals in breast milk on the nursing infant. *J. Clin. Pharmacol.* **33**:213–221 (1993).
2. C. R. Howard and R. A. Lawrence. Drugs and breastfeeding. *Clin. Perinatol.* **26**:447–478 (1999).
3. K. Yoshida, B. Smith, M. Craggs, and R. Kumar. Neuroleptic drugs in breast-milk: a study of pharmacokinetics and of possible adverse effects in breast-fed infants. *Psychol. Med.* **28**:81–91 (1998).
4. A. Lewellyn and Z. N. Stowe. Psychotropic medications in lactation. *J. Clin. Psychiatry* **59**:41–52 (1998).
5. S. Agatonovic-Kustrin, L. H. Ling, S. Y. Tham, and R. G. Alany. Molecular descriptors that influence the amount of drugs transfer into human breast milk. *J. Pharm. Biomed. Anal.* **29**:103–119 (2002).
6. M. S. Meskin and E. J. Lien. QSAR analysis of drug excretion into human breast milk. *J. Clin. Hosp. Pharm.* **10**:269–278 (1985).
7. S. Agatonovic-Kustrin, I. G. Tucker, M. Zecevic, and L. J. Zivanovic. Prediction of drug transfer into human milk from theoretically derived descriptors. *Anal. Chim. Acta* **418**:181–195 (2000).
8. P. Mazzatorta, E. Benfenati, P. Lorenzini, and M. Vighi. QSAR in ecotoxicity: an overview of modern classification techniques. *J. Chem. Inf. Comput. Sci.* **44**:105–112 (2004).
9. S. J. Patankar and P. C. Jurs. Classification of inhibitors of protein tyrosine phosphatase 1B using molecular structure based descriptors. *J. Chem. Inf. Comput. Sci.* **43**:885–899 (2003).
10. J. Souza Jr., F. A. Molfetta, K. M. Honorio, R. H. A. Santos, and A. B. F. da Silva. A study on the anticoronavirus activity of flavonoid compounds (Flavones) by using quantum chemical and chemometric methods. *J. Chem. Inf. Comput. Sci.* **44**:1153–1161 (2004).
11. M. Daszykowski, B. Walczak, Q. S. Xu, F. Daeyaert, M. R. de Jonge, J. Heeres, L. M. H. Koymans, P. J. Lewi, H. M. Vinkers, P. A. Janssen, and D. L. Massart. Classification and regression trees—studies of HIV reverse transcriptase inhibitors. *J. Chem. Inf. Comput. Sci.* **44**:716–726 (2004).
12. V. Vapnik. *Estimation of Dependencies Based on Empirical Data*. Springer, Berlin, 1982.
13. C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Disc.* **2**:121–167 (1998).
14. Y. L. Cun, L. Jackel, L. Bottou, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Müller, E. Säckinger, P. Y. Simard and V. Vapnik. *Learning algorithms for classification: a comparison on handwritten digit recognition, neural networks*. World Scientific, Singapore, 1995, pp. 261–276.
15. V. Blanz, B. Schölkopf, H. Bülthoff, C. Burges, V. Vapnik, and T. Vetter. In: C. V. D. Malsburg, W. V. Seelen, J. C. Vörbrüggen, B. Sendhoff (Eds): *Comparison of view-based object recognition algorithms using realistic 3D models*. Springer. Lect. Notes Comput. Sci. Bochum, 1996 **1112**: 251–256.
16. R. Burbidge, M. Trotter, B. Buxton, and S. Holden. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. Chem.* **26**:5–14 (2001).
17. Y. D. Cai, X. J. Liu, X. B. Xu, and K. C. Chou. Prediction of protein structural classes by support vector machines. *Comput. Chem.* **26**:293–296 (2002).
18. L. Bao and Z. R. Sun. Identifying genes related to drug anti-cancer mechanisms using support vector machine. *FEBS Lett.* **521**:109–114 (2002).
19. C. Y. Zhao, R. S. Zhang, H. X. Liu, C. X. Xue, S. G. Zhao, X. F. Zhou, M. C. Liu, and B. T. Fan. Diagnosing anorexia based on partial least squares, back propagation neural network, and support vector machines. *J. Chem. Inf. Comput. Sci.* **44**:2040–2046 (2004).
20. H. C. Atkinson and E. J. Begg. Prediction of drug distribution into human milk from physicochemical characteristics. *Clin. Pharmacokinet.* **18**:151–167 (1990).
21. W. B. White, J. W. Andreoli, S. H. Wong, and R. D. Cohn. Atenolol in human plasma and breast milk. *Obstet. Gynecol.* **63**:42–44 (1984).
22. L. J. Notarianni, D. Belk, S. A. Aird, and P. N. Bennett. An *in vitro* technique for rapid determination of drug entry into breast milk. *Br. J. Clin. Pharmacol.* **40**:333–337 (1995).
23. H. C. Atkinson and E. J. Begg. Relationship between human milk lipid-ultrafiltrate and octanol–water partition coefficients. *J. Pharm. Sci.* **77**:796–798 (1988).
24. H. C. Atkinson and E. J. Begg. Predictions of drug concentrations in human skim milk from plasma protein binding and acid–base characteristics. *Br. J. Clin. Pharmacol.* **25**:495–503 (1988).
25. E. J. Begg, H. C. Atkinson, and B. Duffull. Prospective evaluation of a model for the prediction of milk: plasma drug concentrations from physicochemical characteristics. *Br. J. Clin. Pharmacol.* **33**:501–505 (1992).
26. A. Tropsha, P. Gramatica, and V. K. Gombar. The importance on being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *Quant. Struct. Act. Relat. Comb. Sci.* **22**:69–76 (2003).
27. K. A. Dimitris, C. Walter, and S. Victor. On the use of neural network ensembles in QSAR and QSPR. *J. Chem. Inf. Comput. Sci.* **42**:903–911 (2002).
28. MDL Information Systems, Inc., ISIS Draw 2.3, <http://www.mdli.com> (accessed 1999–2000).
29. Hypercube, Inc. HyperChem 6.01, <http://www.hyper.com> (accessed 2000).
30. A. R. Katritzky, V. S. Lobanov and M. Karelson. *CODESSA: Training Manual*. University of Florida, Gainesville, FL, 1995.
31. A. R. Katritzky, V. S. Lobanov and M. Karelson. *CODESSA: Reference Manual*. University of Florida, Gainesville, FL, 1994.
32. S. K. Kachigan. *Statistical Analysis*. Radius Press, New York, 1986.
33. R. A. Fisher. The use of multiple measurements in axonomic problems. *Ann. Eugen.* **7**:179–188 (1936).
34. C. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1997.
35. W. J. Wang, Z. B. Xu, W. Z. Lu, and X. Y. Zhang. Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing* **55**:643–663 (2003).
36. F. Luan, C. X. Xue, R. S. Zhang, C. Y. Zhao, M. C. Liu, Z. D. Hu, and B. T. Fan. Prediction of retention time of a variety of volatile organic compounds based on the heuristic method and support vector machine. *Anal. Chim. Acta* **537**:101–110 (2005).
37. J. C. Dearden. Partitioning and lipophilicity in quantitative structure–activity relationships. *Environ. Health Perspect.* **61**:203–228 (1985).
38. A. Kovatcheva, G. Buchbauer, A. Golbraikh, and P. Wolschann. QSAR modeling of α -campholenic derivatives with sandalwood odor. *J. Chem. Inf. Comput. Sci.* **43**:259–266 (2003).
39. N. S. Zefirov, M. A. Kirpichenok, F. F. Izmailov, and M. I. Trofimov. Scheme for the calculation of the electronegativities of atoms in a molecule in the framework of Sanderson's Principle. *Dokl. Akad. Nauk SSSR.* **296**:883–887 (1987).