articles

molecular pharmaceutics

QSAR and Classification Study of 1,4-Dihydropyridine Calcium Channel Antagonists Based on Least Squares Support Vector Machines

Xiaojun Yao,*,† Huanxiang Liu,† Ruisheng Zhang,† Mancang Liu,† Zhide Hu,† A. Panaye,‡ J. P. Doucet,‡ and Botao Fan‡

Department of Chemistry, Lanzhou University, Lanzhou 730000, China, and Université Paris 7-Denis Diderot, ITODYS-CNRS UMR 7086, 1, rue Guy de la Brosse, 75005 Paris, France

Received April 16, 2005

Abstract: The least squares support vector machine (LSSVM), as a novel machine learning algorithm, was used to develop quantitative and classification models as a potential screening mechanism for a novel series of 1,4-dihydropyridine calcium channel antagonists for the first time. Each compound was represented by calculated structural descriptors that encode constitutional, topological, geometrical, electrostatic, quantum-chemical features. The heuristic method was then used to search the descriptor space and select the descriptors responsible for activity. Quantitative modeling results in a nonlinear, seven-descriptor model based on LSSVM with mean-square errors 0.2593, a predicted correlation coefficient (R^2) 0.8696, and a cross-validated correlation coefficient (R_{cv}^2) 0.8167. The best classification results are found using LSSVM: the percentage (%) of correct prediction based on leave one out cross-validation was 91.1%. This paper provides a new and effective method for drug design and screening.

Keywords: QSAR; calcium channel antagonists; least squares support vector machines

1. Introduction

The 1,4-dihydropyridine (DHP) derivatives are an important class of drugs known as calcium antagonists. These drugs act directly on the voltage-dependent calcium channels, localized in the cell membrane, blocking the flux of calcium from the extracellular medium to the cell cytoplasm.[1] Therefore, they can control the calcium-dependent biological events and treat cardiovascular diseases related to calcium channels and have been used in the treatment of a number of cardiovascular disorders such as variant and exertional angina, certain types of cardiac arrhythmias, hypertension, and others.[1] Structure−activity relationship (SAR) and quantitative structure−activity relationship (QSAR) studies on DHP derivatives could identify and provide some insight into what structural features are related to the biological activity of these compounds, provide some instruction for further designing the new calcium antagonists, and narrow the search for future drug compounds.

Structure−activity relationship (SAR) analysis is one of the techniques used to facilitate the search for new drugs. A successful solution to this problem has the potential to provide significant economic benefit via increased process efficiency.[2] QSAR involves modeling a continuous activity for quantitative prediction of the activity of previously unseen compounds. The advances in QSAR studies have widened the scope of rationalizing drug design and the search for the

* Corresponding author. Mailing address: Department of Chemistry, Lanzhou University, Lanzhou 730000, China. Tel: +86-931-891-2578. Fax: +86-931-891-2582. E-mail: xiaojunyao@yahoo.com.

† Lanzhou University.

‡ Université Paris 7-Denis Diderot.

(1) Costa, M. C. A.; Gaudio, A. C.; Takahata, Y. A comparative study of principal component and linear multiple regression analysis in SAR and QSAR applied to 1,4-dihydropyridine calcium channel antagonists (nifedipine analogues). *THEOCHEM* **1997**, *394*, 291−300.

(2) Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. Chem.* **2001**, *26*, 5−14.

10.1021/mp050027v CCC: $30.25
Published on Web 07/20/2005

mechanisms of drug actions. In addition, they are useful in areas such as design of virtual compound libraries, computational-chemical optimization of compounds, and design of combinatorial libraries with appropriate ADME (absorption, distribution, metabolism, and excretion) properties.

Among the investigations of SAR and QSAR, one of the most important factors affecting the quality of the model is the method of building the model. The support vector machine (SVM) is a popular algorithm developed from the machine learning community. Due to its advantages and remarkable generalization performance over many other methods, the SVM has attracted attention and gained extensive applications.[3−12] The reason for SVM's outstanding performance is that it can lead to global models that are often unique by embodying the structural risk minimization (SRM) principle,[13,14] which has been shown to be superior to the traditional empirical risk minimization (ERM) principle. Furthermore, due to their specific formulation, sparse solutions can be found, and both linear and nonlinear regression can be performed. However, finding the final SVM model can be computationally very difficult because it requires the solution of a set of nonlinear equations (quadratic programming). As a simplification of "traditional" SVM, Suykens and co-workers[15] have proposed the use of least squares SVM (LSSVM). LSSVM encompasses advantages similar to those of SVM, but its additional advantage is that it only requires solving a set of linear equations (linear programming), which is much easier and computationally very simple.

Another important factor responsible for the quality of the SAR and QSAR model is the numerical representation (often called molecular descriptor) of the chemical structure. The performance and the accuracy of the results are strongly dependent on the way the structures are represented. The software CODESSA, developed by the Katritzky group, enables the calculation of a large number of quantitative descriptors based solely on the molecular structural information and codes chemical information into mathematical form.[16,17] CODESSA combines diverse methods for quantifying the structural information about the molecule with advanced statistical analysis to establish molecular structure−property/activity relationships. CODESSA has been applied successfully in a variety of QSAR analyses.[18,19]

Several SAR and QSAR studies on DHP derivatives have been reported in recent years.[1,20−25] According to the method

(3) Belousov, A. I.; Verzakov, S. A.; Von Frese J. A flexible classification approach with optimal generalization performance: support vector machines. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 15−25.

(4) Morris, C. W.; Autret, A.; Boddy, L. Support vector machines for identifying organisms−a comparison with strongly partitioned radial basis function networks. *Ecol. Modell.* **2001,** *146*, 57−67.

(5) Liu, H. X.; Zhang, R. S.; Luan, F.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. Diagnosing breast cancer based on support vector machines. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 900−907.

(6) Liu, H. X.; Zhang, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. QSAR study of Ethyl 2-[(3-Methyl-2,5-dioxo(3-pyrrolinyl))-amino]-4-(trifluoromethyl)pyrimidine-5-carboxylate: An Inhibitor of AP-1 and NF-κB Mediated Gene Expression based on support vector machines. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1288−1296.

(7) Liu, H. X.; Zhang, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. Prediction of Isoelectric Point of Amino Acid Based on GA-PLS and SVMs. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 161−169.

(8) Xue, C. X.; Zhang, R. S.; Liu, H. X.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. An Accurate QSPR Study of O−H Bond Dissociation Energy in Substituted Phenols Based on Support Vector Machines.*J. Chem. Inf. Comput. Sci.* **2004**, *44*, 669−677.

(9) Liu, H. X.; Zhang, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. QSAR and Classification models of a novel series of COX-2 selective inhibitors: 1, 5-Diarylimidazoles based on support vector machines. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 389−399.

(10) Yao, X. J.; Panaye, A.; Doucet, J. P.; Zhang, R. S.; Chen, H. F.; Liu, M. C.; Hu, Z. D.; Fan, B. T. Comparative Study of QSAR/QSPR Correlations Using Support Vector Machines, Radial Basis Function Neural Networks and Multiple Linear Regression. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1257−1266.

(11) Yao, X. J.; Panaye, A.; Doucet, J. P.; Chen, H. F.; Zhang R. S.; Fan, B. T.; Liu, M. C.; Hu, Z. D. Comparative classification study of toxicity mechanisms using support vector machines and radial basis function neural networks. *Anal. Chim. Acta* **2005**, *535*, 259−273.

(12) Liu, H. X.; Xue, C. X.; Zhang, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. Quantitative Prediction of logk of Peptides in High-Performance Liquid Chromatography Based on Molecular Descriptors by Using the Heuristic Method and Support Vector Machine. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1979−1986.

(13) Burges, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowledge Discovery* **1998**, *2*, 1−47.

(14) Vapnik, V. *Estimation of Dependences Based on Empirical Data*; Springer: Berlin, 1982.

(15) Suykens, J. A. K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293−300.

(16) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Comprehensive Descriptors for Structural and Statistical Analysis, Reference Manual*, version 2.0; University of Florida: Gainesville, FL, 1994.

(17) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem. Soc. Rev.* **1995**, *24*, 279−287.

(18) Oblak, M.; Randic, M.; Solmajer, T. Quantitative Structure−Activity Relationship of Flavonoid Analogues. 3. Inhibition of $p56^{lck}$ Protein Tyrosine Kinase. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 994−1001.

(19) Katritzky, A. R.; Tatham, D. B. Theoretical Descriptors for the Correlation of Aquatic Toxicity of Environmental Pollutants by Quantitative Structure-Toxicity Relationships. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1162−1176.

(20) Zamponi, G. W.; Stotz, S. C.; Staples, R. J.; Andro, T. M.; Nelson, J. K.; Hulubei, V.; Blumenfeld, A.; Natale, N. R. Unique Structure−Activity Relationship for 4-Isoxazolyl-1,4-dihydropyridines. *J. Med. Chem.* **2003**, *46*, 87−96.

(21) Hemmateenejad, B.; Akhond, M.; Miri, R.; Shamsipur, M. Genetic Algorithm Applied to the Selection of Factors in Principal Component-Artificial Neural Networks: Application to QSAR Study of Calcium Channel Antagonist Activity of 1,4-Dihydropyridines (Nifedipine Analogous). *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1328−1334.

(22) Takahata, Y.; Costa, M. C. A.; Gaudio, A. C. Comparison between Neural Networks (NN) and Principal Component Analysis (PCA): Structure Activity Relationships of 1,4-Dihydropyridine Calcium Channel Antagonists (Nifedipine Analogues). *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 540−544.

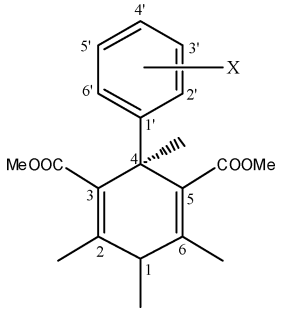
of building the model, these SAR and QSAR models can be divided into two classes, linear models and nonlinear models. Compared with these models, the nonlinear models provided more accurate or almost equivalent results. Compared with traditional neural networks, SVM possesses prominent advantages: (1) A strong theoretical background provides SVM with a high generalization capability and can avoid local minima. (2) SVM always has a solution, which can be quickly obtained by a standard algorithm (quadratic programming). (3) SVM need not determine network topology in advance, which can be automatically obtained when the training process ends. (4) SVM builds a result based on a sparse subset of training samples, which reduce the workload. LSSVM encompasses advantages similar to those of SVM, but its additional advantage is that it requires solving a set of only linear equations (linear programming), which is much easier and computationally very simple. As a novel machine learning algorithm, very few studies about the application of LSSVM in QSAR have been reported to this day.[26]

In the present investigation, LSSVM, as a novel machine learning technique, for the first time, was used to build the SAR and QSAR models of a series of 1,4-dihydropyridine calcium channel antagonists using calculated molecular descriptors from structure alone based on the software CODESSA as inputs. In addition, in order to identify the reliability of the LSSVM, other classification and regression methods such as the linear discriminant analysis (LDA) and the heuristic method (HM) were also used, respectively. The use of quantitative and classification models can augment and narrow the search for future drug compounds and, at the same time, gain some insight into the structural factors that are responsible for their activities. The prediction results were very satisfactory in both regression and classification analyses, which proved LSSVM a powerful and useful tool in drug design and discovery.

2. Computational Methods and Materials

2.1. Data Sets. The structures and biological activity values for 45 antagonists were taken from the literature[1] and are listed in Table 1. The biological activity was expressed

Table 1. Experimental and Predicted Activities ($-\log(IC_{50})$) of 1,4-Dihydropyridine Calcium Channel Antagonists



		$-\log(IC_{50})$		
no.	X	exptl	HM	LSSVM
1	3′-Br	8.89	7.45	7.45
2	2′-CF_3	8.82	8.55	8.57
3	2′-Cl	8.66	7.79	7.78
4	3′-NO_2	8.40	8.19	8.16
5	2′-CH=CH_2	8.35	7.69	7.67
6	2′-NO_2	8.29	8.29	8.28
7	2′-Me	8.22	7.99	7.99
8	2′-Et	8.19	8.27	8.23
9	2′-Br	8.12	7.65	7.65
10	2′-CN	7.80	8.33	8.30
11	3′-Cl	7.80	7.81	7.81
12	3′-F	7.68	7.98	7.97
13	H	7.68	7.25	7.25
14	3′-CN	7.46	7.19	7.18
15	3′-I	7.38	7.67	7.67
16	2′-F	7.37	8.52	8.48
17	2′-I	7.33	7.65	7.65
18	2′-OMe	7.24	6.91	6.92
19	3′-CF_3	7.13	7.84	7.84
20	3′-Me	6.96	7.11	7.11
21	2′-OEt	6.96	7.55	7.54
22	3′-OMe	6.72	7.25	7.25
23	3′-NMe_2	6.05	6.11	6.11
24	3′-OH	6.00	6.28	6.29
25	3′-NH_2	5.70	5.19	5.22
26	3′-OAc	5.22	5.26	5.29
27	3′-OCOPh	5.20	5.28	5.24
28	2′-NH_2	4.40	5.31	5.32
29	4′-F	6.89	6.35	6.36
30	4′-Br	5.40	5.28	5.32
31	4′-I	4.64	5.21	5.24
32	4′-NO_2	5.50	5.62	5.62
33	4′-NMe_2	4.00	3.37	3.41
34	4′-CN	5.46	5.67	5.68
35	4′-Cl	5.09	5.47	5.50
36	2′,6′-Cl_2	8.72	7.91	7.89
37	F_5	8.36	7.84	7.78
38	2′-F,6′-Cl	8.12	8.36	8.32
39	2′3′-Cl_2	7.72	7.36	7.36
40	2′-Cl,5′-NO_2	7.52	7.89	7.85
41	3′,5′-Cl_2	7.03	6.82	6.83
42	2′-OH, 5′-NO_2	7.00	6.93	6.90
43	2′,5′-Me_2	7.00	7.16	7.17
44	2′,4′-Cl_2	6.40	7.18	7.19
45	2′,4′,5′-$(OMe)_3$	3.00	3.12	3.17

(23) Viswanadhan, V. N.; Mueller, G. A.; Basak, S. C.; Weinstein, J. N. Comparison of a Neural Net-Based QSAR Algorithm (PCANN) with Hologram- and Multiple Linear Regression-Based QSAR Approaches: Application to 1,4-Dihydropyridine-Based Calcium Channel Antagonists. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 505−511.

(24) Hemmateenejad, B.; Miri, R.; Akhond, M.; Shamsipur, M. QSAR study of the calcium channel antagonist activity of some recently synthesized dihydropyridine derivatives. An application of genetic algorithm for variable selection in MLR and PLS methods. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 91−99.

(25) Schleifer, K.-J.; Tot, E. CoMFA, CoMSIA and GRID/GOLPE studies on calcium entry blocking1,4-dihydropyridines. *Quant. Struct.-Act. Relat.* **2002**, *21*, 239−248.

(26) Thissen, U.; Üstün, B.; Melssen, W. J.; Buydens, L. M. C. Multivariate Calibration with Least-Squares Support Vector Machines. *Anal. Chem.* **2004**, *76*, 3099−3105.

by IC_{50} (the molar concentration of the drug required to inhibit 50% of the contraction of guinea pig ileum induced by methyl-furmethide). For analysis purposes, $-\log(IC_{50})$ values were used as the dependent variables and are given in Table 1.

2.2. Descriptor Calculation and Feature Selection. All structures were drawn and preoptimized using the MM+ molecular mechanics method within the framework of the Hyperchem program.[27] The preoptimized structures were submitted to the MOPAC6.0 program[28] for further geometry refinement and for the calculation of molecular orbital parameters. The AM1 parametrizations were used to calculate the quantum-chemical molecular descriptors. The output files from MOPAC were transferred to the program CODESSA to calculate various descriptors. In CODESSA there are implemented procedures for the calculation of a large selection of molecular descriptors including a variety of constitutional, topological, geometric, and electrostatic descriptors.[16]

Once molecular descriptors were generated, the heuristic method[16] in CODESSA was used to accomplish the preselection of the descriptors and build the linear model. Its advantages are the high speed and lack of software restrictions on the size of the data set. The heuristic method can either quickly give a good estimation about what quality of correlation to expect from the data, or derive several best regression models. Besides, it will demonstrate which descriptors have bad or missing values, which descriptors are insignificant (from the standpoint of a single-parameter correlation), and which descriptors are highly intercorrelated. This information will be helpful in reducing the number of descriptors involved in the search for the best QSAR/QSPR model.

The heuristic method usually produces correlations 2−5 times faster than other methods with comparable quality.[29] The rapidity of calculations from the heuristic method renders it the first method of choice in practical research. Thus, in the present investigation, we used this method to select structural descriptors and build the linear model.

2.3. Least Squares Support Vector Machine. In recent years, the support vector machine (SVM), based on statistical learning theory, as a powerful new tool for data classification and function estimation, has been developed.[30] SVM maps input data into a high dimensional feature space where it may become linearly separable. Recently SVM has been applied to a wide variety of domains such as pattern recognition and object detection,[13] function estimation,[31] etc. One reason that SVM often performs better than earlier methods is that SVM was designed to minimize structural risk whereas previous techniques were usually based on minimization of empirical risk. So SVM is usually less vulnerable to the overfitting problem. Especially, Suykens and Vandewalle[15] proposed a modified version of SVM called least squares SVM (LSSVM), which resulted in a set of linear equations instead of a quadratic programming problem, which can extend the application of the SVM. To understand LSSVM well, here LSSVM for classification and regression was introduced first.

There exist a number of excellent introductions to SVM, both printed[31−33] and electronically available.[34] The theory of LSSVM has also been described clearly by Suykens et al.[15] For this reason, we will only briefly describe the main idea of LSSVM and the differences between SVM and LSSVM here.

2.3.1. LSSVM for Classification.[15,35,36] Consider a binary classification training sample $\{(x_i,y_i)\}_{i=1,2,...,l}$, where x_i is the vector of input pattern for the ith example and y_i is the corresponding target output. The pattern represented by the subset $y_i = +1$ belongs to class 1, and the pattern represented by the subset $y_i = -1$ belongs to class 2. The original SVM classifier satisfies the following conditions:

$$y_i[w^T\varphi(x)_i + b] \geq 1, \quad i = 1, ..., l \tag{1}$$

where φ: $R^n \rightarrow R^m$ is the feature map mapping the input space to a usually high dimensional feature space where the data points become linearly separable by a hyperplane defined by the pair ($w \in R^m, b \in R$). The classification function is then given by

$$y(x) = \text{sign}\{w^T\varphi(x) + b\} \tag{2}$$

It is usually unnecessary to compute with the feature map, and one only needs to work instead with a kernel function in the original space given by

$$K(x_i,x_j) = \varphi(x_i)^T\varphi(x_j) \tag{3}$$

In the case of noisy data, forcing zero training error will lead to poor generalization. To take account of the fact that some data points may be misclassified, introduce a set of slack variables:

(27) HyperChem. 4.0, Hypercube, 1994.
(28) Stewart, J.P. P. *MOPAC 6.0*; Quantum Chemistry Program Exchange; QCPE, No. 455; Indiana University: Bloomington, IN, 1989.
(29) Katritzky, A. R.; Petrukhin, R.; Jain, R.; Karelson, M. QSPR analysis of Flash points. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1521−1530.
(30) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273−297.
(31) Vapnik, V. *Statistical Learning Theory*; Wiley: New York, 1998.
(32) Schölkopf, B.; Burges, C.; Smola, A. *Advances in Kernel Methods—Support Vector Learning*; MIT Press: Cambridge, MA, 1999.
(33) Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines*; Cambridge University Press: Cambridge, U.K., 2000.
(34) URL: http://www.kernel-machines.org/, 2005.
(35) Viaene, S.; Baesens, B.; Van Gestel, T.; Suykens, J. A. K.; Van den Poel, D.; Vanthienen, J.; De Moor, B.; Dedene, G. Knowledge Discovery in a Direct Marketing Case using Least Squares Support Vector Machines. *Int. J. Intell. Syst.* **2001**, *16*, 1023−1036.
(36) Chua, K. S. Efficient computations for large least square support vector machine classifiers. *Pattern Recognit. Lett.* **2003**, *24*, 75−80.

$$\xi_i > 0, \quad i = 1, ..., l \tag{4}$$

The relaxed separation constraint is given as

$$y_i[w^T\varphi(x)_i + b] \geq 1 - \xi_i, \quad i = 1, ..., l \tag{5}$$

The optimal separating hyperplane can be found by the following minimization problem:

$$\underset{w,b,\xi}{\text{Min}}\, J(w,b) = \frac{1}{2}w^Tw + C\sum_{i=1}^{l}\xi_i \tag{6}$$

subject to the constraints

$$\begin{Bmatrix} y_i[w^T\varphi(x) + b] \geq 1 - \xi_i, & i = 1, ..., l \\ \xi_i \geq 0, & i = 1, ..., l \end{Bmatrix}$$

where C is a regularization parameter used to decide a tradeoff between the training error and the margin. The dual of system 6 via the Karush−Kuhn−Tucker (KKT) condition leads to a well-known convex quadratic programming (QP) problem.

Vapnik's standard SVM classifier formulation was modified by Suykens and Vandewalle into the following LSSVM formulation:

$$\underset{w,b,e}{\text{Min}}\, J(w,b) = \frac{1}{2}w^Tw + \gamma\frac{1}{2}\sum_{i=1}^{l}e_i^2 \tag{7}$$

subject to the equality constraint

$$y_i[w^T\varphi(x_i) + b] = 1 - e_i, \quad i = 1, ..., l$$

We note that the passage from eq 6 to eq 7 involves replacing the inequality constraints by equality constraints and a squared error term (hence least square) similar to ridge regression. The corresponding Lagrange for eq 7 is

$$L(w,b,e,\alpha) = J(w,e) - \sum_{i=1}^{l}\alpha_i\{y_i[w^T\varphi(x_i) + b] - 1 + e_i\} \tag{8}$$

where the α_i are Lagrange multipliers. As was shown in ref 15, the optimality condition leads to the following $(N + 1) \times (N + 1)$ linear system:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + \gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \tag{9}$$

where $Z = [\varphi(x_1)^Ty_1; ...; \varphi(x_l)^Ty_l]$, $Y = [y_1; ...; y_l]$, $\mathbf{1} = [1; ...; 1]$.

Mercer's condition is applied within the matrix ZZ^T:

$$ZZ^T = y_iy_j\varphi(x_i)^T\varphi(x_j) = y_iy_jK(x_i,x_j)$$

Thus, we would only need to use kernel function K in the training algorithm, and would never need to explicitly even know what φ is. The LSSVM classifier is then constructed as follows:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{l}y_i\alpha_iK(\mathbf{x},\mathbf{x}_i) + b\right) \tag{10}$$

In this investigation, for both the classification and the regression task, the radial basis function kernel was used.

2.3.2. LSSVM for Function Estimation.[37,38] For the function estimation problem, given a training data set of l points $\{(x_i,y_i)\}_{i=1,2,...,l}$ with input data $x_i \in R^n$ and output data $y_i \in R$, one considers the following optimization problem in primal weight space:

$$\underset{w,b,e}{\text{Min}}\, J(w,b) = \frac{1}{2}w^Tw + \gamma\frac{1}{2}\sum_{i=1}^{l}e_i^2 \tag{11}$$

subject to the equality constraint

$$y_i = w^T\varphi(x_i) + b + e_i, \quad i = 1, ..., l$$

By comparing eq 11 with eq 7, it can be seen that the only difference between the classification and function estimation problem is the constraint. This difference results in the different Lagrange style for function estimation,

$$L(w,b,e,\alpha) = J(w,e) - \sum_{i=1}^{l}\alpha_i\{w^T\varphi(x_i) + b + e_i + y_i\} \tag{12}$$

The resulting LSSVM model for function estimation becomes

$$f(x) = \sum_{i=1}^{l}\alpha_iK(x,x_i) + b \tag{13}$$

2.3.3. LSSVM Implementation and Computation Environment. In contrast to the Lagrange multipliers, the choice of a kernel and its specific parameters together with γ do not follow from the optimization problem but have to be tuned by the user. These can be optimized by the use of Vapnik−Chervonenkis bounds, cross-validation, an independent optimization set, or Bayesian learning. In this paper, the radial basis function kernel was used as kernel function and cross-validation was used to tune the optimized values of the two parameters γ and σ. For the regression problem, the MSE was used as an error function, and it is computed according to the following equation:

$$\text{MSE} = \frac{\sum_{i=1}^{n}(d_i - o_i)^2}{n}$$

where d_i are the teaching outputs (desired outputs), o_i are the actual outputs, and n is the number of samples.

(37) Suykens, J. A. K.; Vandewalle, J. Chaos control using least-squares support vector machines. *Int. J. Circ. Theor. Appl.* **1999**, *27*, 605−615.

(38) Suykens, J. A. K.; Vandewalle, J.; De Moor, B. Optimal control by least squares support vector machines. *Neural Networks* **2001**, *14*, 23−35.

Table 2. The Linear Model between the Selected Structural Descriptors and Activity ($R^2 = 0.8695$, $R_{cv}^2 = 0.8158$, $F = 35.21$, $s^2 = 0.3082$)

descriptor	coefficient	error	*T*-test value
intercept	2.1551×10	5.0943×10	0.4268
min atomic orbital electronic population	6.6260×10	1.0412×10	6.3637
complementary information content (order 2)	-5.1385×10^{-2}	1.1632×10^{-2}	−4.4174
min n−n repulsion for a C−C bond	-6.3817×10^{-1}	3.8700×10^{-1}	4.2593
HOMO−LUMO energy gap	2.7623	6.4853×10^{-1}	−4.7316
YZ Shadow	-1.1207×10^{-1}	2.3685×10^{-2}	−4.7316
Randic index (order 3)	8.6952×10^{-1}	1.9337×10^{-1}	4.4966
min partial charge for a H atom [Zefirov's PC]	-2.0960×10^{2}	7.8678×10	−2.6640

For the classification problem, the percentage of misclassification was used as an evaluation function.

All calculations implementing LSSVM were performed using the Matlab/C toolbox[39] on a Pentium IV with 256M RAM.

3. Results and Discussion

3.1. Heuristic Method Model. Through the heuristic method in CODESSA, the best linear model with seven parameters was obtained, and it is shown in Table 2. By interpreting of the descriptors in the regression model, it is possible to gain some insight into factors that are likely to govern the activity of these compounds under the same mechanism of interaction.

In the linear model, there are two topological descriptors, one geometric descriptor, three quantum chemical descriptors, and one electrostatic descriptor. The descriptors *Randic index (order 3)* and *complementary information content (order 2)* belong to topological descriptors. The *Randic index*[40] is calculated as a sum of atomic connectivities over molecular paths of certain length (1, 2, ..., *n*); it thus reflects molecular size and branching. The second topological descriptor, *complementary information content (order 2)*,[41] is defined on the basis of the Shannon information theory. Complementary information content in different levels can be calculated as follows:

$$^{k}\mathrm{CIC} = \log_2 (n - {}^{k}\overline{\mathrm{IC}})$$

$${}^{k}\overline{\mathrm{IC}} = -\sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n}$$

where n_i is a number of atoms in the *i*th class and *n* is the total number of atoms in the molecule. The division of atoms into different classes depends on the coordination sphere taken into account. This leads to the indices of different order *k*. In the second level, the atom set is decomposed into equivalence classes using their chemical nature and bonding pattern up to the second-order bonded neighbors. This descriptor can express the internal flexibility of the molecule. In this investigation, a geometric descriptor, YZ Shadow, was involved. It can be defined as the area of shadows of the molecule as projected on the *YZ* plane by the orientation of the molecule in the space along the axes of inertia, which characterizes the size and geometrical shape of the molecule.[16] The two topological descriptors and one geometric descriptor encode the structural characteristics related to connectivity, complexity and shape, proving that the hydrophobic and steric interactions are very important for binding between the antagonists and receptor.

The involved quantum chemical descriptors were *min atomic orbital electronic population*, *HOMO−LUMO energy gap* (eLUMO − eHOMO), and *min n−n repulsion for a C−C bond*. The descriptor *min atomic orbital electronic population*[16] describes the nucleophilicity of the molecule. HOMO and LUMO are the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital, respectively, and eLUMO and eHOMO are the energies of these orbitals, respectively. HOMO−LUMO energy gap is an approximate estimate of the first electron excitation energy in the UV/visible spectra of the molecule or of the band gap between valence and empty zone in solids.[16] Min n−n repulsion for a C−C bond[16] is the minimum nuclear repulsion energy for a C−C bond. Because the main weight atoms are C atoms, this energy describes the nuclear repulsion driven processes in the molecule and may be related to the conformational changes or atomic reactivity in the molecule. These three descriptors indicate the importance of the intramolecular electronic effects and reactivity of a molecule in determining the binding ability between the molecule and receptor.

The last descriptor, min partial charge for an H atom [Zefirov's PC],[16] is an electrostatic descriptor, which is calculated using the approach proposed by Zefirov. This method is based on the Sanderson electronegativity scale and uses the concept that represents the molecular electronegativity as a geometric mean of atomic electronegativities. This

(39) Pelckmans, K.; Suykens, J. A. K.; Van Gestel, T.; De Brabanter, D.; Lukas, L.; Hamers, B.; De Moor, B.; Vandewalle, J. *LS-SVMlab: a Matlab/C Toolbox for Least Squares Support Vector Machines*; Internal Report 02-44, ESATSISTA.; K. U. Leuven: Leuven, 2002.

(40) Katritzky, A. R.; Oliferenko, A. A.; Oliferenko, P. V.; Petrukhin, R.; Tatham, D. B. A General Treatment of Solubility. 1. The QSPR Correlation of Solvation Free Energies of Single Solutes in Series of Solvents. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1794−1805.

(41) Basak, S. C.; Balaban, A. T.; Grunwald, G. D.; Gute, B. D. Topological Indices: Their Nature and Mutual Relatedness. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 891−898.

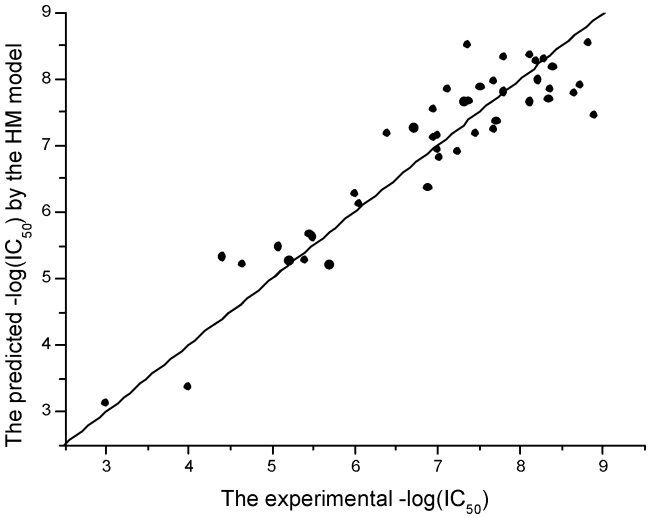


Figure 1. The predicted values of $-\log(IC_{50})$ versus the experimental data by the HM model.

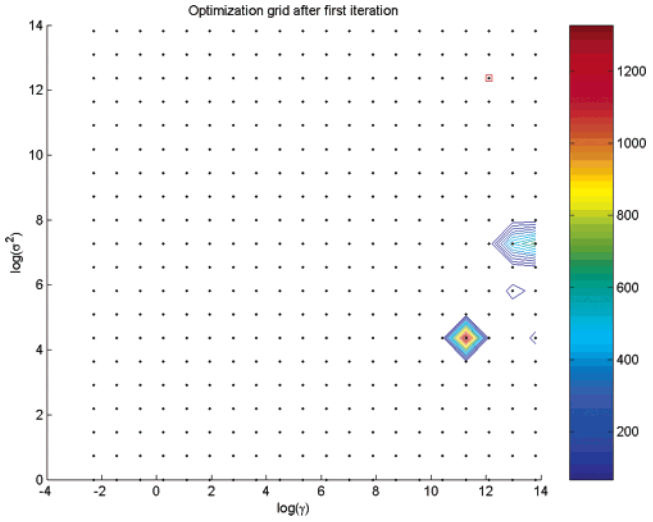


Figure 2. Contour plot of the optimization error for LSSVM when optimizing the parameters σ and γ in the regression problem. The red square indicates the selected optimal settings.

descriptor can encode the hydrogen bond forming ability of a molecule with receptor. In addition, it also can account for the electrostatic interaction between the drug and receptor.

Analysis of the results obtained indicates that the selected molecular descriptors calculated solely from structures can describe the structural features of the compounds responsible for their biological activity. The calculated and experimental values of $-\log(IC_{50})$ of the DHP derivatives by the heuristic method are given in Table 1, and the scatter plot is shown in Figure 1. Table 2 gives the statistical parameters of the heuristic method model: $R^2 = 0.8695$; $F = 35.2096$; $s^2 = 0.3082$.

3.2. Support Vector Regression Model. 3.2.1. Optimizing LSSVM. As discussed in the section 2.3.3, kernel function and its specific parameters together with γ have to be tuned by the user. In this paper, the radial basis function (RBF) kernel was used as the kernel function. Thus γ (the relative weight of the regression error) and σ (the kernel parameter of the RBF kernel) need to be optimized. Here, the optimal parameters are found by an intensive grid search. The result of this grid search is an error surface spanned by the model parameters. A robust model is obtained by selecting those parameters that give the lowest error in a smooth area.

To find the optimal model parameters, a grid search is performed on the basis of leave one out (LOO) cross-validation on the data set. The parameter (σ) of the RBF kernel in the style of σ^2 and the parameter γ were tuned simultaneously in a 20×20 grid ranging from 1 to 1 000 000 and from 0.1 to 1 000 000, respectively. In this way, parameter optimization was performed in different orders of magnitude. Because the grid search has been performed over just two parameters, a contour plot of the optimization error can be visualized easily (Figure 2). This is an advantage of LSSVMs over SVMs in which three parameters have to be optimized. From Figure 2, the optimal parameter settings can now be selected from a smooth subarea with a low prediction error. The selected optimal values of γ and σ^2 are 1.83×10^5 and 2.34×10^5, respectively, which point was marked by the red square in Figure 2.

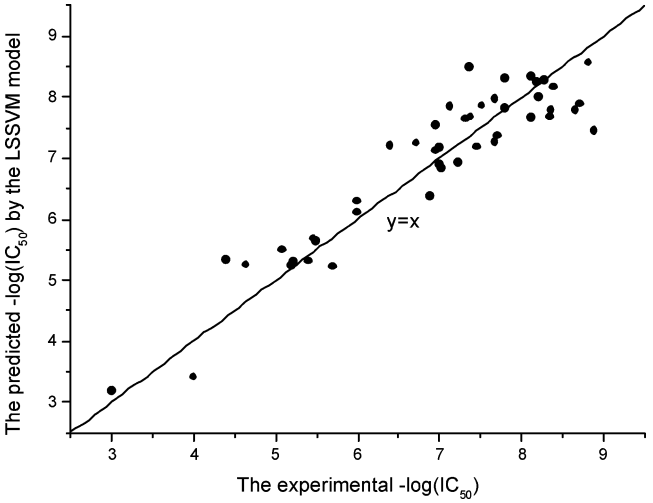


Figure 3. The predicted values of $-\log(IC_{50})$ versus the experimental data by the LSSVM model

3.2.2. Predicted Result of LSSVM. The predicted results of the optimal LSSVM model ($\sigma^2 = 2.34 \times 10^5$, $\gamma = 1.83 \times 10^5$) are shown in Table 1 and Figure 3. The mean square error is 0.0226, and the prediction correlation coefficient (R^2) and cross-validated correlation coefficient (R_{cv}^2) based on LOO cross-validation are 0.8696 and 0.8167, respectively. For the biological activity data of drug with high noise, it can be concluded that the predicted values are in good agreement with the experimental values from the above results. By comparing the results from HM and LSSVM, it can be seen that the performance of the LSSVM model is almost equal to that of the HM model.

3.3. Classification Model. In this investigation, in order to compare with the previous work conveniently, the same strategies for separating the high/low active groups and

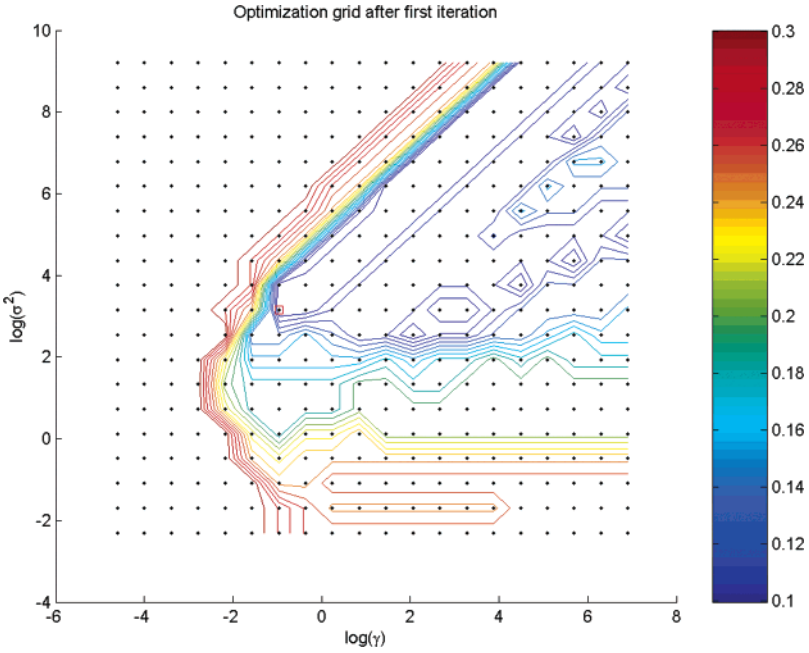


Figure 4. Contour plot of the optimization error for LSSVM when optimizing the parameters σ and γ for the classification problem. The red square indicates the selected optimal settings.

expressing the classification results were adopted. The compounds whose $-\log(IC_{50})$ values are greater than or equal to 6.72 were grouped as high active, and others were grouped as low active. This rule resulted in compounds **1**−**22**, **29**, and **36**−**33** falling into the high active group and compounds **23**−**28**, **30**−**35**, **44**, and **45** belonging to the low active group. The classification results were given in two types, RECALL and LOO (leave one out). In RECALL, the classifier is asked to recall the data used for training the classifier with the training pattern. In LOO, the classifier predicts the activity of a compound from the model developed from the rest of the compounds. This procedure is repeated for each compound in the data set.

In addition, to present the classification results in a more familiar way, we expressed the classification results of different models as ROC (receiver operating characteristic) curve. The ROC curve shows the separation abilities of a binary classifier: by iteratively setting the possible classifier thresholds, the dataset is tested on misclassifications. As a result, a plot is shown. If the plot has a surface of 1 on test data, a perfectly separating classifier is found (on that particular dataset); if the area equals 0.5, the classifier has no discriminative power at all. An ROC curve is a graphical representation of the trade-off between the false negative and false positive rates for every possible cutoff. Equivalently, the ROC curve is the representation of the trade-offs between sensitivity and specificity. By tradition, the plot shows the false positive rate on the *X* axis and 1 − the false negative rate on the *Y* axis.

3.3.1. LDA Model. Linear discriminant analysis is an algorithmically simple and useful classification technique for situations where you want to build a predictive model of group membership based on 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.

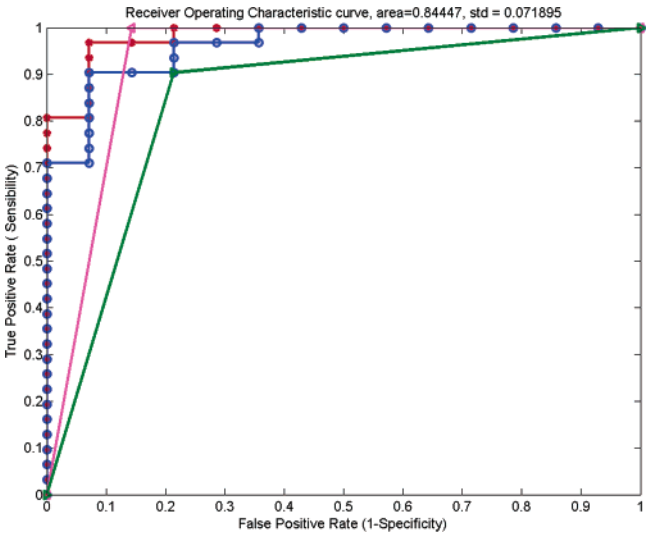


Figure 5. Receiver operating characteristic (ROC) curves.

In this work, the percentages (%) of correct prediction of RECALLand LOO based on LDA were 95.6% and 86.7%, respectively. The ROC curve for RECALL and LOO is given in Figure 5. In Figure 5, the purple line with left triangle was obtained in the RECALL process. The green line with right triangle was based on LOO cross-validation. The areas under the purple line and the green line are 0.9286 and 0.8446, respectively.

Table 3. Comparisons of Different Correlation Models

approach	R^2	R_{cv}^2
ref 23 (MLR)	0.550	0.190
ref 23 (PCANN)	0.730	0.550
ref 25 (CoMFA)	0.872	0.600
ref 25 (CoMSIA)	0.908	0.744
ref 25 (GRID/GOLPE)	0.821	0.620
HM	0.870	0.816
LSSVM	0.870	0.817

3.3.2. LSSVM Classification Model. Similarly to the LSSVM regression, to find the optimal model parameters and prohibit the overfitting of the model, grid searches for the parameter γ (the relative weight of the regression error) and σ (the kernel parameter of the RBF kernel) based on LOO cross-validation of the training set were performed. For the parameters σ and γ, first, wide ranges 0.01−1000 of γ and 0.1−10000 of σ^2 were tuned in a 20 × 20 grid, respectively. The contour plot of the optimization error is given in Figure 4. In Figure 4, the selected optimal values of σ^2 and γ, 0.379269 and 23.3572, are marked by the red square. The corresponding percentages (%) of correct prediction of RECALL and LOO were 93.3% and 91.1%, respectively.

The LSSVM classification results are also expressed as ROC (Figure 5). In Figure 5, the red line with circles and the blue line with circles account for the RECALL and LOO results of LSSVM. The area is respectively 0.9816 and 0.9608 under the red and blue ROC curves. From the ROC curves of LDA and LSSVM, it can be seen that the LSSVM classifier has a stronger separating ability than the LDA classifier.

3.4. Comparison of the Results Obtained by Different Approaches. To test the suitability of the correlation and classification models constructed by LSSVM, we have compared the correlation and classification results obtained by this investigation with those obtained in refs 22, 23, and 25. Table 3 and Table 4 showed the statistical parameters of the results obtained from the four studies for correlation and classification models, respectively, based on the same set of compounds. Generally, the R_{cv}^2 can reflect the quality of a QSAR model more reliably. From Table 3, it can be seen that the HM and LSSVM models in this investigation are most predictive. Among the classification models, in ref 22, PCA (principal components analysis) was used to classify the high/low active compounds. Because PCA is a kind of unsupervised classification technique, this study cannot give the percentage (%) of correct prediction of RECALL and LOO. In addition, generally, the LOO accuracy can reflect the real performance of a classification model. Thus, from Table 3, it can be seen that this LSSVM approach currently constitutes the most accurate method to classify the DHP derivatives into the high/low groups.

Table 4. Comparisons of Different Classification Models

approach	RECALL (%)	LOO (%)	others (%)
ref 22 (PCA)			82.2
ref 23 (BPNN)	88.9	88.9	
LDA	95.6	86.7	
LSSVM	93.3	91.1	

4. Conclusions

The above results indicate that LSSVM is a very promising tool both for function estimation and for classification. The LSSVM, like the conventional SVM, exhibits better overall performance due to embodying the structural risk minimization principle and some advantages over the other techniques of converging to the global optimum, and not to a local optimum. Besides, its additional advantage is that it requires solving a set of only linear equations (linear programming), which is much easier and computationally very simple. The predictive results are satisfied not only for regression but also for classification. Therefore it is a good approach for predicting the expected activity of drugs and aiding drug design. At the same time, the models proposed could identify and provide some insight into what structural features are related to the biological activity of these compounds and provide some instruction for further designing the new highly active 1,4-dihydropyridine calcium channel antagonists.

MP050027V