

Prediction of placenta barrier permeability and reproductive toxicity of compounds in tocolytic Chinese herbs using support vector machine

Fang Lu¹, Yusu He¹, Ludi Jiang, Jiahua Chen, Yilian Cai, Yanling Zhang*

Key Laboratory of TCM-information Engineer of State Administration of TCM, School of Chinese Material Medica, Beijing University of Chinese Medicine, Beijing, 100102, China

e-mail: collean_zhang@163.com (Yanling Zhang)

Keywords: Tocolytic Chinese herbs; Support vector machine; Placenta barrier permeability; Reproductive toxicity; Miscarriage

Abstract. 93 compounds which can permeate the placenta barrier were collected as data set for the construction of support vector regression (SVR) model. Besides, 140 compounds with reproductive toxicity and 170 compounds with no reproductive toxicity were collected as another data set for the construction of support vector classification (SVC) model. 1481 molecular descriptors were calculated to represent the structure characteristics of all the compounds mentioned above by Dragon2.1. CfsSubsetEval valuation method and BestFirst-D1-N5 searching method were used to optimize the subset of molecular descriptors. Then based on the above data, SVR model for prediction the placenta barrier permeability (PBP) and SVC model for prediction the reproductive toxicity were built respectively by using LibSVM program. Both the SVR model and the SVC model obtained better prediction ability. The correlation coefficient (R^2) values of the training set and test set of the optimal SVR model were 0.990 and 0.780. The accuracy, sensitivity, and specificity values of the optimal SVC model were all above 80%. Subsequently, the SVR model was utilized to predict the PBP of the compounds which were collected from 13 commonly used tocolytic Chinese herbs. The compounds with higher permeability were further studied by the SVC model and 15 compounds were classified as positive compounds with reproductive toxicity. The two models constructed in this study might be employed in guiding the application of the tocolytic Chinese herbs in clinical.

Introduction

Traditional Chinese medicine (TCM) has been regarded as the main medical for the treatment of diseases throughout East Asia with a history of 5000 years. It has been widely used to promote health and treat illnesses since then and accepted as a major approach of complementary and alternative medicine in Western world in recent years. According to the theory of TCM [1], miscarriage is defined as “fetal irritability”, “fetal restlessness” or “stirring fetus”, and TCM has unique herbs to reduce the risk of miscarriage.

However, the information about the application of tocolytic Chinese herbs for miscarriage is limited [2] [3]. Besides, tocolytic Chinese herbs may lead toxic effects to the fetus by transferring across the placenta barrier. Therefore it is urgent to investigate whether the active compounds with high placenta barrier permeability (PBP) are reproductive toxicity or not, for the safety of tocolytic Chinese herbs application.

Support vector machine (SVM), of which the contents involved the aspects of the maximum-margin hyperplane, Mercer's kernels and so on, serves as a new artificial intelligence technology [4]. In practice, it consists of two modes: SVC and SVR. The former is a qualitative method, which is focus on the pattern recognition and classification; the latter is a quantitative method for prediction and forecasting [4]. Meanwhile, the time-consuming and high-cost problem caused by traditional toxicological experiments can be alleviated by SVM [5]. In this paper, a drug molecule-based SVR model was built for evaluating the PBP of tocolytic Chinese herbs compounds which derived from the Traditional Chinese Medicine Database (TCMD, version 2009). Then on this basis, a drug molecule-based SVC model was constructed to predict the reproductive toxicity of the active compounds with relatively high PBP. Thus, the SVR and SVC models of tocolytic

compounds were constructed to discover the compounds with reproductive toxicity from Chinese herbs, which may provide guidance for the safety application of tocolytic Chinese herbs in clinical.

Materials and methods

Training and test sets splitting. The PBP of compounds is typically expressed in relative permeability (RP). Antipyrine, a small lipophilic compound that is known to be transported across the placenta via passive diffusion [6], was used as a reference compound to make the data gathered in multiple laboratories comparable.

Then, for the generation of SVR model, 93 compounds that have been reported to have PBP were collected as SVR data set from references [7-12]. To ensure the training set were well representative, the “Kennard-Stone (KS)” program [13] was used to randomly split the SVR data set into a training set containing 73 compounds and a test set with 20 compounds (namely SVR-TS1).

On the other hand, for the generation of SVC model, a data set including 140 reproductive toxicity compounds (positive) and 170 non-reproductive toxicity compounds (negative) were collected from references [14-17]. The data set was also split by KS. 125 positives and 125 negatives were chosen as training set, and 15 positives and 45 negatives were regarded as internal test set (namely SVC-TS1). Then, to further evaluate the SVC models, we constructed another two independent external test sets, namely SVC-TS2 and SVC-TS3. To be specific, SVC-TS2 was comprised of 455 negatives, which were derived from the “approved” drug molecules in DrugBank (<http://www.drugbank.ca/>); SVC-TS3 was comprised of 5 Chinese herbs that had been reported to bear reproductive toxicity [18-22]. There is no overlap between each set mentioned above, and the results of splitting training and test sets are detailed in Table 1.

Table 1. The data structure partitioning results for each model.

The PBP SVR model		The reproductive toxicity SVC model					
Training set	SVR-TS1	Training set+	Training set-	SVC-TS1+	SVC-TS1-	SVC-TS2-	SVC-TS3+
73	20	125	125	15	45	455	5

(+) positive compounds. (-) negative compounds.

Molecular descriptors. Dragon2.1 was used for descriptors calculation for all compounds in all data sets and 1481 molecular descriptors were computed. For feature selection, BestFirst-D1-N5, CfsSubsetEval and 10-fold cross-validation were performed in Weka3.6.10 [23]. Three data normalization methods, including no-normalization, [0, 1] and [-1, 1], were performed for each object. Then, those three data processing results for each study object were used for further study, respectively.

Development of SVM. SVM is an important machine learning method and is rigorously used in both bioinformatics and chemoinformatics. In this study, LibSVM-Faruto-Ultimate (Version2.0), developed by Faruto, was used to run the SVM algorithm, including SVC and SVR. Appropriate kernel function and parameters are crucial and should be chosen while using SVM to settle factual classification problem. In order to map input vectors into a higher dimensional space nonlinearly, RBF Kernel was chosen as the kernel function of both SVC and SVR.

Besides, C and γ are two parameters in RBF kernel, which greatly affect the precision of SVM classifier. In order to ensure the classifier could predict test set accurately, three parameter optimization methods, called Grid Search, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), were used to identify appropriate (C, γ). In addition, the default parameters were also retained for model construction. Thus, 4 parameter selection methods were used to build SVM models based on 3 data processing results.

Data analysis and model validation. Firstly, SVR-TS1 was applied to validate the SVR models so that the relationship between structural variables and placenta barrier RP could be

predicted accurately. The evaluation indicators were presented as follows: the correlation coefficient (R^2) and mean square error (MSE) [8]. A SVR model with a R^2 value of greater than 0.73 and a MSE score of less than 0.15 was preferable [8]. Then, three test sets, including SVC-TS1, SVC-TS2 and SVC-TS3, were employed to assess the performances of the established SVC models. The following evaluation criteria were calculated: accuracy (ACC), sensitivity (SE), and specificity (SP). Computational formulas of three indicators were shown in Equation (1-3):

$$ACC = (TP + TN) / (TP + FN + TN + FP) \quad (1)$$

$$SE = TP / (TP + FN) \quad (2)$$

$$SP = TN / (TN + FP) \quad (3)$$

From the Equation (1-3), TP, TN, FN and FP are the number of true positives, true negatives, false negatives, and false positives, respectively.

Application of the optimal SVR model and SVC model. On the basis of TCMD and the Chinese Pharmacopoeia 2010 [24], the optimal SVR model was employed for evaluating the tocolytic Chinese herbs, which contains 336 natural compounds from 13 commonly used tocolytic Chinese herbs. As antipyrine was utilized as a reference compound, the tocolytic Chinese herbs with the predicted RP values of greater than 0.70 were regarded as relatively high permeability compounds and selected to be further analyzed by the optimal SVC model of reproductive toxicity.

Results and discussion

Molecular descriptors selection. Molecular descriptors of each model were selected by using two feature selection algorithms, namely BestFirst and CfsSubsetEval, in Weak program package. Thus, 70 molecular descriptors were selected for the generation of SVR models, and 17 molecular descriptors were chosen for building the SVC models.

Model generation and validation. 9 SVR models and 12 SVC models were built for each study object based on the selected molecular descriptors. Then, the test sets were used to validate the corresponding models. And the model, which obtained higher evaluation indexes in test set, was selected as the optimal model for further study. [-1, 1] normalization method was applied to normalized data for both SVR and SVC model and the corresponding parameter optimization method was PSO and Grid Search respectively. The specific information of the optimal models for each research object and the validation results were list in Table 2 and 3 respectively. It demonstrated that those models provided a reliable utility for the prediction of compound PBP and reproductive toxicity.

Table 2. The performances of the optimal placenta barrier permeability SVR model.

Data normalization	Parameter optimization			$R^2_{\text{training set}}$	MES _{training set}	$R^2_{\text{SVR-TS1}}$	MES _{SVR-TS1}
[-1, 1]	Method	C	γ	0.990	9.741e-005	0.780	0.011
	PSO	60.172	0.094				

Table 3. The performances of the optimal reproductive toxicity SVC model.

Data normalization	Parameter optimization				ACC			SE	SP
[-1, 1]	Method	C	γ	Training set	SVC-TS1	SVC-TS2	SVC-TS3	SVC-TS1	SVC-TS1
	Grid Search	0.574	0.189	100%	100%	100%	100%	100%	100%

Database search. By using the optimal PBP SVR model, 15 compounds were obtained of which the RP values ≥ 0.70 . Then, the obtained structures were further analyzed by using the optimal reproductive toxicity SVC model. Finally, the 15 compounds all had potential reproductive

toxicity. Also, two compounds from *Fructus Amomi* and *Atractylodes macrocephala* respectively that have higher placenta barrier permeability. *Atractylodes macrocephala* had also been reported with potential reproductive toxicity in pregnant animals within the clinical dose [25]. Most of experimental results were consisted with exiting reference, which suggest that this model can be widely used in prediction of reproduction toxicity of Chinese herbs. The results of SVM-based prediction were list in Table 4.

Table 4. The results of SVR-based prediction.

Tocolytic Chinese herbs	compounds	CAS	Placenta barrier permeability
<i>Artemisia argyi</i>	Moxartenone	182162-13-6	0.74
<i>Artemisia argyi</i>	Stigmasterol	83-48-7	0.71
<i>Atractylodes macrocephala</i>	8 β -Ethoxy atractylenolide III	-	0.76
<i>Atractylodes macrocephala</i>	(+)-Eudesma-4(15),7(11)-dien-8-one	-	0.75
<i>Atractylodes macrocephala</i>	Juniper camphor	473-04-1	0.81
<i>Taxillus chinensis</i>	Lupeol	545-47-1	0.7
<i>Fructus Amomi</i>	Aframodial	115795-58-9	0.82
<i>Fructus Amomi</i>	Bornyl acetate	76-49-3	0.78
<i>Inula helenium</i>	6- α -Acetyl-4-O-oxobedfordiaic acid	-	0.76
<i>Inula helenium</i>	6- α -Acetyl-4-O-oxobedfordiaic methyl ester	-	0.77
<i>Inula helenium</i>	Alantolactone	546-43-0	0.7
<i>Perilla frutescens</i>	Collinsogenin	52936-64-8	0.71
<i>Perilla frutescens</i>	Dihydroperilla alcohol	-	0.75
<i>Perilla frutescens</i>	Perilloside C	146763-94-2	0.72
<i>Perilla frutescens</i>	Perillylglucopyranoside	-	0.71

Conclusion

Placenta barrier, which exists between the mother and her fetus, is an important line of defense against a wide variety of extraneous chemicals which are derived from maternal medication and lifestyle factors (like smoking and drinking). However, as described in early researches, all drug molecules may pass through placenta barrier with different PBP [26] [27]. Accordingly, we built a SVR model to evaluate the PBP of tocolytic herbs compounds.

Giaginis et al. firstly built a quantitative PBP model for prediction based on a multivariate data analysis approach [8]. However, as the correlation coefficient of test set was less than 0.63, Giaginis' model was not good enough to estimate the PBP of compounds. As a result, a better SVR model, of which the correlation coefficients of training set and test set were 0.990 and 0.780 respectively, was constructed in this study, so as to enhance the accuracy of prediction.

Tocolytic Chinese herbs may have beneficial effect on increasing live birth rate and reducing the risk of miscarriage. However, due to the mechanism is not clearly demonstrated among herbal interventions, the safety usage of Chinese herbs for the treatment of miscarriage is urgent to be evaluated. In this paper, the combination of SVC and SVR was firstly used to predict the compounds with reproduction toxicity from tocolytic Chinese herbs. As no specific safety problem was reported in early researches, our findings with potential adverse events should be further investigated. Finally, information gained from this study can be further employed in researching potential natural compounds from tocolytic Chinese herbs.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (No. 81173522) and the Joint Construction Project of the Beijing Municipal Commission of Education.

References

- [1] B. Z. Ma, R. F. Liu, and H. L. Du. *Zhong Yi Fu Ke Xue*, vol. 8, The Science & Technology Press, Shanghai, China, 2006.
- [2] Li, L., Leung, P. C., Chung, T. K. H., Wang, C. C. Systematic Review of Chinese Medicine for Miscarriage during Early Pregnancy , *J. Evid. Based. Complement. Alternat. Med.* 2014, 2014: 753856.
- [3] Yang, G. Y., Luo, H., Liao, X., Liu, J. P. Chinese herbal medicine for the treatment of recurrent miscarriage: a systematic review of randomized clinical trials , *J. BMC Complementary and Alternative Medicine.* 2013,13:320.
- [4] Cortes, C., Vapnik, V. Support-vector networks , *J. MACH. LEARN.* 1995, 20:273-297.
- [5] Jiang, L., He, Y., Zhang, Y. Prediction of hepatotoxicity of traditional Chinese medicine compounds by support vector machine approach , *J. Systems Biology (ISB)*, 2014 8th International Conference on. IEEE. 2014: 27-30.
- [6] Hewitt, M., Madden, J.C., Rowe, P. H., Cronin, M. T. Structure-based modelling in reproductive toxicology: (Q)SARs for the placental barrier , *J. SAR. QSAR. Environ. Res.* 2007, 18:57-76.
- [7] Festag, M., Viertel, B., Steinberg, P., Sehner, C. An in vitro embryotoxicity assay based on the disturbance of the differentiation of murine embryonic stem cells into endothelial cells. II. Testing of compounds , *J. Toxicol. In. Vitro.* 2007, 21(8):1631-40.
- [8] Giaginis, C., Zira, A., Theocharis, S., Tsantili-Kakoulidou, A. Application of quantitative structure-activity relationships for modeling drug and chemical transport across the human placenta barrier: a multivariate data analysis approach , *J. J. Appl. Toxicol.* 2009, 29(8):724-33.
- [9] Scholz, G., Pohl, I., Genschow, E., Klemm, M., Spielmann, H. Embryotoxicity screening using embryonic stem cells in vitro: correlation to in vivo teratogenicity , *J. Cells. Tissues. Organs.* 1999, 165(3-4):203-11.
- [10] Selderslaghs, I. W., Blust, R., Witters, H. E. Feasibility study of the zebrafish assay as an alternative method to screen for developmental toxicity and embryotoxicity using a training set of 27 compounds , *J. Reprod. Toxicol.* 2012, 33(2):142-54.
- [11] Hewitt, M., Madden, J. C., Rowe, P. H., Cronin, M. T. Structure-based modelling in reproductive toxicology: (Q)SARs for the placental barrier , *J. SAR. QSAR. Environ. Res.* 2007, 18(1-2):57-76.
- [12] Li, W., Zhou, J. Yang, S. A prediction model of placenta barrier penetration of drug molecules based on support vector regression , *J. Chem. Res. Appl.* 2011, 23(7):872-876.
- [13] Saptoro, A., Tadé, M. O., & Vuthaluru, H. A modified Kennard-Stone Algorithm for optimal division of data for developing artificial neural network models , *J. Chem. Prod. Process. Model.* 2012, 7(1).
- [14] Festag, M., Sehner, C., Steinberg, P., & Viertel, B. An in vitro embryotoxicity assay based on the disturbance of the differentiation of murine embryonic stem cells into endothelial cells. I: Establishment of the differentiation protocol , *J. Toxicol. In. Vitro.* 2007, 21(8):1619-1630.

- [15] Tin'kov, O. V., Muratov, E. N., Artemenko, A. G., & Kuz'min, V. E. Analysis and Prediction of the Reproductive Toxicity of Organic Compounds of Different Classes using 2D Simplex Representations of Molecular Structure , J. PHARM. CHEM. J. 2013, 47(8):426-432.
- [16] Selderslaghs, I. W., Van Rompay, A. R., De Coen, W., & Witters, H. E. Development of a screening assay to identify teratogenic and embryotoxic chemicals using the zebrafish embryo , J. Reprod. Toxicol. 2009, 28(3):308-320.
- [17] Martin, M. T., Knudsen, T. B., Reif, D. M., Houck, K. A., Judson, R. S., Kavlock, R. J., & Dix, D. J. Predictive model of rat reproductive toxicity from ToxCast high throughput screening, J. Biol. Reprod. 2011, 85(2):327-39.
- [18] Zhou, Y. Research Progress on Toxic Ingredients of Chinese Herbal Medicine with Reproductive Toxicity, J. Chinese. J. Exp. Tradit. Med. Formulae. 2009, 15(8):97-99.
- [19] Hou, X. Alkaloids Constitutents Research and Toxicity Evaluation of *Oxytropis falcate* Bunge , J. Dissertation, Northwest A&F University, Yangling, Shanxi, China, 2008.
- [20] Zhang, J. Study on reproductive toxicity of aconitum to male rats in vivo in vitro, J. Dissertation, Sichuan University, Sichuan, China, 2007.
- [21] Ma, S., Kou, J., Yu, B. Recent Researches on the Toxicity of Saponins from Plants, J. Progress in Pharmaceutical Sciences, 2012, 36(3):110-115.
- [22] Zhang, J. Investigation on methods of male reproductive toxicity induced by Chinese traditional medicine , J. Dissertation, Shanghai Institute of Pharmaceutical Industry, Shanghai, China, 2006.
- [23] Azuaje, F. Review of" Data Mining: Practical Machine Learning Tools and Techniques" by Witten and Frank, J. Biomed. Eng. Online. 2006: 51.
- [24] Chinese Pharmacopoeia Commission. Chinese pharmacopoeia, M. Chemical industry press, Beijing, 2010, 328:547.
- [25] L.Li, L.Y.Tang, Gene C.W.Man, Benson H.Y.Yeung, Clara B.S.Lau, P.C.Leung, and C.C.Wang. Potential reproductive toxicity of Largehead Atractylodes Rhizome, the most commonly used Chinese medicine for threatened miscarriage , J. Human reproduction. 2011, (0):1-9.
- [26] Younglai, E. V., Wu, Y. J., Foster, W. G. Reproductive toxicology of environmental toxicants: emerging issues and concerns, J. Curr. Pharm. Des. 2007, 13:3005-19.
- [27] Myllynen, P., Pasanen, M., Pelkonen, O. Human placenta: a human organ for developmental toxicology research and biomonitoring , J. Placenta. 2005, 26:361-71.