

# The Potential Performance of Artificial Neural Networks in QSTRs for Predicting Ecotoxicity of Environmental Pollutants

Ryo Shoji\*

Department of Chemical Science and Engineering, Tokyo National College of Technology, 1220-2 Kunugida-machi, Hachioji-city, Tokyo 193-0997, Japan

**Abstract:** This review surveys the applications of neural network methodologies to the field of Quantitative Structure-Toxicity Relationships (QSTRs) in environment, and more specifically ecotoxicity. QSTR is one of the methods for predicting hazards of various chemicals and utilizes a computer-based technology such as artificial neural network to predict the toxicity of a chemical solely from its molecular attributes. Many artificial neural network methodologies have been applied to ecotoxicological data for fish, bacteria, protozoa and so on. The results demonstrate the ability of the artificial neural network methodologies to apply non-linear structure-toxicity relationships for the prediction of the corresponding toxicity values for chemicals, which are not part of the training sets. In order to employ an artificial neural network for QSTR, although users must pay attention to over-parameterization, data distribution, the structure and training cycle of neural network, and chance correlation, fine tuned neural network has high performance to predict ecotoxicity of chemicals. In the most of the QSTR studies, the results by artificial neural network modeling gave clearly better prediction of toxicity values compared to the results by multiple linear regression analysis or other commercial QSTR programs.

**Keywords:** Artificial neural networks, ecotoxicity, quantitative structure-toxicity relationships.

## 1. INTRODUCTION

A branch of Chemistry of a great interest nowadays is computer-aided toxicity prediction of chemicals synthesized by human activities [1]. The possibility of predicting toxicity of the chemicals with well-defined properties while avoiding the adverse effects on environment has led to a great effort in basic research [2]. The fundamentals for an effective prediction of toxicity are the so-called quantitative structure-toxicity relationships (QSTRs), a discipline which has become rationalized and systematized very recently. QSTR techniques assume that a relationship between the toxicity of a chemical and its structure exists, and tries to establish simple mathematical relationships to describe – and later, to predict – a given property for a set of chemicals, usually belonging to the same chemical family [3]. QSTR analysis encompasses both the definition of chemical descriptors able to characterize satisfactorily different chemical sets and the statistical treatment, which can be applied to these descriptors in order to improve their predictive capacity. The importance of this subject has led to the apparition of specialized journals (*Quantitative Structure-Activity Relationships*; *Journal of Computer-Aided Molecular Design*; *Journal of Molecular Modeling*; *SAR and QSAR in Environmental Research*; *Current Computer-Aided Drug Design*; etc.).

Recently, environmental water, such as landfill leachate, has been polluted with a number of chemicals produced by various human activities [4]. Chemical analyses do not take into account the overall toxicity caused by multi-component

intakes or possible changes in toxicity caused by chemical interactions and biological metabolisms. Among the chemicals available on the market, few are tested for safety [5]. Animal tests to obtain data on chemical safety generally take too long time and cost too much money to become standard procedure [6]. Bioassays using several ecotoxicity tests are considered to be a rational and promising methodology for evaluating the effect on ecosystem [7]. Large amounts of ecotoxicity data for environmental water or pollutants have been collected by intra-laboratory works [8].

Lack of experimental toxicity data calls for the development of predictive models based on parameters that can be calculated directly from chemical structures and properties [9]. Because prediction of chemical safety by QSTR is considered meaningful for screening, several QSTR prediction systems have been successfully used in the past decade [10]. QSTR methods based on multiple linear regression analysis are often been in the literature [11]. Among the most promising methods is TOPKAT (Toxicity Prediction by Komputer Assisted Technology; Health Designs Inc., USA), which was developed by Enslein *et al.* [12], and is based on multiple linear regression analysis between carcinogenicity data and chemical properties. CASE (Computer Automated Structure Evaluation) is also a promising method because this system can estimate the effect of the body's metabolism with a MULTICASE (Multiple Computer Automated Structure Evaluation) system [13]. COMPACT (Computer-Optimized Molecular Comparative Analysis of Chemical Toxicity), based on the affinity of cytochrome P450 to the Ah receptor, also estimates carcinogenicity quite well [14]. FALS (Fuzzy Adaptive Least Squares) is a novel QSTR system [15]. An example of a methodology for predicting ecotoxicity is the US EPA's Ecological Structure Activity Relationships (ECOSAR) toxicity estimation program [16], which has

\*Address correspondence to this author at the Department of Chemical Science and Engineering, Tokyo National College of Technology, 1220-2 Kunugida-machi, Hachioji-city, Tokyo, 193-0997, Japan; Tel.: +81-426-68-5076; Fax: +81-426-68-5099; E-mail: shoji@tokyo-ct.ac.jp

been criticized for lack of statistical significance [17]. Thus, many QSTR systems have been developed to estimate the ecotoxicity of chemicals, and with an agreement between predicted and experimental values ( $R^2 > 0.70-0.99$ ) [18]. However, this too suffers from the restrictions imposed by the models on which it is based.

Various statistical methods that can be applied to QSTR are listed in Fig. (1). Historically, the most popular method used to QSTR is regression analysis among the various statistical methods [3]. Recently, newer methods have been used to perform QSTRs. Descriptors are statistically selected and attention is devoted mainly to maximize prediction power of the method, though there are hundreds of descriptors that can be employed in such QSTR. A common approach is to generate a number of descriptors and then use an appropriate statistical technique [19]. For all statistical methods, users must be aware that many models are 'black box' models. Only where a mode of action is known or assumed, the relationship within the QSTR will be based on causality rather than correlation. In these latter cases, any conclusions based on the model can be defined by the original datasets where there are a limited number of data and a large variability in the data [19].

The target relationship using multi-linear regression analysis can generate satisfactory results only for very small and narrow chemicals [20]. For the case of large and highly diverse data sets such an approach is unrealistic and an assumption of non-linearity is the natural alternative. On a common basis, because of the complexity between chemical properties and toxicities, it is impossible to verify the mathematical assumptions imposed by most multivariate non-linear statistical methodologies. At least, the linear association between chemical or physical observables and toxicity is not unique [3]. Consequently, a new approach and new tools are needed for the modeling exercise. Therefore, artificial neural networks, which are widely used in various QSAR (Quantitative Structure-Activity Relationship) analyses, offer the most affordable alternative in terms of toxicity prediction.

## 2. ARTIFICIAL NEURAL NETWORK

A widely used alternative to multiple linear regression analysis is the artificial neural network. A brief overview of the artificial neural networks is exposed here. Artificial neural networks constitute a non-linear approach usually

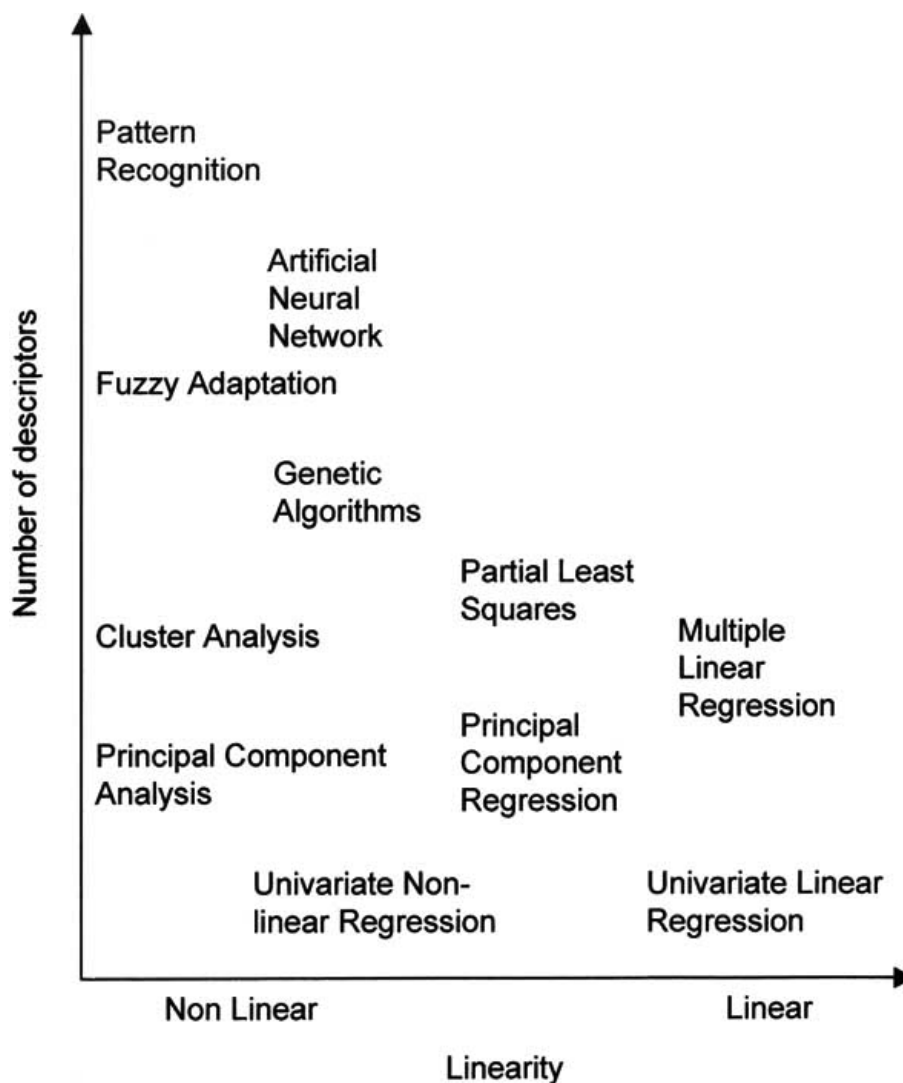
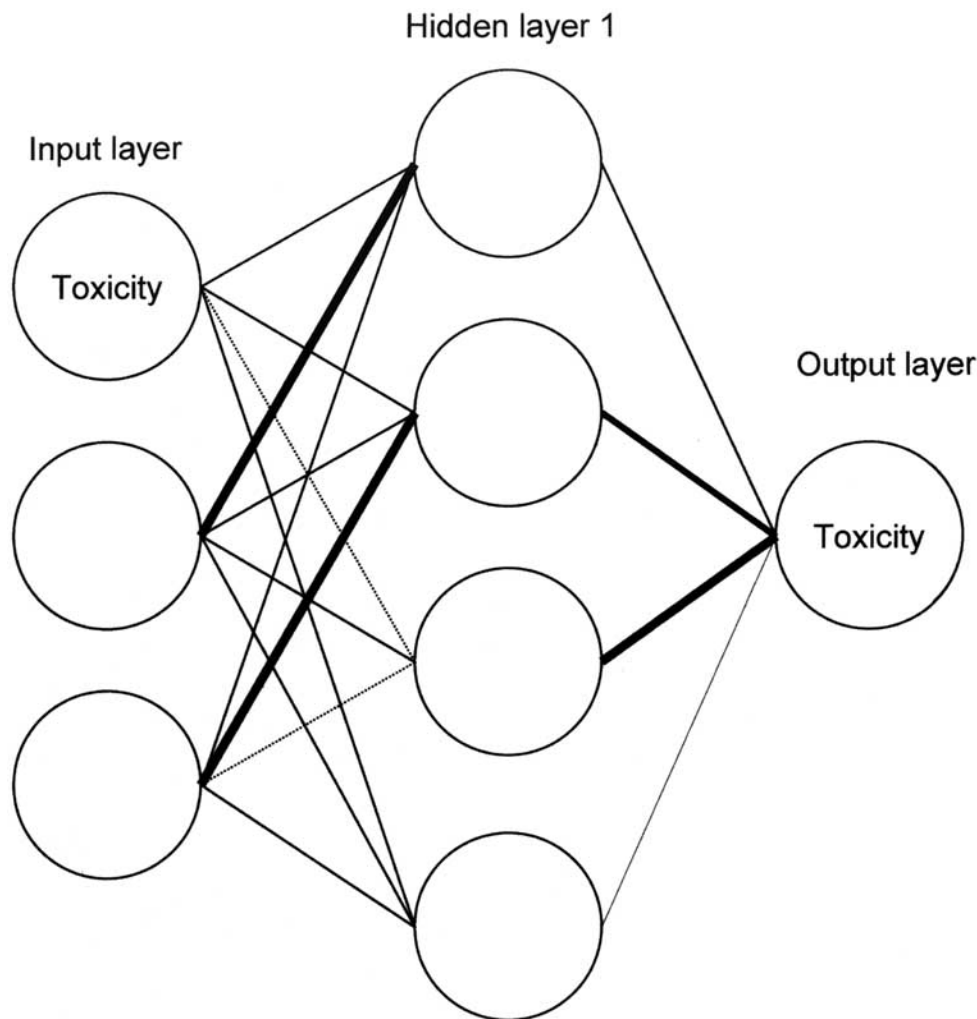


Fig. (1). Statistical methods for QSTRs.



**Fig. (2).** Conceptual diagram of artificial neural network (Four-layer 3-5-4-1 Neural Network).

(An artificial neural network is a computational device consisting of a group of processing elements (neurons) organized in subgroup (layers). Each subgroup may make its independent computations and may pass the results to yet another group.)

employed in QSTR analysis [21]. Fig. (2) shows the conceptual diagram of artificial neural network that tries to simulate the neurological processing of the human brain. A neural network is made up of different layers, divided into three types: an input layer, hidden layers and an output layer. A non-linear transference function is defined from the input data, weighted by coefficients that are adjusted according to the resultant output.

The network adapts as follows: change the weight by an amount proportional to the difference between desired and actual outputs as an equation:

$$\Delta W_i = \eta * (D-Y).I_i$$

where  $\eta$  is the learning rate,  $\Delta$  is the desired output, and  $Y$  is the actual output. This is called the *Perceptron Learning Rule*, and goes back to the early 1960's [22].

The learning of the neural network is supervised through the minimization of an error function defined from the difference between desired and obtained values. The process is repeated by back-propagation until self-consistency is attained. The major advantage of the artificial neural network is its ability to learn from data requiring principal

knowledge of domain problems. In addition, the artificial neural network is very effective in dealing with a large amount of toxicological data [23]. Recently, there has been a significant progress in both biotechnology and chemical engineering in developing the so-called software sensors and analyzers using available measurements to estimate the hard-to-measure variables [24].

The process for developing QSTR based on artificial neural network is illustrated in Fig. (3). The goal is to use specific descriptors and endpoints to develop an understanding of the domain of toxic chemicals, to determine potency, and to distinguish toxic from safe chemicals. QSTRs should have a well-defined and measurable endpoint and should be developed on a diverse dataset, one with a training set that defines the chemical domain to be modeled, and a separate test set for validation [24].

A number of QSTRs based on neural networks have been developed [20, 25]. Highly relevant models have been developed for the prediction of toxicity of chemicals to aquatic species [20]. A recent independent evaluation of several available programs and methods, including ASTER,

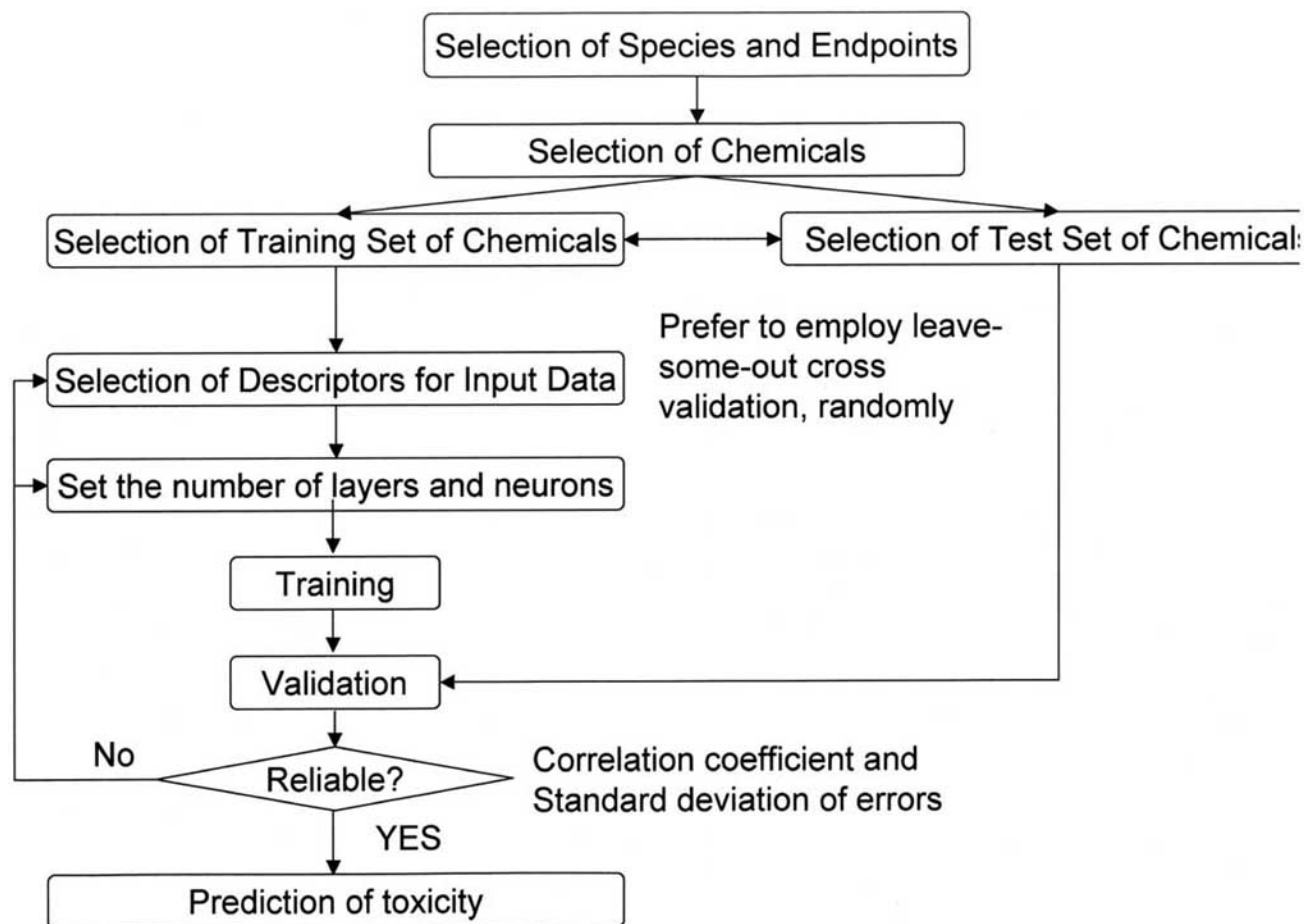


Fig. (3). Flow sheet of developing QSTR based on artificial neural network.

Computational Neural Network (CNN), Optimized Approach based on Structural Indices Set (OASIS), ECOlogical Structure Activity Relationship (ECOSAR), Toxicity Prediction by Komputer Assisted Technology (TOPKAT) and Probabilistic Neural Network (PNN) for the prediction of toxicity of chemicals to fish showed that artificial neural network had the best overall model performance [26]. This may have occurred because the artificial neural network training set had better coverage of the chemical domain of the test set than the other QSTR models.

It should also be mentioned that there has been criticism on the artificial neural network approach, particularly from regulatory agencies, as to their ‘lack of transparency’ and perception as ‘black box’ [23]. In addition, it should be noted that there are very different types of artificial neural networks including those that do and those that do not require iterative optimization of the learning phase for example determination of the optimum structure of artificial neural network to maximize the prediction capability/learning ratio [27]. However, when using identical inputs with proper optimization, artificial neural networks will give reproducible results based on defined mathematical relationship [28]. Generally, QSTR has the following three disadvantages:

1. False correlations may arise through too heavy a reliance being placed on biological data, which, by its nature, is subject to considerable experimental error.
2. Frequently, experiments upon which QSTR analyses depend lack design in the strict sense of experimental design. Therefore the data collected may not reflect the complete property space. Consequently, many QSTR results cannot be used to confidently predict the most likely chemicals of best toxicity.
3. Various physicochemical parameters are known to be cross-correlated. Therefore only variables or their combinations that have little covariance should be used in a QSTR analysis; similar considerations apply when correlations are sought for different sets of biological data.

In order to overcome these disadvantages of QSTR, users must pay attention to the following five points when QSTR is developed by using artificial neural network:

1. Over-parameterization
2. Data distribution
3. Kinds of neural networks

4. Structure and training
5. Chance correlations

### 3. POSSIBLE PROBLEMS OF NEURAL NETWORK PREDICTION SYSTEM

#### 3-1. Effect of Over-Parameterization on Model Robustness

It is important to notice that the artificial neural network models are data intensive because they are trained and not programmed [29]. Therefore, when an artificial neural network model is to be developed, it is important to divide the data into training and test sets so that the robustness of the developed model can be validated.

The most usual way to explore the robustness of a predictive model is through the analysis of the influence of each one of the individual objects that configure the final model in terms of the validation of model performance. In order to validate the model performance, one object of the set is extracted, and the model is recalculated using as training set the remaining  $n-1$  objects. The property is then predicted for the removed element. This process should be repeated for all the objects of the set, obtaining a prediction for every one. This procedure is known as cross-validation by leave-one-out [30]. Obviously, this process can be extended to a larger set of removed elements, yielding the leave-one-out, leave-three-out... or general leave-many-out procedures [31]. These are even more decisive for robustness assessment. Nevertheless, they are rarely reported within current QSTR results.

#### 3-2. Data Distribution

Generally, the model performance increases with an increase in the number of model inputs [22, 27, 31]. It is very important to get an adequate scaling of the experimental data. Most of the toxicities are measured in terms of concentrations or lethal doses, and they can vary within a wide range [8]. Using a logarithmic transformation, which is often employed to reduce the huge differences between data values, usually scales such property types [32]. A smooth and regular property distribution of toxicity values is necessary to get reliable results. Otherwise, if the property is made up of isolated islands of data points, some fitting problems can arise. Thus, apparently satisfactory results can be obtained simply whether the neural network prediction passes through these islands, even if the intercluster differences are not described. These apparently good results can easily mask a poor estimation of the toxicity.

#### 3-3. Kinds of Neural Networks

There are various neural network models for QSTR mapping: (i) back-propagation neural network (BNN) based on minimizing distances and maximizing correlations; (ii) probabilistic neural network (PNN) based on recognizing the distributional characteristics of the population; and (iii) Kohoren neural network (KNN) used to handle classifications based on mode of action [33], and other (iv) computational neural network (CNN). Among various types of artificial neural networks, BNN is probably the most

widely studied neural network. Fine tuned BNN may have an excellent performance, though other neural networks provide a faster approach [34]. A PNN model for the same problem, using the same data pre-processing algorithms, and validated through a complete 20% leave-out cross validation experiment and random selection, is reported in the literature [35]. Particularly, PNN can provide a faster approach though QSTR based on PNN may have some limited application for predicting the relative toxicity of closely related chemicals that possess a common mechanism of action.

#### 3-4. Structure and Training

Before the network can be trained, the network structure (the number of hidden layers and the number of neurons per hidden layer) has to be selected in addition to certain internal network parameters, including the learning rate and momentum, which also have to be selected. Maier and Dandy found that changing the network parameters, learning rate and momentum, has a marked effect on training speed, but does not significantly affect the predicting performance of the model [36]. On the other hand, number of hidden layers and neurons per hidden layer are generally determined by trial and error.

The weight change rule is a development of the perceptron learning rule. The weights are changed by an amount proportional to the error at that unit times the output of the unit feeding into the weight. Running the network consists of

forward pass:

The outputs are calculated and the error at the output units calculated.

backward pass:

The output unit error is used to alter the weights on the output units. Then the error at the hidden nodes is calculated (by *back-propagating* the error at the output units through the weights), and the weights on the hidden nodes altered using these values.

For each data pair to be learned, a forward and backwards passes are performed. This is repeated over and over again until the error is at a low enough level.

#### 3-5. Chance Correlation

In any predictive model, excellent adjustments and even satisfactory predictions can be achieved without the existence of a real relationship between the molecular structure description and the property of the studied set. A method capable to distinguish between a real structure-activity correlation and a chance description is then needed in every case [37]. Although the nature of QSTR based on artificial neural network is such that a causal relationship always may be supposed to exist, via the fundamental equation, in order to evidence the existence of fortuitous correlations, the randomization test must be adopted [38]. If statistically significant models are achieved when using the randomly ordered properties, the model must be considered suspect of random correlation. It probably correlates any data set, due to an excess of either parameters or degrees of freedom [22]. Nevertheless, a detailed analysis is necessary in those cases

where good models are found out for wrong properties, since the routine of generation of permuted vectors might have created an altered vector very similar to the real one.

The usual form to represent the results of a randomization test is by plotting a two-dimensional figure using the correlation and prediction coefficients as axes. Two different geometrical symbols are used to differentiate those points that represent the real property and those that correspond to the permuted ones.

#### 4. EVALUATION OF ROBUSTNESS OF ARTIFICIAL NEURAL NETWORK FOR ECOTOXICITY PREDICTION OF CHEMICALS

Table 1 gives the statistical information of various neural network models for prediction of ecotoxicity. The first attempt to predict the ecotoxicity of chemicals to fish using artificial neural network has been reported by Kaiser *et al.* [39]. The model used a BNN with fifty-one neurons in the input layer, seven neurons in hidden layer, and one neuron in the output layer (51-7-1). The learning involved 400 training cycles, the optimal moment of stopping the training being detected using a complete 20% leave out cross validation experiment using random selection. The correlation between measured and predicted values was over 0.9 with an average square error of 0.333. It gives the explicit system of equations describing a trained neural network based on real data which is a clear proof that neural networks are not black boxes.

It is possible to expand the scope of the toxicity modeling exercise by including model also information

which reflects biological response characteristics of the organisms under study in addition to the information on specific test conditions. There has been such a promising model, specifically on the acute toxicity of pesticides to fish [40]. Devillers and Flatin investigates the acute toxicity to the rainbow trout using detailed information of toxicity experiment for 70 pesticides. It is based on a data set consisting of LC<sub>50</sub> together with the associated bioassay information such as fish weight, exposure time, water temperature and so on. For modeling, the data set was divided into a training set and a test set. The training set grouped all the information concerning 384 of 447 available endpoints, and the test set grouped the information related to the remaining 63 endpoints. The developed model consists of a three-layer 13-6-1 BNN. Additional to the bioassay information, the input to the model includes eight autocorrelation descriptors designed from the hydrogen-suppressed graphs of the molecules. Linear interval compression transforms were applied for the data pre-processing. A small data set grouping the same information for seven additional endpoints from chemicals categorized as industrials was used to validate the training. Computed from the values, the standard deviation of the errors was 0.355 for the training set, and 0.328 for the test set [40].

A QSTR neural network model for the toxicity of organophosphorus insecticides to the larvae is presented in the literature [42]. The data set consists of 180 measured EC<sub>50</sub>, together with experimental information on sediment presence and water temperature and pH. The data set was divided into a training set of 164 records and a test set of 16 records. The identified model was a 10-5-1 BNN, and the

**Table 1. Statistical Data of Various Artificial Neural Network Model for Prediction of Ecotoxicity**

Prediction method	Network Structure	Number of chemicals	Species	Correlation coefficient	Standard deviation of errors	Reference
BNN	51-7-1	419	Fish	0.916	0.596	[39]
BNN	13-6-1	70	Fish	-	0.355	[40]
BNN	10-5-1	180	Invertebrates	-	0.189	[42]
BNN	13-6-1	66	Fish	-	0.352	[44]
BNN	2-3-1	20	Bacteria	-	-	[43]
BNN	5-10-1	604	Bacteria	-	0.216	[48]
BNN	36-26-1	240	Bacteria	-	0.133	[49]
KNN	60-5-5	283	Fish	-	-	[41]
PNN	-	776	Daphnia	0.875	0.558	[43]
ECOSAR	-	776	Daphnia	0.322	1.362	[16]
PNN	-	419	Fish	0.932	0.550	[39]
PNN	-	865	Fish	0.932	0.591	[17]
PNN	-	825	Protozoa	0.936	0.264	[46]
PNN	-	419	Bacteria	0.919	0.360	[17]
CNN	8-6-1	375	Fish	0.866	0.722	[20]
CNN	11-5-1	448	Protozoa	0.550	0.599	[47]

data pre-processing consisted of a classical min/max transformation, and the training was stopped after 500 learning cycles. The standard deviation of the errors produced by the model was 0.104 for the training set, and 0.189 for the test set. There are also various PNN models for the toxicity to *Daphnia magna* [43]. The data set consists of EC<sub>50</sub> values for 776 organic chemicals. The training set was selected by a random selection of 700 of the 776 chemicals. Complete 20% leave-out cross validation experiments based on random selection is reported. On the 700 chemical training set, the Pearson's correlation coefficient between measured and predicted values was 0.875 with an average error of practically zero, a standard deviation of errors of 0.558. Compared to the results by ECOSAR, the correlation coefficient between experimental and values predicted by the ECOSAR was 0.322. From the results, the neural network is more superior to the ECOSAR.

Classification of chemicals according to their toxic action is an important problem of ecotoxicological prediction. Toxic action classification based on learning vector quantization classification network has already been studied in the literature [41]. The data consisted on information on 60 molecular topological descriptors for 283 chemicals, and the target species was the fathead minnow fish. The data set was split into a training set of 220 chemicals and a test set containing the information on the remaining 63 chemicals. The experiment focused on discriminating among uncouplers of oxidative phosphorylation, acetyl cholinesterase (AChE) inhibitors, neurotoxicants, respiratory blockers, narcosis I, narcosis II chemicals. The model consisted of a 60-5-5 learning vector quantization classification neural network. The results indicate the need for additional input-parameters, reflecting physico-chemical properties of the chemicals.

In the most QSTR studies, although Kow (octanol-water partition coefficient) values of chemicals have been adapted as input variable, the Kow values should be generated by using QSPR (Quantitative Structure-Property Relationship) interpreting the molecular structure information in their peculiar ways for many substance including chemicals unknown for Kow values. Therefore, it should be possible to omit the Kow values as variable when including in the list of input variables the type of information on which the computation of Kow relies. A practical implementation of this methodology has been used to perform substance toxicity screening for almost 2000 compounds from the Canadian Domestic Substances List, with results superior to all other available methodologies, including ECOSAR, ASTER, CNN, TOPKAT [39].

## CONCLUSIONS

Developing a quantitative structure activity relationship is difficult. Molecules are typically flexible and it is possible to compute many possibly useful properties that might relate to activity. Early in a research program there are typically few compounds to model. In contrast to the traditional, highly focused and linear QSTR methods, artificial neural networks are able to handle non-linear relationships with large set of chemicals of different complexity and mode of action. This ability is of importance particularly in ecotoxicological field. However,

the successful application of QSTR methodology to predict the ecotoxicity of a wide range of structurally diverse chemicals may not be possible without a better understanding the biological mechanism that underlies the expression of lethality. Besides neural network based QSAR methods, more powerful models, such as Fuzzy Logic, Self Organizing Map, and Bayesian Networks should be developed for the description of the chemical compounds using hundreds of molecular descriptors, and molecular residues. In near future, a hybrid system has been set up for the prediction of pesticide ecotoxicity, integrating neural networks and other expert systems.

## ACKNOWLEDGEMENTS

I thank Mr. Masato Kawakami for his help with the analysis work.

## REFERENCES

- [1] Dorca, R.C.; Robert, D.; Amat, L.; Girones, X.; Besalu, R. *Molecular Quantum Similarity in QSAR and Drug Design*, Springer-Verlag: Berlin Heidelberg, **2000**.
- [2] Walker, J. D. QSARs for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Application, SETAC PRESS: Florida, **2003**.
- [3] Shoji, R.; Miyazaki, T.; Nishimiya, T. *J. Environ. Health Sci.*, **2003**, *A38*, 2807.
- [4] Sakai, Y.; Shoji, R.; Kim, B.S.; Sakoda, A.; Suzuki, M. *Environ. Monitoring and Assessment*, **2001**, *70*, 57.
- [5] Kaminokado, K.; Shoji, R. *J. Ecotechnol Res.*, **2003**, *9*, 129.
- [6] Shoji, R.; Sakoda, A.; Sakai, Y.; Suzuki, M. *Wat. Sci. Technol.*, **2002**, *46*, 355.
- [7] Shoji, R.; Sakai, Y.; Sakoda, A.; Suzuki, M. *Applied Microbiol. Biotechnol.*, **2000**, *54*, 432.
- [8] Shoji, R.; Sakoda, A.; Sakai, Y.; Utsumi, H.; Suzuki, M. *J. Health Sci.*, **2000**, *46*, 493.
- [9] Basak, S.C.; Gute, B.D.; Grunwald, G.D. *SQERED*, **1999**, *10*, 117.
- [10] Shoji, R.; Miyazaki, T.; Nishimiya, T. *Annual Report of Interdisciplinary Res. Inst. Environ. Sci.*, **2002**, *21*, 29.
- [11] Prakash, J.; Nirmalakhandan, N.; Sun, B.; Peace, J. *Wat. Res.*, **1996**, *30*, 1459.
- [12] Enslein, K.; Gombar, U.K. *Mutation Res.*, **1997**, *379*, 514.
- [13] Klopman, G.; Stefan L. R.; Saiakhov, R.D. *Euro. J. Pharm. Sci.*, **2002**, *17*, 253.
- [14] Lewis, D.F.V. *Biochem. Pharmacol.*, **2000**, *60*, 293.
- [15] Moriguchi, I.; Hirano, H.; Hirano, S. *Environ. Health Perspect.*, **1996**, *104*, 1051.
- [16] US EPA, **2000**, [http://www.epa.gov/med/databases/aster\\_example.html](http://www.epa.gov/med/databases/aster_example.html).
- [17] Kaiser, K.L.E.; Dearden, J.C.; Klein, W.; Schultz, T.W. *Water Qual. Res. J. Can.*, **1999**, *34*, 1.
- [18] Schultz, T.W.; Cronin, M.T.D.; Walker, J.D.; Aptula, A.O. *J. Mol. Struct.*, **2003**, *622*, 1.
- [19] Shoji, R.; Miyazaki, T.; Nishimiya, T. *Annual Rep. Interdiscip. Res. Inst. Environ. Sci.*, **2002**, *21*, 1.
- [20] Kaiser, K.L.E. *J. Mol. Struct.*, **2003**, *622*, 85.
- [21] Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*, VCH: Weinheim, **1993**.
- [22] Lenze, B.; Raddatz, J. *Int. J. Neural. Syst.*, **2002**, *12*, 83.
- [23] Turner, J.V.; Maddalena, D.J.; Cutler, D.J. *Int. J. Pharmaceutics*, **2004**, *36*, 17.
- [24] Yuan, B.; Wang, X.Z.; Morris, T. *Waste Management*, **2000**, *20*, 677.
- [25] Niculescu, S.P.; Kaiser, K.L.E.; Schultz, T.W. *Arch. Environ. Contam. Toxicol.*, **2000**, *39*, 289.
- [26] Moore, D.R.J.; Breton, R.L.; MacDonald, D.B. *Environ. Toxicol. Chem.*, **2003**, *22*, 1799.
- [27] Pendharkar, P.C.; Rodger, J.A. *Decision Support Syst.*, **2003**, *36*, 17.
- [28] Lampiner, J.; Vehtari, A. *Neural Netw.*, **2001**, *14*, 257.
- [29] Crowe, E.R.; Vassiliadis, C.S. *Chem. Eng. Prog.*, **1995**, *22*.

- [30] Allen, D.M. *Technometrics*, **1974**, *16*, 125.  
[31] Villmann, Th. *Neurocomput.*, **2002**, *48*, 229.  
[32] Rao, R.P.N. *Neural Networks*, **2002**, *15*, 927.  
[33] Gasteiger, J. [http://www2.chemie.uni-erlangen.de/presentations/acs\\_010401/vortrag-31.html](http://www2.chemie.uni-erlangen.de/presentations/acs_010401/vortrag-31.html)  
[34] Zhang, Z.; Ma, X.; Yang, Y. *Neural Networks*, **2003**, *16*, 995.  
[35] Kaiser, K.L.E.; Niculescu, S.P.; McKinnon, M.B. *Quantitative Structure-Activity Relationships in Environmental Science*, SETAC Press; Pensacola, **1997**.  
[36] Maier, H.R.; Dandy, G.C. *Mathematics and Computers Simulation*, **1997**, *43*, 377.  
[37] Dreiseitl, S.; Ohno-Machado, L.J. *Biomed. Inform.*, **2002**, *35*, 352.  
[38] Wold, S.; Eriksson, L. *Chemometric methods in molecular design*, VCH; New York, **1995**.  
[39] Kaiser, K.L.E.; Niculescu, S.P.; Schuurmann, G., *Water Qual. Res. J. Can.*, **1997**, *32*, 637.  
[40] Devillers, J.; Flatin, J. *SAR QSAR Environ. Res.*, **2000**, *11*, 25.  
[41] Basak, S.C.; Grunward, G.; Host, G.E.; Niemi, G.J.; Bradbury, S.P., *Environ. Toxicol. Chem.*, **1998**, *17*, 1056.  
[42] Devillers, J. *Toxicol. Meth.*, **2000**, *10*, 69.  
[43] Kaiser, K.L.E.; Niculescu, S.P.; Schultz, T.W., *Environ. Toxicol. Chem.*, **2001**, *20*, 420.  
[44] Devillers, J. *Toxicol. Meth.*, **2000**, *10*, 397.  
[45] Gagne, F.; Blaise, C. *Chemosphere*, **1997**, *35*, 1343.  
[46] Niculescu, S.P.; Kaiser, K.L.E.; Schultz, T.W. *Arch. Environ. Toxicol. Chem.*, **2000**, *39*, 289.  
[47] Serra, J.R.; Jurs, P.C.; Kaiser, K.L.E. *Chem. Res. Toxicol.*, **2001**, *14*, 1535.  
[48] Devillers, J.; Domine, D. *SAR QSAR Environ. Res.*, **1999**, *10*, 61.  
[49] Domine, D.; Devillers, J.; Wienke, D.; Buydens, L. *Quant. Struct.-Act. Relat.*, **1996**, *15*, 395.

---

Received: July 17, 2004

Accepted: August 24, 2004