

Artificial Neural Network-Based Equation to Predict the Toxicity of Herbicides on Rats

Mabrouk Hamadache, Salah Hanini, Othmane Benkortbi, Abdeltif Amrane, Latifa Khaouane, Cherif Si Moussa

▶ **To cite this version:**

Mabrouk Hamadache, Salah Hanini, Othmane Benkortbi, Abdeltif Amrane, Latifa Khaouane, et al.. Artificial Neural Network-Based Equation to Predict the Toxicity of Herbicides on Rats. Chemometrics and Intelligent Laboratory Systems, 2016, 154, pp.7-15. 10.1016/j.chemolab.2016.03.007 . hal-01295518

HAL Id: hal-01295518

https://univ-rennes.hal.science/hal-01295518

Submitted on 10 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACCEPTED MANUSCRIPT

Artificial Neural Network-Based Equation to Predict the Toxicity of Herbicides on Rats

Mabrouk Hamadache[1*], Salah Hanini[1], Othmane Benkortbi[1], Abdeltif Amrane[2], Latifa Khaouane[1] and Cherif Si Moussa[1]

[1] *Laboratoire des Biomatériaux et Phénomènes de Transport (LBMPT), Université de Médéa, Quartier Ain D'heb, 26000, MEDEA, Algérie*
[2] *Ecole Nationale Supérieure de Chimie de Rennes, Université de Rennes 1, CNRS, UMR 6226, 11 allée de Beaulieu, CS 50837, 35708 Rennes Cedex 7, France*

L. KHAOUANE : latifa_khaouane@yahoo.fr ; O. BENKORTBI: benkortbi_oth@yahoo.fr ;
C. SI MOUSSA : simoussa_cherif@yahoo.fr ; S. HANINI : s_hanini2002@yahoo.fr ;
A. AMRANE: abdeltif.amrane@univ-rennes1.fr

*****Corresponding author: Mabrouk HAMADACHE;
mhamdeche@yahoo.fr
Tel: + 213 07 78 12 37 50; Fax: +213 25 58 12 53

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

The use of herbicides is increasing around the world. The benefits achieved by the use of these herbicides are indisputable. Despite their importance in agriculture, herbicides can be dangerous to the environment and the human health, depending on their toxicity, and the degree of contamination. Also, it is essential and evident that the risk assessment of herbicides is an important task in the environmental protection. The objective of this work was to investigate and implement an Artificial Neural Network (ANN) model for the prediction of acute oral toxicity of 77 herbicides to rats. Internal and external validations of the model showed high Q^2 and $\overline{r_m^2}$ values, in the range 0.782 – 0.997 for the training and the test. In addition, the major contribution of the current work was to develop artificial neural network-based equation to predict the toxicity of 13 other herbicides; the mathematical equation using the weights of the network gave very significant results, leading to an R^2 value of 0.959. The agreement between calculated and experimental values of acute toxicity confirmed the ability of ANN-based equation to predict the toxicity for herbicides that have not been tested as well as new herbicides.

Key words: acute oral toxicity, ANN-based equation, domain applicability, herbicides, prediction.

1. Introduction

Herbicides are widely used in agriculture. They are indispensable to the farmer in his fight against plant pests and diseases. They are also used to slow the spread of insects. The benefits achieved by the use of herbicides are indisputable. Despite these advantages, several environmental dangers and some potential risk have emerged from the excessive use of these compounds. For nearly fifty years, they have been detected in the water of rivers and groundwater **[**1-10**]**. They are also found in agricultural and animal products (wheat, corn, fruits, vegetables, cereals, tea, fish, milk, eggs, meat, honey and medicinal herbs, etc.) **[**11-14**]**. As a result, this contamination could give rise to serious health and safety problems for consumers.

Herbicides have a major drawback such as toxicity. Long-term exposure to herbicides can cause harm to human life and can disrupt the functioning of various organs in the body. This significant relationship between exposure to herbicides and some chronic diseases has been the subject of several scientific publications. Exposure to these persistent pesticides has been associated with health effects including cancer, headache, skin and eye irritation, immune system problems, stomach, kidney, Parkinson and Alzheimer's disease, reproductive difficulties, birth defects, diabetes, cataracts and anemia [15-22]. As seen, humans and the environment are exposed to hundreds of herbicides. The pollution caused by these compounds has become an important issue affecting the survival and the development of humains. It is evident that the risk assessment for herbicides can provide a precaution against the corresponding pollution. In environmental risk assessment, knowledge of the acute toxicity and chronic toxicity is a basic need [23-25].

Development of *in silico* predictive methods that are designed to reduce and replace the use of animals to predict biological activity of chemical compounds is a widely explored area of predictive toxicology [24]. This pathway is imposed for several reasons: economic considerations, reduction of time constraints, and pressure of public opinion [26]. These methods, which include Quantitative Structure–Activity Relationship (QSAR) has been used in medicinal chemistry and computational toxicology for a long time, find growing applications in chemical risk assessment and are indispensable tools for ecotoxicological risk assessment [27,

ACCEPTED MANUSCRIPT

28]. Of the fact that is a promising technique, an increasing interest in the use of QSAR for environmental risk assessment and for predicting toxicity [29, 30] is observed.

Quantitative structure–activity relationship (QSAR) models are increasingly used in toxicology, ecotoxicology, and pharmacology for predicting the activity of the molecules from their physicochemical properties and/or their structural characteristics. A QSAR model is a mathematical relationship between the chemical's quantitative molecular descriptors and its toxicological, biological, and physicochemical activities. These descriptors are then correlated with a toxicological response of interest through a suitable statistical approach such as linear multiple regression, discriminant analysis and artificial neural networks [31]. The establishment of QSAR models involves a number of steps and conditions: accuracy of the input data, obtain and select the relevant descriptors capable to reflect the structure of the compounds, selection of appropriate statistical tools and checking the validity and stability of the suggested model. Reference books dealing with fundamental concepts of QSAR modeling and their basic concepts for applications in risk assessment are currently available in the literature [32, 33].

QSAR studies conducted by the use of artificial neural network (ANN) modeling approaches have been developed for a large number of toxic endpoints with varying methodologies and varying degrees of success. Their applications encompass both the human health effects and the environmental impact of chemicals [34]. In recent years, researchers have used different modelling techniques such as artificial neural networks (ANN) to reduce the numbers of expensive, complicated and time-consuming tests. Predictive models based on ANN have been studied extensively in many areas of medicine [35]. Advantageously, a neural network (NN) model has a distinctive ability of learning nonlinear functional relationships. It does not require any prior structural knowledge of relationships between important variables and processes to be modeled.

There are many reports about QSAR prediction of pesticides toxicity [36-38]; however, among this abundant literature, studies specifically dedicated to QSAR prediction of herbicides acute oral toxicity appear rather limited. So far, no artificial neural network-based equation has been developed to predict acute oral toxicity of herbicides on rats. Currently, testing for acute oral toxicity is still required in the toxicological assessment of chemicals and agrochemicals worldwide [39]. Consequently, the aim of this study is to develop an ANN-based equation to predict acute oral toxicity of herbicides on rats.

The step one of this work is to develop a QSAR model that could be used to predict oral acute (LD_{50}) toxicity of a diverse set of 77 herbicides on rats. The QSAR model established by using artificial neural networks and molecular descriptors satisfies the guidelines required by the Organisation for Economic Cooperation and Development (OECD). The basic requirements to develop a QSAR model were respected. The first work is to use herbicides with toxicity data with high quality obtained under the same experimental conditions (i.e., the same protocol). Selection of non-redundant and non-correlated descriptors is the second requirement. Third, the statistical tool used to derive the QSAR can be in some cases a source of mistakes and hence the commercial software Statistica was used. Finally, the model is evaluated both in terms of her robustness as well as in terms of her prediction performances and its applicability domain (AD).

The second step of this study is to calculate the oral acute (Lethal Dose: LD_{50}) toxicity of other 13 herbicides based on the developed mathematical equation using the weights of the network. The accuracy of this formula based on ANNs was investigated and the results were very encouraging.

ACCEPTED MANUSCRIPT

2. Materials and Methods

2.1. Rat LD_{50} data

It is well known that high-quality experimental data are essential for the development of high quality QSAR models [40]. If they are unreliable, the model will be unreliable. The rat lethal dose 50 (LD_{50} - rat, male via oral exposure) values were retrieved from Pesticide Properties DataBase (PPDB) [41]. The LD_{50} correspond to the concentration (mg/kg) of pesticide that leads to the death of 50% of rat. The LD_{50} is one way to measure the short-term poisoning potential (acute toxicity) of a material. All values of oral acute toxicity were first converted into mmol/kg body weight and the $1/LD_{50}$ [$(mmol/kg)^{-1}$] as the endpoint was examined. The initial database that included 146 herbicides was rigorously reviewed and "cleaned" by removing pesticides whose LD_{50} was not experimentally determined or whose LD_{50} was not determined in the same experimental conditions. A total of 90 herbicides with experimental data were selected to form the final database and was divided into two sets. The first set with 77 herbicides (Table 1) was dedicated to develop the QSAR model (64 herbicides for training, and 13 herbicides for test set. The second set which included 13 herbicides that had not been used for the development of the QSAR model, was left for the prediction of oral acute LD_{50} based on the developed mathematical formula using the weights and the bias of the network.

2.2. Descriptor calculation

All descriptors were obtained from the online program E-Dragon 1.0 (www.vcclab.org). The structure files of compounds under study, which are the input files for Dragon calculation, cannot be generated in Dragon. The structures have been drawn in SMILES (Simplified Molecular-Input Line-Entry System) notation. SMILES notations were obtained from the Pesticide Properties DataBase (University of Hertfordshire, 2007–2013). Herbicides compounds represented by SMILES format was used as input for calculation of 1666 molecular descriptors with the online software, E-DRAGON. The software converted the molecules from SMILES notation to 3-dimensional structures using the algorithm derived from CORINA [42]. Twenty types of descriptors were calculated by the Dragon software, like: (1) constitutional descriptors; (2) topological descriptors; (3) walk and path counts; (4) connectivity indices; (5) information indices; (6) two dimensional (2D) autocorrelations; (7) edge adjacency indices; (8) Burden eigenvalue descriptors; (9) topological charge indices; (10) eigenvalue-based indices; (11) Randic molecular profiles; (12) geometrical descriptors; (13) RDF descriptors; (14) 3D-MORSE descriptors; (15) WHIM descriptors; (16) GETAWAY descriptors; (17) functional group counts; (18) atom-centered fragments; (19) charge descriptors; and (20) molecular properties.

2.3. Selection of relevant descriptors

An important step in QSAR model is to select robust and informative descriptors from a variety of descriptors. Several methods to simplify a database are used; for example the Principal Component Analysis (PCA), curvilinear component analysis, or the method of Gram-Schmidt orthogonalization can be used. The method used to select the most significant descriptors was described previously [43, 44]. In the first step, invariant descriptors, namely those with absent values (represented by the code ''999''), were manually removed. Next, any descriptor that had identical values for >75% of the samples and any descriptors with a relative standard deviation<0.05 were removed. Finally, half of the descriptors showing an absolute value of the Pearson correlation coefficient > 0.75 were also removed. The number of descriptors obtained after the selection

ACCEPTED MANUSCRIPT

was 76. For relevant descriptors selection, stepwise regression was then used [45]; in this procedure, a variable that entered the model in the earlier stages of selection may be deleted at the later stages. Stepwise addition of further descriptors was continued to find the best multi-parameter regression models with the optimal values of statistical criteria (highest values of correlation coefficient R^2). Stepwise regression were performed by the STATISTICA software (STATISTICA 8.0, Tulsa; StatSoft, Inc, OK, USA.) and XLSTAT software. Eighteen descriptors were selected with stepwise regression. However, it was important to reduce the number of descriptors [46, 47]. Finally, the number of descriptors used to develop the model was 8: HATS1e, HATS1v, ISH, MATS1m, Gats3p, R8u, Gats6m and H-046.

2.4. Model development

There are plenty of different models of neural networks to choose from, each one having its specific properties and advantages for its particular application. One of the most successful and most popular is the feed-forward Multi-Layered Perceptron (MLP) [48-50]. The structure of an MLP consists of one input layer (corresponds to the independent variables: descriptors), one intermediate or hidden layer, and one output layer corresponds to the dependent variable (oral toxicity). Each layer can have a number of neurons, which are connected linearly by weights to the neurons in the neighboring layers. ANN calculations were performed by the STATISTICA software (STATISTICA 8.0, Tulsa; StatSoft, Inc.) to study the structure–activity relationship of various herbicides. A set of 8 descriptors were used as input parameters of the network.

2.5. Model validation

Validation is a crucial and important aspect for determination of reliability of models. There are several approaches of validation including internal validation and external validation. Recent studies [51] indicated that the internal validation is considered to be necessary for model validation. Recently, Roy et al. [52**]** proposed a r_m^2 metrics as additional validation parameters. The r_m^2 metrics can be computed from http://aptsoftware.co.in/rmsquare/.

The most important statistical parameters used in our study to check the performance of the model are the root mean square error (RMS), the determination coefficient (R^2), the cross validated correlation coefficient (Q^2), r_m^2, $r_m'^2$, $\overline{r_m^2}$, Δr_m^2 values for the training and test set, and $r_{m(overall)}^2$ values for the overall set. In the cases where the size of the test set is small, the r_m^2 (test set) may be less reliable and highly dependent on individual test set observations. Thus, the $r_{m(overall)}^2$ statistical may be used, since it is based on both test set and training set predictions. Therefore, the result is based on the prediction of a comparably large number of compounds [53]. The statistical parameters are collected in Eqs. (1) - (8). The terms which are utilized in these equations are defined below:

Y_{obs} Observed (experimental) value of Y

Y_{pred} Predicted Y-value of training set, test set or validation set

n Number of compounds in the data set (training, test, validation)

$\bar{Y}_{obs}$ Average of Y_{obs}

r^2 Squared correlation coefficient between the observed and predicted value of compounds with intercept.

r_0^2 Squared correlation coefficient between the observed and predicted value of compounds without intercept.

$r_0'^2$ It bears the same meaning as r_0^2 , but uses the reversed axes.

ACCEPTED MANUSCRIPT

$$RMS = \sqrt{\frac{\sum \left(Y_{obs} - Y_{pred}\right)^2}{n}} \quad (1)$$

$$R^2 = 1 - \frac{\sum \left(Y_{obs} - Y_{pred}\right)^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2} \quad (2)$$

$$Q^2_{LOO} = 1 - \frac{\sum \left(Y_{obs\,(training)} - Y_{pred\,(training)}\right)^2}{\sum (Y_{obs(training)} - \bar{Y}_{obs(training)})^2} \quad (3)$$

$$Q^2_{test} = 1 - \frac{\sum \left(Y_{obs\,(test)} - Y_{pred\,(test)}\right)^2}{\sum (Y_{obs(test)} - \bar{Y}_{obs(training)})^2} \quad (4)$$

$$r^2_m = r^2 \left(1 - \sqrt{r^2 - r^2_0}\right) \quad (5)$$

$$r'^2_m = r^2 \left(1 - \sqrt{r^2 - r'^2_0}\right) \quad (6)$$

$$\bar{r}^{\,2}_m = \frac{\left(r^2_m + r'^2_m\right)}{2} \quad (7)$$

$$\Delta r^2_m = \left| r^2_m - r'^2_m \right| \quad (8)$$

2.6. Applicability domain

Application domain is defined as the "substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy within the intended application of the model. QSAR models are only valid in the domain they were trained and validated. Extrapolation is dangerous and can lead to grossly erroneous model predictions [54]. The determination of AD is therefore of great importance [55]. The AD is a theoretical region in the space defined by the descriptors of the model and the modeled response for which a given QSAR should make reliable predictions.

There are various approaches for determining AD of QSAR models. Each method has its own merits and flaws. As part of our work, we used for comparison purposes two methods: the first is the leverage approach (Williams plot) which has been largely employed to identify outliers and the compounds residing outside the AD. The second method is a simple statistical approach to define AD of a QSAR model. This approach, which has been reported by Roy et al. [56], is very easy, but she performs well in comparison to the leverage approach.

2.6.1 Applicability domain using leverage approach

The leverage h_{ii} is defined as follows [57]:

$$h_{ii} = \frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \quad (9)$$

where $\boldsymbol{x_i}$ is the descriptor value of the ith object, and $\bar{x}$ is the average value of the descriptor in the training set, and n is the number of substances in the training set. The warning leverage h* is, generally, fixed at 3(p + 1)/n,

ACCEPTED MANUSCRIPT

where n is the total number of samples in the training set and p is the number of descriptors involved in the correlation.

The applicability domain (AD) of QSAR model is defined from the Williams plot. In this plot, the two horizontal lines indicate the limit of normal values for Y outliers (i.e. samples with standardized residuals greater than 3.0 standard deviation units, ±3.0s); the vertical straight lines indicate the limits of normal values for X outliers (i.e. samples with leverage values greater than the threshold value, h > h*). For a sample in the external test set whose leverage value is greater than h*, its prediction is considered unreliable, because the prediction is the result of a substantial extrapolation of the model. Conversely, when the leverage value of a compound is lower than the critical value, the probability of accordance between predicted and experimental values is as high as that for the compounds in the training set [58].

2.6.2 Applicability domain using standardization approach

This approach is a simple method for defining outliers (in the case of the training set) and the compounds residing outside the AD (in the case of the test set). An open access standalone application has also been developed for the calculation of the AD for QSAR models. The software can be accessed from the following link: http://dtclab.webs.com/softwaretools or http://teqip.jdvu.ac.in/QSAR_Tools/. The background theory, the algorithm and methodology and the advantages of the proposed approach are available in literature [56].

3. Results and Discussion

3.1. Selection of relevant descriptors

The selection of the optimum number of descriptors was shown in Fig. 1. The data set of descriptors obtained after selection by the stepwise method was composed of 18 descriptors. However, since the number of herbicides in the training set was 77, it was important to reduce the number of descriptors until the ratio "number of herbicides/predictors" is ≥ 5. To select the most important descriptors and the optimal number, the influences of the number of descriptors on the statistical parameters (R^2, Q^2 and RMSE) were investigated for 1–18 descriptors. The selection of the optimum number of descriptors was shown in Fig.1. Fig.1 shows that beyond 8 descriptors, there is no significant improvement of the statistical parameters. For these reasons, the number of descriptors used to develop the model was 8. The descriptors which obtained from stepwise multiple linear regression were HATS1e, HATS1v, ISH, MATS1m, Gats3p, R8u, Gats6m, H-046, respectively. Order to study the correlation between the selected descriptors, the correlation matrix has been established using the XLSTAT software. The value of the correlation coefficient of each pair of selected descriptors was < 0.639, which means that the selected descriptors were independent.

Descriptors used in our model have been used in previous QSAR models in the literature, but this combination of descriptors has never been mentioned. A QSAR model with high statistical quality for predicting toxicity of phenols was developed by Habibi-Yangjeh [59] with MATS1m and H-046. A QSAR model to predict the toxicities of a diverse set of pharmaceuticals to fish developed by Tugcu [60] employed Gats3p. Moreover, some authors [61-64] found that among the descriptors that affect the toxicity of the compounds studied, a substantial number belong to the categories Getaway descriptors, 2D autocorrelations, and Atom-centered

ACCEPTED MANUSCRIPT

fragments. In our study, the eight relevant descriptors involved in the model also belong to this category. It would seem that the descriptors in this category have major significance in the toxicity of herbicides.

3.2. QSAR modeling

In this investigation, the tanh (hyperbolic tangent) function was used as a transfer function of hidden layer and a logistic function for the output layer. The number of hidden neurons was optimized by trial and error procedure in the training process. One output neuron was used to represent the observed LD_{50}. The network was trained using the BFGS quasi-Newton methods algorithm. To optimize the number of nodes in the hidden layer, several calculations were performed with different numbers of hidden nodes (1–30). The 77 herbicides were divided into two groups: training and test set composed of 64 and 13 herbicides, respectively. This network consisted of 8 inputs and one output for LD_{50}. Then, an ANN with architecture {8-12-1} was generated.

The predictive results from the ANN model for the entire dataset (77 compounds) are obtained and presented in Table 1. Fig. 2 shows the regression line of the model equation, i.e. predicted vs experimental results for the training, and test set highlighted by different symbols. A close correlation between the values predicted by the ANN model and the observed values of toxicity was found.

As can be seen from Table 2, the non-linear ANN model gave good results with higher correlation coefficients R^2, as well as better robustness (Q^2) in training, and test set. In addition to the classical validation parameters (Q^2_{LOO} and Q^2_{test}), different r^2_m values were also checked for both training and test sets. The values of $\overline{r^2_m}$ for the training set (0.991) and the test set (0.753) are greater than 0.5. Furthermore, the Δr^2_m values for both training (0.002) and test (0.058) sets are lower than 0.2 [52]. Moreover, statistically significant results for all the r^2_m metrics indicate that the predicted activity values of all the herbicides are close to the corresponding observed activity data. The model exhibits high predictive ability. An acceptable value (0.993) of r^2_m (overall) implies that the activity data predicted for the test set compounds using the model satisfies the desired range of observed activity data. These results indicated that the ANN not only performed well in model development, but also had excellent prediction and this fact suggested that a non-linear correlation between the acute toxicity and the relevant descriptors. The residuals plot for the observed values of LD_{50} in the training, and prediction sets against their predicted values were determined and studied. The model did not show proportional and systematic error, because the distribution of the residuals on both sides of zero was random.

To see the importance of each variable for the prediction of acute toxicity, a sensitivity analysis was conducted using STATISTICA software. This method, proposed by Garson [65] then taken by Goh [66], provides a quantification of the relative importance of different inputs (variables) on the output of the NN. The contribution of each of the descriptors in the ANN model is as follows: HATS1e (37.5%), HATS1v (14.4%), ISH (13.7%), MATS1m (9.3%), Gats3p (8.8%), R8u 8.5%), Gats6m (5.6%), and H-046 (0.2%).

3.3. Applicability domain

The leverage values (hi) for the studied herbicides are shown in Table1. The applicability domain of the model was analyzed using leverage approach (Williams plot) and standardization approach. As can be seen in Fig. 3, training set compound numbers 47 and 64 are identified as outliers and test set compound numbers 71 are identified as outside the AD by both approaches. However, the training compound 25 is identified outside the AD by the leverage approach but recognized within the AD by standardization approach. On the contrary, the

ACCEPTED MANUSCRIPT

test compound 69 is identified outside the AD by standardization approach but recognized within the AD by leverage approach.

It should be noted that over 95% of the domain was covered by the model when it was applied to predict the acute oral toxicity of 77 herbicides in the training and test set. Thus, these results show that ANN model is complies with the third principle of the OECD. It can be used to predict the acute oral toxicity of herbicides, particularly for those that have not been tested as well as new herbicides.

3.4 Comparison with different models

Following the steps of model validation, our model was compared with a limited number of QSAR models which are available in the literature for predicting the oral acute toxicity of herbicides to rats (Table 3). The evaluation of their advantages and disadvantages is quite difficult, because each published study used different data sets and a different modeling approach (chemical descriptors, algorithms, etc.). In addition, it should be noted that the most of these QSAR models were obtained with structurally similar chemicals such as amide herbicides [67, 68], sulphonylurea and phenylurea herbicides [69, 70] or organophosphorus pesticides [71].

In Table 3, it is possible to observe that unlike our model, no approach for measuring external quality was carried out in the other models with the exception of three models. Can et al. [69] used a single parameter (R^2) while Devillers [71] and Gough et al. [67] have used the root mean square error (RMS). Thus, the comparison was limited to the results obtained for statistics of the internal validation. Again, the number of statistical parameters used for internal validation of this QSAR models is limited. It is possible to observe that all of those models could give high prediction ability (correlation coefficient R^2). However, our model exceeds the previously published models in all statistical indices available for comparison. Indeed, it gives the higher correlation coefficient and the acceptable RSM if compared to the other models. According to these results, the present model can be promisingly used for predicting the toxicity of new herbicides, thus contributing to the risk assessment, saving substantial amounts of money and time.

3.5. Application of artificial neural network-based equation

The architecture of the network developed in this study is a multilayered perceptron {8-12-1} (Fig. 4). The network has eight inputs (X_i, i=1 to 8), one output (Z) and twelve neurons in the hidden layer. The two transfer functions used in this study are hyperbolique tangent and logistic function. Their mathematical definitions are given in Eqs. (10) and (11):

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{10}$$

$$f(x) = \frac{1}{1 + e^{-x}} \tag{11}$$

Each of these twelve input layer's neurons receive one input (X_i, i=1 to 8)) and broadcasts such signal to each one of the hidden layer's neurons. Each hidden neuron computes its transfer function and sends its result (Y_j, j=1 to 12) to the output layer's neuron which finally produces the response of the network (Z). The output signal of each hidden neuron (Yj) is calculated as:

ACCEPTED MANUSCRIPT

$$Y_j = f\left[\sum_{i=1}^{8} w_{i,j} X_i + b_j\right] = \frac{\exp\left(\sum_{i+1}^{8} w_{i,j} X_i + b_j\right) - \exp\left(-\sum_{i=1}^{8} w_{i,j} X_i + b_j\right)}{\exp\left(\sum_{i=1}^{8} w_{i,j} X_i + b_j\right) + \exp\left(-\sum_{i=1}^{8} w_{i,j} X_i + b_j\right)} \quad (12)$$

while the output of the network is given by:

$$Z = f\left[\sum_{j=1}^{12} w_{1,j} Y_J + b_1\right] = \frac{1}{1 + \exp\left(-\sum_{j=1}^{12} w_{1,j} Y_j + b_1\right)}$$

(13)

In Eqs (13) $w_{i,j}$ is the weights of the connections between the input and hidden neurons, X_i are the input variables (relevant descriptors) and b_j is the bias on hidden neuron j. Similarly, $w_{1,j}$ represent the weights of the connections between the hidden and output neuron and b_1 is the bias on the output neuron.

The mathematical formula for predicting the acute oral toxicity of herbicides on rat obtained using the ANN approach is given in equation (13). With this formula, toxicity of 13 herbicides is calculated and carried out for comparison with experimental values (Table 4). The toxicity of each herbicide is calculated by replacing X_i by the values of 8 descriptors of this herbicide, the values of weights and bias obtained. The weights generated for architecture ANN are presented in Tables 5 and 6. Fig. 5 shows the relationship between the observed $1/LD_{50}$ and the calculated (or estimated) $1/LD_{50}$ from the ANN-equation. The best linear fit is indicated by a dashed line. As can be seen from this figure, there is an excellent agreement between the results of our equation and the observed (experimental) data because most of the data points falls close to the 45 degree line (zero error line). The calculated RMS for the regression equation is 0.16 and the value of the R is 0.77.

4. Conclusions

In this study, artificial neural network model was developed for predicting the oral acute toxicity on rats of a series of 77 herbicides based on their molecular structure, represented by 1666 calculated descriptors. Experimental data have been selected from the Pesticide Properties DataBase. The best model was obtained by BFGS quasi-Newton algorithm with {8-12-1} network architecture. The built ANN model was assessed comprehensively (internal and external validations). It showed good values of $R^2 = 0.996$ and $Q^2_{LOO} = 0.996$ for the training set, good values of $R^2 = 0.956$ and $Q^2_{LOO} = 0.977$ for the test set. More than that, the robustness and predictive power of the model was verified by the different r_m^2 values. In addition, estimating the oral acute (LD50) toxicity of 13 other herbicides based on the developed mathematical equation using the weights of the network gave very good results ($R^2 = 0.956$). Based on the comparison with models previously published, the proposed QSAR model achieved good results and provided more than 95% predictions that belong to the applicability domain. According to the obtained results, the ANN model developed and the mathematical formula obtained using the ANN approache for predicting the acute oral toxicity of herbicides to rats gave correct and acceptable results. Therefore, instead of expensive, complicated and time-consuming experiments, it is highly recommended that the ANN can be used for predicting the oral acute toxicity of herbicides to rats, particularly for those that have not been tested as well as new herbicides and thus help reduce the number of animals used for experimental purposes.

ACCEPTED MANUSCRIPT

Conflict of Interests

The authors declare that there is no conflict of interests.

References

[1] C.T. Hansen, C. Ritz, D. Gerhard, J.E. Jensen, J.C. Streibig, Re-evaluation of groundwater monitoring data for glyphosate and bentazone by taking detection limits into account, Sci. Total Environ. 536 (2015) 68–71.

[2] M.J. Hedegaard, E. Arvin, C.B. Corfitzen, H.J. Albrechtsen, Mecoprop (MCPP) removal in full-scale rapid sand filters at a groundwater-based waterworks, Sci. Total Environ. 499 (2014) 257–264.

[3] P. Palma, M. Köck-Schulmeyer, P. Alvarenga, L. Ledo, I.R. Barbosa, M. López de Alda, D. Barceló, Risk assessment of pesticides detected in surface water of the Alqueva reservoir (Guadiana basin, southern of Portugal), Sci. Total Environ. 488–489 (2014) 208–219.

[4] I.A. Idowu, R.M. Alkhaddar, W. Atherton, Possible source term of high concentrations of Mecoprop-p in leachate and water quality: impact of climate change, public use and disposal, Environ. Technol. 35 (2014) 2055-2067.

[5] M.C. Hermosin, M.J. Calderon, M. Real, J. Cornejo, Impact of herbicides used in olive groves on waters of the Guadalquivir river basin (southern Spain), Agric. Ecosyst. Environ. 164 (2013) 229–243.

[6] A. Masiá, J. Campo, P. Vázquez-Roig, C. Blasco, Y. Picó, Screening of currently used pesticides in water, sediments and biota of the Guadalquivir River Basin (Spain), J. Hazard. Mater. 263 (2013) 95–104.

[7] J. Tournebize, E. Passeport, C. Chaumont, C. Fesneau, A. Guenne, B. Vincent, Pesticide de-contamination of surface waters as a wetland ecosystem service in agricultural landscapes, Ecol. Eng. 56 (2013) 51– 59.

[8] B.E. Björklund, G.G. Anskjær, M. Hansen, B. Styrishave, B. Halling-Sørensen, Analysis and environmental concentrations of the herbicide dichlobenil and its main metabolite 2,6-dichlorobenzamide (BAM): A review, Sci. Total Environ. 409 (2011) 2343–2356.

[9] L. Gutowski, O. Olsson, C. Leder, K. Kümmerer, A comparative assessment of the transformation products of S-metolachlor and its commercial product Mercantor Gold and their fate in the aquatic environment by employing a combination of experimental and in silico methods, Sci. Total Environ. 506–507 (2015) 369-379.

[10] Y.Q. Zhao, P. Singleton, S. Meredithand, G.W. Rennick, Current status of pesticides application and their residue in the water environment in Ireland, Int. J. Environ. Stud. 70 (2013) 59–72.

[11] F.A. Oliveira, L.P. Reis, B. Soto-Blanco, M.M. Melo, Pesticides residues in the Prochilodus costatus (Valenciennes, 1850) fish caught in the São Francisco River, Brazil, J. Environ. Sci. Health B. 50 (2015) 398-405.

[12] N. Takakura, P. Sanders, V. Fessard, L. Le Hégarat, In vitro combined cytotoxic effects of pesticide cocktails simultaneously found in the French diet, Food Chem. Toxicol. 52 (2013) 153-162.

[13] C. Ahouangninou, M. Thibaud, P. Edorh, B.B. Sahabi, S. Onil, L. St-Laurent, S. Dion, B. Fayomi, Characterization of Health and Environmental Risks of Pesticide Use in Market-Gardening in the Rural City of Tori-Bossito in Benin, West Africa, J. Environ. Protect. 3 (2012) 241-248.

[14] S. Kang, N. Chang, Y. Zhao, C. Pan, Development of a Method for the Simultaneous Determination of Six Sulfonylurea Herbicides in Wheat, Rice, and Corn by Liquid Chromatography–Tandem Mass Spectrometry, J. Agr. Food Chem. 59 (2011) 9776–9781.

[15] C. Wu, Y. Luo, T. Gui, Y. Huang, Concentrations and potential health hazards of organochlorine pesticides in shallow groundwater of Taihu Lake region, China, Sci. Total Environ. 470–471 (2014) 1047–1055.

[16] M.T. Baltazar, R.J. Dinis-Oliveira de Lourdes, M. Bastos, A.M. Tsatsakis, J.A. Duarte, F. Carvalho, Pesticides exposure as etiological factors of Parkinson's disease and other neurodegenerative diseases-a mechanistic approach, Toxicol. Lett. 230 (2014) 85-103.

[17] L.S. Kjeldsen, M. Ghisari, E.C. Bonefeld-Jørgensen, Currently used pesticides and their mixtures affect the function of sex hormone receptors and aromatase enzyme activity, Toxicol. Appl. Pharmacol. 272 (2013) 453-464.

[18] W. Zheng, H. Yu, X. Wang, W. Qu, Systematic review of pentachlorophenol occurrence in the

ACCEPTED MANUSCRIPT

environment and in humans in China: Not a negligible health risk due to the re-emergence of schistosomiasis, Environ. Int. 42 (2012) 105–116.

[19] W.S. Goldner, D.P. Sandler, F. Yu, J.A. Hoppin, F. Kamel, T.D. Levan, Pesticide use and thyroid disease among women in the Agricultural Health Study, Am. J. Epidemiol. 171 (2010) 455-64.

[20] K. Grote, L. Niemann, B. Selzsam, W. Haider, C. Gericke, M. Herzler, I. Chahoud, Epoxiconazole causes changes in testicular histology and sperm production in the Japanese quail (Coturnix coturnix japonica), Environ. Toxicol. Chem. 27 (2008) 2368-2374.

[21] T. Wigle Donald, C.T. Michelle, D. Krewski, A Systematic Review and Meta-analysis of Childhood Leukemia and Parental Occupational Pesticide Exposure, Environ. Health Perspect. 117 (2009) 1505-1513.

[22] M.P. Montgomery, F. Kamel, T.M. Saldana, M.C.R Alavanja, D.P. Sandler, Incident diabetes and pesticide exposure among licensed pesticide applicators: Agricultural Health Study 1993–2003, Am. J. Epidemiol. 67 (2008) 1235-1246.

[23] A.A. Lagunin, A.V. Zakharov, D.A. Filimonov, V.V. Poroikov, A new approach to QSAR modelling of acute toxicity, SAR QSAR Environ. Res. 18 (2007) 285-298.

[24] D.-S. Cao, J. Dong, N.-N. Wang, M. Wen, B.-C. Deng, W.-B. Zeng, Q.-S. Xu, Y.-Z. Liang, A.-P. Lu, A.F. Chen, In silico toxicity prediction of chemicals from EPA toxicity database by kernel fusion-based support vector machines, Intell. Lab. Syst. 146 (2015) 494-502.

[25] S. Yousefinejad, B. Hemmateenejad, Chemometrics tools in QSAR/QSPR studies: A historical perspective, Chemometr. Intell. Lab. Syst. (2015), http://dx.doi.org/10.1016/j.chemolab.2015.06.016.

[26] J. Devillers, H. Devillers, Prediction of acute mammalian toxicity from QSARs and interspecies correlations, SAR QSAR Environ. Res. 20 (2009) 467-500.

[27] E.M. Hulzebos, R. Posthumus, QSARs: Gate-keepers against risk on chemicals, SAR QSAR Environ. Res. 14 (2003) 285-316.

[28] M. Yang, X. Zhang, Comparative developmental toxicity of new aromatic halogenated DBPs in a chlorinated saline sewage effluent to the marine polychaete Platynereis dumerilii, Environ. Sci. Technol. 47 (2013) 10868-10876.

[29] G. Om Praba, D. Velmurugan, Quantitative structure-activity relationship of some pesticides, Indian J. Biochem. Biophys. 43 (2006) 154-159.

[30] K. Roy, S. Kar, P. Ambure, On a simple approach for determining applicability domain of QSAR models, Chemom. Intell. Lab. Syst. 145 (2015) 22–29.

[31] P. Ruiz, G. Begluitti, T. Tincher, J. Wheeler, M. Mumtaz, Prediction of Acute Mammalian Toxicity Using QSAR Methods: A Case Study of Sulfur Mustard and Its Breakdown Products, Molecules, 17 (2012) 8982-9001

[32] K. Roy, S. Kar, R.N. Das, Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Academic Press, 2015, pp. 1-46.

[33] K. Roy, S. Kar, R.N. Das, A Primer on QSAR/QSPR Modeling: Fundamental Concepts (Springer Briefs in Molecular Science), Springer, 2015, pp. 1-35.

[34] M.T.D. Cronin, The Current Status and Future Applicability of Quantitative Structure-activity Relationships (QSARs) in Predicting Toxicity, Altern. Lab. Anim. 30 (2002) 81-84.

[35] J.R. Ruiz, J. Ramirez-Lechuga, F.B. Ortega, J. Castro-Pinero, J.M. Benitez, A. Arauzo-Azofra, C. Sanchez, M. Sjostrom, M.J. Castillo, A. Gutierrez, M. Zabala, Artificial neural network-based equation for estimating VO_{2max} from the 20 m shuttle run test in adolescents, Artif. Intell. Med. 44 (2008) 233-245.

[36] A. Speck-Planche, V.V. Kleandrova, F. Luan, M.N.D.S. Cordeiro, Predicting multiple ecotoxicological profiles in agrochemical fungicides: A multi-species chemoinformatic approach, Ecotoxicol. Environ. Saf. 80 (2012) 308-313.

[37] F.J. Prado-Prado, F. Borges, L.G. Perez-Montoto, H. Gonzalez-Diaz, Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species, Eur. J. Med. Chem. 44 (2009) 4051-4056.

[38] A.A. Toropov, E. Benfenati, QSAR models of quail dietary toxicity based on the graph of atomic orbitals, Bioorg. Med. Chem. Lett. 6 (2006) 1941-1943.

[39] A. Schrage, K. Hempel, M. Schulz, S.N. Kolle, B. Van Ravenzwaay, R. Landsiedel, Refinement and Reduction of Acute Oral Toxicity Testing: A Critical Review of the Use of Cytotoxicity Data, Altern. Lab. Anim. 39 (2011) 273–295.

[40] M.T.D. Cronin, T.W. Schultz, Pitfalls in QSAR, J. Mol. Struct. 622 (2003) 39-51.

ACCEPTED MANUSCRIPT

[41] PPDB (Pesticide Properties DataBase), *http:// sitem.herts.ac.uk/aeru/footprint/* (accessed 14/05/2014).
[42] C. Nantasenamat, T. Tantimongcolwat, T. Naenna, C. Isarankura-Na-Ayudhya, V. Prachayasittikul, Prediction of selectivity index of pentachlorophenol-imprinted polymers, EXCLI J. 5 (2006) 150-163.
[43] D.A. Konovalov, N. Sim, E. Deconinck, Y.V. Heyden, D. Coomans, Statistical confidence for variable selection in QSAR models, J. Chem. Inf. Model. 48 (2008) 370-383.
[44] A. Srinivas Reddy, S. Kumar, R. Garg, Hybrid-genetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of tipranavir analogs for HIV protease inhibition, J. Mol. Graph. Model. 28 (2010) 852-862.
[45] L. Xu, W.J. Zhang, Comparison of different methods for variable selection, Anal. Chim. Acta 446 (2001) 477-483.
[46] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, QSAR Comb. Sci. 22 (2003) 69–77.
[47] D.M. Hawkins, The problem of overfitting, J. Chem. Inf. Comput. Sci. 44 (2004) 1-12.
[48] Q. Shen, J.H. Jiang, C.X. Jiao, W.Q. Lin, G.L. Shen, R.Q. Yu, Hybridized particle swarm algorithm for adaptive structure training of multilayer feed-forward neural net-work: QSAR studies of bioactivity of organic compounds, J. Comput. Chem. 25 (2004) 1726-1735.
[49] P. Izadyan, M.H. Fatemi, M. Izadyan, Elicitation of the most important structural properties of ionic liquids affecting ecotoxicity in limnic green algae; a QSAR approach, Ecotoxicol. Environ. Saf. 87 (2013) 42-48.
[50] O. Prakash, F. Khan, R.S. Sangwan, L. Misra, ANN-QSAR model for virtual screening of androstenedione C-skeleton containing phytomolecules and analogues for cytotoxic activity against human breast cancer cell line MCF-7, Comb. Chem. High Throughput Screen. 16 (2013) 57-72.
[51] R. Wang, J. Jiang, Y. Pan, H. Cao, Yi. Cui, Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices, J. Hazard. Mater. 166 (2009) 155–186.
[52] P.K. Ojha, I. Mitra, R.N. Das, K. Roy, Further exploring rm2 metrics for validation of QSPR models, Chemometr. Intell. Lab. 107 (2011) 194-205.
[53] I. Mitra, A. Saha, K. Roy, Chemometric QSAR modeling and in silico design of antioxidant NO donor phenols, Sci. Pharm. 79 (2011) 31–57.
[54] J.W. Chen, X.H. Li, H.Y. Yu, Y.N. Wang, X.L. Qiao, Progress and perspectives of quantitative structure-activity relationships used for ecological risk assessment of toxic organic compounds, Sci. China Ser. B-Chem. 51 (2008) 593-606.
[55] OECD principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models, 2009.
[56] K. Roy, S. Kar, P. Ambure, On a simple approach for determining applicability domain of QSAR models, Chemometr. Intell. Lab. 145 (2015) 22-29.
[57] E.M. De Haas, T. Eikelboom, T. Bouwman, Internal and external validation of the long-term QSARs for neutral organics to fish from ECOSAR, SAR QSAR Environ. Res. 22 (2011) 545–559.
[58] T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D.W. Roberts, T.W. Schultz, D.T. Stanton, J.J.M. Van De Sandt, W. Tong, G. Veith, C. Yang, Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships, Altern. Lab. Anim. 33 (2005) 155–173.
[59] A. Habibi-Yangjeh, M. Danandeh-Jenagharad, Application of a genetic algorithm and an artificial neural network for global prediction of the toxicity of phenols to Tetrahymena pyriformis, Monatsh Chem. 140 (2009) 1279–1288.
[60] G. Tugcu, M.S. Türker, M. Vracko, M. Novic, N. Minovski, QSTR modelling of the acute toxicity of pharmaceuticals to fish, SAR QSAR Environ. Res. 23 (2012) 297-310.
[61] H. Du, J. Wang, Z. Hu, X. Yao, X. Zhang, Prediction of fungicidal activities of rice blast disease based on least-squares support vector machines and project pursuit regression, J. Agric. Food Chem. 56 (2008) 10785-10792.
[62] P.R. Duchowicz, J. Marrugo, J.H Erlinda, V. Ortiz, E.A. Castro, R. Vivas-Reyes, QSAR study for the fish toxicity of benzene derivatives, J. Argentine Chem. Soc. 97 (2009) 116-127
[63] N. Tan, P. Li, H. Rao, Z. Li, X. Li, Prediction of the acute toxicity of chemical compounds to the fathead minnow by machine learning approaches, Chemom. Intell. Lab. Syst. 100 (2010) 66-73

ACCEPTED MANUSCRIPT

[64] J. Xu, L. Zhu, D. Fang, L. Wang, S. Xiao, Li. Liu, W. Xu, QSPR studies of impact sensitivity of nitro energetic compounds using three-dimensional descriptors, J. Mol. Graph. Model. 36 (2012) 10-19.
[65] G. D. Garson, Interpreting neural network connection weights, Artif. Intell. Expert 6 (1991) 47-51.
[66] A.T. Goh, Back-propagation neural networks for modeling complex systems, Artif. Intell. Eng. 9 (1995) 143-151.
[67] J.D. Gough, L.H. Hall, Modeling the toxicity of amide herbicides using the electrotopological state, Environ. Toxicol. Chem. 18 (1999) 1069–1075.
[68] D. Zakarya, E.M. Larfaoui, A. Boulaamail, T. Lakhlifi, Analysis of structure–toxicity relationships for a series of amide herbicides using statistical methods and neural network, SAR QSAR Environ. Res. 5 (1996) 269–279.
[69] A. Can, I. Yildiz, G. Guvendik, The determination of toxicities of sulphonylurea and phenylurea herbicides with quantitative structure–toxicity relationship (QSTR) studies, Environ. Toxicol. Pharmacol. 35 (2013) 369–379.
[70] M. Nendza, B. Dittrich, A. Wenzel, W. Klein, Predictive QSAR models estimating ecotoxic hazard of plant protecting agents: target and non-target toxicity, Sci. Total Environ. 109-110 (1991) 527–535.
[71] J. Devillers, Prediction of mammalian toxicity of organophosphorus pesticides from QSTR modeling, SAR QSAR Environ. Res. 15 (2004) 501–510.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

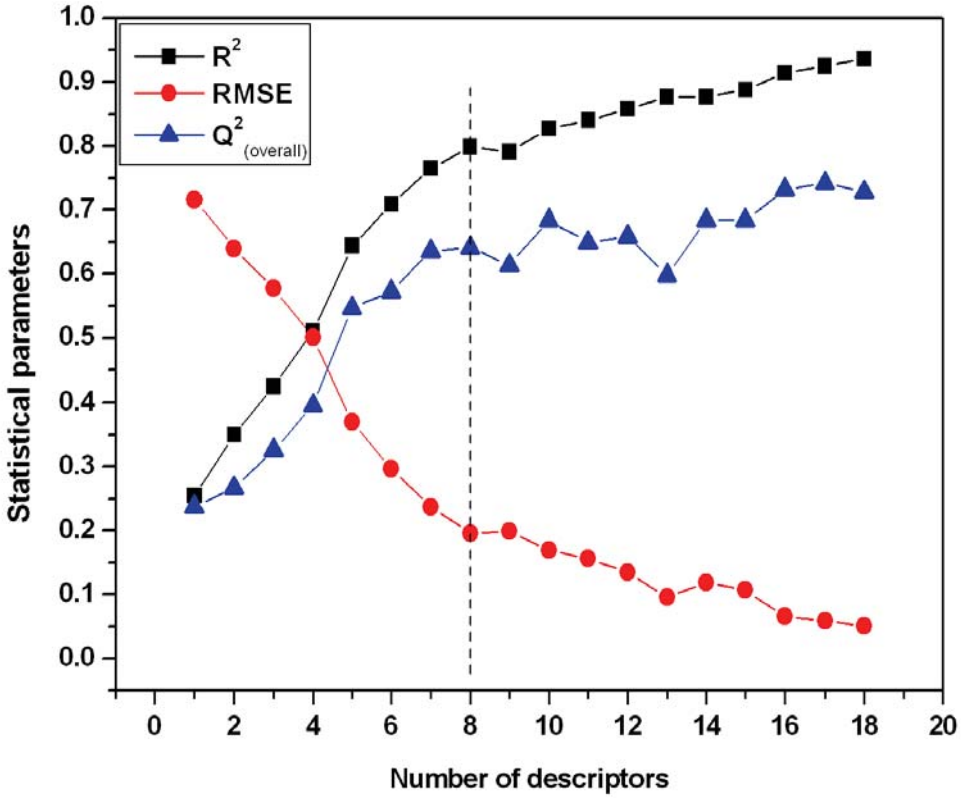


Figure 1

ACCEPT

ACCEPTED MANUSCRIPT

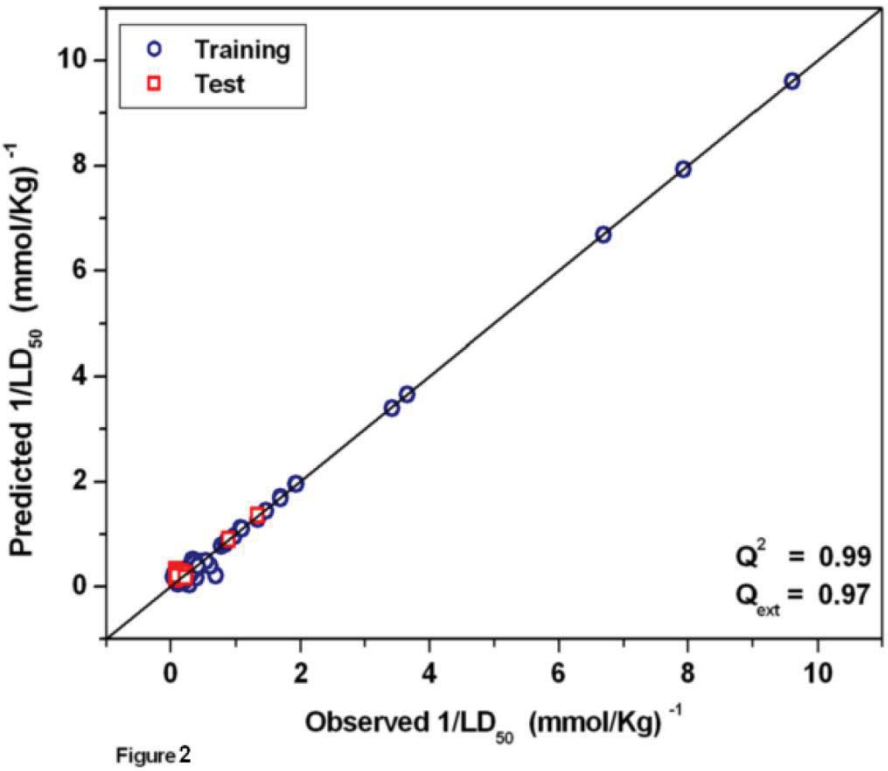


Figure 2

ACCEPTED

ACCEPTED MANUSCRIPT

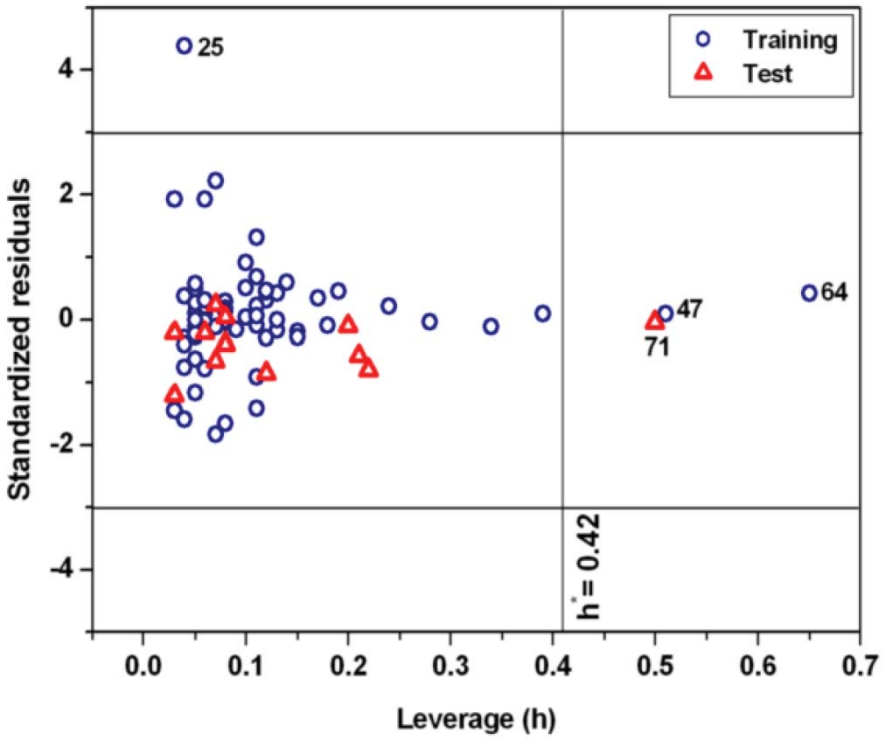


Figure 3

ACCEPTED

ACCEPTED MANUSCRIPT

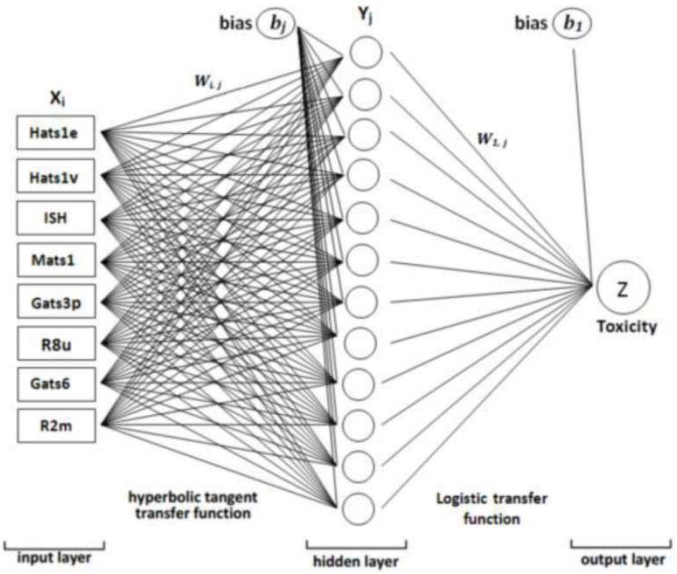


Figure 4

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

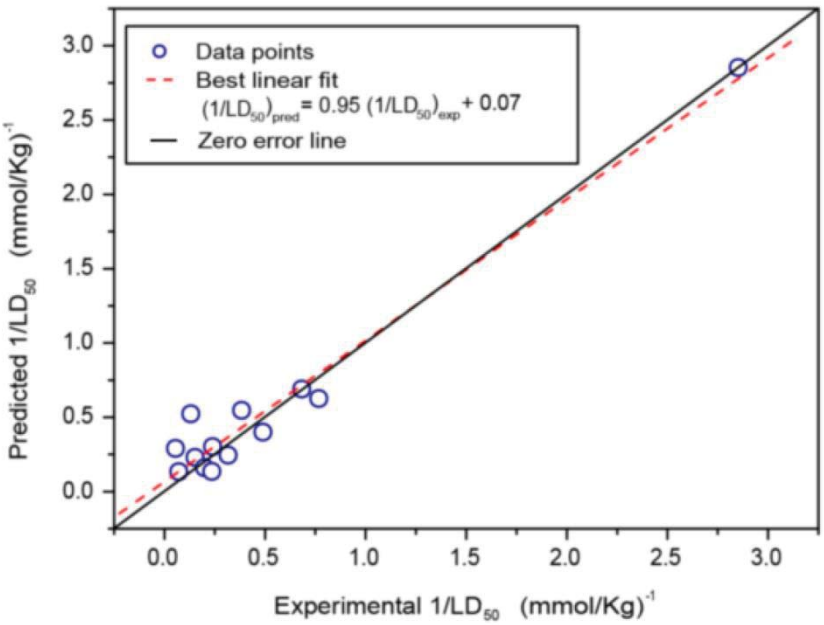


Figure 5

ACCEPTED

ACCEPTED MANUSCRIPT

Table 1
Observed (experimental) values and Artificial Neural Predicted values of oral rat acute toxicity and leverage values for 77 Herbicides.

N°	Compound	CAS number	$1/LD_{50}$ (mmol/Kg)$^{-1}$		Leverage (hi)
			Observed	Predicted	
Training set					
1	2.4-dichlorophenoxyacetic acid	94-75-7	0.47	0.47	0.10
2	2.4-DB	94-82-6	0.28	0.37	0.06
3	4-chlorophenoxy)acetic acid	122-88-3	0.22	0.15	0.14
4	Acifluorfen	50594-66-6	0.26	0.25	0.06
5	Acrolein	107-02-8	1.93	1.95	0.34
6	Alachlor	15972-60-8	0.29	0.23	0.10
7	Aminocyclopyrachlor	858956-08-8	1.94	1.96	0.15
8	Amiprofos-methyl	36001-88-4	0.98	0.95	0.12
9	Anilofos	64249-01-0	0.78	0.78	0.08
10	Aziprotryn	4658-28-0	0.08	0.11	0.12
11	Bensulide	741-58-2	1.47	1.44	0.08
12	Bentazone	25057-89-0	0.48	0.48	0.08
13	Bromoxynil	1689-84-5	3.42	3.39	0.24
14	Bromoxynil heptanoate	56634-95-8	1.34	1.29	0.12
15	Bromoxynil octanoate	1689-99-2	1.69	1.71	0.13
16	Butamifos	36335-67-8	0.53	0.49	0.17
17	Butraline	33629-47-9	0.28	0.14	0.11
18	Chlorthiamide	1918-13-4	0.27	0.20	0.11
19	Clomazone	81777-89-1	0.18	0.22	0.04
20	Cumyluron	99485-76-4	0.32	0.30	0.05
21	Cyanazine	21725-46-2	0.84	0.81	0.06
22	Dicamba	1918-00-9	0.14	0.17	0.15
23	Dichlorprop	120-36-5	0.28	0.04	0.07
24	Difenamide	957-51-7	0.25	0.23	0.08
25	Dimethenamid	87674-68-8	0.69	0.21	0.04
26	Dimethipin	55290-64-7	0.46	0.41	0.13
27	Dinosebe	88-85-7	9.61	9.61	0.13
28	Diquat	2764-72-9	0.86	0.87	0.11
29	DNOC	534-52-1	7.93	7.93	0.28
30	Endothal	145-73-3	3.65	3.66	0.07
31	Ethyl dipropylthiocarbamate	759-94-4	0.21	0.22	0.09
32	Fluazifop-p-butyl	79241-46-6	0.16	0.16	0.06
33	Fluchloraline	33245-39-5	0.23	0.26	0.05
34	Flufenacet	142459-58-3	0.61	0.39	0.03
35	Fomesafen	72178-02-0	0.35	0.43	0.04

ACCEPTED MANUSCRIPT

Table 1 (continued)

N°	Compound	CAS number	$1/LD_{50}$ (mmol/Kg)$^{-1}$		Leverage (hi)
			Observed	Predicted	
Training set (continued)					
36	Glufosinate	51276-47-2	0.11	0.06	0.19
37	Haloxyfop	69806-34-4	1.07	1.11	0.04
38	Hexazinone	51235-04-2	0.15	0.12	0.11
39	Isoproturon	34123-59-6	0.11	0.18	0.05
40	MCPA	94-74-6	0.21	0.11	0.10
41	MCPA-thioethyl	25319-90-8	0.54	0.49	0.05
42	MCPB	94-81-5	0.05	0.21	0.03
43	Metamitrone	41394-05-2	0.17	0.13	0.04
44	Methazole	20354-26-1	0.34	0.51	0.04
45	Metribuzine	21087-64-9	6.70	6.69	0.51
46	Molinate	2212-67-1	0.39	0.17	0.06
47	Monuron	150-68-5	0.19	0.20	0.05
48	Naproanilide	52570-16-8	0.11	0.24	0.05
49	Naptalame	132-66-1	0.16	0.21	0.04
50	Nitrofene	1836-75-5	0.11	0.08	0.05
51	Paraquat	4685-14-7	1.69	1.68	0.39
52	Pebulate	1114-71-2	0.18	0.21	0.05
53	Pethoxamid	106700-29-2	0.30	0.24	0.05
54	Piperophos	24151-93-7	1.09	1.10	0.18
55	Pretilachlore	51218-49-6	0.05	0.23	0.08
56	Propyrisulfuron	570415-88-2	0.12	0.32	0.07
57	Prosulfocarbe	52888-80-9	0.14	0.13	0.11
58	Pyrazoxyfene	71561-11-0	0.25	0.25	0.05
59	Simetryn	1014-70-6	0.28	0.29	0.12
60	Tebutam	35256-85-0	0.04	0.13	0.11
61	Tebuthiuron	34014-18-1	0.35	0.37	0.11
62	Tralkoxydim	87820-88-0	0.35	0.35	0.65
63	Triclopyr	55335-06-3	0.41	0.39	0.08
64	Vernolate	50471-44-8	0.14	0.10	0.06
Test set					
65	Ancymidole	12771-68-5	0.15	0.19	0.03
66	Butachlor	23184-66-9	0.16	0.23	0.08
67	Butroxydim	138164-12-2	0.24	0.24	0.08
68	Dimetachlor	50563-36-5	0.16	0.29	0.07

ACCEPTED MANUSCRIPT

Table 1 (continued)

N°	Compound	CAS number	$1/LD_{50}$ (mmol/Kg)$^{-1}$		Leverage (hi)
			Observed	Predicted	
Test set (continued)					
69	Dimexano	1468-37-7	0.89	0.90	0.50
70	Fluazolate	174514-07-9	0.09	0.33	0.03
71	Mecoprop	7085-19-0	0.18	0.26	0.08
72	Mepiquat	15302-91-7	0.08	0.24	0.22
73	Monalide	7287-36-7	0.09	0.13	0.06
74	Pendimethaline	40487-42-1	0.09	0.26	0.12
75	Prometone	1610-18-0	0.10	0.21	0.21
76	Propanil	709-98-8	0.23	0.18	0.07
77	Tribufos	52-68-6	1.34	1.36	0.20

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2
Performance of MLP-ANN model for herbicides

Statistical parameters	Training set	Test set	Overall set
n	64	13	77
RMS	0.102	0.111	0.104
R^2	0.996	0.956	0.996
Q^2_{LOO}	0.996	--	--
Q^2_{test}	--	0.977	--
r_m^2	0.990	0.782	0.993
$r_m'^2$	0.992	0.724	0.991
$\overline{r_m^2}$	0.991	0.753	0.992
Δr_m^2	0.002	0.058	0.002

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3

Comparison of the results of internal and external validation of our model with previously published models

Models	Internal validation						External validation					
	n	R^2	RMS	Q^2_{LOO}	$\overline{r^2_m}$	Δr^2_m	n	R^2_{ext}	RMS	Q^2_{ext}	$\overline{r^2_m}$	Δr^2_m
Present work	**64**	**0.996**	**0.102**	**0.996**	**0.991**	**0.002**	**13**	**0.956**	**0.111**	**0.977**	**0.753**	**0.058**
Can et al. [68]	20	0.931	0.041				7	0.682				
Devillers [69]	51		0.290				9		0.260			
Gough et al. [70]	50	0.750	0.230				9		0.270			
Zakarya et al. [71]	44	0.850					The external validation was not used					
Nendza et al. [72]	12	0.880	0.170				The external validation was not used					

ACCEPTED MANUSCRIPT

Table 4

Observed (experimental) values of $1/LD_{50}$ and those calculated by Equation (8) for 13 herbicides.

Herbicide	Expérimental $1/DL_{50}$ $(mmol/Kg)^{-1}$	Calculated $1/DL_{50}$ $(mmol/Kg)^{-1}$
2, 4, 5-trichlorophenol	0,24	0.30
Barbane	0,49	0.40
Clethodim	0,32	0.24
Cycluron	0,13	0.52
Desmetryne	0,15	0.23
Diallate	0,68	0.69
Florasulam	0,07	0.13
Halosulfuron-methyl	0,06	0.29
Ioxynil	2,85	2.85
Metolachlor	0,24	0.14
Phenolthiol	0,54	0,35
Propachlore	0,39	0.55
Prosulfuron	0,79	0.62

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 5

Weights and bias between input and hidden layers

j	$W_{j,1}$	$W_{j,2}$	$W_{j,3}$	$W_{j,4}$	$W_{j,5}$	$W_{j,6}$	$W_{j,7}$	$W_{j,8}$	b_j
1	-12,5623	10,4659	6,8128	4,5163	-7,1717	-1,3836	10,555	-1,2162	-0,117
2	2,3897	-2,1796	-0,425	-0,7403	0,8791	-0,7835	-0,7414	-0,13	0,3378
3	12,0091	-8,6073	-3,7229	-4,7296	7,8074	0,5646	-10,1373	0,6098	0,8454
4	7,5259	-7,9819	-3,4047	-2,1069	7,7919	0,3547	-5,1061	3,1768	0,1007
5	5,0956	-4,3988	-2,4442	-1,9796	2,9182	0,0413	-3,9515	0,0441	-0,0232
6	-6,0043	2,8097	3,5531	1,4624	-2,8002	-0,8917	5,1207	-1,1257	0,9697
7	12,3996	-7,2041	-4,7913	-4,5002	8,5899	0,8677	-8,8004	1,0754	1,4131
8	8,7480	-8,0968	-4,0673	-4,0747	4,1031	-0,4312	-5,3961	0,1292	0,3236
9	10,4868	-7,1432	-2,4627	-3,7122	4,4835	0,8243	-8,4937	-0,0287	0,0351
10	-11,7132	5,1074	5,5185	3,5189	-6,0079	-0,9015	9,4268	-1,8093	1,5312
11	7,0875	-1,0575	-2,0678	-3,0068	4,3843	0,4329	-5,1146	-0,6053	2,0435
12	7,2244	-5,1105	-3,5352	-2,5464	4,2691	0,3826	-5,8234	0,4035	0,0393

ACCEPTED

ACCEPTED MANUSCRIPT

Table 6

Weights and bias between hidden and output layers

$W_{1,1}$	$W_{1,2}$	$W_{1,3}$	$W_{1,4}$	$W_{1,5}$	$W_{1,6}$	$W_{1,7}$	$W_{1,8}$	$W_{1,9}$	$W_{1,10}$	$W_{1,11}$	$W_{1,12}$
-1,3592	-0,5671	3,9907	-5,3525	-0,0845	-2,0711	-4,9871	-2,5635	2,9594	4,6692	5,7823	0,2281
b_1 = -2,0559											

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights

Herbicides can be dangerous to the environment and the human health.

The risk assessment of herbicides is crucial

QSAR model for acute oral toxicity of herbicides is developed and proposed.

This model has been developed and validated on the basis of the OECD principles.

Artificial neural network-based equation to predict the toxicity of herbicides was established

ACCEPTED MANUSCRIPT