Rocznik Ochrona Środowiska

Volume 26 Year 2024 ISSN 2720-7501 pp. 264-272

https://doi.org/10.54740/ros.2024.026 open access

Received: October 2023 Accepted: May 2024 Published: July 2024

Regression Quantitative Structure-toxicity Relationship of Pesticides on Fishes

Bowen Yang[1], Limin Dang[2], Cong Chen[3], Mingwang Li[4]*

[1]*College of Materials and Chemical Engineering, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China*

[2]*College of Materials and Chemical Engineering, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China*

https://orcid.org/0000-0001-7625-1410

[3]*College of Materials and Chemical Engineering, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China*

[4]*College of Materials and Chemical Engineering, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China*

[*]*corresponding author's e-mail: l3055155812@163.com*

Abstract: Pesticide usage reaches several million metric tons annually worldwide, and the effects of pesticides on non-target species, such as various fishes in aquatic environments, have resulted in serious concerns. Predicting pesticide aquatic toxicity to fish is of great significance. In this paper, 20 molecular descriptors were successfully used to develop a regression quantitative structure-activity/toxicity relationship (QSAR/QSTR) model for the toxicity $\log LC_{50}$ of a large data set consisting of 1106 pesticides on fishes by using a general regression neural network (GRNN) algorithm. The optimal GRNN model produced correlation coefficients R of 0.8901 (*rms* = 0.6910) for the training set, 0.8531 (*rms* = 0.7486) for the validation set, and 0.8802 (*rms* = 0.6903) for the test set, which are satisfactory compared with other models in the literature, although a large data set of toxicity $\log LC_{50}$ was used in this work.

Keywords: Toxicity; pesticide; QSTR; general regression neural network

1. Introduction

Modern agricultural, residential, commercial and industrial settings are increasingly relying on the use of pesticides such as herbicides, insecticides, nematicides, and fungicides in protecting crops, plants and public health and in controlling overgrowth of insects, fungi, rodents, noxious weeds, etc. Nearly 2.7 million tons of pesticides are used in global agricultural production annually (Isah et al. 2020, Yu & Zeng 2022). The effects of pesticides on non-target species, such as man and aquatic organisms, have resulted in severe concerns (Mo et al. 2022).

Performing a pesticide risk assessment is essential to provide a precaution against environmental pollution. In determining the acute toxicity of pesticides, fish are usually used as laboratory animals. Unfortunately, the experimental tests for acute toxicity to fish are expensive and time-consuming (Yu 2020a, Yu 2021). A quantitative structure-activity/toxicity relationship (QSAR/QSTR) model, being a rapid, cost-effective and ethical alternative, can be used for predicting chemical toxicity (Sullivan et al. 2014, Mit et al. 2022, Masand et al. 2021, Fang et al. 2022), even for chemicals without being synthetized. This methodology is proposed by EU REACH Legislation, ICH M7 guideline, the US FDA and the US EPA to assess the environmental risks of a chemical (Cachot 2014, Schmidt et al. 2021). Some QSTR models have been reported on pesticide aquatic toxicity to fishes.

Toropov et al. (2020) introduced QSTR models for 311 acute toxicity data (pLC_{50}) to Rainbow Trout with the index of ideal correlations. The models have coefficients of determination R^2 being 0.81-0.86 and root mean square (*rms*) errors of 0.55-0.65 for the validation set.

Pandey et al. (2020) considered QSAR modelling of 85 acute fish toxicity (pLC_{50}) of environmental transformation products of pesticides using ten simple 2D descriptors and partial least squares regression. The training and test sets have R^2 higher than 0.73 and mean absolute errors (MAE) lower than 0.57.

Jia et al. (2020) proposed a linear QSAR model for aquatic toxicity (pLC_{50}) of 311 pesticides on Rainbow Trout with molecular weight and 27 norm indices. The model has a coefficient of determination R^2 higher than 0.80 for the training set (249 samples) and test set (62 samples).

Galimberti et al. (2020) established linear QSAR models for small pesticide toxicity Log(EC50) data sets for *Pimephales promelas* and *Oncorhynchus mykiss*. The two models have 12 samples and three descriptors, yielding R^2 of 0.96 and MAE higher than 0.20. However, for linear QSARs, the ratios of the numbers of samples to descriptors are generally greater than 5.

The QSAR models mentioned above focus on a particular fish species and have relatively small data sets of pesticide toxicity to fishes. Thus, these models possess some limitations in application. Li et al. (2017) and Yu & Zeng (2022) reported classification models for large data sets of pesticide toxicity to various fish species. These classification models have a larger applicability domain in predicting the toxic categories of pesticides.

In predicting the physicochemical properties of compounds, QSAR models based on regression analysis are more accurate than classification models, although developing regression models is more difficult than building classification models. This work aims to establish a regression QSAR model for a large data set, including 1106 pesticide toxicity to fishes, by using a general regression neural network (GRNN) algorithm.

2. Materials and Methods

2.1. Data set

Table S1 in Supplementary Material shows 1106 toxicity data (96 *h*, LC50) of organic pesticides on fish species, including *Oncorhynchus mykiss*, *Lepomis macrochirus*, *Pimephales promelas*, *Brachydanio rerio*, *Cyprinodon*, *Cyprinus carpio*, etc., which were reported in the literature (Li et al. 2017, Yu & Zeng 2022). The toxicity data (96 *h*, logLC_{50}) lie from −4.4559 to 4.7324 mg/L. For pesticides of approximately equal molecular weight, a smaller logLC_{50} value suggests the corresponding pesticide molecule possesses higher toxicity to fish. The total data set (1106 organic pesticides) was randomly divided into three sets at the ratio of 70%:15%:15%, which were, respectively, used as the training set (Nos. 1-774 in Table S1 in Supplementary Material), the validation set (Nos. 775-940 in Table S1) and the test set (Nos. 941-1106 in Table S1). QSAR models of logLC_{50} were established with the training set by tuning the model parameters with the validation set. Subsequently, the models were assessed with the test set (Golmohammadi & Safdari 2010).

2.2. Descriptors derivation

The molecular structures were constructed with KingDraw (http://kingdraw.cn/en/index.html) and then optimized with the AM1 method in Gaussian 09 (Revision A.02). Subsequently, these molecules were used as input files for Dragon 6.0 (Talete srl, 2012) to obtain molecular descriptors. After removing those descriptors being a constant or approximately equaling to a constant or whose partial correlation coefficients > 0.90, 773 descriptors were retained for descriptor selection in the next steps (Yu 2023).

2.3. GRNN principle

GRNN can successfully deal with classification and regression prediction by introducing the nonparametric strategy based on Parzen window (Yu 2020a). As is shown in Fig. 1, it consists of four layers: input layer, pattern layer, summation layer and output layer. For the input layer, the number of neurons equals to the dimension of the input vector in the training set. For the pattern layer, the number of neurons equals the number of samples. A transfer function of the *i*th neuron is used to correlate its output with the input variable X and the learning sample X_i, by calculating their Euclid distance:

$$p_i = \exp[-(X - X_i)^T (X - X_i)] / 2\sigma^2] \qquad i = 1, 2, \cdots, n \tag{1}$$

where σ is the SPREAD parameter of the Gaussian function and needs to be adjusted by users.

In the summation layer, two types of neurons are used in summation. One is the denominator node S_A, and the other is the numerator node S_{Nj}. The former is based on an arithmetic sum for the output from the neurons in the pattern layer by setting the connection weights of 1:

$$S_A = \sum_i^n P_i \tag{2}$$

The latter is used for weighted summation with the connection weight *yij* associating the *i*th neuron in the pattern layer with the *j*th neuron in the summation layer:

$$S_{Nj} = \sum_i^n y_{ij} P_i \quad j = 1, 2, \cdots, k \tag{3}$$

In the output layer, the prediction results can be obtained with:

$$y_j = S_{Nj} / S_A \quad j = 1, 2, \cdots, k \tag{4}$$

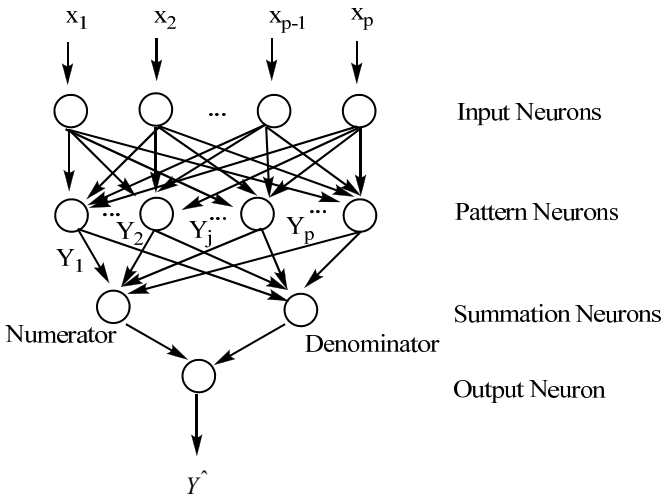

Fig. 1. Network structure of GRNN algorithm

3. Results and Discussion

3.1. Descriptors and toxicity mechanism

Stepwise multiple regression (MLR) analysis in SPSS 19.0 was carried out for 1106 $\log LC_{50}$ of pesticides on fish and 773 molecular descriptors mentioned above. In total, 28 descriptors entered an MLR model when the increment of determination coefficient (ΔR^2) $\geq$ 0.04 was set as the criterion for adding new descriptors. Then, MLR analysis was performed for the 28 descriptors and 774 $\log LC_{50}$ in the training set with the same criterion in introducing new descriptors. In the end, 20 molecular descriptors were obtained and taken as the optimal descriptor subset for developing QSAR models for $\log LC_{50}$ of pesticides on fishes. The physical meaning and toxicity mechanism of descriptors are listed in Table 1, their values are shown in Table S1 in Supplementary Material, and the characteristics of molecular descriptors obtained from the total set are shown in Table 2.

Table 1. The block and physical meaning of descriptors in the GRNN model

Descriptor	Block	Physical meaning and toxicity mechanism
MLOGP2	Molecular properties	MLOGP2 denotes the squared Moriguchi octanol-water partition coefficient. It is related to frequencies of presence (or absence) of some molecular features such as carbon and halogen atoms. A pesticide molecule with a larger MLOGP2 tends to bind lipophilic chemicals and to accumulate in fishes and cause toxicity.
Eig02_AEA(dm)	Edge adjacency indices	It is eigenvalue no. 2 from the augmented edge adjacency matrix weighted by dipole moment and reflects information about edge connectivity in the H-depleted molecular graph. It is related to molecular bond types, group polarity, and molecular size. A larger Eig02_AEA(dm) indicates the molecule has more reaction or binding sites, resulting in toxicity.
CATS2D_09_DD	CATS 2D descriptors	It means CATS2D Donor-Donor at lag 09. Similar to 2D Atom Pairs, CATS 2D descriptors reflect molecular features about potential pharmacophore points, including hydrogen-bond donor (D), hydrogen-bond acceptor (A), positively charged (P), negatively charged (N), and lipophilic (L), in topological distances of 0-9 bonds. A molecule with CATS2D_09_DD>0 means that it has a greater possibility of forming hydrogen bonds, which are related to molecular solubility in the water environment.

Table 1. cont.

Descriptor	Block	Physical meaning and toxicity mechanism
O-058	Atom-centered fragments	It denotes a molecule's number of specific atom types (=O). Ketones (excluding α,β–unsaturated ketones), being non-polar narcosis chemicals, usually have lower toxicity to fish and larger O-058 values.
nR03	Ring descriptors	nR03 is the number of three-member rings. As is known, the three-member ring is less stable than the four-, five-, and six-member rings. Thus, these molecules with three-member rings have higher unspecific reactivity and toxicity.
nRCOOR	Functional group counts	It is the number of esters (aliphatic). It is known that an ester molecule with α-position of a double or triple bond can undergo a Michael type addition of nucleophiles and lead to higher toxicity.
nROR	Functional group counts	nROR is the number of ethers (aliphatic). Generally, linear ethers or monocyclic mono-ethers (excluding epoxides or peroxides) are type narcosis or baseline toxicity.
nCXr=	Functional group counts	It is the number of X on ring C (sp2). Similar to the descriptor nRCOOR, the molecules have unspecific reactivity and possess relatively high toxicity when the substituent X is a good leaving group (e.g. halogen or hydroxyl group).
nArOCON	Functional group counts	nArOCON denotes the number of (thio-) carbmates (aromatic). The molecules with nArOCON groups may undergo the same reaction mechanism as those having nRCOOR and nCXr= groups.
MW	Constitutional descriptors	MW denotes the molecular weight. On the one hand, molecular size influences molecules penetrating the two phospholipid bilayers of the cell membrane. On the other hand, a molecule with a large MW may have more reaction or binding sites, and resulting in toxicity.
H-049	Atom-centered fragments	H-049 equals the number of H attached to C^3(sp3) / C^2(sp2) / C^3(sp2) / C^3(sp) (the superscript represents the formal oxidation number). A molecule with a large The h-049 value may be conducive to forming hydrogen bonds and improving its solubility in the water environment, resulting in low toxicity.
B07[N-N]	2D Atom Pairs	It denotes the presence/absence of N-N at topological distance 7. Similar to the descriptor H-049, the molecules with B07[N-N] >0 (i.e., the presence of N-N at topological distance 7) generally have –NH– group forming hydrogen bonds.
nROH	Functional group counts	nROH is the number of hydroxyl groups. Similar to the descriptors H-049 and B07[N-N], nROH reflects the ability to form hydrogen bonds.

Table 1. cont.

Descriptor	Block	Physical meaning and toxicity mechanism
CATS2D_01_DA	CATS 2D descriptors	It is CATS 2D Donor-Acceptor at lag 1 and can describe compounds with special groups such as >NNH– and –ONH–. The appearance of these groups is conducive to forming hydrogen bonds and to reducing toxicity to fish.
Eta_betaS_A	ETA indices	It is an extended topochemical atom (ETA) index calculated with a sigma average VEM count. This descriptor is related to molecular bulk. As is shown in Table S1, the molecules with large Eta_betaS_A values possess ring groups. Compounds with a leaving group at an α-position of an aromatic bond have benzylic activation, although ring groups may influence molecular penetrability through the cell membrane.
O%	Constitutional descriptors	O% is the per cent of O atoms in a molecule. The compounds containing only C and H (or halogens) have lower O% values and possess type narcosis baseline toxicity. On the other hand, a molecule with large O% indicates that it may possess more flexible single bonds, which is beneficial for molecules to penetrate cell membranes.
DLS_03	Drug-like indices	It is a modified drug-like score from Walters et al. (6 rules). It is related to Moriguchi's logP, the number of H-bond donors/ acceptors, rotatable bond number and molecular weight. A pesticide molecule with a larger DLS_03 has a smaller LC50 value and higher toxicity to fish.
B03[N-P]	2D Atom Pairs	It denotes the presence/absence of N-P at topological distance 3. Organic phosphorus pesticides with relatively high 03[N-P] values can inhibit acetylcholinesterase in vivo and belong to specifically acting chemicals, resulting in high toxicity to fishes.
CATS2D_06_PL	CATS 2D descriptors	CATS2D Positive-Lipophilic at lag 06. On the one hand, the molecules with CATS2D_06_PL >0 have positively charged groups and yield strong polarity, enhancing the solubility of pesticides in the water environment. On the other hand, lipophilic groups in a pesticide molecule can reduce its solubility and lead to high toxicity to fish.
SaasN	Atom-type E-state indices	It is the sum of aasN E-states. Table S1 shows that the pesticide molecules (e.g. Nos. 17, 160, 504, and 748) have large SaasN values and low $\log LC_{50}$. These molecules with special groups, such as triazole and imidazole, can cause toxicity to fish, although the mechanisms of toxicity are complicated (DeLorenzo et al. 2001).

Table 2. Characteristics of molecular descriptors used

Descriptor	Coefficients	Std. Error	t-test	Sig.	VIF
Constant	1.065	0.493	2.160	0.031	/
MLOGP2	-0.016	0.006	-2.596	0.010	2.500
Eig02_AEA(dm)	-0.363	0.061	-5.927	0.000	2.438
CATS2D_09_DD	-0.874	0.238	-3.675	0.000	1.162
O-058	0.578	0.041	14.230	0.000	2.002
nR03	-1.095	0.147	-7.440	0.000	1.169
nRCOOR	-0.550	0.096	-5.749	0.000	1.376
nROR	0.330	0.055	6.030	0.000	1.262
nCXr=	-0.435	0.085	-5.086	0.000	1.048
nArOCON	-0.459	0.142	-3.227	0.001	1.049
MW	-0.004	0.001	-7.389	0.000	4.329
H-049	0.267	0.067	3.997	0.000	1.443
B07[N-N]	0.618	0.175	3.539	0.000	1.181
nROH	0.716	0.070	10.263	0.000	1.573
CATS2D_01_DA	0.636	0.177	3.597	0.000	1.023
Eta_betaS_A	4.616	0.784	5.889	0.000	1.747
O%	-0.034	0.006	-5.355	0.000	1.706
DLS_03	-1.225	0.367	-3.340	0.001	1.295
B03[N-P]	-1.384	0.245	-5.638	0.000	1.271
CATS2D_06_PL	0.259	0.083	3.112	0.002	1.303
SaasN	-0.332	0.100	-3.304	0.001	1.414

3.2. GRNN model

The 20 molecular descriptors in the optimal descriptor subset were used as independent variables, and the toxicity $\log LC_{50}$ was used as the dependent variable to develop GRNN models from 774 pesticides in the training set by applying MATLAB R2014a. The spread parameter σ varying in the range of 0.01-0.15 with the step of 0.01 resulted in *rms* errors for the validation set, which are depicted in Fig. 2. As is shown, the optimal GRNN model with the SPREAD parameter σ being 0.11 has the minimum *rms* error of 0.7486 (log units). Then, 166 $\log LC_{50}$ of pesticides on fishes in the test set were adopted to assess the optimal GRNN model ($\sigma = 0.11$). The calculated $\log LC_{50}$ values are listed in Table S1 in the Supplemental file and shown in Fig. 3.

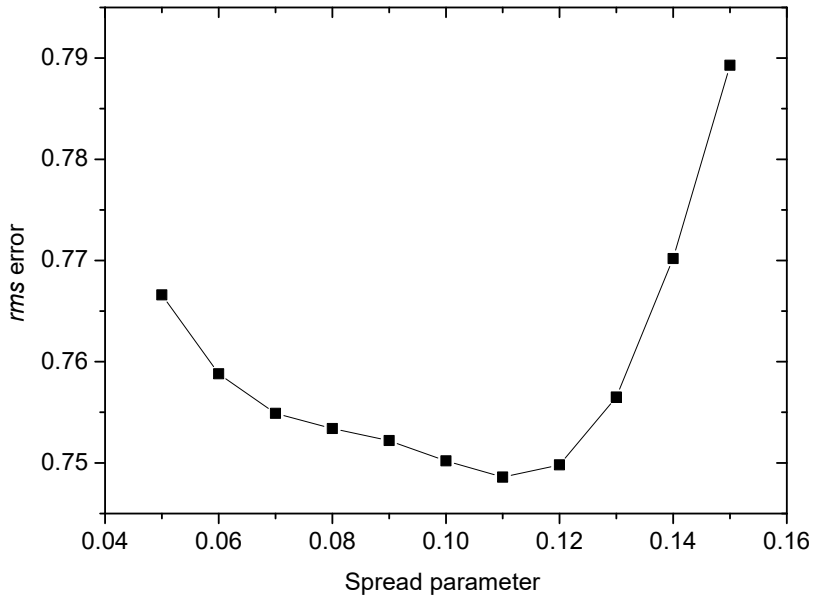

Fig. 2. Relationship between spread parameters and rms errors in the validation set

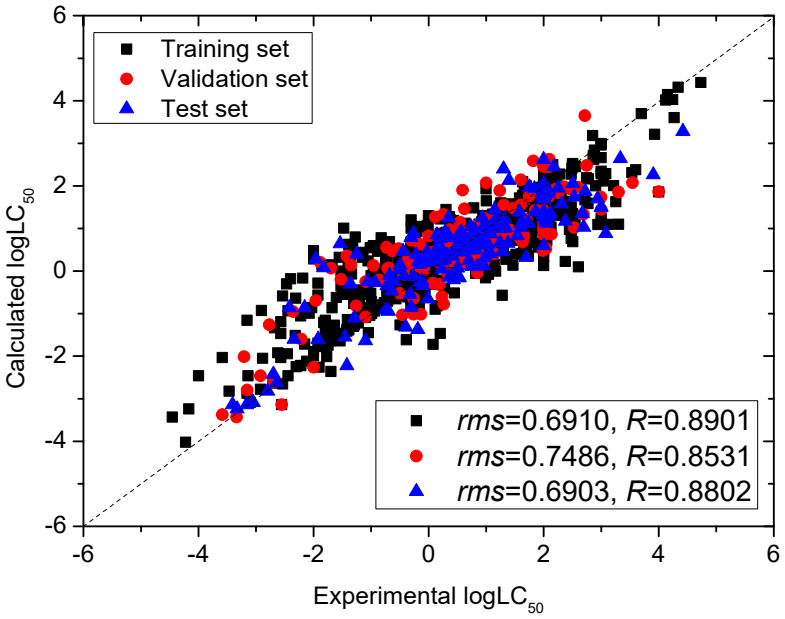

Fig. 3. Relationship between experimental versus calculated $logLC_{50}$ with GRNN model

The optimal GRNN model (σ = 0.11) yielded coefficients of determination R^2 = 0.7922 and *rms* = 0.6910 log units for the training set (774 samples), R^2 = 0.7278 and *rms* = 0.7486 log units for the validation set (166 samples), R^2 = 0.7748 and *rms* = 0.6903 log units for the test set (166 samples). Although the GRNN model dealt with a large dataset of pesticide toxicity $logLC_{50}$ to fishes, it is comparable to the latest similar models from the literature that have the number of samples and R^2 for the training sets being n = 13 and R^2 = 0.839 (Önlü & Saçan 2017), n = 94 and R^2 = 0.79 (Toropov et al. 2017), n = 66 and R^2 = 0.80 (Khan et al. 2019) n = 249 and R^2 = 0.80 (Jia et al. 2020) and n = 233 and R^2 = 0.67 (Toropov et al. 2020).

In addition, the optimal GRNN model produced R^2 = 0.7798 and *rms* = 0.6998 log units for the total set of 1106 pesticides. It is superior to the MLR model based on the same data sets and descriptor set, which has R^2 = 0.5103 and *rms* = 1.0286 log units for the total dataset. Therefore, the 20 molecular descriptors used in the GRNN model are nonlinear with $logLC_{50}$, which indicates that applying the GRNN algorithm to develop QSTRs is reasonable.

Assessing the GRNN model with the test set resulted in an external correlation coefficient q_{ext}^2 = 0.7659 > 0.5; a slope k' (=1.1021) of regression with fix intercept at 0, lying in the range of 0.85-1.15; determination coefficients $R'_0{}^2$ = 0.7740 and R_0^2 = 0.7706, close to the determination coefficient (R^2 = 0.7748) of the test set. Therefore, the development of the GRNN model was successful (Golmohammadi & Safdari 2010).

The optimal GRNN model was further checked with the bias level in prediction errors. There is a systematic error in prediction for a QSAR model if it has any one or more of the following five conditions (Roy, et al. 2017, Yu 2021):

(1) NPE/NNE > 5 or NNE/NPE > 5;
(2) ABS(MPE/MNE)> 2 or ABS(MNE/MPE)> 2;
(3) MAE – ABS(AE) < 0.5 × MAE;
(4) R^2(i^{th} vs (i-1)th residuals) > 0.5 for residuals sorted on Y_{obs};
(5) R^2 (Y vs residuals) > 0.5.

where NPE is the number of positive errors, NNE is the number of negative errors, ABS(x) expresses the absolute value, AE is the average error, MPE is the mean positive error, and MNE is the mean negative error.

Then the following formulas were obtained: NPE/NNE = 94/72 < 5; ABS(MPE/MNE) = 0.5315/0.4918 < 2; MAE – ABS(AE) = 0.5143 – 0.0877 = 0.4266 > 0.5 × 0.5143 = 0.2572; R^2(i^{th} vs (i-1)th residuals) = 0.1691 < 0.5 for residuals sorted on Y_{obs}; R^2(Y vs residuals) = 0.3602 < 0.5. Obviously, the calculation results do not meet any of the above conditions. Thus, the optimal RF model has no systematic error in predictions.

3.3. Applicability domain

Figure 4 shows the Williams plot of the standardized residuals vs. leverages calculated with SPSS 19.0. The prediction points in the domain with absolutes values of standard residual less than 3 and leverages h less than the warning leverage h^* are considered reliable (Yu 2020b). In this paper, the warning leverage is h^* = 0.0814 = 3×(20+1)/774 = 3×(p+1)/n, here p and n are, respectively, the numbers of descriptors and

samples. As shown in Fig. 4, there are 13 samples (Nos. 113, 122, 142, 337, 404, 419, 494, 505, 756, 940, 951, 954 and 1103 in Table S1) with absolute values of standard residuals >3 and leverages *h* less than 0.0814, which suggest that the larger standard residuals may result from the experimental errors for toxicity (pLC50). In addition, there are 32 samples (e.g. Nos. 2 and 4 in Table S1) possessing smaller standard residuals (<3) and higher leverages *h* (> 0.0814), indicating their toxicity (pLC50) can be accurately predicted. However, they have dissimilar structures with other pesticide molecules in the training set.

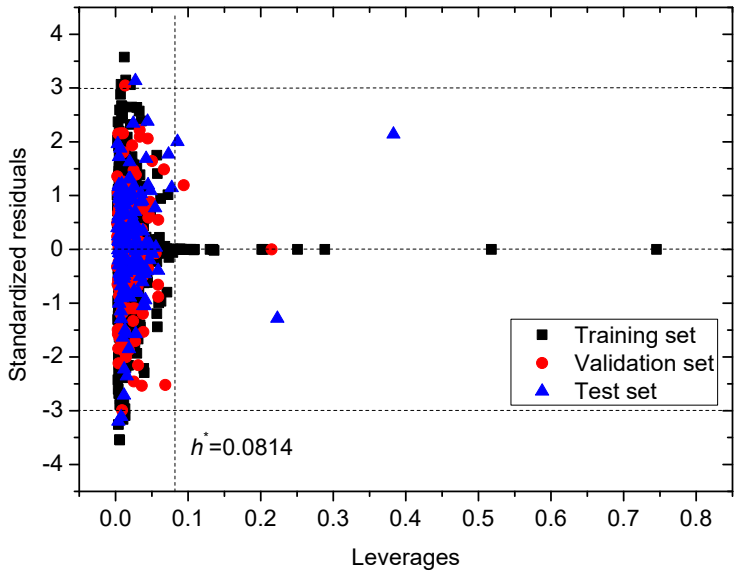

Fig. 4. Williams plot with a warning leverage of 0.0814

4. Conclusions

Although many factors affect pesticide toxicity on fishes, the optimal GRNN model ($\sigma = 0.11$) based on 20 molecular descriptors was successfully developed for toxicity $\log LC_{50}$ of a large data set including 1106 pesticides. The training set (774 pesticides), validation set (166 pesticides) and test set (166 pesticides) yielded correlation coefficients *R* of 0.8901, 0.8531 and 0.8802, respectively. Compared with other QSTR models of toxicity $\log LC_{50}$ on fishes reported in the literature, the optimal GRNN model in this work is accurate. However, a large data set of toxicity $\log LC_{50}$ was used for the model.

Acknowledgements

The authors wish to express their sincere thanks to Professor Xinliang Yu at Hunan Institute of Engineering for his guidance.

Declarations

Conflict of interest: The authors declare no competing interests.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.
Table S1 being molecular descriptors and $\log LC_{50}$ values.

Author contributions

B.W., C.C. and M.L: data curation, software.
L.D.: conceptualization, methodology, writing – original draft, writing – review & editing.
Ethical Approval Bowen Yang declares that he has not received any research grants or honoraria from any commercial companies.
Limin Dang declares that she has not received any research grants or honoraria from any commercial companies.
Cong Chen declares that she has not received any research grants or honoraria from any commercial companies.
Mingwang Li declares that he has not received any research grants or honoraria from any commercial companies.

Human and Animal Rights

This article does not contain any studies with human or animal subjects.

References

Cachot, J. (2014). Assessment of pollution in the Bizerte lagoon (Tunisia) by the combined use of chemical and biochemical markers in mussels, Mytilus gallo-provincialis. *Marine Pollution Bulletin*, *84*, 379-390. https://doi.org/10.1016/j.marpolbul.2014.05.002

DeLorenzo, M.E., Scott, G.I., Ross, P.E. (2001). Toxicity of pesticides to aquatic microorganisms: a review. *Environmental Toxicology and Chemistry*, *20*, 84-98. https://doi.org/10.1002/etc.5620200108

Fang, Z., Yu, X., Zeng, Q. (2022). Random forest algorithm-based accurate prediction of chemical toxicity to *Tetrahymena pyriformis*. *Toxicology*, *480*, 153325. https://doi.org/10.1016/j.tox.2022.153325

Galimberti, F., Moretto, A., Papa, E. (2020). Application of chemometric methods and QSAR models to support pesticide risk assessment starting from ecotoxicological datasets. *Water Research*, *174*, 115583. https://doi.org/10.1016/j.watres.2020.115583

Golmohammadi, H., Safdari, M. (2010). Quantitative structure–property relationship prediction of gas-to-chloroform partition coefficient using artificial neural network. *Microchemical Journal*, *95*(2), 140-151. https://doi.org/10.1016/j.microc.2009.10.019

Jia, Q., Liu, T., Yan, F., Wang, Q. (2020). Norm index–based QSAR model for acute toxicity of pesticides toward *rainbow trout*. *Environmental Toxicology and Chemistry*, *39*(2), 352-358. https://doi.org/10.1002/etc.4621

Isah, H.M., Sawyerr, H.O., Raimi, M.O., Bashir, B.G., Haladu, S., Odipe, O.E. (2020). Assessment of Commonly Used Pesticides and Frequency of Self-Reported Symptoms on Farmers Health in Kura, Kano State, Nigeria. *Journal of Education and Learning Management*, *1*(1), 31-54. http//dx.doi.org/10.46410/jelm.2020.1.1.05

Li, F., Fan, D., Wang, H., Yang, H., Li, W., Tang, Y., Liu, G. (2017). In silico prediction of pesticide aquatic toxicity with chemical category approaches. *Toxicology Research*, *6*, 831. https://doi.org/10.1039/c7tx00144d

Khan, K., Khan, P.M., Lavado, G., Valsecchi, C., Pasqualini, J., Baderna, D., Marzo, M., Lombardo, A., Roy, K., Benfenati, E. (2019). QSAR modeling of Daphnia magna and fish toxicities of biocides using 2D descriptors. *Chemosphere*, *229*, 8-17. https://doi.org/10.1016/j.chemosphere.2019.04.204

Masand, V.H., Zaki, M.E.A., Al-Hussain, S.A., Ghorbal, A.B., Akasapu, S., Lewaa, I., Ghosh, A., Jawarkar, R.D. (2021). Identification of concealed structural alerts using QSTR modeling for *Pseudokirchneriella subcapitata*. *Aquatic Toxicology*, *239*, 105962. https://doi.org/10.1016/j.aquatox.2021.105962

Mit, C., Bado-Nilles, A., Daniele, G., Giroud, B., Vulliet, E., Beaudouin, R. (2022). The toxicokinetics of bisphenol A and its metabolites in fish elucidated by a PBTK model. *Aquatic Toxicology*, *247*, 106174. https://doi.org/10.1016/j.aquatox.2022.106174

Mo, L.-Y., Yuan, B.-K., Zhu, J., Qin, L.-T., Dai, J.-F. (2022). QSAR Models for Predicting Additive and Synergistic Toxicities of Binary Pesticide Mixtures on Scenedesmus Obliquus. *Chinese Journal Structural Chemistry*, *41*(3), 2203166-2203177. https://doi.org/10.14102/j.cnki.0254-5861.2011-3306

Schmidt, S., Schindler, M., Faber, D., Hager, J. (2021). Fish early life stage toxicity prediction from acute daphnid toxicity and quantum chemistry. *SAR and QSAR in Environmental Research*, *32*(2), 151-174. https://doi.org/10.1080/1062936X.2021.1874514

Sullivan, K.M., Manuppello, J.R., Willett, C.E. (2014). Building on a solid foundation: SAR and QSAR as a fundamental strategy to reduce animal testing. *SAR and QSAR in Environmental Research*, *25*, 357-365. https://doi.org/10.1080/1062936X.2014.907203

Önlü, S., Saçan, M.T. (2017). An in silico approach to cytotoxicity of pharmaceuticals and personal care products on the rainbow trout liver cell line RTL-W1. *Environmental Toxicology and Chemistry*, *36*(5), 1162-1169. https://doi.org/10.1002/etc.3663

Pandey, S.K., Ojha, P.K., Roy, K. (2020). Exploring QSAR models for assessment of acute fish toxicity of environmental transformation products of pesticides (ETPPs). *Chemosphere*, *252*, 126508. https://doi.org/10.1016/j.chemosphere.2020.126508

Roy, K., Ambure, P., Aher, R.B. (2017). How important is to detect systematic error in predictions and understand statistical applicability domain of QSAR models?. *Chemometrics and Intelligent Laboratory Systems*, *162*, 44-54. http://dx.doi.org/10.1016/j.chemolab.2017.01.010

Talete srl (2012). DRAGON (software for molecular descriptor calculation) Version 6.0. http://www.talete.mi.it/

Toropov, A.A., Toropova, A.P., Marzo, M., Dorne, J.L., Georgiadis, N., Benfenati, E. (2017). QSAR models for predicting acute toxicity of pesticides in rainbow trout using the CORAL software and EFSA's OpenFoodTox database. *Environmental Toxicology and Pharmacology*, *53*, 158-163. https://doi.org/10.1016/j.etap.2017.05.011

Toropov, A.A., Toropova, A.P., Benfenati, E. (2020). QSAR model for pesticides toxicity to Rainbow Trout based on "ideal correlations". *Aquatic Toxicology*, *227*, 105589. https://doi.org/10.1016/j.aquatox.2020.105589

Yu, X. (2020a). Prediction of chemical toxicity to *Tetrahymena pyriformis* with four descriptor models. *Ecotoxicology and Environmental Safety*, *190*, 110146. https://doi.org/10.1016/j.ecoenv.2019.110146

Yu, X. (2020b). Quantitative structure-toxicity relationships of organic chemicals against *Pseudokirchneriella subcapitata*. *Aquatic Toxicology*, *224*, 105496. https://doi.org/10.1016/j.aquatox.2020.105496

Yu, X. (2021). Support vector machine-based model for toxicity of organic compounds against fish. *Regulatory Toxicology and Pharmacology*, *123*, 104942. https://doi.org/10.1016/j.yrtph.2021.104942

Yu, X., Zeng, Q. (2022). Random forest algorithm-based classification model of pesticide aquatic toxicity to fishes. *Aquatic Toxicology*, *251*, 106265. https://doi.org/10.1016/j.aquatox.2022.106265

Yu, X. (2023). Global classification models for predicting acute toxicity of chemicals towards Daphnia magna. *Environmental Research*, *238*, 117239. https://doi.org/10.1016/j.envres.2023.117239