کاربرد مدل کمی ساختار–سمیت (QSTR) برای پیش بینی سمیت آفت کش های کاربامات با استفاده از روش های محاسباتی و توصیفگرهای مولکولی

سیده آزاده موسوی[۱]، عصمت محمدی نسب[۲و*]، طاهره مومنی اصفهانی[۳]

۱– دانشجوی دکتری شیمی فیزیک، گروه شیمی، واحد اراک، دانشگاه آزاداسلامی، اراک، ایران

۲– استادیار شیمی فیزیک، گروه شیمی،واحد اراک، دانشگاه آزاداسلامی، اراک، ایران

۳– استادیار شیمی تجزیه، گروه شیمی،واحد اراک، دانشگاه آزاداسلامی، اراک، ایران

چکیده

ما در این مطالعه، محاسبات مکانیک کوانتومی را در سطح تئوری تابع چگالی با مجموعه پایه ۶–۳۱G* انجام دادیم تا یک مدل رابطه کمی ساختار–سمیت (QSTR) برای پیش‌بینی دوز کشنده (LD50) مشتقات کاربامات‌ها بسازیم. بهترین توصیفگرهای مولکولی با استفاده از الگوریتم ژنتیک (GA) توسط نرم افزار MATLAB انتخاب شدند. سپس، رابطه بین توصیفگرهای انتخاب شده و $logLD_{50}$ مشتقات کاربامات را با استفاده از مدل های رگرسیون خطی چندگانه گام به گام (BW-MLR) و شبکه عصبی مصنوعی (BP-ANN) مورد مطالعه قرار دادیم. توصیفگرهای RDF010e، WW و R3e برای مدلسازی روش‌های GA-BWMLR و GA-BPANN استفاده شدند. مقایسه نتایج نشان داد که R^2 و Q^2 مدل GA-BPANN برای همه مجموعه ها به طور قابل توجهی بالاتر از مدل GA-BWMLR می باشند. با توجه به مقادیر میانگین مربعات خطای کمتر (MSE)، ریشه میانگین مربع خطا (RMSE)، خطای استاندارد پیش‌بینی (SEP)، و میانگین مطلق انحراف (ADD) مدل GA-BPANN برای مجموعه دادها از دقت بالاتری برای پیش بینی سمیت کاراماتهای مورد مطالعه برخوردار می باشد.

واژه‌های کلیدی: "آفتکش"؛"QSTR"؛"سمیت"؛"کاراماتها"؛" GA-BWMLR"؛"GA-BPANN"

Zapadka M, Kaczmarek M, Kupcewicz B, Dekowski P, Walkowiak A, Kokotkiewicz A, Luczkiewicz M, & Bucinski A, 2019. An application of QSRR approach and multiple linear regression method for lipophilicity assessment of flavonoids, Journal of Pharmaceutical and Biomedical Analysis, 5(164): 681-689.

Zvinavashe E, Du T, Griff T, van den Berg HH, Soffers AE, Vervoort J, Rietjens IM, 2009. Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to Daphnia magna and Cyprinus carpio. *Chemosphere*, *75*(11): 1531-1538.

Pope C, Karanth S, Liu J, 2005. Pharmacology and toxicology of cholinesterase inhibitors: uses and misuses of a common mechanism of action, Environmental toxicology and pharmacology, 19(3): 433-446.

Popoola SI, Adetiba E, Atayero AA, Faruk N, Calafate CT, **2018.** *Cogent Engineering,* 5(1): 1-19.

Pourbasheer E, Ahmadpour S, Zare-Dorabei R, Nekoei MM, 2017. The *Arabian Journal of Chemistry*, 10(1): 33-34.

Randic M, & Basak SC, 2000. *SAR* and *QSAR* in *Environmental*, Res. 11(1): 1-23.

Roberts TR, Hutson DH, Jewess PJ, Lee P, Nicholls PH, 1999. Metabolic pathways of agrochemicals: part 2: insecticides and fungicides. Royal Society of Chemistry, Cambridge, 2.

Roberts TR, Hutson DH, Lee PW, Nicholls PH, 1998. Plimmer, JR. Metabolic pathways of agrochemicals: part 1: herbicides and plant growth regulators. Royal Society of Chemistry, Cambridge, 1.

Roy K, Kar S, Das RN, 2015. Statistical methods in QSAR/QSPR. In A primer on QSAR/QSPR modeling, Springer, Cham, 37.

Saghaie L, Sakhi H, Sabzyan H, Shahlaei M, Shamshirian D, 2013. *Medicinal Chemistry Research*, 22(4): 1679-1688.

Saiz-Urra L, Gonzalez MP, Fall Y, Gomez G, 2007. The *European Journal of Medicinal Chemistry*, 42(1): 64-70.

Sarkhosh M, Khorshidi N, Niazi A, Leardi R, 2014. *Chemometrics* and *Intelligent Laboratory Systems*, 139: 168-174.

Shanmuganathan A, 2016. Artificial neural network modelling: In An introduction. Springer, Cham, 1.

Sigmon K, & Davis TA, 2004. MATLAB primer. Chapman and Hall/CRC.

Souyei B, Seyd AH, Zaiz F, Rebiai A, 2019. Application of inverse QSAR/QSPR analysis for pesticides structures generation, *Acta Chimica* Slovenica, 66 (2): 315-325.

Sun G, Zhang Y, Pei L, Lou Y, Mu Y, Yun J, Li F, Wang Y, Hao Z, Xi S, Li C, 2021. *Ecotoxicology* and *Environmental* Safety, 222: 112525-112538.

Testa B, Mayer JM, Mayer J, 2003. Hydrolysis in drug and prodrug metabolism. John Wiley & Sons.

Thapliyal A, Krishen Khar R, Chandra A, 2018. Artificial neural network modelling of green synthesised silver nanoparticles in bentonite/starch bio-nanocomposite, *Current Nanoscience*, 14(3): 239-251.

Thapliyal A, Krishen Khar R, Chandra A, 2018. *Current Nanoscience*, 14: 239-251.

Todeschini R, & Consonni V, 2008. Handbook of molecular descriptors. John Wiley & Sons.

Toropov AA, Toropova AP, Cappelli CI, Benfenati E, 2015. Model for octanol/water partition coefficient, Fluid Phase Equilibria, 397: 44-49.

Toth G, Bodai Z, Heberger K, 2013. The Journal of Computer-*Aided* Molecular Design, 10: 837-844.

Tripathi M & Singal SK, 2019. Use of principal component analysis for parameter selection for development of a novel water quality index: a case study of river Ganga India, *Ecological Indicators*, 96: 430-436.

Tsuneda T, 2014. Density functional theory in quantum chemistry. Springer, 978-1004.

Tsuneda T, & Hirao K, 2014. Self-interaction corrections in density functional theory, The *Journal of Chemical Physics*, 140(18): 18A5131-13.

Villarrubia G, De Paz JF, Chamoso P, De la Prieta F, 2018. Neurocomputing. 272: 10-16.

Yee LC & Wei YC,2012. Current modeling methods used in QSAR/QSPR. Wiley-VCH Verlag GmbH & Co. KGaA.

Yang L, Wang Y, Chang J, Pan Y, Wei R, Li J, Wang H, 2020. Chemosphere, 258: 127217-127255.

Gaullier C, Baran N, Dousset S, Devau N, Billet D, Kitzinger G, Coisy E, 2019. Wetland hydrodynamics and mitigation of pesticides and their metabolites at pilot-scale, *Journal of Ecological Engineering,* 136: 185-192.

Ghosh AK, & Brindisi M, 2015. Organic carbamates in drug design and medicinal chemistry, *Journal of Medicinal Chemistry,* 58(7): 2895-2940.

Glavanovic S, Glavanovic M, Tomisic V, 2016. *Spectrochimica Acta*, Part A: *Molecular* and *Biomolecular Spectroscopy*, 157: 258-264.

Gupta RC, 2011. Toxicology of organophosphate and carbamate compounds. Academic Press.

Helguera AM, Cabrera Perez MA, Gonzalez MP, 2006. Journal of *molecular modeling*, 12(6): 769-780.

Hocking R, 2013. Methods and Applications of Linear Models: Regression and the Analysis of Variance. New York: John Wiley & Sons.

Http://chem.nlm.nih.gov (Chem ID plus).

Hyndman RJ, & Koehler AB, 2006. The *International Journal* of *Forecasting*, 22: 679-688.

Kawczak P, Belka M, Slawinski J, Baczek T, 2018. Current Pharmaceutical Analysis, 14(1): 35-40.

Kumar V, Chadha N, Tiwari AK, Sehgal N, Mishra AK, 2014. Prospective atom-based 3D-QSAR model prediction, pharmacophore generation, and molecular docking study of carbamate derivatives as dual inhibitors of AChE and MAO-B for Alzheimer's disease, *Medicinal Chemistry Research,* 23(3): 1114-1122.

Kutner MH, Nachtsheim CJ, Neter J, Li W, 2005. Applied Linear Statistical Models. McGraw-Hill.

Leardi R, 2003. Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks, Amsterdam.

Lee S, & Barron MG, 2016. A mechanism-based 3D-QSAR approach for classification and prediction of acetylcholinesterase inhibitory potency of organophosphate and carbamate analogs, The *Journal* of *Computer-Aided Molecular* Design, 30(4): 347-363.

Liu S, Jin L, Yu H, Lv L, Chen CE, Ying GG, 2020. *Science* of the *Total Environment*, 706: 135691-135699.

Lin W, Jiang R, Shen Y, Xiong Y, Hu S, Xu J, Ouyang G, 2018. Sci. Total Environ. 635: 53-59.

Mansour T, & Schork M, 2010. he *Journal* of *Mathematical Chemistry*, 47(1): 72-98.

Mirjalili S, 2019. Evolutionary algorithms and neural networks. Springer, Cham, 780.

Montgomery DC, Peck EA, 2015. Vining, G. G. Introduction to linear regression analysis**.** John Wiley & Sons.

Naik PK, Sindhura Singh T, Singh H, 2009. Quantitative structure–activity relationship (QSAR) for insecticides: development of predictive in vivo insecticide activity models, *SAR* and *QSAR* in *Environmental Research,* 20(5-6): 551-566.

Navabi A, & Momeni Isfahani T, 2021. *Journal* of *Theoretical* and *Computational chemistry*, 19(1): 51-64.

Niazi A, & Leardi R, 2012. Genetic algorithms in chemometrics, Wiley Online Library,

Niazi A, & Leardi R, 2012. Journal of *Chemometrics*, 26(6): 345-351.

Nour H, Abchir O, Belaidi S, Qais FA, Chtita S, Belaaouad S, 2022. 2D-QSAR and molecular docking studies of carbamate derivatives to discover novel potent anti-butyrylcholinesterase agents for Alzheimer's disease treatment, The Bulletin of the Korean Chemical Society, 43(2): 277-292.

Piel C, Pouchieu C, Carles C, Beziat B, Boulanger M, Bureau M, Busson A, Gruber A, Lecluse Y, Migault L, Renier M, 2019. *Environment International*, 130 (104876), 1-12.

Pohanka M, 2012. Acetylcholinesterase based dipsticks with indoxylacetate as a substrate for assay of organophosphates and carbamates, *Analytical Letters,* 45(4): 367-374.

REFERENCES

Adad A, Larif M, Hmammouchi R, Taghki AI, Bouachrine M, Lakhlifi T, 2013. DFT-QSAR models to predict biological activities of carbamates, s-alkylcarbamathioates, ureas and chloroacetamides derivatives, *Journal of* Analytica Chimica *Acta,* 2(2): 105-118.

Ahmadi S, & Ganji S, 2016. Genetic algorithm and self-organizing maps for QSPR study of some N-aryl derivatives as butyrylcholinesterase inhibitors, *Current Drug Discovery Technologies*, 13(4): 232-253.

Ahmadi S, & Habibpour E, 2017. Application of GA-MLR for QSAR modeling of the arylthioindole class of tubulin polymerization inhibitors as anticancer agents**,** *Anti-Cancer Agents* in *Medicinal Chemistry*, 17(4): 552-565.

Amari S, 1990. I. Lecture notes in statistics: In Differential-Geometrical Methods in Statistics, Springer-Verlag, 28.

Amiri R, Djelloul M, Amel B, 2020. QSAR Study of the octanol/water partition coefficient of organophosphorus compounds: The hybrid GA/MLR and GA/ANN approaches, *Journal of* the *Serbian Chemical* Society, 85 (4): 467-480.

Amiri R, Messadi D, Bouakkadia A, 2020. QSAR Study of the octanol/water partition coefficient of organophosphorus compounds: The hybrid GA/MLR and GA/ANN approaches, *Journal of* the *Serbian Chemical Society*, 85(4): 467-480.

Ballantyne B, & Marrs TC, 2017. Clinical and experimental toxicology of organophosphates and carbamates. Elsevier.

Bhhatarai B, Teetz W, Liu T, Oberg T, Jeliazkova N, Kochev N, Pukalov O, Tetko I, Kovarich S, Papa E, Gramatica P, 2011. *Molecular Informatic*, (30): 189–204.

Bora A, Suzuki T, Funar-Timofei S, 2019. Neonicotinoid insecticide design: molecular docking, multiple chemometric approaches, and toxicity relationship with Cowpea aphids, *Environmental Science* and *Pollution Research*, 26(14):14547-14561.

Cappelli CI, Benfenati E, Cester J, 2015. Evaluation of QSAR models for predicting the partition coefficient (log P) of chemicals under the REACH regulation, *Environmental Research,* 143: 26-32.

Chen X, Lin M, Sun L, Xu T, Lai K, Huang M, Lin H, 2019. Detection and quantification of carbendazim in Oolong tea by surface-enhanced Raman spectroscopy and gold nanoparticle substrates, *Food Chemistry,* 293: 271-277.

Chirico N, & Gramatica P, 2011. *Journal* of *Chemical Information* and *Modeling*, 51(9): 2320-2335.

Chirico N, & Gramatica P, 2012. *Journal* of *Chemical Information* and *Modeling,* 52(8): 2044-2058.

Cilimkovic M, 2015. Neural networks and back propagation algorithm. Institute of Technology Blanchardstown, Blanchardstown Road North Dublin. 15(1).

Da Silva IN, Spatti DH, Flauzino RA, Liboni LHB, dos Reis Alves SF, 2017. Artificial neural networks. Cham: Springer International Publishing.

De Simone A, Russo D, Ruda GF, Micoli A, Ferraro M, Di Martino RMC, Ottonello G, Summa M, Armirotti A, Bandiera T, Cavalli A, 2017. Design, synthesis, structure–activity relationship studies, and three-dimensional quantitative structure–activity relationship (3D-QSAR) modeling of a series of o-biphenyl carbamates as dual modulators of dopamine d3 receptor and fatty acid amide hydrolase, *Journal of Medicinal Chemistry,* 60(6): 2287-2304.

Dreyfus G, 2005. Neural networks: methodology and applications. Springer Science & Business Media.

Fourches D, & Ash JR, 2019. 4D-quantitative structure–activity relationship modeling: making a comeback, Expert Opinion on Drug Discovery. 14(12): 1227-1235.

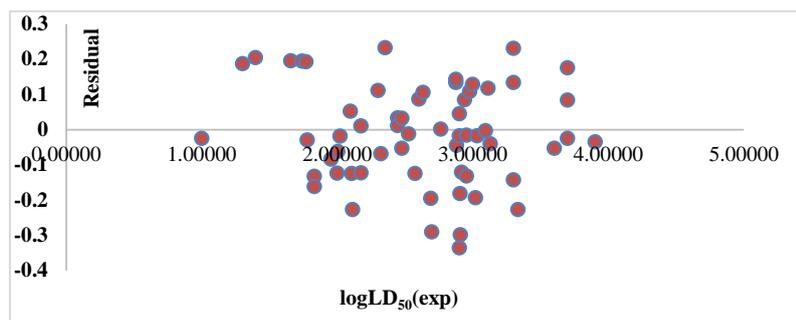**Frisch ML, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Pople JA, 2009.** Gaussian, Inc., Wallingford CT.

31

Fig. (5). The plot of residuals versus experimentals $logLD_{50}$ in GA-BPANN methods.

CONCLUSION

In this study, we performed QSAR using linear and nonlinear analyses for a series of carbamates with a focus on their toxicity. The GA-BPANN with a 3–4–1 (number of input, hidden, and output layer nodes) configuration was developed to predict the $logLD_{50}$ of carbamate derivatives. The QSTR models showed that GA-BPANN model with satisfactory accuracy was appropriate to predict the $logLD_{50}$, as this model had the highest R^2, Q^2_{F1}, Q^2_{F2}, and Q^2_{F3} and the lowest MSE, RMSE, SEP, and ADD. This new approach could be considered as an alternative and practical technique to evaluate the biological activity of pesticides and insecticides and further design of novel carbamates. Regression analysis instates a relation between a dependent variable representing the biologic activity and multiple independent variables, namely the molecular descriptors. According to the modeling descriptors, while the RDF and GETAWAY descriptors contributed positively, the WW descriptor belonging to 2D matrix-based descriptors showed a negative contribution. However, the influence of molecular descriptors is different and varies with the physiochemical information they encode.

Estimating the toxicity of pesticides and insecticides is necessary to identify their harmful effects on humans, animals, plants, and environment. However, the toxicity testing of chemicals is bounded by time, ethical considerations, and financial charges. Therefore, computational methods can be useful to estimate the toxicity of chemicals.
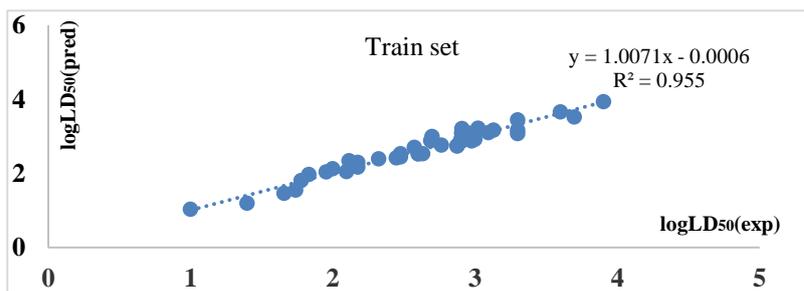
Fig. (2). The plot of predicted versus experimental $logLD_{50}$ of train set in GA-BPANN methods.
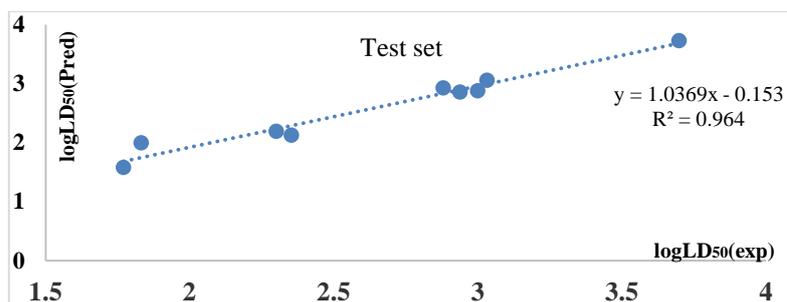
Fig. (3). The plot of predicted versus experimental $logLD_{50}$ of test set in GA-BPANN methods.
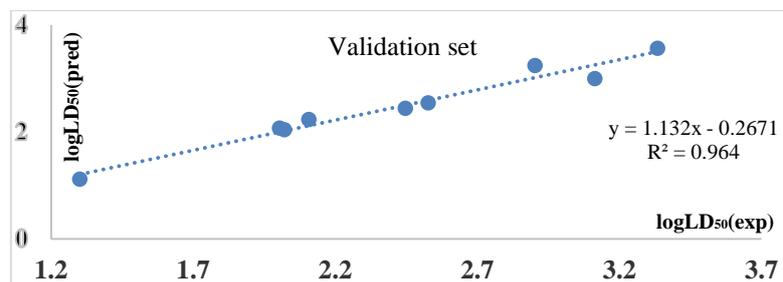

Fig. (4). The plot of predicted versus experimental $logLD_{50}$ of validation set in GA-BPANN methods.

Regular Residuals

The residuals were used to approximate the regularity of supposition. Figure 5 compares the residual values, which is the difference between the observed and predicted values in contrast to the $logLD_{50}$ experimental values of carbamates. The spread of errors in both sides of zero are random. So, the deviations around x-axis had an almost indistinguishable spread. This affirms the adaptability of the GA-BPANN model for the suggested $logLD_{50}$ of studied carbamates.

29

$$Q^2{}_{LOO} = 1 - \frac{PRESS}{TSS} \qquad (equation:4)$$

Where, $PRESS = \sum_{k=1}^{n}(y_k - \hat{y}_k)^2$, $TSS = \sum_{k=1}^{n}(y_k - \bar{y}_k)^2$

$$SEP = \frac{RMSE}{\bar{y}_k} \times 100 \qquad (equation:5)$$

$$AAD = \left(\frac{1}{n} \sum_{k=1}^{n} \left[\frac{|y_k - \hat{y}_k|}{y_k} \right] \right) \qquad (equation:6)$$

Where, y_k, $\hat{y}_k$ and $\bar{y}_k$ show the measured, predicted, and mean values of the dependent variable, respectively.

The comparison between the GA-BWMLR and GA-BPANN

We calculated the values of R^2, MSE, RMSE, $Q^2{}_{LOO}$, SEP, and ADD of the GA-BPANN and GA-MLR models (**Table 6**).

Table 6. The regression parameters and quality of correlation of proposed GA-BWMLR and GA-BPANN methods.

Parameters	GA-BWMLR		GA-BPANN		
	Training set	Test set	Training set	Test set	Validation set
N	45	15	42	9	9
R^2	0.447	0.705	0.9550	0.9646	0.9645
MSE	0.2221	0.1129	0.0185	0.0552	0.0260
RMSE	0.4713	0.3361	0.1359	0.2350	0.1612
$Q^2{}_{LOO}$(data set)	0.458564			0.9486	
SEP (data set)	17.75204			13.9434	
ADD (data set)	15.83351			4.7184	

According to the obtained values for R^2 in both models, 70.5% and 96.46% of the $logLD_{50}$ of test data variations could be determined in terms of one unit difference in $logLD_{50}$ value, respectively. Also, the RMSE=0.3361 for the test set in the GA-BWMLR model was compared with the value of RMSE=0.2350 in the GA-BPANN model.

The GA-BPANN model with higher values of $Q^2{}_{LOO}$, R^2, and lower MSE, RMSE, SEP, and ADD values proved to be more precise. The comparison between the MSE, RMSE, SEP, and ADD values of the GA-BPANN and GA-BWMLR models for training and test sets indicated superiority of the GA-BPANN model over the GA-BWMLR model. The main advantage of the GA-BPANN model was its ability to predict the $logLD_{50}$ for new carbamates; this external prediction ability was evaluated by using the $Q^2{}_{F1}$, $Q^2{}_{F2}$, $Q^2{}_{F3}$, and CCC values. These parameters were calculated as follows (Bhhatarai et al, 2011; Chirico, 2011; Chirico, 2012):

$Q^2{}_{F1}$= 0.9505 , $Q^2{}_{F2}$= 0.9504 , $Q^2{}_{F3}$=0.9542, CCC= 0.9767

The values of $Q^2{}_{F1}, Q^2{}_{F2}, Q^2{}_{F3} \rangle 0.9$ and $CCC > 0.97$ indicated a fitting non-linear relationship between $logLD_{50}$ and selected descriptors. Figures 2, 3, and 4 show the performance of the GA-BPANN model of training, test, and validation sets, respectively. The values of R^2 in these figures indicated a strong correlation between $logLD_{50}$ in contrast to independent variables.

Table 5. Pearson Correlation coefficient between the final descriptors

	WW	RDF010e	R3e
Pearson correlation coefficient			
WW	1.000	-0.499	-0.368
RDF010e	-0.499	1.000	0.097
R3e	-0.368	0.097	1.000

The best linear model contains three molecular descriptors, including WW, RDF010e, and R3e as follows:

$\log LD_{50}$ = 0.210 RDF010e – 5.449E-5 WW + 0.960 R3e + 0.364   (*equation:* 1)

The results indicated that the potential of the WW, RDF010e, and R3e descriptors in modeling $\log LD_{50}$ was better than the other descriptors. These descriptors were classified into RDF, GETAWAY, and 2D matrix-based descriptors, respectively. The RDF010e descriptor is a Radial Distribution Function -010 / weighted by Sanderson electronegativity, and belongs to RDF descriptors (Helguera et al, 2006). The R3e correlates to GETAWAY descriptors, and is R autocorrelation of lag 3 / weighted by Sanderson electronegativity (Saiz-Urra et al, 2007). The WW is hyper-wiener-like index (log function) from topological distance matrix, and belongs to 2D matrix-based descriptors category (Mansour and Schork, 2010). The results showed that RDF010e and R3e descriptors had positive effects, but WW descriptor had negative effects on the $\log LD_{50}$ of studied carbamates.

Validation

The values of predicted and experimental $\log LD_{50}$ of the test set were compared by calculating the MSE and R^2 values, and the models were described by the sum of these two values. Computation of R^2 and MSE of the GA-BPANN models was performed and recorded after every 10 cycles. For each neuron number, the average of the summed MSE was recorded, and the model with the lowest MSE and the highest R^2 values was chosen as the superior model. It was indicated that the hidden layers with four neurons had the lowest MSE and the highest R^2 in the best model, and the best signals were transmitted onto an output layer. The RMSE specifies that the descriptors with less estimated errors have more importance (Hyndman and Koehler, 2006). Other evaluation parameters such as coefficients of determination for cross-validation (leave-one-out-Q^2_{LOO}), the standard error of prediction (SEP), and absolute average deviation (AAD) were used to evaluate the GA-BWMLR and GA-BPANN performances (Roy et al, 2015; Saghaie et al, 2013). The cross-validated correlation coefficient is a scale of the goodness of internal predicting power. The AAD of a data set shows the comparative absolute abnormality from the predicted values (Toth et al, 2013; Popoola et al, 2018). The R^2, RMSE, Q^2_{LOO}, SEP, and ADD parameters were calculated using the following formulas:

$$R^2 = \frac{[\sum_{k=1}^{n}(y_k - \bar{y}_k)(\hat{y}_k - \bar{\hat{y}}_k)]^2}{\sum_{k=1}^{n}(y_k - \bar{y}_k)^2 \sum_{k=1}^{n}(\hat{y}_k - \bar{\hat{y}}_k)^2}$$ (*equation* : 2)

$$RMSE = \sqrt{\frac{1}{n}\sum_{K=1}^{n}(y_K - \hat{y}_K)^2}$$ (*equation* : 3)

Multicollinearity

The linear regression analysis showed that the values of R^2 and correlation coefficient (R) were greater than 0.8, but collinearity needed to be checked. Therefore, the variance inflation factor (VIF) and Pearson's correlation coefficient (PCC) specified for the selected descriptors of the last model. The VIF is the reverse value of $(1-R^2)$, where the R^2 is squared correlation coefficient of multiple regressions between variables in the raised model.

Collinearity occurs when two or more predictor descriptors are inter-correlated. The collinearity is the basis of the VIF value. If the VIF value is between 1 and 10, the related model is acceptable, but if VIF<1 or >10, there is collinearity and a recheck is necessary . (Pourbasheer et al, 2017).

At each step, the descriptor with the highest VIF and its PCC with another descriptor greater than 0.5 is removed. Among 12 extracted descriptors, S0K, Ss, Gnar, X0v, Eeig04r, Espm01r, and RDF010P, which had high VIF and PCC values were removed and modeling was done again. These steps were repeated until 1<VIF<10 and PCC<0.5. The results of PCC, VIF, and tolerance for each descriptor were given in Tables 3 and 4, respectively. The results of PCC between the final three descriptors were given in Table 5.

Table 3. Pearson Correlation coefficient between the molecular descriptors

Pearson correlation coefficient

	R3e	RDF010e	GNar	WW	RDF030p	EEig04r	ClC0	nc	S0k	X0v	SS	ESpm01r
R3e	1.000	-0.125	-0.611	-0.213	-0.160	0.036	-0.525	0.152	0.080	-0.138	-0.395	0.481
RDF010e	-0.125	1.000	-0.339	0.235	0.324	0.200	-0.518	-0.385	0.447	-0.114	-0.622	0.381
GNar	-0.611	-0.339	1.000	-0.147	0.037	-0.078	0.808	0.019	-0.258	0.483	0.795	-0.850
WW	-0.213	0.235	-0.147	1.000	0.069	0.010	0.172	-0.547	-0.163	-0.334	-0.171	0.225
RDF030p	-0.160	0.324	0.037	0.069	1.000	0.229	-0.186	-0.248	0.386	-0.342	-0.148	0.005
EEig04r	0.036	0.200	-0.078	0.010	0.229	1.000	-0.151	0.088	0.471	-0.326	-0.126	-0.315
ClC0	-0.525	-0.518	0.808	0.172	-0.186	-0.151	1.000	-0.095	-0.396	0.263	0.799	-0.735
nc	0.152	-0.385	0.019	-0.547	-0.248	0.088	-0.095	1.000	-0.335	0.125	0.356	-0.283
S0k	0.080	0.447	-0.258	-0.163	0.386	0.471	-0.396	-0.335	1.000	-0.447	-0.566	0.117
X0v	-0.138	-0.114	0.483	-0.334	-0.342	-0.326	0.263	0.125	-0.447	1.000	0.331	-0.302
SS	-0.395	-0.622	0.796	-0.171	-0.148	-0.126	0.799	0.356	-0.566	0.331	1.000	-0.804
ESpm01r	0.481	0.381	-0.850	0.225	0.005	-0.315	-0.735	-0.283	0.117	-0.302	-0.804	1.000

Table 4. collinearity statistical parameters between the molecular descriptors

Name	Collinearity Statistical		Mode1 1	Model 2	Model 3	Model 4
	Tolerance	VIF	VIF	VIF	VIF	VIF
nc	0.021	47.224	22.648	-	-	-
S0K	0.012	85.096	-	-	-	-
Ss	0.011	93.012	-	-	-	-
WW	0.050	19.987	9.064	2.004	1.369	1.170
GNar	0.036	27.922	-	-	-	-
X0v	0.033	29.998	-	-	-	-
ClC0	0.021	47.530	8.506	3.217	-	-
EEig04r	0.045	22.006	-	-	-	-
ESpm01r	0.003	307.258	-	-	-	-
RDF010e	0.270	3.698	1.834	1.930	1.351	1.347
RDF030p	0.384	2.601	-	1.722	1.213	-
R3e	0.336	2.980	1.577	1.566	1.544	1.544

Back propagation artificial neural network method

The back propagation artificial Neural Network (BPANN), as an intelligence model, was built by the Neural Network Toolbox included in MATLAB R2010b (Thapliyal et al, 2018). In this method, MSE was used as the performance function, which consisted of three interrelated layers: an input, one or more hidden layer (s), and an output layer (Dreyfus, 2005; Villarrubia et al, 2018; Cilimkovic, 2015; Da Silva et al, 2017; Shanmuganathan, 2016). To build models, the data subjected to GA-BPANN analysis was randomly divided into a training set of 42 (70%) compounds to make the model, a test set of nine (15%) compounds to estimate an independent variable, and a validation set of nine (15%) compounds. The WW, RDF010e, and R3e descriptors were sent to the input layer where they were consequently passed on to the nodes of the hidden layer for later processing. The neurons number in the hidden layer was optimized by testing the network implementation with different neuron numbers. The neuron number was gently increased from 1 to 10 for each case, and the training step was repeated more than 100 times. Finally, it was signified that the best nonlinear model contained four nodes in the hidden layers. In each training run, 15%−15% of the data were randomly chosen for testing and validation, respectively. The resulting networks were used to predict the $logLD_{50}$ of the test set. Figure 1 shows the structure of the extracted GA-BPANN model.
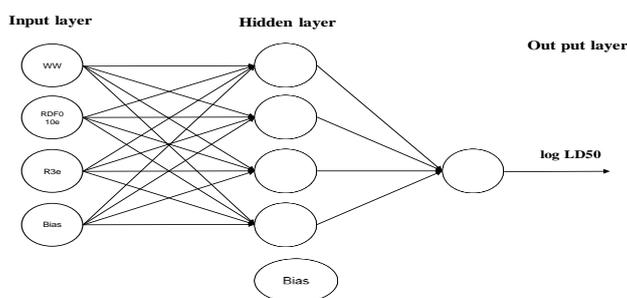
Fig. (1). The structure of GA-BPANN model

25

14	R3e, RDF010e, GNar, R2e, WW, ESpm04d, RDF030p, EEig04r, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, ESpm01r, RDF010u, QXXp	0.921	0.849	0.765	0.094	10.169	0.000
15	R3e, RDF010e, GNar, R2e, WW, ESpm04d, RDF030p, EEig04r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, ESpm01r, RDF010u, QXXp	0.920	0.847	0.769	0.093	10.821	0.000
16	R3e, RDF010e, GNar, WW, ESpm04d, RDF030p, EEig04r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, ESpm01r, RDF010u, QXXp	0.919	0.845	0.771	0.092	11.458	0.000
17	R3e, RDF010e, GNar, WW, ESpm04d, RDF030p, EEig04r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, ESpm01r, RDF010u, QXXp	0.916	0.840	0.770	0.093	11.953	0.000
18	R3e, RDF010e, GNar, WW, ESpm04d, RDF030p, EEig04r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm01r, RDF010u, QXXp	0.914	0.836	0.769	0.093	12.554	0.000
19	R3e, RDF010e, GNar, WW, RDF030p, EEig04r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm01r, RDF010u, QXXp	0.913	0.834	0.773	0.092	13.537	0.000
20	R3e, RDF010e, GNar, WW, RDF030p, EEig04r, CIC0, QXXv, nC, S0K, X0v, Ss, ESpm01r, RDF010u, QXXp	0.911	0.829	0.771	0.092	14.232	0.000
21	R3e, RDF010e, GNar, WW, RDF030p, EEig04r, CIC0, QXXv, nC, S0K, X0v, Ss, ESpm01r, QXXp	0.909	0.826	0.772	0.092	15.246	0.000
22	R3e, RDF010e, GNar, WW, RDF030p, EEig04r, CIC0, QXXv, nC, S0K, X0v, Ss, ESpm01r	0.906	0.820	0.769	0.093	16.147	0.000
23	R3e, RDF010e, GNar, WW, RDF030p, EEig04r, CIC0, nC, S0K, X0v, Ss, ESpm01r	0.903	0.815	0.768	0.093	17.311	0.000

The $logLD_{50}$ values of all the studied compounds (as the dependent variable) and selected descriptors (as the independent variable) were applied.

Then, the data set of 60 compounds was randomly divided into two groups: a training set of 45 (75%) compounds to build the model and a test set of 15 (25%) compounds to evaluate the built model. The SAR model was generated using the GA-BWMLR in SPSS software and considering the error of 0.05. A significance level <0.05 indicated that the association between the $logLD_{50}$ and its predictor variables was statistically meaningful.

All models were evaluated using important statistical parameters such as correlation coefficient (R), squared correlation coefficient (R^2), adjusted squared correlation coefficient (R^2_{Adjust}), Fisher's F-ratio, Durbin–Watson (DW) test, mean square error (MSE), and significant (sig) (Montgomery et al, 2015; Navabi and Momeni Isfahani, 2021; Kawczak et al, 2018). Model 23 showed a relation between the $logLD_{50}$ with R3e, RDF010e, GNar, WW, RDF030p, EEig04r, CIC0, nC, S0K, X0v, Ss, and ESpm01r descriptors. As Table 2 shows, DW=1.067 and F=17.311 values in model 23 indicated no autocorrelation. According to the value of R^2=0.815 in 23st model, 81.5% of $logLD_{50}$ can be described with the best model. Also, the R^2_{Adjust} value illustrated the real effect of applied independent variables on the $logLD_{50}$. Thus, the value of R^2_{Adjust}=0.768 can be used to explain the $logLD_{50}$ variations in terms of the values of 12 selected descriptors.

Table 2. The regression parameters of proposed models in GA-BWMLR method.

Model	Independent variables	R	R^2	R^2_{adj}	MSE	F	Sig
1	R3e, EEig04d, RDF030u, RDF010e, QXXm, GNar, DELS, R2e, RDF045u, BEHe5, WW, ESpm04d, RDF030p, piPC06, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, RDF030e, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm02d, ESpm01r, RDF010u, QXXp	0.927	0.859	0.667	0.134	4.478	0.000
2	R3e, EEig04d, RDF030u, RDF010e, QXXm, GNar, DELS, R2e, RDF045u, BEHe5, WW, ESpm04d, RDF030p, piPC06, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, RDF030e, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.927	0.859	0.680	0.129	4.798	0.000
3	R3e, EEig04d, RDF030u, RDF010e, QXXm, GNar, DELS, R2e, RDF045u, BEHe5, WW, ESpm04d, RDF030p, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, RDF030e, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.927	0.859	0.692	0.124	5.134	0.000
4	R3e, EEig04d, RDF030u, RDF010e, QXXm, GNar, DELS, R2e, RDF045u, BEHe5, WW, ESpm04d, RDF030p, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.927	0.859	0.702	0.120	5.484	0.000
5	R3e, EEig04d, RDF030u, RDF010e, QXXm, GNar, DELS, R2 e, RDF045u, WW, ESpm04d, RDF030p, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.926	0.858	0.712	0.116	5.852	0.000
6	R3e, RDF030u, RDF010e, QXXm, GNar, DELS, R2e, RDF045u, WW, ESpm04d, RDF030p, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.926	0.858	0.721	0.112	6.253	0.000
7	R3e, RDF030u, RDF010e, QXXm, GNar, R2e, RDF045u, WW, ESpm04d, RDF030p, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.926	0.857	0.728	0.109	6.654	0.000
8	R3e, RDF010e, QXXm, GNar, R2e, RDF045u, WW, ESpm04d, RDF030p, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.925	0.856	0.735	0.107	7.050	0.000
9	R3e, RDF010e, QXXm, GNar, R2e, WW, ESpm04d, RDF030p, RDF025u, H2u, EEig04r, IAC, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.925	0.855	0.741	0.104	7.492	0.000
10	R3e, RDF010e, QXXm, GNar, R2e, WW, ESpm04d, RDF030p, RDF025u, H2u, EEig04r, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.924	0.854	0.747	0.102	7.957	0.000
11	R3e, RDF010e, QXXm, GNar, R2e, WW, ESpm04d, RDF030p, RDF025u, EEig04r, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.923	0.852	0.751	0.100	8.427	0.000
12	R3e, RDF010e, GNar, R2e, WW, ESpm04d, RDF030p, RDF025u, EEig04r, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.923	0.852	0.757	0.098	8.985	0.000
13	R3e, RDF010e, GNar, R2e, WW, ESpm04d, RDF030p, EEig04r, EEig03r, CIC0, QXXv, nC, piPC03, S0K, X0v, Ss, ESpm03d, H2e, RDF025e, ESpm01r, RDF010u, QXXp	0.922	0.850	0.761	0.096	9.556	0.000
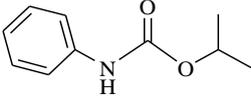
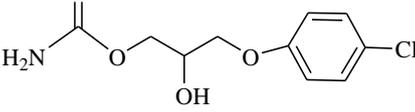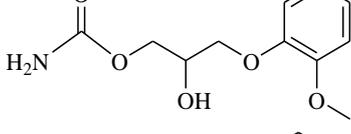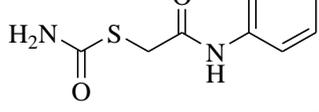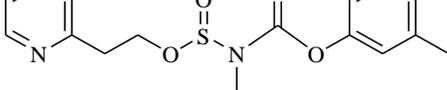| 55 | Propham | | 3.3344 |
56	Carbamic acid, acetylmethyl-, 2-(1,1-dimethylethyl)-2-methyl-1,3-benzodioxol-4-yl ester		3.6020
57	Fenoxycarb		3.6990
58	Felbamate		3.6990
59	Hydroxyphenamate		3.6990
60	Tiocarbazil		3.9031
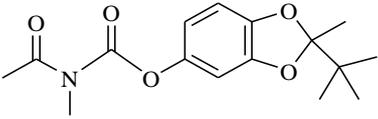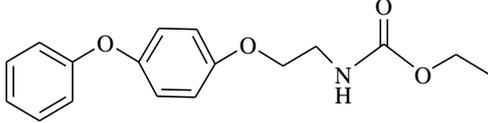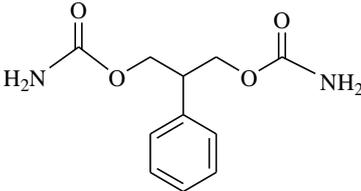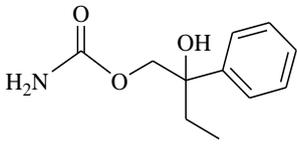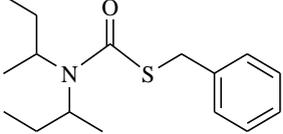
First, the structure of the carbamates derivatives was drawn by GaussView (5.0) software, and optimization was carried out by Gaussian 09 software (Frisch et al, 2009) with Becke, 3-parameter, Lee–Yang–Par (B3LYP) theory level (Tsuneda, 2014) and 6-31G* basis set (Tsuneda and Hirao, 2014) Second, the Dragon software (online 5.4 version) was used to compute the molecular descriptor of studied carbamates derivatives, and traded into the MATLAB (version 2017a) environment (Todeschini and Consonni, 2008). Third, inappropriate descriptors were removed using GA method, and the most suitable ones were extracted for QSTR modeling (Ahmadi and Habibpour, 2017; Ahmadi and Ganji, 2016; Mirjalili, 2019; Glavanovic et al, 2016; Liu et al, 2020). The GA is an accidental method that solves the optimization problems defined by fitness criteria, applying the evolution theories of Darwin, unlike genetic functions such as crossover and mutation (Lin et al, 2018; Niazi and Leardi, 2012; Sarkhosh et al, 2014; Leardi, 2003; Niazi and Leardi, 2012). Finally, the GA-BWMLR technique was performed using the Statistical Package for the Social Sciences (SPSS, version 22) software (Randic and Basak, 2000; Hocking, 2013), and the GA-BPANN model was presented with Neural Network Toolbox in MATLAB software (Sigmon and Davis, 2004).
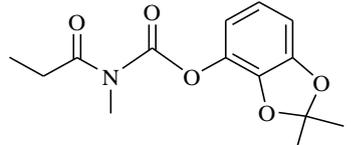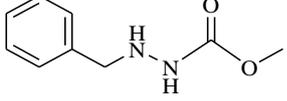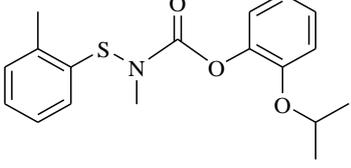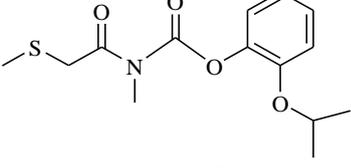
RESULTS AND DISCUSSION
Multiple linear regression method

A total of 3,224 descriptors computed by the Dragon software were traded into the MATLAB program. All inappropriate and repetitive descriptors were removed using the GA, and 34 descriptors were selected for GA-BWMLR analysis. Data analysis resulted in 23 models with 34 to 12 descriptors. Table 2 shows the regression parameters and quality of correlation of proposed models in GA-BWMLR method.
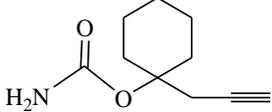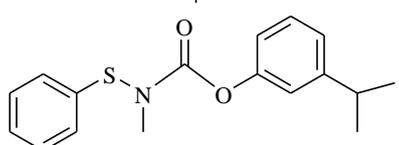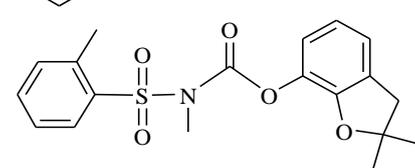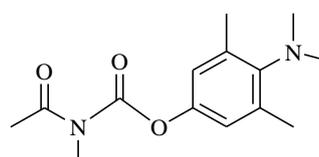
43	Hexapropymate		2.9542
44	Methylpentynol carbamate		2.9542
45	Ethanol, 2-chloro-, 1-carbamate		2.9777
46	Carbamic acid, methyl((phenylthio)sulfinyl)-, 1-naphthalenyl ester		3.0000
47	Mephenesin carbamate		3.0211
48	Carbamic acid, butyl-, 2-chloroethyl ester		3.0334
49	Styramate		3.0944
50	1-Propanol, 3-(3,4-dimethoxyphenyl)-, carbamate (ester)		3.1139
51	Carbamic acid, 3-(o-hydroxyphenyl)propyl ester		3.1300
52	Promacyl		3.3010
53	Carbamic acid, methyl(1-oxobutyl)-, 2,2-dimethyl-1,3-benzodioxol-4-yl ester		3.3010
54	1-Oxa-4-thiaspiro(4.5)decane-2-methanol, carbamate		3.3010
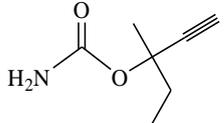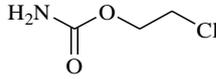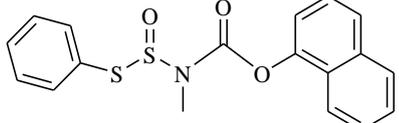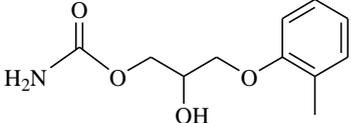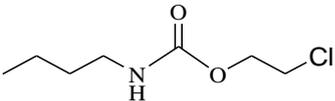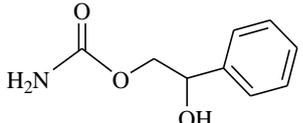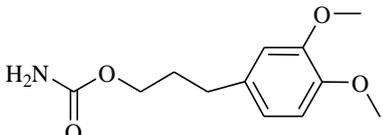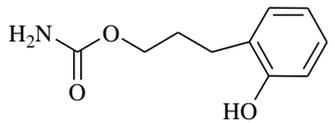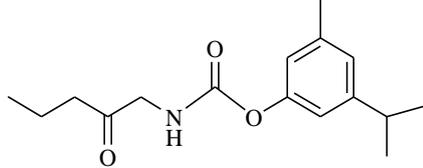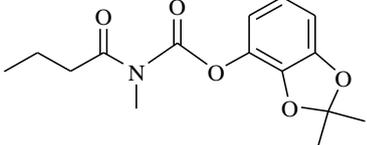
21
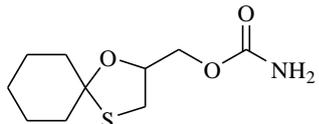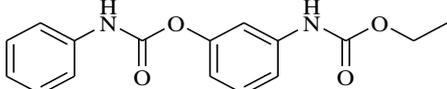
31	Desmedipham		2.6990
32	Carbamic acid, alpha-(chloromethyl)phenethyl ester		2.7634
33	Carbamic acid, methyl((2-pyridinylmethoxy)sulfinyl)-, 3-methylphenyl ester		2.8751
34	Carbamic acid, methyl((3-(2-pyridinyl)propoxy)sulfinyl)-, 3-methylphenyl ester		2.8751
35	Emylcamate		2.8808
36	2-Propyn-1-ol, 3-phenyl-, carbamate (ester)		2.9031
37	Betaquil		2.9031
38	Tybamate		2.9031
39	Chlorphenesin carbamate		2.9069
40	Methocarbamol		2.9095
41	Carbamothioic acid, S-(2-oxo-2-(phenylamino)ethyl) ester		2.9191
42	Carbamic acid, methyl((2-(2-pyridinyl)ethoxy)sulfinyl)-, 3-methylphenyl ester		2.9395
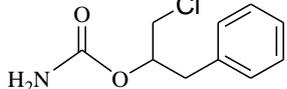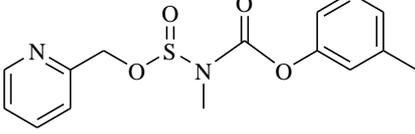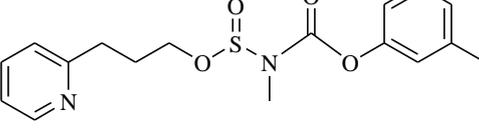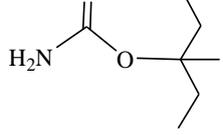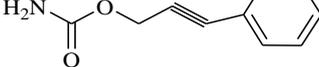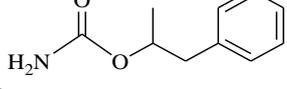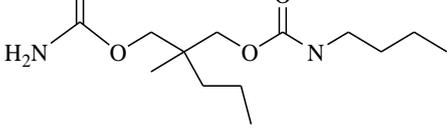
20	Carbamic acid, methyl((phenylmethoxy)thio)-, 3-(1-methylethyl)phenyl ester		2.3222
21	Thiodicarb		2.3541
22	Carbamic acid, methyl(1-oxobutyl)-, 1,3-benzodioxol-4-yl ester		2.4471
23	Carbamic acid, methyl(1-oxopropyl)-, 2,2-dimethyl-1,3-benzodioxol-4-yl ester		2.4471
24	Carbamic acid, acetylmethyl-, 2,2-dimethyl-1,3-benzodioxol-4-yl ester		2.4771
25	Carbamic acid, methyl(phenylthio)-, o-isopropoxyphenyl ester		2.4771
26	Methylpentynol carbamate		2.5276
27	Carbazic acid, 3-benzyl-, methyl ester		2.5740
28	Carbamic acid, methyl((2-methylphenyl)thio)-, o-isopropoxyphenyl ester		2.6020
29	Carbamic acid, methyl ((methylthio)acetyl)-, o-isopropoxyphenyl ester		2.6335
30	Ethinamate		2.6902
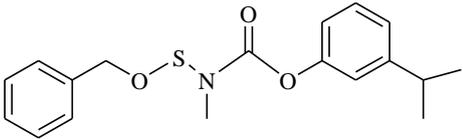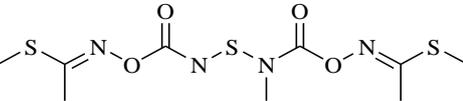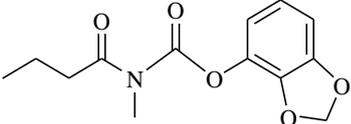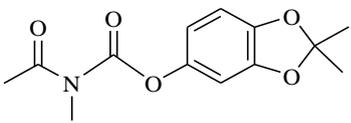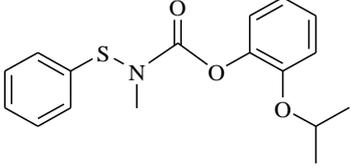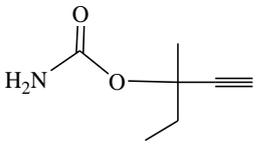
19
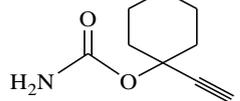
10	Pyrolan (3-Methyl-1-phenyl-1H-pyrazol-5-yl dimethylcarbamate)		1.9542
11	Carbamic acid, methyl((3-(3-pyridinyl)propoxy)sulfinyl)-, 3-(1-methylethyl)phenyl ester		2.0000
12	Trimethacarb		2.0043
13	Carbamic acid, methyl((2-pyridinylmethoxy)sulfinyl)-, 3-(1-methylethyl)phenyl ester		2.0212
14	Carbamic acid, acetylmethyl-, 2-ethyl-2-methyl-1,3-benzodioxol-4-yl ester		2.0969
15	Carbaryl		2.1072
16	Carbamic acid, methyl((2-(2-pyridinyl)ethoxy)sulfinyl)-, 3-(1-methylethyl)phenyl ester		2.1139
17	Carbamic acid, methyl(phenylthio)-, m-isopropylphenyl ester		2.1760
18	N-(2-Toluenesulfonyl) carbofuran		2.1760
19	Carbamic acid, acetylmethyl-, 4-(dimethylamino)-3,5-xylyl ester		2.3010
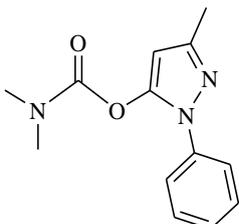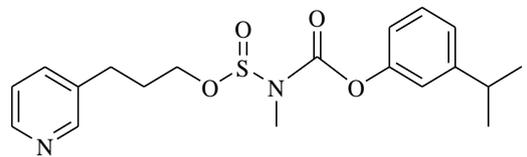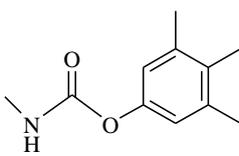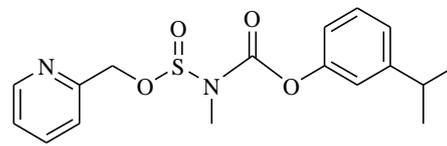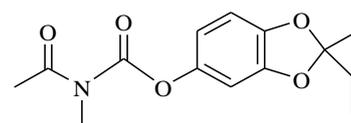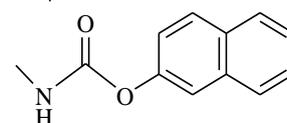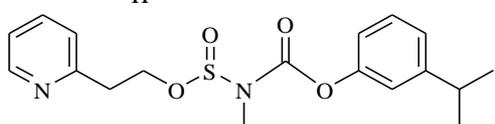
Table 1. The chemical structure of studied carbamates.

NO	Name	Structure	Log LD$_{50}$
1	Methomyl		1.0000
2	Carbamic acid, methyl((pentachlorophenyl)thio)-, 2,3-dihydro-2,2-dimethyl-7-benzofuranyl ester		1.3010
3	Carbamic acid, methyl(phenylthio)-, 2,3-dihydro-2,2-dimethyl-7-benzofuranyl ester		1.3979
4	Xylylcarb		1.6580
5	Carbamic acid, methyl((1-methylethoxy)thio)-, 2,3-dihydro-2,2-dimethyl-7-benzofuranyl ester		1.7404
6	Carbamic acid, methyl((phenylmethoxy)thio)-, 2,3-dihydro-2,2-dimethyl-7-benzofuranyl ester		1.7708
7	Dimetilan		1.7781
8	Pirimicarb		1.8325
9	Pirimicarb		1.8325
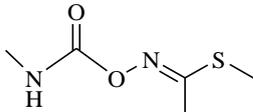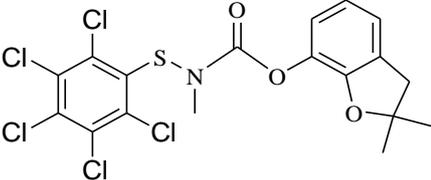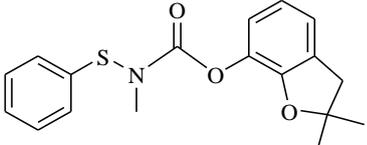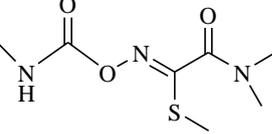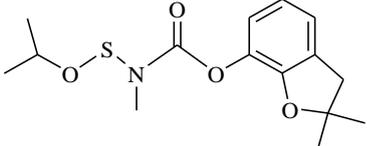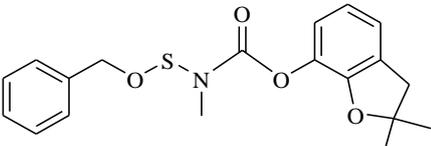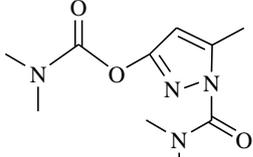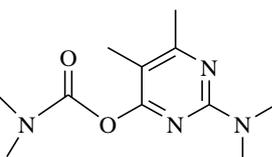
17
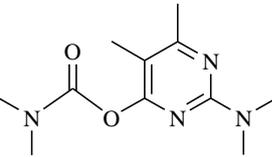
In a research, a large number of physicochemical, thermodynamics, and structural descriptors were calculated for pesticides. Then, several robust QSAR models with high values of coefficient of determination (for training and test sets) were built using a combination of molecular descriptors. Finally, the quantitative structure-toxicity relationship (QSTR) study was used when applying the modeling on toxicological or pharmacokinetic systems (Gupta, 2011; Yang et al, 2020). Bermudez-Saldana and Cronin used multiple linear regression (MLR) and partial least squares (PLS) regression methods to investigate the toxicity of a chemically heterogeneous set of organophosphorus and carbamate pesticides to rainbow trout (Oncorhynchus mykiss Walbaum) (Berudez-Saldana et al, 2009). A structure-activity relationship (SAR) of a series of O-biphenyl carbamates as dual modulators of dopamine receptor and fatty acid amide hydrolase was reported (De Simone et al, 2017). Kumar *et al.* performed atom-based 3D-QSAR model for both targets, including acetylcholinesterase and monoamine oxidase B enzymes, which provided basis for new structural scaffold to serve as building blocks in designing drug-like molecules for Alzheimer's disease (Kumar et al, 2014). Some 2D-QSAR studies used density functional theory (DFT) and Lipinski's descriptors through MLR model to explore the relationships between the structural features of 36 carbamate derivatives and inhibitory activity of butyrylcholinesterase agents (Nour et al, 2022). Also, QSAR analyses were independently performed on data sets belonging to both organophosphates and carbamates of insecticides . (Naik et al, 2009). The wide use of pesticides has received increasing attention in regulatory agencies due to their wide overuse and different adverse effects on all living organisms. In this regard, organizations such as the United States Environmental Protection Agency (EPA) and European Chemicals Agency (ECHA) put laws into effect that pesticides should be fully evaluated before marketing. Thus, techniques based on the QSAR methods using linear models such as MLR (Souyei et al, 2019; Zapadka et al, 2019) PLS (Chen et al, 2019; Gaullier et al, 2019) and principal component analysis (PCA) (Tripathi and Singal, 2019; Amari, 1990) and nonlinear models such as artificial neural networks (ANN) (Adad et al, 2013; Amiri et al, 2020; Bora et al 2019) and k-nearest neighbor (KNN) were successfully developed to model and predict the activity of chemical compounds.

Accordingly, in this research, we evaluated toxicity of pesticides using genetic algorithm (GA) with backward stepwise multiple linear regression (GA-BWMLR) (Kutner et al, 2005) and backpropagation artificial neural networks (GA-BPANN) models (Thapliyal et al, 2018).

EXPERIMENTAL METHOD
The chemical structures of 60 types of carbamates and their lethal dose (LD_{50}) values in mouse were obtained from an available toxicological experimental database (Table 1) (Chem ID plus).

INTRODUCTION

Carbamate derivatives are generally used as insecticides, pesticides, and herbicides. The general formula of carbamate pesticides is RHNCOOR. They have low polarity, chemical reactivity, and high solubility in water (Roberts et al, 1998; Piel et al, 2019). If carbamate pesticides are used properly, they can increase agricultural production and protect humans and animals from insect-vector-mediated diseases. Moreover, pesticides have relatively provided significant benefits by protecting humans from disease threats, and increasing the potency to produce food and fiber. However, overexposure of humans and animals to these compounds often results in poisonings. Recently, the misuse of pesticides has become a major environmental concern. Since pesticides are toxic to living organisms, there is major public concern over the effect of these substances not only on human beings but also on beneficial organisms in the environment. In this respect, it is believed that birds, fish, mice, and other organisms might be affected by pesticides (Testa et al, 2003; Ghosh and Brindisi, 2015). The mechanism of action of the carbamates and organophosphates is related to the inhibition of the enzyme cholinesterase. Carbamate insecticides yield their toxicity by inhibiting acetylcholinesterase (ACHE) enzyme (Roberts et al, 1999). which hydrolyzes the pod synaptic effector, acetylcholine into choline and acetic acid. The inhibition of ACHE leads to the buildup of acelylcholine in the postsynaptic membrane, resulting in a constant nerve stimulation with fatal results.

This stimulation manifests itself by ungovernable movements and paralysis in insects (Ballantyne and Marrs, 2017; Pope et al, 2005; Pohanka, 2012). Quantitative structure-activity relationship (QSAR) technique can provide information about the relationship between chemical structure with biological activity of a compound, which is important in selecting the compound or removing the compound before its synthesis (Lee and Barron, 2016; Sun et al, 2021) especially when experimental testing is not possible for a compound. The QSAR technique is based on the assumption that similar compounds have structurally similar activities. This technique makes a correlation between the activity, such as toxicity of a certain chemical compound, and its structural properties through a definite mathematical algorithm. Then, this relationship can be used in the prediction, interpretation, and assessment of desired activities of new compounds with reducing and rationalizing time. The QSAR study suggests to identify the essential structural features and physicochemical properties in carbamate derivatives (Yee and Wei, 2012). It also provides the possibility to make predictions of designed compounds before the chemical synthesis of novel analogues, and at the same time helps to understand the interactions between functional group of designed molecules and the activity of target molecule (Toropov et al, 2015; Cappelli et al, 2015; Amiri et al, 2020; Fourches and Ash, 2019). Molecular descriptors are the most important components of the QSAR and can be obtained experimentally or through mathematical formulas, such as quantum mechanics and chemical graph theory. These descriptors have different kinds, including constitutional, steric, geometric, electrostatic, quantum chemical, lipophilic, electronic, and topological, which describe the structure of molecules and help to predict the activity and properties of molecules in complex experiences.

15

Applying quantitative structure-toxicity relationship (QSTR) model to predict the toxicity of pesticide carbamates using computation methods and molecular descriptors

S. A. Moosavi[1], E. Mohammadinasab[2*], T. Momeni Isfahani[3]

1. PHD student, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran
2. Assistant Professor, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran
3. Assistant Professor, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran

Abstract

In this study, we performed quantum mechanics computation at density function theory level with 6-31G* basis set to construct a quantitative structure-toxicity relationship (QSTR) model for predicting lethal dose (LD_{50}) pesticide carbamates derivatives. The best molecular descriptors were selected using genetic algorithm (GA) by MATLAB software. Then, we studied the relationship between the selected descriptors and the $logLD_{50}$ of carbamate derivatives using backward-stepwise multiple linear regression (BW-MLR) and backpropagation artificial neural network (BP-ANN) models. The RDF010e, WW, and R3e descriptors were applied for modeling the GA-BWMLR and GA-BPANN models. The comparison of results illustrated that the R^2 and Q^2 of GA-BPANN model for all set were significantly higher than the GA-BWMLR model. The GA-BPANN model was more accurate with lower mean square error (MSE), root-mean-square error (RMSE), standard error of prediction (SEP), and absolute average deviation (ADD) values of data set for predicting the LD_{50} of studied carbamates.

Keywords: **"**Pesticide**"**; **"**QSTR**"**; **"**Toxicity**"**; **"**Carbamates**"**; **"**GA-BWMLR**"**; **"**GA-BPANN**"**