Toxicology in Vitro 72 (2021) 105089

Contents lists available at ScienceDirect

Toxicology in Vitro

journal homepage: www.elsevier.com/locate/toxinvit

In silico prediction of chemical acute contact toxicity on honey bees *via* machine learning methods

Xuan Xu, Piaopiao Zhao, Zhiyuan Wang, Xiaoxiao Zhang, Zengrui Wu, Weihua Li, Yun Tang *, Guixia Liu *

Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

ARTICLE INFO

Keywords:
Acute contact toxicity
Applicability domain
Honey bees
Machine learning
Structural alerts

ABSTRACT

In recent years, the decline of honey bees and the collapse of bee colonies have caught the attention of ecologists, and the use of pesticides is one of the main reasons for the decline. Therefore, ecological risk assessment of pesticides is essential and necessary. *In silico* tools, such as QSAR models can play an important role in predicting physicochemical and biological properties of chemicals. In this study, a total of 54 classification models were developed by combination of 6 machine learning methods along with 9 kinds of molecular fingerprints based on the experimental honey bees acute contact toxicity data (LD_{50}) of 676 structurally diverse pesticides. The best model proposed was SVM algorithm combined with CDK extended fingerprint. The analysis of the applicability domain of the model successfully excluded some extreme molecules. Additionally, 9 structural alerts about honey bees acute contact toxicity were identified by information gain and substructure frequency analysis.

1. Introduction

Honey bees (*Apis mellifera*) (HBs) are the most important pollinators in the global natural ecosystems. Many scientific studies have highlighted the economic and ecological importance of HBs as natural pollinators, not only for crops, but also for forest plants and tropical ecosystems (Spivak et al., 2011; dos Santos et al., 2016). However, in the past decade, many countries have reported a sharp decline in the number of HBs populations, with HBs on farms often mysteriously dying, leading to colony collapse disorder (CCD) (Parmentier et al., 2014; Fevery et al., 2016). The large-scale use of pesticides is an important reason for the decline in the population density of HBs. HBs are frequently exposed to pesticides, because they carry pollen from one plant to another, and HBs are also insects, are sensitive to any poison used to kill pests (Codling et al., 2016). As these serious consequences not only endanger the maintenance of wild plant biodiversity, but also have negative economic effects on agricultural production. Therefore, honey bee toxicity (HBT) assessment has become an essential part of the register of new pesticide (Johnson, 2015).

HBs are non-target test species recommended by the United States Environmental Protection Agency (US.EPA) for terrestrial toxicity assessment of chemicals (USEPA, 2016). According to the principles from Organisation for Economic Co-operation and Development (OECD), the toxicity of chemicals in the HBs should be assessed with 24 or 48 h of acute contact toxicity experimental biostatistics to determine the median lethal concentration that induces 50% death (LC_{50}) or lethal dose that induces 50% death (LD_{50}) (OECD, 1998). However, traditional experimental tests to assess the toxicity of a large number of chemicals to HBs are not only costly and time-consuming, but also have an ethical problem. Therefore, it is urgent to develop alternative methods and tools for assessing the toxicity of HBs (Li et al., 2017). The regulation "Registration, Evaluation and Authorization of Chemicals (REACH)" advocated the use of non-animal testing methods, and in particular quantitative structure activity/toxicity relationship (QSAR/QSTR) methods to reduce the amount and cost of animal testing (Taylor et al., 2014). The establishment of a computational model based on the QSAR/QSTR method is a widely used approach. The QSAR/QSTR model can predict a variety of endpoints and has already been successfully applied for various species and ecotoxicological endpoints (Drgan et al., 2016; Croce et al., 2017; Toropov et al., 2017).

Although HBT is important for the terrestrial toxicological assessment of pesticides, QSTR models specifically designed to predict HBT are quite limited. In 1991, Vighi et al. developed a QSAR model for predicting HBT based on 15 organophosphorus pesticides (Vighi et al., 1991). In 2003, Celli and Maccagnani established a QSAR model for the prediction of HBT with 100 organophosphorus pesticides using a

* Corresponding authors.
E-mail addresses: ytang234@ecust.edu.cn (Y. Tang), gxliu@ecust.edu.cn (G. Liu).

https://doi.org/10.1016/j.tiv.2021.105089
Received 28 November 2020; Accepted 6 January 2021
Available online 11 January 2021

Table 1
The specific pruning criteria for each database.

Database	Species	Route	Unit (LD_{50})	Exposure time	Chemical purity
PPDB[a]	Honey bee (*Apis spp.*)	Contact	μg/bee μg/piece	48 h	–
ECOTOX[b]	*Apis mellifera*	Dermal Topical	μg/bee μg/org	48 h	>90
EFSA[c]	Honey bee	Dermal	μg/bee	48 h	–

[a] PPDB: Pesticide Properties Database.
[b] ECOTOX: The ECOTOXicology konwledgebase of USEPA.
[c] EFSA: European Food Safety Authority.

multilayer feedforward neural network method (Celli and Maccagnani, 2003). Devillers et al. also proposed a QSAR model based on neural network for predicting HBT in the same year (Devillers et al., 2003). In 2010, Cheng et al. developed both the classification and regression models for pesticide toxicity to HBs using five machine approaches (Cheng et al., 2010). In 2014, Singh et al. established global models for qualitative and quantitative HBT prediction based on the experimental toxicity data of 237 structurally diverse pesticides using Probabilistic neural network (PNN) (Singh et al., 2014). In 2017, Como et al. established a series of global models for HBT prediction of pesticides using k-Nearest Neighbor ($k-$NN) algorism based on 256 pesticides with acute contact toxicity data collected from different sources. The same year (Como et al., 2017), Li et al. developed a QSAR model (SubFP_SVM) to predict HBs acute contact toxicity based on Como's work (Li et al., 2017). In 2020, Wang et al. developed a QSAR model for predicting HBT using the Graph attention convolutional neural network (GACNN) based on 900 pesticide data, and the prediction accuracy of the test set reached 0.837, but the 900-pesticide data included acute contact toxicity to bees and acute oral toxicity to bees, as well as toxicity from unknown sources to bees (Wang et al., 2020).

However, almost all of these models had some shortcomings, for example, the amount of data used to build the model was small and the data source was mixed, the predictive performance of some models was not satisfactory, and the applicability domain of most of the models was not analyzed, which limited the use of the model and the scope of research. In this study, we used six machine learning methods combined with nine molecular fingerprints to establish binary classification models for predicting the HBs acute contact toxicity. To ensure the authenticity of the data, we collected all the latest data manually from three public databases. In addition, the applicability domain of the models was analyzed, which helped us exclude some extreme compounds. At the same time, differences of key physical chemical properties and structural characteristics between the chemicals with HBT and those without HBT were analyzed by information gain and substructure frequency analysis.

2. Materials and methods

2.1. Data collection and preparation

In this study, pesticide data on honey bees (*Apis mellifera*) acute contact toxicity were collected manually from three publicly available databases, including the Pesticide Properties Database (PPDB) (http://sitem.herts.ac.uk/aeru/ppdb/en/index.ht m) developed by the Agriculture & Environment Research Unit (AERU) at the University of Hertfordshire, the ECOTOX database (https://cfpub.epa.gov/ecotox/) created by the US. EPA, and the European Food Safety Authority (EFSA) (https://ww w.efsa.europa.eu/en/).

The criteria for data screening are based on guidelines issued by the OECD in 1998, according to which pesticides are administered by contact routes to represent the type of exposure under field conditions. *Apis mellifera* is used as a substitute for assessing risks to HBs. For all collection of pesticides, values of 48 h acute contact exposure based median lethal dose (LD_{50}, μg/bee) for *Apis mellifera* acute contact toxicity are reported (OECD, 1998). The specific pruning criteria for each database were listed in Table 1.

Table 2
Summary of data set used in this study.

Datasets	Positive	Negative	Total number
Training	144	396	540
Test	32	104	136
Total	176	500	676

In order to get a data set with high quality, the data set was carefully prepared by the following steps: (1) removing mixtures; (2) removing inorganic and organometallic compounds; (3) removing all the stereoisomer; (4) salts were converted to their parent forms; (5) removing duplicate compounds. At the same time, we matched the canonical Simplified Molecular Input Line Entry System (SMILES) codes and Chemical Abstracts Service (CAS) registry numbers for all compounds from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/). The SMILES format has the advantage of considering some important molecular features, such as the presence of cycles, cis−/trans-isomerism, and the presence of sp^2 and sp^3 atoms (Toropov et al., 2011).

2.2. Data partitioning

In 2014, the US.EPA classified pesticides into three categories based on the acute contact toxicity endpoint (US. EPA, 2014). The pesticide was classified as practically non-toxic ($LD_{50} \geq 11$ μg/bee), moderately toxic (10.9 μg/bee $\geq LD_{50} \geq 2$ μg/bee), or highly toxic ($LD_{50} < 2$ μg/bee). According to this classification criteria, the dataset included 144 compounds with high toxicity, 32 compounds with moderate toxicity, and 500 compounds with nontoxicity. In order to establish a binary classification model for acute contact toxicity to HBs, we merged the high and moderate toxicity into the toxic dataset (176 compounds), which was classified as the positive dataset. The nontoxicity dataset was classified as the negative dataset (500 compounds).

These 676 compounds were randomly divided into a training set (80%) for training models and a test set (20%) for evaluating the performance of models. The number of compounds in each dataset was listed in Table 2.

2.3. Calculation of molecular fingerprints

In this study, molecular fingerprints were used to represent compounds. Molecular fingerprints have been widely used in chemical toxicity and activity prediction (Alberga et al., 2019), compound similarity search (Vogt and Bajorath, 2013) and virtual screening (Montaruli et al., 2019), because fingerprints are generated directly from chemical structures and can be easily translated into 2D fragments, and each molecule is described as a binary string of structural keys. Nine kinds of molecular fingerprints for all the compounds were calculated by PaDEL-Descriptor software (Yap, 2011). Including the Atom Pair 2D fingerprint (AP2DFP, 780 bits), Estate fingerprint (EstFP, 79 bits), CDK fingerprint (FP, 1024 bits), CDK extended fingerprint (ExtFP, 1024 bits), CDK graph only fingerprint (GraphFP, 1024 bits), Klekota-Roth fingerprint (KRFP, 4860 bits), MACCS fingerprint (MACCSFP, 166 bits), PubChem fingerprint (PubChemFP, 881 bits) and Substructure fingerprint (SubFP, 307 bits).

2.4. Construction of classification models

Machine learning is widely used to build binary classification models. Among a multitude of available modeling algorisms, we applied six machine learning methods. They are the artificial neural network (ANN) (Djemili et al., 2016), C4.5 decision tree (DT) (Quinlan, 1986), k-nearest neighbor (KNN) (Cover and Hart, 1967), Naïve Bayes (NB)

(Watson, 2008), random forest (RF) (Breiman, 2001) and support vector machine (SVM) (Cortes and Vapnik, 1995). All methods are calculated using Python scripts based on the Scikit-Learn package (Pedregosa et al., 2011).

2.5. Assessment of the models

The classification models were assessed by 10-fold cross-validation and external validation (Du et al., 2017). In 10-fold cross-validation, the original training data set was randomly divided into 10 equal-sized subsets. In 10 subsets, nine subsets were used as a new training data set, and the remaining a subset was reserved as verification data for testing the model. Repeat the cross-validation process 10 times, use each of the 10 subsets only once as the verification data, and finally average the 10 results. The external validation was given by the prediction of the test set.

All models were evaluated by the metrics of true positives (TP), true negative (TN), false positive (FP), and false negative (FN). Meanwhile, four measurements of model performance were used, namely classification accuracy (CA), sensitivity (SE), specificity (SP) and F1Score. CA is the total percentage of chemicals that were correctly predicted. SE and SP are two statistical parameters applicable to binary classification. SP denotes the predictive accuracy of the negative chemicals, and SE means the predictive accuracy of the positive chemicals. The F1Score can be interpreted as a weighted average of the precision and recall, where an F1Score reaches its best value at 1 and worst score at 0. The CA, SE, SP and F1Score of the models were calculated using the following equations:

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$SE = \frac{TP}{TP + FN} \quad (2)$$

$$SP = \frac{TN}{TN + FP} \quad (3)$$

$$Recall = SE \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$F1Score = 2*\frac{Precision*Recall}{Precision + Recall} \quad (6)$$

In addition, the value of area under receiver operating characteristic (ROC) curve (AUC) was also calculated for all models (Hanley and McNeil, 1982). The principle is that if the plot has a surface area (AUC) of 1, the classifier is perfect, and if the AUC equals 0.5, the classifier is a useless random one.

2.6. Analysis of applicability domain of models

According to the OECD principles for the QSAR models, all QSAR models must have a defined applicability domain (AD) (OECD, 2007). AD is widely used to express the scope and limitations of models. In this study, a specific AD based on similarity was employed to avoid the prediction for new compounds with substantially different structures from those in the training set (Netzeva et al., 2005; Roy et al., 2015; Sun et al., 2018). We compared the distance between a new molecule and its K-nearest neighbors in the training set based on a threshold of AD, which was calculated as follows:

$$D_T = \bar{\gamma} + Z\sigma \quad (7)$$

where $\bar{\gamma}$ is the average Euclidean distance of each molecule in the training set and its nearest neighbor, σ is the standard deviation of all Euclidean distances, and Z is an arbitrary parameter to stand for the significance level. If the distance of a molecule in the test set from at least one of its k-nearest neighbors in the training set exceeds the calculated threshold D_T, it will be considered as an outlier by the models.

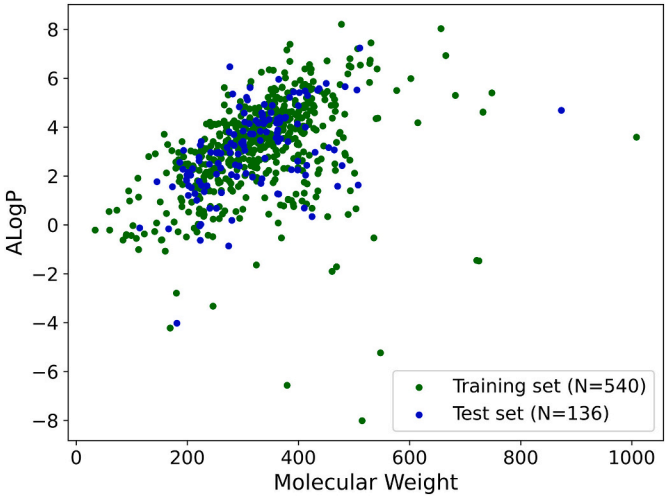


Fig. 1. Chemical space distribution of the training set and test set. The chemical space is defined by molecular weight (MW) as the X-axis, and Ghose-Crippen logKow (ALogP) as the Y-axis. N represents the number of chemicals in different datasets.

2.7. Analysis of structural alerts

Structural alerts (SAs) or privileged substructures are defined as molecular functional groups which can make chemicals toxic. Here, SAs were analyzed using the methods of information gain (IG) and substructure frequency analysis (Shen et al., 2010; Yang et al., 2017; Cao et al., 2018). IG measures the information entropy of a classification system obtained for class prediction by knowing the presence or absence of a pattern in a molecule. If a substructure presented more frequently in chemicals with toxicity than those with nontoxicity, this substructure would be defined as a structural alert (Fan et al., 2018). The presence of SAs can alert investigators to the potential toxicity of test chemicals. The frequency of a fragment (FR) is defined as below:

$$FR = \frac{\left(N^B_{fragment} \times N_{total}\right)}{\left(N_{fragment\_total} \times N_B\right)} \quad (8)$$

where $N^B_{fragment}$ is the number of chemicals containing the fragments in positive chemicals; N_{total} is the total number of compounds in the dataset; $N_{fragment\_toatl}$ is the total number of compounds containing the fragments, and N_B represents the number of positive chemicals.

3. Results

3.1. Data collection and diversity analysis

The chemical spatial distribution of compounds and the diversity of data sets play an important role in the results of classification prediction model. In this study, we collected manually 676 pesticides data from three different publicly available databases; the categories of pesticides include insecticides, herbicides, fungicides, acaricides, and insect growth regulators. The median lethal dose (LD_{50}) values of compounds ranged from 0.0015 μg/bee to above 100,000 μg/bee. The detailed information about the data set was listed in Supplementary Material Table S1.

In order to investigate the chemical spatial distribution, we calculated the molecular weight (MW) and Ghose-Crippen LogKow (ALogP) for the training set and the test set. The values of MW mainly ranged

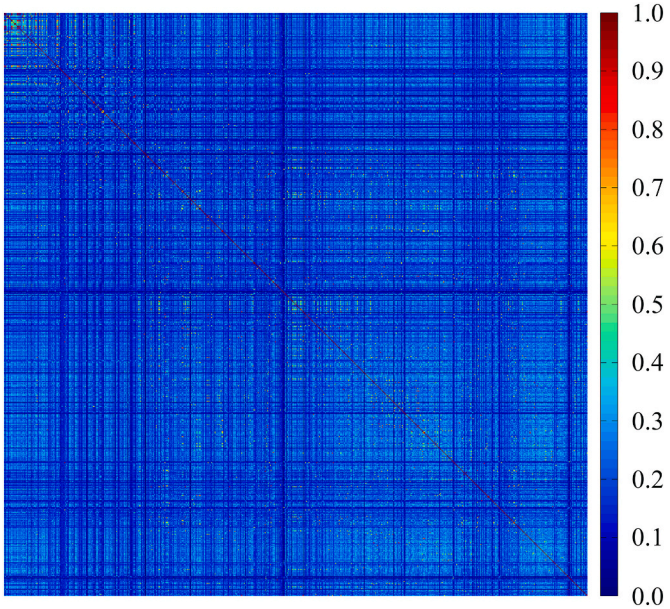


Fig. 2. Heat map of the molecular similarity plotted by the Tanimoto similarity using the CDK extended fingerprint.

from 150 to 600, and ALogP mainly ranged from −1 to 7 (Fig. 1). And the specific MW and ALogP values for each chemical were listed in Supplementary material Table S1. These data illustrated that the 676 pesticides in our data sets covered an enough large chemical space. From the chemical spatial distribution scatter diagrams of the two datasets (Fig. 1), it can be observed that the training set shared a similar chemical space with the test set.

In addition, in order to further explore the chemical diversity of the data sets, the Tanimoto coefficient was also calculated based on CDK extended fingerprint, which has been widely used to evaluate similarities between chemicals (Butina, 1999). The heat maps of the Tanimoto coefficient of data sets were shown in Fig. 2. Red points indicate the high similarity between two chemicals and blue points indicate the low similarity. Meanwhile, the average Tanimoto similarity indexes of the whole data sets was 0.19, indicating that the structures of the chemicals in the data sets were quite diverse.

3.2. Model building and validation

In this study, a total of 54 binary classification models were developed by combination of six machine learning methods along with nine molecular fingerprints. All the models were verified by 10-fold cross-validation and external validation. The detailed values of AUC, CA, SE, SP and F1Score for the 10-fold cross-validation and external validation of models were summarized in Supplementary material Table S2 and Table S3, respectively.

According to AUC values of 10-fold cross-validation, the top five models (Table 3) were ExtFP_SVM, FP_SVM, MACCSFP_ANN, MACCSFP_SVM and GraphFP_SVM. In the top five models, the AUC values ranged from 0.901 to 0.912; the CA values ranged from 0.870 to 0.904; the SE values ranged from 0.727 to 0.799; the SP values ranged from 0.908 to 0.948, and the F1Score values ranged from 0.692 to 0.857.

In order to further assess the robust and predictive ability of the top five models that performed best on 10-fold cross-validation, a test set containing 136 compounds was used. The results of the five models for external validation were also shown in Table 3. As shown in the Table 3, the AUC values ranged from 0.847 to 0.924, the CA values ranged from 0.882 to 0.912, the SE values ranged from 0.735 to 0.794, the SP values ranged from 0.912 to 0.951, and the F1Score values ranged from 0.771 to 0.818. The model performing the best (ExtFP-SVM) in the 10-fold cross-validation was still the best one in the test set, indicating that the models were of the good prediction accuracy and the stable robustness.

3.3. Analysis of applicability domain

According to the principles of the OECD validation for QSAR models, each QSAR prediction model must have a clear applicability domain (AD). In this study, after variable screening, when K = 3 and Z = 0.2, the five metrics of all models improved. As shown in Table 4, taking ExtFP_SVM in test set as an example, the performance of in-domain (ID) compounds is better than that out-of-domain (OD) compounds, which confirmed that the application of AD obviously succeeded in ruling out some extreme molecules. In addition, the AD analysis was carried out for the five best models (ExtFP_SVM, FP_SVM, MACCSFP_ANN, MACCSFP_SVM and GraphFP_SVM) obtained above. More detailed results for AD were given in Supplementary Material Table S4.

3.4. Analysis of the structural alerts

In order to investigate structural features between toxic and nontoxic compounds of HBs, IG and substructure frequency analysis were performed for all data sets based on the CDK extended fingerprint. As shown in Table 5, nine important substructures were identified, which showed a higher frequency in toxic compounds than in nontoxic compounds. The higher the IG value is, the higher the proportion of this substructure in toxic compounds is.

The first and second substructures are phosphoric acid derivatives, which are well-known organophosphorus pesticides. Chemicals containing these fragments inhibit the activity of cholinesterase, leading to

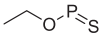
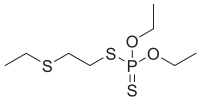
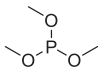
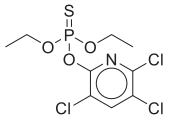
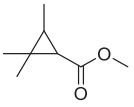
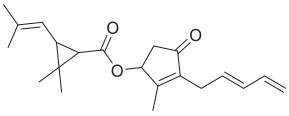
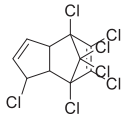
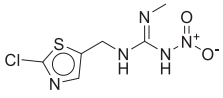
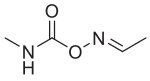
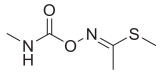
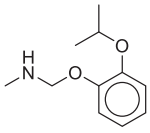
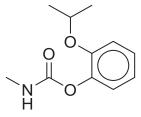
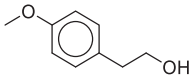
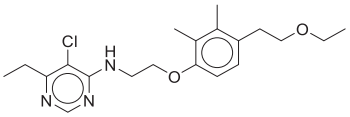
Table 4
Performance of model ExtFP_SVM for in-domain (ID) or out-of-domain (OD) compounds of the test set.

Method	Type (No.)	AUC	CA	SP	SE	F1Score
ExtFP_SVM	ID (113)	0.952	0.912	0.951	0.806	0.833
ExtFP_SVM	OD (23)	0.650	0.870	0.950	0.333	0.400

Table 3
Performance of the best five models for 10-fold cross-validation and external validation.

Data set	Ranking	Method	AUC	CA	SP	SE	F1Score
Training set	1	ExtFP_SVM	0.912	0.890	0.931	0.773	0.857
	2	FP_SVM	0.910	0.904	0.948	0.780	0.692
	3	MACCSFP_ANN	0.910	0.878	0.931	0.727	0.692
	4	MACCSFP_SVM	0.907	0.892	0.925	0.799	0.828
	5	GraphFP_SVM	0.901	0.870	0.908	0.762	0.815
Test set	1	ExtFP_SVM	0.924	0.904	0.951	0.765	0.800
	2	MACCSFP_SVM	0.889	0.912	0.951	0.794	0.818
	3	GraphFP_SVM	0.887	0.882	0.912	0.794	0.771
	4	MACCSFP_ANN	0.867	0.890	0.941	0.735	0.769
	5	FP_SVM	0.847	0.897	0.931	0.794	0.794

Table 5
Nine representative structural alerts (SAs) for HBT identified by information gain (IG) and structural frequency analysis.

ID	SAs	N_P[a]	F_P[b]	N_N[c]	F_N[d]	IG	Representative compound
1		20	3.07	5	0.27	0.021	 **LD_{50}=0.96 μg/bee**
2		11	3.84	0	0	0.032	 **LD_{50}=0.06 μg/bee**
3		30	3.84	0	0	0.091	 **LD_{50}=0.14 μg/bee**
4		7	3.36	1	0.17	0.015	 **LD_{50}=0.53 μg/bee**
5		5	3.20	1	0.23	0.009	 **LD_{50}=0.04 μg/bee**
6		7	3.36	1	0.17	0.015	 **LD_{50}=0.16 μg/bee**
7		5	3.84	0	0	0.015	 **LD_{50}=0.11 μg/bee**
8		6	3.84	0	0	0.018	 **LD_{50}=0.66 μg/bee**
9		8	3.84	0	0	0.023	 **LD_{50}=0.97 μg/bee**

[a]N_P is the number of honey bees toxic compounds with substructure. [b] N_N is the number of honey bees nontoxic compounds with substructure. [c] F_P represents the "frequency of a fragment" in toxic compounds. [d] F_N represents the "frequency of a fragment" in nontoxic compounds.

Table 6
Comparison of our classification model with previous studies for honey bees acute contact toxicity.

No.	Method	Data set	CA	AUC	SP	SE	Reference
1	PNN[a]	237	0.871	–	1.000	0.860	Singh et al., 2014
2	KNN[b]	256	0.840	–	0.860	0.800	Como et al., 2017
3	SVM[c]	256	0.907	0.917	0.936	0.833	Li et al., 2017
4	GACNN[d]	900	0.837	0.838	0.891	0.698	Wang et al., 2020
5	SVM	676	0.904	0.924	0.951	0.765	This study

[a] PNN: Probabilistic Neural Network.
[b] KNN: K-Nearest Neighbor.
[c] SVM: Support Vector Machine.
[d] GACNN: Graph Attention Convolutional Neural Network.

the accumulation of acetylcholine. It can also directly act on cholinergic receptors, thereby affecting the cholinergic receptors and causing neurotoxicity (Casida and Quistad, 2004).

Compounds containing the third substructure are usually pyrethroids, such as lambda-cyhalothrin, tetramethrin and imiprothrin. Pyrethroids are synthetic chemical insecticides whose chemical structures are adapted from the chemical structures of the pyrethrins and act in a similar manner to pyrethrins. They work by altering the nerve function, which causes paralysis in target insect pests, eventually resulting in death (Arthidoro de Castro et al., 2020).

The fourth substructure is allyl chloride, which has been evaluated as a carcinogen. Pesticides that contain this structure are usually organochlorines, such as heptachlor, chlordane, and endrin, which can form central nervous system inhibitors (Bobra et al., 1985; Li et al., 2014).

The fifth substructure is nitramide, which is commonly found in neonicotinoids, such as imidacloprid, clothianidin, and thiamethoxam. Neonicotinoids are neuropalsy agents that, at lethal doses, can cause insects to lose control of their movements, shudder, become paralyzed and die. Under the sublethal concentration, it can cause aphid convulsion, reduce honeydew discharge, and finally die of starvation (Fairbrother et al., 2014; Baines et al., 2017).

In addition, we have discovered four new substructures (compounds 6–9 in Table 5), these important substructures may provide some useful visual warning functions in the field of environmental hazard assessment. If a new compound contains one or more such substructures, it may be a toxic compound to HBs.

4. Discussion

HBT has become an important endpoint in environmental science, especially pesticide registration. In recent years, some studies have also focused on predicting HBT (Li et al., 2017; Cheng et al., 2010; Singh et al., 2014; Como et al., 2017; Wang et al., 2020), It is not appropriate to directly compare our results with previous studies, because different models use different data sets and data description methods. However, a simple comparison of model statistics could provide some basic information about the accuracy of HBT prediction. The detailed comparison results were given in Table 6.

In this study, we developed binary classification models for HBs acute contact toxicity. All models were validated using the 10-fold cross-validation and the external validation. From the performance of 10-fold cross-validation and external validation, different machine learning algorithms and molecular fingerprints have great influence on model's prediction ability. From our results, it was clear that SVM and ANN algorithms were superior to other algorithms, especially SVM algorithm. Among the top five models ranked by AUC values, four models were SVM algorithm, because the SVM algorithm is an excellent kernel-based tool for binary classification and regression (Cortes and Vapnik, 1995), which can deal with high dimensional space problems. So, it is widely used in the establishment of QSAR models. In addition, when SVM algorithm was combined with CDK extended fingerprint, the model performed better, which indicated that the CDK extended fingerprint was more suitable for the SVM algorithm to predict HBT. Besides, MACCS fingerprint is also worthy of attention, as both ANN and SVM algorithms combined with MACCS fingerprint will produce good results. Detailed parameters for building these models were given in Supplementary Material Table S5.

Compared with preciously reported models, to the best of our knowledge, we have collected the largest data on the acute contact toxicity of HBs to develop binary classification prediction models. To ensure the authenticity of the data, we manually collected all the latest data from three public databases (PPDB, ECOTOX, EFSA). The AUC and CA values of the test set of the best model (ExtFP_SVM) are 0.924 and 0.904, respectively. We also analyzed the applicability domain of the model, and excluded some extreme compounds. At the same time, we analyzed the substructures of pesticides that were toxic and non-toxic to HBs, and finally got nine SAs.

Of course, our models also have some shortcomings. The values of SE were all smaller than the SP values due to the imbalance between positive and negative data (toxicity data is less than nontoxicity data), indicating that most pesticides had low toxicity to HBs. Nevertheless, the SE value of our best model was 0.765. In addition, when processing data, since the 2D structures of salts could not be directly converted into the corresponding acids or bases, all inorganic salts, organometallic salts and ammonium salts were removed, so our models could not predict the HBT of salts. In the future, we will combine other *in silico* methods to solve this problem. In conclusion, this study provided critical information and powerful tools for evaluating toxicological properties of pesticides in the environmental hazard assessment.

5. Conclusions

In this study, we used six machine learning methods combined with nine molecular fingerprints to establish 54 binary classification models for prediction of acute contact toxicity on honey bees (*Apis mellifera*). 10-fold cross validation was used to validate the performance of all models and external validation was used to validate the robustness of all models. Based on the five metric values of the training set, SVM algorithm combined with the CDK extended fingerprint was the best model. The AUC value 0.924, the CA value 0.904, the SE value 0.765, the SP value 0.951, and the F1Score value 0.800 were obtained for a test set containing 136 pesticides, which could provide a robust and reliable prediction for the HBT. In addition, through the analysis of the applicability domain of the model successfully excluded some extreme molecules. At the same time, nine SAs were identified by IG and structural frequency analysis. The frequency of these SAs appearing in the toxic compounds was significantly higher than those in the nontoxic compounds. Therefore, these SAs can make researchers pay more attentions to chemicals containing these structural fragments. In summary, our research not only developed a robust and reliable prediction model for the HBs acute contact toxicity of compounds, but also identified some useful substructures for future environmental risk assessment.

Declaration of Competing Interest

There are no conflicts of interest to declare.

Acknowledgments

We gratefully acknowledged the financial supports from the National Key Research and Development Program of China (Grant 2019YFA0904800) and the National Natural Science Foundation of China (Grants 81673356 and 81872800).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tiv.2021.105089.

References

Alberga, D., Trisciuzzi, D., Kamel, M., Mangiatordi, G., Nicolotti, O., 2019. Prediction of acute oral systemic toxicity using a multifingerprint similarity approach. Toxicol. Sci. 167, 484–495. https://doi.org/10.1093/toxsci/kfy255.

Arthidoro de Castro, M.B., Martinez, L.C., Cossolin, J.F.S., Serra, R.S., Serrão, J.E., 2020. Cytotoxic effects on the midgut, hypopharyngeal, glands and brain of *Apis mellifera* honey bee workers exposed to chronic concentrations of lambda-cyhalothrin. Chemosphere 248, 126075. https://doi.org/10.1016/j.chemosphere.2020.126075.

Baines, D., Wilton, E., Pawluk, A., Gorter, M., Chomistek, N., 2017. Neonicotinoids act like endocrine disrupting chemicals in newly-emerged bees and winter bees. Sci. Rep. 7 https://doi.org/10.1038/s41598-017-10489-6.

Bobra, A., Shiu, W.Y., Mackay, D., 1985. Quantitative structure-activity relationships for the acute toxicity of chlorobenzenes to *Daphnia magna*. Environ. Toxicol. Chem. 4, 297–305. https://doi.org/10.1002/etc.5620040305.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/a:1010933404324.

Butina, D., 1999. Unsupervised data base clustering based on Daylight's fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets. J. Chem. Inf. Model. 39, 747–750. https://doi.org/10.1021/ci9803381.

Cao, Q., Liu, L., Yang, H., Cai, Y., Li, W., Liu, G., Lee, P.W., Tang, Y., 2018. In silico estimation of chemical aquatic toxicity on crustaceans using chemical category methods. Environ. Sci. 20, 1234–1243. https://doi.org/10.1039/c8em00220g.

Casida, J., Quistad, G., 2004. Organophosphate toxicology: safety aspects of nonacetylcholinesterase secondary targets. Chem. Res. Toxicol. 17, 983–998. https://doi.org/10.1021/tx0499259.

Celli, G., Maccagnani, B., 2003. Honey bees as bioindicators of environmental pollution. B. Insectol. 56, 137–139.

Cheng, F., Shen, J., Li, W., Lee, P., Tang, Y., 2010. In silico prediction of terrestrial and aquatic toxicities for organic chemicals. Chin. J. Pest. Sci. 12, 477–488. https://doi.org/10.3969/j.issn.1008-7303.2010.04.18.

Codling, G., Al Naggar, Y., Giesy, J.P., Robertson, A.J., 2016. Concentrations of neonicotinoid insecticides in honey, pollen and honey bees (*Apis mellifera* L.) in central Saskatchewan, Canada. Chemosphere 144, 2321–2328. https://doi.org/10.1016/j.chemosphere.2015.10.135.

Como, F., Carnesecchi, E., Volani, S., Dorne, J.L., Richardson, J., Bassan, A., Pavan, M., Benfenati, E., 2017. Predicting acute contact toxicity of pesticides in honeybees (*Apis mellifera*) through a k-nearest neighbor model. Chemosphere 166, 438–444. https://doi.org/10.1016/j.chemosphere.2016.09.092.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/bf00994018.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27. https://doi.org/10.1109/tit.1967.1053964.

Croce, R., Cinà, F., Lombardo, A., Crispeyn, G., Cappelli, C.I., Vian, M., Maiorana, S., Benfenati, E., Baderna, D., 2017. Aquatic toxicity of several textile dye formulations: acute and chronic assays with Daphnia magna and Raphidocelis subcapitata. Ecotoxicol. Environ. Saf. 144, 79–87. https://doi.org/10.1016/j.ecoenv.2017.05.046.

Devillers, J., Hà, M., Delègue, P., Decourtye, A., Budzinski, H., Cluzeau, S., Maurin, G., Ctis, Rillieux, L., Pape, F., 2003. Modeling the acute toxicity of pesticides to *Apis mellifera*. B. Insectol. 56, 103–109.

Djemili, R., Bourouba, H., Amara Korba, M.C., 2016. Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals. Biocybern. Biomed. Eng. 36, 285–291. https://doi.org/10.1016/j.bbe.2015.10.006.

dos Santos, C.F., Acosta, A.L., Dorneles, A.L., dos Santos, P.D.S., Blochtein, B., 2016. Queens become workers: pesticides alter caste differentiation in bees. Sci. Rep. 6, 31605. https://doi.org/10.1038/srep31605.

Drgan, V., Župerl, Š., Vračko, M., Como, F., Novič, M., 2016. Robust modelling of acute toxicity towards fathead minnow (*Pimephales promelas*) using counter-propagation artificial neural networks and genetic algorithm. SAR QSAR Environ. Res. 27, 501–519. https://doi.org/10.1080/1062936x.201 6.1196388.

Du, H., Cai, Y., Yang, H., Zhang, H., Xue, Y., Liu, G., Tang, Y., Li, W., 2017. In silico prediction of chemicals binding to aromatase with machine learning methods. Chem. Res. Toxicol. 30, 1209–1218. https://doi.org/10.1021/acs.chemrestox.7b00 037.

Fairbrother, A., Purdy, J., Anderson, T., Fell, R., 2014. Risk of neonicotinoid insecticides to honeybees. Environ. Toxicol. Chem. 33, 719–731. https://doi.org/10.1002/etc.2527.

Fan, D., Yang, H., Li, F., Sun, L., Di, P., Li, W., Tang, Y., Liu, G., 2018. In silico prediction of chemical genotoxicity using machine learning methods and structural alerts. Toxicol. Res. 7, 211–220. https://doi.org/10.1039/c7tx00259a.

Fevery, D., Houbraken, M., Spanoghe, P., 2016. Pressure of non-professional use of pesticides on operators, aquatic organisms and bees in Belgium. Sci. Total Environ. 550, 514–521. https://doi.org/10.1016/j.scitotenv.2016.01.123.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36. https://doi.org/10.1148/radiology.143.1.7063747.

Johnson, R.M., 2015. Honey bee toxicology. Annu. Rev. Entomol. 60, 415–434. https://doi.org/10.1146/annurev-ento-011613-162005.

Li, X., Chen, L., Cheng, F., Wu, Z., Bian, H., Xu, C., Li, W., Liu, G., Shen, X., Tang, Y., 2014. In silico prediction of chemical acute oral toxicity using multi-classification methods. J. Chem. Inf. Model. 54, 1061–1069. https://doi.org/10.1021/ci5000467.

Li, X., Zhang, Y., Chen, H., Li, H., Zhao, Y., 2017. Insights into the molecular basis of the acute contact toxicity of diverse organic chemicals in the honey bee. J. Chem. Inf. Model. 57, 2948–2957. https://doi.org/10.1021/acs.jcim.7b00476.

Montaruli, Alberga, D., Ciriaco, F., Trisciuzzi, D., Tondo, Mangiatordi, G., Nicolotti, O., 2019. Accelerating drug discovery by early protein drug target prediction based on a multi-fingerprint similarity search. Molecules 24, 2233. https://doi.org/10.3390/molecules24122233.

Netzeva, T., Worth, A., Aldenberg, T., Benigni, R., Cronin, M., Gramatica, P., Jaworska, J., Kahn, S., Klopman, G., Marchant, C., Myatt, G., Jeliazkova, N., Patlewicz, G., Perkins, R., Roberts, D., Schultz, T., Stanton, D., Sandt, J., Tong, W., Yang, C., 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships-the report and recommendations of ECVAM workshop 52. Altern. Lab. Anim. 33, 155–173. https://doi.org/10.1177/026119290503300209.

OECD, 1998. Test No. 214: Honeybees, Acute Contact Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris. https://doi.org/10.1787/9789264070189-en.

OECD, 2007. Guidance Document on the Validation of (Quantitative) Structure Activity Relationship [(Q)SAR] Models, ENV/JM/MONO(2007)2. OECD Publishing, Paris.

Parmentier, L., Meeus, I., Cheroutre, L., Mommaerts, V., Louwye, S., Smagghe, G., 2014. Commercial bumblebee hives to assess an anthropogenic environment for pollinator support: a case study in the region of Ghent (Belgium). Environ. Monit. Assess. 186, 2357–2367. https://doi.org/10.1007/s10661-013-3543-2.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106. https://doi.org/10.1007/bf00116251.

Roy, K., Kar, S., Ambure, P., 2015. On a simple approach for determining applicability domain of QSAR models. Chemom. Intell. Lab. Syst. 145, 22–29. https://doi.org/10.1016/j.chemolab.2015.04.013.

Shen, J., Cheng, F., Xu, Y., Li, W., Tang, Y., 2010. Estimation of ADME properties with substructure pattern recognition. J. Chem. Inf. Model. 50, 1034–1041. https://doi.org/10.1021/ci100104j.

Singh, K.P., Gupta, S., Basant, N., Mohan, D., 2014. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. Chem. Res. Toxicol. 27, 1504–1515. https://doi.org/10.1021/tx500100m.

Spivak, M., Mader, E., Vaughan, M., Euliss, N.H., 2011. The plight of the bees. Environ. Sci. Technol. 45, 34–38. https://doi.org/10.1021/es101468w.

Sun, L., Yang, H., Li, J., Wang, T., Li, W., Liu, G., Tang, Y., 2018. In silico prediction of compounds binding to human plasma proteins by QSAR models. ChemMedChem 13, 572–581. https://doi.org/10.1002/cmdc.201700582.

Taylor, K., Stengel, W., Casalegno, C., Andrew, D., 2014. Experiences of the REACH testing proposals system to reduce animal testing. ALTEX 31, 107–128. https://doi.org/10.14573/altex.1311151.

Toropov, A.A., Toropova, A.P., Martyanov, S.E., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2011. Comparison of SMILES and molecular graphs as the representation of the molecular structure for QSAR analysis for mutagenic potential of polyaromatic amines. Chemom. Intell. Lab. Syst. 109, 94–100. https://doi.org/10.1016/j.chemolab.2011.07.008.

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. Environ. Toxicol. Pharmacol. 53, 158–163. https://doi.org/10.1016/j.etap.2017.05.011.

US. EPA (Environmental Protection Agency), 2014. Guidance for Assessing Pesticide Risks to Bees. Office of Research and Development, National Center for Environmental Assessment, US. EPA, USA. https://www.epa.gov/sites/production/files/2014-06/documents/pollinator_risk_assessment_guidance_06_19_14.pdf.

US. EPA (Environmental Protection Agency), 2016. Guidance on exposure and effects testing for assessing risks to bees. US. EPA, USA. https://www.epa.gov/sites/production/files/2016-07/documents/guidance-exposure-effects-testing-assessing-risks-bees.pdf.

Vighi, M., Garlanda, M.M., Calamari, D., 1991. QSARs for toxicity of organophosphorous pesticides to Daphnia and honeybees. Sci. Total Environ. 109-110, 605–622. https://doi.org/10.1016/0048-9697(91)90213-x.

Vogt, M., Bajorath, J., 2013. Similarity searching for potent compounds using feature selection. J. Chem. Inf. Model. 53, 1613–1619. https://doi.org/10.1021/ci4003206.

Wang, F., Yang, J., Wang, M., Jia, C., Shi, X., Hao, G., Yang, G., 2020. Graph attention convolutional neural network model for chemical poisoning of honey bees' prediction. Sci. Bull. 65, 1184–1191. https://doi.org/10.1016/j.scib.2020.0 4.006.

Watson, P., 2008. Naïve Bayes classification using 2D pharmacophore feature triplet vectors. J. Chem. Inf. Model. 48, 166–178. https://doi.org/10.1021/ci7003 253.

Yang, H., Li, X., Cai, Y., Wang, Q., Li, W., Liu, G., Tang, Y., 2017. In silico prediction of chemical subcellular localization via multi-classification methods. Med. Chem. Commun. 8, 1225–1234. https://doi.org/10.1039/c7md00074j.

Yap, C.W., 2011. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 32, 1466–1474. https://doi.org/10.1002/jcc.21707.