Artificial Intelligence in the Life Sciences 1 (2021) 100013

Contents lists available at ScienceDirect

Artificial Intelligence in the Life Sciences

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

Research article

BeeToxAI: An artificial intelligence-based web app to assess acute toxicity of chemicals to honey bees

José T. Moreira-Filho[a,1], Rodolpho C. Braga[b,1], Jade Milhomem Lemos[a], Vinicius M. Alves[c], Joyce V.V.B. Borba[a], Wesley S. Costa[d], Nicole Kleinstreuer[e], Eugene N. Muratov[c,f], Carolina Horta Andrade[a], Bruno J. Neves[a,*]

[a] LabMol – Laboratory for Molecular Modeling and Drug Design, Faculty of Pharmacy, Universidade Federal de Goiás, Goiás 74605-170, Brazil
[b] InsilicAll Inc., São Paulo 04363-090, Brazil
[c] Laboratory for Molecular Modeling, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, NC 27599, United States of America
[d] Centro Universitário de Anápolis, UniEVANGÉLICA, Goiás 75083-515, Brazil
[e] National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, NIEHS, Durham, NC 27560, United States of America
[f] Department of Pharmaceutical Sciences, Federal University of Paraiba, Paraíba 58059-900, Brazil

ARTICLE INFO

Keywords:
Apis mellifera
Artificial intelligence
Pollinators
Ecotoxicology
Machine learning
Predictive modeling

ABSTRACT

Chemically induced toxicity is the leading cause of recent extinction of honey bees. In this regard, we developed an innovative artificial intelligence-based web app (BeeToxAI) for assessing the acute toxicity of chemicals to *Apis mellifera*. Initially, we developed and externally validated QSAR models for classification (external set accuracy ~91%) through the combination of Random Forest and molecular fingerprints to predict the potential for chemicals to cause acute contact toxicity and acute oral toxicity to honey bees. Then, we developed and externally validated regression QSAR models ($R^2 = 0.75$) using Feedforward Neural Networks (FNNs). Afterward, the best models were implemented in the publicly available BeeToxAI web app (http://beetoxai.labmol.com.br/). The outputs of BeeToxAI are: toxicity predictions with estimated confidence, applicability domain estimation, and color-coded maps of relative structure fragment contributions to toxicity. As an additional assessment of BeeToxAI performance, we collected an external set of pesticides with known bee toxicity that were not included in our modeling dataset. BeeToxAI classification models were able to predict four out of five pesticides correctly. The acute contact toxicity model correctly predicted all of the eight pesticides. Here we demonstrate that BeeToxAI can be used as a rapid new approach methodology for predicting acute toxicity of chemicals in honey bees.

1. Introduction

Pesticides play an essential role in protecting plants and reducing large-scale agricultural crop losses from insects and pathogens [1,2]. However, several harmful effects of pesticides on aquatic and terrestrial ecosystems have been described in the literature, particularly towards non-target species such as fishes, earthworms, birds, and bees [3,4]. In the past few years, there has been increased concern regarding the impact of pesticides on bees [5–9]. During foraging or nectar, pollen, and water collection, bees inadvertently may be contaminated with a wide array of pesticides [10]. Very often, contaminated bees carry these hazardous chemicals back to the hive, potentially inducing sub-lethal or lethal effects for the entire colony [10,11].

The drastic extinction of bees is a serious threat to global food security and planetary ecosystem stability [12,13]. For this reason, scientific advisory bodies and government agencies have employed standardized protocols to test the acute toxicity of active pesticide ingredients against adult honey bees (*Apis mellifera*) [14–17]. The United States Environmental Protection Agency (EPA) Pollinator Risk Assessment Guidance

Abbreviations: ACC, accuracy; AD, applicability domain; AUC, area under receiver operating characteristic curve, D_s, Dice similarity; D_T, applicability domain threshold; κ, Cohen's kappa; LD_{50}, median lethal dose that induces death in 50% of the population; MACCS, Molecular ACCess System; MCC, Matthews correlation coefficient; ML, machine learning; NPV, negative predictive value; OCHEM, Online Chemical Modeling Environment; OECD, Organization for Economic Cooperation and Development PPV, positive predictive value; QSAR, Quantitative StructureToxicity/Activity Relationship; RF, Random Forest; SE, sensitivity; SMILES, Simplified Molecular Input Line Entry Specification SP, specificity; SVM, Support Vector Machines; Tc, Tanimoto coefficient; US EPA, United States Environmental Protection Agency's; 5FCV, 5-fold cross-validation.

* Corresponding author.
E-mail address: brunoneves@ufg.br (B.J. Neves).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ailsci.2021.100013
Received 27 October 2021; Received in revised form 9 November 2021; Accepted 11 November 2021
Available online 14 November 2021

provides both a short and long-term strategy to assess the risks posed against honey bees by pesticides [16]. These animals can be exposed to pesticide residues by indirect contact on plant surfaces, via oral intake with contaminated water or food, or by direct contact during their application in standard farming practice. If the proposed use pattern of a pesticide indicates a possible exposure of honey bees, acute contact and oral toxicity studies are necessary for pesticide registration [16].

The acute toxicity tests examine the effects of a pesticide after a short-term exposure (24–96 h) using the median lethal dose that induces death in 50% of the population (LD_{50}). The predominant exposure routes are through contact (i.e., direct spray) or oral incubation (i.e., consumption of nectar and pollen). Both the contact and oral toxicity tests report toxicity values as μg/bee. If a pesticide's active ingredient is classified as highly or moderately toxic (LD_{50} <11 μg/bee), it would require complementary studies to determine acute oral toxicity and foliage test using the final product [18]. The US EPA estimates that studies required for pesticide registration cost around $ 13,400 for honey bee acute toxicity, with a semi-field study on honey bees reaching $133,300 per chemical [19].

Understanding the costly and time-consuming characteristics of experimental assays, *in silico* models have emerged as a practical solution to be employed as screening tools and within integrated testing strategies to avoid animal testing as well as reduce cost and chemical waste [20]. Computational strategies comply with the "3R's" principle of replacement, refinement, and reduction of animal testing [21]. The more simplistic approaches are based on the concept that compounds sharing structural similarity (read-across)[22] or certain substructures (structural alerts) [23] have increased probability to share the same toxicological properties [22]. However, there has been a growing concern that structural alerts disproportionally flag too many chemicals as toxic, which questions their reliability as toxicity markers [23]. Therefore, their use in regulatory toxicology has been hampered by the lack of transparency and interpretability [23].

Consequently, Quantitative Structure-Activity/Toxicity Relationship (QSAR/QSTR) models have been developed as alternative methods to experimental tests and rule-based methods. QSAR modeling reveals relationships between structural properties of chemical compounds and corresponding biological/toxicological properties [22–24]. These characteristics confer more accurate predictions of the toxicity of untested chemicals from their chemical structures [23]. Currently, the most modern QSAR methodologies are developed using artificial intelligence methods, such as machine learning (ML) and deep learning (DL) algorithms [25,26]. ML is a growing field of artificial intelligence that uses different statistical techniques to enable computers to learn from chemical and biological or toxicological data without being explicitly programmed for this task [27]. These algorithms are capable of capturing the complex nonlinear relationships between the relevant descriptors (i.e., mathematical representations of molecules' properties) and the observed properties/toxicities [26,28].

Although QSAR is widespread in the ecotoxicology field, there are few freely available QSAR-based tools with a graphical interface to assess the oral and acute contact toxicity of chemicals to honey bees [29,30]. Similarly, no QSAR model has been reported to assess acute oral toxicity of chemicals to honey bees. In addition, a critical analysis reveals that the vast majority of the published models do not comply with the OECD principles [31], as well as best practices for data curation[32–34] and QSAR modeling [35]. The main drawbacks of previous QSAR studies includes: (i) lack of evidence on data curation and duplicate analysis [29,36–38]; (ii) use of an undefined or confusing endpoint [29]; (iii) lack of AD estimation [29,37,39]; and (iv) lack of mechanistic interpretation [29,36–39]. For example, although the models developed by Wang et al. [29]. and implemented in the BeeTox web app have satisfactory predictive performance, estimates of the applicability domain (DA) and mechanistic interpretation of the predictions were not provided. In addition, we found more than 60 duplicates in its dataset, which can lead to overestimation of the model. Thus, their reliability for assessing chemically-induced acute toxicity for honey bees is not assured.

Hence, this manuscript describes the development and application of an easily accessible, open-source, public-facing web application (BeeToxAI: http://beetoxai.labmol.com.br/) to democratize access to these predictive QSAR models for a wide range of stakeholders, including regulators, regulated industry, research scientists, and the public. BeeToxAI is the first web app for the prediction of acute contact and oral toxicity in honey bees fully compliant with the stringent predictive modeling practices [35] and OECD guidelines [31].

2. Material and methods

2.1. Datasets

A dataset of compounds containing experimental acute toxicity data for honey bees (*A. mellifera*) was collected from the scientific literature [40–52], as well as the US EPA's Ecotox database [53], EFSA's OpenFoodTox database [54], and Online Chemical Modeling Environment (OCHEM) database [55]. Data integration resulted in an uncured dataset consisting of 2543 pesticide and pesticide-like compounds representing different classes (e.g., insecticides, herbicides, fungicides) and with a wide spectrum of toxicity mechanisms. Then, compounds with a median lethal dose ($LD_{50,}$ μg/bee) for adult honey bee recorded after 48 h were categorized into toxic and nontoxic using a threshold of 11 μg/bee as described in the US EPA test guidelines [16]. Compounds were divided into two independent datasets according to honey bee exposure type (contact and oral). A brief description of the datasets is presented below:

- Contact exposure dataset (File S1): 615 compounds with LD_{50} data for contact exposure of honey bees. It consisted of 229 toxic compounds with LD_{50} ≤11 μg/bee and 386 nontoxic compounds (LD_{50} >11 μg/bee).
- Oral exposure dataset (File S2): 211 compounds with LD_{50} data for contact exposure of honey bees. It consisted of 93 toxic compounds with LD_{50} ≤11 μg/bee and 118 nontoxic compounds (LD_{50} >11 μg/bee).

2.2. Data curation

All chemical structures and correspondent LD_{50} data were carefully standardized using Standardizer v.16.9.5.0 (ChemAxon, Budapest, Hungary) according to the protocols proposed by Fourches and colleagues [32–34]. Briefly, explicit hydrogens were added, whereas salts, mixtures, polymers, and organometallic compounds were removed. In addition, specific chemotypes such as aromatic rings and nitro groups were normalized. Then, we performed the analysis and exclusion of duplicates. Distinct criteria were employed, as follows:

- *Classificatory QSAR models:* (*i*) if duplicates presented discordance in toxicological outcomes (e.g., toxic vs nontoxic), both entries would be excluded; and (*ii*) if the reported outcomes of the duplicates were the same, one entry would be retained in the dataset and the other excluded. After duplicate removal, the contact exposure dataset had 382 compounds (toxic: 112, nontoxic: 269), while the oral exposure dataset had 169 compounds (toxic: 71, nontoxic: 98).
- *Regression QSAR models*: (*i*) duplicates were inspected visually, (*ii*) if duplicates presented discordant potencies, both entries would be excluded; and (*iii*) if the reported potencies were similar, an average of the values was calculated, and one entry was retained in the dataset. Subsequently, the LD_{50} values were converted to negative logarithmic (−log) units (pLD_{50}) at μM range. At the end of this process, the contact exposure dataset had 218 compounds, while the oral exposure dataset had 142 compounds.

2.3. Chemical space analysis

The chemical space formed by toxic and nontoxic compounds for *A. mellifera* and 2044 commercial pesticides (untested against *A. mellifera*) collected from Pesticide Product Information System Database [56] was analyzed by plotting a similarity map and was generated using OSIRIS DataWarrior software v.05.02.01 [57]. The similarity map uses a Rubberbanding Forcefield approach, which translates similarity (vertices) between compounds (nodes). The approach involves the following steps: (*i*) randomly positioning all compounds in 2D space; (*ii*) calculating the similarity matrix between all compounds using Tanimoto coefficients (Tc) and FragFP descriptors; (*iii*) location of most similar neighbors (Tc >0.8) to be considered for every compound; and (*iv*) stepwise relocation of all compounds to ensure similar molecules were located close to each other [57].

2.4. Classification models

The models were developed in Python v.3.6[58] following best practices for QSAR modeling [35], and are fully compliant with OECD principles for validation of QSAR modeling for regulatory purposes, i.e., a defined endpoint, an unambiguous algorithm, a defined domain of applicability, appropriate measures of goodness-of-fit, robustness and predictivity, and a mechanistic interpretation if possible [31].

2.4.1. Molecular fingerprints

Molecular ACCess System (MACCS) keys, Morgan (ECFP-like), and FeatMorgan (FCFP-like) fingerprints were calculated in the open-source cheminformatics software RDKit [59] with a diameter of 4–8 and bit vector of 2048 bits [60]. ECFPs are circular fingerprints that captures highly specific atomic information enabling the representation of a large set of precisely defined structural features [61], whereas FCFP fingerprints captures functional features (i.e., hydrogen-bond donor and acceptors, aromatic, halogen, and basic and acid groups) [62].

2.4.2. Dataset splitting and 5-fold cross-validation

The general flowchart for splitting contact and oral datasets of compounds is shown in Supplementary Fig. S1. Initially, datasets were split into modeling sets (80% of compounds) and external sets (20% of compounds) using the random distribution approach. The modeling sets were used to generate QSAR models through a 5-fold external cross-validation (5FCV) approach, while the external sets were used to assess the predictive power of models. The structural diversity of the modeling and external set compounds was evaluated using the similarity maps. These plots show that external set compounds occupy the same chemical space as the modeling set compounds. The modeling sets were subjected to 5FCV approach where five subsets of equal size were generated. Together, four subsets (80% of the modeling set) were used to build the QSAR models (training set) and the remaining subset of 20% (test set) was employed to evaluate the robustness of the QSAR model. QSAR models were developed five times, allowing each of the five subsets to be used as a momentary test set. The 5FCV modeling procedure was followed by an evaluation of the model performance using the external holdout sets.

2.4.3. QSAR modeling and hyperparameter optimization

QSAR models were developed using the Support Vector Machine (SVM) [63] and Random Forest (RF) [64] algorithms implemented in Scikit-learn v.0.24.2 [65]. Since the performance of ML is closely related to its hyperparameters, the models were optimized using a Bayesian approach implemented in Scikit-Optimize v.0.7.4 [66]. Details of hyperparameters explored in this work are available in the Supporting Information. The Bayesian optimization may be defined as follows:

$$\mathrm{P}\left(f\mathrm{D}_{1:t}\right) \propto \mathrm{P}\left(\mathrm{D}_{1:t}f\right)\mathrm{P}(f) \tag{1}$$

where, x_i is the *i*th sample, and $f(\mathrm{x}_i)$ is the observation of the objective function at x_i. The observations $\mathrm{D}_{1:t} = \{\mathrm{x}_{1:t}, f(\mathrm{x}_{1:t})\}$ are accumulated. The prior distribution is combined with the likelihood function $\mathrm{P}(\mathrm{D}_{1:t}|f)$ of overserving $\mathrm{D}_{1:t}$ given model f multiplied by the prior probability of $\mathrm{P}(f)$. In doing so, Bayesian optimization finds hyperparameters that maximize the objective function (G-mean score) by building a surrogate function (probabilistic model) based on past evaluation hyperparameters of the objective [66,67]. The geometric (G)-mean was selected as the scorer since it measures the balance between classification performances on both the majority (nontoxic) and minority (toxic) classes.

2.4.4. Threshold-moving

The QSAR models were calibrated using a threshold-moving approach implemented in Scikit-learn v.0.24.2 [65]. This approach uses different probability thresholds in the range of 0 to 1 obtained via the Receiver Operating Characteristic (ROC) curve to find the threshold with the largest G-mean value. Thus, it is easier to predict the minority class examples accurately. To use the G-mean value as the model boundary class, we implemented a Python class to overlay toxicity prediction and its probability in the Scikit-learn framework. Subsequently, probability values were scaled to estimate the confidence of predictions as follows:

$$x_{normalized} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2}$$

where $\min(x)$ denotes the minimum of the range of your measurement, $\max(x)$ denotes the maximum of the range of your measurement, and $x \in [\min(x), \max(x)]$ denotes your measurement to be scaled. Probabilities x are on the interval [0,1], with $x = \min(x)$ mapped to 0 and $x = \max(x)$ mapped to 1.

2.4.5. Assessment of model performance

The internal and external predictive performances of QSAR models were evaluated using accuracy (ACC), sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC). These metrics were calculated as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{N}} \tag{3}$$

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{5}$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{6}$$

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}} \tag{7}$$

$$\mathrm{MCC} = \frac{\mathrm{TP\ x\ TN} - \mathrm{FP\ x\ FN}}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{8}$$

$$\mathrm{AUC} = \sum_{\mathrm{i}} \left[\left(\mathrm{SE}_{i+1}\right)\left(\mathrm{SP}_{i+1} - \mathrm{SP}_{\mathrm{i}}\right)\right] \tag{9}$$

where N represents the number of compounds, TP and TN represent the number of true positives and true negatives, and FP and FN represent the number of false positives and false negatives, respectively.

In addition to the above model evaluation metrics, Cohen's kappa (κ) was used to measure the agreement between experimental data and model predictions [68]. This statistical parameter is calculated by the following equations:

$$\mathrm{Pr}\,(a) = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{N}} \tag{10}$$

$$\mathrm{Pr}\,(\mathrm{e}) = \frac{(\mathrm{TP} + \mathrm{FP})\ \mathrm{x}\ (\mathrm{TP} + \mathrm{FN}) + (\mathrm{TN} + \mathrm{FN})\ \mathrm{x}\ (\mathrm{TN} + \mathrm{FP})}{\mathrm{N}} \tag{11}$$

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (12)$$

where Pr(e) is the hypothetical probability of chance agreement, and Pr(a) represents the relative observed agreement between the predicted classification of the model and the known classification.

2.4.6. Applicability domain (AD)

Any QSAR model needs to demonstrate not only good accuracy but also characterize the reliability of external predictions. To address the latter, the AD of models was estimated using Dice similarity (D_S) between a compound under prediction (A) and the training set compounds (B) [69]. The Dc is calculated by the following equation:

$$D_S = \frac{2\,|A \cap B|}{|A| + |B|} \quad (13)$$

where a set with vertical bars on either side refers to the cardinality of the set, i.e., the number of bits (fingerprints) in that set. The ∩ is used to represent the intersection of two sets (bits that are common to both sets). Then, an AD threshold (D_T) was defined to estimate the reliability of external predictions:

$$D_T = \bar{y} + Z\sigma \quad (14)$$

where $\bar{y}$ is the average D_S of the compounds under prediction and training set compounds, σ is the standard deviation of D_S, and Z is an arbitrary parameter to control the significance level. We set the default value of this parameter Z at 0.5. If the compound distance exceeds the D_T, the prediction may be considered less trustworthy [70].

2.5. Regression models

Then, regression models based on Feedforward Neural Networks (FNNs) were developed using Keras (https://keras.io/), and Tensorflow (www.tensorflow.org) as backend. Initially, the datasets were split into modeling sets (80% of compounds) and test sets (20% of compounds) using the random distribution approach. Then, regression models were developed using ECFP4 fingerprints and pLD_{50} values at μM range. The architecture of the FNNs was optimized according to the following combinations: layer type (dense), number of hidden layers (3–7), activation functions (ReLU, Elu, Selu), output layer function (sigmoid), model optimizer (Adam). The "mean squared error" was used as a loss function. The "mean absolute error" was used as a parameter to judge the performance of the models. The following hyperparameters were used for further FNN training: the number of epochs (1–200), dropout (0.001), and batch size (5–30). Baseline comparison of models was performed using Support Vector Regression (SVM) [63] and RF [64] algorithms implemented in Scikit-learn v.0.24.2 [65]. The predictive performance of regression models was evaluated using correlation coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE) [71]. These metrics were calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{test}} (Y_{obs} - Y_{pred})^2}{\sum_{i=1}^{n_{test}} (Y_{obs} - \bar{Y}_{train})^2} \quad (15)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_{test}} (Y_{obs} - Y_{pred})^2}{n_{test}}} \quad (16)$$

$$MAE = \frac{\sum_{i=1}^{n_{test}} \left|Y_{obs} - Y_{pred}\right|}{n_{test}} \quad (17)$$

In the above equations, Y_{obs} represents experimental pLD_{50} value, Y_{pred} represents the predicted pLD_{50} value, n_{train} and n_{test} are the number of compounds in training and test set, respectively, and $\bar{Y}_{train}$ is the average of experimental values of the training set.

2.6. Model interpretation

Contribution maps[72,73] were generated from QSAR models to visualize the fragments and atoms contributing to acute contact and oral toxicities. Here, the "weight" of an atom was considered as a predicted-probability difference (classification models) or pLD_{50} difference (regression models) obtained when the bits in the fingerprint corresponding to the atom are removed. Then, the normalized weights were used to color the atoms in a topography-like map in which green indicating negative contribution for toxicity (i.e., predicted probability or pLD_{50} increases when the bits are removed), and red indicating a positive contribution for toxicity (i.e., predicted probability or pLD_{50} decreases when the bits are removed) [73].

2.7. Model implementation

The BeeToxAI web app was implemented on a multiplatform framework with technology ready to support large-scale demands on microservice serverless and Kubernetes environments [74]. The backend serving makes deploying new algorithms and experiments easy while keeping the same server architecture and the APIs. The APIs work with JSON infrastructure and permit out-of-the-box integration with other computer software (e.g., KNIME or a custom front-end). The main libraries integrated are Python [58], RDKit [59], Scikit-learn [65], uWSGI [75], JavaScript [76], Flask [77], Matplolib [78], and Seaborn [79]. BeeToxAI also includes the JSME molecule editor written in JavaScript [80], which is supported by the most popular web browsers. Java or Flash plugins are not required to use the app. Also, our back and front-end use GitLab [81], CI/CD for Continuous integration (CI), Continuous Delivery (CDE), and Continuous Deployment (CD). The application's code base is hosted in a Git repository[82] and, to every push, runs a pipeline of scripts to build, test, validate, and deploying your application to production.

3. Results and discussion

In the present study, we integrated and carefully compiled the largest collection of compounds with acute toxicity data (LD_{50}) for adult *A. mellifera*. The compounds were divided into "contact" (File S1) and "oral" (File S2) datasets according to the type of bee exposure during experimental acute toxicity assays. A threshold of 11 μg/bee was used to categorize compounds into toxic (≤11 μg/bee) and nontoxic (>11 μg/bee) as described in the US EPA test guidelines [16].

Subsequently, we carefully curated contact and oral datasets using standard protocols [32–34]. Data curating represents a crucial step for building predictive QSAR models. It has been recognized that genotype differences among the 26 recognized subspecies of *A. mellifera* can directly impact the response to chemicals [83]. Unfortunately, subspecies information is not provided for most chemicals deposited in public domain databases. In addition, considerable differences often appear when toxicity tests are performed by different laboratories [84] and when different colonies of a single subspecies are tested in the same laboratory [85]. On the other hand, the same compound may be registered multiple times in the modeling and external sets. QSAR models built with datasets containing duplicates will have low accuracy if toxicity outcomes are dissimilar or overoptimistic performances if outcomes are identical [34]. Despite this, data curation procedures have not been uniformly applied to the development of some QSAR models [29,36–38].

The number of compounds in contact and oral datasets is shown in Table 1. After data curation and duplicate removal, the contact exposure dataset had 382 compounds (toxic: 113, nontoxic: 269) while the oral exposure dataset had 169 compounds (toxic: 71, nontoxic: 98). The datasets were further divided (Fig. S1) into modeling (80%) and external (20%) sets. So, 305 compounds of the contact dataset were used for the model development while the remaining 77 compounds (20%) were used to validate the models. Similarly, 135 compounds of the oral

Table 1
Distribution of chemicals in the modeling and external validation set of contact and oral exposure datasets.

Datasets	Classification models		Regression models
	Toxic	Nontoxic	
Contact dataset			
Modeling set	90	215	174
External set	23	54	44
Total	113	269	218
Oral dataset			
Modeling set	57	78	114
External set	14	20	28
Total	71	98	142

dataset were used for model development while the remaining 34 compounds were used to validate the models. The overall quality of dataset splitting is shown in Supplementary Fig. S2, which indicated that external set compounds are relatively distributed in all regions of the chemical space of modeling set compounds.

On the other hand, for the datasets to develop regression models, after data curation and duplicate removal, the contact exposure dataset had 218 compounds while the oral exposure dataset had 142 compounds. After splitting the datasets into modeling (80%) and external (20%) sets, 174 compounds of the contact dataset were used for the model development and 44 compounds were used to validate the models. In the same manner, 114 compounds of the oral dataset were used for model development and 28 compounds were used to validate the models.

3.1. Chemical space analysis

The analysis of chemical space was performed by using contact and oral exposure datasets and 2044 pesticides collected from the Pesticide Product Information System Database [56]. The analysis has been performed by plotting both datasets separately against the pesticide database using similarity maps (Fig. 1) [57].

As shown in Fig. 1, both datasets are structurally diverse, containing smaller clusters of similar compounds (black circles), and covers all regions of the chemical space of pesticides. This finding correlates well with our previous observation that a broad range of chemical categories (drugs, industrial use, pesticides, cosmetics) have similar structures and that a property of a compound depends on its chemical structure and not on its industrial class [86].

When analyzing the outcomes of the contact dataset (Fig. 1a), it was found that most of the toxic and nontoxic compounds do not share the same clusters, and similar characteristics were observed in the oral dataset (Fig. 1d). This analysis shows that both datasets have few toxicity cliffs (i.e., structurally similar compounds with a large difference in toxicity) [87–89]. Representative compounds of clusters 1–3 from the contact dataset are shown in Fig. 1b. Cluster 1 contains toxic pyrethroids; cluster 2 – toxic organothiophosphates; cluster 3 – nontoxic sulfonylureas. The representative compounds of clusters 1–3 from the oral dataset are shown in Fig. 1d. Cluster 1 contains toxic phosphonothioates; cluster 2 – toxic pyrethroids; cluster 3 – nontoxic azoles.

3.2. Performance of classification models

A total of 20 classification models were developed by a combination of two ML methods (RF and SVM) along with three fingerprint sets: MACCS, FCFP and ECFP (diameter 4: FCFP4, ECFP4; diameter 8: FCFP8, ECFP8). The statistical characteristics of the QSAR models developed for acute contact toxicity and acute oral toxicity are summarized in Tables S1 and S2. Briefly, ACC values ranged between 0.81–0.89; SE = 0.64–0.82; SP = 0.81–0.98, κ = 0.58–0.74; and MCC = 0.59–0.75. The model built using ECFP4 + RF demonstrated the best internal performance among all other models developed for acute contact toxicity (ACC = 0.89; SE = 0.70; SP = 0.96; and κ = 0.71). The model built using ECFP4 + SVM demonstrated the best internal performance among models developed for acute oral toxicity (ACC = 0.87; SE = 0.75; SP = 0.96; and κ = 0.74).

Furthermore, the external sets were used to evaluate the predictivity of the QSAR models. Because external set compounds were not involved in model building, the resulting performance reflects the ability of the models to predict the toxicity of new compounds. The results indicate that models built using FCFP4 + RF, ECFP4 + RF and ECFP8 + RF (Table S1) showed the best external predictivity among all other models developed for acute contact toxicity (ACC = 0.90; SP = 0.98; κ = 0.73). However, these models have a limited ability to correctly predict toxic compounds (SE = 0.70). The model developed using MACCS + RF (Table S2) showed the best external predictivity among all other models developed for acute oral toxicity (ACC = 0.88; SE = 0.86; SP = 0.90; and κ = 0.76), suggesting that this model has a higher accuracy to sort the toxicological potential of new compounds.

3.3. Threshold-moving calibration for imbalanced classification

To increase prediction confidence without losing data, i.e., without needing to balance the data, we tried threshold-moving calibration of probability estimates [90]. Mechanistically, classification models also output a continuous value as the probability of a given case belonging to a output class. The probabilities can be interpreted as the likelihood or confidence of a given case belonging to each class. Here, classification models were trained independently using acute contact and oral datasets to distinguish toxic vs. nontoxic compounds. Usually, predicted probabilities values less than 0.5 are assigned to the class of nontoxic compounds and values greater than or equal to 0.5 are assigned to the class of toxic compounds. However, QSAR models developed for classification using imbalanced data usually provide poor probability estimates (<0.5) for the minority class [91,92]. In view of this, different probability thresholds in a range between 0 and 1 were explored to find the optimal threshold that reflects the best performance. Details of statistical performances of calibrated models for acute contact toxicity and acute oral toxicity are summarized in Tables S3 and S4, respectively. In general, the threshold-moving led to significant improvements in statistical performance of these QSAR models (Fig. 2a). As shown in Table 2 and Fig. 2a, changing the threshold from 0.5 to 0.32 improved the ACC (+4%), MCC (+14%), and κ (+16%) of the FCFP4 + SVM model developed for acute contact toxicity. Based on this, the adjusted probability threshold obtained for this model was kept as the final model for the prediction of acute contact toxicity of new compounds.

The performance of the acute oral toxicity models was also investigated after threshold-moving calibration, although they were generated using a dataset with a similar ratio of toxic and nontoxic compounds (1:1.4). The statistical characteristics of calibrated models developed for acute oral toxicity are shown in Tables 2 and S4. According to the radar plot (Fig. 2b), the threshold-moving calibration did not lead to apparent improvements in the internal and external performances of these models. Consequently, threshold-moving calibration was not used as adjusting parameters for predicting acute oral toxicity of new compounds.

From the statistical point of view, our modeling approach enabled the development of externally predictive classification models. However, building QSAR models trained with small datasets of compounds must always be considered delicate, since it may suffer from various deficiencies like inconsistent classifications for chemicals outside AD. Furthermore, mechanistic interpretation of these models may be challenging given the sum of a plethora of toxicological mechanisms, each involving different biochemical pathways concurring to the final effect.

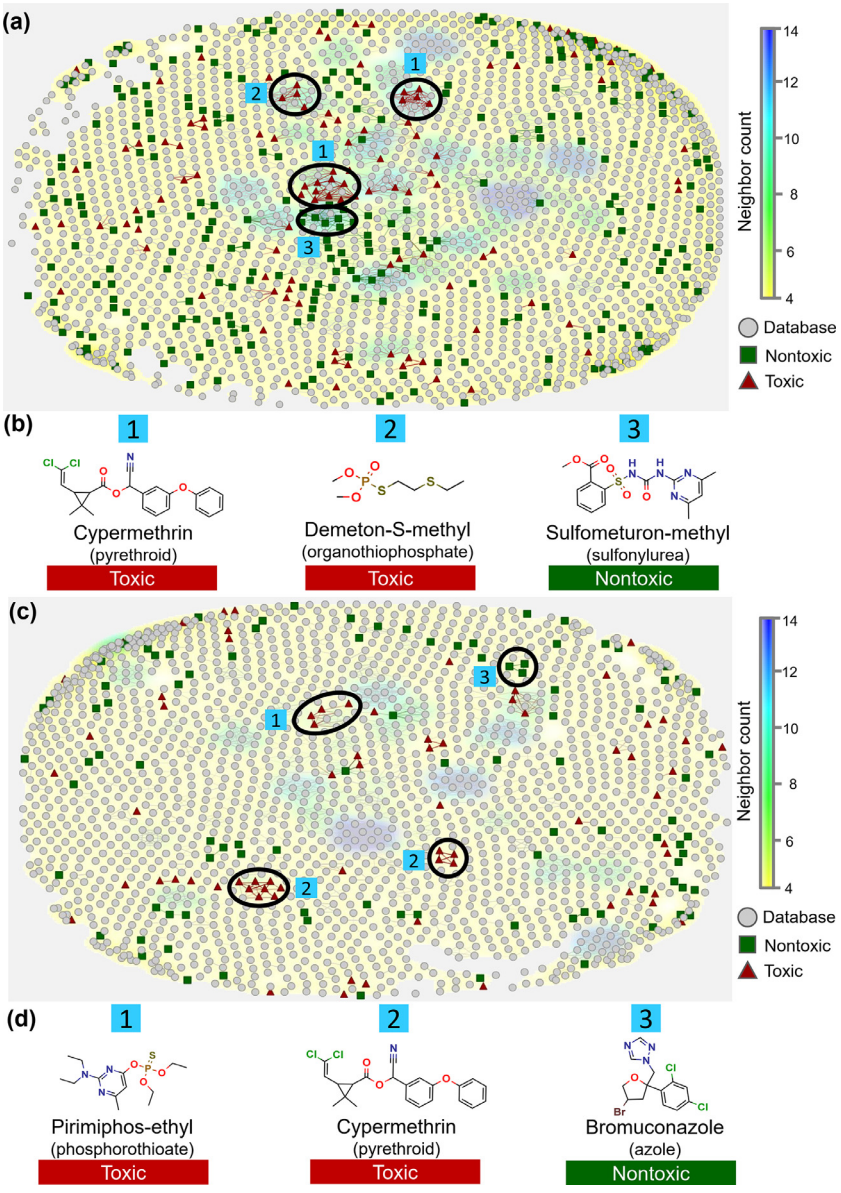

Fig. 1. Structural distribution of toxic and nontoxic compounds from contact (a) and oral (c) datasets along with the chemical space of pesticides. Three clusters of highly similar compounds are highlighted by black circles and numbered. (b) and (d) denotes representative compounds of clusters 1–3 highlighted in contact and oral datasets, respectively. Red triangles represent toxic compounds, green squares represent nontoxic compounds, and gray circles represent pesticides collected from Pesticide Product Information System Database. Compounds with a Tanimoto coefficient >0.8 are connected by vertices. The color scheme in the background represents the number of neighbors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Statistical characteristics of best QSAR models developed for acute contact toxicity and acute oral toxicity.

Fingerprint	Method	PT	Set	ACC	SE	SP	PPV	NPV	κ	MCC	AUC	Coverage
Acute contact toxicity models												
FCFP4	SVM	0.32[a]	Modeling	0.87	0.78	0.90	0.77	0.91	0.68	0.68	0.84	0.65
			External	0.91	0.78	0.96	0.90	0.91	0.77	0.78	0.87	0.70
ECFP8	RF	0.37[a]	Modeling	0.88	0.78	0.93	0.81	0.91	0.71	0.71	0.85	0.64
			External	0.90	0.74	0.96	0.89	0.90	0.74	0.75	0.85	0.67
ECFP4	RF	0.26[a]	Modeling	0.86	0.80	0.89	0.75	0.91	0.68	0.68	0.84	0.68
			External	0.90	0.83	0.93	0.83	0.93	0.75	0.75	0.88	0.74
Acute oral toxicity models												
MACCS	RF	0.50[b]	Modeling	0.82	0.74	0.88	0.82	0.82	0.63	0.63	0.81	0.75
			External	0.88	0.86	0.90	0.86	0.90	0.76	0.76	0.88	0.85
ECFP8	SVM	0.50[b]	Modeling	0.84	0.75	0.91	0.86	0.84	0.68	0.68	0.83	0.66
			External	0.85	0.86	0.85	0.80	0.89	0.70	0.70	0.85	0.79
FCFP4	RF	0.50[b]	Modeling	0.84	0.67	0.96	0.93	0.80	0.65	0.67	0.81	0.68
			External	0.85	0.64	1.00	1.00	0.80	0.68	0.72	0.82	0.79

RF, Random Forest; FCFP4, functional-class fingerprints with diameter 4; ECFP4, extended-connectivity fingerprints with diameter 4; ECFP8, extended-connectivity fingerprints with diameter 8; SVM, Support Vector Machine; PT, probability threshold; ACC, accuracy; SE, sensitivity; SP, specificity; PPV, positive predictive value; NPV, negative predictive value; κ, Cohen's kappa; MCC, Matthews correlation coefficient; AUC, area under ROC curve; Coverage, ratio of test set or external set compounds within the applicability domain. [a] Statistical results obtained after threshold-moving calibration. [b] Statistical results obtained from default probability threshold.

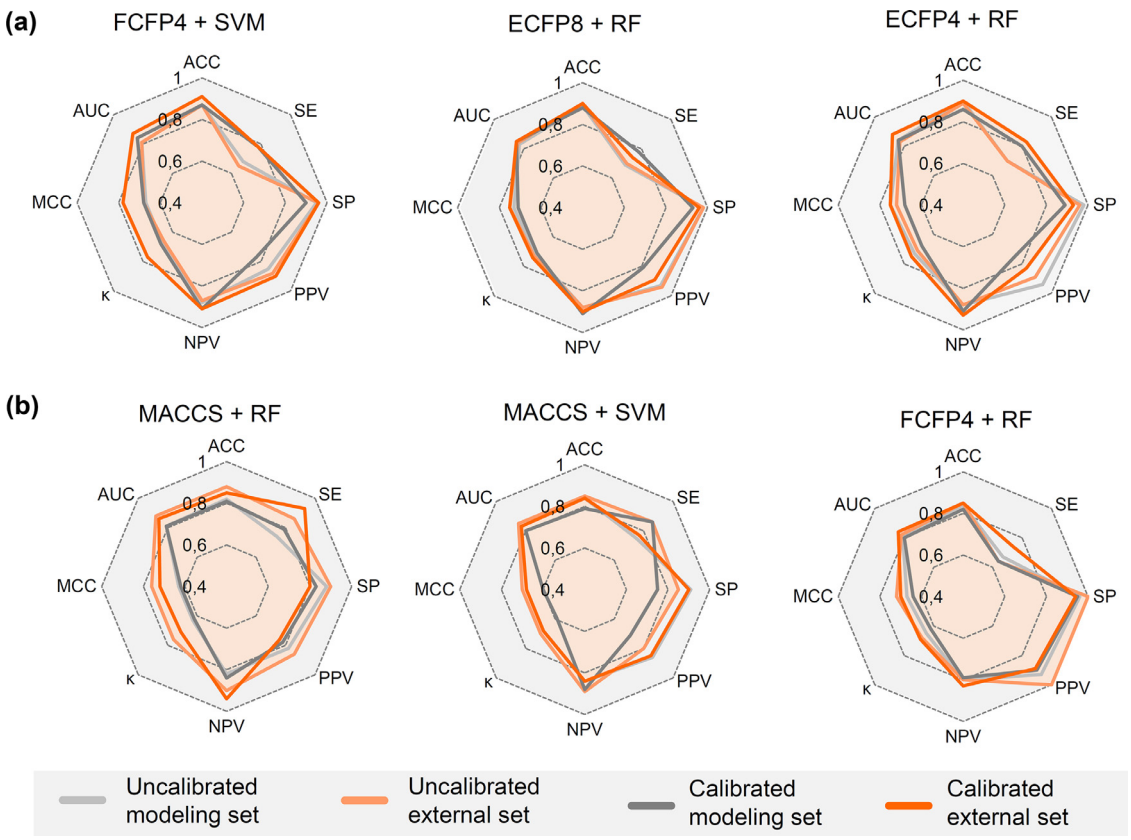

Fig. 2. Comparison of the best QSAR models developed for (a) acute contact toxicity and (b) acute oral toxicity across multiple metrics using standard probability values and threshold-moving calibration. ACC: accuracy; SE: sensitivity; SP: specificity; PPV: positive predictive value; and NPV: negative predictive value; κ: Cohen's kappa; MCC: Matthews correlation coefficient; AUC: area under ROC curve. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Performance of regression models

Following the results from the former section, we developed regression models using FNNs aiming to predict pLD_{50} values for acute contact and acute oral toxicities on honey bees. We evaluated the combination of different FNN configurations and hyperparameters for both acute contact and acute oral datasets. The best performance was determined based on the best balance between R^2 and RMSE. The tested hyperparameters and their respective influence on model performance are shown in Fig. 3. Initially, we assessed different number of hidden layers for each dataset, since depth has a large impact on the model complexity and can lead to overfitting of models. As we can see in Fig. 3a,b, five hidden layers showed the best results for the acute contact toxicity model, while six hidden layers improved the performance for acute oral toxicity model. In addition, ReLU activation function showed the best results among all tested functions (Fig. 3c,d). In addition, batch sizes of 20 and 15 showed the best performances for the acute contact and acute oral models, respectively (Fig. 3e,f).

After the finding of the best combinations, we found that most predictive acute contact model (R^2 = 0.75, RMSE = 0.39, and MAE = 0.32) was generated using five hidden layers, decreasing the number of neurons in the subsequent hidden layer [512, 256, 128, 16, and 4, respectively (Fig. 3g)], the ReLU activation function, batch size of 20, and 199 epochs. On the other hand, the best acute oral model (R^2 = 0.75, RMSE = 0.68, and MAE = 0.53) was generated using six hidden layers, decreasing the number of neurons in the subsequent hidden layer [512, 256, 128, 64, 32, and 16, respectively (Fig. 3h)], the ReLU activation function, batch size of 15, and 142 epochs. We also trained RF and SVR methods using ECFP4 fingerprints as baseline models to examine dataset modelability. SVR algorithm advantages in modeling nonlinear problems [63], whereas RF attracts much interest in QSAR studies because it is not sensitive to the hyperparameters [64]. For both acute toxicity endpoints, regression models based on optimized FNNs ($R^2 \geq 0.75$) showed superior performance over RF and SVR ($R^2 \leq 0.41$) models, indicating that modeled datasets do not have easily distinguishable patterns and are not biased.

3.5. Comparison with publicly available models

3.5.1. Classification models

Several classification methods have been developed to assess chemical-induced acute toxicity on honey bees, and thus a comparison of their statistical performances is shown in Table 3.

Overall, the classification models reported in this study (ACC = 0.91, MCC = 0.78) showed higher performance compared with those generated by Como et al. (ACC = 0.84, MCC = 0.67) [37], Venko et al. (ACC = 0.77, MCC = 0.48) [38], Wang et al. (ACC = 0.83, MCC = 0.59) [29], Li et al. (ACC = 0.90, MCC = 0.76) [39], and Singh et al. (ACC = 0.87 and 0.89 for classification and multi-class classification, respectively) [36]. On the other hand, our model showed statistical performance comparable with QSAR developed by Carnesecchi et al. [93]. (ACC = 0.90, MCC = 0.78). Nonetheless, it is noteworthy that these comparisons should not be interpreted rigorously, as different compositions and sizes of the training and test sets were used to build models. In addition, Carnesecchi et al. [30]. developed externally predictive QSAR models to assess the acute contact toxicity of mixtures to honey bees (R^2 = 0.89), and the nature of combined synergic/non-synergic toxicity (ACC = 0.96, MCC = 0.90). Although these models may contribute to fill part of the gaps in toxicological assessment in honey bees, it is not suitable to compare them with our active ingredient-based models, since

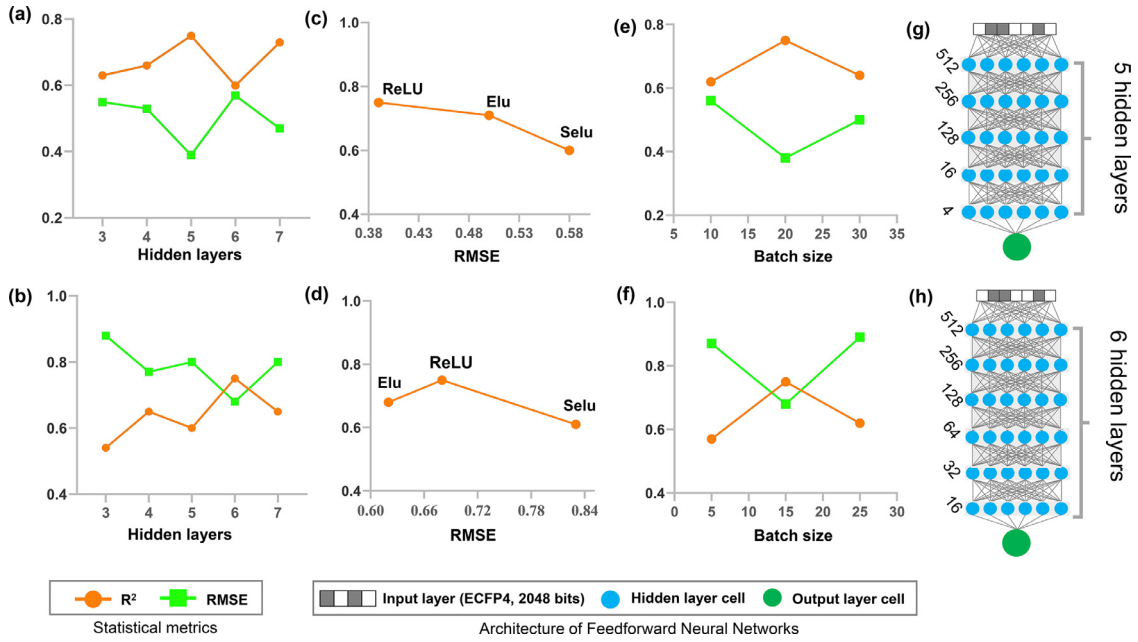

Fig. 3. Different architectures and hyperparameters evaluated for both acute contact and acute oral regression models. (a) number of hidden layers assessed and their R^2 and RMSE results for the acute contact external set. (b) number of hidden layers tested and their R^2 and RMSE results for acute oral external set. (c) activation functions tested and their R^2 results for acute contact external set. (d) activation functions assessed and their R^2 results for acute oral external set. (e) batch sizes evaluated and their R^2 and RMSE results for acute contact external set. (f) batch sizes evaluated and their R^2 and RMSE results for acute oral dataset. (g and h) number of cells in hidden layers of the best models for acute contact and oral toxicities, respectively.

Table 3
Comparison of the external performances of classification models developed here with publicly available models.

Model	Type	Method	Descriptor	*n* test	ACC	SE	SP	MCC	Ref.	Year
Models developed in this work										
Acute contact toxicity	Classification	SVM	FCFP4	77	0.91	0.78	0.96	0.78	–	–
Acute oral toxicity	Classification	RF	MACCS	34	0.88	0.86	0.90	0.76	–	–
Literature models										
Venko et al.	Classification	CPANN	Dragon	22	0.77	0.75	0.79	0.48	[38]	2018
Singh	Classification	PNN	CDK	36	0.87	0.85	1.00	–	[36]	
et al.	Multi-class classification	PNN	CDK	59	0.89	0.85	0.91	–		2014
Li et al.	Classification	SVM	SubFP	43	0.90	0.83	0.93	0.76	[39]	2017
Como et al.	Classification	k-NN	VEGA	41	0.84	0.80	0.86	0.67	[37]	2017
Carnesecchi et al.	Classification	k-NN	Dragon	83	0.90	0.93	0.85	0.78	[93]	2020
Wang et al.	Classification	GACNN	Undirected graph	90	0.83	0.69	0.89	0.59	[29]	2020

Note: RF, Random Forest; FCFP4, functional-class fingerprints with diameter 4; SVM, Support Vector Machine; PNN, Probabilistic Neural Network; k-NN, k-Nearest Neighbor; CPANN, Counter-Propagation Artificial Neural Network; GACNN, Graph Attention Convolutional Neural Network; MACCS, Molecular ACCess System keys; CDK, Chemistry Development Kit; SubFP, Substructure fingerprint; ACC, accuracy; SE, sensitivity; SP, specificity; MCC, Matthews correlation coefficient.

they were developed using a small dataset ($n = 120$) of binary mixtures [30].

3.5.2. Regression models

A comparison of predictive performances of our regression models with publicly available models is shown in Table 4. Again, a rigorous comparison with previous models reported in the literature is not feasible due to different compositions of the training and test sets. Despite that, the best model developed for acute contact toxicity presented lower error on test set compared to the model developed by Hamadache et al. (RMSE = 0.71) [94], Carnesecchi et al. (RMSE = 0.71) [93], and Toropov and Benfenati (RMSE = 0.68) [95], while the best model developed for acute oral toxicity presented comparable statistical performance with them. On the other hand, the best model developed for acute contact toxicity showed comparable statistical performance with the models developed by Devillers et al. (RMSE = 0.39) [96] and Singh et al. (RMSE = 0.33) [36], Although our models presented higher errors on the test set compared to the model of Dulin et al. (RMSE = 0.218), their number of compounds on the test set ($n = 6$) could overestimate the external performance of this model. Mukherjee et al. [97]. did not calculate the RMSE metric of their model. Compared to Mukherjee et al. [97]. work ($R^2 = 0.666$), we have modeled a more structurally representative dataset of chemicals that allowed us obtain more predictive models ($R^2 = 0.75$).

3.6. The BeeToxAI usage

The most predictive classification and regression models for acute contact and oral toxicity were implemented in the BeeToxAI web app (http://beetoxai.labmol.com.br/). The BeeToxAI has an intuitive user interface (Fig. 4), in which the user may draw a compound of interest in the "molecular editor" box or directly paste the Simplified Molecular Input Line Entry Specification (SMILES) string of the queried chemical structure. After hitting the "Submit Analysis" button, the user will receive the classification outcomes (e.g., toxic, nontoxic) using the best classification models developed for acute contact and oral toxicities. Are

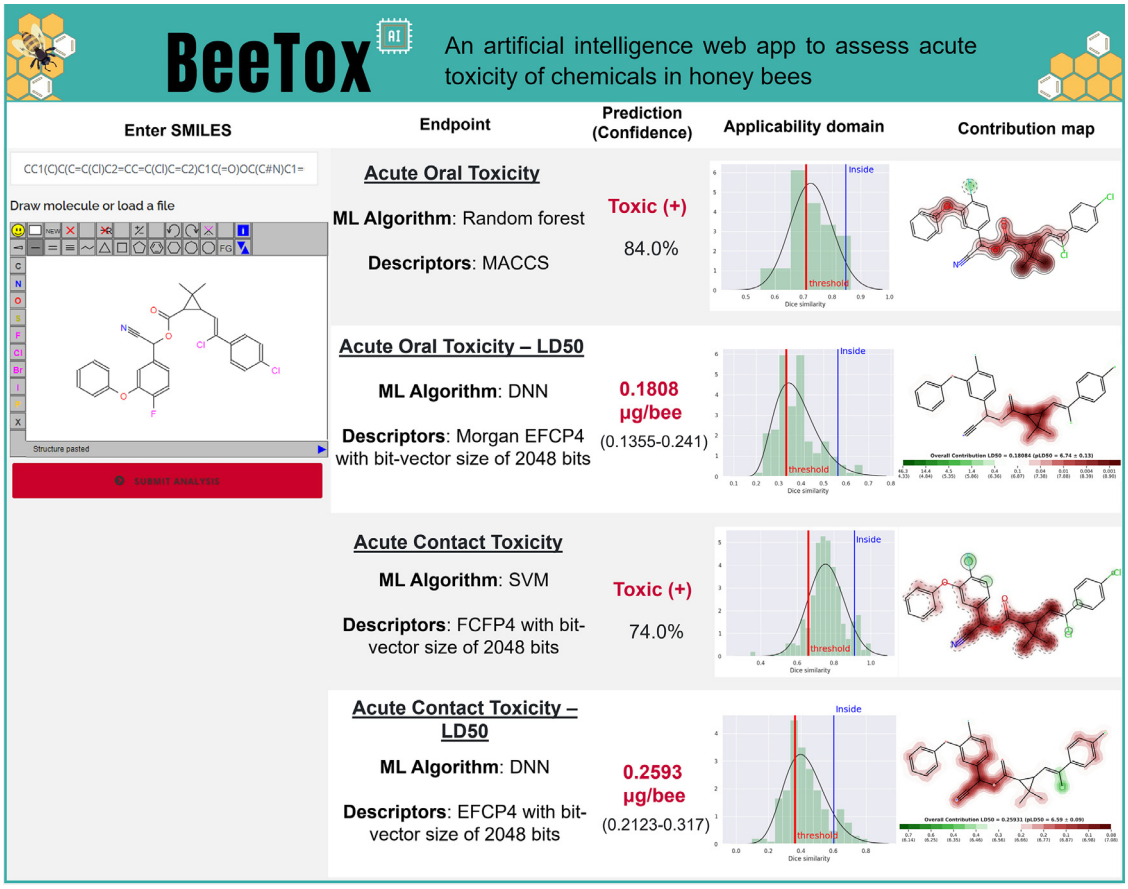

Fig. 4. User interface of BeeToxAI. The query chemical can be drawn in the "molecular editor" box or directly by pasting the SMILES strings. After hitting the "Submit Analysis" button, the user will receive predicted values and LD_{50} for acute contact and oral toxicities, predicted probability values, AD estimates, and color-coded maps of fragment contributions to toxicity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Comparison of the external performance of publicly available vs our regression models.

Model	*n* test	R^2	RMSE	Ref.	Year
Models developed in this work					
Acute contact toxicity	44	0.75	0.39	–	–
Acute oral toxicity	28	0.75	0.68	–	–
Literature models					
Hamadache et al.	16	0.96	0.71	[94]	2018
Devillers et al.	11	0.94	0.39	[96]	2002
Singh et al.	47	0.86	0.33	[36]	2014
Dulin et al.	6	0.85	0.218	[98]	2012
Carnesecchi et al.	25	0.74	0.71	[93]	2020
Toropov and Benfenati	20	0.72	0.68	[95]	2007
Roy et al.	23	0.66	–	[97]	2021

also showed predicted probability values, which are useful for estimating the confidence of classification outcomes [70]. When a compound is classified as toxic by acute contact or oral toxicity classification model, then the prediction by the respective regression model is activated and the predicted LD_{50} is displayed on the screen. All predictions are followed by the AD estimates and mechanistic interpretation using color-coded maps of fragment contribution [72,73]. For the maps, atoms or fragments promoting positive toxicity contributions are highlighted in red, while structural moieties decreasing the toxicity are highlighted in green. Also, the predicted pLD_{50} is shown on the respective contribution map.

3.7. Case study using BeeToxAI

In order to carry out an additional statistical validation of the BeeToxAI app, we compiled and prepared a list of 8 additional pesticides (not included in the QSAR datasets) with honey bee contact and/or oral LD_{50} data used in the United States from 1992 to 2014 (Table S5). Then, we used BeeToxAI to predict these pesticides' contact and oral toxicity potential to honey bees (Figs. 5 and 6).

According to the results of BeeToxAI (Figs. 5 and 6), the acute oral toxicity classification model correctly classified four out of five pesticides (three of the listed pesticides did not have experimental data on acute oral toxicity) and the acute contact toxicity classification model correctly classified all of the eight pesticides. The pesticide incorrectly classified by the acute oral toxicity classification model was correctly predicted by the respective regression model. These results corroborate with the high external predictive power reported above. On the other hand, two compounds were erroneously predicted by the acute contact toxicity regression model. Most of the incorrect predictions were outside of the model AD. Therefore, the analysis of predictions using BeeToxAI should be cautious when considering chemicals outside the DA, since the classification and regression models were trained using small datasets.

All predictions are accompanied by color-coded maps of fragment contributions to toxicity (Figs. 5 and 6). These contribution maps estimate the weight of single fragments or atoms to toxicity by removing them from the structure and calculating the difference between the predicted values for the initial structure and the structure with a removed fragment. This enables estimation of the contributions of whole fragments (functional groups, scaffolds, linkers) as well as their combined

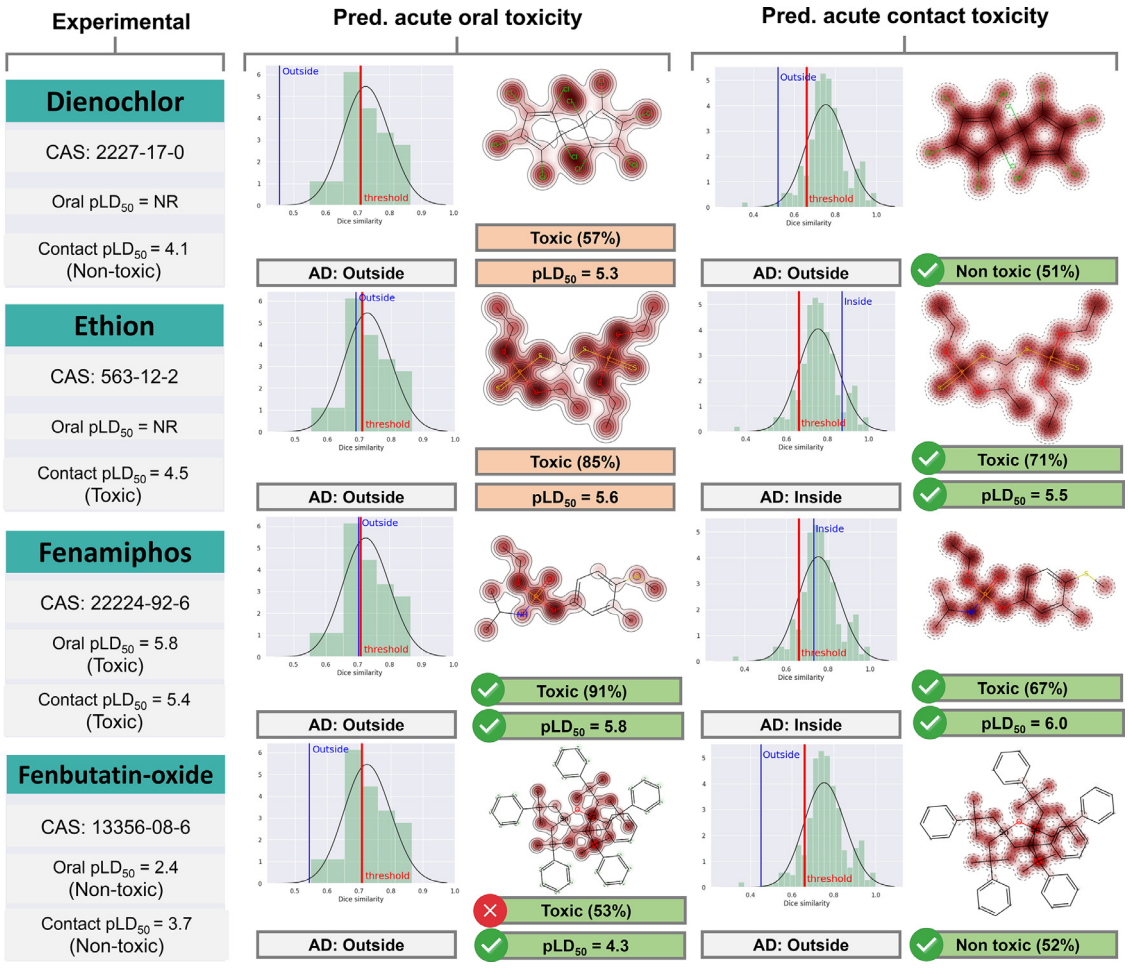

Fig. 5. Experimental and predicted toxicity of four pesticides from the external set and structural fragments' contribution to toxicity. NR = Not Reported. Fragments increasing the toxicity are highlighted in red and fragments decreasing the toxicity are colored in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

effect if two or more groups are removed simultaneously from the structure [73].

Overall, contribution maps from BeeToxAI allow the user to analyze the individual contribution of each fragment for acute oral toxicity and contact toxicity. In addition, the AD and prediction probabilities provide further context to the model outputs that allow for estimates of overall confidence. Other existing models do not provide this degree of mechanistic interpretation [29,36–39,93]. Mechanistic interpretation of predictions reported here can lead to such additional benefits as (i) elicitation of structural alerts; (ii) insights on the mode of action; and (iii) utility to chemists who want to modify the molecule [99,100]. BeeToxAI indicated the significant contribution of several fragments for toxicity, including phosphinates (see fenamiphos acute oral toxicity map, Fig. 5) and phosphonates (see oxydemeton-methyl and phosphamidon maps, Fig. 6). In view of this, we overcame an inherent trade-off between predictive performance and interpretability that has been traditionally assumed [101], implementing highly predictive and interpretable QSAR models in BeeToxAI.

4. Conclusion

In this work, we developed and rigorously validated classification and regression QSAR models that accurately predict acute contact and oral toxicities of chemicals to honey bees. Benchmarking with existing computational tools demonstrated superior or comparable performance. In addition, we implemented a visual algorithm that facilitates model interpretation. These models were implemented in the BeeToxAI web app – a fast, reliable, and user-friendly tool for the assessment of acute chemical toxicity to honey bees. Users can make predictions using rigorously and externally validated computational models that fulfill all the OECD principles for developing and validating QSAR models for regulatory purposes. The web app does not require any prior knowledge of programming or computational skills for its utilization. The predictions for a single compound take only a few seconds.

Furthermore, BeeToxAI provides the user with the following outcomes: (i) toxic/nontoxic classification for acute contact toxicity and acute oral toxicity honey bees endpoints; (ii) confidence of the predictions; (iii) applicability domain estimation; and (iv) color-coded contribution maps illustrating the relative contribution of chemical fragments for toxicity. The web app, designed to openly share the aforementioned predictive models, is freely available to the public at http://beetoxai.labmol.com.br/. We propose these models as a valuable contribution to the scientific community that enables regulators and regulated companies to rapidly evaluate the risk of chemical harm to honey bees (*Apis mellifera)* for the registration of the new pesticides. Future directions of the BeeToxAI include the implementation of Quantitative Activity-Activity Relationships (QAAR) [102], Mode of Action (MoA) predictors, multi-task (species) models, and read-across (nearest neighbors) [23]. In this context, the ongoing BeeToxAI project aims to implement new predictive models to assess the acute toxicity of active ingredients and mixtures [30] against multiple stages (adult and larvae) of honey bees and bumble bees.

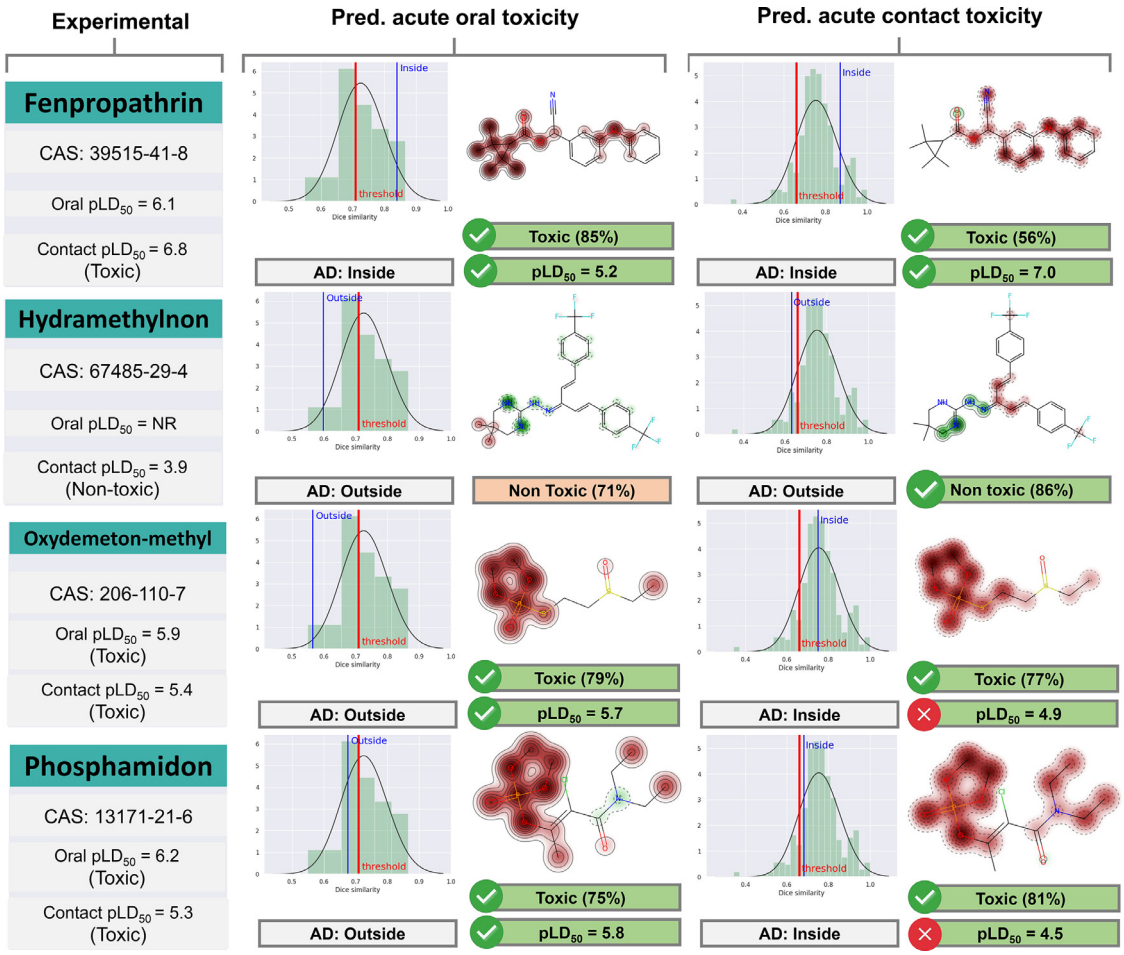

Fig. 6. Experimental and predicted toxicity of four pesticides from the external set and structural fragments' contribution to toxicity. NR = Not Reported. Fragments increasing the toxicity are highlighted in red and fragments decreasing the toxicity are colored in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rodolpho C. Braga is CTO of InsilicAll. The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

José T. Moreira-Filho: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Rodolpho C. Braga:** Conceptualization, Formal analysis. **Jade Milhomem Lemos:** Methodology, Formal analysis. **Vinicius M. Alves:** Formal analysis. **Joyce V.V.B. Borba:** Methodology, Formal analysis. **Wesley S. Costa:** Formal analysis. **Nicole Kleinstreuer:** Supervision. **Eugene N. Muratov:** Supervision. **Carolina Horta Andrade:** Conceptualization, Supervision. **Bruno J. Neves:** Conceptualization, Supervision.

Funding sources

B.J.N. is supported by CNPq (grant 425119/2018–1). E.N.M. is supported by NIH (grants 1U01CA207160 and GM5105946). R.C.B. is supported by AWS Amazon sponsorship. V.M.A. thanks the Lush Prize. C.H.A. is supported by CNPq (grant 400760/2014–2). C.H.A. also thanks the "L'Oréal-UNESCO-ABC Para Mulheres na Ciência" and "L'Oréal-UNESCO International Rising Talents" for the awards and fellowships received, which partially funded this work. **Notes** R.C.B. is CTO of InsilicAll Inc. The other authors declare they have no actual or potential competing financial interests.

Acknowledgment

The authors appreciate the financial support from the Brazilian funding agencies, CNPq, FAPEG, FAPESP, and CAPES. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research and ChemAxon for providing us with an academic license for their software. The authors gratefully thank Ms. Zoe Sessions for her kind help with editing the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ailsci.2021.100013.

References

[1] Oerke EC. Crop losses to pests. J Agric Sci 2006;144(1):31–43. doi:10.1017/S0021859605005708.
[2] Frische T, Egerer S, Matezki S, Pickl C, Wogram J. 5-Point programme for sustainable plant protection. Environ Sci Eur 2018;30(1):8. doi:10.1186/s12302-018-0136-2.

[3] Daam MA, Chelinho S, Niemeyer JC, Owojori OJ, De Silva PMCS, Sousa JP, van Gestel CAM, Römbke J. Environmental risk assessment of pesticides in tropical terrestrial ecosystems: test procedures, current status and future perspectives. Ecotoxicol Environ Saf 2019;181:534–47. doi:10.1016/j.ecoenv.2019.06.038.

[4] Cuevas N, Martins M, Costa PM. Risk assessment of pesticides in estuaries: a review addressing the persistence of an old problem in complex environments. Ecotoxicology 2018;27(7):1008–18. doi:10.1007/s10646-018-1910-z.

[5] Osborne JL. Bumblebees and pesticides. Nature 2012;491(7422):43–5. doi:10.1038/nature11637.

[6] Rhodes CJ. Pollinator decline – an ecological calamity in the making? Sci Prog 2018;101(2):121–60. doi:10.3184/003685018X15202512854527.

[7] Butler DEU. Expected to vote on pesticide ban after major scientific review. Nature 2018;555(7695):150–1. doi:10.1038/d41586-018-02639-1.

[8] Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL. Patterns of widespread decline in North American bumble bees. Proc Natl Acad Sci 2011;108(2):662–7. doi:10.1073/pnas.1014743108.

[9] Fairbrother A, Purdy J, Anderson T, Fell R. Risks of neonicotinoid insecticides to honey bees. Environ Toxicol Chem 2014;33(4):719–31. doi:10.1002/etc.2527.

[10] Banks JE, Banks HT, Myers N, Laubmeier AN, Bommarco R. Lethal and sublethal effects of toxicants on bumble bee populations: a modeling approach. Ecotoxicology 2020;29(3):237–45. doi:10.1007/s10646-020-02162-y.

[11] Thorbek P, Campbell PJ, Thompson HM. Colony impact of pesticide-induced sublethal effects on honey bee workers: a simulation study using BEEHAVE. Environ Toxicol Chem 2017;36(3):831–40. doi:10.1002/etc.3581.

[12] Sponsler DB, Grozinger CM, Hitaj C, Rundlöf M, Botías C, Code A, Lonsdorf EV, Melathopoulos AP, Smith DJ, Suryanarayanan S, Thogmartin WE, Williams NM, Zhang M, Douglas MR. Pesticides and pollinators: a socioecological synthesis. Sci Total Environ 2019;662:1012–27. doi:10.1016/j.scitotenv.2019.01.016.

[13] Ollerton J. Pollinator diversity: distribution, ecological function, and conservation. Annu Rev Ecol Evol Syst 2017;48(1):353–76. doi:10.1146/annurev-ecolsys-110316-022919.

[14] OECD. Test No. 213: honey bees, acute oral toxicity test. 2021. doi:10.1787/9789264070165-en. 1998. (acessed May 14, 2020).

[15] Organisation for Economic Co-operation and Development Test No. 214: honey bees, acute contact toxicity test; 1998. (accessed May 14, 2020). doi:10.1787/9789264070189-en.

[16] U.S. Environmental Protection Agency Guidance for assessing pesticides risks to bees; 2014 https://wwwepagov/pollinator-protection/pollinator-risk-assessment-guidance (accessed Apr 13, 2020).

[17] U.S. Environmental Protection Agency Honey bee toxicity testing frequently asked questions; 2021 https://wwwepagov/pesticides/new-frequently-asked-questions-honey-bee-toxicity-testing-registrants-and-contract (accessed Jul 13, 2020).

[18] U.S. Environmental Protection Agency Ecological Effects Test Guidelines OCSPP 850.3030: Honey Bee Toxicity of Residues on Foliage; 2012. https://nepis.epa.gov/Exe/ZyNET.exe/P100IRF8.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2011+Thru+2015&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery= (accessed May 17, 2020).

[19] U.S. Environmental Protection Agency. Cost estimates of studies required for pesticide registration. 2019. https://www.epa.gov/pesticide-registration/cost-estimates-studies-required-pesticide-registration (accessed Jul 13, 2020).

[20] Myatt GJ, Ahlberg E, Akahori Y, Allen D, Amberg A, Anger LT, Aptula A, Auerbach S, Beilke L, Bellion P, Benigni R, Bercu J, Booth ED, Bower D, Brigo A, Burden N, Cammerer Z, Cronin MTD, Cross KP, Custer L, Dettwiler M, Dobo K, Ford KA, Fortin MC, Gad-McDonald SE, Gellatly N, Gervais V, Glover KP, Glowienke S, Van Gompel J, Gutsell S, Hardy B, Harvey JS, Hillegass J, Honma M, Hsieh JH, Hsu CW, Hughes K, Johnson C, Jolly R, Jones D, Kemper R, Kenyon MO, Kim MT, Kruhlak NL, Kulkarni SA, Kümmerer K, Leavitt P, Majer B, Masten S, Miller S, Moser J, Mumtaz M, Muster W, Neilson L, Oprea TI, Patlewicz G, Paulino A, Lo Piparo E, Powley M, Quigley DP, Reddy MV, Richarz AN, Ruiz P, Schilter B, Serafimova R, Simpson W, Stavitskaya L, Stidl R, Suarez-Rodriguez D, Szabo DT, Teasdale A, Trejo-Martin A, Valentin JP, Vuorinen A, Wall BA, Watts P, White AT, Wichard J, Witt KL, Woolley A, Woolley D, Zwickl C, Hasselgren C. *In silico* toxicology protocols. Regul Toxicol Pharmacol 2018;96(April):1–17. doi:10.1016/j.yrtph.2018.04.014.

[21] Russell W, Burch R. The principles of humane experimental technique. 1st ed. London: Methuen & Co Ltd; 1959.

[22] Cronin MTD. Chapter 1. An introduction to chemical grouping, categories and read-across to predict toxicity. In: Chemical toxicity prediction: category formation and read-across. The Royal Society of Chemistry; 2013. p. 1–29. doi:10.1039/9781849734400-00001.

[23] Alves VM, Muratov EN, Capuzzi SJ, Politi R, Low Y, Braga RC, Zakharov AV, Sedykh A, Mokshyna E, Farag S, Andrade CH, Kuz'min VE, Fourches D, Tropsha A. Alarms about structural alerts. Green Chem 2016;18(16):4348–60. doi:10.1039/C6GC01492E.

[24] Jaworska JS, Comber M, Auer C, Van Leeuwen CJ. Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. Environ Health Perspect 2003;111(10):1358–60. doi:10.1289/ehp.5757.

[25] Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, Poroikov V, Oprea TI, Baskin II, Varnek A, Roitberg A, Isayev O, Curtalolo S, Fourches D, Cohen Y, Aspuru-Guzik A, Winkler DA, Agrafiotis D, Cherkasov A, Tropsha A. QSAR without borders. Chem Soc Rev 2020;49(11):3525–64. doi:10.1039/D0CS00098A.

[26] Gini G, Zanoli F, Roy K. Machine learning and deep learning methods in ecotoxicological QSAR modeling. In: Ecotoxicological QSARs. New York, NY: Humana Press; 2020. p. 111–49. editor. doi:10.1007/978-1-0716-0150-1_6.

[27] Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, Hickey AJ, Clark AM. Exploiting machine learning for end-to-end drug discovery and development. Nat Mater 2019;18(5):435–41. doi:10.1038/s41563-019-0338-z.

[28] Mitchell JBO. Machine learning methods in chemoinformatics. Wiley Interdiscip Rev Comput Mol Sci 2014;4(5):468–81. doi:10.1002/wcms.1183.

[29] Wang F, Yang JF, Wang MY, Jia CY, Shi XX, Hao GF, Yang GF. Graph attention convolutional neural network model for chemical poisoning of honey bees' prediction. Sci Bull 2020;65(14):1184–91. doi:10.1016/j.scib.2020.04.006.

[30] Carnesecchi E, Toropov AA, Toropova AP, Kramer N, Svendsen C, Dorne J, Lou, Benfenati E. Predicting acute contact toxicity of organic binary mixtures in honey bees (A. mellifera) through innovative QSAR models. Sci Total Environ 2020;704:135302. doi:10.1016/j.scitotenv.2019.135302.

[31] Organisation for Economic Co-operation and Development. Principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationship models, 2004, https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf (accessed Apr 27, 2020).

[32] Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and qsar modeling research. J Chem Inf Model 2010;50(7):1189–204. doi:10.1021/ci100176x.

[33] Fourches D, Muratov E, Tropsha A. Curation of chemogenomics data. Nat Chem Biol 2015;11(8):535. doi:10.1038/nchembio.1881.

[34] Fourches D, Muratov E, Tropsha A. Trust, but verify II: a practical guide to chemogenomics data curation. J Chem Inf Model 2016;56(7):1243–52. doi:10.1021/acs.jcim.6b00129.

[35] Tropsha A. Best practices for QSAR model development, validation, and exploitation. Mol Inform 2010;29(6–7):476–88. doi:10.1002/minf.201000061.

[36] Singh KP, Gupta S, Basant N, Mohan D. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. Chem Res Toxicol 2014;27(9):1504–15. doi:10.1021/tx500100m.

[37] Como F, Carnesecchi E, Volani S, Dorne JL, Richardson J, Bassan A, Pavan M, Benfenati E. Predicting acute contact toxicity of pesticides in honey bees (Apis mellifera) through a k-nearest neighbor model. Chemosphere 2017;166:438–44. doi:10.1016/j.chemosphere.2016.09.092.

[38] Venko K, Drgan V, Novič M. Classification models for identifying substances exhibiting acute contact toxicity in honey bees (Apis mellifera). SAR QSAR Environ Res 2018;29(9):743–54. doi:10.1080/1062936X.2018.1513953.

[39] Li X, Zhang Y, Chen H, Li H, Zhao Y. Insights into the molecular basis of the acute contact toxicity of diverse organic chemicals in the honey bee. J Chem Inf Model 2017;57(12):2948–57. doi:10.1021/acs.jcim.7b00476.

[40] Iwasa T, Motoyama N, Ambrose JT, Roe RM. Mechanism for the differential toxicity of neonicotinoid insecticides in the honey bee, Apis mellifera. Crop Prot 2004;23(5):371–8. doi:10.1016/j.cropro.2003.08.018.

[41] Laurino D, Porporato M, Patetta A, Manino A. Toxicity of neonicotinoid insecticides to honey bees: laboratory tests. Bull Insectology 2011;64(1):107–13.

[42] Sanchez-Bayo F, Goka K. Pesticide residues and bees – a risk assessment. PLoS One 2014;9(4):e94482. doi:10.1371/journal.pone.0094482.

[43] Atkins EL, Kellum D. Comparative morphogenic and toxicity studies on the effect of pesticides on honey bee brood. J Apic Res 1986;25(4):242–55. doi:10.1080/00218839.1986.11100725.

[44] Thompson HM. Assessing the exposure and toxicity of pesticides to bumblebees (Bombus Sp.). Apidologie 2001;32(4):305–21 (Celle). doi:10.1051/apido:2001131.

[45] Hu YT, Wu TC, Yang EC, Wu PC, Lin PT, Wu YL. Regulation of genes related to immune signaling and detoxification in apis mellifera by an inhibitor of histone deacetylation. Sci Rep 2017;7(1):41255. doi:10.1038/srep41255.

[46] Pohorecka K, Szczęsna T, Witek M, Miszczak A, Sikorski P. The exposure of honey bees to pesticide residues in the hive environment with regard to winter colony losses. J Apic Sci 2017;61(1):105–25. doi:10.1515/jas-2017-0013.

[47] Como F, Carnesecchi E, Volani S, Dorne JL, Richardson J, Bassan A, Pavan M, Benfenati E. Predicting acute contact toxicity of pesticides in honey bees (Apis mellifera) through a k-nearest neighbor model. Chemosphere 2017;166:438–44. doi:10.1016/j.chemosphere.2016.09.092.

[48] Tsvetkov N, Samson-Robert O, Sood K, Patel HS, Malena DA, Gajiwala PH, Maciukiewicz P, Fournier V, Zayed A. Chronic exposure to neonicotinoids reduces honey bee health near corn crops. Science 2017;356(6345):1395–7 *(80-.)*. doi:10.1126/science.aam7470.

[49] Mullin CA, Frazier M, Frazier JL, Ashcraft S, Simonds R, VanEngelsdorp D, Pettis JS. High levels of miticides and agrochemicals in North American apiaries: implications for honey bee health. PLoS One 2010;5(3):e9754. doi:10.1371/journal.pone.0009754.

[50] Decourtye A, Devillers J, Genecque E, Menach K, Le, Budzinski H, Cluzeau S, Pham-Delègue MH. Comparative sublethal toxicity of nine pesticides on olfactory learning performances of the honey bee Apis mellifera. Arch Environ Contam Toxicol 2005;48(2):242–50. doi:10.1007/s00244-003-0262-7.

[51] Bovi TS, Zaluski R, Orsi RO. Toxicity and motor changes in Africanized honey bees (Apis mellifera L.) exposed to fipronil and imidacloprid. An Acad Bras Cienc 2018;90(1):239–45. doi:10.1590/0001-3765201820150191.

[52] Badawy MEI, Nasr HM, Rabea EI. Toxicity and biochemical changes in the honey bee Apis mellifera exposed to four insecticides under laboratory conditions. Apidologie 2015;46(2):177–93 (Celle). doi:10.1007/s13592-014-0315-0.

[53] U.S. Environmental Protection Agency. Ecotoxicology database (ECOTOX), 2019, https://cfpub.epa.gov/ecotox/ (accessed Apr 13, 2020).

[54] Dorne JLou, Richardson J, Kass G, Georgiadis N, Monguidi M, Pasinato L, Cappe S, Verhagen H, Robinson T. Editorial: OpenFoodTox: EfSA's open source toxicological database on chemical hazards in food and feed. EFSA J 2017;15(1):e15011. doi:10.2903/j.efsa.2017.e15011.

[55] Sushko I, Novotarskyi S, Körner R, Pandey AK, Rupp M, Teetz W, Brandmaier S, Abdelaziz A, Prokopenko VV, Tanchuk VY, Todeschini R, Varnek A, Marcou G, Ertl P, Potemkin V, Grishina M, Gasteiger J, Schwab C, Baskin II, Palyulin VA, Radchenko EV, Welsh WJ, Kholodovych V, Chekmarev D, Cherkasov A, Aires-de-Sousa J, Zhang QY, Bender A, Nigsch F, Patiny L, Williams A, Tkachenko V, Tetko IV. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J Comput Aided Mol Des 2011;25(6):533–54. doi:10.1007/s10822-011-9440-2.

[56] U.S. Environmental Protection Agency. Pesticide product information system database, 2020, https://www.epa.gov/ingredients-used-pesticide-products (accessed Apr 13, 2020).

[57] Sander T, Freyss J, von Korff M, Rufener C. DataWarrior: an open-source program for chemistry aware data visualization and analysis. J Chem Inf Model 2015;55(2):460–73. doi:10.1021/ci500588j.

[58] Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009. Python Software Foundation. Python language reference http://www.python.org(accessed Apr 13, 2020)

[59] Landrum, G. RDKit: Open-source cheminformatics software, 2010, http://www.rdkit.org/ (accessed Jul 11, 2020).

[60] Riniker S, Landrum GA. Open-source platform to benchmark fingerprints for ligand-based virtual screening. J Cheminform 2013;5(1):26. doi:10.1186/1758-2946-5-26.

[61] Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model 2010;50(5):742–54. doi:10.1021/ci100050t.

[62] Gobbi A, Poppinger D. Genetic optimization of combinatorial libraries. Biotechnol Bioeng 1998;61(1):47–54. doi:10.1002/(SICI)1097-0290(199824)61:1<47::AID-BIT9>3.0.CO;2-Z.

[63] Vapnik VV. The nature of statistical learning theory. 2nd ed. New York: Springer; 2000.

[64] Breiman LEO. Random forests. Mach Learn 2001;45:5–32. doi:10.1023/A:1010933404324.

[65] 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. Scikit-learn: machine learning in Python. J Mach Learn Res 2012;12:2825–30. doi:10.1007/s13398-014-0173-7.2.

[66] Wu J, Chen XY, Zhang H, Xiong LD, Lei H, Deng SH. Hyperparameter optimization for machine learning models based on Bayesian optimization. J Electron Sci Technol 2019;17(1):26–40. doi:10.11989/JEST.1674-862X.80904120.

[67] Jiang D, Lei T, Wang Z, Shen C, Cao D, Hou T. ADMET evaluation in drug discovery. 20. Prediction of breast cancer resistance protein inhibition through machine learning. J Cheminform 2020;12(1):16. doi:10.1186/s13321-020-00421-y.

[68] Cohen J. A coefficient of agreement of nominal scales. Educ Psychol Meas 1960;20:37–46. doi:10.1177/001316446002000104.

[69] Sørensen TJ. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyzes of the vegetation on danish commons. 5th ed. I kommission hos E. Munksgaard; 1948.

[70] Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. J Comput Aided Mol Des 2003;17(2–4):241–53. doi:10.1023/a:1025386326946.

[71] Consonni V, Ballabio D, Todeschini R. Comments on the definition of the Q 2 parameter for QSAR validation. J Chem Inf Model 2009;49(7):1669–78. doi:10.1021/ci900115y.

[72] Neves BJ, Braga RC, Alves VM, Lima MNdo N, Cassiano GC, Muratov EN, Costa FTM, Andrade CH. Deep learning-driven research for drug discovery: tackling malaria. PLoS Comput Biol 2019 *in press*.. doi:10.1371/journal.pcbi.1007025.

[73] Riniker S, Landrum GA. Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods. J Cheminform 2013;5(9):1–7. doi:10.1186/1758-2946-5-43.

[74] FAUN Publication. Benefits Of Kubernetes For Microservices Architecture. 2019 https://faun.pub/benefits-of-kubernetes-for-microservices-architecture-a04704f0d3a0 (accessed Mar 1, 2020).

[75] Unbit. The uWSGI project, 2016, https://uwsgi-docs.readthedocs.org, (accessed Apr 13, 2020).

[76] Pluralsight. JavaScript, 2016, https://www.javascript.com/ (accessed May 17, 2020).

[77] Pallets. Flask web development, one drop at time, 2010, https://flask.palletsprojects.com/en/1.1.x/ (accessed May 17, 2020).

[78] Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007;9(3):90–5. doi:10.1109/MCSE.2007.55.

[79] Waskom, M., Botvinnik, O., Kunt, D.O., Hobson, P., Lukauskas, S., Gemperline, D.C., Augspurger, T., Halchenko, Y., Cole, J.B., Warmenhoven, J., de Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunt, G. mwaskom/seaborn: v0.8.1. 2017 https://github.com/mwaskom/seaborn/tree/v0.8.1 (accessed Mar 1, 2020).

[80] Bienfait B, Ertl P. JSME: a free molecule editor in javascript. J Cheminform 2013;5(1):24. doi:10.1186/1758-2946-5-24.

[81] GitLab. The DevOps Platform. 2020 https://about.gitlab.com/ (accessed Jul 2, 2020).

[82] Git https://git-scm.com/ (accessed Aug 2, 2020).

[83] Laurino D, Manino A, Patetta A, Porporato M. Toxicity of neonicotinoid insecticides on different honey bee genotypes. Bull Insectology 2013;66(1):119–26.

[84] Blacquière T, Smagghe G, van Gestel CAM, Mommaerts V. Erratum to: neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. Ecotoxicology 2012;21(5):1581. doi:10.1007/s10646-012-0890-7.

[85] Suchail S, Guez D, Belzunces LP. Characteristics of imidacloprid toxicity in two apis mellifera subspecies. Environ Toxicol Chem 2000;19(7):1901–5. doi:10.1002/etc.5620190726.

[86] Alves VM, Muratov EN, Zakharov A, Muratov NN, Andrade CH, Tropsha A. Chemical toxicity prediction for major classes of industrial chemicals: is it possible to develop universal models covering cosmetics, drugs, and pesticides? Food Chem Toxicol 2018;112:526–34. doi:10.1016/j.fct.2017.04.008.

[87] Stumpfe D, Bajorath J. Exploring activity cliffs in medicinal chemistry. J Med Chem 2012;55(7):2932–42. doi:10.1021/jm201706b.

[88] Wassermann AM, Wawer M, Bajorath J. Activity landscape representations for structure–activity relationship analysis. J Med Chem 2010;53(23):8209–23. doi:10.1021/jm100933w.

[89] Stumpfe D, Hu Y, Dimova D, Bajorath J. Recent progress in understanding activity cliffs and their utility in medicinal chemistry. J Med Chem 2014;57(1):18–28. doi:10.1021/jm401120g.

[90] Zakharov AV, Peach ML, Sitzmann M, Nicklaus MC. QSAR modeling of imbalanced high-throughput screening data in pubchem. J Chem Inf Model 2014;54(3):705–12. doi:10.1021/ci400737s.

[91] Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. In: Proceedings of the 22nd international conference on machine learning - ICML '05. ACM Press; 2005. p. 625–32. doi:10.1145/1102351.1102430.

[92] Wallace BC, Dahabreh IJ. Class probability estimates are unreliable for imbalanced data (and how to fix them). In: Proceedings of the IEEE 12th International Conference on Data Mining; 2012. p. 695–704. IEEE. doi:10.1109/ICDM.2012.115.

[93] Carnesecchi E, Toma C, Roncaglioni A, Kramer N, Benfenati E, Dorne JLCM. Integrating QSAR models predicting acute contact toxicity and mode of action profiling in honey bees (A. mellifera): data curation using open source databases, performance testing and validation. Sci Total Environ 2020;735:139243. doi:10.1016/j.scitotenv.2020.139243.

[94] Hamadache M, Benkortbi O, Hanini S, Amrane A. QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (Apis mellifera L.). Environ Sci Pollut Res 2018;25(1):896–907. doi:10.1007/s11356-017-0498-9.

[95] Toropov AA, Benfenati E. SMILES as an alternative to the graph in QSAR modeling of bee toxicity. Comput Biol Chem 2007;31(1):57–60. doi:10.1016/j.compbiolchem.2007.01.003.

[96] Devillers J, Pham-Delègue MH, Decourtye A, Budzinski H, Cluzeau S, Maurin G. Structure-toxicity modeling of pesticides to honey bees. SAR QSAR Environ Res 2002;13(7–8):641–8. doi:10.1080/1062936021000043391.

[97] Mukherjee RK, Kumar V, Roy K. Chemometric modeling of plant protection products (PPPs) for the prediction of acute contact toxicity against honey bees (A. mellifera): a 2D-QSAR approach. J Hazard Mater 2021;423:127230 PB. doi:10.1016/j.jhazmat.2021.127230.

[98] Dulin F, Halm-Lemeille MP, Lozano S, Lepailleur A, Sopkova-de Oliveira Santos J, Rault S, Bureau R. Interpretation of honey bees contact toxicity associated to acetylcholinesterase inhibitors. Ecotoxicol Environ Saf 2012;79:13–21. doi:10.1016/j.ecoenv.2012.01.007.

[99] Polishchuk P. Interpretation of quantitative structure–activity relationship models: past, present, and future. J Chem Inf Model 2017;57(11):2618–39. doi:10.1021/acs.jcim.7b00274.

[100] Sheridan RP. Interpretation of QSAR models by coloring atoms according to changes in predicted activity: how robust is it? J Chem Inf Model 2019;59(4):1324–37. doi:10.1021/acs.jcim.8b00825.

[101] Marchese Robinson RL, Palczewska A, Palczewski J, Kidley N. Comparison of the predictive performance and interpretability of random forest and linear models on benchmark data sets. J Chem Inf Model 2017;57(8):1773–92. doi:10.1021/acs.jcim.6b00753.

[102] Bouhedjar K, Benfenati E, Nacereddine AK. Modeling quantitative structure activity–activity relationships (QSAARs): auto-pass-pass, a new approach to fill data gaps in environmental risk assessment under the REACH regulation. SAR QSAR Environ Res 2020;31(10):785–801. doi:10.1080/1062936X.2020.1810770.