SAR and QSAR in Environmental Research

ISSN: 1062-936X (Print) 1029-046X (Online) Journal homepage: http://www.tandfonline.com/loi/gsar20

Machine learning-based models to predict modes of toxic action of phenols to Tetrahymena pyriformis

J. A. Castillo-Garit, G. M. Casañola-Martin, S. J. Barigye, H. Pham-The, F. Torrens & A. Torreblanca

To cite this article: J. A. Castillo-Garit, G. M. Casañola-Martin, S. J. Barigye, H. Pham-The, F. Torrens & A. Torreblanca (2017) Machine learning-based models to predict modes of toxic action of phenols to Tetrahymena pyriformis, SAR and QSAR in Environmental Research, 28:9, 735-747, DOI: 10.1080/1062936X.2017.1376705

To link to this article: http://dx.doi.org/10.1080/1062936X.2017.1376705

Published online: 12 Oct 2017.

SAR AND QSAR IN ENVIRONMENTAL RESEARCH, 2017
VOL. 28, NO. 9, 735–747
https://doi.org/10.1080/1062936X.2017.1376705

Taylor & Francis
Taylor & Francis Group

Machine learning-based models to predict modes of toxic action of phenols to *Tetrahymena pyriformis*

J. A. Castillo-Garit[a,b], G. M. Casañola-Martin[c], S. J. Barigye[d], H. Pham-The[e], F. Torrens[f] and A. Torreblanca[b]

[a]Unidad de Toxicología Experimental, Universidad de Ciencias Médicas de Villa Clara, Santa Clara, Villa Clara, Cuba; [b]Departament de Biología Funcional i Antropología Física, Universitat de València, Burjassot, Spain; [c]Departamento de Química Física, Facultad de FarmaciaUnidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Universitat de València, Spain; [d]Department of Chemistry, McGill University, Montréal, Québec, Canada; [e]Hanoi University of Pharmacy, Hoan Kiem, Hanoi, Vietnam; [f]Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, Valencia, Spain

ABSTRACT

The phenols are structurally heterogeneous pollutants and they present a variety of modes of toxic action (MOA), including polar narcotics, weak acid respiratory uncouplers, pro-electrophiles, and soft electrophiles. Because it is often difficult to determine correctly the mechanism of action of a compound, quantitative structure-activity relationship (QSAR) methods, which have proved their interest in toxicity prediction, can be used. In this work, several QSAR models for the prediction of MOA of 221 phenols to the ciliated protozoan *Tetrahymena pyriformis*, using Chemistry Development Kit descriptors, are reported. Four machine learning techniques (ML), *k*-nearest neighbours, support vector machine, classification trees, and artificial neural networks, have been used to develop several models with higher accuracies and predictive capabilities for distinguishing between four MOAs. They showed global accuracy values between 95.9% and 97.7% and area under Receiver Operator Curve values between 0.978 and 0.998; additionally, false alarm rate values were below 8.2% for training set. In order to validate our models, cross-validation (10-folds-out) and external test-set were performed with good behaviour in all cases. These models, obtained with ML techniques, were compared with others previously reported by other researchers, and the improvement was significant.

ARTICLE HISTORY
Received 26 July 2017
Accepted 1 September 2017

KEYWORDS
Molecular descriptor; machine learning technique; QSAR; mode of toxic action; phenol derivative; pollutant

Introduction

Phenolic compounds are environmental pollutants that exhibit their toxicity via different mechanisms of toxic actions (MOAs) [1]. The interactions between toxicants and living systems are poorly understood; that is why determining the correct MOA of a chemical compound is not easy [2]. This problem becomes more pronounced when the mechanistic classes are unevenly distributed, as in the case of 221 phenols investigated in the present report. Phenols have been of interest to environmental toxicologists because of the increasing use

CONTACT J. A. Castillo-Garit jacgarit@yahoo.es, juan.castillo@uv.es

of them in various industries, such as paper manufacturing and resin production [3,4]. Others, such as chlorophenols used as bactericides, fungicides, and herbicides [5], are very toxic due to its hydrophobicity and capability of persistence in the environment [6,7]. Most synthetic phenolic compounds are toxic and it is not surprising that a lot of them are classified as hazardous pollutants [4], and some of these products have been catalogued as potentially hazardous by the Environment Protective Agency [8,9].

It is recognized that substitution of a phenolic ring may result in a number of different mechanisms of toxic action. These range from non-reactive polar narcosis, through to respiratory uncoupling of oxidative phosphorylation, directly attack nucleophilic sites mediated by soft electrophiles, as well as metabolic activation before undergoing respective covalent interactions, this last one a distinctive feature of compounds classified as proelectrophiles [10]. There are different methods of classifying compounds according to MOAs, the most frequent and conventional one is the rules-based method [11]; because of its limitations [2], attention has been focused on the use of quantitative approach based on statistical classification in predicting MOAs [12–14]. Quality toxicity data is required to formulate and validate high-quality quantitative structure–activity relationship (QSAR) studies [15], for example the database published by Aptula et al. [12] to develop descriptor-based classification models. Several studies, based on various statistical methods, for separating phenols into these four MOAs, had been derived for this dataset, using quantum mechanical, whole molecule, and empirically based molecular descriptors with known or pre-assigned mechanism in a learning database [12,13,16–19].

Both Aptula et al. [12] and Ren [13], in their investigations, have used linear discriminant analysis (LDA). However, other authors have used regression methods [18] and neural networks [19–21], but to a minor amount. Aptula et al. [12] had problems with correctly developing models for assigning the correct MOA using linear discriminant analysis (LDA). Thus, the use of machine learning (ML) approaches, such as support vector machine (SVM) [22], artificial neural network (ANN) [23,24], classification trees (CTs) [25] and *k*-nearest neighbours (*k*-NNs) [26,27], provide an alternative to LDA method to develop models with better predictive capabilities. Therefore, such techniques are known to enable the analysis of more complex, non-linear relationships and, in general, offer greater performance than their statistical forebears [28].

On the other hand, our research group has developed a novel scheme to generate molecular descriptors, based on the application of discrete mathematics and linear algebra theory. This approach has been successfully applied to the prediction of several physical, physicochemical, chemical, pharmacokinetical, toxicological, as well as biological properties [29–38]. Bearing in mind that mentioned above, and in response to increased performance demands in modelling MOAs, the main aim of the present report was to make use of different supervised ML techniques to perform QSAR models for improving the prediction of toxic modes of action of phenols from molecular structures.

Materials and methods

Dataset

The MOA dataset of Aptula et al. [12] was identified as suitable for developing classification models. The database contained 153 polar narcotic (MOA-1), 18 weak acid respiratory uncoupling (MOA-2), 27 pro-electrophilic (MOA-3), and 23 soft-electrophilic (MOA-4) phenols

derivatives. The whole dataset contains 221 phenols classified into four MOAs, which were assigned following structural rules developed earlier using growth inhibition assays with *Tetrahymena pyriformis* [39].

Mechanisms of toxic action (MOA) of the 221 phenols

Distinct substituent features of the phenolic ring explained a variety of MOAs present in this interesting chemical class. The phenols classified as weak acid respiratory uncoupling are those dependent on the system pH, which undergo heterolytic dissociation, and, in the case of eukaryotic cells, impair the pH and electrochemical gradient across the inner mitochondrial membrane preventing ATP final formation. Electrofilic modes of actions include soft electrophiles and pro-electrophiles; their toxicity can be attributed to the alkylation of essential protein thiol or amino groups, or to oxidative stress produced by free radical formation, and in the second case covalent interactions with electron-rich sites take place after initial biotransformation to more reactive form [10]. An important MOA by means of the one which phenols act as toxicants to aquatic organisms is polar narcosis. It is an unspecific membrane irritation caused by non-covalent van der Waals type interactions of xenobiotics accumulated in lipid tissues [16]. The most important difference between electrophilic mechanisms and polar narcosis lies in the greater specificity for the covalent bonding to high electronic density sites for the first ones, contrary to the non-specificity of the non-covalent interactions with lipid components in the case of the latter [31].

Molecular descriptors

For this study, we employed the Chemistry Development Kit (CDK https://sourceforge.net) [40] to calculate the molecular descriptors. After that, those MDs with values constant or near to constant were deleted and not used in further analysis. Consequently, we finally used 157 descriptors to perform the attribute selection, which permits us to obtain the ML-based models.

Machine learning approaches

The 'No-free-lunch' theorem [41] demonstrates that it is not possible to find an algorithm being better in behaviour for any problem. In several cases, a selection of descriptors is the only essential condition for developing a general system. The next step involves defining the best computational method to develop robust QSAR models. The present report deals with some of the most common classification techniques and how they behave for prediction of the MOA of phenol derivatives. Four classification algorithms were applied: support vector machines using sequential minimal optimization (SVM-SMO) [42], multilayer perceptron (MLP) [23,43], tree classification derived from artificial intelligence (J48) [44] and instance-based algorithm (IBk) [27,45]. The models were developed using Waikato Environment for Knowledge Analysis (WEKA), version 3.6 [46].

Attribute selection

One of the problems of ML process is to select attributes from a large list of candidates to describe the data. For instance, not all of the 157 selected CDK molecular descriptors are needed for representing features of the depiction of MOAs. Usually the addition of irrelevant

or distracting attributes to a dataset 'confuses' the system [46]. Attribute selection is normally done by searching the space of attribute sub-set most appropriate for the prediction. In the case of non-linear ML, using supervised techniques generally can be filter or wrapper method(s) [47]. In the present report, the scheme of attribute selection wrapper implemented in WEKA was applied to select the attribute sub-sets for each ML technique. It evaluates each sub-set using the ML algorithm that will ultimately be employed for learning. It means the learning algorithm is wrapped into the selection procedure. At each step, this option includes in the model set the feature whose addition to the model results in the smallest error (computed as the error rate) and continues until the specified stopping condition is met.

Performance criteria

There are many approaches to evaluate the performance of classification models. In a well-known report [48] a unified overview of methods that are currently used was given, as well as the advantages and disadvantages of each approach. In general, parameters derived from a confusion (contingency) matrix of the actual vs predicted class are one of the most used constants. In the present report, as performance criteria, we have selected the accuracy (Ac) (global good classification), that is the rate of total number of predictions that were corrected, and the rate of false positives (also called false alarm rate), that was estimated as the ratio of negative incorrectly classified instances [46].

Another way to evaluate the performance of a classifier is by the Receiver Operator Curve (ROC) [49–51]. An ROC graph is a technique for visualizing, organizing, and selecting classifiers based on their performance. The ROC graphs are two-dimensional graphs in which true positive rate (tp_{rate}, also called hit rate and recall) is plotted on the *y*-axis, and fp_{rate} is plotted on the *x*-axis by means of the variation of decision threshold. An ROC graph depicts relative trade-off between benefits true positive (TP) and costs false positive (FP) [51]. An indicator of the quality of the classifier is the area under the ROC curve, abbreviated AURC [51–54]. Since the AURC is a portion of the area of the unit square, its value will always be between 0 and 1. The AURC has an important statistical property: the AURC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [54]. The closer this area is to 1, the closer the behaviour of the classifier as a perfect classification model (100% of tp_{rate}, with 0% of fp_{rate}).

Model validation procedure

Internal validation of ML models

In the development of each model, a 10-fold cross-validation (CV) procedure was used. In this procedure, compounds of the training set (TS) were randomly divided into 10 sub-sets. Nine sub-sets were used as novel 'TS' to develop a ML model, and the holdout set was used to 'predict' the performance of the fitted model. This process is repeated 10 times on different 'TSs', so that, in the end, every instance has been used exactly once for testing. This procedure is used to perform the selection of parameters and to avoid over-fitting.

External test set

Once the final models based on the TS of 221 phenols were established, an additional dataset, extracted from a publication using the same MOA assignment rules based on *T. pyriformis*

assays, was used to test the predictive capability of models [55]. After the removal of duplicates and compounds with another MOA (pro-redox cycler) not covered by the TS, a prediction set (PS) of 21 compounds was obtained for this goal. This set consisted of 16 polar narcotics, one respiratory uncoupler and four soft electrophiles; the PS is never used in the development of the model, but it is used to test the predictive power of the final model. In order to assess the applicability domain (AD) of the models, we use Ambit-Discovery software [56,57].

Results and discussion

Development of the classification models to predict the MOAs

The ML approach consists of programming computers to optimize a performance criterion by using example data or past experience. The optimized criterion can be the accuracy provided by a predictive model and the value of a fitness or evaluation function in an optimization problem. Therefore, such techniques are known to enable the analysis of more complex, non-linear relationships [28], for example; Schultz and Netzeva [58] have declared that toxicity is intrinsically a non-linear phenomenon. In the present work, the results for each ML technique used to develop various models to predict MOAs are given. Here, only the best model for each ML approach is displayed. The performance of the classifiers for the TS is provided in Table 1 and the performance for the 10-fold CV study is given in Table 2 .

J48

In the present study, a decision tree for classification of phenols into the four MOAs is inducted by the algorithm C 4.5 (J48). The C 4.5 is an algorithm used to generate a decision tree developed by Quinlan [44], an extension of Quinlan's earlier ID3 algorithm. The basic methodology of divide-and-conquer described in CART [59] is also used in C 4.5. The differences are in the tree structure, the splitting criteria, the pruning method, and the way missing values are handled [44]. Finally, a J48 tree was developed by Weka, with nine attributes. The minimum number of instances per leaf was four. The tree contained 21 nodes in total; 11 of them were terminal ones. This model correctly classified 95.93% of the compounds in the TS, with an AURC of 0.978 and fp_{rate} of 8.2%. In the modelling of individual MOA classes (see Table 1), polar narcotics achieved the best results using all data as TS (Ac of 100%). In the CV

Table 1. Performance of QSAR classifiers using all data (221 compounds) as TS.

	MOA 1 {153}[b]			*MOA 2 {18}*[b]			*MOA 3 {27}*[b]			*MOA 4 {23}*[b]		
Method[a]	fp_{rate}*	*AURC*	*Ac**	fp_{rate}*	*AURC*	*Ac**	fp_{rate}*	*AURC*	*Ac**	fp_{rate}*	*AURC*	*Ac**
J48 (9)	11.8	0.975	100.0	0.5	0.989	94.4	0.0	0.973	81.5	0.0	0.997	87.0
IBk (8)	8.8	0.999	100.0	0.5	1.000	94.4	0.0	0.998	81.5	0.0	1.000	95.7
MLP (8)	1.5	0.985	98.0	0.0	1.000	100.0	1.5	0.954	92.6	0.5	0.996	100.0
SVM-SMO (9)	2.9	0.981	98.7	0.5	0.998	100.0	1.0	0.969	92.6	0.0	0.998	95.7
LDA[c] (5)	nr	nr	93.5	nr	nr	77.8	nr	nr	77.8	nr	nr	82.6
LDA[c] (6)	nr	nr	94.1	nr	nr	66.7	nr	nr	81.5	nr	nr	82.6

[a]Between parentheses is the number of attributes in the model, J48, IBk, MLP, SVM-SMO, and LDA.
[b]Between curly brackets numbers of compounds for each MOA.
[c,d]Models perform on five (log K_{ow}, pK_a, E_{LUMO}, E_{HOMO}, N_{Hdon}) and six (log K_{ow}, pK_a, E_{LUMO}, E_{HOMO}, N_{Hdon}, N_{Hacc}) molecular descriptors by Aptula et al. [12].
*All values are expressed as a percentage (%).; nr, not reported.

experiment, the model predicted properly an average of 90.50% of the chemicals with an fp_{rate} of 11.7% (for other details see Table 2).

IBk

Another model was developed by instance-based learning (IBL) algorithms; IBL algorithms are derived from the nearest neighbour pattern classifier [26]. They are highly similar to edited nearest-neighbour algorithms [60–62], which also save and use only selected instances to generate classification predictions. The best sub-set of attributes contained eight selected molecular descriptors. Euclidean distance was used for finding nearest neighbours. No distance weighting was applied. The optimal number of nearest neighbours was determined by 'trial and error' test and achieved a value of two. This model for the TS showed an accuracy of 96.83%, an AURC of 0.999, and a fp_{rate} of 6.1%. However, in CV test an average of 88.23% of the chemicals was appropriately predicted, this was the lowest value of all ML models (see Table 2).

MLP

The ANN model was obtained with the sigmoid function as an activation function. The number of hidden nodes was selected by 'trial and error' strategy, ranking the results of performance. The ANN with the configuration of 8-6-4 was achieved. The model was performed on eight input variables, an average of 97.74% of the compounds were correctly classified, with an AURC of 0.984 and fp_{rate} of 1.3%. Between the different MOAs, uncouplers and soft electrophiles achieved the best results using all data as TS (Ac: 100%) (see more details in Table 1). This model achieved good results in the CV test, properly predicted an average of 94.57% of the chemicals with only a fp_{rate} of 3.6% (see Table 2).

SVM-SMO

All kernels implemented in WEKA were tested. For each kernel, different parameters were examined by 'trial and error' strategy, ranking the results of performance. The configuration that yields the highest ranking was selected. After all, we found the radial basis function (RBF) kernel produced the best results in modelling MOA. The model was based on nine independent variables. For RBF kernel, we ran the experiments with gamma = (1–4) and ranked the obtained results. Finally, gamma = 3.8 that ended the RBF kernel in $K(x_i, x_j) = e^{(-3.8|x_i - x_j|^2)}$ yielded the highest ranking with a C value of 20. For this model, the value of accuracy was 97.74%, with an AURC of 0.982 and fp_{rate} of 2.2%. The uncouplers and

Table 2. Performance of the ML-based QSAR classifiers in the 10-fold-cross-validation.

	MOA 1 {153}[b]			*MOA 2 {18}*[b]			*MOA 3 {27}*[b]			*MOA 4 {23}*[b]		
Method [a]	fp_{rate}*	*AURC*	*Ac**	fp_{rate}*	*AURC*	*Ac**	fp_{rate}*	*AURC*	*Ac**	fp_{rate}*	*AURC*	*Ac**
J48 (9)	16.2	0.940	96.7	0.0	0.957	77.8	0.0	0.891	63.0	5.1	0.932	91.3
IBk (8)	27.9	0.957	96.1	0.5	0.964	77.8	3.1	0.933	44.4	0.0	0.976	95.7
MLP (8)	4.4	0.965	96.1	0.5	0.997	94.4	3.1	0.917	85.2	1.0	0.993	95.7
SVM-SMO (9)	17.6	0.893	95.4	1.5	0.936	83.3	3.1	0.868	70.4	0.0	0.994	87.0

[a]Between parentheses is the number of attributes in the model, J48, IBk, MLP, SVM-SMO, and LDA.
[b]Between curly brackets numbers of compounds for each MOA.
*All values are expressed as a percentage (%).

polar narcotics achieved the best degree of classification (100% and 98.7%, correspondingly), AURC of 0.998 and 0.981, as well as values of fp_{rate} of 0.5% and 2.9% (see Table 1). Interestingly, in the CV test, 90.50% of the compounds were correctly predicted with an fp_{rate} of 12.7% (more details in Table 2).

Comparison between ML models and with previous reports

First, we developed a comparison between our ML-based models according with their general performance for the TS. The best accuracy values were achieved with MLP and SVM-SMO models (97.74%) followed by IBk model (96.83%) and J48 model with 95.93%; as can be seen the difference between them is lower than 2% and, also, they are rather better than the results of Aptula et al. [12] (89.10%) using LDA-based models12. The parameter AURC for the models shows values between 0.999 (IBk) and 0.982 (SVM-SMO). On the other hand, the fp_{rate} was always lower than 10%; for the models developed with J48 and IBk the achieved values were 8.2% and 6.1%, respectively. A better performance for this parameter was obtained with SVM-SMO (2.2%) and MLP (1.3%) models. As we pointed out above, all the models showed good behaviour in the CV experiment. Unfortunately, these other parameters were not reported in the previous work.

Here, taking into account the correct assignment of each MOA per compound, a comparative analysis for the most significant differences between our ML-based models and the two LDA models is carried out. As can be appreciated in Table 1, the ML techniques presented a very good performance. The higher accuracies for polar narcotics were obtained using J48 and IBk with 100%, followed by the other ML models (over 98%) and LDA models with 93.5% and 94.1%. The correct identification of uncouplers by ML-models (all above 94.4%) is striking with respect to Aptula et al.'s [12] results using LDA (under 78%). Although the predictions for the pro-electrophiles group were identical for two of our models (J48 and IBk) and the best LDA model 81.5%; the models developed with MLP and SVM-SMO showed slightly better values of 92.6%. The last MOA (soft electrophiles) was also well predicted, the MLP model shows a 100% accuracy followed by IBk and SVM-SMO with 95.7%, J48 with 87% and, finally, both LDA models with 82.6%. The best performance of the present study corresponds to MLP and SVM-SMO models when all the classes, despite the severely skewed distribution of them, achieved accuracies over 92.6% and a fp_{rate} below 1.5% (except SVM-SMO for polar narcotics with fp_{rate} = 2.9%). In the case of MLP and SVM-SMO in general, the difference in predictive performance with respect to J48 and IBk is rather greater, with a better overall performance for the pro-electrophiles and very similar for polar narcotics. The ML-based models, for the four groups, show better accuracies than Aptula et al.'s [12] LDA-models, with the exception of pro-electrophiles for J48 and IBk (idem).

We have discussed and demonstrated that, by making use of several ML techniques, it is possible to construct models with better predictive capabilities compared with the models previously obtained by Aptula et al. [12]. As a result, we concluded that ML-based showed better performance in discriminating between the four MOAs.

Now, although this topic is not an objective of our work, we are going to give a brief and general explanation about the descriptors used to develop the ML-based models. In Table 3 you can see the names of the descriptors that were used in each model, following the classification proposed by the CDK program. The direct interpretation of ML-based models is not an easy task, mainly due to the fact that sometimes they do n'ot give an equation or

Table 3. Descriptors used in the development of ML-based models.

Classification[a]	*J48*	IBk	*MLP*	*SVM-SMO*
Constitutionals	nAtomP			ALogP
topologicals	ATSc4	WTPT-3	VP-0	ATSm4
	SP-4	WTPT-4	WTPT-3	WTPT-5
	WTPT-5	WTPT-5	SCH-7	Kier3
	Kier2	C2SP2	VC-3	MDEO-11
	WPOL	C3SP2	MDEC-33	SPC-4
	fragC			
Electronics	FPSA-1	DPSA-3 nHBDon	FPSA-3RNCG	FNSA-1nHBDon
Hybrids	BCUTw-1l			BCUTp-1l
Geometricals		GRAV-1	GRAV-6	

[a]Classification given by CDK programme.

because they take into account non-linear relationships that make this issue difficult. However, we decided to give our considerations related to the information that the different types of descriptors give from molecular structure.

The problems that we try to solve are to classify the phenol derivatives in one of the four MOAs; taking a closer look at these we can see that they are driven for different structural characteristics. For instance, taking the two more different MOAs among them as examples, the MOA-1 (polar narcotics) should be well described by descriptors related with hydrophobicity, molecular size, structural features (like functional groups), among others. On the other hand, to describe respiratory uncouplers, MOA-2 descriptors capable of obtain the electronic characteristics (like charge, polar surface area, etc.) of the molecule will be more useful.

As can be seen in Table 3, all ML models contains descriptors classified as electronics (FPSA-1, NHBDon, RNCG, etc.), but also contains topological descriptors related with electronic characteristics of the compounds like, for example, autocorrelation descriptor of charge (ATSc4) and Wiener polarity number (WPOL). Other descriptors that describe molecular characteristics are also present in the ML models, for example, ALogP to describe hydrophobicity as well as related topological descriptor such as: fragC, a descriptor of fragment complexity and other descriptors of chain or path, like WTPT-x, CxSPx, VP-0, SPC-4, etc. Others important structural features are the shape, size and spatial distribution of the atoms, among them the geometrical descriptors. GRAV-x encodes this information (present in the models developed with IBk and MLP); also several topological descriptors give information about these features, for instance some Kappa shape indices (Kier), the molecular distance edge descriptors (MDEC-33 and MDEO-11), as well as the hybrid indices based in Burden eigenvalues (BCUT), which not only uses a 2D approach, but also a 3D approach, to account for geometric and inter-atomic distances. As we pointed out, each ML-based model uses several molecular descriptors that codify different structural features of the molecules, which are important to describe the four MOAs of phenolic compounds.

Testing the predictive power of classification models using an external test set

According to Golbraikh and Tropsha [63], the only way to establish a reliable model is by means of external validation with some data that did not take part in the training, and, hence, it can measure the prediction ability and check the chance correlation. In this sub-section, we compare the behaviour of ML-based QSAR models obtained in the present report. The predictions of all ML models in the external test set were used for this purpose. In this work,

Table 4. Testing the predictive power of classification models.

	MOA 1 (16)			MOA 2 (1)			MOA 4 (4)		
Method	$fp_{rate\text{-}P}$*	$AURC_P$	Ac_P*	$fp_{rate\text{-}P}$*	$AURC_P$	Ac_P*	$fp_{rate\text{-}P}$*	$AURC_P$	Ac_P*
J48	0.0	1.000	100.0	5.0	0.950	100.0	0.0	0.853	0.75
IBk	0.0	0.969	93.8	0.0	1.000	100.0	5.9	0.963	0.75
MLP	0.0	1.000	87.5	0.0	1.000	100.0	5.9	1.000	100.0
SVM-SMO	0.0	1.000	93.8	0.0	1.000	100.0	5.9	0.971	100.0

*All values are expressed as a percentage (%).; Between parentheses is the number of compounds for each MOA.

Table 5. MOA prediction results (using the best models trained acting all data as TS) employing an external test set.

		MOA[b]				
No	*Name*	*MOA*[a]	*J48*	*IBk*	*MLP*	*SVM-SMO*
1	2-Ethylhexyl-4-hydroxybenzoate	1	1	1	1	1
2	4-Propyloxyphenol	1	1	1	1	1
3	5-Methyl-2-nitrophenol	1	1	4	4	4
4	2,2′,4,4′-Tetrahydroxybenzophenone	1	1	1	3	1
5	3-Hydroxydiphenylamine	1	1	1	1	1
6	Benzyl-4-hydroxyphenyl ketone	1	1	1	1	1
7	2-Hydroxybenzophenone	1	1	1	1	1
8	2-Hydroxydiphenylmethane	1	1	1	1	1
9	Butyl-4-hydroxybenzoate	1	1	1	1	1
10	*n*-Pentyloxyphenol	1	1	1	1	1
11	Isoamyl-4-hydroxybenzoate	1	1	1	1	1
12	4-Cyclohexylphenol	1	1	1	1	1
13	4-(4-Bromophenyl)phenol	1	1	1	1	1
14	Nonyl-4-hydroxybenzoate	1	1	1	1	1
15	4-Bromo-2-fluoro-6-nitrophenol	2	2	2	2	2
16	2-Bromo-2′-hydroxy-5′-nitroacetanilide	4	2	2	4	4
17	2-Fluoro-4-nitrophenol	4	4	4	4	4
18	4-Fluoro-2-nitrophenol	4	4	4	4	4
19	4-Bromo-2-nitrophenol	4	4	4	4	4
20	4-(4-Hydroxyphenyl)-2-butanone	1	1	1	1	1
21	Benzyl-4-hydroxybenzoate	1	1	1	1	1

[a]MOA as reported Cronin et al. [55].
[b]MOA predicted by the best model of each technique.

in addition to the learning set used to fit the models, an external test set selected from a previous report by Cronin et al. [64], in which the assigned mechanism of action was performed according to Aptula et al. [12], was used to validate our models. We also remove duplicates and those compounds with a MOA (pro-redox cycler) not covered by the TS of 221 phenol derivatives. We performed an assessment of the applicability domain and found that all compounds of the test set were inside of the AD of the models.

The models developed with J48 and SVM-SMO showed an accuracy of 95.24% (only one compound misclassified) for the test set, while the models developed with MLP and IBk achieved an accuracy of 90.48% (misclassifying two compounds). Tables 3 and 4 give the performance of models for each MOA and the observed and predicted class for each compound in the external test set, correspondingly. The parameter fp_{rate} was in all cases lower than 6%; according to these results we can say that our four ML-based models have a good predictive power. Notice that three of these four models (IBk, MLP and SVM-SMO) fail to predict the MOA of 5-methyl-2-nitrophenol, which is classified according to the rules as polar

narcotic, but is classified by the models as soft electrophile. In this sense, we can say that this compound has certain characteristics in their structure; for example, it has a nitro group (as soft electrophile) but lacks of halogen substituent [12], so it is classified by the rules as a polar narcotic. Something similar happens with 2-bromo-2'hydroxy-5'nitroacetanilide compound soft electrophile according to assigned MOA by rules, which is wrongly predicted by the models J48 and IBk as a respiratory uncoupler, but which has more than one halogen group [12].

A general inspection of these results demonstrates that all models achieved rather good performance in the external test set, which means that the classification models are reliable. Although a larger database must be evaluated in future works, finally the results evidenced that the models present good predictive capabilities (for more details see Table 5). All these results validate our models, making them useful tools in the prediction of ecotoxicological potential of phenols and related compounds.

Conclusions

Prediction of the mechanism of toxic action of phenols attracts great scientific interest, because the increasing use of these compounds with various intentions makes its ecotoxicity evaluation necessarily. In the prediction, the MOAs of new compounds '*in silico*' approaches emerged to be especially worthwhile in both terms of financial cost and time consumption by exploring the usefulness of QSAR methods. In this report, we made use of the CDK's descriptors and different ML approaches as J48, IBk, MLP and SVM-SMO to find QSAR models that can describe the mechanism of toxic action, classifying the chemicals into polar narcotics, respiratory uncouplers, pro-electrophiles and soft electrophiles. All ML-based QSAR models showed in general very good performances in TS, CV study and the external test set. These ML techniques were demonstrated to be better than LDA models previously developed by Aptula et al. [12]. The contribution of this work in uncouplers and pro-electrophiles correct identification using ML techniques is striking with respect to our and Aptula et al.'s [12] models using LDA. Finally, the contribution of this report is rather encouraging because it represents better results for modelling the mechanism of toxic action of phenols using ML models and could increase the practicality of data mining procedures of chemical databases for the discovery or identification of the mechanism of toxic action of new phenol derivatives.

Acknowledgements

J.A. Castillo-Garit and G.M. Casañola-Martin thank the program 'Estades Temporals per an Investigadors Convidats' for a fellowship to work at Valencia University in 2013. F. Torrens acknowledges support from the Spanish Ministerio de Economía y Competitividad (Project No. BFU2013-41648-P) and EU ERDF.

Disclosure statement

The authors confirm that this article content has no conflicts of interest. All authors are working for public academic institutions.

References

[1] R. Garg, A. Kurup, and C. Hansch, *Comparative QSAR: On the toxicology of the phenolic OH moiety*, Crit. Rev. Toxicol. 31 (2001), pp. 223–245.

[2] S.P. Bradbury, *Predicting modes of toxic action from chemical structure: An overview*, SAR QSAR Environ. Res. 2 (1994), pp. 89–104.

[3] H.H.P. Fang, T. Chen, Y.Y. Li, and H.K. Chui, *Degradation of phenol in wastewater in an upflow anaerobic sludge blanket reactor*, Water Res. 30 (1996), pp. 1353–1360.

[4] N. Caza, J.K. Bewtra, N. Biswas, and K.E. Taylor, *Removal of phenolic compounds from synthetic wastewater using soybean peroxidase*, Water Res. 33 (1999), pp. 3012–3018.

[5] U.G. Ahlborg, and T. Thunberg, *Chlorinated phenols: Occurrence, toxicity, metabolism, and environmental impact*, Crit. Rev. Toxicol. 7 (1980), pp. 1–35.

[6] E. Argese, C. Bettiol, A. Ghelli, R. Todeschini, and P. Miana, *Submitochondrial particles as toxicity biosensors of chlorophenols*, Environ. Toxicol. Chem. 14 (1995), pp. 363–368.

[7] M.P. Moulton and T.W. Schultz, *Comparisons of several structure-toxicity relationship for chlorophenols*, Aquat. Toxicol. 8 (1986), pp. 121–128.

[8] C.M. Auer, J.V. Nabholz, and K.P. Baetcke, *Mode of action and the assessment of chemical hazards in the presence of limited data: Use of structure-activity relationships (SAR) under TSCA, Section 5*, Environ. Health Perspect. 87 (1990), pp. 183–197.

[9] T. Ivanciuc, O. Ivanciuc, and D.J. Klein, *Posetic quantitative superstructure/activity relationships (QSSARs) for chlorobenzenes*, J. Chem. Inf. Model. 45 (2005), pp. 870–879.

[10] S. Spycher, B.I. Escher, and J. Gasteiger, *A quantitative structure–activity relationship model for the intrinsic activity of uncouplers of oxidative phosphorylation*, Chem. Res. Toxicol. 18 (2005), pp. 1858–1867.

[11] C.L. Russom, S.P. Bradbury, S.J. Broderius, D. Hammermeister, and R.A.E. Drummond, *Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow* (Pimephales promelas), Environ. Toxicol. Chem. 16 (1997), pp. 948–967.

[12] A.O. Aptula, T.I. Netzeva, I.V. Valkova, M.T.D. Cronin, and T.W. Schultz, *Multivariate discrimination between modes of toxic action of phenols*, Quant. Struct. Act. Relat 21 (2002), pp. 12–22.

[13] S. Ren, *Determining the mechanisms of toxic action of phenols to* Tetrahymena pyriformis, Environ. Toxicol. 17 (2002), pp. 119–127.

[14] S. Ren, *Use of molecular descriptors in separating phenols by three mechanisms of toxic action*, Quant. Struct.-Act. Relat. 21 (2002), pp. 486–492.

[15] S.P. Bradbury, C.L. Russom, G.T. Ankley, T.W. Schultz, and J.D. Walker, *Overview of data and conceptual approaches for derivation of quantitative structure-activity relationships for ecotoxicological effects of organic chemicals*, Environ. Toxicol. Chem. 22 (2003), pp. 1789–1798.

[16] G. Schuurmann, A.O. Aptula, R. Kuhne, and R.U. Ebert, *Stepwise discrimination between four modes of toxic action of phenols in the* Tetrahymena pyriformis *assay*, Chem. Res. Toxicol. 16 (2003), pp. 974–987.

[17] S. Ren, P.D. Frymier, and T.W. Schultz, *An exploratory study of the use of multivariate techniques to determine mechanisms of toxic action*, Ecotoxicol. Environ. Saf. 55 (2003), pp. 86–97.

[18] S. Ren, and H. Kimand, *Comparative assessment of multiresponse regression methods for predicting the mechanisms of toxic action of phenols*, J. Chem. Inf. Comput. Sci. 43 (2003), pp. 2106–2110.

[19] S. Spycher, E. Pellegrini, and J. Gasteiger, *Use of structure descriptors to discriminate between modes of toxic action of phenols*, J. Chem. Inf. Model. 45 (2005), pp. 200–8.

[20] S.J. Enoch, M.T.D. Cronin, T.W. Schultz, and J.C. Madden, *An evaluation of global QSAR models for the prediction of the toxicity of phenols to* Tetrahymena pyriformis, Chemosphere 71 (2008), pp. 1225–1232.

[21] G. Melagraki, A. Afantitis, H. Sarimveis, O. Igglessi-Markopoulou, and A. Alexandridis, *A novel RBF neural network training methodology to predict toxicity to* Vibrio fischeri, Mol. Divers. 10 (2006), pp. 213–221.

[22] J. Guo, H. Chen, Z. Sun, and Y. Lin, *A novel method for protein secondary structure prediction using dual-layer SVM and profiles*, Proteins 54 (2004), pp. 738–743.

[23] J. Zupan and J. Gasteiger, *Neural Networks for Chemistry and Drug Design*, Wiley-VCH Publishers, Weinheim, Germany, 1999.
[24] S. Anzali, G. Barnickel, M. Krug, J. Sadowski, A. Teckentrup, and M. Wagener, *The use of self-organizing neural networks in drug design*, in *3D QSAR in Drug Design*, H. Kubinyi, G. Folkers and Y.C. Martin, eds., Kluwer/ESCOM, Dordrecht, The Nederlands, 1998, pp. 273–299.
[25] J.R. Quinlan, *Induction of decision trees*, Machine Learn. 1 (1986), pp. 81–106.
[26] T. Cover and P. Hart, *Nearest neighbor pattern classification*, IEEE Trans. Inform. Theory 13 (1967), pp. 21–27.
[27] D. Aha and D. Kibler, *Instance-based learning algorithms*, Machine Learn. 6 (1991), pp. 37–66.
[28] P. Larrañaga, B. Calvo, R. Santana, C. Bielza, J. Galdiano, I. Inza, J.A. Lozano, R. Armañanzas, G. Santafé, A. Pérez, and V. Robles, *Machine learning in bioinformatics*, Brief. Bioinform. 7 (2006), pp. 86–112.
[29] Y. Marrero-Ponce, E. Martínez-Albelo, G. Casañola-Martín, J.A. Castillo-Garit, Y. Echevería-Díaz, V. Zaldivar, J. Tygat, J. Borges, R. García-Domenech, F. Torrens, and F. Pérez-Giménez, *Bond-based linear indices of the non-stochastic and stochastic edge-adjacency matrix. 1. Theory and modeling of ChemPhys properties of organic molecules*, Mol. Div. 14 (2010), pp. 731–753.
[30] J.A. Castillo-Garit, O. Del Toro-Cortés, V.V. Kouznetsov, C.O. Puentes, A.R. Romero Bohórquez, M.C. Vega, M. Rolón, J.A. Escario, A. Gómez-Barrio, Y. Marrero-Ponce, F. Torrens, and C. Abad, *Identification in silico and in vitro of novel trypanosomicidal drug-like compounds*, Chem. Biol. Drug Des. 80 (2012), pp. 38–45.
[31] Y. Brito-Sánchez, J.A. Castillo-Garit, H. Le-Thi-Thu, Y. González-Madariaga, F. Torrens, Y. Marrero-Ponce, and J.E. Rodríguez-Borges, *Comparative study to predict toxic modes of action of phenols from molecular structures*, SAR QSAR Environ. Res. 24 (2013), pp. 235–251.
[32] J.A. Castillo-Garit, O. del Toro-Cortés, M.C. Vega, M. Rolón, A. Rojas de Arias, G.M. Casañola-Martin, J.A. Escario, A. Gómez-Barrio, Y. Marrero-Ponce, F. Torrens, and C. Abad, *Bond-based bilinear indices for computational discovery of novel trypanosomicidal drug-like compounds through virtual screening*, Eur. J. Med. Chem. 96 (2015), pp. 238–244.
[33] G.M. Casanola-Martin, H. Le-Thi-Thu, Y. Marrero-Ponce, J.A. Castillo-Garit, F. Torrens, F. Perez-Gimenez, and C. Abad, *Analysis of proteasome inhibition prediction using atom-based quadratic indices enhanced by machine learning classification techniques*, Lett. Drug Des. Discovery 11 (2014), pp. 705–711.
[34] J.A. Castillo-Garit, Y. Marrero-Ponce, S.J. Barigye, R. Medina-Marrero, M.G. Bernal, J.M.G.d.l. Vega, F. Torrens, V.J. Arán, F. Pérez-Giménez, R. García-Domenech, and R. Acevedo-Barrios, *In silico antibacterial activity modeling based on the TOMOCOMD-CARDD approach*, J. Braz. Chem. Soc. 26 (2015), pp. 1218–1226.
[35] Y. Cañizares-Carmenate, K. Mena-Ulecia, Y. Perera-Sardiña, F. Torrens, and J.A. Castillo-Garit, *An approach to identify new antihypertensive agents using Thermolysin as model: In silico study based on QSARINS and docking*, Arab. J. Chem. Available at https://doi.org/10.1016/j.arabjc.2016.10.003.
[36] J.A. Castillo-Garit, C. Abad, G.M. Casañola-Martin, S.J. Barigye, F. Torrens, and A. Torreblanca, *Prediction of aquatic toxicity of benzene derivatives to* Tetrahymena pyriformis *according to OECD principles*, Curr. Pharm. Des. 22 (2016), pp. 5085–5094.
[37] G.M. Casañola-Martin, H. Le-Thi-Thu, Y. Marrero-Ponce, J.A. Castillo-Garit, F. Torrens, A. Rescigno, C. Abad, and M. Tareq, *Hassan Khan, Tyrosinase enzyme: 1. An overview on a pharmacological target*, Curr. Top. Med. Chem. 14 (2014), pp. 1494–1501.
[38] J.A. Castillo-Garit, G.M. Casañola-Martin, H. Le-Thi-Thu, H. Pham-The, and S.J. Barigye, *A simple method to predict blood-brain barrier permeability of drug- like compounds using classification trees*, Med. Chem. 13 (2017), p.https://doi.org/10.2174/1573406413666170209124302. Available online since October 10, 2016.
[39] T.W. Schultz, *TERATOX:* Tetrahymena pyriformis *population grow impairment endpoint- A surrogate for fish lethality*, Toxicol. Meth. 7 (1997), pp. 289–309.
[40] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, and E. Willighagen, *The Chemistry Development Kit (CDK): An open-source java library for chemo- and bioinformatics*, J. Chem. Inf. Comput. Sci. 43 (2003), pp. 493–500.

[41] D.H. Wolpert and W.G. Macready, *No free lunch theorems for optimization*, IEEE Trans. Evol. Comput. 1 (1997), pp. 67–82.
[42] John C. Platt, *Fast training, of support vector machines using sequential minimal optimization*, in *Advances in Kernel Methods* Bernhard Schölkopf, Christophe J. C. Burges and Alexander J. Smola, eds., MIT Press, Cambridge, MA, USA, 1999, pp. 185–208.
[43] R.P. Lippmann, *An introduction to computing with neural nets*, ACM SIGARCH Comput. Arch. News 16 (1988), pp. 7–25.
[44] R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA, 1993.
[45] K. Vanhoof, *Performance measures*, in *Conceptual Course on Data Mining*, Universidad Central 'Martha Abreu' de Las Villas, Santa Clara, Cuba, 2006, p. 101.
[46] I.H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers, San Francisco, CA, USA, 2005.
[47] G. John, R. Kohavi, and K. Pfleger, *Irrelevant features and the subset selection problem*, in *Machine Learning: Proceeding of Eleventh International Conference* William W. Cohen and Haym Hirsh, Morgan Kaufman, San Francisco, CA, USA, 1994, pp. 121–129.
[48] B. Pierre, S. Brunak, Y. Chauvin, C.A.F. Andersen, and N. Henrick, *Assessing the accuracy of prediction algorithms for classification: An overview*, Bioinformatics 16 (2000), pp. 412–424.
[49] J.W. Shavlik (ed.), *Analysis and Visualization of Classifier Performance: Comparison Under Imprecise Class and Cost Distribution*, AAAI Press, Menlo Park, CA, August 14–17, 1997.
[50] J.W. Shavlik (ed.), *The Case Against Accuracy Estimation for Comparing Induction Algorithms*, Morgan Kaufmann, San Francisco, CA, USA, 1998.
[51] T. Fawcett, *ROC graphs: Notes and practical considerations for researchers*, Machine Learn 31 (2004), pp. 1–38.
[52] C.X. Ling, J. Huang, and H. Zhang, *AUC: A statistically consistent and more discriminating measure than accuracy*, Proceedings of 18th International Conference on Artificial Intelligence (IJCAI-2003), Acapulco, Mexico, August 09–15, (2003), pp. 519–524.
[53] A.P. Bradley, *The use of the area under the ROC curve in the evaluation of machine learning algorithms*, Pattern Recogn. 30 (1997), pp. 1145–1159.
[54] J.A. Hanley and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve*, Radiology 143 (1982), pp. 29–36.
[55] M.T.D. Cronin, A.O. Aptula, J.C. Duffy, T.I. Netzeva, P.H. Rowe, I.V. Valkova, and T.W. Schultz, *Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to Tetrahymena pyriformis*, Chemosphere 49 (2002), pp. 1201–1221.
[56] N. Nikolova-Jeliazkova, *AMBIT: Building blocks for a future (Q)SAR decision support system* (2006). Available at https://ambit.acad.bg.
[57] J. Jaworska, N. Nikolova-Jeliazkova, and T. Aldenberg, *QSAR applicability domain estimation by projection of the training set in descriptor space: A review*, ATLA 33 (2005), pp. 445–459.
[58] T.W. Schultz and T.I. Netzeva, *Development and evaluation of QSARs for ecotoxic endpoints: the benzene response-surface model for* Tetrahymena *toxicity.*, in *Modelling Enviromental Fate and Toxicity*, M.T. Cronin and D. Livingstone, eds., CRC Press, Boca Raton, FL, 2004, pp. 265–284.
[59] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, *Classification and Regression Trees*, Wadsworth, Monterey, CA, 1984.
[60] G. Gates, *The reduced nearest neighbor rule*, IEEE Trans. Inform. Theory 18 (1972), pp. 431–433.
[61] P. Hart, *The condensed nearest neighbor rule*, IEEE Trans. Inform. Theory 14 (1968), pp. 515–516.
[62] B.V. Dasarathy, *Nosing around the neighborhood: A new system structure and classification rule for recognition in partially exposed environments*, Pattern Anal. Machine Intell. 2 (1980), pp. 67–71.
[63] A. Golbraikh, and A. Tropsha, *Beware of q2!*, J. Mol. Graphics Model. 20 (2002), pp. 269–76.
[64] M.T. Cronin, A.O. Aptula, J.C. Duffy, T.I. Netzeva, P.H. Rowe, I.V. Valkova, and T.W. Schultz, *Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to* Tetrahymena pyriformis, Chemosphere 49 (2002), pp. 1201–1221.