

Quantitative Structure-Activity Relationship study for pesticides by means of classification techniques

Estudio de la Relación Cuantitativa Estructura-Actividad de pesticidas mediante técnicas de clasificación

Fernando Cárdenas^{1*}, Piercosimo Tripaldi², Cristian Rojas³

¹Grupo de Investigación en Biotecnología y Ambiente (INBIAM). Escuela de Ingeniería Ambiental. Universidad Politécnica Salesiana. Calle Vieja 12-30 y Elia Liut, Cuenca-Ecuador.

²Laboratorio UDALAB, Facultad de Ciencia y Tecnología, Universidad del Azuay. Av. 24 de Mayo 7-77 y Hernán Malo, Apartado postal 01.01.981. Cuenca-Ecuador.

³Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), Universidad Nacional de la Plata. Diagonal 113 y calle 64, C.C. 16, Suc. 4 (1900), La Plata-Argentina.

*Autor principal/Corresponding author, e-mail: lcardenas@ups.edu.ec

Editado por/Edited by: Cesar Zambrano, Ph.D.

Recibido/Received: 15/09/2014. Aceptado/Accepted: 20/10/2014.

Publicado en línea/Published on Web: 19/12/2014. Impreso/Printed: 19/12/2014.

Abstract

The aim of this work was the comparison between k -Nearest Neighbors (k -NN) and Counterpropagation Artificial Neural network (CP-ANN) classification methods for modeling the toxicity of a set of 192 organochlorinated, organophosphates, carbamates, and pyrethroid pesticides measured as effective concentration (EC50). The EC50 values were divided into three classes, i.e. low, intermediate, and high toxicity. The 4885 molecular descriptors were calculated using the Dragon software, and then were simultaneously analyzed through k -NN classification analysis coupled with Genetic Algorithms - Variable Subset Selection (GA-VSS) technique. The models were properly validated through an external test set of compounds. The results clearly suggest that 3D-descriptors did not offer relevant information for modeling the classes. On the other hand, k -NN showed better results than CP-ANN.

Keywords. Pesticides, k -NN, CP-ANN, GA-VSS, QSAR Theory.

Resumen

El objetivo de este trabajo fue la comparación entre los métodos de clasificación del vecino más cercano (k -NN) y las redes neuronales artificiales de contrapropagación (CP-ANN) para modelar la toxicidad de un conjunto de 192 pesticidas organoclorados, organofosforados, carbamatos y piretroides, medidos como Concentración Efectiva (EC50) y que fueron divididos en tres clases, es decir, baja, intermedia y alta toxicidad. Se calcularon 4885 descriptores moleculares usando el programa DRAGON, los que fueron simultáneamente analizados mediante el método k -NN acoplado con la técnica de selección de variables de los Algoritmos Genéticos (GA-VSS). Los modelos fueron apropiadamente validados mediante un subconjunto de predicción. Los resultados claramente sugieren que los descriptores 3D no ofrecen información relevante para modelar las clases. Por otro lado, k -NN muestra mejores resultados que CP-ANN.

Palabras Clave. Pesticidas, k -NN, CP-ANN, GA-VSS, Teoría QSAR.

Introducción

Los pesticidas son un grupo muy importante de compuestos químicos que son utilizados en la agricultura para la protección contra las plagas. Su aplicación anual estimada es de 4 millones de toneladas pero sólo el 1 % de estos compuestos llegan a ser efectivos en su aplicación, es decir, el 99 % de sus componentes activos son

liberados y estos pueden causar diferentes afectaciones a los componentes ambientales y a los seres vivos. Varias investigaciones han estudiado la toxicidad de los pesticidas y se ha demostrado que los mismos pueden ser peligrosos si se usan indiscriminadamente. La toxicidad de los pesticidas depende en gran medida de la estructura química que poseen, metabolismo de los organismos, dosis, grado de descomposición, modo de en-

trada al organismo, entre otros; razón por la cual es de suma importancia determinar su toxicidad [1].

Para este fin se han estudiado y desarrollado diversos ensayos en organismos vivos. La ventaja fundamental de estas pruebas es su realización en condiciones rigurosas y controladas de laboratorio, permitiendo establecer de manera cuantitativa la relación entre la exposición (dosis) y su efecto (respuesta) [2]. Dentro de las desventajas de realizar estas pruebas es la detección de los efectos en dosis bajas; esto requiere de un número elevado de ejemplares lo que convierte al proceso inviable económicamente [3]. En un estudio de la Comisión Europea se estimó que las pruebas toxicológicas de 30.000 sustancias constituyeron un costo total de 2 billones de euros hasta el año 2013 [4]; asimismo, se ha estimado que para realizar estas pruebas fueron necesarios el uso de más de 12 millones de animales vertebrados que fueron sacrificados para este propósito [3]. Esto claramente indica que existen varias limitaciones en los ensayos de toxicidad como: restricciones económicas, consideraciones éticas y la presión pública para reducir las pruebas toxicológicas en animales [4].

Debido a estas limitaciones, se han desarrollado nuevos métodos que permiten reducir el número de animales usados en la experimentación, ahorrar tiempo y reducir el costo de los ensayos [5]. Dentro de estas estrategias se encuentran los modelos matemáticos de predicción basados en la teoría QSAR (Relación Cuantitativa Estructura-Actividad) [6]. Estos modelos fueron descritos inicialmente por Hansch y su equipo en 1964 [7]. Ellos establecieron que moléculas semejantes tendrán un comportamiento similar tanto en las propiedades fisicoquímicas como biológicas. Por lo tanto, estas sustancias pueden inducir una determinada respuesta en el organismo, lo cual indica el efecto y la función característica de las mismas. El postulado de Hansch afirma que la actividad está en función de propiedades lipofílicas (L), propiedades electrónicas (E), propiedades estéricas (S) y, finalmente cualquier otro tipo de propiedad molecular (M) necesaria para describir el efecto biológico estudiado. Estos modelos son basados en una serie de aplicaciones matemáticas y estadísticas que tienen como premisa encontrar un modelo matemático cuantitativo que relacione la estructura molecular y una actividad definida. Para lograr este objetivo los modelos QSAR se apoyan en técnicas quimiométricas que permiten analizar de manera exhaustiva los datos [8].

La Quimiometría es una ciencia que utiliza técnicas matemáticas y estadísticas multivariantes sobre datos químicos, permitiendo obtener la máxima información para entender un problema [9]. Estas técnicas analizan sistemas complejos de información, los cuales están formados por matrices compuestas por varias partes interconectadas y entrelazadas cuyos vínculos crean información adicional no visible antes del análisis. Con los avances de la informática se han podido desarrollar técnicas que permiten utilizar toda la información disponible en las bases de datos; minimizando el tiempo y el

costo del análisis y maximizando la calidad de la respuesta. Esto permite tener un mejor entendimiento del problema al estudiar todos los datos disponibles usando técnicas tales como análisis de componentes principales, análisis de conglomerados, métodos de regresión, métodos de clasificación, métodos de selección de variables, redes neuronales, diseño experimental, entre otros [8].

Las técnicas de clasificación [8, 9] tienen como objetivo la construcción de modelos discretos sobre la base de un cierto número de variables independientes y una clase como respuesta cualitativa. Existe una vasta gama de métodos de clasificación, entre los cuales se tiene el método del vecino más cercano k -NN [8] y las redes neuronales artificiales de contrapropagación (CP-ANN) [10, 11]. Debido a la dimensionalidad de las bases de datos es necesario realizar una selección previa de las variables (descriptores) más significativas que permitirán predecir la toxicidad de los pesticidas. Para obtener un conjunto óptimo de variables necesarias para la construcción del modelo se usan diversas técnicas, entre ellas, la metodología de selección de variables de los algoritmos genéticos (GA-VSS) [12–14], esta metodología se basa en el proceso genético evolutivo de los organismos vivos y simula el proceso de la selección natural y la supervivencia.

Duchowicz y Castro [6] presentaron un artículo de revisión de las diversas aplicaciones de la teoría QSAR/QSPR en el estudio de pesticidas desarrolladas durante la década 2003-2013. En ella mostraron estudios de propiedades tales como Factor de Bioconcentración, Coeficiente de Adsorción Suelos/Sedimentos, Toxicidad, Solubilidad Acuosa y Partición Aire a Agua; constituyendo un resumen detallado de estudios específicos en este campo. Dadas estas premisas, el objetivo principal fue el desarrollar un modelo QSAR para modelar la Concentración Efectiva 50 (EC₅₀) medida en *Daphnia magna*, usando un conjunto de 192 pesticidas. Se aplicó el método de selección de variables GA-VSS acoplado con k -NN. Finalmente se compararon las técnicas de clasificación k -NN y CP-ANN. Estos modelos permitirán predecir la actividad biológica de nuevas moléculas en términos discretos, permitiendo ahorrar tiempo, recursos económicos y reducir la experimentación de laboratorio sobre animales.

Materiales y Métodos

Base de datos experimental

El dominio químico analizado involucra 192 moléculas de pesticidas de cuatro tipos: carbamatos, piretroides organoclorados y organofosforados, reportados en Pesticide Properties Data Base (PPDB) [15, 16]. La actividad experimental reportada es la Concentración Efectiva 50 (EC₅₀) medida en *Daphnia magna*, la cual representa la concentración en mg/L que producirá la reducción del 50 % de la población en estudio luego de 48 horas [2].

Parámetro	Clase 1	Clase 2	Clase 3
Valor umbral	5,3-4,65mg/L	<4,65-0,10mg/L	<0,10-0,000001mg/L
Toxicidad	Baja	Mediana	Alta
Número de objetos	64	65	63

Tabla 1: Intervalo de valores de EC_{50} para la construcción de las clases y número de pesticidas por cada clase.

Cálculo de los descriptores moleculares

Un problema crucial en los estudios QSAR es el encontrar una representación estructural conveniente y por este motivo se usan los descriptores moleculares como una caracterización estructural. Los descriptores moleculares son el resultado de un procedimiento lógico y matemático que transforma la información química codificada dentro de una representación simbólica de una molécula en un número útil o el resultado de algún experimento estandarizado [17]. Para tal efecto, las estructuras moleculares fueron diseñadas y optimizadas en el programa Hyperchem [18] por el método de los campos de fuerza de la mecánica molecular (MM+), seguido por el método semiempírico PM3. Se usó el algoritmo del gradiente conjugado en la versión Polak-Ribiere hasta que la desviación cuadrática media sea menor que $0.01\text{kcal.}(\text{\AA}\cdot\text{mol})^{-1}$. Seguidamente, se calcularon 4885 descriptores moleculares usando el programa DRAGON versión 6.0 [19]. Esta base de datos de descriptores incluyen 29 familias: 0D- descriptors (constitutional indices), 1D- descriptors (functional group counts, atom-centred fragments, molecular properties), 2D- descriptors (ring descriptors, topological indices, walk and path counts, connectivity indices, information indices, 2D matrix-based descriptors, 2D autocorrelations, Burden eigenvalues, P_VSA-like descriptors, edge adjacency indices, CATS 2D, 2D atom pairs, atom-type E-state indices, ETA indices, drug-like indices), and 3D-descriptors (Randic molecular profiles, geometrical descriptors, RDF descriptors, 3D- MoRSE descriptors, WHIM descriptors, GETAWAY descriptors, charge descriptors, 3D matrix-based descriptors, 3D autocorrelations, 3D atom pairs). Se excluyeron descriptores con valores constantes y casi constantes, además de los que están correlacionados al 95 %. De esta manera se obtuvieron 1126 descriptores que nos permitieron relacionar la estructura con una actividad biológica conocida que en nuestro estudio es la toxicidad EC_{50} .

Asignación de clases

Para evitar errores en la modelización de la respuesta debidos al rango y la variabilidad de los datos dentro del mismo; los valores de EC_{50} fueron normalizados realizando la siguiente transformación:

$$EC_{50}^{nor} = \log \frac{1}{EC_{50}} \quad (1)$$

Posteriormente, los valores de EC_{50}^{nor} obtenidos se dividieron en tres clases. Para ello se calcularon los percentiles 33 y 66 para representar los valores de frontera

para las clases, es decir, clase 1 o toxicidad baja (0-33 % de los datos), toxicidad media o clase 2 (34-66 % de los datos) y clase 3 o toxicidad alta (67-100 % de los datos). Las características de cada una de las clases se presentan en la Tabla 1.

Desarrollo de los Modelos

Selección de los descriptores en clasificación

Otro de los temas a considerar en la teoría QSAR es la selección de los descriptores moleculares más útiles a partir de un gran conjunto de datos experimentales. Existen varias metodologías de selección de variables, dentro de las cuales GA-VSS han sido ampliamente utilizados en Quimiometría y QSAR [13, 20–40]. En resumen, GA-VSS consisten en un ciclo iterativo de creación de conjuntos (cromosomas) de variables originales que generan poblaciones viables las cuales son las más óptimas para la construcción del modelo final. Los cálculos GA-VSS se realizaron usando el GA toolbox para MatLab [41] provisto por el Milano Chemometrics and QSAR Research Group (University of Milano - Bicocca).

Método de clasificación del vecino más cercano (*k*-Nearest Neighbors)

k-NN es un método de clasificación no paramétrico que para realizar la clasificación se basa sobre el concepto de analogía. El método se fundamenta en la elección de una distancia (generalmente la Euclidea) y en la selección oportuna de los *k* objetos más cercanos a la molécula en estudio [42]. Luego se enumera cuántos de estos *k* objetos caen en cada clase y la molécula es atribuida a la clase más numerosa; esta evaluación se extiende a todos los objetos estudiados hasta determinar a qué clase corresponde cada uno de ellos y en ese momento se termina el algoritmo de clasificación. *k*-NN no provee un modelo matemático, es decir, una función analítica aplicable sucesivamente a los objetos desconocidos a ser clasificados [9]. En la práctica es necesario probar diversos valores de *k* que oscilan entre 3 y 10 hasta obtener el valor de clasificación más óptimo. Este es un algoritmo que posee una gran simplicidad y por lo tanto ha sido ampliamente aplicado a diversos estudios de clasificación debido a su rendimiento en la respuesta obtenida [42].

Método de clasificación de las Redes Neuronales Artificiales de Contrapropagación (Counterpropagation Artificial Neural network)

CP-ANNs son uno de los métodos más populares de las Redes Neuronales Artificiales [10, 11]. Ellas imitan la acción de una red biológica de neuronas donde cada

Descriptor	Descripción	Bloque	Sub-bloque
B01[C-N]	Presence/absence of C - N at topological distance 1		Binary atom pairs of order 1
B02[C-N]	Presence/absence of C - N at topological distance 2		Binary atom pairs of order 2
B04[C-P]	Presence/absence of C - P at topological distance 4	2D atom pairs	Binary atom pairs of order 4
B04[C-N]	Presence/absence of C - N at topological distance 4		Binary atom pairs of order 4
F02[C-N]	Frequency of C - N at topological distance 2		Frequency atom pairs of order 2
O-059	Al-O-Al	Atom-centred fragments	Basic descriptors
CATS2D_02_DA	Donor-acceptor at lag 02	CATS 2D	Basic descriptors
CATS2D_02_DL	CATS2D Donor-lipophilic at lag 02		
nP	Number of phosphorous atoms	Constitutional indices	Basic descriptors
SssNH	Sum of ssNH E-states atom-type	E-state indices	E-State sums
GATS1s	Geary autocorrelation of lag 1 weighted by I-state	2D Autocorrelations	Geary autocorrelations

Tabla 2: Descripción de los descriptores calculados por Dragon considerados en los modelos k -NN y CP-ANN.

neurona acepta diferentes señales de las neuronas vecinas y las procesa para un número de iteraciones definidas. Se las puede considerar una extensión de los mapas de Kohonen, también llamado Mapas de Auto-Organización (SOMs). CP-ANNs consiste en dos capas, una capa Kohonen y una capa de salida (también denominada capa de Grosberg). La primera capa es usualmente caracterizada por ser un espacio toroidal cuadrado o hexagonal, que consiste en una cuadrícula de $N \times N$ neuronas; donde N es el número de neuronas para cada lado del espacio. Cada neurona contiene tantos elementos (pesos) como el número de variables de entrada. Los pesos de cada neurona son aleatoriamente inicializados entre 0 y 1 y actualizados en función de los vectores de entrada, es decir, las muestras para un determinado número de tiempo (denominado épocas). Ambos parámetros deben ser definidos por el usuario [39]. El uso de CP-ANNs se ha incrementado en diferentes aspectos de la química y actualmente pueden ser consideradas una herramienta importante dentro del análisis multivariable. Para los cálculos, se usó utilizó el Kohonen and CPANN Toolbox [43, 44] también implementado en MatLab.

Validación del modelo

Para propósitos de validación externa, la base de datos de 192 pesticidas y 1126 descriptores fue dividida en dos grupos de manera aleatoria, considerando el 70 % de las moléculas para la calibración (training set) y las restantes para la predicción (test set). El parámetro que sintetiza el resultado de un método de clasificación es la tasa de no error (non-error rate, $NER\%$) definido de la siguiente manera:

$$NER = \frac{\sum_g c_{gg}}{n} \quad (2)$$

Donde c_{gg} son los elementos diagonales de la matriz de confusión y n es el número total de datos. Los valores de NER tienen un intervalo que van de 0 a 1 o de 0 a 100 ($NER\%$), donde valores cercanos a 1 o 100, respectivamente, representan el mejor resultado de clasificación [9]. Se usó el subconjunto de calibración para realizar la validación cruzada extrayendo aleatoriamente el 20 % de moléculas (leave-more-out) y obtener el NER_{cv} ; mientras que, con el subconjunto de predicción se ha obtenido el NER_{pred} .

Resultados y Discusión

En la Tabla 1S (material suplementario) se indica la información de los pesticidas (tipo y nombre), su valor de toxicidad, las clases asignadas de acuerdo a los valores presentados en la Tabla 1. Se indican también las cla-

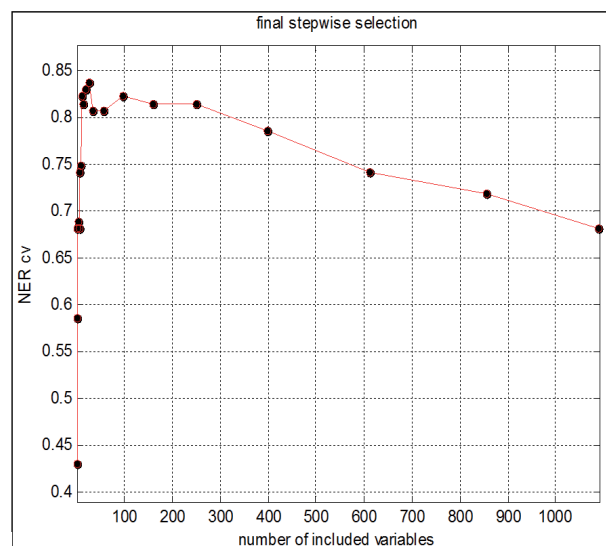


Figura 1: Relación entre el número de descriptores incluidos en los modelos y la tasa de no error en validación cruzada (NER_{cv}).

Clases verdaderas	Clases Asignadas <i>k</i> -NN			Clases Asignadas CP-ANN		
	Clase 1	Clase 2	Clase 3	Clase 1	Clase 2	Clase 3
Clase 1	33	3	3	30	6	3
Clase 2	4	36	8	5	37	6
Clase 3	2	3	43	3	5	40

Tabla 3: Matrices de confusión para los métodos *k*-NN y CP-ANN obtenidos a partir del subgrupo de calibración.

ses asignadas por los dos métodos de clasificación, las cuales se usan para el cálculo del NER. La partición de la base de datos fue aleatoria, manteniendo el 70 % de moléculas en el subgrupo de calibración y las restantes en el de predicción, las cuales serán usadas para la validación externa del modelo.

Sobre el subconjunto de calibración, se aplicó el algoritmo *k*-NN acoplado con la técnica de selección de variables de los Algoritmos Genéticos (GA-VSS), optimizando como parámetro de clasificación la tasa de no error en validación cruzada (NER_{cv}). El resultado de la optimización del NER_{cv} respecto al número de variables incluidas en el modelo se muestra en la Fig. 1.

Del algoritmo de selección de variables se obtuvo un modelo constituido por 11 descriptores moleculares, buscando cumplir el principio de parsimonia (Occam's Razor Principle) [6]. En la tabla 2 se muestra la información de los mejores descriptores con los cuales se construyeron y validaron los modelos *k*-NN y CP-ANN.

Estos 11 descriptores son los que mejor describen la relación entre la estructura molecular y la toxicidad de los pesticidas en las tres clases asignadas oportunamente. Para la construcción del mapa de Kohonen hexagonal se usaron los siguientes parámetros optimizados: número de neuronas de 14×14 , número de épocas de 150 y la frecuencia en un valor de 0.2. El mapa obtenido se muestra en la Fig. 2.

El gráfico que es de forma toroidal muestra claramente un espacio de clase definido por cada color, es decir,

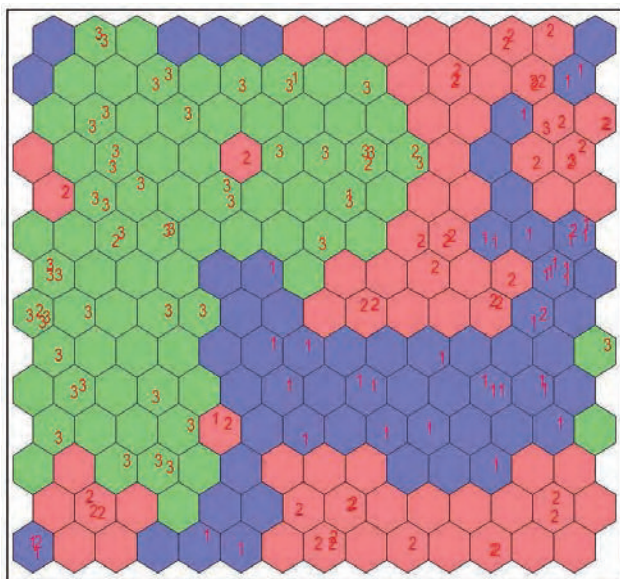


Figura 2: Mapa de Kohonen obtenido para el modelo de clasificación CP-ANN para el subconjunto de calibración.

azul para la clase 1 (toxicidad baja), rojo para la clase 2 (toxicidad media) y verde para la clase 3 (toxicidad alta). Se observa que existen pesticidas que son mal clasificados y son los que determinan la tasa de error del modelo ($ER=1-NER$). En este mapa se indican únicamente los resultados de asignación de los pesticidas del subconjunto de calibración, a partir del cual se obtuvieron los valores de cada clase. Para la construcción de modelo de clasificación del *k*-NN se optimizó previamente el número de los *k* vecinos, usándose valores comprendidos entre 3 y 7. El valor de *k* que muestra la mejor respuesta, es decir, la menor tasa de no error fue $k=4$ ($NER=0.82$). Con las clases asignadas por cada método de clasificación (ver Tabla 1S), se construyeron las matrices de confusión para interpolación, tanto para el *k*-NN y CP-ANN, cuyos resultados se muestran en la Tabla 3.

A partir de estas matrices se calculó el NER_{cal} según la ecuación 2. Los mejores resultados se obtienen para el método CP-ANN, mientras que en validación cruzada *k*-NN resulta ser más estable (ver Tabla 4).

Posteriormente, se usó el subconjunto de predicción para realizar la validación externa de los modelos. Se calcularon las clases asignadas por cada método (ver Tabla 1S) y con estos valores se construyeron las matrices de confusión que se presentan en la Tabla 5, que servirán para establecer los parámetros predictivos de los modelos.

Dado que *k*-NN muestra mejores resultados respecto a CP-ANN se procedió a realizar un análisis más exhaustivo del modelo, calculando el NER_{cal} para cada una de las clases, como se observa en la Tabla 6. El modelo *k*-NN es bastante estable para separar apropiadamente los pesticidas de las tres clases, de manera significativa los de toxicidad alta (89 %), seguido por los de toxicidad baja (84 %) y algo menor para los de toxicidad media (75 %).

Asimismo, se calcularon los valores de Sensibilidad (S_n) y Especificidad (S_p) [9] (ver Tabla 6). La sensibilidad se define como la capacidad de una clase para representar los objetos de aquella clase, indicando que la clase 3 tiene la mayor sensibilidad seguida de la clase 1; mientras que, la clase menos sensible es la 2 debido a que se encuentra en la frontera entre las clases de pesticidas de

Método	NER_{cal}	NER_{cv}	NER_{pred}
<i>k</i> -NN	0.88	0.83	0.70
CP-ANN	0.91	0.79	0.67

Tabla 4: Valores de la tasa de no error alcanzados para los modelos de clasificación en calibración, validación cruzada y predicción.

Clases verdaderas	Clases Asignadas <i>k</i> -NN			Clases Asignadas CP-ANN		
	Clase 1	Clase 2	Clase 3	Clase 1	Clase 2	Clase 3
Clase 1	12	2	3	11	3	3
Clase 2	1	11	8	2	10	8
Clase 3	1	2	17	1	2	17

Tabla 5: Matrices de confusión para los métodos *k*-NN y CP-ANN obtenidos a partir del subgrupo de predicción.

baja y alta toxicidad, lo que generaría que varios de estos vengan clasificados erróneamente en alguna de estas dos clases. Por otro lado, la especificidad define la capacidad de una clase para aislar de las otras clases los objetos de aquella clase. La clase 2 es la más específica a pesar de ser la menos sensible; por otro lado, la clase 1 resulta ser la más sensible y específica y la clase tres es la que tiene la menor especificidad.

Según la estructura molecular de los pesticidas, al existir mayor cantidad de estructuras cíclicas y heterocíclicas aumenta su toxicidad [4]. En este estudio los pesticidas clasificados en la clase de baja toxicidad tienen la característica de ser moléculas de cadena lineal; mientras que, las moléculas de la clase 2 tienen una estructura cíclica aromática como característica principal y los compuestos de la clase 3 son estructuras cíclicas y heterocíclicas en donde se encuentran más de dos anillos aromáticos.

El modelo encontrado y validado presenta una buena capacidad predictiva ($NER_{cv}=0,83$) a pesar que el mismo integra pesticidas de diversa estructura química. En un estudio similar publicado en 2002, Ferreira [45] desarrolló un modelo QSAR con un $NER_{cv}=0,71$, utilizando 50 pesticidas, lo cual representa un 26 % de compuestos respecto a los 192 considerados en el presente estudio. El presente modelo presenta mejores parámetros tanto en calibración, validación cruzada y predicción.

En los estudios QSAR son muy apreciados los modelos continuos, dado que la predicción devuelve un valor de toxicidad (modelo cuantitativo), en tanto que los modelos discretos muestran el valor de una clase de toxicidad (modelo cualitativo) [46]. Sin embargo, los modelos discretos permiten enfrentar estudios donde la variabilidad de la respuesta y la heterogeneidad de las fuentes de medición constituyen un detonante para que un modelo continuo fracase y donde un modelo de clasificación puede ser usado como herramienta útil. De hecho, los métodos de clasificación son apropiados en los estudios QSAR para modelar diversos tipos de respuestas, tales como compuestos activos/no activos o compuestos de toxicidad baja, mediana y alta [47].

En definitiva, dado que no se consideran dentro de los modelos los descriptores 3D, se tiene la ventaja de reducir el costo computacional que involucra el determinar

Clase	NER_{cal}	Sn	Sp
1	0,84	0.85	0.85
2	0,75	0.75	0.86
3	0,89	0.90	0.80

Tabla 6: Tasa de no error, Sensibilidad y Especificidad para cada clase en el modelo de clasificación *k*-NN.

la geometría óptima de menor energía, que también podría limitar la aplicación futura de los modelos *k*-NN y CP-ANN debido a inconsistencias con la generación de estructuras 3D [48].

Conclusiones

Se presenta una aplicación satisfactoria de la teoría QSAR para el desarrollo de los modelos de clasificación *k*-NN y CP-ANN para la predicción de la toxicidad (EC_{50} medida en *Daphnia magna*, en un ensayo de duración de 48 horas) de 192 pesticidas definida discretamente en tres clases (baja, media y alta toxicidad). Los modelos muestran buena estabilidad en calibración, validación cruzada y predicción; sin embargo, entre los dos, el modelo *k*-NN muestra los mejores parámetros de calidad ($NER_{cal}=0.88$, $NER_{cv}=0.83$ y $NER_{pred}=0.70$), pudiéndose usar para modelar esta actividad y predecir el comportamiento de pesticidas no considerados y aún no sintetizados. Este modelo complementa resultados previamente publicados; sin embargo, constituye un modelo más general y predictivo. Finalmente, los descriptores 3D no ofrecen información relevante para modelar las clases y en este contexto, los métodos QSAR independientes de la conformación continúan a emerger como una alternativa para el desarrollo de los modelos, basados únicamente en descriptores topológicos y constitucionales de los compuestos.

Agradecimientos

Cristian Rojas agradece la Beca Doctoral otorgada por la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) de la República del Ecuador. Asimismo, los autores agradecen al Milano Chemometrics and QSAR Research Group de la Università degli Studi di Milano - Bicocca por permitir el uso del algoritmo de selección de variables GA Toolbox.

Referencias

- [1] Jaramillo, B.; Martelo, I.; Duarte, E. 2013. "Toxicidad Aguda de Pesticidas Organofosforados y Análisis de la Relación Cuantitativa de Estructura Actividad (QSAR)". *Bioteología en el Sector Agropecuario y Agroindustrial*, 11:76-84.
- [2] Castillo Morales, G. 2004. "Ensayos toxicológicos y métodos de evaluación de calidad de aguas; estandarización, intercalibración, resultados y aplicaciones". *IDRC/IMTA*.

- [3] Mazzatorta, P.; Smiesko, M.; Lo Piparo, E.; Benfenati, E. 2005. "QSAR model for predicting pesticide aquatic toxicity". *Journal of chemical information and modeling*, 45:1767–1774.
- [4] Zvinavashe, E.; Du, T.; Griff, H. T. van den Berg; Soffers, A.; Vervoort, J.; Murk, A.; Rietjens, I. 2009. "Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to *Daphnia magna* and *Cyprinus carpio*". *Chemosphere*, 75: 1531–1538.
- [5] Toropov, A.; Benfenati, E. 2006. "QSAR models for *Daphnia* toxicity of pesticides based on combinations of topological parameters of molecular structures". *Bioorganic & medicinal chemistry*, 14:2779–2788.
- [6] Duchowicz, P.; Castro, E. 2013. "The Importance of the QSAR-QSPR Methodology to the Theoretical Study of Pesticides". *International Journal of Chemical Modeling*, 5:35–50.
- [7] Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Sellassie, C. 2002. "Chem-bioinformatics: comparative QSAR at the interface between chemistry and biology". *Chemical reviews*, 102:783–812.
- [8] Frank, I.; Todeschini, R. 1994. "The data analysis handbook". Elsevier.
- [9] Todeschini, R. 1998. "Introduzione alla chemiometria". Edises, Napoli, Italia.
- [10] Zupan, J.; Novič, M.; Gasteiger, J. 1995. "Neural networks with counter-propagation learning strategy used for modelling". *Chemometrics and intelligent laboratory systems*, 27:175–187.
- [11] Zupan, J.; Novič, M.; Ruisánchez, I. 1997. "Kohonen and counterpropagation artificial neural networks in analytical chemistry". *Chemometrics and intelligent laboratory systems*, 38:1–23.
- [12] Leardi, R. 2001. "Genetic algorithms in chemometrics and chemistry: a review". *Journal of Chemometrics*, 15: 559–569.
- [13] Leardi, R. 2003. "Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks". Elsevier.
- [14] Leardi, R. 2007. "Genetic algorithms in chemistry". *J. Chromatogr. A*, 1158:226–233.
- [15] University of Hertfordshire. 2013. "PPDB: Pesticide Properties DataBase". <http://sitem.herts.ac.uk/aeru/ppdb/en/index.htm>.
- [16] Lewis, K.; Green, A. 2011. "The Pesticide Properties Database". *Chemistry International*.
- [17] Todeschini, R.; Consonni, V. 2009. "Molecular Descriptors for Chemoinformatics". WILEY-VCH, Weinheim.
- [18] Hypercube, Inc. 2014. "HyperChem". <http://www.hyper.com>.
- [19] TALETE. 2014. "DRAGON, Software for Molecular Descriptor Calculation". <http://www.talete.mi.it/>.
- [20] Cho, S.; Hermsmeier, M. 2002. "Genetic algorithm guided selection: variable selection and subset selection". *Journal of chemical information and computer sciences*, 42:927–936.
- [21] Fan, Y.; Shi, L.; Kohn, K.; Pommier, Y.; Weinstein, J. 2001. "Quantitative structure-antitumor activity relationships of camptothecin analogues: cluster analysis and genetic algorithm-based studies". *Journal of medicinal chemistry*, 44:3254–3263.
- [22] Fernandez, M.; Caballero, J.; Fernandez, L.; Sarai, A. 2011. "Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM)". *Molecular diversity*, 15: 269–289.
- [23] Gao, H. 2001. "Application of BCUT metrics and genetic algorithm in binary QSAR analysis". *Journal of chemical information and computer sciences*, 41:402–407.
- [24] Ghosh, P.; Bagchi, M. 2009. "QSAR modeling for quinoline derivatives using genetic algorithm and simulated annealing based feature selection". *Current medicinal chemistry*, 16:4032–4048.
- [25] Hemmateenejad, B. 2004. "Optimal QSAR analysis of the carcinogenic activity of drugs by correlation ranking and genetic algorithm-based PCR". *Journal of Chemometrics*, 18:475–485.
- [26] Hemmateenejad, B.; Akhond, M.; Miri, R.; Shamsipur, M. 2003. "Genetic algorithm applied to the selection of factors in principal component-artificial neural networks: application to QSAR study of calcium channel antagonist activity of 1, 4-dihydropyridines (nifedipine analogues)". *Journal of chemical information and computer sciences*, 43:1328–1334.
- [27] Hemmateenejad, B.; Miri, R.; Akhond, M.; Shamsipur, M. 2002. "QSAR study of the calcium channel antagonist activity of some recently synthesized dihydropyridine derivatives. An application of genetic algorithm for variable selection in MLR and PLS methods". *Chemometrics and intelligent laboratory systems*, 64:91–99.
- [28] Hoffman, B.; Kopajtic, T.; Katz, J.; Newman, A. 2000. "2D QSAR modeling and preliminary database searching for dopamine transporter inhibitors using genetic algorithm variable selection of Molconn Z descriptors". *Journal of medicinal chemistry*, 43:4151–4159.
- [29] Leardi, R.; Lupiáñez González, A. 1998. "Genetic algorithms applied to feature selection in PLS regression: how and when to use them". *Chemom. Intell. Lab. Syst.*, 41:195–207.
- [30] Leardi, R.; Seasholtz, M.; Pell, R. 2002. "Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data". *Analytica Chimica Acta*, 461:189–200.
- [31] Li, T.; Mei, H.; Cong, P. 1999. "Combining nonlinear PLS with the numeric genetic algorithm for QSAR". *Chemometrics and intelligent laboratory systems*, 45: 177–184.

- [32] Saripinar, E.; Geçen, N.; Şahin, K.; Yanmaz, E. 2010. "Pharmacophore identification and bioactivity prediction for triaminotriazine derivatives by electron conformational-genetic algorithm QSAR method". *European journal of medicinal chemistry*, 45:4157–4168.
- [33] Sutherland, J.; O'brien, L.; Weaver, D. 2003. "Spline-fitting with a genetic algorithm: A method for developing classification structure-activity relationships". *Journal of chemical information and computer sciences*, 43:1906–1915.
- [34] Taha, M.; Qandil, A.; Zaki, D.; AlDamen, M. 2005. "Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling". *European journal of medicinal chemistry*, 40:701–727.
- [35] Turner, D.; Willett, P. 2000. "Evaluation of the EVA descriptor for QSAR studies: 3. The use of a genetic algorithm to search for models with enhanced predictive properties (EVA_GA)". *Journal of Computer-Aided Molecular Design*, 14:1–21.
- [36] Wanchana, S.; Yamashita, F.; Hashida, M. 2003. "QSAR analysis of the inhibition of recombinant CYP 3A4 activity by structurally diverse compounds using a genetic algorithm-combined partial least squares method". *Pharmaceutical research*, 20:1401–1408.
- [37] Weber, L. 1998. "Applications of genetic algorithms in molecular diversity". *Current Opinion in Chemical Biology*, 2:381–385.
- [38] Pavan, M.; Consonni, V.; Gramatica, P.; Todeschini, R. 2006. "New QSAR Modelling Approach Based on Ranking Models by Genetic Algorithms - Variable Subset Selection (GA-VSS)". in: R. Brüggemann, L. Carlsen (Eds.) *Partial Order in Environmental Sciences and Chemistry*, Springer Berlin Heidelberg: 181–217.
- [39] Ballabio, D.; Vasighi, M.; Consonni, V.; Kompany-Zareh, M. 2011. "Genetic algorithms for architecture optimisation of counter-propagation artificial neural networks". *Chemometrics and intelligent laboratory systems*, 105:56–64.
- [40] Devillers, J. 2012. "Genetic Algorithms in Molecular Modeling (Principles of QSAR and Drug Design)". *Academic Press, New York*.
- [41] The MathWorks, Inc. 2014. "Matlab". *Masachussetts, USA*, <http://www.mathworks.com>.
- [42] Gemperline, P. 2012. "Practical guide to chemometrics". *CRC press*.
- [43] Ballabio, D.; Consonni, V.; Todeschini, R. 2009. "The Kohonen and CP-ANN toolbox: a collection of MATLAB modules for self organizing maps and counter-propagation artificial neural networks". *Chemometrics and intelligent laboratory systems*, 98:115–122.
- [44] Ballabio, D.; Vasighi, M. 2012. "A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies". *Chemometrics and intelligent laboratory systems*, 118:24–32.
- [45] Ferreira, M. 2002. "Multivariate QSAR". *Journal of the Brazilian Chemical Society*, 13:742–753.
- [46] Benfenati, E. 2011. "Quantitative Structure-Activity Relationships (QSAR) for pesticide regulatory purposes". *Elsevier*.
- [47] Gramatica, P. 2007. "Principles of QSAR models validation: internal and external". *QSAR Comb. Sci*, 26: 694–701.
- [48] Brandmaier, S.; Peijnenburg, W.; Durjava, M.; Kolar, B.; Gramatica, P.; Papa, E.; Bhatarai, B.; Kovarich, S.; Cas-sani, S.; Roy, P. 2014. "The QSPR-THESAURUS: The online platform of the CADASTER project". *Alternatives to laboratory animals: ATLA*, 42:13–24.

Material Suplementario

Id.	Tipo	Nombre	EC₅₀	Clase_{EXP}	Clase_{k-NN}	Clase_{CP-ANN}
1	Fosforados	Acefato	67.2	1	1	1
2	Fosforados	Ampropylfos	100	1	1	1
3*	Fosforados	Anilofos	56	1	1	2
4	Fosforados	Azametifos	0.00067	1	1	1
5	Fosforados	Azinfos-etil	0.0002	3	2	2
6*	Fosforados	Azinfos-metil	0.0011	3	3	3
7	Fosforados	Bensulida	0.58	1	1	1
8	Fosforados	Bromofos	0.0086	2	2	2
9*	Fosforados	Bromofos-etil	0.0086	2	1	2
10	Fosforados	Butonato	0.006	2	1	1
11	Fosforados	Cadusafos	0.00075	3	3	3
12	Fosforados	Cianofos	0.097	2	2	2
13	Fosforados	cloretoxifos	0.00041	3	3	3
14	Fosforados	Clorfenvinfos	0.00025	3	3	3
15	Fosforados	Clorpirifos	0.0001	3	1	2
16	Fosforados	clorpirifos-metil	0.0006	3	3	3
17*	Fosforados	Demeton-o	0.0104	2	2	2
18	Fosforados	Demeton-s	0.0104	2	2	2
19	Fosforados	Demeton-S-metil	0.023	2	2	2
20*	Fosforados	Diazinon	0.001	3	2	3
21	Fosforados	Diclofention	0.0011	3	3	3
22	Fosforados	Diclorvos	0.00019	1	1	2
23*	Fosforados	Dicrotofos	0.013	2	2	2
24	Fosforados	Dimethylvinphos	0.002	3	3	3
25*	Fosforados	Dimetoato	2	1	1	1
26*	Fosforados	Dioxation	0.00035	3	3	3
27*	Fosforados	Disulfoton	0.013	2	2	1
28	Fosforados	Edifenfos	0.000032	3	3	3
29	Fosforados	EPN	0.00006	3	2	2
30	Fosforados	Etion	0.000056	3	3	3
31	Fosforados	Etoprop	0.2	1	2	2
32	Fosforados	Etrimfos	0.004	2	2	1
33	Fosforados	Fenamifos	0.0019	3	3	3
34	Fosforados	Fenitrotion	0.0086	2	2	2
35	Fosforados	Fenkaptona	0.0013	3	3	3
36	Fosforados	Fonofos	0.0023	3	3	3
37*	Fosforados	Forato	0.004	2	2	2
38	Fosforados	Formotion	16.1	1	2	2
39	Fosforados	Fosalon	0.00074	3	3	3
40	Fosforados	Fosamina	152	1	1	1
41	Fosforados	Fosfamidon	0.008	2	2	2
42	Fosforados	Fosmet	0.002	3	3	3
43*	Fosforados	Fostiazato	0.282	1	1	2
44	Fosforados	Foxim	0.00081	3	3	3
45	Fosforados	Heptenofos	0.0022	3	2	2
46	Fosforados	Iodofenfos	0.0016	3	3	3
47*	Fosforados	Iprobenfos	0.2	1	1	1
48	Fosforados	Isazofos	0.5	1	1	1
49*	Fosforados	Isofenfos	0.0039	2	3	2
50	Fosforados	Isoxation	0.0052	2	2	2
51*	Fosforados	Malation	0.0007	3	2	3
52	Fosforados	Mefosfolan	0.0003	3	3	3
53	Fosforados	Metamidofos	0.27	1	1	1
54*	Fosforados	Metidation	0.0064	3	2	3
55	Fosforados	Mevinfos	0.00016	3	2	2
56*	Fosforados	Monocrotofos	0.023	2	2	2

Tabla 1S: Nombres químicos, familia de pertenencia, Concentración Efectiva 50 y clases de los 192 pesticidas. Las moléculas del subgrupo de predicción se indican con *.

Id.	Tipo	Nombre	EC₅₀	Clase_{Exp}	Clase_{k-NN}	Clase_{CP-ANN}
57*	Fosforados	Naled	0.00035	3	3	3
58	Fosforados	Omatoato	0.022	2	2	2
59	Fosforados	Oxidemeton metil	0.11	2	2	2
60	Fosforados	Paratión	0.0025	2	2	2
61	Fosforados	Paratión-metil	0.0073	2	2	2
62	Fosforados	Piperophos	0.0033	2	2	2
63	Fosforados	Pirimifos-etil	0.0025	2	1	1
64	Fosforados	Pirimifos-metil	0.00021	1	2	2
65	Fosforados	Profenofos	0.5	1	1	1
66	Fosforados	Protiofos	0.014	2	2	2
67	Fosforados	Quinalfos	0.00066	3	2	2
68*	Fosforados	Sulfotep	0.002	3	2	
69	Fosforados	Sulprofos	0.00083	3	3	3
70	Fosforados	Tebupirimfos	0.00008	3	3	3
71	Fosforados	Temefos	0.00001	3	1	2
72*	Fosforados	Terbufos	0.00031	3	2	3
73*	Fosforados	Tetraclorvinfos	0.002	3	3	3
74	Fosforados	Tetraetil pyrofosfato	0.055	2	1	1
75*	Fosforados	Tiometon	8.2	1	1	2
76	Fosforados	Triazofos	0.0026	2	2	2
77	Fosforados	Tribufos	0.12	2	2	2
78	Fosforados	Triclorfon	0.00096	3	2	2
79	Fosforados	Vamidotion	0.19	1	2	2
80	Carbamatos	Alanycarb	9.4	1	1	1
81	Carbamatos	Aldicarb	0.42	1	3	2
82*	Carbamatos	Aldoxicarb	0.28	1	1	2
83*	Carbamatos	Aminocarb	0.19	1	1	1
84	Carbamatos	Asulam	59.7	1	1	1
85*	Carbamatos	Barban	0.3	1	1	1
86	Carbamatos	Bendiocarb	0.03	2	2	2
87	Carbamatos	Benfuracarb	0.01	2	2	2
88*	Carbamatos	Benthiavalicarb isopropyl	10	1	1	1
89	Carbamatos	Butilato	158.6	1	1	1
90	Carbamatos	Butocarboxim	3.2	1	1	1
91	Carbamatos	Butoxicarboxim	0.5	1	1	1
92	Carbamatos	Carbaril	0.0064	2	2	2
93	Carbamatos	Carbetamida	81	1	1	1
94	Carbamatos	Carbofurano	0.0094	2	2	2
95*	Carbamatos	Carbosulfan	0.0015	3	2	3
96	Carbamatos	Chinometionato	0.12	2	3	3
97	Carbamatos	Cicloato	24	1	1	1
98	Carbamatos	Clororofam	2.6	3	3	3
99	Carbamatos	Dazomet	19	1	1	1
100	Carbamatos	Desmedifam	0.45	1	1	1
101*	Carbamatos	Dietofencarb	23	1	1	1
102*	Carbamatos	Dimepiperato	40	1	1	1
103	Carbamatos	EPTC	14	1	1	1
104	Carbamatos	Etiofencarb	0.22	1	1	1
105	Carbamatos	Fenmedifam	0.41	1	1	1
106*	Carbamatos	Fenobucarb	0.1	2	2	2
107	Carbamatos	Fenotiocarb	6.7	1	1	1
108	Carbamatos	Fenoxicarb	0.5	1	1	1
109*	Carbamatos	Ferbam	0.09	2	1	3
110	Carbamatos	Furatiocarb	0.0018	3	3	3
111*	Carbamatos	Iprovalicarb	19.8	1	1	1
112	Carbamatos	Isolan	0.0125	2	2	2

Continuación Tabla 1S: Nombres químicos, familia de pertenencia, Concentración Efectiva 50 y clases de los 192 pesticidas. Las moléculas del subgrupo de predicción se indican con *.

Id.	Tipo	Nombre	EC₅₀	Clase_{Exp}	Clase_{k-NN}	Clase_{CP-ANN}
113	Carbamatos	Isoprocarb	0.024	2	2	2
114*	Carbamatos	Metiocarb	0.008	2	3	3
115	Carbamatos	Metomilo	0.0076	2	3	1
116	Carbamatos	Mexacarbato	0.0018	3	2	2
117	Carbamatos	Molinato	14.9	1	3	2
118	Carbamatos	Orbencarb	2.88	1	3	2
119	Carbamatos	Oxamilo	0.319	1	3	2
120	Carbamatos	Pebulato	5.9	1	1	1
121	Carbamatos	Pirimicarb	0.017	2	2	2
122*	Carbamatos	Profam	23	1	3	2
123*	Carbamatos	Propamocarb	106	1	1	1
124	Carbamatos	Propoxur	0.15	1	1	2
125	Carbamatos	Prosulfocarb	0.51	1	1	1
126*	Carbamatos	Tiobencarb	1.1	1	1	2
127	Carbamatos	Tiodicarb	0.027	2	2	2
128	Carbamatos	Tiofanox	0.31	1	1	1
129*	Carbamatos	Tiram	0.011	2	3	1
130*	Carbamatos	Trialato	0.091	2	3	3
131*	Carbamatos	Trimetacarb	30	1	1	2
132*	Carbamatos	Vernolato	1.8	1	1	1
133	Carbamatos	XMC	0.055	2	2	2
134*	Carbamatos	Ziram	0.048	2	2	2
135	Clorados	Aldrin	0.028	2	2	2
136	Clorados	Alfaendosulfan	0.44	1	2	2
137*	Clorados	Alpha-hexaclorociclohexano	0.37	1	2	2
138*	Clorados	Bromociclen	0.7	1	2	
139	Clorados	Clortion	4.2	1	1	1
140	Clorados	Clorbenzilato	0.01	2	2	2
141*	Clorados	Clordano	0.59	1	1	1
142*	Clorados	Clordecon	0.03	2	2	2
143	Clorados	DDT	0.005	2	2	2
144	Clorados	Dicofol	0.14	1	1	2
145	Clorados	Dienoclor	1.2	3	3	3
146*	Clorados	Endosulfan	0.44	1	1	2
147	Clorados	Endrin	0.0042	2	2	2
148	Clorados	Epiclorhidrina	20.4	1	1	1
149	Clorados	Fomesafen	22.1	1	1	1
150	Clorados	Heptaclor	0.042	2	2	2
151	Clorados	Isobenzan	0.008	2	2	2
152	Clorados	Lindano	1.6	1	1	1
153	Clorados	Metoxiclor	0.00078	3	2	2
154	Clorados	Mirex	0.1	2	2	2
155	Clorados	Pentaclorofenato	0.45	1	1	1
156	Clorados	TDE	0.009	2	2	2
157*	Piretroides	Acrinatrina	0.000022	3	3	3
158	Piretroides	Aletrina	0.021	2	2	2
159	Piretroides	Alfacipermetrina	0.0003	3	3	3
160	Piretroides	Beta-ciflutrina	0.00029	3	3	3
161	Piretroides	Beta-cipermetrina	0.00026	3	3	3
162	Piretroides	Bifentrin	0.00011	3	3	3
163*	Piretroides	Bioaletrina	0.0356	2	2	1
164*	Piretroides	Bioresmetrin	0.0008	3	2	3
165	Piretroides	Ciflutrin	0.00016	3	3	3
166	Piretroides	Cihalotrin lambda	0.00036	3	3	3
167	Piretroides	Cihalotrina	0.38	1	1	1
168	Piretroides	Cipermetrin	0.0003	3	3	3

Continuación Tabla 1S: Nombres químicos, familia de pertenencia, Concentración Efectiva 50 y clases de los 192 pesticidas. Las moléculas del subgrupo de predicción se indican con *.

Id.	Tipo	Nombre	EC₅₀	Clase_{EXP}	Clase_{k-NN}	Clase_{CP-ANN}
169	Piretroides	Cicloprotrin	10	1	1	1
170*	Piretroides	Cyphenothrin	0.00043	3	3	3
171*	Piretroides	Deltametrin	0.00056	3	1	3
172	Piretroides	Empentrin	0.02	2	2	2
173	Piretroides	Esfenvalerato	0.0009	3	3	3
174*	Piretroides	Etofenprox	0.0012	3	3	3
175	Piretroides	Fenotrin	0.0043	2	2	2
176*	Piretroides	Fenpropatrin	0.00053	3	3	3
177	Piretroides	Fenvalerato	0.00003	3	3	3
178*	Piretroides	Flucitrinato	0.0083	2	2	2
179	Piretroides	Fluvalinate	0.074	2	2	2
180*	Piretroides	Gamma-cihaltrina	0.00045	3	3	3
181*	Piretroides	Halfenprox	0.000031	3	3	3
182	Piretroides	Imiprothrin	0.051	2	2	2
183*	Piretroides	Metofluthrin	0.0047	2	2	1
184	Piretroides	Permetrin	0.0006	3	3	3
185	Piretroides	Praletrin	0.0062	2	2	2
186*	Piretroides	Resmetrin	0.0037	2	2	1
187	Piretroides	Tau-fluvalinate	0.0089	2	2	2
188	Piretroides	Teflutrín	0.00007	3	3	3
189	Piretroides	Tetrametrina	0.045	2	2	2
190	Piretroides	Tralometrín	0.000038	3	3	3
191	Piretroides	Transfluthrin	0.0017	3	3	3
192	Piretroides	Zeta-cipermetrin	0.00014	3	3	3

Continuación Tabla 1S: Nombres químicos, familia de pertenencia, Concentración Efectiva 50 y clases de los 192 pesticidas. Las moléculas del subgrupo de predicción se indican con *.