UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE INGENIERÍA AMBIENTAL

Tesis previa la obtención del Título de:
Ingeniera Ambiental

TÍTULO

"APLICACIÓN DE MÉTODOS QSAR Y QUIMIOMÉTRICOS PARA DETERMINAR LA RELACIÓN ESTRUCTURA – ACTIVIDAD DE PESTICIDAS EN MAMÍFEROS (RATA)"

AUTORA:
Cajamarca Cango Mónica Elizabeth

DIRECTOR:
Blgo. Luis Fernando Cárdenas Calle MSc

Cuenca, Febrero del 2015

CERTIFICACIÓN

Certifico, que la presente investigación Intitulada "APLICACIÓN DE MÉTODOS QSAR Y QUIMIOMÉTRICOS PARA DETERMINAR LA RELACIÓN ESTRUCTURA – ACTIVIDAD DE PESTICIDAS EN MAMÍFEROS (RATA)", fue desarrollada por Cajamarca Cango Mónica Elizabeth, con cédula de identidad número 110455152-6, bajo mi supervisión y tutoría.

Cuenca, Febrero del 2015

Blgo.. Luis Fernando Cárdenas Calle MSc

DIRECTOR DE TESIS

DECLARATORIA DE RESPONSABILIDAD

Los conceptos desarrollados, analizados y las conclusiones del presente trabajo son de exclusiva responsabilidad Cajamarca Cango Mónica Elizabeth.

A través de la presente declaración cedemos los derechos de propiedad intelectual correspondiente a este trabajo, a la Universidad Politécnica Salesiana, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la Normativa Institucional Vigente.

Cuenca, Febrero del 2015

Mónica Elizabeth Cajamarca Cango

AGRADECIMIENTOS

Quiero expresar mis sinceros agradecimientos a:

Mis padres y hermana por su apoyo permanente y haber confiado en mis capacidades, siempre.

Biólogo Fernando Cárdenas mi director de tesis, que me incentivo a desarrollar este tema de investigación, que inicialmente me parecía muy complicado.

Me gustaría también agradecer al Dr. Piercosimo Tripaldi, quien me brindó su apoyo y me compartió sus conocimientos para alcanzar el éxito en esta investigación.

DEDICATORIA

A mi amado hijo Francisco

A mis padres Olga y Luis

A mi hermana Priscilla

Ustedes son mi motivación, mi fuerza… son los pilares de mi vida.

RESUMEN

En el presente trabajo de investigación se describe el análisis de la relación: estructura de las moléculas de pesticidas y su actividad biológica (toxicidad) sobre mamíferos, apoyándose en técnicas quimiométricas.

En primera instancia se construyó una base de datos formada por 159 moléculas de pesticidas (organofosforados y carbamatos), y su respectiva respuesta biológica sobre mamíferos (rata) expresada en Concentración Inhibitoria (IC_{50}).

Las moléculas seleccionadas para este estudio se digitalizaron y optimizaron en el programa computacional HyperChem, luego fueron ingresadas al software DRAGON donde se calcularon un total de 254 descriptores moleculares (variables), por cada molécula.

Los valores de la respuesta biológica fueron tratados aplicando el método k-means, separándolos en dos clases según su actividad, la clase 1 de baja toxicidad y la clase 2 de alta toxicidad.

Luego se procedió con la selección de las variables óptimas mediante la aplicación de algoritmos genéticos y, el algoritmo *k*-NN para la asignación de las clases. Los algoritmos genéticos seleccionaron 2 variables de las 254 iniciales para la aplicación del modelo CART. Las variables elegidas fueron nN que es el número de nitrógenos y ATSC2e la correlación de electronegatividad.

Los resultados obtenidos de la aplicación del modelos fueron: con el algoritmo k-NN se obtuvo una tasa de no error o NER de 0,90 para la asignación de las clases, lo que indica que tiene una buena capacidad de asignación es del 90%.

En predicción se obtuvo un valor NER de 0,99 lo que se traduce como una capacidad predictiva del 99%, a pesar de que para construirlo se usaron moléculas de diferente origen químico.

Palabras clave: QSAR, Quimiometría, Pesticidas, Algoritmo Genético, KNN.

ÍNDICE DE CONTENIDOS

CERTIFICACIÓN II

DECLARATORIA DE RESPONSABILIDAD III

AGRADECIMIENTOS IV

DEDICATORIA V

RESUMEN VI

ÍNDICE DE CONTENIDOS VIII

JUSTIFICACIÓN 1

OBJETIVOS 3

Objetivo General 3

Objetivos específicos 3

CAPITULO I 4

Pesticidas 4

1.1. Definición 4

1.2. Clasificación de los pesticidas 4

1.3. Toxicidad de los pesticidas 8

1.4. Efectos en el Ambiente de los pesticidas 9

1.5. Ensayos de Toxicidad 10

CAPITULO II 13

Análisis QSAR 13

2.1. QSAR (Quantitative Structure-Activity Relationships) 13

2.2. Descriptores moleculares 14

2.3. Quimiometría 15

2.4. Algoritmos Genéticos (AG) 16

CAPÍTULO III 18

Metodología 18

3.1. Construcción de la Base de Datos 18

3.2. Digitalización y Optimización de las moléculas de pesticidas 19
3.3. Cálculo de los Descriptores Moleculares 21
3.4. Auto-escalado 21
3.5. Asignación de las Clases 22
3.6. Selección de las variables 22
3.7. Selección Training – Test 23
CAPITULO IV 24
RESULTADOS 24
4.1. Resultados 24
4.2. Validación del modelo 28
CONCLUSIONES 30
RECOMENDACIONES 32
ANEXOS 33
ANEXO 1: MOLÉCULAS DE PESTICIDAS Y SU ACTIVIDAD BIOLÓGICA 33
ANEXO 2: MOLÉCULAS DEL TRAINING SET 41
ANEXO 3: MOLÉCULAS DEL TEST SET 44
ANEXO 4: MOLÉCULAS ASIGNADAS A CADA CLASE 46
BIBLIOGRAFÍA 49

JUSTIFICACIÓN

Los pesticidas sintéticos aparecieron entre 1930 y 1940; el primero de los ellos se usó en la segunda guerra mundial con el fin de proteger a los soldados estadounidenses de vectores transmisores de enfermedades. Desde entonces hasta la actualidad se han creado más sustancias con el fin de controlar organismos que causan daño a los cultivos agrícolas o a la salud del hombre. Pero con el pasar del tiempo los efectos colaterales indeseables del uso de los pesticidas se han hecho evidentes, en las afecciones que provocan al ser humano y al ambiente. (RAMÍREZ & LACASAÑA, 2001).

Para saber cuál es la toxicidad de los pesticidas se realizan lo que se conoce como ensayos de toxicidad, desde la década de los 60's alrededor de todo el mundo se han realizado varios de estos ensayos, cuyo principal objetivo es regulador es decir evitar que ingresen al mercado sustancias químicas que pongan en riesgo la salud de la población mundial. (VINARDELL MARTÍNEZ - HIDALGO, 2007).

En los ensayos de toxicidad se pueden determinar diferentes aspectos: toxicidad aguda, subaguda, crónica, neurotoxicidad, carcinogénesis, mutagénesis, alteraciones en los cromosomas, etc; se utilizan animales como sujetos de prueba. (VINARDELL MARTÍNEZ - HIDALGO, 2007). Estas pruebas experimentales de la toxicidad de los productos (farmacéuticos, pesticidas, cosméticos, etc), ha provocado reacciones diversas. Por un lado los defensores de los animales que se oponen al sufrimiento a los que estos son sometidos, por otro la exigencia de garantizar la salud humana y finalmente los productores que expresan su molestia por el elevado coste de los estudios toxicológicos. (CAMEAN, 1995).

Por los motivos mencionados anteriormente, la comunidad científica se ha esforzado en encontrar alternativas a los ensayos con animales, que cumplan con los postulados del principio de las tres R que se refieren a: reemplazar los animales por otros métodos que no impliquen su uso, reducir el número si es estrictamente necesario su uso, y refinar las técnicas para aminorar su sufrimiento. (VINARDELL MARTÍNEZ

- HIDALGO, 2007). Dentro de estos métodos alternativos tenemos las técnicas QSAR (Quantitative Structure Activity Relationships) y CSAR (Classification Structure Activity Relationship), cuyo objetivo primordial es construir modelos matemáticos que relacionan la estructura química y la actividad biológica de una serie de compuestos análogos. (MAZZATORTA, 2010).

La presente investigación pretende entonces encontrar un modelo matemático mediante el uso de técnicas QSAR y métodos quimiométricos que nos permita relacionar la estructura molecular de los pesticidas y la toxicidad de los mismos; esto nos permitirá encontrar nuevas estructuras químicas de pesticidas que posean una actividad biológica definida. Estos métodos son una alternativa a los ensayos tradicionales de toxicidad, buscando así reducir las pruebas en animales, el ahorro de tiempo y dinero, y la fomentación de la investigación científica en esta área.

OBJETIVOS

OBJETIVO GENERAL

Encontrar un modelo matemático que describa la relación entre los pesticidas y la respuesta biológica del organismo en estudio (toxicidad).

Objetivos específicos

- Elaborar una base de datos de los pesticidas a ser utilizados.
- Modelar la estructura molecular de los pesticidas a estudiar mediante diferentes métodos quimiométricos.
- Identificar el modelo matemático que relacione la estructura molecular de los pesticidas y la respuesta biológica del organismo en estudio.
- Validar el modelo encontrado.

CAPITULO I

PESTICIDAS

1.1. Definición

La FAO (Food and Agriculture Organization) define a los pesticidas o plaguicidas como “cualquier sustancia o mezcla de sustancias destinadas a prevenir, destruir o controlar cualquier plaga, incluyendo los vectores de enfermedades humanas o de los animales, las especies de plantas o animales indeseables” (FAO, 2003).

También se los puede definir como la sustancia o sustancias utilizadas con la finalidad de matar, inhibir o altera el ciclo de vida de los organismos a los que se aplica. (VÁZQUEZ M. , 2001).

1.2. Clasificación de los pesticidas

Los pesticidas pueden clasificarse de acuerdo a ciertos criterios: por su origen siendo estos naturales y sintéticos, por su formulación o presentación comercial se clasifican en: polvos, líquidos, gases y comprimidos; por el tipo de plaga que ataca se agrupan en insecticidas, fungicidas, herbicidas, molusquicidas, rodenticidas, nematicidas, ovicidas, larvicidas y acaricidas; por su persistencia es decir su capacidad de permanecer en el ambiente son: no persistentes, persistentes y permanentes; por su naturaleza química tenemos entre los principales: organofosforados, carbamatos, organoclorados y piretroides. Por la relevancia de esta última clasificación para el presente estudio se profundizará a continuación.

1.2.1. Por su naturaleza química, este criterio de clasificación permite agrupar a los pesticidas de acuerdo a un criterio uniforme, permitiendo desarrollar relaciones de estructura – actividad, estructura – toxicidad, entre otras; en los principales tenemos:

1.2.1.1. Organofosforados.

Estos compuestos que se derivan de los ácidos forfórico, fosfónico y fosfortoico. Éstos pesticidas son solubles en agua, poco persistentes en el ambiente. (JUÁREZ, RINCÓN Sánchez, & MARTÍNEZ, 2009). Su modo de acción es inhibir la acetilcolinesterasa1 lo que provoca una acumulación de acetilcolina2 ocasionando una interrupción en la transmisión de los impulsos nerviosos. (FERRER, 2003).

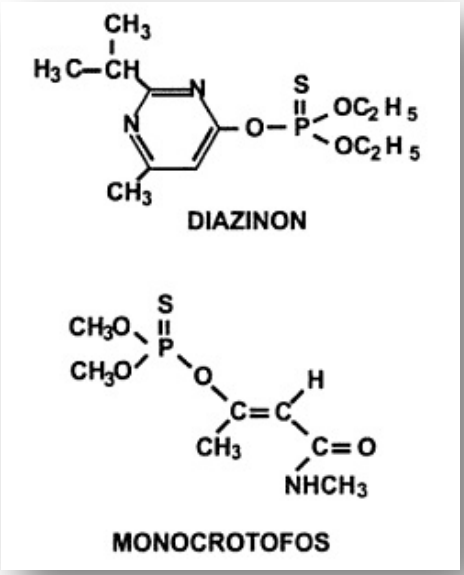

Ilustración 1 *Ejemplo de pesticidas organofosforados*

Fuente: VÁZQUEZ, A. (2005). *Golfo de México: contaminación e impacto ambiental : diagnóstico y tendencias*.

1.2.1.2. Carbamatos.

Son esteres N-metilados y N,N-dimetilados del ácido carbámico; pueden ser insecticidas herbicidas fungicidas nematicidas o molusquicidas de acuerdo a los sustituyentes de la molécula. Su modo de acción al igual que el de los organofosforados radica en la capacidad de inhibir la acetilcolinesterasa. (VÁZQUEZ A. , 2005).

---

[1]Enzima que descompone por hidrólisis el neurotransmisor acetilcolina. Tomado de SÁNCHEZ, y SALCEDA, *Enzimas polifuncionales: el caso de la Acetilcolinesterasa*, p.44-45, 2008.

[2]Neurotransmisor específico en las sinapsis del sistema nervioso somático. Tomado de http://www.biopsicologia.net/el-proyecto/n3-participacion-plastica-y-funcional/28-13-acetilcolina

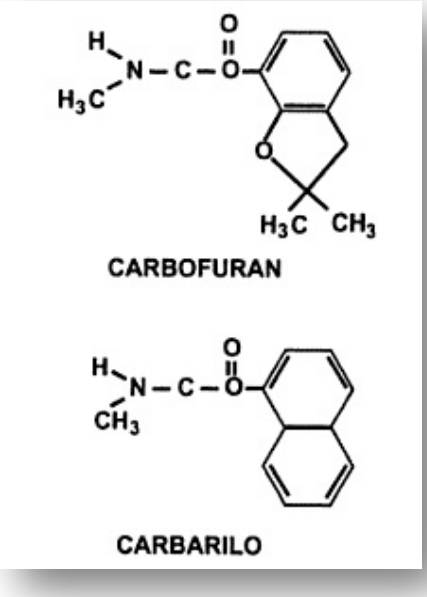

Ilustración 2: *Ejemplo de pesticidas de la familia carbamatos*

Fuente: VÁZQUEZ, A. (2005). *Golfo de México: contaminación e impacto ambiental : diagnóstico y tendencias*.

1.2.1.3. Organoclorados.

Son compuestos física y químicamente estables debido a que su estructura corresponde a los hidrocarburos clorados, por esta razón son solubles en disolventes orgánicos, insolubles en agua y no volátiles, se biodegradan lentamente y son persistentes en el ambiente. Su forma de acción es como venenos estomacales y de contacto. (RAMÍREZ & LACASAÑA, 2001).

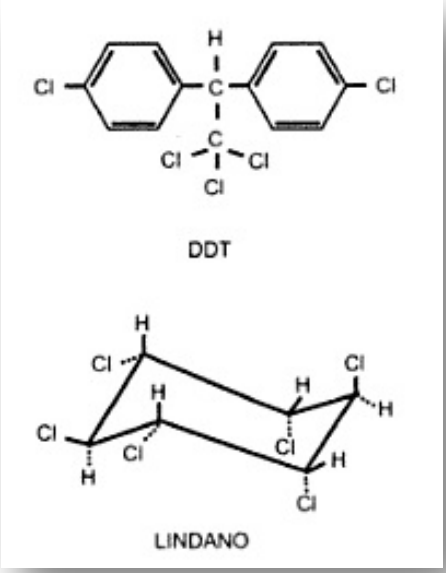

Ilustración 3: *Ejemplo de pesticida de la familia organoclorados*

Fuente: VÁZQUEZ, A. (2005). *Golfo de México: contaminación e impacto ambiental : diagnóstico y tendencias*.

1.2.1.4. Piretroides.

Son plaguicidas que en sus inicios se obtenían del polvo de la flor de crisantemo, en los años 50 fue sintetizado por primera vez. (RAMÍREZ & LACASAÑA, 2001). Estos compuestos son insolubles en agua, lipofílicos3, de fácil degradación por microorganismos; su modo de acción es sobre el sistema nervioso interfiriendo con los impulsos que este transmite. (VÁZQUEZ A. , 2005). A continuación en Ilustración 4 un ejemplo.

[3] Compuestos que tienen afinidad por las grasas.

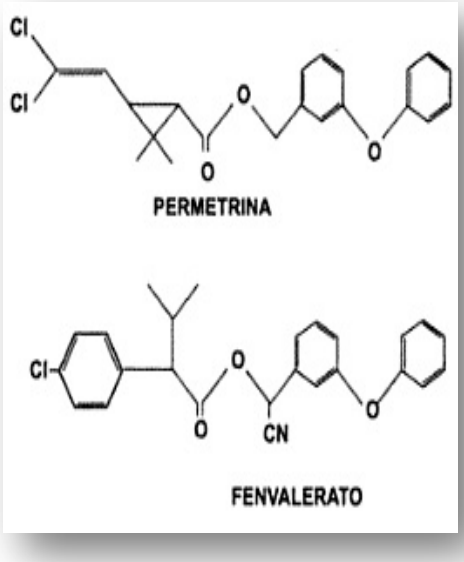

Ilustración 4: *Ejemplo de pesticidas de la familia piretroides.*

Fuente: VÁZQUEZ, A. (2005). *Golfo de México: contaminación e impacto ambiental : diagnóstico y tendencias*.

1.3. Toxicidad de los pesticidas

Se define como toxicidad "a la actividad tóxica, concreta y específica, vinculada a la estructura química de una sustancia exógena al organismo (xenobiótico) por su interacción con moléculas endógenas (receptor)". (BELLO & LÓPEZ de CERAIN, 2008).

Se la puede definir también como la capacidad que posee el pesticida de producir daños a un organismo o su biocenosis[4]. (DÍAZ, BULUS, & PICA, 2004). Esta toxicidad dependerá de situaciones como: propiedades químicas y concentración de la sustancia; en el caso específico de los pesticidas la OMS (2009) los ha agrupado en 5 clases conforme a su concentración (DL_{50} o CL_{50}) [5], como se puede observar en la Tabla 1.

[4]También es conocida como comunidad biótica o comunidad ecológica, es un conjunto de organismos de todas las especies coexistentes dentro de un espacio definido que se llama biotopo. Tomado de http://www.ecologiahoy.com/biocenosis.

[5] WHO, *The WHO recommended classification of pesticides by hazard and guidelines to classification*, p. 3-5, 2009.

Tabla 1: *DL_{50} para rata, expresado en mg/kg peso corporal*

Clase	Toxicidad	Vía Oral	Vía Dérmica
Clase IA	Extremadamente peligrosos	< 5	< 50
Clase IB	Altamente peligrosos	5 – 50	50 – 200
Clase II	Moderadamente peligrosos	50 – 2000	200 – 2000
Clase III	Ligeramente peligrosos	> 2000	> 2000
U	Improbable de presentar peligro agudo	5000 o mayor	

Fuente: WHO, The WHO recommended classification of pesticides by hazard and guidelines to classification, 2009.

1.4. Efectos en el Ambiente de los pesticidas

Los pesticidas tienen efectos indeseables sobre el medio ambiente que se puede agrupar de la siguiente manera:

1.4.1.1. Efectos adversos a corto plazo en el ambiente cercano, sus efectos se evidencian en el componente abiótico con contaminación en suelo, aire, agua superficial y subterránea del lugar donde se aplicaron, en el componente biótico afectan a organismos que no se deseaban afectar; en corto plazo afectan el equilibrio fisiológico de todos los organismos expuestos incluso el hombre. (MILLA & PALOMINO, 2002).

1.4.1.2. Efectos adversos a largo plazo en el ambiente cercano, además del daño inmediato, la aplicación de pesticidas de tipo persistente provoca efectos a largo plazo como: desarrollo de resistencia a los compuesto por parte de las plagas, transferencia acelerada de residuos de plaguicidas a la cadena trófica, aparición de nuevas plagas; en la parte abiótica la contaminación de los

suelos puede volverse irreversible y con ello se contaminen fuentes de agua. (MILLA & PALOMINO, 2002).

1.4.1.3. Efectos adversos a largo en el ambiente lejano, los efectos que se han evidenciado: es la presencia de pesticidas en la zonas polares donde no se usan, la biomagnificación a través de la cadenas tróficas, presencia de estos compuestos en alimentos de origen animal en los tejidos de las personas y en la leche materna, y extinción de especies. (ALBERT, 2005).

1.5. Ensayos de Toxicidad

Los ensayos de toxicidad son una herramienta de diagnóstico que evalúan la respuesta o efectos que puede causar la aplicación de un agente externo (pesticidas, metales pesados, muestras ambientales contaminadas con químicos, fámarcos, etc.) a un organismo en condiciones definidas. (CASTILLO Morales, 2004).

Los estudios de toxicidad con experimentación biológica son de los siguientes tipos: *in vivo* realizados en animales o vegetales, *in vitro* utilizando solo partes o tejidos y, estudios conjuntos con varias especies (incluido el hombre). (REPETTO & REPETTO, 2009).

La valoración de la toxicidad de las sustancias se hace principalmente con fines reguladores, es decir que se hace a sustancias que van a salir al o que ya estan en el mercado con el fin de evitar que al tener contacto con el hombre provoquen un problema de salud pública. Entre los principales ensayos de toxicología reguladora tenemos:

- **Toxicidad aguda**, consiste en la determinacion de la dosis que provoca la muete del 50% de la población.

- **Irritación ocular**, colocar la sustancia directamente en el ojo y observar los efectos.
- **Irritación dérmica**, este determinar el grado de inflamación y/o enrojecimiento de la piel despues de aplicar el producto.
- **Ensayos de mutagénesis, carcinogénesis, genotoxicidad** estos consisten en la inducir alteraciones a nivel celular y de cadena de adn. (VINARDELL MARTÍNEZ - HIDALGO, 2007)

Este tipo de estudios fundamenta su uso de los animales como sujetos de prueba en el hecho de que las pruebas alternativas no poseen la capacidad de predecir cual será el efecto de una sustancia al momento de entrar en un sistema de organos más complejo.

1.5.1. Alternativas a los ensayos con animales

Según lo expuesto en el Sexto Informe de los estados miembros de la Unión Europea sobre las estadísticas relativas al número de animales usados con fines científicos (SEGURA & D'ORS, 2012), la cantidad de animales que se usan con este fin superan los 12 millones de individuos, repartidos de la siguiente manera: 80% son conejos y roedores de los cuales el 59% fueron ratones, aves con 6%, anfibios reptiles y peces 9,6%.

La búsqueda de alternativas a la experimentación con animales en los últimos años ha incrementado debido a razones éticas, económicas y científicas, aunque el término “alternativa” fue mencionado ya en el año de 1959 por Rusell y Burch en su libro “The Principles of Humane Experimental Technique”, en el cual se expresa el principio de las 3R: reemplazar, reducir y refinar; que se describen brevemente a continuación. (VINARDELL MARTÍNEZ - HIDALGO, 2007).

1.5.1.1. Refinamiento

Constituye en mejorar las técnicas que se aplican en los estudios toxicológicos de tal manera que contribuyan a reducir la angustia, malestar o sufrimiento del animal mediante el uso de fármacos analgésicos en función de la especie, tamaño, tipo de dolor Se refiere también a mejorar las condiciones de vida de los sujetos de prueba mediante la implementación de lugares de alojamiento óptimos, mantenerlos libres de patógenos. (MENEAU, 2014)

1.5.1.2. Reducir

Consiste en reducir el número de animales utilizados en la experimentación, esta sugiere como primera opción realizar un análisis completo sobre la información existente para evitar que los ensayos sean duplicados; otra opción incluida es el uso de órganos o tejidos de animales que fueron sacrificados para otro estudio. (MENEAU, 2014).

1.5.1.3. Reemplazar

Este postulado consiste en la no utilización de animales en ningún tipo de procedimiento experimental, para lo que se sugieren usar métodos alternativos conocidos como sistemas *in silico*[6], que se basan en la aplicación de modelos matemáticos o de simulación. (MENEAU, 2014); entre estos los basados en el análisis estadístico de la relación estructura química – actividad conocidos como QSAR por sus siglas en inglés (Quantitative Structure-Activity Relationships).

[6] *In silico* se refiere a cualquier aplicación de tecnologías informáticas, como algoritmos, sistemas y análisis o minería de datos», señala el profesor Norbert Graf, director de la Clínica de Hematología y Oncología Pediátricas del Hospital Universitario de Sarre (Alemania) e investigador sénior del proyecto ACGT («Ensayos clínico-genómicos avanzados en cáncer». Tomado de http://cordis.europa.eu/result/rcn/86061_es.html.

CAPITULO II

ANÁLISIS QSAR

2.1. QSAR (Quantitative Structure-Activity Relationships)

Es una metodología de análisis que busca relacionar matemáticamente las características fisicoquímicas de una serie de compuestos y su actividad biológica cuantificada, permitiéndonos explicar la actividad de esos compuestos e incluso predecir la de aquellos que aún no han sido sintetizados o probados. (ZAVALA, GONZÁLEZ, & CHACON, 2013).

En este tipo de técnicas la actividad biológica se interpreta como una función de los diferentes aspectos de la estructura química de la molécula, lo que Hansh expreso matemáticamente en la siguiente fórmula: (HANSCH, 2002).

$$\ln A = f_h(X_h) + f_e(X_e) + f_s(X_s) + cte \quad \textbf{(1)}$$

Donde A es la respuesta biológica que depende de las funciones $f_h(X_h) + f_e(X_e) + f_s(X_s) + cte$, que pueden ser cualquier propiedad de la molécula, ya sea lipofílica, electrónica, estérica o de cualquier tipo. (HANSCH, 2002)

Según Todeschini (2009), QSAR se basa en la suposición de que la estructura de una molécula es responsable de sus características físicas, químicas y biológicas. De lo cual se desprende el hecho de que los modelos QSAR puedan predecir la actividad biológica de sustancias nuevas basándose en la estructura de compuestos similares. (TODESCHINI & CONSONNI, 2009)

La metodología QSAR maneja una gran cantidad de datos y que a su vez pueden ser complejos, por lo cual se apoya para su análisis en la quimiometría, la cual será explicada más adelante.

2.2. Descriptores moleculares

Los descriptores o índices moleculares son números obtenidos mediante un proceso lógico-matemático de transformación de la información química contenida en la representación simbólica de una molécula. (CRUZ, 2009).

Se los utiliza para modelar varias propiedades de las sustancias químicas basadas en diferentes teorías como: la química cuántica, teoría de la información, química orgánica, y la teoría de grafos. Los primeros descriptores moleculares obtenidos fueron el Índice de Winner y Número Platt a partir de la teoría de los grafos en el año de 1947, en 1960 se propusieron nuevos descriptores sobre la misma teoría y partir de esta fecha se iniciaron los estudios sistemáticos de los descriptores moleculares. (TODESCHINI & CONSONNI, 2009).

Los descriptores moleculares son de gran importancia en ciencias como química, farmacología, toxicología; pues cada uno de ellos contiene información de cada pequeña parte de la molécula, y como expresan Todeschini y Consonni (2009) la importancia de estos se puede evidenciar en la cantidad de descriptores propuestos por varios investigadores que hasta la fecha suman 3000. (TODESCHINI & CONSONNI, 2009)

Estos índices pueden clasificarse de forma general basados en las dimensiones que abarcan, así: 0D son descriptores constitucionales estos describen la conformación molecular son independientes de la fórmula o estructura, entre estos tenemos: peso molecular, número de átomos; 1D son descriptores unidimensionales estos consideran fragmentos de las moléculas como subconjuntos de átomos; 2D son descriptores bidimensionales que se derivan de la teoría de los grafos; 3D son descriptores tridimensionales que se basan en la conformación tridimensional de la molécula, y de su configuración espacial este tipo de descriptor describe propiedades estereoquímicas de la molécula. (CRUZ, 2009).

2.3. Quimiometría

Es una disciplina científica que nació alrededor de los años 70, que utiliza conocimientos matemáticos y estadísticos para aplicarlos a procesos químicos y obtener información relevante de dichos procesos. (MONGAY, 2011)

La Quimiometría ha desarrollado métodos de clasificación y regresión para poder obtener modelos fiables, ya sea para reproducción datos experimentales conocidos como para la predicción de datos desconocidos. (TODESCHINI & CONSONNI, 2009).

2.3.1. Métodos de Clasificación

Los métodos quimiométricos de clasificación usados en esta investigación fueron métodos de clasificación no paramétricos, que se mencionan a continuación.

2.3.1.1. *k*-NN

Es un método de clasificación no paramétrico, que clasifica al objeto nuevo en la categoría a la que sus vecinos más próximos pertenecen. Éste elige una distancia y selecciona los *k* objetos más cercanos a la molécula en estudio. (FRANK & TODESCHINI, 2006)

Este método tiene ventajas como: con un solo modelo se puede clasificar nuevos objetos independientemente del número de clases, puede clasificar objetos nuevos independientemente de la forma de las clases. (JIMÉNEZ, 2013).

2.3.1.2. CART

Es un método de clasificación no paramétrico que consiste en la segmentación binaria recursiva, es decir que divide repetidamente los datos; en cada división los datos se dividen en dos subgrupos, de acuerdo algún criterio, las particiones se siguen realizando hasta que haya la menor cantidad de datos. (BOUREL, 2012).

Este algoritmo desarrollado por Breiman en 1984, tiene como mayor ventaja que es fácil de interpretar, es un método flexible, las variables a analizar pueden ser varios tipos como categóricas nominales discretas continuas. (SALINAS & VALCARCEL, 2002)

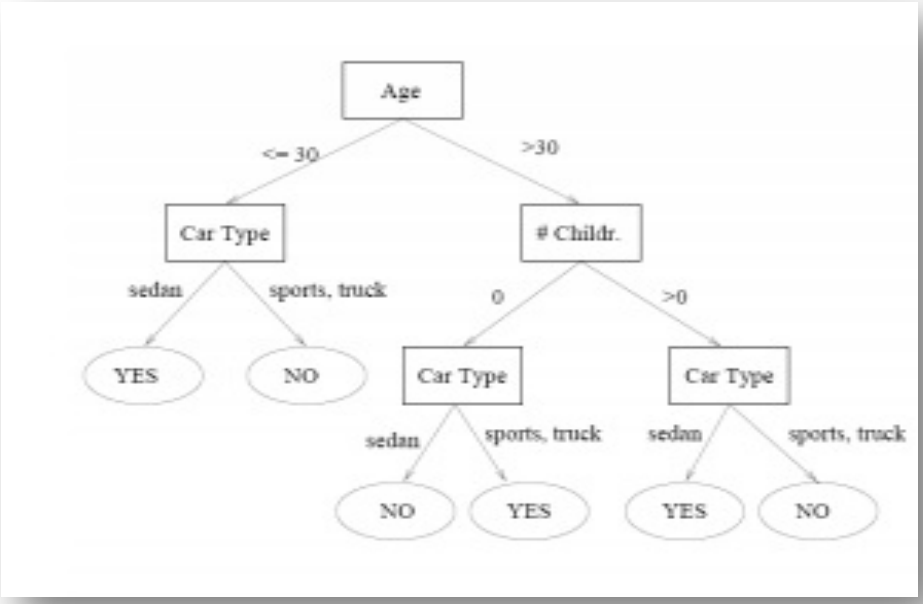

Ilustración 5*: Ejemplo de CART.*

Fuente: *SERNA, (2009), Comparación de Árboles de Regresión y Clasificación y Regresión Logística tomado de Dobra 2002*

2.4. Algoritmos Genéticos (AG)

Se puede definir a los algoritmos genéticos como un algoritmo de búsqueda que se basa en el proceso de la selección natural y supervivencia del más apto, imitando los postulados de Darwing (1859). (RODRÍGUEZ - PIÑERO, 2003).

2.4.1. Ventajas de los Algoritmos Genéticos

Entre las ventajas de los AG tenemos:

- Son individualmente paralelos lo que les permite explorar los espacios de soluciones en varias direcciones.
- Funcionan bien en la resolución de problemas donde el espacio de soluciones es vasto.
- Pueden manipular muchos parámetros de forma simultánea. (ROMÁN, MANZANAREZ, & TAPIA, 2012).

De acuerdo a Román, Manzanarez y Tapia (2012), los AG son una técnica firme que puede tratar exitosamente una gran variedad de problemas y encontrar soluciones aceptables. (ROMÁN, MANZANAREZ, & TAPIA, 2012).

CAPÍTULO III

METODOLOGÍA

En el presente capítulo se describirá cada uno de los procesos realizados para el desarrollo del modelo, empezando por la construcción de la base de datos, seguido de la digitalización y optimización de las moléculas, el cálculo de los descriptores moleculares, selección de las variables y obtención el modelo.

3.1. Construcción de la Base de Datos

Para la elaboración de la base de datos se utilizó la información proporcionada por la Universidad Hertfordshire de Inglaterra disponible en su página web http://sitem.herts.ac.uk/aeru/footprint/es/index.htm., aquí encontramos información de parámetros fisicoquímicos y ecotoxicológicos de los pesticidas. [7]

En la página mencionada anteriormente existen alrededor de 1150 pesticidas, de donde se seleccionaron un total de 159 sustancias con su respectiva información de la dosis letal media (DL_{50}[8]) sobre mamíferos (ratas), necesaria para el desarrollo del modelo. En el ANEXO 1 se presenta la matriz con las moléculas elegidas y su actividad biológica.

Con la finalidad de que la información de la respuesta biológica de los pesticidas no contenga números extremadamente pequeños o grandes, fueron matemáticamente normalizados mediante la siguiente fórmula:

$$respuesta\ biológica = \log\left(\frac{1}{LD_{50}}\right) \quad \textbf{(2)}$$

[7] University of Hertfordshire (2013) The Pesticide Properties DataBase (PPDB) developed by the Agriculture & Environment Research Unit (AERU), University of Hertfordshire, 2006-2013.
[8] Dosis requerida para matar la mitad de la población que reciben la sustancia en estudio. *Toxicología ambiental*. México: Universidad Autónoma de Aguascalientes. (2009).

Posteriormente se calculó la IC_{50}[9] mediante de la fórmula:

$$respuesta\ biológica\ (IC_{50}) = 10^{\left(\log\frac{1}{LD_{50}}\right)} \quad \textbf{(3)}$$

Con estos valores realizamos la asignación hipotética de las clases para la obtención del modelo.

3.2. Digitalización y Optimización de las moléculas de pesticidas

Para la digitalización de las estructuras moleculares de los pesticidas seleccionados se utilizó el programa HYPERCHEM 8.0. (Hypercube, 2008), que nos permite visualizar y analizar las estructuras de las moléculas.

La optimización de las moléculas se realizó con una serie procedimientos apoyados en las herramientas incorporadas en el programa; en primer lugar se optimizaron geométricamente con mecánica molecular en la herramienta MM+, como se muestra en las Ilustración 7 y 8.

[9]Indica la cantidad de un medicamento en particular u otra sustancia que se necesita para inhibir un proceso biológico dado por la media. *Toxicología ambiental*. México: Universidad Autónoma de Aguascalientes. (2009)

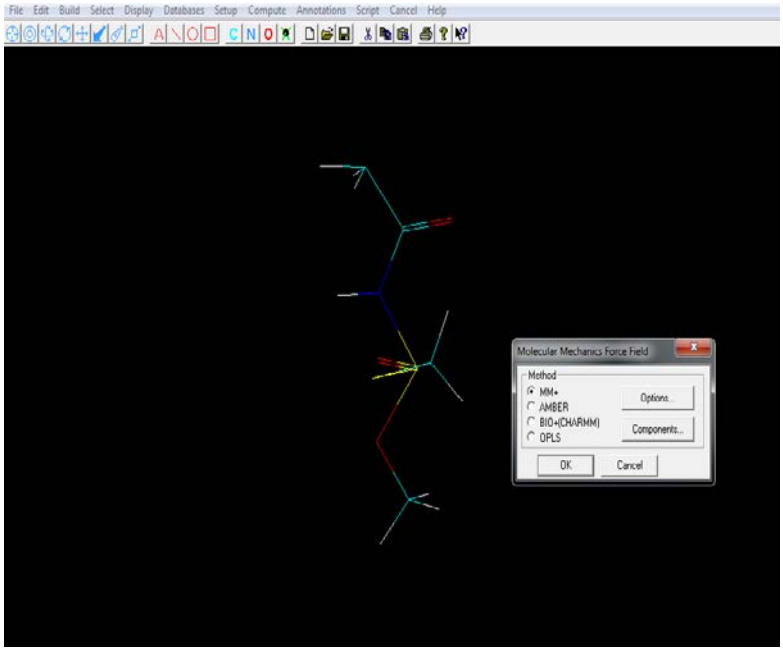

***Ilustración 6**: Uso de la herramienta MM+*

***Fuente:** Software Hyperchem 8.0.*

Luego de esto se refinaron con el algoritmo Polak-Ribiere hasta que el valor de la desviación RMS (raíz cuadrada media) sea menor a 0,001 kcal/[Ãmol], este proceso se realizó con cada una de las 159 moléculas.

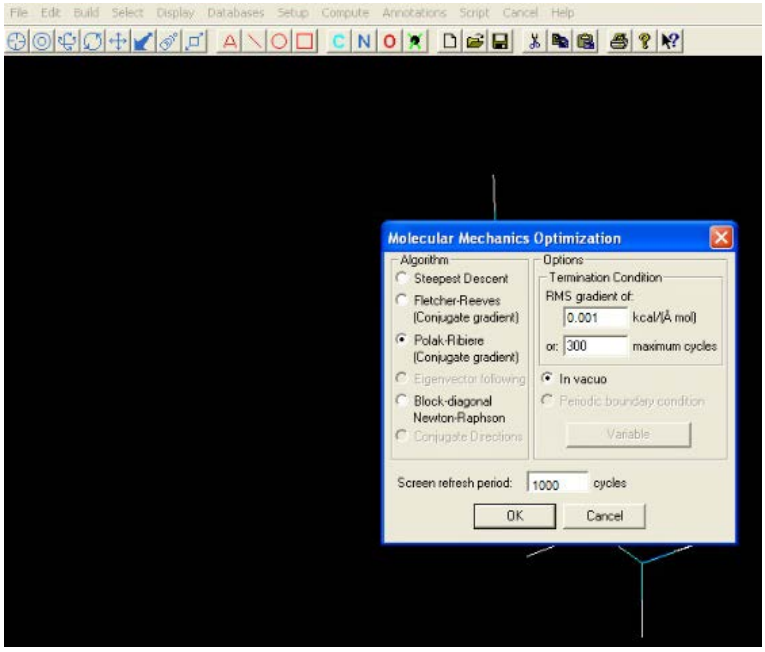

***Ilustración 7:** Ajuste del Algoritmo para optimización*

***Fuente:** Software Hyperchem 8.0.*

3.3. Cálculo de los Descriptores Moleculares

Una vez digitalizadas y optimizadas las moléculas de los pesticidas se ingresaron al programa DRAGON 6.0, que calculó 1502 descriptores moleculares para cada una de las moléculas.

Durante el proceso de cálculo el software eliminó aquellos descriptores constantes para evitar información redundante, descriptores con información faltante, o que están correlacionados con más del 95%. (CONSONNI, 2009). Así la matriz inicial de 1502 DM se reduce a 159 moléculas por 254 descriptores, es decir una matriz principal de 40386 datos.

Tabla 2: Dimensión de la matriz de los datos

	Descriptores Moleculares			
Molécula	**D 1**	**D 2**	**D 3**	**D 254**
1	X1	X2	X3	X 254
2				
3				
N	Xn	Xn	Xn	Xn
159		159*254 =		**40386 datos**

Nota. Reducción de la dimensión de la base de datos. Fuente: La autora

3.4. Auto-escalado

El auto-escalado de los datos de la matriz se realizó con la finalidad que los datos contenidos en la matriz de trabajo sean homogéneos, para esto se aplicó la siguiente fórmula:

$$\text{Matriz autoescalada} = \left(columna\ descriptor - \frac{\bar{x}\ descriptor}{\sigma\ descriptor}\right) \quad \textbf{(3)}$$

3.5. Asignación de las Clases

Para el proceso de asignación de las clases la respuesta biológica (IC_{50}) de las moléculas de pesticidas, fueron sometidas a un proceso de clasificación mediante el método del algoritmo k-means este utiliza la media ponderada de los datos para su agrupación en dos clases: los de clase 1 de baja toxicidad, y los de clase 2 de alta toxicidad.

3.6. Selección de las variables

El proceso de selección de los descriptores moleculares óptimos (variables) para la construcción del modelo se realizó mediante la aplicación de Algoritmos Genéticos incorporados en el abanico de herramientas "ga_toolbox" de Matlab®, creadas por Milano Chemometrics and QSAR Research Group de la Universidad Milano Bicocca (Italia)[10].

En Matlab se utilizó el *k-NN* o vecinos cercanos, este se programó con los parámetros necesarios para obtener el modelo de clasificación idóneo, que se indican en la Tabla 3.

Tabla 3: *Parámetros de programación del Algoritmo Genético*

Método	KNN
Escalado	None
Grupos de cross-validación	5
Tipo de cross-validación	Ventanas venecianas
Número de cromosomas	30
Probabilidad de mutación	0.01
Probabilidad de cruce	0.5
Repeticiones	100
Número de evaluaciones	350

Nota. Fuente: La autora

[10] http://michem.disat.unimib.it/chm/contacts/contacts.htm

Este proceso de selección de variables se realizó tres veces, y para la obtención de modelo final se unieron en un solo conjunto todos los descriptores seleccionados anteriormente y se les aplicó nuevamente los algoritmos genéticos y el método de clasificación CART.

3.7. Selección Training – Test

En el mismo software Matlab se realizó la partición de los datos, que se escogen totalmente al azar, separándolos en dos grupos: 60% para el Training Set que son los usados para construir el modelo y corresponden a un total de 96 datos y, el 40% para Test Set que son los datos usados para la validación del modelo obtenido con los datos anteriores. Aquí se consideraron únicamente los descriptores que fueron agrupados después de las tres repeticiones realizadas en el paso anterior.

CAPITULO IV

RESULTADOS

A continuación se presentan los resultados obtenidos de la aplicación de la metodología para la obtención de un modelo que relacione la estructura-actividad de las moléculas de los pesticidas.

4.1. Resultados

Una vez terminado el cálculo del modelo con los datos del Training set que se muestran en el ANEXO 2, se procede a evaluar los parámetros de clasificación, esta evaluación se basa en el NER que se ha obtenido, si este es satisfactorio se acepta el modelo. En la Ilustración 9 se muestra el NER% en función de las variables elegidas.

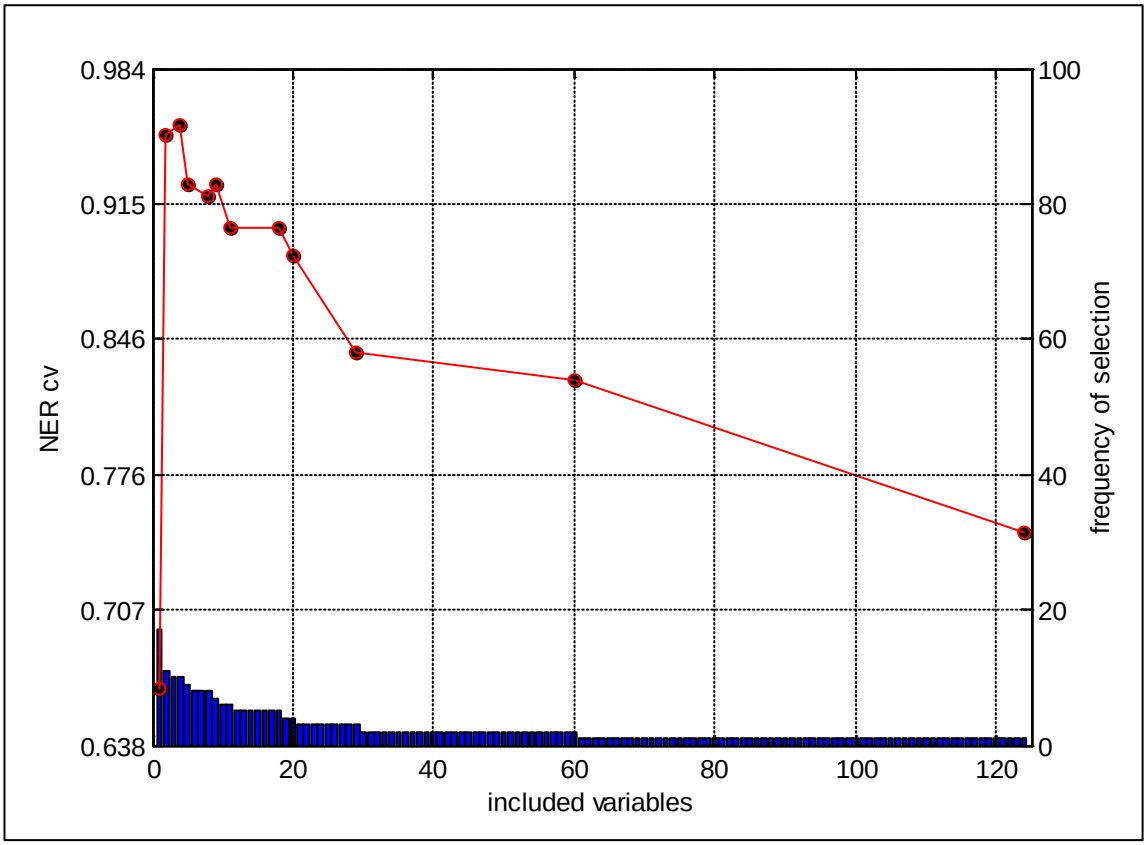

***Ilustración 8**: Relación entre el número de descriptores considerados en la clasificación y la calidad del modelo en predicción*

***Fuente:** Matlab R2013a*

El modelo se obtuvo con 2 variables de las 254 iniciales, que fueron seleccionadas mediante la aplicación de los algoritmos genéticos y luego se aplicó el método de CART para la construcción del modelo. Con este obtuvimos un valor NER de 0.99; es decir que la capacidad del modelo en predicción es del 99%.

Las variables elegidas por el modelo se presentan a continuación en la Ilustración 10, y estas fueron la primera y segunda variable que corresponden a los descriptores: nN que son el número de nitrógenos y, ATSC2e indica autocorrelación de electronegatividad de las moléculas, respectivamente.

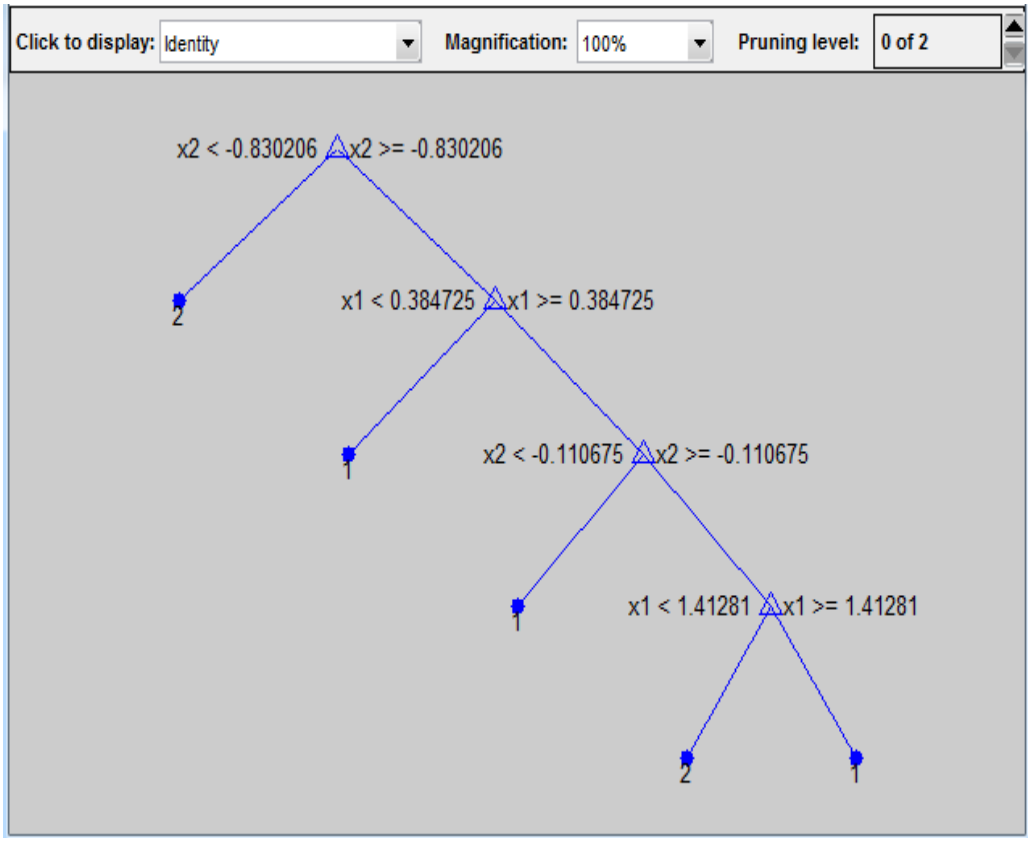

Ilustración 9: *Modelo de árbol obtenido.*

Fuente*: Autora*

Después procedemos a calcular y graficar los centroides de cada descriptor molecular seleccionado para cada una de las clases, esto nos permite observar el comportamiento de los descriptores y su aporte a las clases. Lo podemos observar en la Tabla 4 e Ilustración 10.

Tabla 4: *Cálculo de centroides de cada descriptor molecular elegido*

Descriptores / Clases	nN	ATSC2e
Clase 1	-0,06078011	0,25574524
Clase 2	0,18481897	-0,59660113

Nota. Fuente: Autora

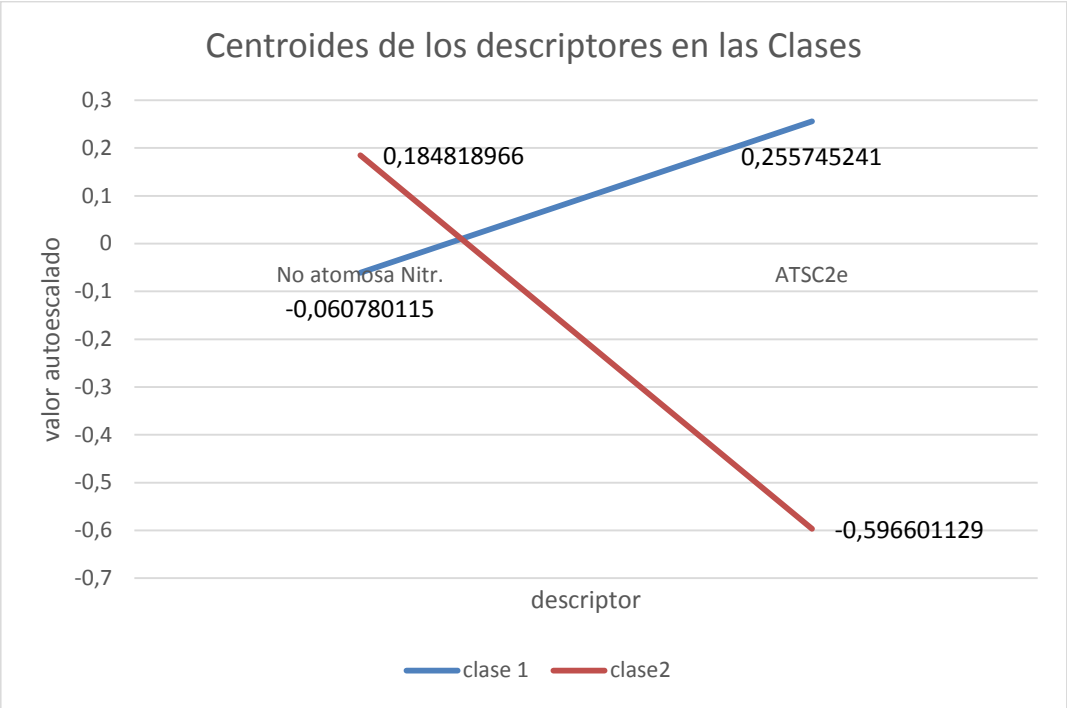

Ilustración 10: *Comportamiento de los centroides.*

Fuente: *Autora*

A partir de la Ilustración 11 se puede decir que las moléculas de pesticidas asignadas a la clase 1 tienen menor número de átomos de nitrógeno y un alto valor de auto-correlación de electronegatividad; mientras que las de clase 2 poseen mayor número de átomos de nitrógeno y un bajo valor de auto-correlación.

Se procedió a evaluar la asignación las moléculas del Training Set en las clases preestablecidas obtenidas mediante el método de clasificación CART con el apoyo

del software Matlab, las cuales quedaron de la siguiente manera: 62 en la clase 1 y, 34 en la clase 2.

En la Tabla 5 tenemos la matriz de confusión que nos facilita la visualización de los objetos que fueron asignados correctamente a las clases.

Tabla 5: *Matriz de confusión del modelo*

Asignadas / Real	clase 1	clase 2
clase 1	56	4
clase 2	6	30

Nota. Fuente: Autora

De lo anterior decimos que los valores que se encuentran en la diagonal son aquellos que se están correctamente asignados, y los que están fuera de ella han sido asignados a una clase pero deberían pertenecer a una otra. Esto también nos permite obtener el NER del modelo en la asignación de las clases, lo que hace sumando los valores de la diagonal y dividirlos para el número total de objetos en la matriz, que para este caso nos da un valor de 0,90.

Se obtuvo además otros parámetros de calidad del modelo: especificidad y sensibilidad. La especificidad se refiere a la pureza de la clase es decir la capacidad del modelo para aislar las moléculas de una clase de las otras; y la sensibilidad es la capacidad de representar las moléculas en cada clase (TODESCHINI R. , 2003). Estos parámetros se pueden observar en la Tabla 6.

Tabla 6: *Parámetros de calidad del modelo obtenidos con el método CART*

Clase	**Especificidad**	**Sensibilidad**
1	1,00	0,98
2	0,98	1,00

Nota. Fuente: Autora

De los valores expresados en la Tabla 6 decimos que la especificidad del modelo es mayor para la clase 1 con un 100% en relación a la clase 2 tiene 98%, pero también que el modelo es mucho más sensible para la clase 2 con un 100% de sensibilidad frente a un 98% de la clase 1.

4.2. Validación del modelo

Para la validación del modelo se usaron los datos del Test set que se muestran en el ANEXO 3, se aplicó el mismo procedimiento usado para el modelo; Aquí obtuvimos un valor del NER de 0,90 es decir que la capacidad del predicción del modelo es del 90%.

Los resultados obtenidos por el modelo en la asignación de las clases fue: 44 moléculas a la clase 1 y, 19 a la clase 2; que se asignaron al igual que en la evaluación del modelo por el método de clasificación CART, esta información se puede observar en la Tabla 7. El NER obtenido para la clasificación de los datos de validación fue de 0,92.

Tabla 7: *Matriz de confusión de validación del modelo*

Asignadas / Real	clase 1	clase 2
clase 1	39	0
clase 2	5	19

Nota. Fuente: Autora

Al igual que en la evaluación del modelo se obtuvieron otros parámetros de calidad: especificidad y sensibilidad, la información que nos proporciona la Tabla 8 nos indica que la especificidad del modelo es de 79% para la clase 1 y de 100% para la clase 2, mientras que la sensibilidad es de 100% para la clase 1 y 79% para la clase 2.

Tabla 8: *Parámetros de calidad del modelo obtenidos por CART*

Clase	**Especificidad**	**Sensibilidad**
1	0,79	1,00
2	1,00	0,79

Nota. Fuente: Autora

CONCLUSIONES

- De las 254 variables calculadas para describir la toxicidad, luego de aplicar la selección de variables mediante la aplicación de algoritmos genéticos, se seleccionaron dos variables que describen la toxicidad en el presente estudio. Esto nos evidencia que los algoritmos genéticos son una herramienta potente para la selección de variables en estudios de toxicidad.

- El NER encontrado con el modelo CART tuvo un valor de 0,99 lo que significa que la capacidad de predicción del modelo es buena. El modelo se obtuvo con 2 variables que son: nN y ATSC2e

- De acuerdo a lo expresado en la matriz de confusión podemos concluir que el modelo es más sensible para las moléculas de clase 2 con un 100% de sensibilidad.

- Con respecto al análisis de los centroides concluimos que se asignan a clase 1 aquellas moléculas cuyo número de átomos de nitrógenos sea menor y su valor de correlación de electronegatividad sea alto, recordando que la clase 1 representa a los pesticidas de baja toxicidad. Los que se asignan a la clase 2 por su parte tienen un mayor número de átomos de nitrógeno y un valor de autocorrelación de electronegatividad bajo, a la clase 2 corresponden los pesticidas de alta toxicidad.

- Para este modelo se utilizó el método de clustering k-means para asignar a las 159 en dos clases en función de su respuesta biológica (toxicidad), la clase 1 de baja toxicidad y la clase 2 de alta toxicidad.

- El modelo encontrado tanto en training y test set resulta muy eficiente, lo que permitiría ingresar moléculas al modelo y calcular su toxicidad.

- La construcción de este tipo de modelos para predecir toxicidad reemplaza el uso de animales de animales de experimentación, evitando repercusiones éticas, además la reducción costes de experimentación y problemas ambientales.

- Las herramientas quimiométricas son de gran ayuda para resolver problemas ambientales como el del presente estudio.

- El modelo construido en esta investigación es aplicable a pesticidas que pertenezcan a las familias: organofosforados y carbamatos; pues debido a la falta de la información de la respuesta biológica de las familias de organoclorados y piretroides no pudieron ser incluidas en este investigación

RECOMENDACIONES

- Se recomienda aumentar el número de moléculas al modelo para observar su comportamiento.

- Ampliar la base de datos de la respuesta biológica (toxicidad) de los pesticidas pertenecientes a otras familias, para que puedan ser incluidas dentro de los modelos.

- Utilizar moléculas externas para comprobar con ellas la validez del modelo, que sean de las mismas familias que las del entrenamiento.

- En esta investigación para la clasificación se utilizó el método de k-NN o k-vecinos cercanos, se recomienda utilizar otro método de clasificación que permita que la capacidad del modelo mejore.

- Impulsar el uso de la metodología QSAR y las herramientas Quimiométricas en el área Ambiental, útiles en la determinación de sustancias peligrosas.

ANEXOS

ANEXO 1: MOLÉCULAS DE PESTICIDAS Y SU ACTIVIDAD BIOLÓGICA

ID	Familia de Pesticida	Nombre	Actividad Biológica	IC50
			Mamíferos - Agudos oral LD_{50} (mg kg^{-1})	
1	Fosforados	Acefato	945	0,0010582
2	Fosforados	Anilofos	472	0,00211864
3	Fosforados	Azametifos	1180	0,00084746
4	Fosforados	Azinfos-etil	12	0,08333333
5	Fosforados	Azinfos-metil	9	0,11111111
6	Fosforados	Bensulida	270	0,0037037
7	Fosforados	bromofos	1600	0,000625
8	Fosforados	bromofos-etil	52	0,01923077
9	Fosforados	Butonato	1100	0,00090909
10	Fosforados	Cadusafos	30,1	0,03322259
11	Fosforados	carbofenotion	10	0,1
12	Fosforados	chlorthiophos	10	0,1
13	Fosforados	Cianofos	610	0,00163934
14	Fosforados	cloretoxifos	44	0,02272727
15	Fosforados	clorfenvinfos	12	0,08333333
16	Fosforados	Clorfoxim	2500	0,0004
17	Fosforados	Clormefos	7	0,14285714

18	Fosforados	clorpirifos	64	0,015625
19	Fosforados	clorpirifos-metil	2814	0,00035537
20	Fosforados	Crotoxifos	66	0,01515152
21	Fosforados	demeton-o	2,5	0,4
22	Fosforados	demeton-s	2,5	0,4
23	Fosforados	demeton-S-metil	40	0,025
24	Fosforados	Diazinon	1139	0,00087796
25	Fosforados	diclofention	172	0,00581395
26	Fosforados	Diclorvos	80	0,0125
27	Fosforados	Dicrotofos	17	0,05882353
28	Fosforados	dimethylvinphos	97,5	0,01025641
29	Fosforados	Dimetoato	245	0,00408163
30	Fosforados	dioxabenzophos	125	0,008
31	Fosforados	dioxation	23	0,04347826
32	Fosforados	disulfoton	2,6	0,38461538
33	Fosforados	Edifenfos	150	0,00666667
34	Fosforados	EPN	14	0,07142857
35	Fosforados	Etion	208	0,00480769
36	Fosforados	etoprop	32,9	0,03039514
37	Fosforados	etrimfos	1800	0,00055556
38	Fosforados	Fenamifos	6	0,16666667
39	Fosforados	fenclorfos	500	0,002
40	Fosforados	Fenitrotion	330	0,0030303

41	Fosforados	fenkaptona	44	0,02272727
42	Fosforados	fonofos	6,8	0,14705882
43	Fosforados	forato	2	0,5
44	Fosforados	Formotion	365	0,00273973
45	Fosforados	Fosacetim	3,7	0,27027027
46	Fosforados	Fosalon	120	0,00833333
47	Fosforados	Fosamina	5000	0,0002
48	Fosforados	fosfamidon	7	0,14285714
49	Fosforados	fosmet	113	0,00884956
50	Fosforados	Fostiazato	57	0,01754386
51	Fosforados	foxim	2000	0,0005
52	Fosforados	Heptenofos	96	0,01041667
53	Fosforados	iodofenfos	2330	0,00042918
54	Fosforados	iprobenfos	680	0,00147059
55	Fosforados	Isazofos	27	0,03703704
56	Fosforados	isofenfos	28	0,03571429
57	Fosforados	isoxation	112	0,00892857
58	Fosforados	leptophos	43	0,02325581
59	Fosforados	malation	1778	0,00056243
60	Fosforados	mecarbam	36	0,02777778
61	Fosforados	mefosfolan	9	0,11111111
62	Fosforados	merfos	910	0,0010989
63	Fosforados	metamidofos	30	0,03333333

64	Fosforados	metidation	25	0,04
65	Fosforados	mevinfos	3,5	0,28571429
66	Fosforados	monocrotofos	14	0,07142857
67	Fosforados	naled	83	0,01204819
68	Fosforados	ometoato	50	0,02
69	Fosforados	oxidemeton metil	48	0,02083333
70	Fosforados	paratión	2	0,5
71	Fosforados	paratión-metil	3	0,33333333
72	Fosforados	piperophos	324	0,00308642
73	Fosforados	piraclofos	237	0,00421941
74	Fosforados	piridafention	769	0,00130039
75	Fosforados	pirimifos-etil	140	0,00714286
76	Fosforados	pirimifos-metil	1414	0,00070721
77	Fosforados	profenofos	358	0,0027933
78	Fosforados	protioate	8	0,125
79	Fosforados	protiofos	925	0,00108108
80	Fosforados	quinalfos	71	0,01408451
81	Fosforados	Sulfotep	5	0,2
82	Fosforados	sulprofos	176	0,00568182
83	Fosforados	tebupirimfos	1,3	0,76923077
84	Fosforados	Temefos	4204	0,00023787
85	Fosforados	terbufos	1,3	0,76923077
86	Fosforados	tetraclorvinfos	4000	0,00025

87	Fosforados	tetraethyl pyrophosphate	1,12	0,89285714
88	Fosforados	thionazin	3,5	0,28571429
89	Fosforados	tiometon	40	0,025
90	Fosforados	triazofos	66	0,01515152
91	Fosforados	tribufos	234	0,0042735
92	Fosforados	triclorfon	212	0,00471698
93	Fosforados	vamidotion	64	0,015625
94	Carbamatos	Alanycarb	330	0,0030303
95	Carbamatos	Aldicarb	0,93	1,07526882
96	Carbamatos	aldoxicarb	27	0,03703704
97	Carbamatos	aminocarb	30	0,03333333
98	Carbamatos	Asulam	5000	0,0002
99	Carbamatos	Barban	527	0,00189753
100	Carbamatos	bendiocarb	34	0,02941176
101	Carbamatos	benfuracarb	205	0,00487805
102	Carbamatos	benthiavalicarb isopropyl	500	0,002
103	Carbamatos	bufencarb (Ethylpropyl)	61	0,01639344
104	Carbamatos	bufencarb (Methylbuthyl)	61	0,01639344
105	Carbamatos	Butilato	3500	0,00028571
106	Carbamatos	butocarboxim	153	0,00653595
107	Carbamatos	butoxicarboxim	458	0,00218341
108	Carbamatos	Carbaril	614	0,00162866
109	Carbamatos	carbetamida	1718	0,00058207

110	Carbamatos	carbofurano	7	0,14285714
111	Carbamatos	carbosulfan	101	0,00990099
112	Carbamatos	carbutilato	3000	0,00033333
113	Carbamatos	chinometionato	5000	0,0002
114	Carbamatos	cicloato	2000	0,0005
115	Carbamatos	clororofam	4200	0,0002381
116	Carbamatos	dazomet	415	0,00240964
117	Carbamatos	desmedifam	5000	0,0002
118	Carbamatos	dialato	395	0,00253165
119	Carbamatos	dietofencarb	5000	0,0002
120	Carbamatos	dimepiperato	946	0,00105708
121	Carbamatos	dimexano	240	0,00416667
122	Carbamatos	dioxacarb	25	0,04
123	Carbamatos	EPTC	916	0,0010917
124	Carbamatos	esprocarb	2000	0,0005
125	Carbamatos	etiofencarb	200	0,005
126	Carbamatos	fenmedifam	8000	0,000125
127	Carbamatos	fenobucarb	620	0,0016129
128	Carbamatos	fenotiocarb	1200	0,00083333
129	Carbamatos	fenoxicarb	10000	0,0001
130	Carbamatos	ferbam	4000	0,00025
131	Carbamatos	furatiocarb	53	0,01886792
132	Carbamatos	iprovalicarb	5000	0,0002

133	Carbamatos	Isolan	10,8	0,09259259
134	Carbamatos	isoprocarb	403	0,00248139
135	Carbamatos	methasulfocarb	112	0,00892857
136	Carbamatos	metiocarb	19	0,05263158
137	Carbamatos	metolcarb	268	0,00373134
138	Carbamatos	metomilo	30	0,03333333
139	Carbamatos	mexacarbato	14	0,07142857
140	Carbamatos	molinato	483	0,00207039
141	Carbamatos	orbencarb	800	0,00125
142	Carbamatos	Oxamilo	2,5	0,4
143	Carbamatos	pebulato	1120	0,00089286
144	Carbamatos	pirimicarb	142	0,00704225
145	Carbamatos	profam	5000	0,0002
146	Carbamatos	promecarb	35	0,02857143
147	Carbamatos	propoxur	50	0,02
148	Carbamatos	prosulfocarb	1820	0,00054945
149	Carbamatos	thiocarboxime	12,6	0,07936508
150	Carbamatos	tiobencarb	560	0,00178571
151	Carbamatos	tiocarbazil	10000	0,0001
152	Carbamatos	tiodicarb	50	0,02
153	Carbamatos	tiofanox	8,5	0,11764706
154	Carbamatos	Tiram	1800	0,00055556
155	Carbamatos	trialato	1100	0,00090909

156	Carbamatos	trimetacarb	566	0,00176678
157	Carbamatos	vernolato	1500	0,00066667
158	Carbamatos	XMC	542	0,00184502
159	Carbamatos	ziram	267	0,00374532

***Fuente**: Autora*

ANEXO 2: MOLÉCULAS DEL TRAINING SET

17	242	27	173	72
nN	ATSC2e	N%	SM1_Dz(i)	MAXDN
-0,12931939	-0,11899325	-0,32809105	-0,00831327	0,25201674
0,89876978	1,3824562	1,11369601	1,42329614	1,09614588
1,92685896	-0,21465346	1,31030334	0,74338725	0,2270425
1,92685896	0,0432132	2,03119687	0,74338725	0,23303632
-1,15740857	0,43417234	-1,14728825	0,09303961	0,65659934
-1,15740857	-0,63056744	-1,14728825	-1,50326824	-0,00871428
-1,15740857	0,3676261	-1,14728825	-0,25747243	0,65659934
-0,12931939	0,21789707	0,16342726	-0,1265583	0,50475599
-1,15740857	1,6028906	-1,14728825	0,60825007	0,75150144
0,89876978	0,28860245	0,78601713	0,73071814	0,798453
-0,12931939	0,82513147	-0,03318006	0,81517887	0,74650659
-1,15740857	1,97305404	-1,14728825	0,78139458	1,48874086
-1,15740857	0,31355728	-1,14728825	-1,05562636	0,26800025
-1,15740857	0,19294223	-1,14728825	-1,05562636	0,68457049
-1,15740857	0,33435298	-1,14728825	-1,05562636	0,64660965
0,89876978	0,4508089	0,49110614	0,43510557	0,59666118
-1,15740857	1,44900243	-1,14728825	0,599804	1,3079274
-1,15740857	-0,51827066	-1,14728825	-2,01847871	-0,14857
-1,15740857	-0,9300255	-1,14728825	-1,6721897	0,65759831
-0,12931939	-0,48499755	-0,1970195	0,31263751	0,49776321
-1,15740857	-0,67631797	-1,14728825	-1,6721897	0,24802086
0,89876978	1,11211211	0,72048136	0,81940191	0,61264469
-0,12931939	0,44664976	-0,36085894	-0,1265583	1,125116
-1,15740857	-1,44575883	-1,14728825	-2,37321378	-0,49820929
-1,15740857	-0,53490722	-1,14728825	-2,01847871	-0,15156691
-0,12931939	0,29276159	0,09789149	0,09726264	0,2270425
0,89876978	-0,57233948	0,72048136	0,63781133	0,80444682
-0,12931939	0,51735513	-0,22978739	0,41821343	0,27399406
-0,12931939	0,20541965	0,78601713	0,51534327	1,95226266
-0,12931939	-0,09403841	-0,1970195	-0,36727139	0,84440559
-1,15740857	1,34502394	-1,14728825	0,29574536	1,28694904
-1,15740857	0,46744545	-1,14728825	-0,23213421	0,6056519
-0,12931939	-0,41845131	-0,42639472	-0,1265583	0,49176939
0,89876978	0,58390137	1,27753545	0,43088254	0,18009094
-1,15740857	2,21428414	-1,14728825	0,78139458	1,40782434

-1,15740857	1,79421103	-1,14728825	0,8067328	1,40682537
-0,12931939	0,6504476	0,22896304	0,31263751	0,92032727
-1,15740857	0,44664976	-1,14728825	-0,51507767	0,79345815
-0,12931939	0,76274438	-0,13148373	0,70960295	0,64461171
-0,12931939	0,87504115	0,09789149	0,70960295	0,60665087
-0,12931939	-0,14394809	-0,49193049	-0,36727139	0,20306723
0,89876978	0,13055514	0,49110614	0,73071814	1,02921493
0,89876978	0,58806051	0,52387403	0,81940191	0,74051277
1,92685896	0,84176803	1,04816023	0,92920086	0,63162511
1,92685896	1,0164519	1,37583911	0,92920086	0,59366427
-0,12931939	-0,37270077	-0,22978739	-0,36727139	0,15611567
-1,15740857	-0,73870507	-1,14728825	-0,76423683	0,20606414
0,89876978	0,3468304	0,78601713	0,43510557	0,61664057
-1,15740857	2,14773791	-1,14728825	-0,23635725	0,69955503
0,89876978	0,25532933	0,39280248	0,43510557	0,62762923
-1,15740857	-0,46420185	-1,14728825	-2,01847871	-0,09462565
-1,15740857	4,1566024	-1,14728825	0,99254641	1,74048115
-1,15740857	-0,45588357	-1,14728825	-2,01847871	-0,18653084
-1,15740857	-1,15877818	-1,14728825	-2,70261064	-0,1126071
-1,15740857	1,57793576	-1,14728825	0,88697049	1,56166563
-0,12931939	0,76274438	-0,1970195	0,09726264	0,79745403
1,92685896	0,27612503	0,78601713	1,25015164	-0,72397638
0,89876978	-0,68463625	0,98262446	0,40554432	-0,85683931
-0,12931939	-0,56818034	0,16342726	0,49422809	-0,7459537
0,89876978	0,18878309	-0,03318006	1,31772023	-0,80289496
-0,12931939	-0,8676384	-0,1970195	-0,16034259	-0,8798156
-0,12931939	-0,8676384	-0,1970195	-0,16034259	-0,87881663
-0,12931939	-1,22948356	-0,26255528	-0,92471222	-1,54013438
0,89876978	-0,78861475	1,37583911	0,19439249	-0,76992897
0,89876978	-0,04828787	1,17923179	0,97565426	1,1381026
0,89876978	-0,38517819	-0,03318006	0,60402704	-0,9287651
1,92685896	-0,17306207	1,24476756	1,25015164	-0,77592278
-0,12931939	-1,26691582	-0,1970195	-0,92471222	-1,58808491
0,89876978	0,00994008	0,58940981	1,14879876	-0,69800317
-0,12931939	-1,31682549	-0,13148373	-0,15189652	-1,34733328
-0,12931939	-0,46836099	-0,32809105	0,6800417	-0,82487229
-0,12931939	-1,28355238	-0,29532317	-0,92471222	-1,53414056
-0,12931939	-1,2544384	-0,36085894	-0,92471222	-1,49118488
-0,12931939	-0,9092298	-0,06594795	-0,40105568	-0,88181354

-0,12931939	-1,12966421	-0,22978739	-0,40105568	-1,42824981
0,89876978	-0,49331583	0,13065938	0,79406369	-0,73696298
-0,12931939	-0,95913948	-0,03318006	-0,16034259	-0,87681869
-0,12931939	0,01825836	0,0651236	0,27885322	1,4417893
-0,12931939	-0,87179754	-0,06594795	-0,40105568	-0,86982591
-0,12931939	-0,98409432	0,26173093	-0,16034259	-0,86483106
0,89876978	-0,79277389	2,12950053	0,19439249	-0,76193721
0,89876978	-0,65136314	0,78601713	0,40554432	-0,86183415
-0,12931939	-1,31682549	-0,03318006	-0,92471222	-1,58608697
-0,12931939	-1,38337173	-0,13148373	-0,51930071	-1,42425393
-0,12931939	-1,25859754	-0,1970195	-0,92471222	-1,55611789
2,95494813	0,0848046	2,58825096	1,35150452	-0,86083518
-0,12931939	-1,03816313	0,09789149	-0,16034259	-0,89380117
-0,12931939	-0,89259324	-0,13148373	-0,16034259	-0,87382179
-0,12931939	-1,2752341	-0,29532317	-0,92471222	-1,51416118
1,92685896	-0,76781905	2,94869772	0,71804903	-0,71198874
2,95494813	0,61717448	2,22780419	1,50775688	-0,59810623
0,89876978	-0,78445561	0,91708868	0,19439249	-0,77492382
-0,12931939	-0,9508212	-0,13148373	0,18594641	-1,29238997
-0,12931939	-0,92170722	-0,03318006	-0,16034259	-0,87182385
-0,12931939	-0,96329862	0,09789149	-0,16034259	-0,86982591
0,89876978	-1,22116528	1,47414278	-2,45345148	-2,14950572

***Fuente**: Autora*

ANEXO 3: MOLÉCULAS DEL TEST SET

17	242	27	173	72
nN	ATSC2e	N%	SM1_Dz(i)	MAXDN
-0,129319393	-0,26456314	0,49110614	-0,1265583	0,74850453
-0,129319393	0,21789707	-0,4591626	-0,13078134	1,4288027
1,926858956	0,33019384	2,03119687	2,28901865	-0,17654114
-1,157408568	2,23923898	-1,14728825	1,23325949	1,65656772
-1,157408568	-0,60977174	-1,14728825	-1,6721897	0,35990543
-1,157408568	1,54050351	-1,14728825	0,88697049	1,56066666
-1,157408568	-0,43924701	-1,14728825	-1,19498657	-0,06965141
-0,129319393	0,82513147	0,26173093	0,81517887	0,70854575
-1,157408568	1,35334222	-1,14728825	0,599804	1,43779342
-0,129319393	1,88987124	-0,09871584	0,89541656	1,38984289
-1,157408568	1,64864114	-1,14728825	0,88697049	1,52270582
-0,129319393	-0,09403841	0,22896304	-0,36727139	0,08818575
-1,157408568	0,08064546	-1,14728825	-0,77690594	0,32194459
-1,157408568	1,54882179	-1,14728825	-0,23635725	0,32194459
-1,157408568	0,68372072	-1,14728825	-1,30900855	0,08219194
-1,157408568	0,66708416	-1,14728825	0,30419144	0,64561068
-0,129319393	0,9332691	-0,03318006	0,70960295	0,62663026
-1,157408568	-0,50163411	-1,14728825	-1,04295725	0,05921564
-0,129319393	1,97721318	-0,26255528	1,16146787	1,51870994
-0,129319393	0,12223686	-0,09871584	0,09726264	0,30696005
0,898769782	0,14303255	0,78601713	0,43510557	0,74051277
-1,157408568	0,0432132	-1,14728825	-0,77690594	0,90833963
1,926858956	1,16618093	1,6707501	1,19525217	0,66958595
-0,129319393	0,36346696	-0,22978739	0,31263751	0,63661996
-1,157408568	0,10144116	-1,14728825	0,30841447	1,32191297
-1,157408568	1,91066694	-1,14728825	0,96298515	0,93131593
-0,129319393	0,44249062	-0,32809105	0,09726264	0,20606414
-0,129319393	0,59221965	-0,09871584	-0,36727139	1,0761665
-1,157408568	-1,5372599	-1,14728825	-3,28961273	-1,61905296
-0,129319393	0,48824115	1,04816023	0,09726264	0,98226337
-0,129319393	1,72766479	0,03235571	0,89541656	1,37985319
-1,157408568	0,03073578	-1,14728825	-0,2532494	1,01023451
-1,157408568	-0,68463625	-1,14728825	-2,01847871	0,0672074
-1,157408568	2,47215081	-1,14728825	-0,00409024	0,55470447
-1,157408568	0,86256373	-1,14728825	0,79828672	1,08116134

0,898769782	0,44664976	1,17923179	0,43510557	0,53872095
1,926858956	0,7544261	1,57244644	0,92920086	0,65260347
0,898769782	-0,77197819	1,37583911	0,19439249	-0,75694237
0,898769782	-0,03581046	1,17923179	0,97565426	1,20703149
0,898769782	0,11807772	1,47414278	0,97565426	1,80641313
-0,129319393	-0,14394809	-0,03318006	0,6800417	-0,76893
1,926858956	-0,28951798	0,81878502	1,73580085	-0,68201966
-0,129319393	-1,03400399	0,09789149	-0,16034259	-0,85683931
0,898769782	-0,62224916	0,85155291	0,79406369	-0,66603615
-0,129319393	-0,83436529	-0,09871584	0,27885322	-0,82487229
0,898769782	-1,63292011	1,96566109	-0,50240856	-1,36131886
-0,129319393	-0,91338894	0,09789149	0,17750034	-0,83885786
0,898769782	-1,22532442	2,29333997	-1,03873421	-2,31533464
-1,157408568	-0,9508212	-1,14728825	-2,07337818	-1,75691074
-0,129319393	0,25532933	-0,03318006	0,6800417	-0,79989805
-0,129319393	-1,28355238	-0,09871584	-0,92471222	-1,5461282
0,898769782	0,11391858	0,58940981	1,14879876	-0,69900214
-0,129319393	-0,9092298	-0,13148373	-0,16034259	-0,88381148
-0,129319393	-0,54738464	-0,36085894	0,6800417	-0,87681869
1,926858956	-0,72622765	1,50691066	-3,08268394	-1,20048478
0,898769782	0,63381104	0,09789149	1,31772023	-0,70299802
1,926858956	-0,46420185	1,9328932	0,90386264	-0,91677747
1,926858956	-0,19385777	2,4899473	1,08123018	-0,60410005
-0,129319393	-0,77197819	-0,06594795	0,27885322	-0,8048929
-0,129319393	-1,38337173	-0,13148373	-0,51930071	-1,44723023
-0,129319393	-1,23780184	-0,39362683	-0,92471222	-1,5101653
0,898769782	-1,22948356	1,57244644	-1,65529756	-2,14550984
-0,129319393	-1,25859754	-0,1970195	-0,92471222	-1,56311068

***Fuente**: Autora*

ANEXO 4: MOLÉCULAS ASIGNADAS A CADA CLASE

ID MOLÉCULA	MOLÉCULA DE PESTICIDA	CLASE
3	Azametifos	2
5	Azinfos-metil	1
7	bromofos	2
12	carbofenotion	2
13	cianofos	1
14	cloretoxifos	1
16	Clorfoxim	2
17	clormefos	1
18	clorpirifos	1
21	demeton-o	1
23	demeton-S-metil	1
24	diazinon	2
28	dimethylvinphos	1
29	dimethylvinphos	2
30	dioxabenzophos	1
31	dioxation	1
32	disulfoton	1
33	edifenfos	1
34	EPN	1
35	EPN	2
36	etoprop	1
37	etrimfos	1
38	fenamifos	1
39	fenclorfos	1
40	fenitrotion	1
42	fonofos	2
43	forato	1
44	forato	2
45	fosacetim	1
47	fosalon	2
48	fosfamidon	1
49	fosmet	1
50	fostiazato	1
51	foxim	1
52	heptenofos	1
54	iprobenfos	1
56	isofenfos	1

60	mecarbam	1
61	mefosfolan	1
65	mevinfos	1
68	monocrotofos	2
69	oxidemeton metil	1
70	paratión	1
71	paratión-metil	1
73	piraclofos	1
74	piridafention	2
75	pirimifos-etil	1
77	profenofos	1
78	protioate	1
79	protiofos	1
82	quinalfos	2
84	sulprofos	2
87	tetraethyl pyrophosphate	1
88	thionazin	2
89	tiometon	1
90	triazofos	1
91	tribufos	2
95	aldicarb	1
96	aldoxicarb	1
97	aldicarb	2
98	asulam	2
100	bendiocarb	1
102	bendiocarb	2
106	butocarboxim	1
109	carbetamida	1
111	carbosulfan	1
113	chinometionato	2
114	cicloato	2
116	dazomet	2
118	dialato	1
123	EPTC	2
124	esprocarb	1
125	etiofencarb	2
126	fenmedifam	2
128	fenotiocarb	1
130	fenotiocarb	2
131	furatiocarb	1
133	isolan	1
134	isoprocarb	2
136	metiocarb	2

137	metolcarb	2
139	mexacarbato	1
140	molinato	1
141	orbencarb	2
142	oxamilo	1
143	pebulato	2
145	profam	2
146	promecarb	1
149	thiocarboxime	1
151	tiocarbazil	1
152	tiodicarb	1
154	tiram	1
155	trialato	2
156	trimetacarb	2
157	vernolato	2
158	XMC	2

***Fuente**: Autora*

BIBLIOGRAFÍA

ALBERT, L. B. *Impacto Ambiental de los plaguicidas en los ecosistemas costeros, Golfo de México Contaminación e Impacto Ambiental: diagnóstico y Tendencias.* Universidad Nacional Autónoma de México, Instituto Nacional de Ecología, 420-530. (2005).

ALMEIDA, W. *Fundamentos toxicológicos de los plaguicidas.* p.63-64. (1982).

BELLO, J., & LÓPEZ de CERAIN, A. *Fundamentos de ciencia toxicológica*. Ediciones Díaz de Santos. (2008).

BOUREL, M. *Métodos de agregación de modelos y aplicaciones* . Memoria de Trabajos de Difusión Científica y Técnica(10), p.19-32. (2012).

CAMEAN, A. *Toxicología avanzada*. Obtenido de http://books.google.es/books?hl=en&lr=lang_es&id=opad2FFk9g0C&oi=fnd&pg=PA37&dq=ensayos+de+toxicidad+de+pesticidas+y+sus+problemas+%C3%A9ticos&ots=V6-..(1995).wRpMeOZ&sig=jwVsN2b1X1sr1rS4wtXlOX6D3so#v=onepage&q=ensayos%20de%20toxicidad%20de%20pesticidas%20y%20sus%. (1995).

CASTILLO Morales, G. *Ensayos toxicológicos y métodos de evaluación de calidad de aguas. Estandarizazción, intercalibración, resultados y aplicaciones.* p.17-20. México. (2004).

CONSONNI, V. *Molecular descriptors and chemometrics: a powerful combined tool for pharmaceutical, toxicological and environmental problems*. p.89-93. Departament of Environmental Sciences, Uviersite of Milano - Bicocca. (2009).

CRUZ, M. *Métodos de correlación estructura-actividad multiobjetivos aplicados al desarrollo racional de fármacos en Cuba.* p.34-37. Editorial Universitaria. Obtenido de http://www.ebrary.com. (2009).

DÍAZ, M., BULUS, G., & PICA, Y. *Métodos estadísticos para el análisis de resultados de toxicidad.* Ensayos Toxicológicos Y Métodos de Evaluación de Calidad de Aguas, p.100 - 101. (IDRC, Ed.). (2004).

ESCALONA, J., CARRASCO, R., & PADRÓN, J. *Introducción al diseño racional de fármacos*. Editorial Universitaria. (2008).

FAO. *Código Internacional de Conducta para la Distribución y Utilización de Plaguicidas (Versión Revisada).* p.7. Obtenido de http://www.fao.org/docrep/006/Y4544S/y4544s00.htm#Contents. (2003).

FERRER, A. *Intoxicación por plaguicidas*. *26*, p.155-171. (2003).

FRANK, I., & TODESCHINI, R. *The Data Analysis Handbook.* p.334-350. Elsevier Science. (2006).

HANSCH, C. *Chem-Bioinformatics: Comparative QSAR at the interface between Chemistry and Biology.* Chem. Rev(102), p.583-812. (2002).

Hypercube, I. *HyperChem (Software for molecular modeling)*, 8.0. (2008).

JIMÉNEZ, V. *Diseño QSAR para inhibidores de acetilcolinesterasa.* p.10-17. Loja, Ecuador: Universidad Particular de Loja. (2013).

JUÁREZ, F. J., RINCÓN Sánchez, A. R., & MARTÍNEZ, R. R. *Toxicología Ambiental.* Universidad Autónoma de Aguascalientes. Obtenido de http://www.ebrary.com. (2009).

LI, B., BAKSHI, B., & GOEL, P. *Other Methods in Nonlinear Regression*. p.465-467. (2009).

MAZZATORTA, P. *QSAR Model for Predicting Pesticide Aquatic Toxicity*. Istituto di Ricerche Farmacolgiche "Mario Negri", 1767-1774. Milano, Italia. (2010).

MENEAU, R. *Métodos Alternativos en Toxicología.* Revista CENIC Ciencias Biológicas, *45*(1), p.18-28. (2014).

MILLA, O. M., & PALOMINO, W. R. *Niveles de colinesterasa sérica en agricultores de la localidad de Carapongo (Perú) y determinación de residuosde plagicidas inibidores de la Acetilcolinesterasa en frutas y hortalizas cultivadas*. (2002).

MONGAY, C. *Quimiometría. Universitat de València.* (2011).

RAMÍREZ, J., & LACASAÑA, M. *Plaguicidas: clasificación, uso, toxicología y medición de la exposición.* Arch Prev Riesgos Labor, *4*(2), p.67 -75. (2001).

REPETTO, M., & REPETTO, G. *Toxicología fundamental*. Ediciones Díaz de Santos. (2009).

RODRÍGUEZ - PIÑERO, P. T. *Introducción a los algoritmos genéticos y sus aplicaciones.* Universidad Rey Juan Carlos, Servicio de Publicaciones. (2003).

ROMÁN, H., MANZANAREZ, J. C., & TAPIA, J. *Optimizacion con Algoritmos Genéticos: Aplicando Cómputo Multi-Núcleo.* Editorial Academia Española. (2012).

SALINAS, J., & VALCARCEL, V. *Aplicación del Algoritmo CART como una técnica*. Universidad Nacional Agraria La Molina. Obtenido de http://repositorio.lamolina.edu.pe/handle/123456789/828. (2002).

SÁNCHEZ, G., & SALCEDA, R. *Enzimas Polifuncionales: El caso de la Acetilcolinesterasa. 27(2)*. (2008).

SEGURA, R., & D'ORS, A. *Animales de experimentación y dermatología: entre la ética y la seguridad.* Más Dermatol. doi:10.5538/1887-5181.2012.18.15. (2012).

SERNA, S. *Comparación de Árboles de Regresión y Clasificación y Regresión Logística*. (2009).

SOLANO, M. *Bioensayo de toxicidad por Microtox®:Movilización de metales pesados en residuos y suleos industriales afectados por la hidrometalurgia del cinc*, 239 - 241. Murcia: Universidad de Murcia. (2005).

TODESCHINI, R. *Introduzione alla Chemiometria.* Napoli: EdiSeS. (2003).

TODESCHINI, R., & CONSONNI, V. *Chemometric in QSAR*. p.129-164. (2009).

VÁZQUEZ, A. *Golfo de México: contaminación e impacto ambiental : diagnóstico y tendencias.* (2005).

VÁZQUEZ, M., *Avances en seguridad alimentaria.* Altaga. (2001).

VEGA, S. *Toxicología I: evaluación epidemiológica de riesgos causados por agentes químicos ambientales*. *4(19)*, p.69. (1985).

VINARDELL MARTÍNEZ - HIDALGO, M. P. A*lternativas a la experimentación animal en toxicología: Situación actual*. *13*(1), p.41 - 52. Recuperado el 20 de Noviembre de 2014, de http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S1726-569X2007000100005. (2007).

WHO. *The WHO recommended classification of pesticides by hazard and guidelines to classification*. p.3-5. (2009).

YOLIS, E., BRITOS, P., PERICHISKY, G., & GARCÍA - MARTÍNEZ, R. *Algoritmos genéticos aplicados a la categorización automática de documentos.* Revista Eletrônica de Sistemas de Informação. doi:10.5329. (2003).

ZAVALA, A., GONZÁLEZ, J., & CHACON, L. *Estudio QSAR por algoritmos genéticos de reconocedores de surco del DNA.* BiolóGicas Revista De La DES Ciencias BiolóGico Agropecuarias Universidad Michoacana De San

NicoláS De Hidalgo, 12(2), p.108-115. Obtenido de http://biologicas.umich.mx/index.php/biologicas/article/view/86/86. (2013).