



**FACULTAD DE MEDICINA**

**Trabajo de titulación previo a la obtención de título de médico**

**Métodos de Correlación Estructura-Actividad Aplicados a Efectos  
Adversos de Fármacos Cardiovasculares, Cuenca, 2018**

**Autores:**

**Geovanny Andrés Torres Soliz**

**Javier Iván Jara Jiménez**

**Director:**

**Dr. Johann Radax**

**Asesor metodológico:**

**Dr. Piercosimo Tripaldi Cappelletti**

**Cuenca, 19 de septiembre de 2019.**

### Abstract

The study of drug development has focused on its effectiveness, without regard to possible adverse effects. However, the drugs' pharmacological, toxicological, pharmacokinetic and pharmaco-dynamic properties should be considered for structural study to apply a rearrangement of the sequential paradigm in the development of new drugs to achieve greater usefulness and greater effectiveness for the patient. For this reason, the QSAR methodology was applied in this correlational study, which has the ability to simultaneously analyze multiple properties by modifying different components involved in the structure-property relationship. After molecule selection and weighting of cardiac side effects, genetic algorithms were applied to calculate the most relevant molecular descriptors. Classification was carried out by using the N3 ("N-nearest neighbor") method. The final mathematical model showed satisfactory sensitivity and specificity (91% and 83%, respectively). In addition, its external validity was verified. The predictive value of the model reached 0.46. This pilot study has led to the development of a mathematical model for the prediction of cardiac adverse effects, which can be used as a basis for future studies within the same research line.



A handwritten signature in purple ink is located in the bottom right corner, above the text "Translated by Geovanny Torres".

Translated by

Geovanny Torres

**Key words:**

QSAR, adverse reactions, mathematical model, prediction, molecular engineerin

**Lista de abreviaciones:**

BNN = Binned Nearest Neighbours (vecinos más cercanos desechados)

C.A.S. = Chemical Abstracts Service (servicio de resúmenes químicos)

C-PANN = Redes Neuronales para identificación y predicción de series de tiempo

GA-TOOLBOX = Algoritmo genético toolbox) (Genetic Algorithm Toolbox)

KNN = k Nearest Neighbours (k vecinos más cercanos)

MedDRA = Medical Dictionary for Regulatory Activities (diccionario médico para actividades regulatorias)

N.E.R = No error rate (porcentaje de los clasificados correctos)

N3 = N-Nearest Neighbours (N vecinos más cercanos)

OLS = Ordinary least squares (Mínimos cuadrados ordinarios)

PCR = Principal component regression (Componente principal de regresión)

PLS = Partial least squares regression (regresión de mínimos cuadrados parciales)

QSAR = Quantitative structure–activity relationship (relación cuantitativa estructura-actividad)

S.M.I.L.E.S. = Simplified Molecular Input Line Entry Specification (especificación de introducción lineal molecular simplificada)

## Introducción

Los fármacos cardiovasculares son un grupo de medicamentos que actúan en el funcionamiento correcto del corazón y de la circulación sanguínea evitando de esta manera el desarrollo o agravamiento de las distintas patologías de índole cardiovascular. Ciertos fármacos son usados solo cuando aparecen los síntomas de una determinada patología, como es el caso de la angina de pecho, no obstante el resto de fármacos se debe tomar a diario y con una frecuencia y dosis establecida por el médico de acuerdo a la gravedad de la patología y así lograr efectos beneficiosos en el paciente; La dosis y frecuencia puede ser regulada por el médico de acuerdo a como responda al tratamiento y hayan reducido los síntomas (1).

No hay que dejar a un lado que los fármacos cardiovasculares pueden llegar a tener efectos secundarios que se manifiestan por la aparición de nuevos síntomas de origen cardiovascular o de otro origen sistémico. El incumplimiento del tratamiento o su abandono es una de las principales causas del empeoramiento de la enfermedad y de la aparición de efectos adversos (1). Los mecanismos de acción y efectos adversos están relacionados con la composición molecular del fármaco mismo de manera fisiológica dichos fármacos tienen la capacidad de corregir u agravar una determinada condición médica, produciéndose diferentes reacciones en el paciente, desde hipersensibilidad, hasta aquellas de consecuencias fatales (2).

La noción de una conexión entre estructura molecular y acción fisiológica o farmacológica de una sustancia lleva más de 150 años de antigüedad (3).

El método QSAR (quantitative structure–activity relationships) se define como una ecuación matemática donde se correlaciona la actividad biológica con descriptores moleculares de característica química (4), además de ser una herramienta utilizada en los últimos años para simplificar el proceso mencionado, en el que básicamente se establece que existe una respuesta biológica de acuerdo a la estructura de una molécula (5).

Es por ello que este método es usado como una herramienta computacional, pero sus principales beneficiarios y los jueces finales son químicos medicinales (6). Dado que siempre existe el desafío de encontrar la interfaz y el equilibrio adecuados entre la experiencia computacional y el dominio experimental respectivo. Los descriptores juegan un papel crucial en el desarrollo de cualquier modelo QSAR ya que representan cuantitativamente la información química codificada. No sólo ayudan en la derivación de una correlación matemática entre información de la estructura química y la respuesta de interés, sino que también la exploración de un proceso bioquímico (7).

Es por ello que tendemos a pensar que el mayor nivel de éxito se habilitará a través de la colaboración entre científicos computacionales y experimentales que tienen un conocimiento profundo de sus respectivos campos pero que también han hecho esfuerzos para comprender y desarrollar el conocimiento práctico del campo complementario. La cantidad de datos que se pueden obtener y generar, contribuye a tener al alcance la posibilidad de desarrollar predictores más exactos en lo que hace referencia a actividad o acción química de una molécula, permitiendo desarrollar modelos de la mencionada actividad y así mediante una regresión, predecir efectos adversos a futuro ahorrando tiempo de investigación y recursos al personal, además de brindar un direccionamiento adecuado de la investigación de determinada molécula (4).

Dada la composición química de una molécula, es posible obtener mediante regresión los posibles efectos adversos y beneficiosos que presentara esta molécula (8); basándonos en este principio es que la metodología QSAR aparece como la herramienta necesaria para brindar una directriz en el desarrollo de fármacos (5).

Diversos modelos en metodología y aplicación del QSAR han permitido la incursión en diferentes áreas (5) como, por ejemplo en la química de los alimentos (9), como protagonizado por integrantes de nuestra Universidad.

Este estudio a nivel molecular aspira aportar otra alternativa en el desarrollo a futuro de fármacos mediante el desarrollo de un modelo QSAR , representadas por sus

descriptores moleculares que permite identificar las causas moleculares de los efectos adversos producidos por fármacos utilizados en patología de índole cardiovascular y, de esta forma, proveer un instrumento de predicción útil para el desarrollo farmacológico y de esta manera simplificar la complejidad del mismo, basado en la composición de la molécula y la consecuente actividad biológica (5), obteniendo beneficios en tiempo y recursos económicos y mostrando un lineamiento adecuado de hacia donde se necesita llegar o que efectos beneficiosos o adversos se podrían llegar a presentar (7).

## Materiales y Métodos

### Lista de moléculas

Seleccionamos los fármacos para la investigación – cardiovasculares -- de la base de datos SIDER 4.1: Side Effect Resource (10). Posteriormente se procedió con la obtención de cada uno de ellos, tanto del C.A.S. como del S.M.I.L.E.S., usando la base de datos de PubChem (11).

El CAS de un fármaco es un código numérico que permite identificar la molécula activa a nivel científico.

Los SMILES de las moléculas son códigos que permiten calcular propiedades moleculares de forma automática.

Tabla No.1: Codificación C.A.S. y S.M.I.L.E.S

Farmaco	C.A.S.	S.M.I.L.E.S.
Acetaminophen	103-90-2	<chem>CC(=O)NC1=CC=C(C=C1)O</chem>
Alfentanil	69049-06-5	<chem>CCC(=O)N(C1=CC=CC=C1)C2(CCN(CC2)CCN3C(=O)N(N=N3)CC)COC</chem>
Almotriptan	154323-57-6	<chem>CN(C)CCC1=CNC2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3</chem>

Los datos recopilados para obtener los efectos adversos también fueron obtenidos de la base de datos online SIDER 4.1: Side Effect Resource (10). Los términos de

efectos adversos recopilados se transformaron en formato MedDRA (12) para garantizar su estandarización, se usó el programa MedDRA Desktop Browser Version Beta 4.0.0.97 (2015). Como valores exponentes de las intensidades del efecto adverso se utilizó el porcentaje de usuarios que experimentaron el efecto. Adicionalmente se pesaron los valores por un factor 5 si los efectos secundarios eran graves y 10 para los efectos secundarios muy preocupantes de acuerdo al criterio de los autores, este criterio ha sido propuesto por los mismos por falta de referencias en literatura, siendo este el primero dentro de su línea de estudio.

### Cálculo de descriptores

Una vez controladas todas las moléculas para eliminar eventuales dobles o errores, se han calculado los descriptores moleculares hasta los bidimensionales con el programa DRAGON 7 versión 7.0.10, 2017 ([https://chm.kode-solutions.net/products\\_dragon.php](https://chm.kode-solutions.net/products_dragon.php))

Inicialmente se ha intentado construir un modelo de regresión OLS utilizando los distintos algoritmos genéticos para la selección de las variables significativas (es decir, los descriptores significativos, tomando en cuenta que Dragon 7.0 tiene la capacidad de calcular cerca de 3822 descriptores moleculares por molécula), pero se han obtenido modelos inapropiados y no validos en predicción.

Debido a esta razón se ha decidido construir un modelo de clasificación, utilizando de forma secuencial el modelo N3 acoplado a un algoritmo genético de selección del mejor conjunto de descriptores, para posteriormente usar un método de clasificación basado en mapas neurales auto organizativas de Kohonen (13).

### Clasificación mediante N3

El método de clasificación N3 (14) es una modificación del método de los K vecinos más cercanos (KNN) (15), el mismo que considera para la clasificación solo la información proporcionada por los K vecinos más cercanos del objeto a clasificarse.

A diferencia de KNN, el método N3 considera, para la clasificación, también los demás objetos.

La información de los demás objetos es ordenada en forma decreciente y se toma la contribución en forma exponencialmente negativa en función de la distancia del n-1 ésimo objeto desde el objeto a clasificarse. Además, la contribución es pesada por un peso que contiene un exponente alfa a optimizarse. El mejor valor de alfa será aquel que minimiza el error de clasificación.

Para la aplicación de este método, se ha utilizado un toolbox de MATLAB, el toolbox N3-BNN para MATLAB, que ha sido descrito por los autores del método (<http://michem.disat.unimib.it/chm/download/n3info.htm>).

#### Umbral de clase

En un análisis previo se intentó construir un modelo de regresión entre los descriptores y las respuestas; sin embargo, todos los métodos utilizados: OLS, PCR, PLS, han producido modelos de mala calidad y absolutamente incapaces de proporcionar predicciones confiables. Por esta razón se ha optado por construir un modelo de clasificación que pueda predecir; si la molécula presenta problemas cardiacos leves (clase 1) o graves (clase 2). Se ha fijado el umbral entre las dos clases en el 66-avo percentil de la distribución de las respuestas que corresponde a un valor de 2.7. Las moléculas con respuesta inferior a 2.7 constituyen la clase 1 (97 moléculas), mientras que las con respuesta superior o igual a 2.7 constituyen la clase 2 (49 moléculas).

Una vez establecidas las dos clases de actividad se ha subdividido el conjunto inicial de los datos (146 moléculas X 1072 descriptores) en dos sub conjuntos: el conjunto

de entrenamiento (*training set*) para entrenar el algoritmo, tomando (103 moléculas) y el conjunto ensayo (*test set*) para validarlo en predicción, tomando (43 moléculas). Esta subdivisión se ha realizado utilizando una rutina MATLAB que respeta la numerosidad inicial de las dos clases elaborada por los autores.

### Selección de los descriptores

Teniendo una base de datos constituida por 146 moléculas y 1072 descriptores, resulto necesario seleccionar los descriptores y de esta manera eliminar los redundantes y conservar únicamente los que permiten elaborar un modelo predictivo.

La selección se ha realizado usando un algoritmo genético en la versión de Leardi (16) implementado en un utilitario ga-toolbox de MATLAB.

### Dominio de Aplicación

El modelo QSAR es de tipo local, esto quiere decir que tiene validez únicamente en el lugar establecido por los valores de los descriptores de training set, fuera de este lugar los resultados son extrapolados y la predicción no es segura. Explicado en otros términos el modelo QSAR trata de reducir y confinar en cuanto a su aplicabilidad ya que toman sus estructuras químicas, propiedades físicas, mecanismos de acción y efectos adversos (como es en nuestro caso) para poder realizar una predicción confiable únicamente cuando las moléculas establecidas por el modelo matemático presenten semejanzas estructurales con las utilizadas para la elaboración del modelo (9).

Con el objetivo de determinar la valoración de la aplicación se ha empleado un sistema que se basa en las distancias, es decir se calcula la distancia de una molécula externa: test set u otra, tomando su trayecto desde un punto establecido en el lugar de los descriptores del training set. Se compara el trayecto con una

predefinida distancia umbra; La estimación de este se establece para denotar la segregación de los lugares más densamente pobladas del espacio de los datos originales (17), el cálculo se procedió utilizando KNN con 5 vecinos para definir las clases y distancia de Mahalanobis (18) para aproximarse más a las circunstancias de cálculo de N3. Las moléculas de test set que no forman parte del dominio, se excluyen del mismo.

### Cálculo del modelo

Una vez seleccionadas las variables significativas y definido el dominio se procedió al cálculo del modelo QSAR aplicando el algoritmo N3 sobre el training set. Posteriormente sobre el training set se ha aplicado el algoritmo GA\_PANN de Kohonen implementado sobre Matlab.

### Evaluación de los Modelos

Los parámetros de calidad usados para el training set como para el test set que ha sido utilizado ha sido: el N.E.R en una validación cruzada (*cross validation*), en otras palabras, se explica como la razón entre los casos clasificados acertadamente y los casos totales.

Además, se ha calculado para el training set la sensibilidad ( $S_n$ ) a saber, la «habilidad de separación en clases», y la especificidad ( $S_p$ ), la «capacidad de producir clases puras» de la clasificación de los péptidos en las dos clases.

Figura No. 1: Formula Sensibilidad y Especificidad del modelo empleado

$$S_p = \frac{TN}{TN + FP}; \quad (\text{ec. No 1}) \qquad S_n = \frac{TP}{TP + FN} \quad (\text{ec. No 2})$$

Se acota que TN y TP se refiere a el número de verdaderos negativos y verdaderos positivos respectivamente, mientras que FN y FP son el número de falsos negativos y falsos positivos respectivamente.

El modelo ha sido validado aplicándolo al test set externo, obteniendo también en este caso el N.E.R, Sp y Sn

## Resultados

Tabla No. 2: Descriptores moleculares, su descripción y tipo

Descriptor molecular	Descripción	Original inglés	Tipo
CATS2D_05_AL	Presencia de un sitio aceptor lipofilitico cada 5 átomos	CATS2D Acceptor-Lipophilic at lag 05	Descriptores CATS (Chemically Advanced Template Search) 2D
SpMin7_Bh(m)	Valor propio más pequeño de la matriz de la carga pesada por la masa. Derivado de una matriz de conectividad modificada. Relacionado con el potencial de ionización.	Smallest eigenvalue n. 7 of Burden matrix weighted by mass	Valores propios de la carga
B10[N-O]	Presencia/ausencia de grupos N-O a	Presence/absence of N - O at	Pares de átomos 2D

	una distancia topológica de 10.	topological distance 10	
B04[C-O]	Presencia/ausencia de grupos C-O a una distancia topológica de 4	Presence/absence of C - O at topological distance 4	Pares de átomos 2D
F10[C-F]	presencia/ausencia de grupos C-F a una distancia topológica de 10.	Frequency of C - F at topological distance 10	Pares de átomos 2D

Los descriptores seleccionados para el training set han sido: CATS2D\_05\_AL, SpMin7\_Bh(m), B10[N-O], B04[C-O], F10[C-F]; el modelo de cinco descriptores tiene un N.E.R cross-validado (ventanas venecianas de 10 grupos de cancelación) de 0.8, el valor de alfa es de 0.75.

Tabla No 3. Matriz de sensibilidad y especificidad del modelo de clasificación obtenido.

	Sn	Sp
Class1(no act.)	0.863	0.731
Class2 (activ)	0.731	0.863

Validación

Validación externa

Se aplicó el modelo a las moléculas del training set externo con el objetivo de obtener los estándares de validación sobre los péptidos no usados en la construcción del modelo.

El N.E.R en predicción externa fue de 0.46; en tanto que los valores de sensibilidad y precisión obtenidos se presentan en la tabla 4.

Tabla No 4. Sensibilidad y especificidad del modelo obtenido en predicción externa.

	Sn	Sp
Class 1	0.38	0.53
Class 2	0.53	0.38

Es evidente que el modelo obtenido no predice las actividades de nuevas moléculas de forma satisfactoria. Por tal motivo se ha aplicado el algoritmo C-pann con un mapa de 8X8 neuronas en celdas hexagonales obteniendo para el training set un N.E.R de 0.88.

La sensibilidad, la especificidad y la precisión del modelo se reportan en la tabla No. 5

Tabla No 5. Matriz de sensibilidad y especificidad del modelo de clasificación obtenido.

	Sn	Sp
Class1(no act.)	0.83	0.91
Class2 (activa)	0.91	0.83

El N.E.R en predicción externa fue de 0.67; mientras que los valores de sensibilidad y especificidad obtenidos se presentan en la tabla No. 6

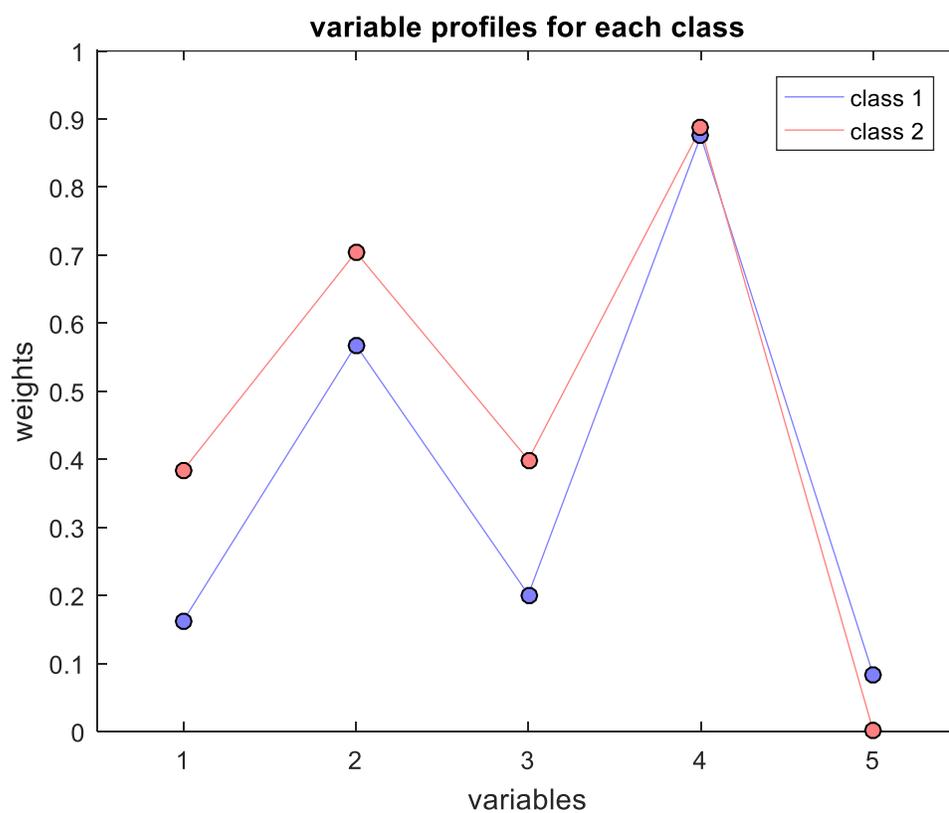
Tabla No. 6 Sensibilidad y especificidad del modelo obtenido en predicción externa

	Sn	Sp
Class 1	0.62	0.70
Class 2	0.70	0.62

Este último modelo es mucho mejor en predicción; ya que clasifica de mejor manera los de la clase 2 que los de la clase 1, lo que permite identificar con mayor seguridad y confianza a las futuras moléculas que producirán problemas cardíacos severos.

En la figura No. 2 se presenta el diagrama de los centroides tanto de la clase 1 como de la clase 2

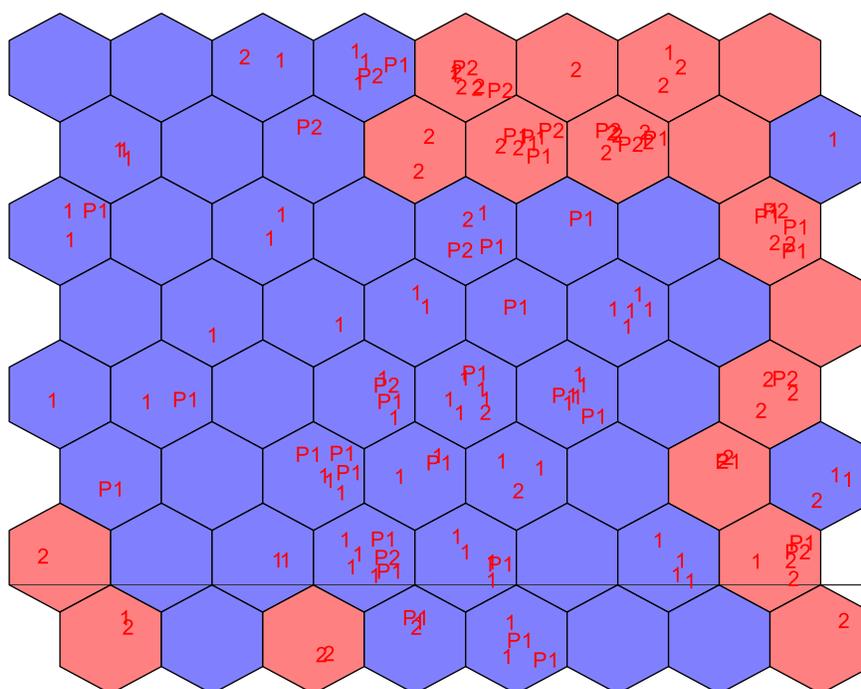
Figura No. 2 Diagrama de los centroides de las variables en las dos clases.



En el grafico anterior se puede interpretar que las moléculas de la clase 1, son las que presentan en menor porcentaje problemas cardiacos, además tienen pesos promedio de los descriptores CATS2D\_05\_AL, SpMin7\_Bh(m), B10[N-O], B04[C-O] inferiores a los de las moléculas de la clase 2 que presentan en mayor porcentaje problemas cardiacos. Salvo en el caso del descriptor F10[C-F] en donde la situación es inversa.

En la figura No. 3 se presenta el mapa de Kohonen de la capa auto organizativa, en donde se puede observar que tanto la clase 1 como la clase 2 están bien delimitadas, los mismos que también indican los parámetros de calidad.

Figura No. 3 Mapa de Kohonen.



## Discusión

La importancia de realizar estudios de este tipo se justifica por la elevada presencia de efectos adversos de gravedad severa que se pueden presentar en el tratamiento de enfermedades cardiovasculares (2). Es importante señalar que no existen estudios QSAR con el mismo propósito, por lo que las comparaciones no se pueden realizar. El presente estudio, sin embargo, se propone ser una herramienta útil en direccionar una investigación farmacéutica, pues facilita la identificación previa de aquellos fármacos que con alta probabilidad puedan desarrollar efectos adversos. Esto permite guiar la realización de posibles estudios a futuro. Otra ventaja constituye el cuantioso ahorro de recursos que esto generaría en el desarrollo de nuevos fármacos.

El presente estudio plantea un modelo predictivo para la identificación de aquellos grupos de fármacos sobre el umbral descrito en los resultados, en nuestro caso, es de 2,7.

Para la elaboración de nuestro modelo predictivo se trabajó con 5 descriptores moleculares. Los descriptores principales, que nos brinda más información acerca de los efectos adversos, son CATS2D\_05\_AL, SpMin7\_Bh(m), B10[N-O]. Los mismos que indican que las moléculas con un alto potencial de producir efectos adversos presentan más grupos con los mismos potenciales de ionización a la distancia de cinco enlaces (véase tabla No. 2).

Tanto la validación externa, la sensibilidad y la especificidad resultaron adecuadas en el presente estudio. (véase tabla No. 5). Este resultado puede sentar el fundamento para futuros modelos más precisos y válidos, cuando se disponga de una base de datos de efectos secundarios más completa. También la inclusión de otras familias de fármacos en el estudio puede mejorar el potencial predictivo. Los valores de significancia estadística demuestran que este modelo es útil en la predicción, constituyendo una herramienta potencial en el campo de la toxicidad farmacéutica.

Las fortalezas del presente estudio son:

- Estudio novedoso, por primera vez realizado a nivel mundial en la universidad del Azuay
- Solo existe 1 estudio similar realizado en el mismo laboratorio
- Relativamente sencillo de ejecutar
- No trabaja con sujetos de prueba

Las debilidades son:

- Problemas en la comparación por falta de estudios similares.
- El número total de moléculas estudiadas es limitado.
- La base de datos de síntomas adversos reportados no es completa.
- La ponderación toma como base en el criterio de los autores, ya que no existen trabajos previos que puedan servir como guía.

## **Conclusiones**

Finalmente, señalamos la importancia de desarrollar un modelo predictivo de los diferentes efectos adversos que pueden provocar los distintos fármacos, por lo que constituye una herramienta importante y útil previo a la investigación y desarrollo de los mismos. Además, permite direccionar de manera adecuada en el momento del estudio y realización de determinado fármaco que probablemente presente una alta probabilidad de producir efectos tóxicos, evitando así tomar una inadecuada dirección en la realización del mismo; teniendo como resultado y ventaja el ahorro de tiempo y cuantiosos recursos económicos que conlleva la investigación de un fármaco en particular.

Nuestro estudio fundamenta bases para posteriores investigaciones con la metodología QSAR en cuanto a efectos adversos farmacológicos. Alentamos a la posibilidad futura de investigar y mejorar las herramientas para obtener estudios minuciosos y precisos; El origen de nuestro estudio parte de una base de datos que recopila información de efectos adversos presentados y reportados en pacientes. A futuro se debería emplear esfuerzos para mejorar este punto de partida, lo que

mejoraría la precisión y exactitud de futuros modelos. Otro aspecto que merece mención es el limitado número de fármacos que nos brindan las diferentes bases de datos de efectos secundarios, por lo que se espera se incremente la cantidad de los mismos para así obtener una base con datos más amplia y con mayor información, que arroje resultados más satisfactorios.

Es por ello que nuestro estudio es útil como modelo de predicción de efectos adversos y que además busca brindar ayuda a gran parte de la medicina y de la sociedad, disminuyendo el grado y riesgo de provocar un daño, desde el momento del desarrollo molecular de los distintos fármacos.

Además, vale destacar que esta clase de estudio fue realizado por primera vez a nivel mundial y ha sido llevado a cabo por autores de nuestra universidad y que se espera sea publicado.

### **Agradecimiento**

Agradecemos a nuestros padres por su apoyo absoluto, a nuestro director de tesis, asesor metodológico, ya que sin su guía y apoyo el proyecto no hubiese podido salir adelante.

### **Bibliografía**

1. Freire RB. Fármacos cardiovasculares.
2. Werner J Pichler M. Features, Drug hypersensitivity: Classification and clinical [Internet]. Features, Drug hypersensitivity: Classification and clinical. 2019. Available from: [https://www.uptodate.com/contents/drug-hypersensitivity-classification-and-clinical-features?search=mecanismo de accion de farmacos&source=search\\_result&selectedTitle=2~150&usage\\_type=default&display\\_rank=2](https://www.uptodate.com/contents/drug-hypersensitivity-classification-and-clinical-features?search=mecanismo%20de%20accion%20de%20farmacos&source=search_result&selectedTitle=2~150&usage_type=default&display_rank=2)
3. Hansch, C., Maloney, P. P., Fujita, T., & Muir RM. © 1962 Nature Publishing Group. In: Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients Nature. 1962. p. 194(4824), 178–180.

4. Ensuncho AE, Robles JR, Figueredo SF. Modelado de las relaciones cuantitativas estructura-actividad (QSAR) de los derivados 5-(nitroheteroaril)-1,3,4-tiadiazol con actividad leishmanicida. *Inf Tecnol.* 2017;28(2):191–202.
5. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR modeling: Where have you been? Where are you going to? *J Med Chem.* 2014;57(12):4977–5010.
6. Ishikawa T, Hirano H, Saito H, Sano K, Ikegami Y, Yamaotsu N, et al. Quantitative Structure-Activity Relationship (QSAR) Analysis to Predict Drug-Drug Interactions of ABC Transporter ABCG2. *Mini-Reviews Med Chem.* 2012;12(6):505–14.
7. Lozano-aponte J, Scior T, Joachim P, Qsar D. ¿Qué sabe Ud. acerca de...QSAR? *Rev Mex Ciencias Farm.* 2012;43(2):82–4.
8. Rasyid H, Purwono B, Armunanto R. Quantitative structure activity relationship (QSAR) based on electronic descriptors and docking studies of quinazoline derivatives for anticancer activity. *Orient J Chem.* 2018;34(5):2361–9.
9. Rojas villa christian xavier, Procesamiento CE, Cer DEM, Refractarios M, Sistema DEL, Nicol F, et al. Universidad nacional de la plata. 2017;1–53.
10. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. 2016;44(October 2015):1075–9.
11. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update : improved access to chemical data. 2019;1–8.
12. Dictionary TM, Activities R. Understanding MedDRA The Medical Dictionary for Regulatory Activities What is ICH ?
13. Melssen W, Wehrens R, Buydens L. Supervised Kohonen networks for classification problems. 2006;83:99–113.
14. Todeschini R, Ballabio D, Cassotti M, Consonni V. N3 and BNN : Two new similarity based classification methods in comparison with other classifiers . 2015;
15. Zhang S, Deng Z, Cheng D, Zhu X. Author ' s Accepted Manuscript Efficient k NN Classification Algorithm for Big Data Reference : To appear in : Neurocomputing. Neurocomputing [Internet]. 2015; Available from: <http://dx.doi.org/10.1016/j.neucom.2015.08.112>

16. Leardi R. Genetic algorithms applied to feature selection in PLS regression : how and when to use them. 1998;
17. Sahigara F, Mansouri K, Ballabio D, Mauri A, Consonni V, Todeschini R. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. 2012;4791–810.
18. Ballabio D, Vasighi M, Consonni V, Kompany-zareh M. Chemometrics and Intelligent Laboratory Systems Genetic Algorithms for architecture optimisation of Counter-Propagation Artificial Neural Networks. Chemom Intell Lab Syst [Internet]. 2011;105(1):56–64. Available from: <http://dx.doi.org/10.1016/j.chemolab.2010.10.010>