Aplicación de Machine Learning para la predicción de antagonistas del receptor de histamina RH3 como potenciales candidatos terapéuticos

Carlos Marcelo Benitez [1], Natacha Soledad Represa [1* [0000-0002-7066-4799]], Ricardo Di Pasquale [1 [0000−0003−0549−9595]], María Betina Comba [2 [0000-0001-5365-548X]], Vanina Medina [3] and María Marta Zanardi [2 [0000-0002-7145-5358]]

[1] Laboratorio de Ciencias de Datos e Inteligencia Artificial, Facultad de Ingeniería y Ciencias Agrarias, Pontificia Universidad Católica Argentina, Av. Alicia Moreau de Justo 1300 (C1107AAZ). C.A.B.A., Argentina
* natacharepresa@uca.edu.ar

[2] Instituto de Investigaciones en Ingeniería Ambiental, Química y Biotecnología Aplicada (INGEBIO). Facultad de Química e Ingeniería del Rosario, Pontificia Universidad Católica Argentina, Av. Pellegrini 3314, (S2002QEO) Rosario, Argentina.

[3] Instituto de Investigaciones Biomédicas, Facultad de Ciencias Médicas, Pontificia Universidad Católica Argentina, Av. Alicia Moreau de Justo 1300 (C1107AAZ). C.A.B.A., Argentina

Abstract. Los antagonistas del receptor de histamina H3 (RH3) emergen como potenciales fármacos para diversos trastornos neurológicos como Alzheimer o Parkinson, además de su reciente evaluación en el cáncer de mama triple negativo. En este contexto, identificar nuevos ligandos antagonistas del RH3 es de gran interés por su amplio espectro de aplicaciones. Este estudio emplea técnicas de machine learning, específicamente, de regresión con el algoritmo Gradient Boosting (XGBoost), para predecir la afinidad de compuestos orgánicos antagonistas por el RH3 (pKi) utilizando descriptores moleculares. Se recopiló una base de datos con 831 compuestos antagonistas con valores de pKi conocidos, a partir de los cuales se generaron representaciones SMILES y se calcularon 1173 descriptores moleculares de baja dimensionalidad. La base se dividió en conjuntos de entrenamiento (665 registros) y testeo (166 registros). Se entrenaron y evaluaron 10 modelos diferentes, aplicando validación cruzada K-fold=5. El modelo más destacado alcanzó un MSE de 0.54 y un MAE de 0.50 en el conjunto de entrenamiento, y un MSE de 0.76 y un MAE de 0.53 en el conjunto de prueba, con un RMSE de 0.72 y 0.87, respectivamente. Este abordaje quimioinformático propone una metodología eficaz para el cribado virtual de potenciales ligandos antagonistas del RH3, acelerando el descubrimiento de nuevos compuestos terapéuticos.

Keywords: machine learning, receptor histamina RH3, cribado virtual

1 Introducción

La histamina es una amina biogénica crucial en procesos fisiológicos y patológicos, mediando sus efectos a través de cuatro subtipos de receptores acoplados a proteínas G, entre los que destaca el receptor H3 (RH3). El RH3 se expresa casi exclusivamente en el sistema nervioso central, donde regula la síntesis y liberación de histamina y otros neurotransmisores [1]. Por ello, los antagonistas del RH3 representan una prometedora clase de fármacos para el tratamiento de trastornos neurológicos, neuropsiquiátricos y cáncer de mama triple negativo [2]. Identificar nuevos ligandos antagonistas del RH3 con adecuadas propiedades farmacocinéticas y seguridad es un área de intensa investigación [3].

Las técnicas quimioinformáticas ofrecen un enfoque racional para el descubrimiento de nuevos ligandos del RH3. Mediante el modelado de relaciones cuantitativas estructura-actividad (QSAR), es posible explorar las características estructurales que gobiernan la afinidad de compuestos orgánicos hacia esta diana biológica [4]. En esta línea, el presente estudio propone desarrollar modelos de aprendizaje automático para predecir la afinidad de unión (pKi) de antagonistas del RH3 a partir de descriptores moleculares, facilitando así el cribado virtual de potenciales ligandos.

2 Metodología

Se recopiló una colección de 831 compuestos orgánicos con valores experimentales de afinidad de unión (pKi) hacia el receptor H3 de histamina, abarcando diversas familias químicas. Cada compuesto se representó mediante su notación SMILES (Simplified Molecular Input Line Entry Specification) y se calcularon descriptores moleculares de baja dimensionalidad con la librería Mordred de Python, capturando propiedades fisicoquímicas, geométricas, topológicas y electrostáticas relevantes. El conjunto de datos se dividió aleatoriamente en entrenamiento (80%) y prueba (20%). Se entrenaron 10 modelos de regresión XGBoost con distintos hiperparámetros, utilizando validación cruzada K-fold=5 en el conjunto de entrenamiento. El rendimiento de los modelos se evaluó mediante el Error Cuadrático Medio (MSE), la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Medio (MAE) en los conjuntos de entrenamiento y prueba. Las métricas en el conjunto de prueba permitieron seleccionar el modelo con mejor capacidad predictiva.

3 Resultados y Discusión

Se realizó un análisis exploratorio de la variable respuesta (afinidad de unión expresada como pKi). La variable se encuentra dentro del rango: 1.1 - 10.6, con media: 8.15, mediana: 8.30 y una desviación estándar: 1.137. Respecto a las variables descriptoras empleadas, se calcularon 1173 descriptores moleculares por molécula. Con estos datos, se entrenaron y evaluaron 10 modelos de regresión empleando el algoritmo XGBoost con distintos hiperparámetros, cuya performance se observa en la **Tabla 1**.

Table 1. Métricas de rendimiento de los modelos en los conjuntos de entrenamiento y prueba.

Modelo	Entrenamiento			Prueba		
	MSE	MAE	RMSE	MSE	MAE	RMSE
1	0.726	0.634	0.849	1.042	0.691	1.021
2	0.993	0.748	0.988	9.203	2.925	3.034
3	0.686	0.570	0.828	0.792	0.533	0.890
4	46.736	6.753	6.836	44.968	6.592	6.706
5	9.664	3.008	3.108	0.598	0.849	0.921
6	13.276	3.017	3.644	44.964	6.591	6.705
7	1.193	0.798	1.090	46.594	6.718	6.826
8	0.562	0.499	0.742	0.845	0.600	0.919
9	0.529	0.500	0.722	0.763	0.529	0.874
10	0.844	0.667	0.912	0.517	0.733	0.856

El modelo 9 sobresale por su precisión y capacidad de generalización. Con una tasa de aprendizaje (eta = 0.055488) que promueve un ajuste efectivo y progresivo, y una regularización L1 (alpha = 0.060220) que favorece la simplicidad del modelo, este enfoque reduce la complejidad y mejora la robustez. El alto submuestreo de columnas (colsample_bytree = 0.811337) añade variabilidad, y un número elevado de iteraciones de boosting (num_round = 220) permite un aprendizaje exhaustivo. Esta configuración de hiperparámetros explica su rendimiento superior, resultando en un modelo altamente eficaz y confiable para predecir sobre nuevos datos, consiguiendo predecir el 75% de los datos de prueba con menos del 8% de error.

La relación entre valores predichos y observados (Fig. 1) revela anomalías por debajo de 6. Esto se debe a que la base de datos presenta valores insuficientes de pKi para ese rango de afinidad, lo que resulta en un sesgo en el entrenamiento del modelo y limita su precisión en este rango. La presencia de estos valores anómalos resalta la necesidad de un conjunto de datos más completo para mejorar la capacidad predictiva del modelo.

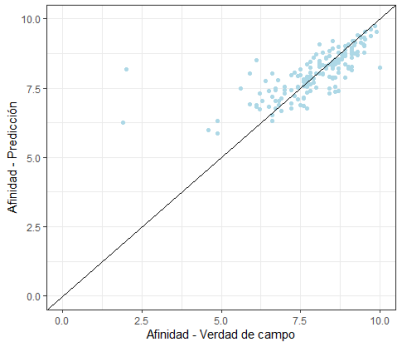

Fig. 1. Gráfica de dispersión entre los valores de afinidad predicha y observada

4 Conclusiones

En este estudio se desarrollaron y evaluaron 10 modelos XGBoost para predecir la afinidad de unión (pKi) de antagonistas del receptor H3 de histamina, usando 831 compuestos y sus descriptores moleculares, logrando modelos con notable desempeño predictivo. Para el mejor modelo se obtuvo un RMSE de 0.874, en el conjunto de prueba, siendo este un modelo robusto y generalizable. Sin embargo, la gráfica de dispersión señaló una limitación en la predicción de compuestos con baja afinidad, atribuible a la falta de datos en ese rango específico en la base de datos. Para superar este obstáculo y mejorar el rendimiento general, se propone la incorporación de descriptores 3D adicionales y la creación de modelos especializados según la familia química. Esta estrategia no solo refinará la capacidad predictiva de los modelos, sino que también proveerá una herramienta computacional más precisa para la identificación de nuevos candidatos terapéuticos contra el receptor H3 de histamina, optimizando así el proceso de descubrimiento de fármacos.

Contribuciones de los Autores. C. M. Benitez realizó el análisis de los datos y se encargó de generar las visualizaciones de los datos, N. S. Represa coordinó la tarea entre los miembros del equipo y redactó el manuscrito original, R. Di Pasquale se encargó del desarrollo de software y de la revisión del manuscrito, M. B. Comba, V. Medina y M. M. Zanardi son quienes llevan adelante la investigación sobre el receptor de histamina RH3, realizaron la recolección de datos y fueron responsables de la conceptualización del estudio.

Referencias

1. Rahman, S. N.; McNaught-Flores, D. A.; Huppelschoten, Y.; da Costa Pereira, D.; Christopoulos, A.; Leurs, R.; Langmead, C. J. ACS Chem. Neurosci. **14**, 645−656 (2023).
2. Ospital, I.A., Delgado, M.A.T., Nicoud, M.B., Corrêa, M.F., Fernandes, G.A.B., Andrade, I.W., Lauretta, P.; Vivot, R.M.; Comba, M.B.; Zanardi, M.M.; Speisky, D. Therapeutic potential of LINS01 histamine H3 receptor antagonists as antineoplastic agents for triple negative breast cancer. Biomedicine & Pharmacotherapy, **174**, p.116527 (2024).
3. Moriwaki, H.; Tian, Y. S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. J. Cheminform. **10** (1), 1–14 (2018).