



## DESARROLLO DE UN SISTEMA DE BÚSQUEDA INTELIGENTE DE INTERACCIONES ENTRE FÁRMACOS Y GENES EN TEXTOS CIENTÍFICOS UTILIZANDO APRENDIZAJE PROFUNDO.

Autores: Bertinetti, Juan y Ramírez, Darién

Director: Dr. Cristian Yones

Fecha de presentación: -- -- ----

## Motivación

La medicina de precisión consiste en un modelo médico que propone la personalización de las decisiones médicas, prácticas y tratamientos para cada paciente de forma individual teniendo como sustento la genómica personalizada, la cual consiste en la utilización de fármacos dirigidos a moléculas específicas que intervienen en la enfermedad del individuo. Para esto es de vital importancia para los médicos conocer cómo diferentes fármacos interactúan con distintos genes y de qué forma.

A medida que mejoran las tecnologías de secuenciación de próxima generación, la cantidad de información generada por la comunidad científica es demasiada para ser procesada y clasificada manualmente. Dado que los buscadores tradicionales de texto sólo encuentran los trabajos que contienen los términos de búsqueda sin arrojar luz sobre las relaciones entre ellos, sumado a que los intentos de automatizar el trabajo de extraer relaciones entre genes y fármacos en textos no brindan los resultados deseados, se manifiesta la necesidad de un sistema integral de búsqueda inteligente que permita realizar búsquedas específicas sobre estos conceptos.

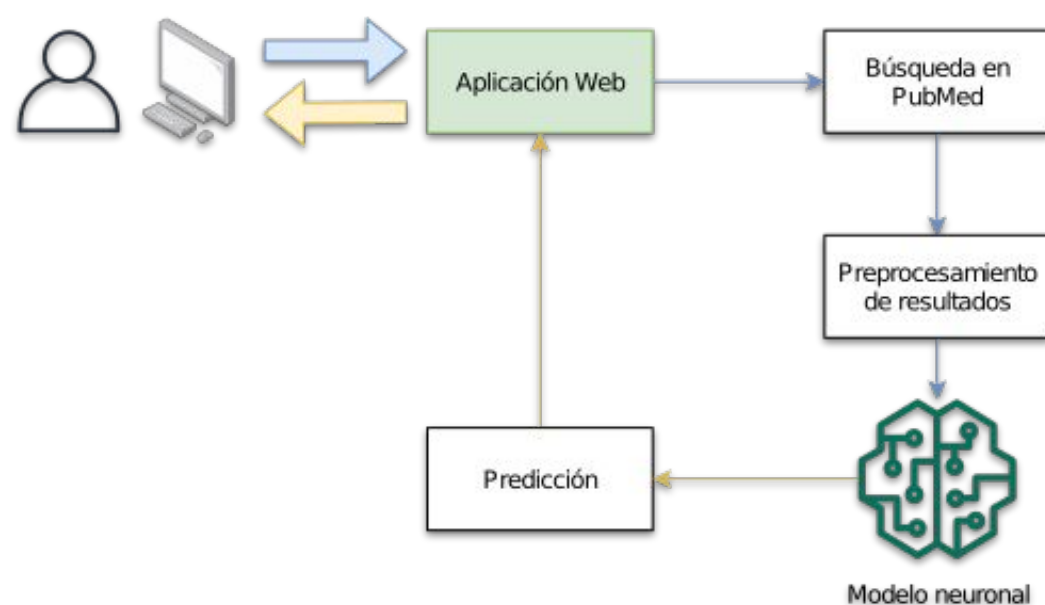


Diagrama de flujo de información del sistema.

## Objetivos

### Objetivo general:

- Desarrollar un sistema de búsqueda inteligente de interacciones entre fármacos y genes en textos científicos utilizando aprendizaje profundo.

### Objetivos específicos:

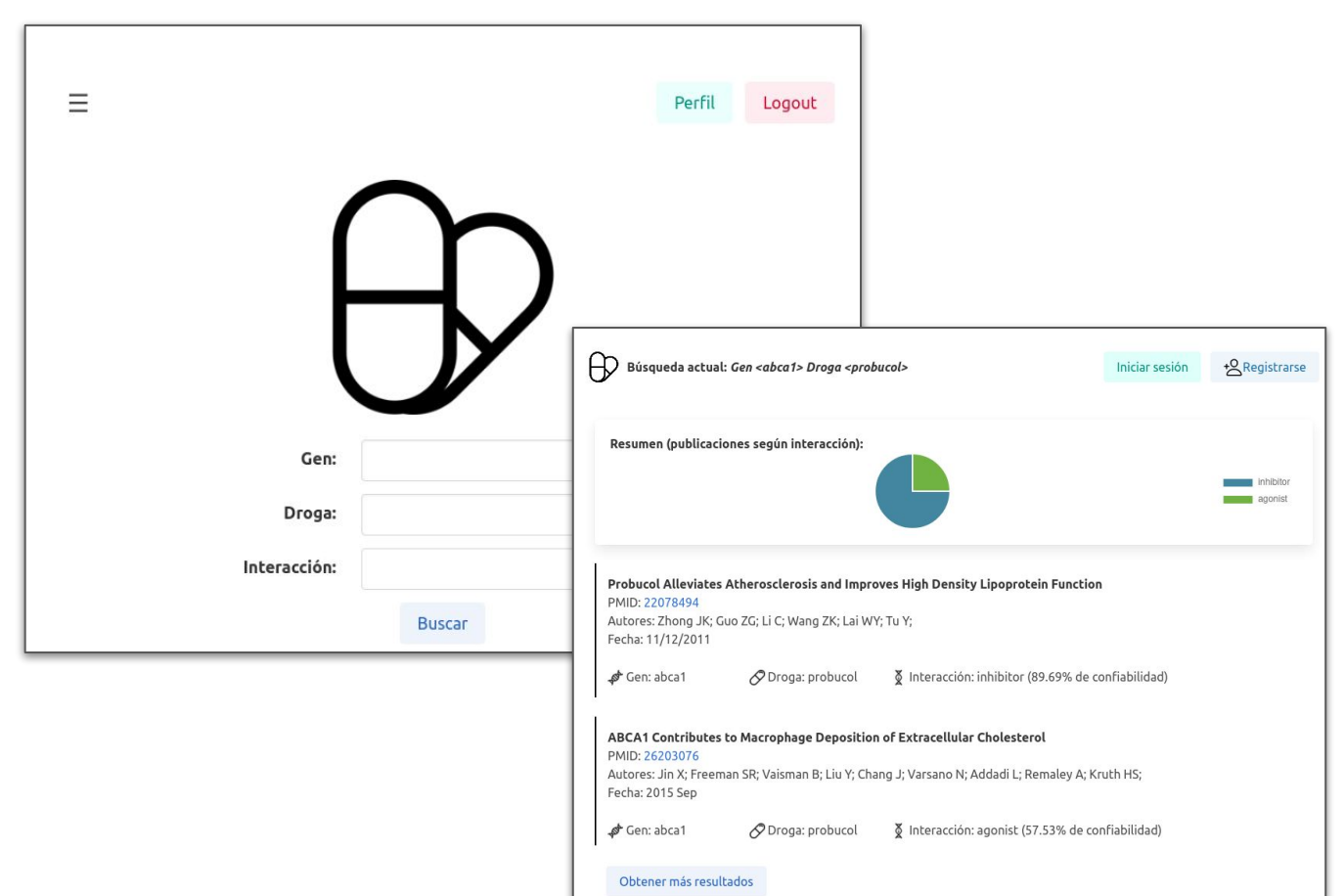
- Crear una base de datos de publicaciones científicas etiquetadas que permita el entrenamiento de modelos de aprendizaje automático.
- Desarrollar un motor de búsqueda de interacciones fármaco-gen mediante redes neuronales convolucionales (CNN) y Transformers.
- Desarrollar una aplicación web que permita a usuarios realizar búsquedas de interacciones fármaco-gen.

## Desarrollo

Inicialmente se confeccionó una base de datos de *papers* etiquetados, mediante *web scraping* a partir de la información de DGIdb (*Drug Gene Interaction database*, <https://dgidb.org>). Con esta base de datos se entrenaron múltiples modelos neuronales, basados en dos tipos de redes diferentes: CNN y Transformers. Los mismos se evaluaron y compararon utilizando diferentes métricas, escogiendo el de mejor desempeño.

Finalmente, se desarrolló una aplicación web para buscar artículos médicos, que utiliza el modelo seleccionado como motor de inferencia (predicción) de interacciones entre genes y fármacos. Los artículos científicos son obtenidos en tiempo real desde PubMed.

La aplicación además brinda funcionalidad adicional al usuario, como historial de búsquedas, búsquedas guardadas, y posibilidad de marcar artículos como favoritos.



Pantalla principal y de resultados de la aplicación web.

## Resultados-Conclusiones

Se puede afirmar que las redes de aprendizaje profundo, específicamente las redes neuronales convolucionales y Transformers, son útiles para realizar procesamiento del lenguaje natural (NLP) dando buenos resultados en la tarea de identificación de interacciones entre fármacos y genes.

La base de datos de publicaciones etiquetadas que se creó puede ser utilizada para futuros proyectos, relacionados o no a este, donde se requieran entrenar otros modelos.

Por último, la aplicación web constituye junto al modelo neuronal finalmente seleccionado, una herramienta útil para profesionales de la salud dentro del campo de la medicina de precisión.