

Universidad de La Habana  
Facultad de Matemática y Computación



# Propuesta para un modelo de aprendizaje de máquinas para el reposicionamiento de medicamentos

Autor:

**Carmen Irene Cabrera Rodríguez**

Tutores:

**Dr. Yudiivián Almeida Cruz**  
**Lic. Lázaro Raúl Iglesias Vera**

Trabajo de Diploma  
presentado en opción al título de  
Licenciado en Ciencia de la Computación

Fecha

[github.com/cicr99/thesis](https://github.com/cicr99/thesis)

Dedicación

# Agradecimientos

Agradecimientos

# Opinión del tutor

Opiniones de los tutores

# Resumen

Resumen en español

# Abstract

Resumen en inglés

# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Estado del Arte</b>	<b>3</b>
1.1. Reposicionamiento de medicamentos . . . . .	3
1.2. Enfoques computacionales para el reposicionamiento de medicamentos	4
1.2.1. Clasificación de los enfoques computacionales . . . . .	5
1.3. Grafos de conocimiento . . . . .	6
<b>2. Propuesta</b>	<b>7</b>
<b>3. Detalles de Implementación y Experimentos</b>	<b>8</b>
<b>Conclusiones</b>	<b>9</b>
<b>Recomendaciones</b>	<b>10</b>
<b>Bibliografía</b>	<b>11</b>

# Índice de figuras



## Ejemplos de código

# Introducción

Desde hace varios años, la disminución de la productividad de la investigación y el desarrollo farmacéutico se ha identificado como un problema importante por su insostenibilidad en el tiempo. Esto es consecuencia tanto de lo que se tarda en introducir un nuevo fármaco, como los recursos económicos que se deben destinar para ello. Según el *Tufts Center for the Study of Drug Development (CSDD)*, el costo de desarrollar un nuevo medicamento recetado con aprobación de comercialización es de aproximadamente 2600 millones de dólares y solo uno de cada veinte medicamentos obtiene la aprobación para ensayos clínicos [1].

La disminución de la productividad se traduce en un mayor riesgo de desarrollo y en un aumento de los precios de los medicamentos cuando llegan al mercado. Esto plantea una amenaza, no solo para la industria farmacéutica, sino también para la accesibilidad de medicamentos innovadores por parte de los pacientes.

## Motivación

Una de las estrategias utilizadas para la optimización del proceso de investigación y desarrollo farmacéutico es conocida como reposicionamiento de medicamentos. El **reposicionamiento de medicamentos** (también llamado reutilización de medicamentos, reperfilado, redirección o redescubrimiento de drogas [2]) consiste en la identificación de nuevos propósitos terapéuticos para medicamentos ya aprobados, más allá del alcance de su uso terapéutico original. Esta técnica ofrece varias ventajas sobre el desarrollo de fármacos completamente nuevos, donde se incluyen la posibilidad de acelerar el proceso de descubrimiento y reducir las tasas de fallas en las fases de desarrollo clínico y prueba [3]. En particular, evita las evaluaciones de seguridad en modelos preclínicos y humanos, si estas se han completado ya para la indicación original y existe compatibilidad de dosis con la nueva indicación. Por consiguiente, lleva a costos generales de desarrollo potencialmente más bajos.

## Antecedentes

## Problemática

Dadas las desventajas que presenta el desarrollo de nuevos medicamentos por la vía regular, los enfoques para optimizar dicho proceso se encuentren bajo un escrutinio intenso. En el caso de la reutilización de medicamentos, tradicionalmente, las historias de éxito han resultado, principalmente, de hallazgos en gran medida oportunistas y fortuitos [4]. Con vistas a hacer de este un proceso más sistemático, en los últimos años, se han desarrollado diversos enfoques computacionales. Sin embargo, ninguno de ellos ha llegado a ser totalmente efectivo.

## Objetivo

### Objetivo general

El diseño e implementación de una propuesta de modelo de aprendizaje computacional que posibilite la reutilización de medicamentos, aplicando específicamente, técnicas de aprendizaje profundo.

### Objetivos específicos

- Realizar un análisis del estado del arte en materia de reposicionamiento computacional de medicamentos.
- Identificación de bases de datos con información sobre los medicamentos, así como sobre drogas ya reposicionadas, que puedan ser utilizadas en la propuesta.
- Diseñar e implementar un prototipo de modelo de aprendizaje de máquina a partir de las bases de datos y los vectores de análisis seleccionados, que sea capaz de predecir si una droga puede ser reposicionada con un determinado fin.
- Plantear una metodología de evaluación para el modelo que permita el análisis de los resultados una vez aplicado.

## Estructura

# Capítulo 1

## Estado del Arte

### 1.1. Reposicionamiento de medicamentos

La reutilización de medicamentos, también conocida como reposicionamiento, es una estrategia que busca nuevos tratamientos médicos entre los medicamentos con licencia ya existentes, en lugar del desarrollo de nuevas estructuras moleculares [2]. Esta técnica ofrece numerosas ventajas en contraste, en términos de desarrollo [5]:

- Disponibilidad de datos de farmacocinética, farmacodinámica y posología
- Conocimiento de seguridad y toxicidad, incluidos los efectos adversos
- Experiencia clínica derivada de las indicaciones originales
- Bajo costo, en particular para el caso de medicamentos genéricos con múltiples fabricantes
- Comprensión de los mecanismos de acción y/o *targets* moleculares.

Existen numerosos ejemplos de reposicionamiento de medicamentos de forma exitosa en el pasado, como es el caso de los siguientes [6, 7]:

**Aspirina** Comercializada inicialmente en 1899 como analgésico, es reposicionada por primera vez en la década de 1980 como un fármaco antiagregante plaquetario para prevenir eventos cardiovasculares

**Talidomida** Prescrito en 1950 para el tratamiento de las náuseas y el insomnio en mujeres embarazadas y posteriormente aprobado para el tratamiento de la lepra.

**Citrato de sildenafil** Desarrollado originalmente como un medicamento antihipertensivo, y luego reutilizado por Pfizer y comercializado como Viagra para el tratamiento de la disfunción eréctil.

**Dimetilfumarato** Durante años, solo se conoció como un inhibidor de moho para proteger el cuero y como causa de alergias. Se ha comercializado como medicamento desde 1994, para tratar la psoriasis. Teniendo en cuenta su actividad antiinflamatoria, se propuso su uso en otra enfermedad autoinmune, la esclerosis múltiple. Se comercializó para esta última indicación a partir de 2013.

**Finasterida** Originalmente recetado para tratar el agrandamiento de la próstata y luego, para curar la pérdida de cabello.

**Clorpromazina** Desarrollada en un principio como antihistamínico, se comercializó posteriormente como sedante y antiemético.

**Azidotimidina** Usada en el tratamiento del cáncer, ahora se reutiliza como una terapia contra el VIH

Como se puede apreciar, la reutilización de medicamentos no constituye una idea nueva en el campo de la medicina; sin embargo, como estrategia explícita de desarrollo, ha tenido auge en los últimos años. De hecho, datos obtenidos de *PubMed* muestran que el número de publicaciones relacionadas con el reposicionamiento de medicamentos ha aumentado de forma exponencial desde 2004 [8]. Esto se debe, en parte, a la acumulación de grandes volúmenes de datos que permiten la implementación de enfoques computacionales. Las fuentes de información populares incluyen: registros de salud electrónicos, análisis de asociación de todo el genoma o perfiles de expresión génica, estructuras compuestas y otros datos de perfiles fenotípicos. La tabla (AÑADIR REFERENCIA A TABLA) muestra algunas de las bases de datos utilizadas con mayor frecuencia en la actualidad [9].

## 1.2. Enfoques computacionales para el reposicionamiento de medicamentos

El problema principal en el reposicionamiento de medicamentos es la detección de nuevas relaciones entre medicamentos y enfermedades. Para abordar dicho problema, se han desarrollado una variedad de enfoques que incluyen enfoques computacionales, enfoques experimentales biológicos y enfoques mixtos. [10]

Dado el crecimiento de la cantidad de datos biomédicos y farmacéuticos a gran escala disponibles públicamente, los enfoques computacionales que utilizan minería de datos, aprendizaje automático y análisis de redes se vuelven cada vez más importantes [11]. Se pudiera asumir que los ensayos experimentales biológicos son más confiables y predictivos en comparación; sin embargo, las condiciones de experimentación, la cantidad de concentraciones probadas y la cantidad de repeticiones pueden influir en

la precisión de los resultados. Además, en contraste, los enfoques computacionales son de menor costo y presentan menos limitaciones. [12]

### 1.2.1. Clasificación de los enfoques computacionales

Los enfoques computacionales existentes pueden ser clasificados de diferentes formas de acuerdo a distintos puntos de vista. Por ejemplo, algunos investigadores los agrupan de acuerdo a las redes biológicas utilizadas: redes de interacción molecular, redes reguladoras de genes, redes metabólicas, redes de interacción entre proteínas, entre otras [13]. Otra clasificación, que se enfoca en los modelos y metodologías utilizadas durante el reposicionamiento, es: métodos basados en datos y basados en hipótesis [14], aunque esta clasificación general puede ser aplicada a cualquier enfoque computacional. De acuerdo a las metodologías centrales utilizadas, se han dividido también en tres categorías: enfoques basados en redes, enfoques de minería de texto y enfoques semánticos [15]. Este trabajo abordará la última clasificación mencionada.

#### Enfoques basados en redes

Los métodos basados en redes se pueden dividir en: enfoques de agrupamiento (*clustering*) y enfoques de propagación [10].

El propósito de los enfoques de agrupamiento basados en redes es el descubrimiento de nuevas relaciones de tipo fármaco-enfermedad o fármaco-objetivo en grupos más pequeños dentro de una red más grande. Están basados en la idea de que entidades biológicas (enfermedades, medicamentos, proteínas, etc.) que pertenecen al mismo módulo de redes, comparten características similares. Por tanto, utilizan algoritmos de *cluster* para encontrar diferentes subredes (también conocidas como grupos o cliques).

Por otra parte, los enfoques de propagación están basados en el flujo de conocimiento adquirido previamente a través de las diferentes capas de una red. Estos parecen ser bastante efectivos para la identificación de interconexiones de interés [16]. De acuerdo a la forma en que se trata la red para la obtención de información, los métodos basados en propagación se separan igualmente en dos categorías: enfoques locales, centrados en secciones de la red, y enfoques globales, que examinan la red en su totalidad. Los enfoques de propagación local toman en cuenta una cantidad de información limitada de la red, por consiguiente, fallan en algunos casos a la hora de realizar predicciones correctas [16]. Por el contrario, los segundos presentan un mejor desempeño.

#### Enfoques basados en minería de textos

Los enfoques basados en minería de textos explotan la gran cantidad de literatura disponible, que es filtrada posteriormente para retener solo las fuentes relevantes y

extraer de ellas conocimiento para un conjunto de términos biológicos de interés. El origen de estos métodos en el campo médico se describe como el método  $.^A B C^B$  [17]. La idea consiste en que al examinar los conceptos A, B y C, donde el concepto A está relacionado con B y B con C, es posible que exista también una conexión entre A y C. Basados en este modelo, varios métodos de minería de texto han sido propuestos para encontrar posibles relaciones enfermedad-medicamento en la literatura. Igualmente, el desarrollo de técnicas de procesamiento de lenguaje natural (*NLP - Natural Language Processing*), ha influido en el incremento de la implementación de herramientas para llevar a cabo este enfoque (véase [18, 19])

Un estudio reciente por Andronis et al. [20] resume varios enfoques y fuentes en el campo de minería de textos para el reposicionamiento de medicamentos.

Las herramientas de minería de texto reducen la complejidad temporal del reposicionamiento de medicamentos y ayuda a los investigadores a verificar sus resultados experimentales devolviendo masivas cantidades de relaciones de entidades biológicas. Sin embargo, presenta ciertas desventajas, como es el problema de cobertura limitada. Este se debe a que entidades biomédicas o relaciones parcialmente importantes tales como mutaciones, objetivos y fármacos no son considerados.

## Enfoques semánticos

### 1.3. Grafos de conocimiento

## Capítulo 2

### Propuesta



## Capítulo 3

# Detalles de Implementación y Experimentos

# Conclusiones

Conclusiones

# Recomendaciones

Recomendaciones

# Bibliografía

- [1] Alcimed. «Artificial Intelligence at the service of drug repositioning.» (2021), dirección: %5Chref%7Bhttps://www.alcimed.com/en/alcim-articles/artificial-intelligence-at-the-service-of-drug-repositioning/%7D%7BAlcimed%20Articles%7D (visitado 24-06-2022) (vid. pág. 1).
- [2] S. D. Langedijk J Mantel-Teeuwisse AK, «Drug repositioning and repurposing: terminology and definitions in literature,» *Drug Discov Today*, vol. 20, págs. 1027-1034, 2015 (vid. págs. 1, 3).
- [3] M. L. Pantziarka P Bouche G, «The repurposing drugs in oncology (ReDO) project,» *Ecancermedicalscience*, 2014 (vid. pág. 1).
- [4] E. P. Pushpakom S Iorio F, «Drug repurposing: progress, challenges and recommendations,» *Nat Rev Drug Discov*, vol. 18, págs. 41-58, 2019 (vid. pág. 2).
- [5] e. a. Pantziarka P Verbaanderd C, «ReDO\_DB: the repurposing drugs in oncology database,» *Ecancermedicalscience*, vol. 12, pág. 886, 2018 (vid. pág. 3).
- [6] e. a. Jourdan J Bureau R, «Drug repositioning: a brief overview,» *Journal of Pharmacy and Pharmacology*, vol. 72:9, págs. 1145-1151, 2020 (vid. pág. 3).
- [7] A. P, «Advantages and Challenges in Drug Re-Profiling,» *Journal of Pharmacovigilance*, vol. S2:e002, 2015 (vid. pág. 3).
- [8] S. P. Shah R, «Repurposing old drugs in oncology: Opportunities with clinical and regulatory challenges ahead,» *Journal of Clinical Pharmacy and Therapeutics*, vol. 44:1, págs. 6-22, 2018 (vid. pág. 4).
- [9] e. a. Li J Zheng S, «A survey of current trends in computational drug repositioning,» *Brief Bioinform*, vol. 17:1, págs. 2-12, 2016 (vid. pág. 4).
- [10] e. a. Xue H Li J, «Review of Drug Repositioning Approaches and Resources,» *International Journal of Biological Sciences*, vol. 14:10, págs. 1232-1244, 2018 (vid. págs. 4, 5).
- [11] R. A. Tamer N. Jarada Jon G. Rokne, «A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions,» *Journal of Cheminformatics*, vol. 12:1, pág. 46, 2020 (vid. pág. 4).

- [12] O. J. Oprea T, «Computational and Practical Aspects of Drug Repositioning,» *Assay Drug Dev Technol*, vol. 13:6, págs. 299-306, 2015 (vid. pág. 5).
- [13] e. a. Shahreza M Ghadiri N, «A review of network-based approaches to drug repositioning,» *Brief Bioinform*, vol. 19:5, págs. 878-892, 2018 (vid. pág. 5).
- [14] e. a. Zou J Zheng M, «Advanced systems biology methods in drug discovery and translational biomedicine,» *Biomed Res Int.*, 2013 (vid. pág. 5).
- [15] P. G. Koromina M Pandi M, «Rethinking Drug Repositioning and Development with Artificial Intelligence, Machine Learning, and Omics,» *OMICS A Journal of Integrative Biology*, vol. 23:11, págs. 539-548, 2019 (vid. pág. 5).
- [16] e. a. Emig D Ivliev A, «Drug Target Prediction and Repositioning Using an Integrated Network-Based Approach,» *Plos One*, vol. 8, 2013 (vid. pág. 5).
- [17] e. a. Weeber M Klein H, «Using concepts in literature-based discovery: Simulating Swanson’s Raynaud–fish oil and migraine–magnesium discoveries,» *Journal of the American Society for Information Science and Technology*, vol. 52:7, págs. 548-557, 2001 (vid. pág. 6).
- [18] A. W. Fleuren WW, «Application of text mining in the biomedical domain,» *Methods*, vol. 74, págs. 97-106, 2015 (vid. pág. 6).
- [19] V. A. Krallinger M Erhardt RA-A, «Text-mining approaches in molecular biology and biomedicine,» *Drug discovery today*, vol. 10, págs. 439-445, 2005 (vid. pág. 6).
- [20] e. a. Andronis C Sharma A, «Literature mining, ontologies and information visualization for drug repurposing,» *Brief Bioinform*, vol. 12:4, págs. 357-368, 2011 (vid. pág. 6).