

© 2022 Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Zaragoza.

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

TIP Revista Especializada en Ciencias Químico-Biológicas, 25: 1-17, 2022.

<https://doi.org/10.22201/fesz.23958723e.2022.450>

## Uso de la inteligencia artificial en la investigación para el reposicionamiento de fármacos

Kevin Samael Olascoaga-Del Angel <sup>1,2</sup>, Mina Konigsberg-Fainstein <sup>2</sup>,  
Jaime Pérez-Villanueva <sup>3,a</sup> y Norma Edith López Díaz-Guerrero <sup>2,b</sup>

<sup>1</sup>Posgrado en Biología Experimental, DCBS, <sup>2</sup>Laboratorio de Bioenergética y Envejecimiento Celular, Departamento de Ciencias de la Salud, Universidad Autónoma Metropolitana Iztapalapa, Ciudad de México 09340, México. <sup>3</sup>Departamento de Sistemas Biológicos, División de Ciencias Biológicas y de la Salud, Universidad Autónoma Metropolitana-Xochimilco, Ciudad de México 04960, México. E-mails: <sup>b</sup>norm@xanum.uam.mx, <sup>a</sup>jp villanueva@correo.xoc.uam.mx

### RESUMEN

El desarrollo de algoritmos para su utilización en la inteligencia artificial y la disponibilidad masiva de datos biomédicos han impulsado y acelerado la investigación, con el resultado de descubrimientos y de tratamientos novedosos utilizando fármacos patentados y aprobados para el tratamiento de diferentes enfermedades. Este proceso se conoce como reposicionamiento de fármacos (REFA), y puede ser abordado por una rama de la inteligencia artificial (IA) conocida como aprendizaje automático (AA). El aprendizaje automático se basa en un conjunto de algoritmos que, combinados con técnicas computacionales bien establecidas en el campo del descubrimiento de fármacos, han sido capaces de dar lugar, con una alta eficacia, a nuevas propiedades y relaciones farmacológicas anteriormente desconocidas. Así, se han identificado nuevos blancos y tratamientos potenciales contra diversos padecimientos como el cáncer, y las enfermedades neurodegenerativas e infecciosas, entre otras. El objetivo de este trabajo de revisión es contribuir a la literatura en español sobre el uso de la inteligencia artificial y el aprendizaje de máquina en la investigación para el reposicionamiento de fármacos.

**Palabras clave:** reposicionamiento de fármacos, inteligencia artificial, descubrimiento de fármacos, aprendizaje automático, quimioinformática.

### Artificial intelligence in drug repurposing research

### ABSTRACT

The development of sophisticated artificial intelligence algorithms and the massive availability of biomedical data have driven accelerated research to discover novel treatments using drugs previously developed and approved for other uses. This process is known as drug repositioning (DR), and it can be addressed by a branch of artificial intelligence (AI) known as machine learning (ML). Machine learning is based on a set of algorithms that, combined with well-established computational techniques in the field of drug discovery, have been able to infer new previously unknown pharmacological properties and relationships with high efficiency. Thus, new targets and potential treatments against various diseases such as cancer and neurodegenerative and infectious diseases have been identified. The objective of this review is to contribute to the literature in Spanish on the use of Artificial Intelligence and machine learning in drug repositioning research.

**Keywords:** drug repurposing, artificial intelligence, drug discovery, machine learning, chemoinformatics.

## INTRODUCCIÓN

**E**l desarrollo de fármacos es un proceso complejo que consta de varias etapas y donde participan diversas áreas del conocimiento. Actualmente, un estudio de este tipo puede tomar entre 10 y 15 años, tener un costo promedio de \$1,350 millones de dólares (DiMasi, Grabowski & Hansen, 2016) y, aun así, tener una probabilidad de fallar del 90% al 96% (Hingorani *et al.*, 2019); sin embargo, las investigaciones continúan para mejorar los métodos que generen resultados positivos. El desarrollo de fármacos se divide en dos grandes fases: la primera es la investigación que comprende la identificación, caracterización y validación del blanco o blancos moleculares. No obstante, en los últimos 15 años el interés por desarrollar fármacos no se ha limitado a compuestos altamente selectivos (un solo blanco), en contraste, un enfoque de búsqueda de compuestos dirigidos a múltiples dianas terapéuticas ha adquirido fuerza (Ramsay, Popovic-Nikolic, Nikolic, Uliassi & Bolognesi, 2018). El siguiente paso es la identificación y la optimización de las moléculas con actividad biológica que hasta hace 10 años era realizado principalmente por el ensayo masivo de moléculas sobre modelos celulares o animales (cribado fenotípico); sin embargo, este proceso se agilizó con los avances en los métodos computacionales, principalmente en el campo de la inteligencia artificial (IA). Lo anterior se traduce en un ahorro de recursos y tiempo, además de producir mejores resultados como fármacos con menores efectos secundarios (Paul *et al.*, 2021); La segunda etapa se centra en los ensayos clínicos, la formulación, el proceso comercial y el registro del medicamento novedoso (Parvathaneni, Kulkarni, Muth & Gupta, 2019) (Figura 1).

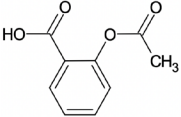
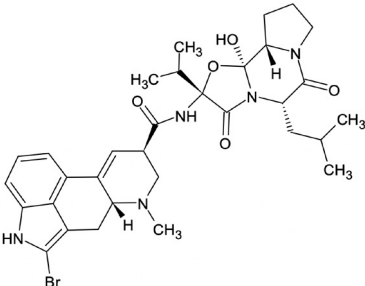
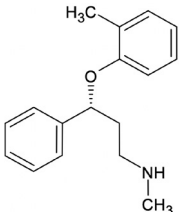
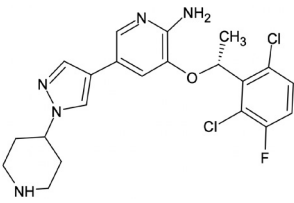
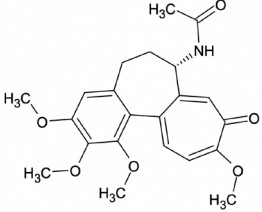
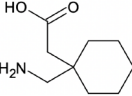
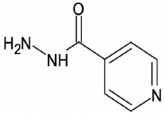
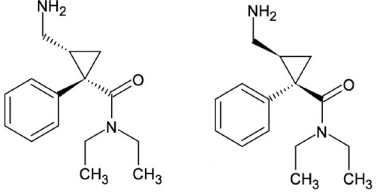
Con el objetivo de reducir el tiempo de investigación y desarrollo (I+D), surgió una estrategia conocida como reposicionamiento de fármacos (REFA), que consiste en encontrarles distintas aplicaciones terapéuticas, diferentes a las descritas originalmente. El mejor ejemplo del REFA es el Sildenafil (Viagra), que originalmente fue desarrollado para tratar la angina de pecho, pero posteriormente resultó ser efectivo para tratar la disfunción eréctil (Pushpakom *et al.*, 2019). En la Tabla I se muestran algunos ejemplos de fármacos exitosamente reposicionados.

Las principales ventajas del REFA son la reducción del tiempo de investigación e inversión económica, así como el descubrimiento de tratamientos menos costosos. Por ejemplo, la combinación de Sirolimus y Vorinostat con un costo de \$7,696 dólares, tienen una efectividad para tratar el linfoma de Hodgking (un tipo de cáncer del sistema linfático) similar al Brentuximab, recientemente aprobado por la Administración de Alimentos y Medicamentos de los Estados Unidos (FDA) cuyo costo es de \$13,683 dólares (Kato *et al.*, 2015). Sin embargo, el REFA también tiene ciertas desventajas y retos que superar, concernientes principalmente a la dosificación y a la seguridad como, por ejemplo: a) con el nuevo uso se debe de tener una dosis terapéutica cercana a la prescripción original con probados beneficios y riesgos mínimos, b) también implica tener la capacidad de dirigirlo a otros órganos diana y por ende desarrollar nuevas fórmulas o mecanismos de liberación, incluso se tiene una pobre integración con las ciencias toxicológicas y farmacéuticas, c) existen dificultades legales y de financiamiento; el REFA disminuye el margen



Figura 1. Proceso del desarrollo de un fármaco.

Tabla I. Casos exitosos de reposicionamiento de fármacos.

| Fármaco                | Estructura                                                                          | Indicación original                                 | Nueva indicación                        | Referencia                                       |
|------------------------|-------------------------------------------------------------------------------------|-----------------------------------------------------|-----------------------------------------|--------------------------------------------------|
| Ácido acetilsalicílico |    | Dolor e inflamación                                 | Cáncer de próstata                      | (Joshi, Murphy, Olaniyi & Bryant, 2021)          |
| Bromocriptina          |    | Enfermedad de Parkinson                             | Diabetes Mellitus                       | (Ooi & Lye, 2018)                                |
| Atomoxetina            |   | Trastorno de déficit de atención con hiperactividad | Alzheimer                               | (Levey <i>et al.</i> , 2021)                     |
| Crizotinib             |  | Linfoma                                             | Cáncer de pulmón de células no pequeñas | (Chuang & Neal, 2015)                            |
| Colchicina             |  | Gota                                                | Pericarditis                            | (Thompson, 2019)                                 |
| Gabapentina            |  | Epilepsia                                           | Dolor neuropático                       | (Wiffen <i>et al.</i> , 2017)                    |
| Isoniazida             |  | Tuberculosis                                        | Cáncer de próstata                      | (Lv <i>et al.</i> , 2018)                        |
| Milnacipran            |  | Depresión                                           | Fibromialgia                            | (Cording, Derry, Phillips, Moore & Wiffen, 2015) |

de ganancias de las empresas farmacéuticas debido a que los fármacos económicos pueden ser reutilizados en lugar de los fármacos novedosos y costosos, desalentando la inversión para su investigación y d) la protección de los derechos intelectuales e industriales representa grandes retos (Oprea *et al.*, 2011).

El REFA emplea aproximaciones experimentales como el ensayo de afinidad utilizando, por ejemplo, técnicas de proteómica como la cromatografía de afinidad y la espectrometría de masas que permiten validar blancos farmacológicos, mecanismos de acción y sitios de unión específicos (Pushpakom *et al.*, 2019). Otra metodología es el cribado fenotípico *in vitro* donde se usan cultivos celulares o *in vivo* empleando modelos animales sobre los que se prueban moléculas y se seleccionan aquellas capaces de alterar el fenotipo del sistema en estudio. La ventaja de este método es que puede ser automatizado y ensayar miles de moléculas de manera simultánea, incluso, a diferentes rangos de concentraciones. Sin embargo, es muy costoso, aunque el más empleado, ya que más del 50% de las moléculas pequeñas aprobadas por la FDA han sido identificadas mediante esta técnica (Moffat, Vincent, Lee, Eder & Prunotto, 2017; Rudrapal, Khairnar & Jadhav, 2020).

De manera importante, el REFA también se auxilia de las ventajas de los métodos por computación para acelerar y refinar el proceso de descubrimiento de nuevas funciones para los fármacos. Entre las técnicas más empleadas se encuentra la búsqueda por similitud química o molecular de compuestos, que se basa en consideraciones topológicas, farmacofóricas o fragmentos estructurales de las representaciones moleculares 2D y 3D (Willett, Barnard & Downs, 1998). Otra estrategia, es el acoplamiento molecular que emplea la estructura de una diana farmacológica para encontrar la posición energéticamente más favorable de interacción en la que se encuentra un complejo ligando-receptor (Meng, Zhang, Mezei & Cui, 2011). La dinámica molecular es otro método ampliamente empleado que sigue un principio parecido al acoplamiento molecular, sin embargo, esta técnica considera flexibles tanto al ligando como al receptor, además de tomar en cuenta factores como iones y moléculas de agua presentes en el ambiente celular y el tiempo que duran éstos unidos al complejo asociado (Liu *et al.*, 2018). Más información relacionada con los métodos clásicos en el descubrimiento de fármacos se puede encontrar en el trabajo de Saldívar-González, Prieto-Martínez & Medina-Franco (2017).

Otra forma de abordar la investigación en el REFA cuando no se tiene un blanco definido es a través de las firmas transcriptómicas en plataformas como Cmap o LINCS L1000 que usan perfiles de expresión generados por microarreglos o secuenciación de RNAm. De éstos, los perfiles de perturbación transcriptómica son los más comunes y consisten en determinar los cambios en la expresión de diversos genes cuando son sometidos

a estímulos externos como el silenciamiento por RNA de interferencia (siRNA) o por fármacos (Lamb *et al.*, 2006; Subramanian *et al.*, 2017). El análisis de las redes biológicas del tipo interacción proteína-proteína, co-expresión, regulación genética, o también el análisis de redes farmacológicas, permite visualizar e inferir relaciones e interacciones fármaco-blanco, fármaco-fármaco y fármaco-enfermedad (Morrow, Tian & Zhang, 2010; Emig *et al.*, 2013). Por último, es importante mencionar la minería de datos y el aprendizaje automático (AA) (March-Vila *et al.*, 2017) que pertenecen a un área de las ciencias de la computación de la mencionada IA, que ha ganado popularidad en los últimos años debido a su capacidad de análisis masivo de información, gracias al desarrollo de un equipo de cómputo de alto rendimiento y diversos algoritmos entrenados con una enorme cantidad de datos biomédicos, químicos y biológicos actualizados (Sidey-Gibbons & Sidey-Gibbons, 2019).

El uso de la IA en el REFA se ha discutido ampliamente en la literatura en inglés (Tanoli, Vähä-Koskela & Aittokallio, 2021; Urbina, Puhl & Ekins, 2021; Koromina, Pandi & Patrinos, 2019; Mohapatra *et al.*, 2020), sin embargo, existen pocos trabajos en español sobre este tema (Saldívar-González & Medina-Franco, 2021).

En esta revisión también nos enfocamos en describir el uso y las aplicaciones actuales de los algoritmos del AA mostrando que por sí solos y en combinación con las metodologías computacionales clásicas han mejorado y acelerado el proceso del REFA al descubrir una nueva utilidad en el tratamiento de otras enfermedades. Presentamos algunos ejemplos de éxito en enfermedades infecciosas y crónico-degenerativas.

#### APRENDIZAJE AUTOMÁTICO

La IA es un campo de la computación cuyo objetivo es lograr que las máquinas tomen decisiones a través de la experiencia y la adaptación por medio de la simulación de los procesos de razonamiento humano. Los orígenes de la IA se remontan a 1943 con el trabajo de McCulloch y Pitts, donde describían un modelo matemático de aprendizaje basado en el comportamiento de las neuronas humanas (Muthukrishnan *et al.*, 2020), sin embargo, el término IA fue introducido por primera vez en 1956 por John McCarthy (Kaul, Enslin & Gross, 2020). La IA ha sido empleada en diversas áreas como la ingeniería, economía, industria, etc. Pero fue hasta el año 2000 que comenzó a usarse de manera intensiva en las ciencias médicas y químico biológicas gracias al desarrollo de equipo de cómputo de altas prestaciones, así como la disposición masiva de datos clínicos, químicos y biológicos depositados en bases de datos públicas, que ayudaron a mejorar y acelerar el desarrollo y el descubrimiento de nuevos fármacos, métodos de síntesis, vías de señalización y blancos farmacológicos (Kononenko, 2001; Gasteiger, 2020). La IA procesa datos mediante algoritmos que emplean herramientas matemáticas tan diversas como el álgebra

lineal, la estadística, la teoría de la probabilidad, y el análisis numérico para razonar y resolver problemas, hacer inferencias basadas en eventos pasados, aprender nuevas tareas, clasificar e identificar patrones complejos, y extraer conocimientos de grandes volúmenes de datos (Baştanlar & Ozuysal, 2014).

El AA se centra en el desarrollo de sistemas que aprenden mediante la identificación de patrones y en función de los datos disponibles con el objetivo de hacer predicciones. El aprendizaje se realiza mediante la construcción de modelos, que son un conjunto de operaciones y parámetros de entrada y salida de datos, y para entrenarlos se utilizan diversos algoritmos. Un flujo de trabajo clásico del AA consiste en cuatro pasos: (1) *El preprocesamiento de datos*; en donde se limpia, transforma y reducen los datos con el objetivo de filtrar o eliminar los de baja calidad, reducir el ruido, remover valores anormales o duplicados para mejorar la calidad de los resultados; entre las técnicas comúnmente utilizadas se encuentra la normalización, fusión y reducción de datos, estandarización e imputación de valores perdidos (Malley, Ramazzotti & Wu, 2016). (2) *La extracción de características*; se selecciona un subconjunto de las características totales de los datos originales mediante la reducción de la dimensionalidad usando técnicas como análisis de componentes principales (PCA) (Jolliffe & Cadima, 2016), incrustación de vecinos estocásticos distribuidos en t (t-SNE) (Linderman & Steinerberger, 2019) y escalamiento multidimensional (MDS) (Mugavin, 2008), etc., con el objetivo de eliminar características altamente correlacionadas, redundantes o irrelevantes, reduciendo el tamaño y la complejidad del modelo para hacer el entrenamiento más fácil y rápido (Saeys, Inza & Larrañaga, 2007). (3) *El entrenamiento del modelo*; en el que se utilizan dos tipos principales de algoritmos llamados aprendizaje supervisado y no supervisado. En el aprendizaje supervisado, que es el más utilizado, los modelos se entrenan mediante datos previamente etiquetados y su resultado está predefinido. Algunos ejemplos de aprendizaje automático supervisado incluyen algoritmos de regresión, empleados para modelar la relación entre una variable dependiente y una o más variables predictivas. Entre los algoritmos de regresión, están: a) el lineal que determina la tendencia de un conjunto de datos continuos, b) el de logística que se utiliza para datos discretos y c) la logarítmica que es una generalización de la regresión lineal para datos multivariados (Doan & Kalita, 2015). Las máquinas de soporte vectorial o Support Vector Machines (SVM) son un conjunto de algoritmos de aprendizaje supervisado, un método de clasificación y regresión binaria basado en espacios p-dimensionales llamados hiperplanos que son subespacios de una sola dimensión que no pasan por el origen, y se aplican en problemas de clasificación múltiple (Hearst, Dumais, Osuna, Platt & Scholkopf, 1998). Los árboles de decisión son algoritmos predictivos del valor de una variable en función de otras variables de entrada; las predicciones son realizadas mediante métodos probabilísticos como el de bosques aleatorios o Random Forest (RF) que son

un conjunto de árboles de decisión en donde cada árbol ve una porción de los datos de entrenamiento, y al final, al combinar los datos de cada árbol aumenta el poder de generalización mientras disminuye el error. Los algoritmos de aprendizaje no supervisado son capaces de identificar procesos y patrones complejos, sin la necesidad de entrenarlos con datos etiquetados o resultados predefinidos. Entre estos algoritmos encontramos k-medias, tSNE, PCA y Linear Discriminant Analysis (LDA). La mayoría de estos algoritmos funcionan a través de la reducción de la dimensionalidad que implica transformaciones lineales (Dash, Nayak & Mishra, 2021). En la Figura 2 se resumen los principales algoritmos del AA. (4) *La evaluación del modelo*; es necesario evaluar los modelos de aprendizaje para determinar su desempeño y realizar predicciones acertadas con futuros datos. Al llevar a cabo esta validación se separan del modelo los datos de los que ya se conoce una respuesta, y se etiquetan con valores reales como comprobación de la capacidad de generalización. Entre las métricas empleadas para evaluarlos encontramos: exactitud, precisión y exhaustividad que cuantifican la tasa de falsos positivos y negativos, puntuación F1 (media armónica de precisión y exhaustividad), curva de características operativas del receptor (ROC), índice Jaccard y coeficiente de Ginni (Handelman *et al.*, 2019).

Los algoritmos del AA han acelerado el descubrimiento de fármacos, sin embargo, existen algunas limitaciones a la hora de entrenar los modelos, como la falta de información experimental de blancos farmacológicos novedosos que dificulta la capacidad de extrapolación de los datos. También en el área de la quimioinformática se complica el agrupamiento de moléculas (espacio químico) ocasionado por el uso de descriptores moleculares incorrectos, o que otorgan poca información (Baskin, 2021). Otra limitación intrínseca del AA es la generación de modelos con poca capacidad de generalización (sobreajuste), o modelos incapaces de predecir e identificar patrones (infraajuste). Estas limitaciones se originan principalmente por la mala calidad, poca cantidad y variedad de datos, así como a un mal preprocesamiento, ajuste de parámetros y elección del algoritmo para entrenar el modelo (Patel, Shukla, Huang, Ussery & Wang, 2020).

## IDENTIFICACIÓN DE BLANCOS TERAPÉUTICOS

La identificación de blancos adecuados es una tarea crítica y complicada debido a la etiología de las enfermedades. Para abordar este problema, diversos autores han desarrollado estrategias para identificar y priorizar blancos realmente efectivos. Ferrero, Dunham & Sanseau (2017) desarrollaron un modelo de clasificación que predice nuevos blancos terapéuticos a partir de datos de asociación gen-enfermedad. Para construir este modelo, se utilizaron cuatro algoritmos: SVM, RF, máquina de potenciación del gradiente o Gradient Boosting Machine (GBM) y redes neuronales artificiales o Artificial Neural Networks (ANNs) entrenados con datos de asociación gen-enfermedad provenientes de bases de

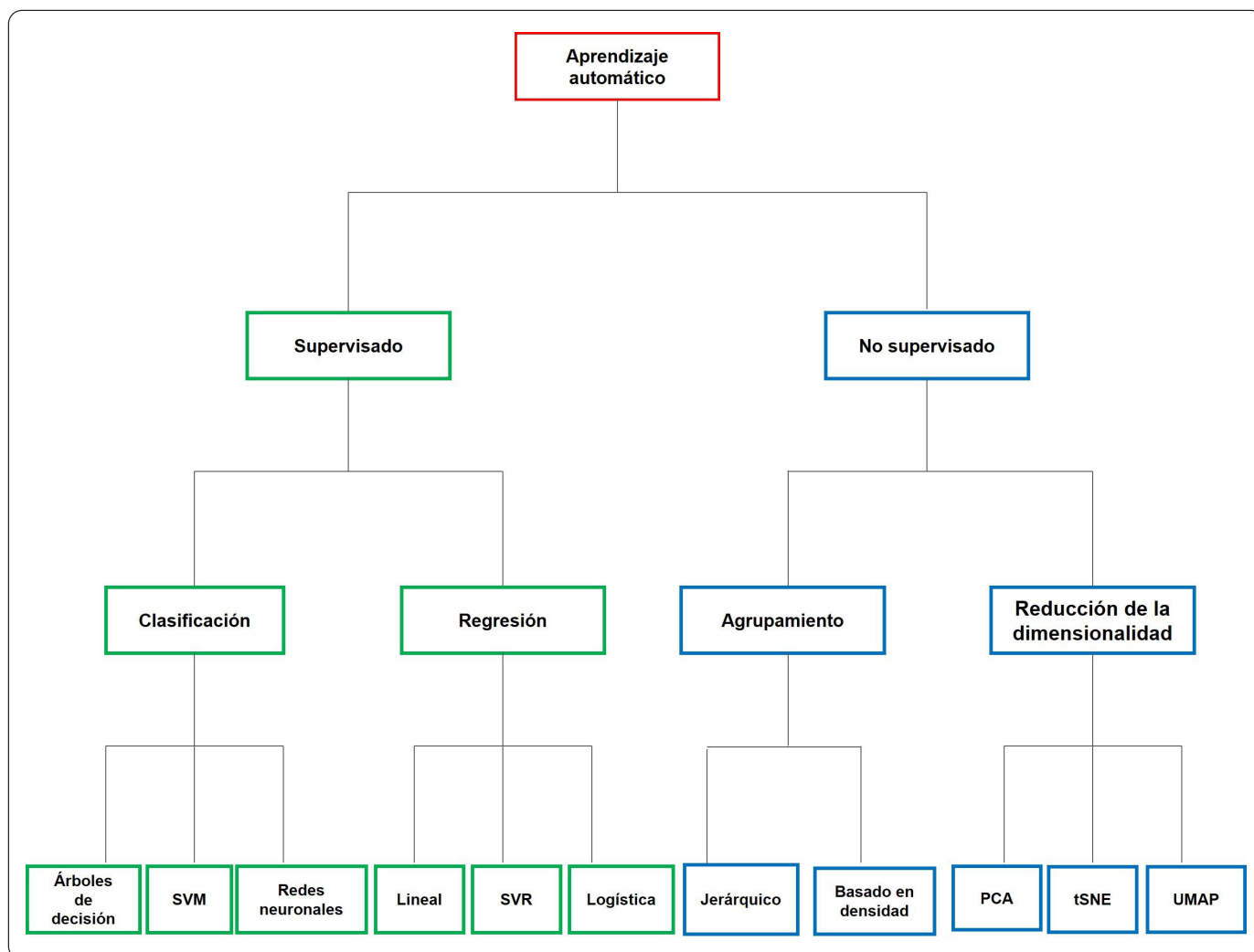


Figura 2. Resumen de los principales algoritmos de aprendizaje automático.

datos públicas. Los datos de entrada que se utilizaron fueron mutaciones, expresión del RNA, vías de señalización, modelos animales y asociaciones genéticas. Los autores encontraron que los cuatro algoritmos tenían una eficacia similar cercana al 70% (Ferrero *et al.*, 2017). Por otra parte, Dezső & Ceccarelli (2020) desarrollaron y entrenaron más de 10 mil modelos utilizando RF para la predicción de blancos terapéuticos oncológicos novedosos. Tomando en cuenta un total de 277 blancos conocidos de fármacos oncológicos aprobados y en fase experimental para la construcción de los modelos, se utilizaron 70 características entre las que destacan la localización subcelular, las modificaciones postraduccionales, la clasificación enzimática, las regiones PEST (regiones ricas en residuos de prolina (P), glutamato/aspartato (E/D), serina (S) y treonina (T) flanqueadas por aminoácidos básicos), la estructura secundaria, la accesibilidad del solvente y la especificidad de tejido, así como la información topológica de redes de interacción proteína-proteína. Después de la validación

y selección del mejor modelo, se obtuvo una eficacia del 73%. Mamoshina *et al.* (2018) desarrollaron diversos modelos predictivos para encontrar blancos asociados con la progresión de la edad y hacer estudios dirigidos al reposicionamiento de fármacos contra estos blancos responsables del envejecimiento. Utilizando datos de expresión provenientes de experimentos con microarreglos depositados en los repositorios de Gene Expression Omnibus (GEO) y en ArrayExpress, se evaluó la eficacia de cinco algoritmos distintos: k vecinos más cercanos (k-nn) – 58%, RF – 66%, ElasticNet – 83%, SVM – 87% y selección de características profundas – 80%. El uso de estos algoritmos les permitió identificar un total de veinte blancos, de los que algunos ya tienen fármacos que interactúan con ellos.

## IDENTIFICACIÓN DE FÁRMACOS

### Modelos basados en medidas de similitud

Como se mencionó anteriormente, el REFA utiliza diversos enfoques computacionales para el proceso de descubrimiento

de nuevas funciones de los fármacos ya conocidos. Muchos métodos se han combinado con el uso del AA y su eficacia ha mejorado. Son tres los principales métodos que se han visto beneficiados por el AA basados en: (1) el ligando que a su vez abarca: (a) las relaciones cuantitativas estructura-actividad o Quantitative Structure–Activity Relationship (QSAR) y (b) el de búsqueda por similitud; (2) la estructura como el acoplamiento molecular, y (3) los quimiogenómicos (Kim *et al.*, 2020).

Gottlieb, Stein, Ruppín & Sharan (2011) desarrollaron PREDICT, un modelo capaz de identificar asociaciones fármaco-gen-enfermedad utilizando medidas de similitud y regresión logística, así como datos de entrenamiento de perfiles de expresión génica, redes de interacción proteína-proteína, y análisis de funciones ontológicas. El modelo tiene una eficacia cercana al 90%. Otro modelo de predicción con un concepto similar llamado SPACE también utiliza medidas de similitud y asociación entre diversas fuentes de datos para predecir la clasificación terapéutica de los fármacos aprobados para inferir nuevas funciones (Liu *et al.*, 2015). Utilizando datos de Cmap, Napolitano *et al.* (2013) combinaron datos de similitud de perturbaciones transcriptómicas inducidas por fármacos, asociaciones gen-enfermedad y estructura química de los fármacos en un modelo multiclase o Semi-Supervised Multiclass (SVM). Kim, Choi & Nam (2019) utilizaron algoritmos de clasificación lineales y no lineales para encontrar asociaciones fármaco-enfermedad de productos naturales; los algoritmos lineales utilizados fueron regresión logística y el SVM con núcleo lineal. Los no lineales utilizados fueron los RF, regresión polinomial y el SVM con núcleo no lineal. Un hallazgo interesante fue que los algoritmos de clasificación no lineal dieron mejores resultados con una precisión cercana al 90%. Al entrenar los distintos modelos se utilizaron doce características llamadas puntuaciones de similitud entre las que se encuentran similitudes de: estructura química, efectos secundarios, blancos farmacológicos, términos ontológicos, ontológica de fenotipo y términos ontológicos de genes relacionados con enfermedades, etc. Los modelos basados en similitud entrenados usando algoritmos de regresión fueron de los primeros en ser implementados y son ampliamente usados (Li & Lu, 2012; Luo *et al.*, 2016).

### Modelos basados en redes

Una red es una abstracción matemática de la conexión entre dos objetos. Las redes se representan típicamente como grafos, los cuales están compuestos de nodos y aristas. Cada nodo representa un objeto y está unido a otros nodos mediante una arista que significa algún tipo de relación. En el REFA es común encontrar redes fármaco-enfermedad, fármaco-fármaco, fármaco-proteína, proteína-proteína y proteína-enfermedad, etc. Este tipo de modelos se han utilizado desde hace tiempo, pero recientemente se han combinado con el uso de algoritmos de AA lo cual aumenta de manera significativa la eficacia de las redes. Zhang *et al.* (2018) desarrollaron un método de inferencia

basado en la similitud topológica de red o Network Topological Similarity-Based Inference Method (NTSIM), que es un modelo de inferencia basado en redes bipartitas de asociaciones fármaco-enfermedad capaz de predecir nuevas asociaciones fármaco-blanco con gran exactitud. El modelo fue construido utilizando un perceptrón multicapa o Multilayer Perceptron (MLP) y RF. Le & Nguyen-Ngoc (2018) construyeron una red de tres capas utilizando similitud de enfermedades y fármacos, para inferir nuevas asociaciones fármaco-enfermedad mediante un modelo de aprendizaje conocido como mínimos cuadrados recursivos o Recursive Least Squares (RLS), que es un método de aprendizaje semi-supervisado capaz de predecir nuevas asociaciones con un 97% de eficacia. Himmelstein *et al.* (2017) desarrollaron Hetionet, un modelo capaz de predecir más de 200 mil nuevas relaciones fármaco-enfermedad. Este sistema multi-redes integra millones de estudios biomédicos utilizando veinticuatro tipos de relaciones diferentes como enfermedades, genes, anatomía, vías de señalización, procesos biológicos, funciones moleculares, componentes celulares, clases farmacológicas, efectos adversos y síntomas, entre otros. El modelo fue entrenado con un algoritmo de regresión logística y mostró una eficacia cercana al 85%.

### Modelos basados en aprendizaje profundo

El aprendizaje profundo (AP) es un subconjunto de algoritmos y técnicas de AA que han revolucionado el campo de la IA y el descubrimiento de fármacos. El AP hace un uso intensivo de estructuras abstractas que implican múltiples transformaciones lineales a través de diversas capas de ANNs compuestas por neuronas artificiales interconectadas a través de enlaces, de los que salen valores multiplicados por un valor de peso determinado por la función de activación. Posteriormente, estos valores se propagan a través de las neuronas aprendiendo cómo regular la ganancia y la pérdida de información. Las arquitecturas principales y más populares son la red neuronal profunda o Deep Neural Network (DNN), la red de creencia profunda o Deep Belief Network (DBN), la red neuronal recurrente o Recurrent Neural Network (RNN), la red de memoria a largo plazo o Long Short-Term Memory Network (LSTM) y la red neuronal convolucional o Convolutional Neural Network (CNN). Una de las principales ventajas que puede ofrecer el AP frente al AA en el descubrimiento de fármacos es que éste no necesita el proceso de selección de características, evitando de esta manera el trabajo con conjuntos de datos excesivamente grandes y redundantes. La aplicación del AP en esta área, incluso, permite el descubrimiento de nuevas características en modelos de predicción fármaco-enfermedad (Schmidhuber, 2015). Aliper *et al.* (2016) utilizaron un modelo basado en una DNN que entrenaron con datos de perturbación transcriptómica de 678 fármacos sobre las líneas celulares A549, MCF-7 y PC-3. El primer logro notable de este modelo fue la capacidad de clasificar cada tipo de fármaco con una exactitud diez veces superior a un modelo basado en un SVM. Además, el modelo pudo reconocer fármacos antineoplásicos, fármacos para el sistema

nervioso central y de uso dermatológico. Wei, Zhang, Huang & Fang (2019) desarrollaron un algoritmo de incrustación de redes o Network Embedding (NE) utilizando datos de similitud estructural, interacción fármaco-enfermedad, efectos adversos y datos de enfermedades mendelianas para descubrir nuevas características latentes en los fármacos conocidos, teorizaron que podrían descubrirse nuevos usos, por lo que fueron utilizados posteriormente para entrenar modelos de un SVM con una exactitud superior al 90%. Otro flujo de trabajo diseñado para encontrar nuevas características útiles para estudios sobre reposicionamiento de fármacos fue propuesto por otro grupo de investigación; este método tiene una eficacia del 94%, y consiste de cinco pasos, (1) Representación de cuatro características de los fármacos (estructura, enzimas, proteínas relacionadas y niveles de expresión génica) con las que se entrenó una DNN, (2) Transformación de dos características de enfermedades (fenotipo y genotipo) usando PCA, (3) Construcción de matrices de similitud fármaco-fármaco utilizando las características de los fármacos, (4) Construcción de matrices de similitud enfermedad-enfermedad basadas en las características de las enfermedades, y (5) Construcción de matrices de asociación usando las dos matrices de similitud anteriores (Moridi, Ghadirinia, Sharifi-Zarchi & Zare-Mirakabad, 2019). Donner, Kazmierczak & Fortney (2018) idearon un método de embebido profundo con datos de perturbación transcriptómica; utilizando valores estandarizados de expresión de los genes de referencia, entrenaron una DNN con múltiples capas para efectuar diversas tareas, entre ellas el reposicionamiento basado en similitud de perturbación. Este tipo de red puede reducir el ruido en los perfiles de perturbación y dar una precisión del 90%. DeepAction es un procedimiento para encontrar nuevas interacciones fármaco-blanco con una efectividad del 98%. El modelo consiste en extraer características de las moléculas usando descriptores moleculares además de las características geométricas y topológicas; también extrajeron características de los blancos identificados de cada molécula. Utilizando un modelo de regresión de operador de selección y contracción mínima absoluta (LASSO), seleccionaron las características más importantes para entrenar una CNN (Hasan Mahmud *et al.*, 2020). Otro modelo llamado deepDR combina análisis de redes con el AP; en total, utilizaron 10 tipos de redes, entre ellas: fármaco-enfermedad, fármaco-efectos adversos, fármaco-blanco y siete redes distintas fármaco-fármaco. El modelo fue entrenado con un codificador automático profundo multimodal para aprender características de alto nivel y convertirlas en representaciones de baja dimensión utilizadas en un Autoencoder Variacional (VAE) con el objetivo de reposicionar fármacos contra las enfermedades de Parkinson y de Alzheimer (Zeng *et al.*, 2019). Otra aproximación es el método basado en el AP para predecir indicaciones de enfermedades candidatas relacionadas con fármacos llamado CGARDP. Este modelo, entrena una CNN con datos de similitud de fármacos y una unidad recurrente cerrada (GRU)

con datos de similitud de enfermedades; posteriormente estas dos RNN fueron combinadas y se utilizaron para su entrenamiento en cincuenta millones de parámetros. Un hecho interesante es que este modelo es capaz de realizar predicciones de asociaciones fármaco-enfermedades realistas en comparación con otros modelos de redes neuronales (Xuan, Zhao, Zhang, Ye & Zhang, 2019). Un sistema distinto fue propuesto por Liu, Wei & Zhang, 2021, que consiste en utilizar datos longitudinales de observaciones provenientes de pacientes heterogéneos; algunos de estos datos son encuestas de salud, historiales médicos electrónicos, actividades de reclamación y facturación que fueron emulados para cada fármaco candidato a reposicionamiento. Utilizando métodos de inferencia causal y AP, estimaron los efectos para cada fármaco utilizando una RNN de memoria a largo plazo o Long Short-Term Memory (LSTM) entrenada con datos de noventa millones de pacientes alrededor del mundo, lo que les permitió estimar nuevos efectos en diversos fármacos.

### **Casos de estudio con fines de reposicionamiento guiado por inteligencia artificial en diversas enfermedades**

#### **Enfermedades del sistema nervioso central**

En la Figura 3, se muestran las estructuras químicas de algunos fármacos que actualmente se encuentran en investigación para ser reposicionados con diferentes usos. Utilizando una librería química de 1,280 fármacos aprobados por la FDA y modelos de inferencia Bayesiana, Ekins y colaboradores encontraron que Nicardipina y Nivaldipina, fármacos originalmente utilizados como vasodilatadores e hipotensivos, son buenos candidatos para tratar el síndrome de Pitt-Hopkins causado por una rara mutación genética en el gen que codifica el factor de transcripción TCF4, lo que ocasiona diversos trastornos parecidos al síndrome de espectro autista. Además, encontraron otros 55 fármacos potenciales para tratar esta afección, todos relacionados con la inhibición de canales de voltaje implicados en ataques epilépticos que son una comorbilidad común en este síndrome (Ekins *et al.*, 2019). La segunda enfermedad más común del sistema nervioso central (SNC) en personas mayores es el Parkinson, que actualmente no tiene cura y los tratamientos son limitados; por esta razón, Sebastián-Pérez y colaboradores utilizaron los RF y una RNN conocida como multiperceptrón en combinación con modelos QSAR para identificar inhibidores putativos de la proteína LRRK2. El gen que codifica a esta proteína se encuentra comúnmente mutado en pacientes con Parkinson y se ha relacionado con una exacerbada respuesta inflamatoria y aumento del estrés oxidante en el cerebro (Sebastián-Pérez *et al.*, 2019). Otro grupo utilizó RF, SVM y K-vecinos más próximos o K-Nearest-Neighbor (KNN) y una RNN para hallar nuevos fármacos contra la ansiedad, la depresión y la esquizofrenia; estos algoritmos fueron empleados para encontrar nuevas vías de señalización que estuvieran relacionadas con estas patologías. Además, utilizaron datos de expresión génica que los llevó a proponer a la Apigenina, que es un antioxidante,

y a la Metformina, que es un fármaco para la diabetes, como nuevos candidatos en el tratamiento de la depresión/ansiedad. En cuanto a tratar la esquizofrenia propusieron al ácido acetil salicílico comúnmente utilizado para el dolor, y a la Genisteína que es un fitoestrógeno con propiedades anticancerígenas y antioxidantes (Zhao & So, 2019).

### Cáncer

El cáncer es la segunda causa de muerte a nivel mundial y es un problema grave para los sistemas de salud, pues a pesar que existen numerosas terapias, las células cancerosas son

capaces de mutar y desarrollar mecanismos de defensa contra los fármacos antineoplásicos (Mattiuzzi & Lippi, 2019). Por esta razón, numerosos grupos de investigación alrededor del mundo realizan esfuerzos constantes para encontrar novedosas terapias contra esta enfermedad, siendo el REFA uno de los enfoques mayormente utilizados. Muchos de los modelos de AA y AP antes expuestos fueron desarrollados originalmente para reposicionar fármacos contra el cáncer. Ese es el caso de la aplicación de un sofisticado algoritmo de AP conocido como autocodificador generativo adversario o Adversarial Auto Encoder (AAE) implementado en un modelo de siete

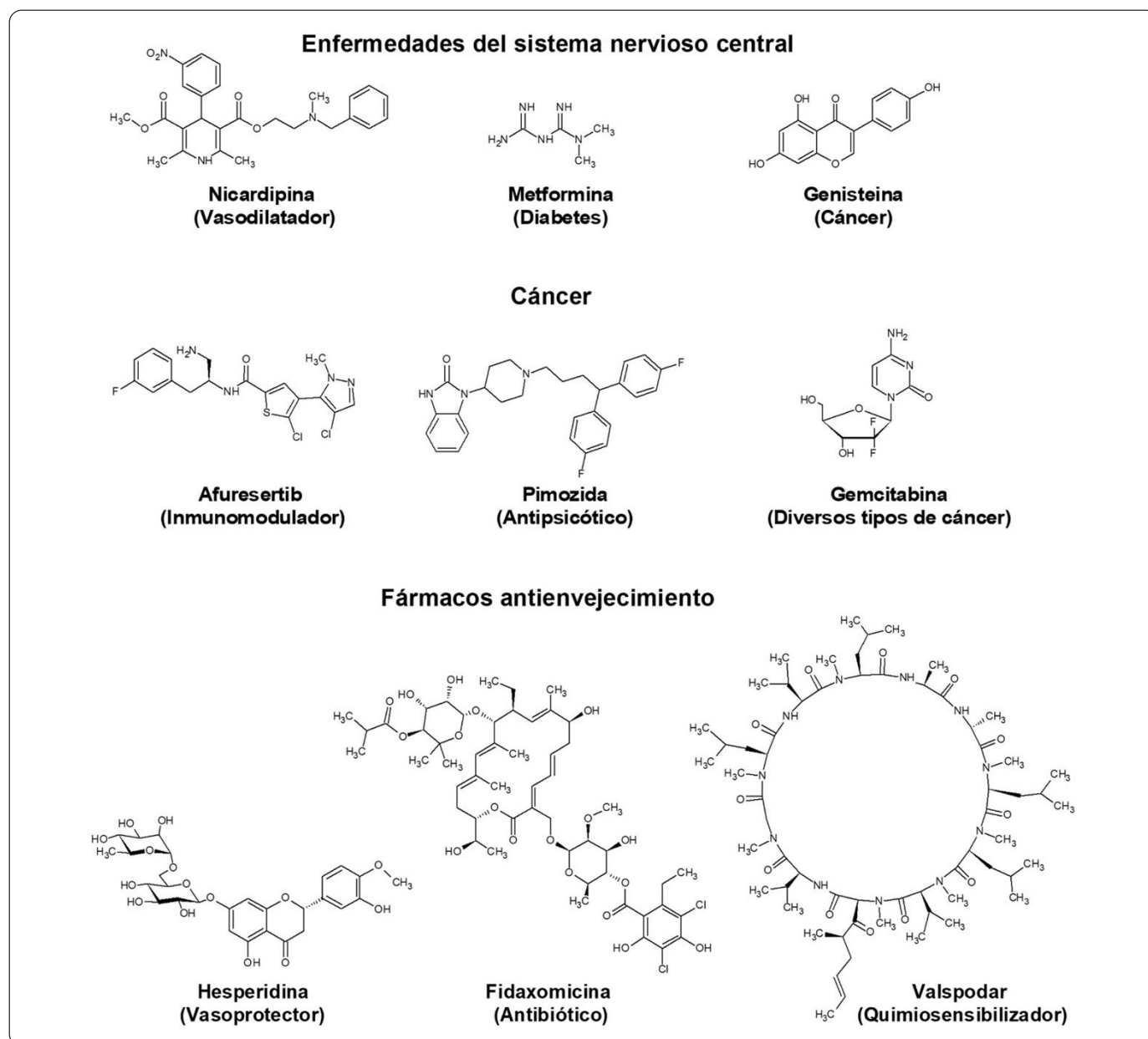


Figura 3. Estructura química de diversos fármacos bajo investigación identificados mediante algoritmos de AA que podrían ser reposicionados.

capas que sirve como discriminador para encontrar nuevas características que permitan descubrir fármacos con actividad anticancerígena dentro de una colección de 72 millones de compuestos. En total, se encontraron 632 nuevos compuestos de los cuales algunos fueron sometidos a ensayos de citotoxicidad, descubriendo efectos significativos, contra algunas líneas de células cancerosas, por parte de los fármacos que previamente habían sido reportados con actividad antimicrobiana, agentes quelantes y antivirales (Kadurin *et al.*, 2017). Deshmukh, Chandra, Singh, Siddiqi & Banerjee (2017) se concentraron en hallar inhibidores de FEN1, una proteína expresada en diversos tipos de cáncer; para esto, combinaron cribado virtual consenso con los RF y una SVM seleccionando los cinco primeros hits entre una base de datos de 53,000 compuestos identificando a JFD00950 como inhibidor novedoso de FEN1, proteína que al ser inhibida ocasiona la detención del crecimiento tumoral. La combinación del cribado fenotípico y el AA se utilizó para generar idTRAX, un modelo para identificar nuevos inhibidores de cinasas usados para tratar el cáncer de mama triple negativo, un tipo de cáncer que hasta el momento tiene pocas opciones terapéuticas y bajo índice de supervivencia (Gupta *et al.*, 2020). El modelo identificó seis nuevos inhibidores de cinasas entre los que destacan Afuresertib y Uprosertib previamente utilizados como moduladores del sistema inmunológico; además, descubrieron nuevos blancos asociados a este tipo de cáncer que no habían sido identificados previamente por métodos experimentales (Gautam, Jaiswal, Aittokallio, Al-Ali & Wennerberg, 2019). Li *et al.* (2020) encontraron el efecto de la Gemcitabina, un agente quimioterapéutico originalmente utilizado para tratar el cáncer de páncreas, y la Pimozida, un fármaco antipsicótico usado en el tratamiento del síndrome de Tourette, como nuevos agentes citotóxicos contra el cáncer de pulmón de células no pequeñas. Este descubrimiento fue posible gracias al uso de datos de expresión, perturbaciones transcriptómicas e información química de las moléculas con los que se entrenó un modelo de DNN; sus resultados fueron validados en las líneas celulares pulmonares transformadas A549 y H157.

## Enfermedades infecciosas

### Infecciones virales: el caso SARS-CoV-2

Desde la emergencia sanitaria provocada por el brote del virus SARS-CoV-2 en 2019, y la carencia de vacunas y tratamientos efectivos contra este patógeno, surgió el interés en la investigación para reposicionar fármacos contra la enfermedad COVID-19. La investigación ha sido impulsada por técnicas de alto rendimiento combinadas con algoritmos de IA que han permitido encontrar y proponer una gran variedad de tratamientos prometedores contra el virus. Delijewski & Haneczok (2021) utilizaron un conjunto de 229,000 moléculas inactivas y otras 435 eficaces contra diversos SARS-CoV para entrenar un método de aprendizaje de árbol impulsado por gradiente (XGBoost). El modelo fue alimentado con datos de similitud estructural determinada por

huellas moleculares. Con esta aproximación, estos autores proponen el fármaco Zafirlukast, que se utiliza para tratar el asma, ahora para actuar contra el SARS-CoV-2. Utilizando el mismo conjunto de datos, pero con un enfoque ligeramente distinto, Kumar, Loharch, Kumar, Ringe & Parkesh (2021) combinaron técnicas quimioinformáticas como el análisis de núcleos base, la exploración del espacio químico y las redes de similitud molecular con diversos algoritmos del AA, con las que identificaron al Fluconazol, un fármaco con acción antifúngica, y al Pranlukast, que se utiliza para tratar la congestión nasal y la rinitis, como potenciales inhibidores del SARS-CoV-2. Feng *et al.* (2020) desarrollaron un modelo de AA basado en funciones de puntuación tomando en cuenta energías de unión provenientes del cribado virtual, además de similitud y agrupamiento de fármacos. Este modelo llamado MCCS, identificó diversos fármacos potenciales para el tratamiento contra el SARS-CoV-2 como son el: Lopinavir y Tenofovir, que normalmente son utilizados para tratar el VIH, Peramivir y Zanamivir que son usados en el tratamiento contra la influenza, y el Sofosbuvir que es un antiviral contra el virus de la hepatitis C, así como los fármacos antidiabéticos Tolazamida, Gliburida y Acarbosa. Otro trabajo desarrolló una plataforma de predicción de potenciales inhibidores de este coronavirus utilizando bases de datos de compuestos activos contra otros virus como SARS, VIH e influenza; también se incluyeron datos de inhibidores de proteasas de diversos virus. Los investigadores calcularon diversos descriptores moleculares con los que se entrenó una DNN que identificó hasta ochenta moléculas con potencial para combatir el SARS-CoV-2 de las que ocho, Bedaquilina (antituberculoso), Brequinar (inmunosupresor), Celecoxib (antiinflamatorio no esteroideo), Clofazimina (antimicrobiano), Conivaptan (tratamiento para niveles bajos de sodio), Gemcitabina (antibiótico), Tolcapona (tratamiento del Parkinson) y Vismodegib (carcinoma de células basales), mostraron actividad *in vitro* contra el virus de la peridontitis infecciosa felina, un coronavirus felino muy similar al que genera la enfermedad por COVID-19 (Ke *et al.*, 2020). Otro modelo bastante novedoso fue propuesto por Loucera *et al.* (2020), que consiste en combinar algoritmos del AA con un modelado mecanicista de los circuitos de transducción de señales relacionados con la infección por SARS-CoV-2. Entre las características seleccionadas para construir el modelo están los datos de expresión de múltiples tejidos, las vías de señalización, los blancos de fármacos, la respuesta inflamatoria, la actividad inmune, la defensa antiviral, la endocitosis, la replicación, la bioenergética y las interacciones virus-hospedero. Los resultados arrojaron la posible actividad antiviral del inmunosupresor Sirolimus, el antibiótico Ciclosporina, el agente antimalárico Hidroxicloroquina, y el fármaco para tratar diversas enfermedades autoinmunes Etanercept, además de otros cinco fármacos contra el coronavirus. Diversos autores han propuesto otros modelos y aproximaciones guiadas por la IA para investigaciones enfocadas al REFA que sirvan para controlar la pandemia o reducir la mortalidad por el virus

(Zhou, Wang, Tang, Nussinov & Cheng, 2020; Mohamed, Yazdanpanah, Saghazadeh & Rezaei, 2021; Mohanty, Harun Ai Rashid, Mridul, Mohanty & Swayamsiddha, 2020).

### **Infecciones bacterianas**

El abuso y el mal uso de los antibióticos ha generado la aparición de bacterias multi-resistentes llamadas super bacterias. Este problema emergente se perfila como una causa de alta mortalidad para los siguientes años convirtiéndose en un problema sanitario y pandémico. Para tratar de encontrar soluciones tempranas, varios grupos de investigación han propuesto el enfoque del REFA para identificar agentes que sensibilicen de nuevo a las bacterias ante los fármacos, o para encontrar tratamientos bactericidas novedosos (Crunkhorn, 2020; Udaondo & Matilla, 2020; Fanelli *et al.*, 2020). Stokes *et al.*, (2020) utilizaron información de similitud molecular y espacio químico de más de cien millones de moléculas e identificaron a la Halicina, un fármaco originalmente investigado para tratar la diabetes, como un nuevo antibiótico con actividad bactericida de amplio espectro, convirtiéndose éste en el primer antibiótico descubierto utilizando la IA. Empleando métodos de agrupamiento jerárquico basado en árboles usando análisis lineal discriminante o Linear Discriminant Analysis (LDA) combinados con modelos QSAR, Suay-García *et al.* (2020) proponen que 168 fármacos previamente aprobados por la FDA podrían tener actividad bactericida contra *Escherichia coli*.

### **Infecciones parasitarias**

En cuanto a las enfermedades parasitarias se han obtenido resultados prometedores en la investigación para el REFA. Es el caso de DeepMalaria, un modelo de redes neuronales convolucionales gráficas o Graph Convolutional Neural Network (GCNN), identificó nuevos compuestos con capacidad para el tratamiento de la malaria dentro de una colección de 211 moléculas con una exactitud del 74%. Entre los nuevos compuestos propuestos se encontraron: el Omeprazol (úlcera péptica), la Loperamida (gastroenteritis), y el Amlodipino (presión arterial) (Keshavarzi Arshadi, Salem, Collins, Yuan & Chakrabarti, 2020). En otro proyecto utilizaron modelos de inferencia bayesiana para identificar 56 moléculas como fármacos potenciales contra la esquistosomiasis, una enfermedad generada por los parásitos del género *Schistosoma* que afecta al sistema venoso portal humano; entre los fármacos candidatos a reposicionamiento se encontraron: el Itraconazol (antifúngico), el Eletriptán (migraña aguda), la Rutecarpina (antiinflamatorio) y el Trilostano (síndrome de Cushing) (Zorn *et al.*, 2021). Otros algoritmos del AP y el AA han sido aplicados para identificar fármacos con el fin de reposicionarlos contra la leishmaniasis y la tripanosomiasis (Winkler, 2021).

### **Fármacos antienviejecimiento**

El envejecimiento es un fenómeno biológico complejo, progresivo y universal con una disminución de la capacidad de respuesta de los organismos ante agentes estresores,

aumentando de esta manera la acumulación de daño celular que conduce a la fragilidad y finalmente a la muerte. El envejecimiento es el principal factor de riesgo para la aparición y progresión de diversas enfermedades crónico-degenerativas como el Alzheimer, el Parkinson, la diabetes, la artritis y el cáncer, etc. (Rattan, 2006; López-Otín, Blasco, Partridge, Serrano & Kroemer, 2013). En los últimos años, los métodos computacionales para buscar fármacos capaces de retrasar el deterioro asociado al envejecimiento y las enfermedades crónico-degenerativas han ganado popularidad (Mallikarjun & Swift, 2016; Dönertaş, Fuentealba, Partridge & Thornton, 2019), algunos fármacos candidatos a reposicionamiento para evitar y ralentizar el envejecimiento se muestran en la Figura 3. Tomando en cuenta las características de la ontología de genes (GO), Barardo *et al.*, (2017) construyeron un modelo usando RF para identificar fármacos y extender la vida del gusano *Caenorhabditis elegans*; que previamente han mostrado aumentar la esperanza de vida de especies como *Drosophila melanogaster*, *Mus musculus*, *Rattus norvegicus* y *Mytilina brevispina*, etc. El modelo determinó que existe un rango del 63% al 68% de probabilidad de que los siguientes fármacos extiendan la vida del gusano: Valspodar (quimiosensibilizador), Ganirelix (control de la ovulación) y Bortezomib (mieloma múltiple). Otra interesante aplicación del algoritmo RF para hallar fármacos capaces de extender la vida fue publicada por Kapsiani & Howlin (2021) que con una exactitud superior al 80% lograron identificar los vasculoprotectores Diosmina y Hesperidina, además de los antibióticos Fidaxomicina, Rifapentina y Clortetraciclina como potenciales candidatos a ser reposicionados, ya que extendieron la vida de *Caenorhabditis elegans* hasta un 60%.

### **CONCLUSIONES**

Según Drugbank, que es la base de datos más popular para el análisis y el descubrimiento de fármacos *in silico* (<https://go.drugbank.com/>), existen 4,108 fármacos aprobados para uso humano, y 8,142 moléculas se encuentran en ensayos clínicos que podrían ser aprobadas en los siguientes años (Wishart *et al.*, 2018). Cada uno de estos fármacos es candidato para ser reposicionado, y para encontrar nuevos efectos se han desarrollado modelos sofisticados basados en algoritmos de IA, que lejos de reemplazar a los métodos computacionales y quimioinformáticos clásicos, les han dado un nuevo aire complementándolos e incluso muchas veces mejorándolos. Esto ha sido posible gracias a diversos algoritmos del AA y el AP entrenados con una enorme cantidad de datos biomédicos, químicos y experimentales. Lo que debe resaltarse es el aumento en la eficacia de los modelos gracias a las técnicas del AP que son capaces de construir modelos más robustos y menos redundantes que los algoritmos del AA. El uso de estas técnicas ha demostrado buenos resultados para identificar características y asociaciones previamente desconocidas que han servido para encontrar nuevos efectos en diferentes fármacos.

## PERSPECTIVAS

Los algoritmos del AP han ganado popularidad en los estudios enfocados al REFA, principalmente el uso de ANN. Aunque existen ciertas arquitecturas de red que han sido utilizadas, existen aún más por ser aplicadas en este campo como son las redes generativas y de confrontación, el aprendizaje incremental y las redes bidireccionales (Wang, Zhao & Pourpanah, 2020). Aunque los algoritmos del AA son los que más se utilizan en los estudios dirigidos al reposicionamiento, otras técnicas de la IA podrían ser igualmente aplicadas para mejorar las existentes; como es el caso de los algoritmos genéticos utilizados en programas de acoplamiento molecular (Gardiner *et al.*, 2001), lógica difusa para el descubrimiento de mecanismos de acción de fármacos (Sproule, Naranjo & Türksen, 2002), así como la aplicación de modelos de procesamiento de lenguaje natural (Öztürk, Özgür, Schwaller, Laino & Ozkirmli, 2020).

En esta revisión nos enfocamos en resumir los algoritmos y métodos derivados de la IA empleados en la investigación con la finalidad de encontrar fármacos candidatos a reposicionamiento. Algunos fármacos cuyo reposicionamiento fue señalado por técnicas computacionales, que indirectamente usan la IA, como el cribado virtual basado en el ligando y el receptor se encuentran actualmente en ensayos clínicos. Los casos más interesantes son los tratamientos potenciales contra el COVID-19, ya que fármacos como el Telmisartán (antihipertensivo), el Favipiravir (tratamiento VIH), y el Nafamostat (anticoagulante) han mostrado resultados prometedores al inhibir la entrada del virus o evitar cuadros graves de la enfermedad en humanos (Aronskyy, Masoudi-Sobhanzadeh, Cappuccio & Zaslavsky, 2021), lo que confirma la capacidad del REFA para tratar de mantener la salud humana.

## AGRADECIMIENTOS

Este trabajo fue apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) FORDECYT-PRONACES / 263957/2020. Kevin Samael Olascoaga Del Ángel es becario del CONACYT.

## REFERENCIAS

Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P. & Zhavoronkov, A. (2016). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Molecular Pharmaceutics*, **13**(7), 2524–2530. <https://doi.org/10.1021/acs.molpharmaceut.6b00248>

Aronskyy, I., Masoudi-Sobhanzadeh, Y., Cappuccio, A. & Zaslavsky, E. (2021). Advances in the computational landscape for repurposed drugs against COVID-19. *Drug Discovery Today*, **26**(12), 2800–2815. <https://doi.org/10.1016/j.drudis.2021.07.026>

Barardo, D. G., Newby, D., Thornton, D., Ghafourian, T., de Magalhães, J. P. & Freitas, A. A. (2017). Machine learning for predicting lifespan-extending chemical compounds.

*Aging*, **9**(7), 1721–1737. <https://doi.org/10.18632/aging.101264>

Baskin, I. I. (2021). Practical constraints with machine learning in drug discovery. *Expert opinion on drug discovery*, **16**(9), 929–931. <https://doi.org/10.1080/17460441.2021.1887133>

Baştanlar, Y. & Ozuysal, M. (2014). Introduction to machine learning. *Methods in Molecular Biology (Clifton, N.J.)*, **1107**, 105–128. [https://doi.org/10.1007/978-1-62703-748-8\\_7](https://doi.org/10.1007/978-1-62703-748-8_7)

Cording, M., Derry, S., Phillips, T., Moore, R. A. & Wiffen, P. J. (2015). Milnacipran for pain in fibromyalgia in adults. *Cochrane Database of Systematic Reviews*, **2015**(10), CD008244. <https://doi.org/10.1002/14651858.CD008244.pub3>

Crunkhorn, S. (2020). Predicting novel antibacterial agents. *Nature Reviews Drug Discovery*, **19**(4), 238. <https://doi.org/10.1038/d41573-020-00033-z>

Chuang, J. C. & Neal, J. W. (2015). Crizotinib as first line therapy for advanced ALK-positive non-small cell lung cancers. *Translational Lung Cancer Research*, **4**(5), 639–641. <https://doi.org/10.3978/j.issn.2218-6751.2015.03.06>

Dash, S. S., Nayak, S. K. & Mishra, D. (2021) A Review on Machine Learning Algorithms. In: Mishra, D., Buyya, R., Mohapatra, P., Patnaik, S. (eds) Intelligent and Cloud Computing. Smart Innovation, Systems and Technologies, Vol. **153**. Springer, Singapore. [https://doi.org/10.1007/978-981-15-6202-0\\_51](https://doi.org/10.1007/978-981-15-6202-0_51)

Delijewski, M. & Haneczok, J. (2021). AI drug discovery screening for COVID-19 reveals zafirlukast as a repurposing candidate. *Medicine in Drug Discovery*, **9**, 100077. <https://doi.org/10.1016/j.medidd.2020.100077>

Deshmukh, A. L., Chandra, S., Singh, D. K., Siddiqi, M. I. & Banerjee, D. (2017). Identification of human flap endonuclease 1 (FEN1) inhibitors using a machine learning based consensus virtual screening. *Molecular BioSystems*, **13**(8), 1630–1639. <https://doi.org/10.1039/c7mb00118e>

Dezső, Z. & Ceccarelli, M. (2020). Machine learning prediction of oncology drug targets based on protein and network properties. *BMC Bioinformatics*, **21**(1), 104. <https://doi.org/10.1186/s12859-020-3442-9>

DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, **47**, 20–33. <https://doi.org/10.1016/j.jhealeco.2016.01.012>

Doan, T. & Kalita, J. (2015). Selecting machine learning algorithms using regression models. 2015 IEEE International Conference on Data Mining Workshop (ICDMW). <https://doi.org/10.1109/icdmw.2015.43>

Dönertaş, H. M., Fuentealba, M., Partridge, L. & Thornton, J. M. (2019). Identifying potential ageing-modulating drugs *in silico*. *Trends in Endocrinology and Metabolism: TEM*, **30**(2), 118–131. <https://doi.org/10.1016/j.tem.2018.11.005>

Donner, Y., Kazmierczak, S. & Fortney, K. (2018). Drug repurposing using deep embeddings of gene expression

- pfiles.
- Molecular Pharmaceutics*
- ,
- 15**
- (10), 4314–4325.
- <https://doi.org/10.1021/acs.molpharmaceut.8b00284>
- Ekins, S., Gerlach, J., Zorn, K. M., Antonio, B. M., Lin, Z. & Gerlach, A. (2019). Repurposing approved drugs as inhibitors of K<sub>v</sub>7.1 and Na<sub>v</sub>1.8 to treat Pitt Hopkins syndrome. *Pharmaceutical Research*, **36**(9), 137. <https://doi.org/10.1007/s11095-019-2671-y>
- Emig, D., Ivliev, A., Pustovalova, O., Lancashire, L., Bureeva, S., Nikolsky, Y. & Bessarabova, M. (2013). Drug target prediction and repositioning using an integrated network-based approach. *PLoS ONE*, **8**(4), e60618. <https://doi.org/10.1371/journal.pone.0060618>
- Fanelli, U., Pappalardo, M., Chinè, V., Gismondi, P., Neglia, C., Argentiero, A., Calderaro, A., Prati, A. & Esposito, S. (2020). Role of artificial intelligence in fighting antimicrobial resistance in pediatrics. *Antibiotics*, **9**(11), 767. <https://doi.org/10.3390/antibiotics9110767>
- Feng, Z., Chen, M., Xue, Y., Liang, T., Chen, H., Zhou, Y., Nolin, T. D., Smith, R. B. & Xie, X. Q. (2020). MCCS: a novel recognition pattern-based method for fast track discovery of anti-SARS-CoV-2 drugs. *Briefings in Bioinformatics*, bbaa260. Advance online publication. <https://doi.org/10.1093/bib/bbaa260>
- Ferrero, E., Dunham, I. & Sanseau, P. (2017). *In silico* prediction of novel therapeutic targets using gene-disease association data. *Journal of Translational Medicine*, **15**(1), 182. <https://doi.org/10.1186/s12967-017-1285-6>
- Gardiner, E. J., Willett, P. & Artymiuk, P. J. (2001). Protein docking using a genetic algorithm. *Proteins*, **44**(1), 44–56. <https://doi.org/10.1002/prot.1070>
- Gasteiger, J. (2020). Chemistry in times of artificial intelligence. *ChemPhysChem*, **21**(20), 2233–2242. <https://doi.org/10.1002/cphc.202000518>
- Gautam, P., Jaiswal, A., Aittokallio, T., Al-Ali, H. & Wennerberg, K. (2019). Phenotypic screening combined with machine learning for efficient identification of breast cancer-selective therapeutic targets. *Cell Chemical Biology*, **26**(7), 970–979. e4. <https://doi.org/10.1016/j.chembiol.2019.03.011>
- Gottlieb, A., Stein, G. Y., Rupp, E. & Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Molecular Systems Biology*, **7**, 496. <https://doi.org/10.1038/msb.2011.26>
- Gupta, G. K., Collier, A. L., Lee, D., Hoefer, R. A., Zheleva, V., Siewertsz van Reesema, L. L., Tang-Tan, A. M., Guye, M. L., Chang, D. Z., Winston, J. S., Samli, B., Jansen, R. J., Petricoin, E. F., Goetz, M. P., Bear, H. D. & Tang, A. H. (2020). Perspectives on triple-negative breast cancer: current treatment strategies, unmet needs, and potential targets for future therapies. *Cancers*, **12**(9), 2392. <https://doi.org/10.3390/cancers12092392>
- Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Huang, S., Brooks, M., Lee, M. J. & Asadi, H. (2019). Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *AJR. American Journal of Roentgenology*, **212**(1), 38–43. <https://doi.org/10.2214/AJR.18.20224>
- Hasan Mahmud, S. M., Chen, W., Jahan, H., Dai, B., Din, S. U. & Dzisoo, A. M. (2020). DeepACTION: A deep learning-based method for predicting novel drug-target interactions. *Analytical Biochemistry*, **610**, 113978. <https://doi.org/10.1016/j.ab.2020.113978>
- Hearst, M., Dumais, S., Osuna, E., Platt, J. & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, **13**(4), 18–28. <https://doi.org/10.1109/5254.708428>
- Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., Green, A., Khankhanian, P. & Baranzini, S. E. (2017). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*, **6**, e26726. <https://doi.org/10.7554/eLife.26726>
- Hingorani, A. D., Kuan, V., Finan, C., Kruger, F. A., Gaulton, A., Chopade, S., Sofat, R., MacAllister, R. J., Overington, J. P., Hemingway, H., Denaxas, S., Prieto, D. & Casas, J. P. (2019). Improving the odds of drug development success through human genomics: modelling study. *Scientific Reports*, **9**(1), 18911. <https://doi.org/10.1038/s41598-019-54849-w>
- Jolliffe, I. T. & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, **374**(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Joshi, S. N., Murphy, E. A., Olaniyi, P. & Bryant, R. J. (2021). The multiple effects of aspirin in prostate cancer patients. *Cancer Treatment and Research Communications*, **26**, 100267. <https://doi.org/10.1016/j.ctarc.2020.100267>
- Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K. & Zhavoronkov, A. (2017). The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget*, **8**(7), 10883–10890. <https://doi.org/10.18632/oncotarget.14073>
- Kapsiani, S. & Howlin, B. J. (2021). Random forest classification for predicting lifespan-extending chemical compounds. *Scientific Reports*, **11**(1), 13812. <https://doi.org/10.1038/s41598-021-93070-6>
- Kato, S., Moulder, S. L., Ueno, N. T., Wheler, J. J., Meric-Bernstam, F., Kurzrock, R. & Janku, F. (2015). Challenges and perspective of drug repurposing strategies in early phase clinical trials. *Oncoscience*, **2**(6), 576–580. <https://doi.org/10.18632/oncoscience.173>
- Kaul, V., Enslin, S. & Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointestinal Endoscopy*, **92**(4), 807–812. <https://doi.org/10.1016/j.gie.2020.06.040>
- Ke, Y. Y., Peng, T. T., Yeh, T. K., Huang, W. Z., Chang, S. E., Wu, S. H., Hung, H. C., Hsu, T. A., Lee, S. J., Song, J. S., Lin, W. H., Chiang, T. J., Lin, J. H., Sytwu, H. K. & Chen, C. T. (2020). Artificial intelligence approach fighting COVID-19 with repurposing drugs. *Biomedical*

- Journal*, **43**(4), 355–362. <https://doi.org/10.1016/j.bj.2020.05.001>
- Keshavarzi Arshadi, A., Salem, M., Collins, J., Yuan, J. S. & Chakrabarti, D. (2020). Deepmalaria: artificial intelligence driven discovery of potent antiplasmodials. *Frontiers in Pharmacology*, **10**, 1526. <https://doi.org/10.3389/fphar.2019.01526>
- Kim, E., Choi, A. S. & Nam, H. (2019). Drug repositioning of herbal compounds via a machine-learning approach. *BMC Bioinformatics*, **20**(Suppl 10), 247. <https://doi.org/10.1186/s12859-019-2811-8>
- Kim, H., Kim, E., Lee, I., Bae, B., Park, M. & Nam, H. (2020). Artificial intelligence in drug discovery: a comprehensive review of data-driven and machine learning approaches. *Biotechnology and Bioprocess Engineering*, **25**(6), 895–930. <https://doi.org/10.1007/s12257-020-0049-y>
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, **23**(1), 89–109. [https://doi.org/10.1016/s0933-3657\(01\)00077-x](https://doi.org/10.1016/s0933-3657(01)00077-x)
- Koromina, M., Pandi, M. T. & Patrinos, G. P. (2019). Rethinking drug repositioning and development with artificial intelligence, machine learning, and omics. *Omics: A Journal of Integrative Biology*, **23**(11), 539–548. <https://doi.org/10.1089/omi.2019.0151>
- Kumar, A., Loharch, S., Kumar, S., Ringe, R. P. & Parkesh, R. (2021). Exploiting cheminformatic and machine learning to navigate the available chemical space of potential small molecule inhibitors of SARS-CoV-2. *Computational and Structural Biotechnology Journal*, **19**, 424–438. <https://doi.org/10.1016/j.csbj.2020.12.028>
- Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S. & Golub, T. R. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**(5795), 1929–1935. <https://doi.org/10.1126/science.1132939>
- Le, D. H. & Nguyen-Ngoc, D. (2018). Drug repositioning by integrating known disease-gene and drug-target associations in a semi-supervised learning model. *Acta Biotheoretica*, **66**(4), 315–331. <https://doi.org/10.1007/s10441-018-9325-z>
- Levey, A. I., Qiu, D., Zhao, L., Hu, W. T., Duong, D. M., Higginbotham, L., Dammer, E. B., Seyfried, N. T., Wingo, T. S., Hales, C. M., Gámez Tansey, M., Goldstein, D., Abrol, A., Calhoun, V. D., Goldstein, F. C., Hajjar, I., Fagan, A. M., Galasko, D., Edland, S. D., Hanfelt, J., Lah, J. J. & Weinshenker, D. (2021). A phase II study repurposing atomoxetine for neuroprotection in mild cognitive impairment. *Brain*, en prensa, <https://doi.org/10.1093/brain/awab452>
- Li, B., Dai, C., Wang, L., Deng, H., Li, Y., Guan, Z. & Ni, H. (2020). A novel drug repurposing approach for non-small cell lung cancer using deep learning. *PLoS ONE*, **15**(6), e0233112. <https://doi.org/10.1371/journal.pone.0233112>
- Li, J. & Lu, Z. (2012). A new method for computational drug repositioning using drug pairwise similarity. *IEEE International Conference on Bioinformatics and Biomedicine*, **2012**, 1–4. <https://doi.org/10.1109/BIBM.2012.6392722>
- Linderman, G. C. & Steinerberger, S. (2019). Clustering with t-SNE, provably. *SIAM Journal on Mathematics of Data Science*, **1**(2), 313–332. <https://doi.org/10.1137/18m1216134>
- Liu, R., Wei, L. & Zhang, P. (2021). A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data. *Nature Machine Intelligence*, **3**(1), 68–75. <https://doi.org/10.1038/s42256-020-00276-w>
- Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H. & Yao, X. (2018). Molecular dynamics simulations and novel drug discovery. *Expert Opinion on Drug Discovery*, **13**(1), 23–37. <https://doi.org/10.1080/17460441.2018.1403419>
- Liu, Z., Guo, F., Gu, J., Wang, Y., Li, Y., Wang, D., Lu, L., Li, D. & He, F. (2015). Similarity-based prediction for anatomical therapeutic chemical classification of drugs by integrating multiple data sources. *Bioinformatics*, **31**(11), 1788–1795. <https://doi.org/10.1093/bioinformatics/btv055>
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. (2013). The hallmarks of aging. *Cell*, **153**(6), 1194–1217. <https://doi.org/10.1016/j.cell.2013.05.039>
- Loucera, C., Esteban-Medina, M., Rian, K., Falco, M. M., Dopazo, J. & Peña-Chilet, M. (2020). Drug repurposing for COVID-19 using machine learning and mechanistic models of signal transduction circuits related to SARS-CoV-2 infection. *Signal Transduction and Targeted Therapy*, **5**(1), 290. <https://doi.org/10.1038/s41392-020-00417-y>
- Luo, H., Wang, J., Li, M., Luo, J., Peng, X., Wu, F. X. & Pan, Y. (2016). Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics*, **32**(17), 2664–2671. <https://doi.org/10.1093/bioinformatics/btw228>
- Lv, Q., Wang, D., Yang, Z., Yang, J., Zhang, R., Yang, X., Wang, M. & Wang, Y. (2018). Repurposing antitubercular agent isoniazid for treatment of prostate cancer. *Biomaterials Science*, **7**(1), 296–306. <https://doi.org/10.1039/c8bm01189c>
- Malley, B., Ramazzotti, D. & Wu, J. T., Data pre-processing. 2016 Sep 10. In: Secondary analysis of electronic health records [Internet]. Cham (CH): Springer; 2016. Chapter 12. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK543629/>. <https://doi.org/10.1007/978-3-319-43742-2>
- Mallikarjun, V. & Swift, J. (2016). Therapeutic manipulation of ageing: repurposing old dogs and discovering new tricks. *EBioMedicine*, **14**, 24–31. <https://doi.org/10.1016/j.ebiom.2016.11.020>

- Mamoshina, P., Volosnikova, M., Ozerov, I. V., Putin, E., Skibina, E., Cortese, F. & Zhavoronkov, A. (2018). Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. *Frontiers in Genetics*, **9**, 242. <https://doi.org/10.3389/fgene.2018.00242>
- March-Vila, E., Pinzi, L., Sturm, N., Tinivella, A., Engkvist, O., Chen, H. & Rastelli, G. (2017). On the integration of *in silico* drug design methods for drug repurposing. *Frontiers in Pharmacology*, **8**, 298. <https://doi.org/10.3389/fphar.2017.00298>
- Mattiuzzi, C. & Lippi, G. (2019). Current cancer epidemiology. *Journal of Epidemiology and Global Health*, **9**(4), 217–222. <https://doi.org/10.2991/jegh.k.191008.001>
- Meng, X. Y., Zhang, H. X., Mezei, M. & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, **7**(2), 146–157. <https://doi.org/10.2174/157340911795677602>
- Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nature Reviews Drug Discovery*, **16**(8), 531–543. <https://doi.org/10.1038/nrd.2017.111>
- Mohamed, K., Yazdanpanah, N., Saghadzadeh, A. & Rezaei, N. (2021). Computational drug discovery and repurposing for the treatment of COVID-19: A systematic review. *Bioorganic Chemistry*, **106**, 104490. <https://doi.org/10.1016/j.bioorg.2020.104490>
- Mohanty, S., Harun Ai Rashid, M., Mridul, M., Mohanty, C. & Swayamsiddha, S. (2020). Application of artificial intelligence in COVID-19 drug repurposing. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, **14**(5), 1027–1031. <https://doi.org/10.1016/j.dsx.2020.06.068>
- Mohapatra, S., Nath, P., Chatterjee, M., Das, N., Kalita, D., Roy, P. & Satapathi, S. (2020). Repurposing therapeutics for COVID-19: Rapid prediction of commercially available drugs through machine learning and docking. *PLoS ONE*, **15**(11), e0241543. <https://doi.org/10.1371/journal.pone.0241543>
- Moridi, M., Ghadirinia, M., Sharifi-Zarchi, A. & Zare-Mirakabad, F. (2019). The assessment of efficient representation of drug features using deep learning for drug repositioning. *BMC Bioinformatics*, **20**(1), 577. <https://doi.org/10.1186/s12859-019-3165-y>
- Morrow, J. K., Tian, L. & Zhang, S. (2010). Molecular networks in drug discovery. *Critical Reviews<sup>TM</sup> in Biomedical Engineering*, **38**(2), 143–156. <https://doi.org/10.1615/critrevbiomedeng.v38.i2.30>
- Mugavin M. E. (2008). Multidimensional scaling: a brief overview. *Nursing Research*, **57**(1), 64–68. <https://doi.org/10.1097/01.NNR.0000280659.88760.7c>
- Muthukrishnan, N., Maleki, F., Ovens, K., Reinhold, C., Forghani, B. & Forghani, R. (2020). Brief history of artificial intelligence. *Neuroimaging Clinics of North America*, **30**(4), 393–399. <https://doi.org/10.1016/j.nic.2020.07.004>
- Napolitano, F., Zhao, Y., Moreira, V. M., Tagliaferri, R., Kere, J., D'Amato, M. & Greco, D. (2013). Drug repositioning: a machine-learning approach through data integration. *Journal of Cheminformatics*, **5**(1), 30. <https://doi.org/10.1186/1758-2946-5-30>
- Ooi, C. P. & Lye, M. S. (2018). Bromocriptine for type 2 diabetes mellitus. *The Cochrane Database of Systematic Reviews*, **2018**(7), CD010319. <https://doi.org/10.1002/14651858.CD010319.pub2>
- Oprea, T. I., Bauman, J. E., Bologa, C. G., Buranda, T., Chigaev, A., Edwards, B. S., Jarvik, J. W., Gresham, H. D., Haynes, M. K., Hjelle, B., Hromas, R., Hudson, L., Mackenzie, D. A., Muller, C. Y., Reed, J. C., Simons, P. C., Smagley, Y., Strouse, J., Surviladze, Z., Thompson, T., Ursu, O., Waller, A., Wandinger-Ness, A., Winter, S. S., Wu, Y., Young, S. M., Larson, R. S., Willman, C. & Sklar, L. A. (2011). Drug Repurposing from an Academic Perspective. *Drug discovery today. Therapeutic Strategies*, **8**(3-4), 61–69. <https://doi.org/10.1016/j.ddstr.2011.10.002>
- Öztürk, H., Özgür, A., Schwaller, P., Laino, T. & Ozkirimli, E. (2020). Exploring chemical space using natural language processing methodologies for drug discovery. *Drug Discovery Today*, **25**(4), 689–705. <https://doi.org/10.1016/j.drudis.2020.01.020>
- Parvathaneni, V., Kulkarni, N. S., Muth, A. & Gupta, V. (2019). Drug repurposing: a promising tool to accelerate the drug discovery process. *Drug Discovery Today*, **24**(10), 2076–2085. <https://doi.org/10.1016/j.drudis.2019.06.014>
- Patel, L., Shukla, T., Huang, X., Ussery, D. W. & Wang, S. (2020). Machine learning methods in drug discovery. *Molecules*, **25**(22), 5277. <https://doi.org/10.3390/molecules25225277>
- Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K. & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discovery Today*, **26**(1), 80–93. <https://doi.org/10.1016/j.drudis.2020.10.010>
- Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Williams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D. & Pirmohamed, M. (2019). Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, **18**(1), 41–58. <https://doi.org/10.1038/nrd.2018.168>
- Ramsay, R. R., Popovic-Nikolic, M. R., Nikolic, K., Uliassi, E. & Bolognesi, M. L. (2018). A perspective on multi-target drug discovery and design for complex diseases. *Clinical and Translational Medicine*, **7**(1), 3. <https://doi.org/10.1186/s40169-017-0181-2>
- Rattan, S. I. (2006). Theories of biological aging: genes, proteins, and free radicals. *Free Radical Research*, **40**(12), 1230–1238. <https://doi.org/10.1080/10715760600911303>
- Rudrapal, M., J. Khairnar, S. & G. Jadhav, A. (2020). Drug repurposing (DR): an emerging approach in drug discovery.

- Drug Repurposing - Hypothesis, Molecular Aspects and Therapeutic Applications*, **1(1)**. <https://doi.org/10.5772/intechopen.93193>
- Saeyns, Y., Inza, I. & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, **23(19)**, 2507–2517. <https://doi.org/10.1093/bioinformatics/btm344>
- Saldívar-González, F. & Medina-Franco, J. (2021). Nuevos usos a medicamentos: el reposicionamiento de fármacos y su importancia en crisis sanitarias. *Entorno UAPAP*, **14**, 1-45.
- Saldívar-González, F., Prieto-Martínez, F. D. & Medina-Franco, J. L. (2017). Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educación Química*, **28(1)**, 51-58. <https://doi.org/10.1016/j.eq.2016.06.002>
- Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Networks*, **61**, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Sebastián-Pérez, V., Martínez, M. J., Gil, C., Campillo, N. E., Martínez, A. & Ponzoni, I. (2019). QSAR modelling to identify LRRK2 inhibitors for Parkinson's disease. *Journal of Integrative Bioinformatics*, **16(1)**, 20180063. <https://doi.org/10.1515/jib-2018-0063>
- Sidey-Gibbons, J. A. M. & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*, **19(1)**, 64. <https://doi.org/10.1186/s12874-019-0681-4>
- Sproule, B. A., Naranjo, C. A. & Türksen, I. B. (2002). Fuzzy pharmacology: theory and applications. *Trends in Pharmacological Sciences*, **23(9)**, 412–417. [https://doi.org/10.1016/s0165-6147\(02\)02055-2](https://doi.org/10.1016/s0165-6147(02)02055-2)
- Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R. & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. *Cell*, **180(4)**, 688–702.e13. <https://doi.org/10.1016/j.cell.2020.01.021>
- Suay-García, B., Falcó, A., Bueso-Bordils, J. I., Anton-Fos, G. M., Pérez-Gracia, M. T. & Alemán-López, P. A. (2020). Tree-based QSAR model for drug repurposing in the discovery of new antibacterial compounds against *Escherichia coli*. *Pharmaceutics*, **13(12)**, 431. <https://doi.org/10.3390/ph13120431>
- Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., Lahr, D. L., Hirschman, J. E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I. C., Lam, D., Liberzon, A., Toder, C., Bagul, M., Orzechowski, M., Enache, O. M., Piccioni, F., Johnson, S. A., Lyons, N. J., Berger, A. H., Shamji, A. F., Brooks, A. N., Vrcic, A., Flynn, C., Rosains, J., Takeda, D. Y., Hu, R., Davison, D., Lamb, J., Ardlie, K., Hogstrom, L., Greenside, P., Gray, N. S., Clemons, P. A., Silver, S., Wu, X., Zhao, W. N., Read-Button, W., Wu, X., Haggarty, S. J., Ronco, L. V., Boehm, J. S., Schreiber, S. L., Doench, J. G., Bittker, J. A., Root, D. E., Wong, B. & Golub, T. R. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, **171(6)**, 1437–1452.e17. <https://doi.org/10.1016/j.cell.2017.10.049>
- Tanoli, Z., Vähä-Koskela, M. & Aittokallio, T. (2021). Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opinion on Drug Discovery*, **16(9)**, 977–989. <https://doi.org/10.1080/17460441.2021.1883585>
- Thompson, P. L. (2019). Colchicine in cardiovascular disease: repurposing an ancient gout drug. *Clinical Therapeutics*, **41(1)**, 8–10. <https://doi.org/10.1016/j.clinthera.2018.11.014>
- Udaondo, Z. & Matilla, M. A. (2020). Mining for novel antibiotics in the age of antimicrobial resistance. *Microbial Biotechnology*, **13(6)**, 1702–1704. <https://doi.org/10.1111/1751-7915.13662>
- Urbina, F., Puhl, A. C. & Ekins, S. (2021). Recent advances in drug repurposing using machine learning. *Current Opinion in Chemical Biology*, **65**, 74–84. <https://doi.org/10.1016/j.cbpa.2021.06.001>
- Wang, X., Zhao, Y. & Pourpanah, F. Recent advances in deep learning. *International Journal of Machine Learning and Cybernetics*, **11**, 747–750 (2020). <https://doi.org/10.1007/s13042-020-01096-5>
- Wei, X., Zhang, Y., Huang, Y. & Fang, Y. (2019). Predicting drug–disease associations by network embedding and biomedical data integration. *Data Technologies and Applications*, **53(2)**, 217–229. <https://doi.org/10.1108/dta-01-2019-0004>
- Wiffen, P. J., Derry, S., Bell, R. F., Rice, A. S. C., Tölle, T. R., Phillips, T. & Moore, R. A. (2017). Gabapentin for chronic neuropathic pain in adults. *Cochrane Database of Systematic Reviews*, **6(6)**, CD007938. <https://doi.org/10.1002/14651858.CD007938.pub4>
- Willett, P., Barnard, J. & Downs, G. M. (1998). Chemical similarity searching. *Journal of Chemical Information and Computer Sciences*, **38(6)**, 983–996. <https://doi.org/10.1021/ci9800211>
- Winkler, D. A. (2021). Use of artificial intelligence and machine learning for discovery of drugs for neglected tropical diseases. *Frontiers in Chemistry*, **9**, 614073. <https://doi.org/10.3389/fchem.2021.614073>
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C. & Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, **46(D1)**, D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
- Xuan, P., Zhao, L., Zhang, T., Ye, Y. & Zhang, Y. (2019). Inferring drug-related diseases based on convolutional neural network

- and gated recurrent unit. *Molecules*, **24**(15), 2712. <https://doi.org/10.3390/molecules24152712>
- Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R. & Cheng, F. (2019). deepDR: a network-based deep learning approach to *in silico* drug repositioning. *Bioinformatics*, **35**(24), 5191–5198. <https://doi.org/10.1093/bioinformatics/btz418>
- Zhang, W., Yue, X., Huang, F., Liu, R., Chen, Y. & Ruan, C. (2018). Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network. *Methods*, **145**, 51–59. <https://doi.org/10.1016/j.ymeth.2018.06.001>
- Zhao, K. & So, H. C. (2019). Drug repositioning for schizophrenia and depression/anxiety disorders: a machine learning approach leveraging expression data. *IEEE Journal of Biomedical and Health Informatics*, **23**(3), 1304–1315. <https://doi.org/10.1109/JBHI.2018.2856535>
- Zhou, Y., Wang, F., Tang, J., Nussinov, R. & Cheng, F. (2020). Artificial intelligence in COVID-19 drug repurposing. *The Lancet. Digital Health*, **2**(12), e667–e676. [https://doi.org/10.1016/S2589-7500\(20\)30192-8](https://doi.org/10.1016/S2589-7500(20)30192-8)
- Zorn, K. M., Sun, S., McConnon, C. L., Ma, K., Chen, E. K., Foil, D. H., Lane, T. R., Liu, L. J., El-Sakkary, N., Skinner, D. E., Ekins, S. & Caffrey, C. R. (2021). A machine learning strategy for drug discovery identifies anti-schistosomal small molecules. *ACS Infectious Diseases*, **7**(2), 406–420. <https://doi.org/10.1021/acsinfecdis.0c00754>