

Aplicación de la Inteligencia Artificial en la Industria Farmacéutica

ALEJANDRO JIMÉNEZ CÁCERES

FACULTAD DE FARMACIA (UNIVERSIDAD DE SEVILLA)

TFM 2022 - 2023





Aplicación de la Inteligencia Artificial en la Industria Farmacéutica

Trabajo Fin de Máster

2022-2023

Máster Universitario en Especialización
Profesional en Farmacia: Especialidad
Industria Farmacéutica (IF)

Autor: Alejandro Jiménez Cáceres

Tutora: M.^a Matilde Durán Lobato

Revisión bibliográfica

Resumen

Este trabajo de fin de máster se centra en estudiar la aplicación de la inteligencia artificial (IA) en la industria farmacéutica. La IA ha demostrado ser una herramienta prometedora en este campo, ya que permite el análisis y procesamiento eficiente de grandes cantidades de datos, acelerando así el proceso de descubrimiento de nuevos medicamentos y mejorando la eficiencia de los procesos de producción existentes. El trabajo comienza presentando los conceptos fundamentales de la IA y sus subconjuntos, como el aprendizaje automático, las redes neuronales artificiales y el aprendizaje profundo. Se explican las características y aplicaciones de cada uno de estos enfoques, destacando cómo se utilizan en la industria farmacéutica.

Se profundiza en el descubrimiento de fármacos asistido por IA. Se explora cómo se utilizan sus subconjuntos para analizar grandes bases de datos de compuestos químicos y seleccionar aquellos con mayor potencial para convertirse en medicamentos. Además, se examina el uso de redes neuronales artificiales en la predicción de la actividad y toxicidad de los compuestos, lo que permite reducir los costos y el tiempo necesarios para llevar un fármaco al mercado.

El trabajo también aborda la optimización de los procesos de producción farmacéutica mediante el uso de IA. Se destacan las ventajas de emplear algoritmos de aprendizaje automático para el control y monitoreo de los procesos, la detección de anomalías y la mejora de la calidad.

- **Palabras clave:** Inteligencia artificial, aprendizaje automático, aprendizaje profundo, redes neuronales artificiales, industria farmacéutica, base de datos.

Abstract

This work focuses on studying the application of artificial intelligence (AI) in the pharmaceutical industry. AI has proven to be a promising tool in this field, as it enables the efficient analysis and processing of large amounts of data, thus accelerating the process of discovering new drugs and improving the efficiency of existing production processes. The paper begins by presenting the fundamental concepts of AI and its subsets, such as machine learning, artificial neural networks and deep learning. The characteristics and applications of each of these approaches are explained, highlighting how they are used in the pharmaceutical industry.

This is followed by an in-depth look at AI-assisted drug discovery. It explores how machine learning and deep learning techniques are used to analyze large databases of chemical compounds and select those with the greatest potential to become drugs. In addition, it examines the use of artificial neural networks in predicting the activity and toxicity of compounds, thus reducing the costs and time required to bring a drug to market.

The paper also addresses the optimization of pharmaceutical production processes using AI. The advantages of employing machine learning algorithms for process control and monitoring, anomaly detection and quality improvement are highlighted.

- **Keywords:** Artificial intelligence, Machine Learning, Deep Learning, Artificial Neural Networks, Pharmaceutical Industry, Database.

ÍNDICE.

LISTADO DE ABREVIATURAS	1
ÍNDICE DE FIGURAS	2
1. Antecedentes	3
2. Objetivos	7
3. Metodología	8
4. Resultados y discusión	9
4.1. Redes neuronales artificiales	9
4.1.1. Aplicación de ANNs en procesos industriales previos:.....	10
4.1.2. Aplicación de ANNs en procesos industriales posteriores:.....	12
4.2. ‘Machine Learning’	15
4.2.1. ML para la definición de objetivos:	17
4.2.2. ML para el estudio de imágenes.	19
4.2.3. ML para el HTVS.	20
4.2.4. ML para el diseño de fármacos basado en estructuras.	21
4.2.5. ML en la interacción fármaco-diana.	24
4.2.6. ML y otras AI en química cuántica.	24
4.2.7. ML y otras AI para propiedades fisicoquímicas.	26
4.3. ‘Deep Learning’ :	27
4.3.1. Predicción del plegamiento de proteínas a partir de la secuencia.	29
4.3.2. LBVS:	30
4.4. Modelos AI para el diseño ‘de novo’	31
4.5. Modelos AI para la predicción de interacciones proteína-proteína :	32
5. Conclusión :.....	34
6. Bibliografía :.....	35

LISTADO DE ABREVIATURAS.

Abreviaturas empleadas	Significado	Traducción
AI	Artificial Intelligence	Inteligencia Artificial
ANN	Artificial Neural Network	Red Neuronal Artificial
API	Active Pharmaceutical Ingredient	Principio Activo Farmacéutico
CNN	Convolutional Neural Network	Red Neuronal Convolutacional
DBN	Deep Belief Network	Red de Creencias Profundas
DFT	Density Functional Theory	Teoría del Funcional de la Densidad
DL	Deep Learning	Aprendizaje Profundo
DT	Decision Tree	Arbol de Decisión
GA	Genetic Algorithm	Algoritmo Genético
GPR	Gaussian Process Resegion	Regresión del Proceso Gaussiano
HTVS	High Throughput Virtual Slit	Cribado Virtual de Alto Rendimiento
ML	Machine Learning	Aprendizaje Automático
MLP	Multilayer Perceptron	Perceptrón Multicapa
MPC	Model Predictive Control	Modelo de Control Predictivo
NN	Neuronal Network	Red Neuronal
PAT	Process Analytical Technology	Tecnología Analítica de Procesos
PLS	Partial Least Squares Regression	Regresión de Minimos Cuadrados Parciales
PSO	Particle Swarm Optimization	Optimización por Enjambre de Partículas
QSAR	Quantitative Structure Activity Relationships	Relación Cuantitativa Estructura-Actividad
RF	Random Forest	Arbol Aleatorio
RNN	Recurrent Neural Network	Red Neuronal Recurrente
SVM	Support-Vector Machines	Máquinas de Vectores de Soporte

ÍNDICE DE FIGURAS

Figura 1. Clasificación de la AI, sus subconjuntos, y herramientas más empleadas	3
Figura 2. Arquitectura básica de una ANN	9
Figura 3. Comparativa de programación general frente a ML	16
Figura 4. Ejemplo de estructura de un DL	28
Figura 5. Ejemplo de uso de DL: Representación de un tóxico por características jerárquicamente relacionadas	29

1. Antecedentes

La Inteligencia Artificial (AI, Artificial Intelligence) está revolucionando muchos sectores al realizar operaciones que requieren inteligencia humana habitualmente. La AI contribuye a flujos de trabajo científicos y de ingeniería complejos simulando, complementando o aumentando la inteligencia de forma eficiente y precisa. El objetivo final sería programar la inteligencia en las máquinas aprendiendo de las experiencias y adaptándose a cambios del entorno para simular los procesos humanos de toma de decisiones y razonamiento (Muthukrishnan et al., 2020). John McCarthy la definió por primera vez como "la ciencia y la ingeniería de máquinas inteligentes" (Hamet & Tremblay, 2017). El interés por la AI no se despertó realmente hasta mediados de la

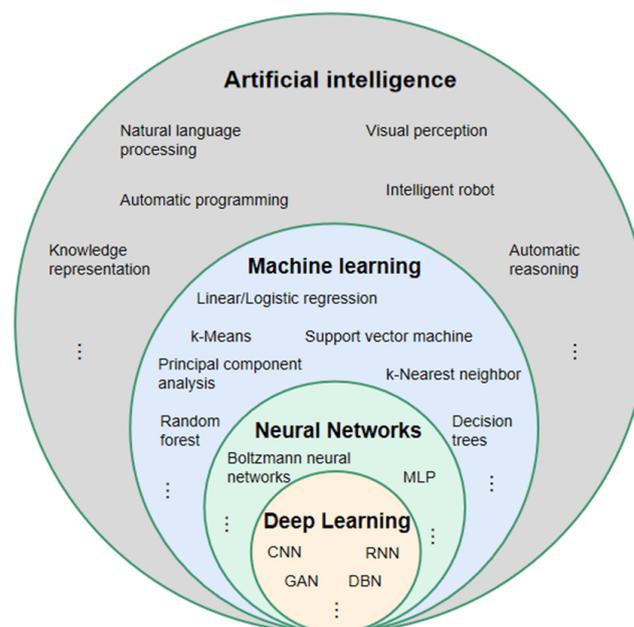


Figura 1. Clasificación de la AI, sus subconjuntos, y herramientas más empleadas (S. Li et al., 2021).

década de 1990 al aumentar la potencia de cálculo y poder apoyar el desarrollo de redes neuronales. La revolución de los microordenadores y la ley de Moore describieron avances que experimentaron los ordenadores de la época y que permitieron sustituir a las máquinas Lisp tradicionales (Muthukrishnan et al., 2020). La ley de Moore es un modelo tecno-económico que ha permitido a la industria tecnológica duplicar el rendimiento y la funcionalidad de la electrónica digital aproximadamente cada dos años con un coste, potencia y superficie fijos gracias a los avances en la litografía del silicio que han permitido el crecimiento exponencial de la miniaturización y del rendimiento (Shalf, 2020). Las máquinas Lisp tradicionales limitaban el desarrollo de la AI, su lenguaje Lisp no ha sido el más dominante en la industria durante la mayor parte de su historia. En lugar

de eso, lenguaje como C, Java, Python y otros han tenido mayor adopción y popularidad. Hay varios factores que podrían explicar el por qué dicho lenguaje de programación no fue dominante, entre ellas destacan su limitada curva de aprendizaje, la falta de herramientas y bibliotecas, sus requisitos de rendimiento y sus intereses comerciales (Takeuchi, 2002).

La representación del conocimiento, el procesamiento del lenguaje natural, la concepción visual, el razonamiento automático, el aprendizaje automático, los robots inteligentes, la programación automática y otros campos se han enriquecido con el desarrollo de la AI. El término AI se ha convertido en un término paraguas (Zhou et al., 2020) que engloba a otros subconjuntos tecnológicos (Figura 1). El concepto de aprendizaje automático (ML, 'Machine Learning'), un subcampo de la AI, es una tecnología utilizada para hacer realidad la AI. La idea principal es utilizar algoritmos para analizar y aprender de los datos, luego hacer predicciones y tomar decisiones sobre los acontecimientos del mundo tangible (Erickson et al., 2017). Esto difiere de los programas de software convencionales que están codificados para resolver tareas específicas (Foster et al., 2014). Las categorías de algoritmos incluyen algoritmos de aprendizaje supervisado, como los métodos de clasificación y regresión, algoritmos de aprendizaje no supervisado, como el análisis de conglomerados (Jafari et al., 2020), y algoritmos de aprendizaje semisupervisado, como las regresiones para datos de neuroimagen de alta dimensionalidad (Bzdok et al., 2018). Mediante un procedimiento denominado ingeniería de características, a menudo un experto selecciona o genera un conjunto de características informativas para construir modelos predictivos. La accesibilidad de una gran cantidad de información y de potencia computacional ha dado lugar a una oleada de algoritmos de éxito (Muthukrishnan et al., 2020). El aprendizaje profundo (DL, 'Deep Learning') es una herramienta algorítmica para el ML (Anwar et al., 2018). Se deriva de una red neuronal artificial (ANN, Artificial Neural Network), que simula el modo en que el cerebro humano procesa la información (Lecun et al., 2015), y utiliza el método de descenso de gradiente y el algoritmo de retropropagación para corregir automáticamente sus propios parámetros, haciendo que la red se ajuste mejor a los datos (Kriegeskorte & Golan, 2019; Zipser & Andersen, 1988). Comparada con la ANN tradicional, el DL tiene una capacidad de ajuste más potente gracias a un mayor número de niveles neuronales (Manisha et al., 2020). Según los diferentes escenarios, el DL incluye una variedad de modelos de redes neuronales, como las redes neuronales convolucionales (CNN, Convolutional Neural Network) con

potentes capacidades de procesamiento de imágenes (Soffer et al., 2019), las redes neuronales recurrentes (RNN, Recurrent Neural Network), que procesan principalmente muestras de series temporales (Deng et al., 2019; Y. Li et al., 2021), y las redes de creencias profundas (DBN, Deep Belief Network), que pueden expresar en profundidad los datos de entrenamiento (Harbin et al., 2015). En la última década, los métodos basados en CNN han ganado popularidad en el ámbito del análisis de imágenes médicas (Anwar et al., 2018; Ko et al., 2020; Skrede et al., 2020). En los estudios de imágenes de CNP utilizando modelos DL, se adoptaron CNN en casi todos los estudios. Hay tres conceptos que engloban las CNN: campos receptivos locales, reparto de pesos y submuestreos espaciales/temporales (Lecun et al., 2015). Una CNN tiene tres capas: una convolucional, una de agrupamiento y una totalmente conectada. Tras introducir una imagen, ésta se convierte en una matriz monocanal en el caso de ser una imagen sin color (presentada en escala de grises), o en una matriz de tres canales en el caso a color. El proceso consiste en un escaneo de la imagen llevada a cabo por núcleos de convolución (pertenecientes a la capa convolucional) obteniendo una matriz denominada característica convolucionada. Tras esto, se lleva a cabo un proceso denominado ‘pooling’, que consiste en una reducción de la dimensión espacial. Así, los procesos de convolución y ‘pooling’ de una CNN se repiten muchas veces, y la “profundidad” de un modelo DL se materializa en el número de convoluciones (Goodfellow et al., 2016)

Para la Pharma 4.0 (como subconjunto de la Industria 4.0), la AI y el ML han surgido como herramientas versátiles para abordar varias tareas emergentes, por ejemplo, el análisis de macrodatos o el desarrollo de gemelos digitales (objeto físico, proceso o sistema reflejado con exactitud). Las agencias reguladoras médicas también se muestran cada vez más abiertas a los enfoques de AI/ML. Como ejemplo, la Agencia Danesa de Medicamentos (DKMA) ha publicado recientemente una lista de cuestiones a tener en cuenta para desarrollar y aplicar modelos basados en ML en áreas reguladas por GxP (Good x Practices). Por otro, la Administración de Alimentos y Medicamentos de EE.UU. (FDA) ha publicado un plan de acción para establecer una “Buena práctica de ML” (U.S. Food and Drug Administration, 2021). Aunque se trata de dispositivos médicos basados en ML, el enfoque (por ejemplo, cómo evaluar la solidez, el sesgo y el rendimiento en el mundo real) podría generalizarse para otras aplicaciones de ML (Nagy et al., 2022). Ante este escenario, el presente trabajo pretende explorar las aplicaciones de la AI en la industria farmacéutica.

2. Objetivos.

La aplicación de la AI en la industria farmacéutica representa una oportunidad para mejorar diversos aspectos, desde el descubrimiento de nuevos fármacos hasta la optimización de procesos de fabricación. Este trabajo de investigación se enfocará en analizar cómo los subconjuntos de la AI están siendo utilizados en la industria farmacéutica. Así, se definen los siguientes objetivos de este trabajo:

- Analizar el estado actual de la industria farmacéutica en relación con la adopción de la AI y sus subconjuntos, como las ANN, el ML y el DL.
- Investigar las principales aplicaciones y casos de uso de la AI en la industria farmacéutica, incluyendo la identificación y desarrollo de fármacos, el descubrimiento de nuevas moléculas, la optimización de procesos de fabricación y control de calidad, la personalización de tratamientos médicos y la mejora de la eficiencia operativa.
- Realizar un análisis comparativo de las diferentes técnicas y algoritmos utilizados en las ANN, el ML y el DL en el contexto de la industria farmacéutica, identificando las fortalezas y limitaciones de cada enfoque y su aplicabilidad en diferentes escenarios.
- Investigar el alcance de dichas herramientas y conocer si en un proceso concreto, la combinación de estas ofrece mejores resultados.

3. Metodología.

Para la elaboración de este trabajo de fin de máster bibliográfico se ha procedido a la búsqueda de artículos a partir de 'SCOPUS'. Esta es una base de datos de citas y resúmenes de artículos de revistas científicas de Elsevier, en activo desde 2004. Esta base de datos permite realizar búsquedas avanzadas con filtros específicos con el fin de obtener artículos precisos en función del objetivo. En este caso, siendo el principal objetivo obtener información acerca de la AI en la Industria Farmacéutica, el abanico de posibilidades fue bastante amplio, debido a que, en esta última década, el empleo de la IA en el campo de la investigación farmacéutica está en auge, y necesitó hacer uso de varios filtros para identificar y seleccionar documentos recientes que pudieran recoger la información más pertinente al respecto.

- Para el apartado de antecedentes se utilizaron los siguientes filtros: Article title: history artificial intelligence; Year: 2020 – 2023, Document type: Review; Subject Area: 'Biochemistry, Genetics and Molecular Biology' y 'Medicine'. De los cuales se obtuvieron tres resultados concretos.
- Para el cuerpo del contenido, se utilizaron los siguientes filtros: Article title: artificial intelligence pharmaceutical (industry OR industries), Year: 2019 – 2023, Document type: ALL; Subject Area: ALL. Obteniendo un total de 8 resultados, priorizando aquellos que tenían mayor índice de impacto.

En esta última búsqueda, los artículos de mayor impacto permitieron identificar casos muy concretos de aplicación de la AI en diferentes campos relacionados con el ámbito farmacéutico.

4. Resultados y discusión

4.1. Redes neuronales artificiales

Las redes neuronales artificiales (ANN), desarrolladas en la década de 1960, no comenzaron a expandirse de manera progresiva hasta principios de la década de 1980, junto con el lanzamiento y el desarrollo de la tecnología informática moderna. Desde entonces, las redes neuronales se han utilizado exitosamente en numerosos campos. El uso de la AI, como las ANNs, se han aplicado en ciencias farmacéuticas para generar y refinar formas de dosificación en un área cada vez más creciente de descubrimiento de conocimientos (Kumar Ananthu et al., 2021).

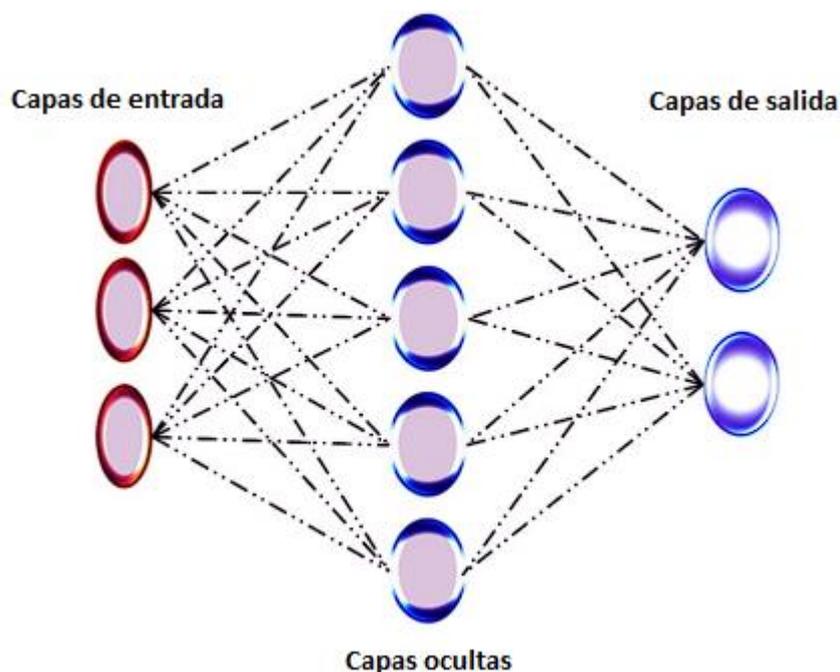


Figura 2. Arquitectura básica de una ANN (Selvaraj et al., 2021).

Las ANN son modelos informáticos favorablemente reconocidos, desarrollados a partir del cerebro del *Homo Sapiens* y sus tendencias de conexión en red (Zador, 2019). El caso más sencillo muestra una red totalmente conectada o *feed-forward* que conforma el diagrama de cálculo con tres capas (de entrada, oculta y de salida) (Alzahab et al., 2021). La unidad única de cómputo por capas, llamada neurona, funciona como una transformación no lineal de los datos de entrada. Esta información se propaga por capas y recibe la salida de la capa anterior (Figura 2). La modelización molecular y el diseño de

fármacos se basan principalmente en las ANN. Resuelve la complejidad asociada a los modelos estadísticos utilizados en el cribado virtual de alto rendimiento (HTVS), la relación cuantitativa estructura-actividad (QSAR) y los estudios farmacocinéticos y farmacodinámicos (Ekins et al., 2007). En cuanto a los valores numéricos que determinan la salida, las ANN tienen un excelente rendimiento en la interpretación de relaciones no lineales y predicen el proceso de éxito en la búsqueda de medicamentos (Ahmed et al., 2020). Un subconjunto de las ANN son las redes neuronales directas, o perceptrón multicapa (MLP), en la que la información pasa a través de la red sin bucles de retorno a las capas anteriores (Nagy et al., 2022).

Las ANN son muy fiables para mejorar la eficacia y el descubrimiento de fármacos basados en dianas. Tienen una enorme capacidad para la investigación compleja y las relaciones no lineales, por lo que estas ANN reciben el nombre alternativo de “cerebro modelo digitalizado” (Henry & Wlodkowic, 2019). Las aplicaciones de las ANN son muy importantes para las ciencias, la tecnología, la ingeniería y la medicina (STEM, Science Technology, Engineering and Medicine). Especialmente en el modelado molecular y las ciencias farmacéuticas, la aplicación de las ANN marca la tendencia al proporcionar una alta fiabilidad de los resultados (Pesapane et al., 2020). En función del proceso industrial farmacéutico la aplicación de las ANN varía.

4.1.1. Aplicación de ANNs en procesos industriales previos:

- **Síntesis:**

La síntesis de compuestos orgánicos es el primer paso en la fabricación farmacéutica industrial para producir el API (Active pharmaceutical ingredient). Las ANN han encontrado aplicaciones en la optimización de parámetros del proceso para mejorar el resultado y la eficacia de las reacciones y explicar las relaciones no lineales entre los parámetros deseados y los datos espectroscópicos (Nagy et al., 2022). Varios estudios se han ocupado de la optimización de síntesis utilizando ANNs. Por ejemplo (Valizadeh et al., 2009) trató de pronosticar y optimizar el rendimiento de la producción de glucosamina a partir de quitina mediante el algoritmo genético (GA), la optimización por enjambre de partículas (PSO) y las ANNs como herramientas de los métodos de AI. Estos dos primeros métodos son modelos heurísticos de AI, es decir, "reglas aproximadas" que no garantizan

encontrar siempre la mejor solución, pero son eficientes y efectivas en muchos casos. Los métodos GA y PSO fueron ligeramente mejores que el modelo ANN. Pero en la validación de los modelos, el modelo MLP fue mejor que los modelos obtenidos por los métodos GA y PSO. De hecho, el modelo MLP presentó mayor ‘no linealidad’ que los demás, lo que conduce a un modelo más generable con menor error en la validación. Un menor error en la fase de prueba implica un modelo más preciso y fiable para otros casos desconocidos, que es la principal característica de las ANN.

La RNN podría perfectamente sustituir a un modelo de planta real (Wong et al., 2018) o a un modelo de espacio de estados (Baranilingesan, 2021) en los algoritmos de control. Esto podría ser beneficioso para predecir la dinámica del proceso dentro de un modelo de control predictivo (MPC) de una síntesis farmacéutica continua, ya que la demanda computacional se reduce significativamente en comparación con los modelos mecanicistas. Además, las ANN pueden beneficiarse enormemente del entorno rico en datos de la fabricación asistida por las tecnologías analíticas de procesos (PAT). Otro estudio basado en fermentación, la glucosa y la concentración de ácido glucónico se determinaron a partir de mediciones espectroscópicas de infrarrojos por transformada de Fourier (FT-IR) (Franco et al., 2006), utilizando una red multicapa *feed-forward* con 15 muestras de calibración. La ANN superó a la regresión clásica por mínimos cuadrados parciales (PLS). También se cuantificaron simultáneamente fenol y clorofenoles empleando un modelo ANN a partir de los espectros UV-Vis (Hasani & Moloudi, 2008). En este caso, se utilizaron las puntuaciones de los componentes principales para comprimir los espectros para el entrenamiento de redes.

- **Cristalización:**

La cristalización es clave para conectar la síntesis del API y los pasos posteriores de la formulación, ya que proporciona un API cristalino sólido, lo que influye en diferentes parámetros como el rendimiento, la pureza, la capacidad de fabricación posterior e incluso la biodisponibilidad del producto final (Nagy et al., 2022). El ML puede estimar resultados de cristalización basándose en datos históricos del proceso. Un ejemplo es el método desarrollado por (Velásco-Mejía et al., 2016), basado en ANN y GA de 54 cristalizaciones industriales por lotes. Utilizaron nueve descriptores y modelizaron la densidad cristalina como resultado, lo que permitió identificar los parámetros más críticos y, tras la optimización realizada, se obtuvo una mejora sustancial del producto. En otro

trabajo, se pudo experimentar con el espacio de diseño de un proceso de co-cristalización a partir de 25 experimentos y cuatro variables de entrada (Shaikh et al., 2021). Las variables de entrada para conformar la ANN fueron el tiempo de molienda, el tamaño del recipiente y el tipo y contenido de excipiente. Con el entrenamiento de la ANN, esta pudo predecir una tasa de crecimiento cristalino más precisa que con la regresión no lineal múltiple (Vasanth Kumar et al., 2008).

Las ANN se han utilizado para extraer información de herramientas PAT (Tecnología analítica de procesos) ricas en datos, como las imágenes microscópicas en línea. Una ResNet (Residual Network), un tipo de CNN, ha demostrado su eficacia en la clasificación de cristales detectados en imágenes, que se utilizó para la clasificación de la contaminación con una precisión superior al 98% (Salami et al., 2021). Esta técnica en línea permite identificar trazas de polimorfos no deseados y utilizarse así, con el fin de aumentar la pureza del producto. Además, la tasa de crecimiento también puede predecirse midiendo la dispersión del tamaño de las partículas mediante el análisis de imágenes en línea basado en CNNs (S. Chen et al., 2019).

Otro uso de las ANN en este contexto es el control de la cristalización. Por ejemplo, en una RNN diagonal (Brereton, 2012), para el control de la temperatura y el novel o la determinación del perfil de temperatura óptimo para la reducción de finos. Para esto último, (Paengjuntuek et al., 2012) generaron datos para el entrenamiento de la ANN, luego calcularon la concentración de la solución y el volumen del cristal utilizando los datos de temperatura y concentración de la disolución en puntos temporales anteriores. La ANN predijo el estado de optimización y proporcionó así, un mayor rendimiento de control que las metodologías convencionales. Además, uno de los aspectos más importantes es que el entrenamiento de estas tiene un coste computacional menor que los modelos convencionales.

4.1.2. Aplicación de ANNs en procesos industriales posteriores:

- **Mezclado:**

La correcta ejecución de la mezcla de polvos garantiza principalmente la distribución homogénea de los componentes en la fabricación de formas farmacéuticas sólidas. Las técnicas de ML se han aplicado en varias ocasiones para ayudar al análisis de datos

inmediato de la concentración de API durante un proceso de mezcla y predecir el comportamiento de los polvos en varios escenarios (Nagy et al., 2022).

A partir de los años 2000, la predicción de la concentración de API en polvos mediante ANNs basadas en espectros de infrarrojo cercano (NIR) (B. Wang et al., 2009) llegó a ser tan eficaz como la regresión PLS. Por otro lado, pueden predecir el tiempo necesario para lograr una mezcla homogénea. Por ejemplo, (Tewari et al., 2017) utilizaron espectroscopía NIR, ANNs y otros métodos de análisis multivariados para técnicas de detección de puntos finales de mezcla en línea. Por otro lado, Mujumdar et al. (Mujumdar et al., 2007) crearon un modelo con el que simularon la mezcla de dos fracciones de partículas con diferentes tamaños bajo diversos parámetros. Los resultados simulados se utilizaron como conjunto de datos de entrenamiento para predecir la concentración media de mezcla, un parámetro que describe la eficacia del proceso de mezcla basado en la amplitud de las oscilaciones del mezclador, su frecuencia, el tamaño de partículas y el número de ciclos.

En general, las ANNs son una opción excelente cuando varios parámetros de funcionamiento tienen una relación no lineal. Podría también aplicarse a datos de proceso en los que los efectos de ciertos factores aparecen posteriormente (Nagy et al., 2022). Por ejemplo, la composición de una mezcla en un mezclador continuo puede predecirse mediante una RNN basándose en el caudal másico de los flujos de material de entrada y en la distribución del tiempo de residencia del sistema (Y. Chen et al., 2020).

- **Granulado:**

La granulación es una técnica de aglomeración de partículas que resulta esencial para garantizar la procesabilidad posterior, influyendo en la calidad (uniformidad del producto y la disolución) del producto final. La granulación se aplica por vía seca o húmeda en aparatos de alto cizallamiento, lecho fluido, compactador de rodillos, etc (Nagy et al., 2022).

La AI se ha manifestado durante estos años atrás creando modelos de ANNs para predecir la calidad del producto basándose en los parámetros del proceso de lecho fluidizado (Petrović et al., 2011), granulación húmeda de alto cizallamiento (Sampat & Ramachandran, 2021) y granulación seca (Turkoglu et al., 1999). Por ejemplo, (Kesavan & Peck, 1996) emplearon las ANNs para predecir el tamaño de las partículas, la velocidad

de flujo, la densidad aparente y la densidad a granel para modificar tanto la granulación de alto cizallamiento como la granulación fluida. En los nodos de entrada se dispusieron el tipo y porcentaje de diluyente, el tipo de equipo de granulación y la cantidad y el método de adición del aglutinante. Los atributos críticos de calidad (CQA) del producto pudieron predecirse con buena precisión, y la ANN funcionó mejor que el análisis de regresión multilineal. Además, las ANN también facilitaron el escalado de los procesos de granulación en lecho fluido húmedo (Millen et al., 2019). Por otro lado, (Korteby et al., 2018) demostraron con un proceso de granulación fluida por fusión en caliente que la relevancia de las variables de entrada independientes del modelo ANN podía determinarse cuando se combinaban con la ecuación de Garson. Esta ecuación determina la importancia relativa de las variables no lineales en la predicción de una variable lineal. Identificaron que el tamaño de partícula del aglutinante tuvo el mayor impacto en las propiedades de los gránulos finales, seguido del grado de viscosidad y contenido del aglutinante. Así, la ANN combinó los aspectos positivos de los modelos inspirados en los primeros principios y en los datos, proporcionando como salida información sobre el efecto de los factores. En la granulación en seco, se pudo predecir la distribución del tamaño del gránulo obtenido tras la molienda. (Kazemi et al., 2017) emplearon una ANN para predecir la distribución del tamaño del gránulo (GSD) estudiando las propiedades de las cintas de la compactación en rollo y cómo afectan al GSD. Otros empleos de ANN en granulación en seco fue el de (Inghelbrecht et al., 1997) sobre la friabilidad de la cinta o la densidad de la cinta (Sajjia et al., 2017).

La AI también tiene la capacidad de analizar inmediatamente los datos generados por los sensores utilizados como herramientas PAT en el proceso de granulación para, controlar el contenido de API o el contenido de humedad residual.

- **Compresión/recubrimiento:**

En la mayoría de los procesos de compresión, el comprimido es la creación individual del producto final. Garantizar que cada uno de los comprimidos que recibe el paciente cumple los estrictos requisitos de la calidad es esencial. Las tecnologías de modelado predictivo y PAT ofrecen ayuda para lograr este objetivo.

Al desarrollar un proceso de tableteado, conocer cómo se comporta la mezcla de polvo comprimido es clave. La fluidez de la mezcla debe ser lo suficientemente buena como para que, cada vez que se llene la matriz, se desplace en su interior una masa de polvo

casi idéntica (Nagy et al., 2022). Por otro lado, (Kachrimanis et al., 2003) utilizaron una ANN FF-BP (Feed-Forward Back Propagation ANN) para predecir la velocidad de flujo de polvos a través de un orificio circular. Utilizaron propiedades como la densidad aparente, la densidad compactada, el diámetro de las partículas, la relación de aspecto, la esfericidad, etc. Las predicciones de caudal obtenidas fueron más precisas que las de la ecuación de caudal de Jones y Pilpel, expresión que determina cómo varía la velocidad de reacción con respecto a la concentración de un reactante.

La flexibilidad de las ANN permite procesar eficazmente cualquier tipo de señal. Por ejemplo, los termopares sirven como herramientas PAT. (Korteby et al., 2016) colocaron tres conjuntos de termopares en el interior de un granulador cónico de lecho fluidizado. Registraron temperaturas en diferentes condiciones, ya que la dispersión de la temperatura dentro del granulador puede influir en las propiedades del gránulo; por tanto, comprender su dependencia de los parámetros de fabricación puede contribuir a la creación de un proceso más fiable. Los datos que se obtuvieron permitieron entrenar a la ANN, la cual proporcionó predicciones muy precisas para los casos de prueba. Cabe esperar, que dicho modelo pueda formar parte de un control de calidad específico con el tiempo.

4.2. 'Machine Learning'.

El 'Machine Learning', también conocido como aprendizaje automático, es un área de investigación en AI que se dedica a desarrollar algoritmos y modelos que brindan a los ordenadores adquirir conocimiento a través del análisis de datos, sin requerir instrucciones explícitas de programación. Los modelos de ML utilizan técnicas estadísticas y matemáticas para reconocer patrones en los datos y realizar predicciones o tomar decisiones (Figura 3). Es el arquetipo básico que implica múltiples campos basados en metodologías y algoritmos para reconocer patrones dentro de los datos. La automatización implica el uso de DL y ML, pero mantienen diferencias. El ML se clasifica en varios tipos, y el DL es el subcampo del ML que utiliza redes artificiales que interconectan elementos (Ben-Bassat et al., 2018).

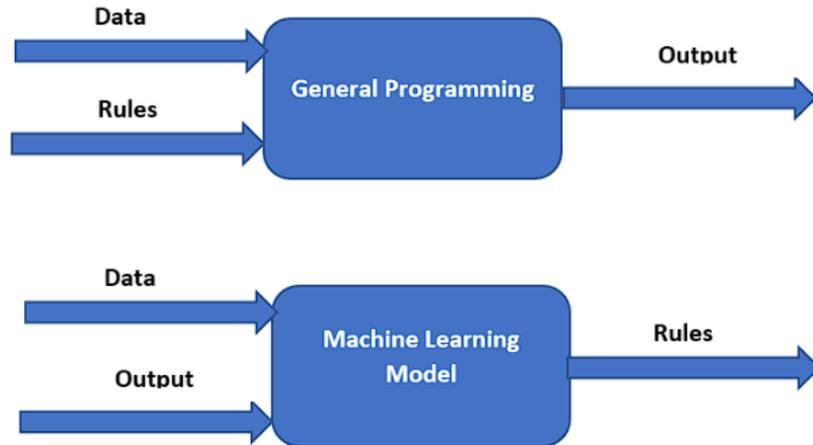


Figura 3. Comparativa de programación general frente a ML (SQLRelease, n.d.)

El ML es equivalente a las neuronas biológicas humanas e imita la transmisión de impulsos electrónicos. Este es un modelo matemático que muestra patrones subyacentes disponibles en los datos y la información, y se aplica para aprender métodos de predicción de datos futuros. La fuerza del ML se debe a resolver cuestiones matemáticas complejas y se utiliza en diversos campos de la biología moderna. En términos de métodos de ML generalizados, se ha aplicado para predecir con precisión el conjunto de datos no vistos en la elección del método para optimizar el rendimiento en situaciones complicadas. En el contexto del análisis de sensibilidad, los modelos de ML pueden aprender los patrones y dependencias subyacentes entre las variables de inicio y las respuestas de salida. Pueden capturar relaciones complejas y no lineales que no son evidentes mediante métodos de fuerza bruta (basados en un enfoque de prueba y error), que permiten identificar qué variables de entrada tienen un mayor impacto en los resultados del sistema y a comprender la relación entre las variables y los resultados. Al entrenarse con datos históricos y aprender de los patrones en los datos, los algoritmos de ML pueden proporcionar información sobre la sensibilidad de las salidas del modelo ante cambios en las variables de entrada. (Graupe & Vern, 2001).

El aprendizaje por refuerzo (Reinforcement Learning) es otra categoría principal del método ML basada en la interacción de un agente con un entorno. El objetivo del mismo es la obtención de los beneficios a largo plazo a partir de la toma óptima de acciones en respuesta a la información ambiental (Agatonovic-Kustrin & Beresford, 2000). El AutoDock, software empleado para el acoplamiento molecular, que es el proceso de predecir cómo una molécula se une o se acopla a una proteína objetivo, los métodos ML están bien preparados y entrenados en relación con interacciones específicas de

proteína/ADN, residuos de aminoácidos del sitio activo, pequeñas moléculas o fármacos, en vez de proporcionar funciones de puntuación como la *'Drug score'* (métrica que clasifica y compara diferentes compuestos en términos de eficacia, seguridad y demás propiedades relevantes) (Närhi et al., 2018; Ravindranath et al., 2015). El ML se ocupa de múltiples aplicaciones en diversas etapas de cribados virtuales de alto rendimiento (HTVS, High-Throughput Virtual Screening) para el descubrimiento de fármacos, en relación con el conocimiento de su función e interacciones y la identificación de típicos en el entendimiento de los nuevos fármacos y *leads* (o molécula líder, se denomina así a los compuestos químicos iniciales que muestran actividad biológica prometedora frente a una diana terapéutica). Los métodos ML analizan la información de los sitios de unión de los receptores, en conjunto con la homología de los mismos y la información disponible de la enfermedad (Lavecchia & Giovanni, 2013; You et al., 2019).

4.2.1. ML para la definición de objetivos:

El descubrimiento de fármacos está altamente relacionado con el proceso de identificación de proteínas diana con aspectos causales de la fisiopatología. Si se malinterpreta la información relacionada con la proteína objetivo, esto puede afectar en cómo se modula la información sobre la enfermedad por parte de las herramientas de AI. Por tanto, la selección correcta de la proteína objetivo es crucial en un proceso general de descubrimiento y desarrollo (Schenone et al., 2013; Singh et al., 2018). La respuesta exitosa del API en un ensayo clínico aleatorio se tendrá en cuenta para evaluar su eficiencia y, en última instancia, para identificar los objetivos más prometedores del API en el desarrollo de medicamentos. El algoritmo ML predice los sucesos, eventos y problemas biológicos no observados (Lima et al., 2016). Los genes susceptibles de fármacos, también conocidos como genes farmacogenéticos o variantes genéticas relacionadas con la respuesta a los fármacos, son variaciones genéticas que pueden influir en la manera en que una persona responde a un medicamento. Estas variaciones pueden afectar la eficacia, seguridad y tolerabilidad de un fármaco en un individuo.

Cada individuo tiene un conjunto único de genes que determinan su respuesta a los medicamentos. Algunas variantes genéticas pueden hacer que una persona sea más sensible a un fármaco en particular, lo que significa que pueden experimentar efectos secundarios adversos o que el medicamento puede no ser eficaz para ellos. La ML se ha aplicado en este ámbito, por ejemplo, (Costa et al., 2010) desarrollaron un modelo

informático para prever los genes asociados a enfermedades y los genes susceptibles de fármacos a escala de todo el genoma. Este modelo ha sido muy utilizado para reducir los laboriosos procedimientos experimentales e identificar las posibles dianas moleculares de los fármacos relacionadas con los mecanismos de la enfermedad en cuestión. Este modelo se adapta en relación con los fundamentos biológicos del genoma de acuerdo para cada caso específico.

Las principales características de clasificación son:

- La esencialidad del gen.
- La aparición de mutaciones.
- Interacciones de la proteína codificada con otras proteínas.

El análisis meta-clasificador, que es un enfoque dentro del campo del ML, ha arrojado ciertos resultados en relación con la recuperación de genes patógenos o, genes causantes de enfermedades, conocidos y genes *druggables* (es decir, genes que son considerados como objetivos viables para el desarrollo de fármacos o terapias) desconocidos. El análisis meta-clasificador logró identificar el 65% de genes patógenos de todos los conocidos actualmente. Además, identificó el 78% para los genes *druggables* desconocidos, lo cual significa que identificó con precisión ese porcentaje de genes que tienen el potencial de ser blanco para futuros fármacos, a pesar de que previamente eran desconocidos como tal.

Estos resultados sugieren que el análisis meta-clasificador tiene un buen desempeño en la identificación de genes mórbidos y genes *druggables*, lo que puede tener implicaciones importantes en el descubrimiento de nuevos enfoques terapéuticos y el desarrollo de medicamentos. (Qi et al., 2006; Zhang et al., 2017). El árbol de decisión (DT) y las reglas descubiertas inspeccionaron los parámetros que incluyen la localización en membrana y la regulación de múltiples factores de transcripción para identificar rasgos biológicos (Sakkiah et al., 2017). Esto también contribuye a la comprensión y el diseño de los principios de biosistemas mediante la aplicación de los métodos de ingeniería inversa. (Doane & Elemento, 2017; Liu, 2015). (Volk et al., 2020) utilizaron técnicas de ML para analizar y comprender los patrones y procesos en el nivel molecular y celular relacionados con el ADN, las proteínas y las vías bioquímicas y metabólicas. Estos métodos pueden

ayudar a comprender mejor la forma en que las comunidades genómicas y celulares funcionan y enfrentan desafíos. (Jeon et al., 2014) propusieron y crearon un método de análisis utilizando SVM para analizar datos genómicos y sistemáticos con el objetivo de identificar proteínas específicas relacionadas con el cáncer de mama, cáncer de páncreas y cáncer de ovario. El enfoque se centró en distinguir proteínas asociadas a patologías basándose en su similitud con otras proteínas conocidas (homólogos) o en la probabilidad de que se unan a fármacos. Esto puede ayudar a identificar posibles blancos terapéuticos y comprender mejor las características moleculares asociadas con estos tipos de cáncer. (Mamoshina et al., 2018) utilizaron un enfoque similar al mencionado anteriormente para identificar objetivos terapéuticos enfocados en el tejido muscular y en comprender la base molecular del envejecimiento humano. Utilizando un modelo SVM (máquinas de vectores de soporte) con un núcleo lineal y una selección profunda de características, se buscó identificar los genes de expresión que están vinculados con el proceso de envejecimiento. Este enfoque puede ayudar a identificar posibles objetivos terapéuticos relacionados con el envejecimiento y comprender mejor los mecanismos moleculares involucrados en este proceso.

4.2.2. ML para el estudio de imágenes.

Los enfoques de ML también se utilizan para investigación en la automatización o adquisición de imágenes robóticas. Por ejemplo, (Ljosa et al., 2013), estudiaron inhibidores potentes de ligandos del receptor β_2 adrenérgico mediante un ensayo de unión radiológica utilizando nuevas moléculas de bajo peso molecular. En este caso, se realizó un cribado para identificar moléculas con capacidad para interferir con el ligando radiomarcado y su afinidad de unión. El efecto agudo de estas moléculas puede causar alteraciones en la resonancia de plasma en la superficie (SPR) detectada en los receptores. Este efecto permite seleccionar inhibidores del receptor para ser procesados en la etapa de optimización de las moléculas líderes (Aristotelous et al., 2013). Sin embargo, este proceso es laborioso y largo, por lo que métodos alternativos como el cribado fenotípico son de gran interés. En este contexto, se aplicaron análisis basados en ML para identificar fenotipos complejos que tienden a aumentar la eficacia de las moléculas inhibitoras (Lee & Berg, 2013; Swinney & Lee, 2020). Otra técnica, la imagen avanzada, es un mecanismo aplicable para identificar los fenotipos y la perturbación causada por las moléculas inhibitoras, mejorando así las predicciones (Scheeder et al., 2018). En términos más generales, la obtención de imágenes puede estar compuesta por dos campos:

- El cribado fenotípico, el cual se enfoca en fenotipos predefinidos de moléculas de señalización intracelular asociadas con el mecanismo de la enfermedad
- Las diversas estructuras subcelulares como anticuerpos, agentes infecciosos o químicos, y marcadores fluorescentes que permiten estudiar sus respuestas.

4.2.3. ML para el HTVS.

El HTVS es esencial en la búsqueda de nuevos medicamentos y se integra con nuevos métodos para actualizaciones tecnológicas. Por lo tanto, se utilizan métodos AI/ML para identificar y seleccionar los objetivos terapéuticos más relevantes y prometedores, que luego se llevan a cabo en experimentos posteriores (Zhavoronkov et al., 2020).

(Valentini et al., 2014) desarrollaron una combinación de redes génicas funcionalmente diferentes con métodos basados en núcleos (kernels) para clasificar el orden de los genes. Más tarde, (Ferrero et al., 2017) trabajaron con datos asociativos objetivo-enfermedad disponibles en bases de datos para predecir nuevos objetivos terapéuticos.

El método SVM fue desarrollado por Vladimir Vapnik y su equipo en los años 90. Vladimir Vapnik, junto con Alexey Chervonenkis, fue el precursor en el desarrollo de las bases conceptuales de los SVM y su uso en ML y la clasificación de patrones. Los SVM son un tipo de algoritmo de aprendizaje supervisado utilizado para problemas de clasificación y regresión. Su objetivo principal es encontrar la mejor separación posible entre diferentes clases de datos utilizando un hiperplano en un espacio de alta dimensión. El SVM busca un límite o frontera para separar las muestras de diferentes clases de manera óptima. El margen máximo, o distancia entre las muestras de diferentes clases, se logra seleccionando este límite, también conocido como hiperplano (Vapnik, 1999). (Arabfard et al., 2019) utilizaron métodos de aprendizaje no supervisado como Naïve Bayes, Spy y Rocchio-SVM para predecir y clasificar alrededor de 3000 objetivos asociados con un gen del envejecimiento, categorizando los genes humanos según esta implicación en el envejecimiento. Por otro lado, (Selvaraj, Vierra, et al., 2021) también estudiaron detalladamente la correlación entre los objetivos terapéuticos y las enfermedades en el descubrimiento de fármacos. El proceso posterior implica encontrar

una molécula de bajo peso molecular apta para intervenir en el mecanismo de la enfermedad. Generalmente, se diseña un candidato a fármaco adecuado que se presenta a empresas farmacéuticas o se deposita en grandes bibliotecas de compuestos.

El algoritmo de clasificación SVM (Ranking SVM) es una función de recuperación de aprendizaje que emplea métodos de clasificación por pares para ordenar de forma adaptativa los resultados en función de su "relevancia" para una consulta específica. Esta clasificación se asemeja a otros métodos de muestreo como la discriminación lineal, y el hiperplano es la estrategia principal para clasificar la dependencia de las condiciones límite (Burbidge et al., 2001). Particularmente, siendo la clasificación esencial, se maximiza la longitud entre la instancia y el hiperplano y solo un pequeño subconjunto de instancias de entrenamiento define el límite (Maltarollo et al., 2019).

Como se ha mencionado anteriormente, el SVM permite analizar gran cantidad de datos; como ejemplo, es capaz de analizar casi 36000 compuestos del conjunto de datos sobre SIDA del NCI (Instituto Nacional del Cáncer) (Geppert et al., 2008; Wilton et al., 2006). Por otro lado, otro estudio muy interesante fue el realizado por (Franke et al., 2005), ellos examinaron 2.7 millones de moléculas de bajo peso molecular de una base de datos de ligandos. El objetivo del estudio era encontrar moléculas con actividad antagonista de Cox-2 (Ciclooxigenasa-2). Para llevar a cabo este análisis, se utilizó un conjunto de entrenamiento compuesto por 94 compuestos de referencia que actúan como antagonistas de Cox-2. Posteriormente, se estudiaron varias moléculas nuevas y, como resultado, solo se encontraron tres moléculas con una actividad potente como antagonistas de Cox-2. Lo destacado del estudio es que una de las moléculas encontradas mostró perfiles inhibidores más potentes que Celecoxib y Rofecoxib, antagonistas de Cox-2 de referencia utilizados en el tratamiento de edemas y el dolor.

4.2.4. ML para el diseño de fármacos basado en estructuras.

El proceso de descubrimiento de medicamentos comienza con la identificación de una diana objetivo, su evaluación y la búsqueda de fármacos candidatos adecuados (Patidar et al., 2016; Sharma et al., 2018). La selección de dianas desempeña un papel imperial en la patología de las enfermedades, evaluando la capacidad farmacológica de las moléculas líderes y priorizando las dianas candidatas (Reddy & Singh, 2014; Sliwoski et al., 2014). Sin embargo, debido a la compleja naturaleza de las enfermedades humanas, el proceso

de selección de dianas necesita métodos exhaustivos para analizar datos heterogéneos y comprender el mecanismo a nivel molecular de los distintos tipos de enfermedades, y que también ayuden a identificar los cambios específicos en los pacientes (Subramanian et al., 2020).

Se han aplicado métodos avanzados como el AI/ML para superar estos retos. Por ejemplo, los códigos DL (DL Codes) son fragmentos de código o bibliotecas de software que implementan algoritmos y modelos de DL. Estos códigos suelen estar escritos en lenguajes de programación como Python, utilizando bibliotecas populares como TensorFlow, PyTorch o Keras, que proporcionan herramientas y funciones para crear y entrenar DNN (Cova & Pais, 2019). Uno de los métodos ML utilizados en el diseño basado en estructuras es el SOM-based Prediction of Drug Equivalence Relationships (SPiDER), un enfoque utilizado en el campo de la farmacología y el descubrimiento de medicamentos para predecir la relación de equivalencia entre diferentes medicamentos. Esta herramienta aplica ampliamente el algoritmo de red neuronal (NN) para discretizar los vectores iniciales en mapas de características de forma no supervisada (Brereton, 2012). En la predicción de las interacciones entre fármacos y proteínas, se utilizan descriptores moleculares para comparar la similitud entre una molécula de fármaco y moléculas de referencia en una "neurona". Sin embargo, estas predicciones basadas en la similitud de descriptores no permiten identificar de manera definitiva la proteína objetivo con la cual el fármaco interactúa. Aunque las predicciones pueden sugerir qué proteína puede ser el objetivo del fármaco, se requieren investigaciones adicionales para confirmar de manera concluyente la diana del fármaco y su interacción precisa con una proteína específica. Este método de predicción se realiza con el conjunto disponible de descriptores CATS2 (Chemically Advanced Template Search 2) para crear plantillas o patrones que describen características comunes en una clase de compuestos, para identificar, por ejemplo, en este caso, los átomos que sean hidrógeno (Cherkasov et al., 2014). Recientemente, SPiDER se ha aplicado en el diseño de fármacos *'de novo'*, también conocido como diseño racional de fármacos, referido a la creación de nuevas moléculas de fármacos desde cero, sin basarse en compuestos existentes o estructuras previamente conocidas. SPiDER se aplica para identificar las nuevas estructuras químicas que se ajustan a un espacio farmacológico definido para la síntesis de moléculas de bajo peso molecular y su validación experimental (Yang et al., 2019).

Otro método conocido para la predicción de la farmacología en red es DECRYPT (DEcRyPT, por sus siglas en inglés, Design, Evaluate, Control, and Predict Together). Este método puede aplicarse tanto de manera individual como en combinación con SPiDER, utilizando el algoritmo de Random Forest (RF) y descriptores CATS2 (Lind & Anderson, 2019). Mediante este método, los predictores definidos por los usuarios analizan de forma independiente una porción del conjunto de entrenamiento antes de la agregación de las salidas de resultados. Un ejemplo específico es la aplicación de β -lapachone como una molécula específica en el estudio de la 5-lipoxigenasa como objetivo terapéutico. DEcRyPT se utilizó para analizar cómo β -lapachone interactúa con la enzima 5-lipoxigenasa, lo que puede tener implicaciones en el diseño y desarrollo de nuevos fármacos para abordar enfermedades relacionadas con esta enzima (Rodrigues et al., 2018). Después de identificar un objetivo ideal, una estrategia medicinal novedosa se basa en encontrar moléculas de bajo peso molecular adecuadas que alteren la estructura nativa del objetivo (Garscha et al., 2016). La biología contemporánea, especialmente en la investigación del cáncer, depende en gran medida de estas pequeñas moléculas y de procesos innovadores en la investigación y desarrollo de medicamentos.

Es necesario identificar moléculas que tengan similitudes estructurales con el ligando y poder realizar modificaciones funcionales para interactuar adecuadamente con la molécula diana. Algunas dianas terapéuticas pueden carecer de un dominio descrito para la unión del ligando, pueden activarse sin la presencia del ligando, pueden tener múltiples ligandos o pueden tener un ligando desconocido (Park et al., 2020). Estas razones explican la posibilidad de que moléculas de bajo peso molecular muestren reactividad cruzada con otros receptores, por ello, estos son factores a considerar. Por ejemplo, en el campo de la oncología, existen algunos ejemplos de enfoques para el desarrollo de medicamentos que superen las limitaciones de interacción o estructurales de la molécula líder a la diana de interés (Freedman et al., 2017). Estos avances o enfoques, registrados en diferentes bases de datos, permiten a metodologías como el *Virtual Screening of Potential Inhibitors* (Cribado Virtual de Inhibidores Potenciales), identificar compuestos químicos que puedan tener actividad inhibidora contra una diana molecular específica, como una enzima, una proteína o un receptor (Šmak et al., 2021). Como ejemplos de empleo de la metodología destacamos por ejemplo su uso para encontrar inhibidores contra la proteína NS3 del virus del Zika (Sahoo et al., 2016), o el estudio por parte de (Ruan et al., 2021) de búsqueda de fármacos contra el Covid-19.

4.2.5. ML en la interacción fármaco-diana.

Se utilizan varios modelos de ML para evaluar las cavidades de las dianas farmacológicas donde interactúan los fármacos mediante el estudio de diferentes características. El *'Surface Cavity Recognition and Evaluation'* (SCREEN) es uno de los servidores web de ML conocidos construido sobre la base del clasificador RF para analizar las propiedades estructurales, geométricas y fisicoquímicas de las interacciones farmacológicas a los diferentes sitios activos, incluido especialmente el sitio activo de interés (Bruno et al., 2019). La superficie del sitio activo de la proteína, el espacio de asignación y la geometría de la forma de las cavidades del sitio activo son esenciales en el proceso de clasificación (Selvaraj, Chandra, et al., 2021). Estudios previos han aplicado el método SVM para predecir los sitios activos en las proteínas diana basándose en sus propiedades fisicoquímicas disponibles en las secuencias proteicas (Q. Li & Lai, 2007).

4.2.6. ML y otras AI en química cuántica.

La mecánica cuántica (QM) o los métodos híbridos QM/mecánica molecular (QM/MM) son otras estrategias que también se emplean para predecir la interacción proteína-ligando en el proceso de descubrimiento de medicamentos (Selvaraj, Dinesh, et al., 2021). Estos métodos reúnen sistemas cuánticos para la simulación del sistema a nivel atómico y ofrecen mayores niveles de precisión que los otros métodos clásicos disponibles (Aminpour et al., 2019).

En los métodos clásicos de MM (mecánica molecular), se considera la función de energía simple basada en las coordenadas atómicas implicadas. Recientemente, la implementación de métodos de AI al cálculo de QM ofrece una precisión mucho mayor debido a la inclusión de cargas y mayor rapidez que el método de MM clásico (Kubař & Elstner, 2013). En la AI, los métodos están altamente entrenados para ofrecer las energías reales disponibles en las coordenadas atómicas y a velocidades superiores con respecto a los métodos clásicos de MM. Generalmente, estos métodos son utilizados para predecir las propiedades eléctricas basados en simulaciones atómicas.

Por el contrario, los métodos DL se aplican ampliamente para la predicción de la energía de moléculas pequeñas, por ejemplo, las energías potenciales basadas en la teoría del funcional de la densidad (DFT) derivadas de la química cuántica se calculan incluso para

moléculas de gran tamaño. Las energías potenciales derivadas de la DFT derivadas de la química cuántica se calculan incluso para una gran base de datos que contiene 2 millones de cristales de elpasolita, cristales con una estructura cúbica que consisten en haluros de metales alcalinos y metales alcalinotérreos. Estos cristales tienen aplicaciones en detectores de radiación, dispositivos ópticos y láseres, así como en la investigación científica. La precisión del modelo ML mejora al aumentar el tamaño de los datos de la muestra y alcanza 0,1 eV/átomo para el cálculo DFT entrenado en 10.000 moléculas (Tkatchenko, 2020). El enfoque ML con propiedades QM permite un análisis más detallado y preciso a nivel atómico. La integración de la regresión Kernel Ridge (KRR), la regresión del proceso gaussiano (GPR) y las NN son técnicas que se utilizan para modelar y predecir las propiedades moleculares y su relación con la actividad farmacológica, herramientas con alto potencial para el desarrollo de fármacos.

Teniendo en cuenta la fiabilidad de la baja complejidad numérica y la alta precisión de los algoritmos, ML ofrece un uso cómodo, atractivo y supone una alternativa para los cálculos ab initio, es decir conjunto de técnicas y métodos computacionales utilizados para resolver ecuaciones cuánticas que describen el comportamiento de los electrones en un sistema atómico o molecular, y DFT (Huang et al., 2007). Con la notable capacidad de comprender la complejidad en los datos, los métodos integrados ML son los más eficientes para calcular los campos de fuerza y de QM empíricos. Sin embargo, los modelos ML predichos dependen de la calidad y cantidad de los datos entrenados. Las NN modelan eficaz y eficientemente los datos de referencia con alta flexibilidad, fiabilidad, rentabilidad y permiten orientar una gran cantidad de datos de referencia para entrenar las ML para una predicción precisa.

Recientemente, (Smith et al., 2016) desarrollaron ANAKIN-ME o ANI. Este es un software para la predicción precisa de las energías moleculares utilizando NN. Se utiliza en el campo de la química computacional y la modelización molecular para predecir las energías de las moléculas con alta precisión. Se desarrolló para evaluar un extenso conjunto de datos que contiene 22 millones de conformaciones de moléculas de bajo peso molecular. Este método tiene la capacidad de predecir las energías incluso en moléculas de alto peso molecular. Se destaca que el rendimiento y la aplicabilidad del sistema han sido demostrados en moléculas de hasta 70 átomos. En particular, ANI-1 (versión más limitada de ANI que utiliza menor cantidad de parámetros y aproximaciones) se destaca

por su capacidad para predecir la energía en función del tamaño molecular externo, utilizando el error cuadrático medio (RMSE) en comparación con la energía DFT a temperatura ambiente. Posteriormente (Koohy, 2017), informó sobre el uso de un sistema orgánico extenso, entrenado con moléculas pequeñas fragmentadas y con la adición de la teoría del funcional de la densidad (DFT) para sistemas más grandes

4.2.7. ML y otras AI para propiedades fisicoquímicas.

Para descubrir nuevos medicamentos, la detección temprana de las propiedades fisicoquímicas, especialmente en lo relativo a su biodisponibilidad y farmacocinética, de una molécula líder resulta indispensable para reducir el riesgo en la fase primaria. Hay numerosos estudios basados en la AI para la predicción de estas propiedades concretas. Por ejemplo, (Duvenaud et al., n.d.) utilizó una combinación de ANN-CNN para predecir la solubilidad de una molécula principal mediante la extracción de información de un grafo molecular con un rendimiento predictivo muy eficaz. Más tarde (Coley et al., 2017), superaron el método de Duvenaud et al para poder predecir la solubilidad de una molécula principal en medio acuoso.

La biodisponibilidad es un factor clave, la predicción de esta se aplica en la orientación de la optimización de las moléculas en desarrollo. Los estudios de perfiles ADME/T (Absorción, Distribución, Metabolismo, Excreción y Toxicidad) permiten obtener una serie de parámetros utilizados en farmacología y ciencias farmacéuticas para evaluar y predecir el comportamiento de un fármaco en el organismo. La aplicación de técnicas ML para estos perfiles se orienta a la elaboración de modelos predictivos mediante la extracción de datos de entrenamiento y que predicen los valores pKa de las moléculas líderes (F. Yamashita et al., 2002). La aplicación de diferentes modelos ML a partir de datos clínicos ADME/T es realmente útil para el desarrollo de predicciones de toxicidad. Antes de que se apruebe el uso clínico de un compuesto en desarrollo, se requieren estos perfiles para determinar su efectividad y garantía. Tras el conocimiento de la absorción, la distribución es uno de los factores más importantes. Actualmente, el uso de matrices PLS, MR y DT son útiles para conocer la distribución del fármaco y la evaluación de este (Castillo-Garit et al., 2008).

Por otro lado, en relación con la toxicidad, los estudios de toxicidad hepática y renal se han llevado a cabo tradicionalmente para predecir el perfil de toxicidad de los fármacos mediante el conocimiento experto basado en reglas y de alerta estructural, es decir, de identificación de características estructurales específicas en un compuesto químico que están asociadas a la toxicidad (Milanetti et al., 2016). Por lo tanto, en los últimos años se han implementado métodos de DL para manejar automáticamente los diversos rasgos característicos químicos en la predicción de la toxicidad. Un ejemplo, una CNN de codificación de grafos moleculares (MGE-CNN), con el modelo de predicción de toxicidad oral aguda, obtiene resultados muy precisos (Xu et al., 2017). Otra herramienta esencial, es la base de datos Tox21, la cual comprende más de 12.000 muestras de entrenamiento y más de 600 muestras de prueba que representan compuestos químicos. Hay ‘características densas’ que representan descriptores químicos, como peso molecular, solubilidad o área superficial de la molécula, y cientos de miles de ‘características dispersas’ que representan subestructuras químicas. Los métodos ML pueden usar datos dispersos, densos o combinarlos. Para cada muestra hay 12 etiquetas binarias que representan el resultado (activo/inactivo) de 12 experimentos toxicológicos diferentes (Mayr et al., 2016). Junto con Tox21, la herramienta XenoSite también resulta de gran utilidad, se trata de una herramienta de predicción para la bioquímica de moléculas pequeñas, es decir, sobre el metabolismo humano *in vivo* (Matlock et al., 2015).

Otro método basado en la predicción de la farmacocinética es el del metaclasificador basado en DT, este combina las predicciones de varios DT para obtener una predicción más precisa y robusta. Cada árbol de decisión es un modelo individual que toma características específicas del fármaco, como su estructura molecular, propiedades fisicoquímicas y otros descriptores relevantes, y genera una predicción sobre su comportamiento en términos de perfil de ADME (M. Gao et al., 2019).

4.3. ‘Deep Learning’:

El método DL, también conocido como método de aprendizaje profundo, ya mencionado anteriormente en este trabajo, representa la NN poseyendo múltiples capas ocultas, y es muy utilizado por su flexibilidad para aprender funciones arbitrariamente complejas. Puede aprender tanto como sea posible con los datos adecuados y la inversión de tiempo

computacional, proporcionando salidas altamente fiables. Las múltiples capas ocultas en los modelos DL ofrecen un acceso flexible para aprender módulos arbitrariamente complejos (Figura 4), que proporcionan directamente neuronas adecuadas y conjuntos entrenados.

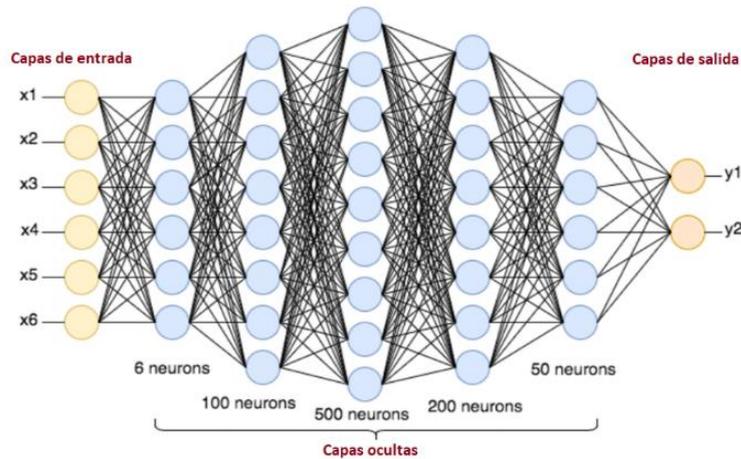


Figura 4. Ejemplo de estructura de un DL (Bahi & Batouche, 2018)

El DL aplica un algoritmo de retropropagación y un método de optimización basado en gradientes que permite el entrenamiento de NN un ejemplo como el de la Figura 5, dando como resultado una diferenciación de extremo a extremo (Sakellarpoulos et al., 2019). Además, las redes *feed-forward* interrelacionan capas, arquitectura convencional y convolucional que avanzan hacia el desarrollo de varios dominios y tipos de datos. La tendencia basada en la lectura de datos, los productos que utilizan algoritmos avanzados o complejos y el hardware computacional están impulsando la metodología DL (Del Fiol et al., 2018; R. Yamashita et al., 2018). Desde un punto de vista algorítmico, las redes DL con más capas han experimentado con frecuencia la desaparición de gradientes, lo que dificulta el aprendizaje eficiente de modelos. Las redes profundas entrenadas eficientemente se mejoran significativamente con los nuevos esquemas de inicialización de métodos, la función de activación neuronal y los métodos de optimización basados en gradientes (Poggio et al., 2020).

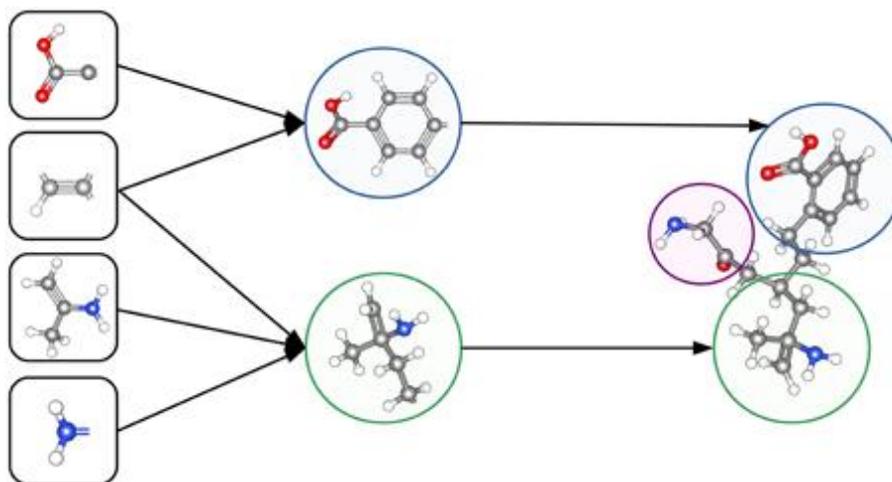


Figura 5. Ejemplo de uso de DL: Representación de un toxicóforo por características jerárquicamente relacionadas (Mayr et al., 2016)

4.3.1. Predicción del plegamiento de proteínas a partir de la secuencia.

Gran parte de los trastornos patológicos están asociados a disfunciones de las proteínas. De ahí que la comprensión de los aspectos estructurales proteicos desempeñe un papel esencial en el enfoque SBDD (Structure-Based Drug Design, Diseño de Fármacos Basado en Estructura). Este puede utilizarse para identificar las moléculas potencialmente activas frente a las proteínas diana de los fármacos. Se estima que predecir y analizar la estructura tridimensional de todas las proteínas diana mediante métodos *in vitro* será un proceso rentable y que requerirá mucho tiempo (Aparoy et al., 2012). Por lo tanto, el desarrollo de algoritmos apropiados para predecir la estructura modelo de una proteína es muy importante en la actualidad. Al utilizar algoritmos de predicción de estructuras de proteínas, se pueden obtener modelos aproximados de la estructura de una proteína de manera más rápida y económica en comparación con los métodos experimentales. Esto ahorra tiempo valioso en el proceso de investigación y desarrollo, ya que permite realizar estudios preliminares, evaluar interacciones proteína-ligando, diseñar fármacos y realizar estudios de simulación sin la necesidad de obtener la estructura experimentalmente (Schmidt et al., 2014). Aunque se disponga de la secuencia de la proteína, sigue siendo un problema aún no resuelto esperar una predicción *de novo* precisa de las estructuras 3D correspondientes (Fang et al., 2019). Debido a la presencia de características avanzadas, las técnicas DL se aplican a la predicción de la estructura secundaria, la información topológica, la determinación de los ángulos de torsión con respecto al eje de la hélice, y los contactos entre residuos. Otros métodos también se utilizaron para combinar las estructuras unidimensionales y bidimensionales, y las CNN se utilizaron para predecir los

contactos entre residuos. Las características avanzadas de DL pueden estudiar con precisión la asociación entre la secuencia y la estructura de la proteína a través del enfoque de extracción de características. Actualmente, sigue siendo un objetivo predecir con precisión la estructura 3D, y el DL parece mostrar un punto culminante del desarrollo futuro. El modelado homológico y los métodos *de novo* se utilizan tradicionalmente para predecir la estructura 3D de proteínas diana y se han vuelto más precisos y sofisticados (Baker & Sali, 2001). Sin embargo, la reciente valoración de la estructura de proteínas con herramientas de AI se desarrolla y realiza para predecir la estructura 3D de la proteína. Por ejemplo, la herramienta de AI AlphaFold se utiliza para la estructura 3D de la proteína diana, que predice con precisión unas 25 estructuras de 45 estructuras de datos (Selvaraj et al., 2021).

4.3.2. LBVS:

Los métodos de cribado virtual basados en ligandos (LBVS) se aplican cuando se desconoce o se carece de información sobre la estructura 3D de la diana proteica farmacológica (Selvaraj, et al., 2021). Se opone al método de acoplamiento disponible (*Docking*) predeterminando la orientación de unión del ligando, los volúmenes excluidos y los requisitos de espacio de unión de la diana farmacológica (Suryanarayanan et al., 2013). El principio básico del LBVS es que moléculas líderes estructuralmente similares a compuestos activos conocidos tienen una actividad de referencia similar. En el LBVS, se requiere información de los compuestos conocidos, y basándose en ella, se predice el conjunto de moléculas activas a partir del conjunto de moléculas de prueba en los ensayos funcionales, incluso sin información de la diana de la estructura proteica (Ferreira et al., 2015). En este caso, los métodos LBVS se utilizan eficazmente para encontrar moléculas líderes de gran potencial evaluando las estructuras similares a las moléculas activas. Recientemente, se ha implementado el ML en los métodos LBVS para encontrar dichas moléculas y aumentar la capacidad predictiva de los modelos LBVS (Carpenter & Huang, 2018). El objetivo crucial de los métodos ML es avanzar en la predicción de las moléculas activas frente a una diana proteica específica utilizando el conjunto de datos entrenados de entrada que discrimina las moléculas líderes de una enorme base de datos de compuestos no farmacológicos. Para superar estos problemas, los investigadores han aplicado los modelos SVM, la arquitectura bayesiana y las ANNs. Por ejemplo, (Stokes et al., n.d.) han identificado con éxito varios antibióticos nuevos utilizando redes

convolucionales gráficas (GCN), realizadas por modelos ML en la predicción de propiedades moleculares (Vatansever et al., 2021). Los investigadores han llevado a cabo un HTVS a gran escala de datos aplicando el modelo GCN y han identificado un nuevo antibiótico prometedor, la halicina. La potencia de los métodos DL en los enfoques de descubrimiento de fármacos ha dado lugar a técnicas extraordinarias para encontrar la nueva molécula líder con DL y la plataforma Spark-H2O Python para LBVS (Vink et al., 2021). En los métodos basados en ligandos, se aplican modelos QSAR para comprender los valores estadísticos de las propiedades fisicoquímicas de las moléculas pequeñas representadas por descriptores moleculares, junto con su actividad biológica. Estos modelos QSAR desempeñan un papel crucial en la optimización de moléculas pequeñas. Además, ofrecen una evaluación teórica primaria de los rasgos característicos esenciales relacionados con la actividad inhibidora, la selectividad de unión y la toxicidad de las moléculas líderes. El enfoque QSAR reduce drásticamente el número de moléculas piloto que deben someterse a experimentos *in vitro* e *in vivo* (T. Wang et al., 2017). Este método puede evaluarse basándose en un modelo de regresión o clasificación que depende de la estrategia computacional. La implementación del enfoque AI/ML en el modelo QSAR se ha utilizado ampliamente en los últimos años. En 2012, se desarrolló un método de vanguardia con múltiples DNN para predecir la precisión en un 15% sobre el método RF de referencia. Desde que se implementó en el modelo QSAR, este método basado en RF se utilizó para el descubrimiento de fármacos (Meftahi et al., 2018). Recientemente, (Zakharov et al., 2016) desarrollaron un modelo QSAR combinado con DNNs multitarea para modelar la predicción QSAR a gran escala para mejorar la precisión y la predicción sobre el concepto de modelos QSAR. Los enfoques ML con integración de conjuntos se combinan con varios modelos básicos disponibles para superar la debilidad de cada modelo de aprendizaje y tienen como objetivo mejorar la tarea de rendimiento de QSAR.

4.4. Modelos AI para el diseño ‘*de novo*’.

Se han introducido varios programas informáticos y metodologías en el diseño de fármacos ‘*de novo*’ para generar nuevas moléculas potentes sin la información de los compuestos de referencia. Desgraciadamente, estos métodos ‘*de novo*’ no son de aplicación generalizada en comparación con otros métodos de cribado basados en estructuras. En este método, se llega a compuestos difíciles de sintetizar. El auto

codificador variacional (Variational Autoencoder, VAE), tipo de modelo generativo en el campo del ML, tiene dos NN: la red codificadora y la red decodificadora (Lin et al., 2020). La red codificadora traduce activamente la estructura química de los compuestos en su notación SMILE a un vector continuo de valor real. El proceso de retrotraducción de la molécula (proceso dominante) prevalece sobre los cambios conformacionales ligeros, ya que estos últimos son poco probables o despreciables en la estructura química del compuesto. En otro estudio, se comparó el rendimiento del autocodificador de variaciones con el autocodificador adversarial (Grisoni et al., 2020). Estos decodificadores adversariales tienen un modelo generativo para la producción de estructuras químicas novedosas (Pogány et al., 2019). Por ejemplo, la predicción de estructuras novedosas en combinación con el modelo *in silico* muestra compuestos más activos contra el receptor de dopamina tipo 2. De forma similar, (Kadurin et al., 2017) utilizaron una red generativa adversarial (GAN) y sugirieron compuestos con propiedades anticancerígenas más potentes y eficaces (Kell et al., 2020).

4.5. Modelos AI para la predicción de interacciones proteína-proteína:

La evaluación de las interacciones proteína-proteína (PPI) es funcionalmente esencial para varios procesos y enfermedades biológicos. STRING es una de las bases de datos de PPI más utilizadas, con más 1.000 millones de PPI (Protein-Protein Interaction) obtenidas mediante métodos computacionales y experimentales. La interfaz del PPI se predice mediante sitios de unión proteína-proteína con un alto número de aminoácidos (Rao et al., 2014). Puede plantearse como una nueva clase de dianas farmacológicas tradicionales, incluyendo los GPCR (G Protein-Coupled Receptor), es decir, proteínas de membrana que actúan como receptores en la superficie celular, receptores nucleares, canales iónicos y quinasas. Por ejemplo, en la base de datos proteína-proteína (iPPI-DB) se han registrado más de 1700 inhibidores no peptídicos que actúan bloqueando uniones proteína-proteína (Legrain & Selig, 2000). Esto amplía el espacio diana y fomenta el desarrollo de pequeñas moléculas inhibitoras potentes. Como ejemplo, el servidor web eFindSite, predice la interfaz de interacción proteína-proteína basándose en las plantillas. Se utilizan métodos de AI como los modelos SVM y NBS para predecir las características basadas en residuos y secuencias. Basándose en este principio, se pueden utilizar varios métodos de acoplamiento proteína-proteína (ZDOCK, SymmDock) para predecir la interfaz de las

PPI (Ding & Kihara, 2018). La predicción de los cambios conformacionales cuando dos proteínas se unen es un gran reto debido a la complejidad estructural y dinámica involucrada en la formación de la interfaz de interacción. El desafío radica en predecir con precisión estos cambios conformacionales en el contexto de la PPI. La estructura de una proteína en su forma aislada puede ser diferente de su estructura cuando está unida a otra proteína. Estos cambios pueden ser sutiles o drásticos, y capturarlos de manera precisa y confiable es un desafío en el campo del acoplamiento proteína-proteína. Por ello, los métodos DL extraen eficazmente las características de secuencia más relevantes para la predicción de la interfaz PPI, lo que muestra mejoras en comparación con otros métodos ML como SVM. Es crucial identificar y predecir con precisión las regiones de la interfaz donde ocurre la unión entre las proteínas. Estas regiones de unión se denominan 'sitios de unión' o 'puntos calientes'. Los puntos calientes son aquellos residuos o regiones específicas en las proteínas que contribuyen de manera significativa a la estabilidad y especificidad de la interacción proteína-proteína. La identificación de estos puntos calientes en la interfaz de la PPI es importante porque pueden representar posibles dianas terapéuticas para el diseño de fármacos o terapias, ya que estos puntos calientes son áreas de la interfaz que contribuyen en gran medida a la afinidad y la estabilidad de la interacción, y dirigir fármacos a estos puntos calientes puede interferir con la interacción proteína-proteína y tener un efecto terapéutico (Gao & Coley, 2020).

5. Conclusión:

En los últimos años, la AI se ha utilizado con frecuencia en las industrias farmacéuticas y biomédicas. Estas industrias adoptan en gran medida varias herramientas basadas en AI para llevar a cabo procesos más eficientes y automatizados que integran decisiones predictivas y basadas en datos. Proporciona una cantidad significativa de bases de datos que se emplean para sucesivos modelos entrenados. Los datos deben ser fiables y disponibles en gran cantidad para garantizar el análisis farmacéutico. Por otro lado, la falta de personas calificadas para manejar las plataformas basadas en AI es otro reto que impide la adopción plena de la AI en las industrias farmacéutica y biomédica. Con el análisis enfocado en la AI aplicada a la industria farmacéutica que se ha llevado a cabo en este trabajo, se podría concluir que la AI:

- Analiza grandes cantidades de datos y buscar patrones que puedan ayudar a identificar nuevos medicamentos.
- Analiza cómo los compuestos químicos interactúan con las dianas terapéuticas del cuerpo humano.
- Analiza los procesos de producción y hacen sugerencias para mejorar la eficiencia y reducir los costos.
- Desarrollar fármacos desde cero sin necesidad de material de referencia.
- Permite la combinación de sus herramientas hasta horizontes ilimitados en el campo farmacéutico para impulsar avances significativos en la investigación y desarrollo de fármacos.

6. Bibliografía:

- Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J Pharm Biomed Anal*, 22(5), 717–727. [https://doi.org/10.1016/s0731-7085\(99\)00272-1](https://doi.org/10.1016/s0731-7085(99)00272-1)
- Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, 2020. <https://doi.org/10.1093/database/baaa010>
- Alzahab, N. A., Apollonio, L., Di Iorio, A., Alshalak, M., Iarlori, S., Ferracuti, F., Monteriù, A., & Porcaro, C. (2021). Hybrid Deep Learning (hDL)-Based Brain-Computer Interface (BCI) Systems: A Systematic Review. *Brain Sciences*, 11(1), 75. <https://doi.org/10.3390/brainsci11010075>
- Aminpour, M., Montemagno, C., & Tuszynski, J. A. (2019). An Overview of Molecular Modeling for Drug Discovery with Specific Illustrative Examples of Applications. *Molecules*, 24(9), 1693. <https://doi.org/10.3390/molecules24091693>
- Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., & Khan, M. K. (2018). Medical Image Analysis using Convolutional Neural Networks: A Review. *Journal of Medical Systems*, 42(11), 1–13. <https://doi.org/10.1007/S10916-018-1088-1/METRICS>
- Aparoy, P., Kumar Reddy, K., & Reddanna, P. (2012). Structure and Ligand Based Drug Design Strategies in the Development of Novel 5- LOX Inhibitors. *Current Medicinal Chemistry*, 19(22), 3763–3778. <https://doi.org/10.2174/092986712801661112>
- Arabfard, M., Ohadi, M., Rezaei Tabar, V., Delbari, A., & Kavousi, K. (2019). Genome-wide prediction and prioritization of human aging genes by data fusion: a machine learning approach. *BMC Genom*, 20(1), 832. <https://doi.org/10.1186/s12864-019-6140-0>
- Aristotelous, T., Ahn, S., Shukla, A. K., Gawron, S., Sassano, M. F., Kahsai, A. W., Wingler, L. M., Zhu, X., Tripathi-Shukla, P., Huang, X. P., Riley, J., Besnard, J., Read, K. D., Roth, B. L., Gilbert, I. H., Hopkins, A. L., Lefkowitz, R. J., & Navratilova, I. (2013). Discovery of beta2 adrenergic receptor ligands using biosensor fragment screening of tagged wild-type receptor. *ACS Med Chem Lett*, 4(10), 1005–1010. <https://doi.org/10.1021/ml400312j>
- Bahi, M., & Batouche, M. (2018). Deep Learning for Ligand-Based Virtual Screening in Drug Discovery. *Proceedings - PAIS 2018: International Conference on Pattern Analysis and Intelligent Systems*. <https://doi.org/10.1109/PAIS.2018.8598488>
- Baker, D., & Sali, A. (2001). Protein Structure Prediction and Structural Genomics. *Science*, 294(5540), 93–96. <https://doi.org/10.1126/science.1065659>
- Baranilingesan, I. (2021). Optimization algorithm based Elman neural network controller for continuous stirred tank reactor process model. *Curr Sci.*, 120(8), 1324–1333. <https://doi.org/10.18520/cs/v120/i8/1324-1333>
- Ben-Bassat, I., Chor, B., & Orenstein, Y. (2018). A deep neural network approach for learning intrinsic protein-RNA binding preferences. *Bioinformatics*, 34(17), i638–i646. <https://doi.org/10.1093/bioinformatics/bty600>

- Brereton, R. G. (2012). Self organising maps for visualising and modelling. *Chem Cent J*, 6(Suppl 2), S1. <https://doi.org/10.1186/1752-153x-6-s2-s1>
- Bruno, A., Costantino, G., Sartori, L., & Radi, M. (2019). The In Silico Drug Discovery Toolbox: Applications in Lead Discovery and Optimization. *Current Medicinal Chemistry*, 26(21), 3838–3873. <https://doi.org/10.2174/0929867324666171107101035>
- Burbidge, R., Trotter, M., Buxton, B., & Holden, S. (2001). Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput Chem*, 26(1), 5–14. [https://doi.org/10.1016/s0097-8485\(01\)00094-8](https://doi.org/10.1016/s0097-8485(01)00094-8)
- Bzdok, D., Krzywinski, M., & Altman, N. (2018). Points of significance: Machine learning: Supervised methods. *Nature Methods*, 15(1), 5–6. <https://doi.org/10.1038/NMETH.4551>
- Carpenter, K. A., & Huang, X. (2018). Machine Learning-based Virtual Screening and Its Applications to Alzheimer’s Drug Discovery: A Review. *Current Pharmaceutical Design*, 24(28), 3347–3358. <https://doi.org/10.2174/1381612824666180607124038>
- Castillo-Garit, J. A., Marrero-Ponce, Y., Torrens, F., & García-Domenech, R. (2008). Estimation of ADME Properties in Drug Discovery: Predicting Caco-2 Cell Permeability Using Atom-Based Stochastic and Non-stochastic Linear Indices. *Journal of Pharmaceutical Sciences*, 97(5), 1946–1976. <https://doi.org/10.1002/jps.21122>
- Chen, S., Liu, T., Xu, D., Huo, Y., & Yang, Y. (2019). Image based Measurement of Population Growth Rate for L-Glutamic Acid Crystallization. *2019 Chinese Control Conference (CCC)*, 2019-July, 7933–7938. <https://doi.org/10.23919/CHICC.2019.8866441>
- Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R., & Ierapetritou, M. (2020). Digital twins in pharmaceutical and biopharmaceutical manufacturing: a literature review. *Processes*, 8(9), 1088. <https://doi.org/10.3390/pr8091088>
- Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz’Min, V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., & Tropsha, A. (2014). QSAR modeling: where have you been? Where are you going to? *J Med Chem*, 57(12), 4977–5010. <https://doi.org/10.1021/jm4004285>
- Coley, C. W., Barzilay, R., Green, W. H., Jaakkola, T. S., & Jensen, K. F. (2017). Physical Property Prediction. *Journal of Chemical Information and Modeling*, 57, 1757–1772. <https://doi.org/10.1021/acs.jcim.6b00601>
- Costa, P. R., Acencio, M. L., & Lemke, N. (2010). A machine learning approach for genome-wide prediction of morbid and druggable human genes based on systems-level data. *BMC Genom*, 11(Suppl 5), S9. <https://doi.org/10.1186/1471-2164-11-s5-s9>
- Cova, T. F. G. G., & Pais, A. A. C. C. (2019). Deep learning for deep chemistry: optimizing the prediction of chemical patterns. *Front Chem*, 7, 809. <https://doi.org/10.3389/fchem.2019.00809>

- Del Fiol, G., Michelson, M., Iorio, A., Cotoi, C., & Haynes, R. B. (2018). A Deep Learning Method to Automatically Identify Reports of Scientifically Rigorous Clinical Research from the Biomedical Literature: Comparative Analytic Study. *Journal of Medical Internet Research*, *20*(6), e10281. <https://doi.org/10.2196/10281>
- Deng, Y., Bao, F., Dai, Q., Wu, L. F., & Altschuler, S. J. (2019). Scalable analysis of cell-type composition from single-cell transcriptomics using deep recurrent learning. *Nature Methods*, *16*(4), 311–314. <https://doi.org/10.1038/S41592-019-0353-7>
- Ding, Z., & Kihara, D. (2018). Computational Methods for Predicting Protein-Protein Interactions Using Various Protein Features. *Current Protocols in Protein Science*, *93*(1). <https://doi.org/10.1002/cpps.62>
- Doane, A. S., & Elemento, O. (2017). Regulatory elements in molecular networks. *Wiley Interdiscip Rev Syst Biol Med*, *9*(3). <https://doi.org/10.1002/wsbm.1374>
- Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., & Adams, R. P. (n.d.). *Convolutional Networks on Graphs for Learning Molecular Fingerprints*.
- Ekins, S., Mestres, J., & Testa, B. (2007). *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British Journal of Pharmacology*, *152*(1), 9–20. <https://doi.org/10.1038/sj.bjp.0707305>
- Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. *Radiographics*, *37*(2), 505–515. <https://doi.org/10.1148/RG.2017160130>
- Fang, C., Shang, Y., & Xu, D. (2019). Prediction of Protein Backbone Torsion Angles Using Deep Residual Inception Neural Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *16*(3), 1020–1028. <https://doi.org/10.1109/TCBB.2018.2814586>
- Ferreira, L., dos Santos, R., Oliva, G., & Andricopulo, A. (2015). Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*, *20*(7), 13384–13421. <https://doi.org/10.3390/molecules200713384>
- Ferrero, E., Dunham, I., & Sanseau, P. (2017). In silico prediction of novel therapeutic targets using gene-disease association data. *J Transl Med*, *15*(1), 182. <https://doi.org/10.1186/s12967-017-1285-6>
- Foster, K. R., Koprowski, R., & Skufca, J. D. (2014). Machine learning, medical diagnosis, and biomedical engineering research - commentary. *BioMedical Engineering Online*, *13*(1). <https://doi.org/10.1186/1475-925X-13-94>
- Franco, V. G., Perín, J. C., Mantovani, V. E., & Goicoechea, H. C. (2006). Monitoring substrate and products in a bioprocess with FTIR spectroscopy coupled to artificial neural networks enhanced with a genetic-algorithm-based method for wavelength selection. *Talanta*, *68*(3), 1005–1012. <https://doi.org/10.1016/j.talanta.2005.07.003>
- Franke, L., Byvatov, E., Werz, O., Steinhilber, D., Schneider, P., & Schneider, G. (2005). Extraction and visualization of potential pharmacophore points using support vector machines: Application to ligand-based virtual screening for COX-2 inhibitors. *Journal of Medicinal Chemistry*, *48*(22), 6997–7004. <https://doi.org/10.1021/JM050619H/ASSET/IMAGES/MEDIUM/JM050619HN0001.GIF>

- Freedman, J. D., Hagel, J., Scott, E. M., Psallidas, I., Gupta, A., Spiers, L., Miller, P., Kanellakis, N., Ashfield, R., Fisher, K. D., Duffy, M. R., & Seymour, L. W. (2017). Oncolytic adenovirus expressing bispecific antibody targets T-cell cytotoxicity in cancer biopsies. *EMBO Molecular Medicine*, 9(8), 1067–1087. <https://doi.org/10.15252/emmm.201707567>
- Gao, W., & Coley, C. W. (2020). The Synthesizability of Molecules Proposed by Generative Models. *Journal of Chemical Information and Modeling*, 60(12), 5714–5723. <https://doi.org/10.1021/acs.jcim.0c00174>
- Garscha, U., Voelker, S., Pace, S., Gerstmeier, J., Emini, B., Liening, S., Rossi, A., Weinigel, C., Rummler, S., Schubert, U. S., Scriba, G. K. E., Çelikoğlu, E., Çalışkan, B., Banoglu, E., Sautebin, L., & Werz, O. (2016). BRP-187: a potent inhibitor of leukotriene biosynthesis that acts through impeding the dynamic 5-lipoxygenase/5-lipoxygenase-activating protein (FLAP) complex assembly. *Biochem Pharmacol*, 119, 17–26. <https://doi.org/10.1016/j.bcp.2016.08.023>
- Geppert, H., Horváth, T., Gärtner, T., Wrobel, S., & Bajorath, J. (2008). Support-vector-machine-based ranking significantly improves the effectiveness of similarity searching using 2D fingerprints and multiple reference compounds. *J Chem Inf Model*, 48(4), 742–746. <https://doi.org/10.1021/ci700461s>
- Goodfellow, I., Bengio, Y., & Courville, A. (n.d.). *Deep Learning*.
- Graupe, D., & Vern, B. (2001). On the inter-relations between artificial and physiological neural networks. *Neurol Res*, 23(5), 482–488. <https://doi.org/10.1179/016164101101198875>
- Grisoni, F., Moret, M., Lingwood, R., & Schneider, G. (2020). Bidirectional Molecule Generation with Recurrent Neural Networks. *Journal of Chemical Information and Modeling*, 60(3), 1175–1183. <https://doi.org/10.1021/acs.jcim.9b00943>
- Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism: Clinical and Experimental*, 69, S36–S40. <https://doi.org/10.1016/J.METABOL.2017.01.011>
- Hasani, M., & Moloudi, M. (2008). Application of principal component-artificial neural network models for simultaneous determination of phenolic compounds by a kinetic spectrophotometric method. *J Hazard Mater.*, 157(1), 161–169. <https://doi.org/10.1016/j.jhazmat.2007.12.096>
- Henry, J., & Wlodkovic, D. (2019). Towards High-Throughput Chemobehavioural Phenomics in Neuropsychiatric Drug Discovery. *Marine Drugs*, 17(6), 340. <https://doi.org/10.3390/md17060340>
- Huang, L., Massa, L., & Karle, J. (2007). Kernel Energy Method: The Interaction Energy of the Collagen Triple Helix. *Journal of Chemical Theory and Computation*, 3(4), 1337–1341. <https://doi.org/10.1021/ct7000649>
- Inghelbrecht, S., Remon, J. P., Fernandes De Aguiar, P., Walczak, B., Massart, D. L., Van De Velde, F., De Baets, P., Vermeersch, H., & De Backer, P. (1997). Instrumentation of a roll compactor and the evaluation of the parameter settings by neural networks. *Int J Pharm.*, 148(1), 103–115. [https://doi.org/10.1016/s0378-5173\(94\)04837-1](https://doi.org/10.1016/s0378-5173(94)04837-1)

- International Conference on Intelligent Computing and Internet of Things 2015 Harbin, Ha er bin shi fan da xue, Institute of Electrical and Electronics Engineers, International Conference on Intelligent Computing and Internet of Things 2015.01.17-18 Harbin, & ICIT 2015.01.17-18 Harbin. (2015). *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things ICIT 2015 : January 17-18, 2015, Harbin, China*.
- Jafari, M., Wang, Y., Amiryousefi, A., & Tang, J. (2020). Unsupervised Learning and Multipartite Network Models: A Promising Approach for Understanding Traditional Medicine. *Frontiers in Pharmacology*, *11*, 563852. <https://doi.org/10.3389/FPHAR.2020.01319/BIBTEX>
- Jeon, J., Nim, S., Teyra, J., Datti, A., Wrana, J. L., Sidhu, S. S., Moffat, J., & Kim, P. M. (2014). A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. *Genome Med*, *6*(7), 57. <https://doi.org/10.1186/s13073-014-0057-7>
- Kachrimanis, K., Karamyan, V., & Malamataris, S. (2003). Artificial neural networks (ANNs) and modeling of powder flow. *Int J Pharm.*, *250*(1), 13–23. [https://doi.org/10.1016/s0378-5173\(02\)00528-8](https://doi.org/10.1016/s0378-5173(02)00528-8)
- Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., & Zhavoronkov, A. (2017). druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. *Molecular Pharmaceutics*, *14*(9), 3098–3104. <https://doi.org/10.1021/acs.molpharmaceut.7b00346>
- Kazemi, P., Khalid, M. H., Gago, A. P., Kleinebudde, P., Jachowicz, R., Szlęk, J., & Mendyk, A. (2017). Effect of roll compaction on granule size distribution of microcrystalline cellulose-mannitol mixtures: computational intelligence modeling and parametric analysis. *Drug Des Devel Ther.*, *11*, 241–251. <https://doi.org/10.2147/dddt.s124670>
- Kell, D. B., Samanta, S., & Swainston, N. (2020). Deep learning and generative methods in cheminformatics and chemical biology: navigating small molecule space intelligently. *Biochemical Journal*, *477*(23), 4559–4580. <https://doi.org/10.1042/BCJ20200781>
- Kesavan, J. G., & Peck, G. E. (1996). Pharmaceutical granulation and tablet formulation using neural networks. *Pharm Dev Technol.*, *1*(4), 391–404. <https://doi.org/10.3109/10837459609031434>
- Ko, W. Y., Siontis, K. C., Attia, Z. I., Carter, R. E., Kapa, S., Ommen, S. R., Demuth, S. J., Ackerman, M. J., Gersh, B. J., Arruda-Olson, A. M., Geske, J. B., Asirvatham, S. J., Lopez-Jimenez, F., Nishimura, R. A., Friedman, P. A., & Noseworthy, P. A. (2020). Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram. *Journal of the American College of Cardiology*, *75*(7), 722–733. <https://doi.org/10.1016/J.JACC.2019.12.030>
- Koohy, H. (2017). The rise and fall of machine learning methods in biomedical research. *F1000Research*, *6*. <https://doi.org/10.12688/F1000RESEARCH.13016.2>
- Korteby, Y., Kristó, K., Sovány, T., & Regdon, G. (2018). Use of machine learning tool to elucidate and characterize the growth mechanism of an in-situ fluid bed melt

- granulation. *Powder Technol.*, 331, 286–295. <https://doi.org/10.1016/j.powtec.2018.03.052>
- Korteby, Y., Mahdi, Y., Azizou, A., Daoud, K., & Regdon, G. (2016). Implementation of an artificial neural network as a PAT tool for the prediction of temperature distribution within a pharmaceutical fluidized bed granulator. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences*, 88, 219–232. <https://doi.org/10.1016/J.EJPS.2016.03.010>
- Kriegeskorte, N., & Golan, T. (2019). Neural network models and deep learning. *Current Biology*, 29(7), R231–R236. <https://doi.org/10.1016/J.CUB.2019.02.034>
- Kubař, T., & Elstner, M. (2013). A hybrid approach to simulation of electron transfer in complex molecular systems. *Journal of The Royal Society Interface*, 10(87), 20130415. <https://doi.org/10.1098/rsif.2013.0415>
- Kumar Ananthu, M., Kumar Chintamaneni, P., Basha Shaik, S., Thadipatri, R., & Mahammed, N. (2021). Artificial Neural Networks in Optimization of Pharmaceutical Formulations. *Saudi J Med Pharm Sci*, 7(8), 368–378. <https://doi.org/10.36348/sjmps.2021.v07i08.004>
- Lavecchia, A., & Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Curr Med Chem*, 20(23), 2839–2860. <https://doi.org/10.2174/09298673113209990001>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/NATURE14539>
- Lee, J. A., & Berg, E. L. (2013). Neoclassic drug discovery: the case for lead generation using phenotypic and functional approaches. *J Biomol Screen*, 18(10), 1143–1155. <https://doi.org/10.1177/10870571113506118>
- Legrain, P., & Selig, L. (2000). Genome-wide protein interaction maps using two-hybrid systems. *FEBS Letters*, 480(1), 32–36. [https://doi.org/10.1016/S0014-5793\(00\)01774-9](https://doi.org/10.1016/S0014-5793(00)01774-9)
- Li, Q., & Lai, L. (2007). Prediction of potential drug targets based on simple sequence properties. *BMC Bioinformatics*, 8(1), 353. <https://doi.org/10.1186/1471-2105-8-353>
- Li, S., Deng, Y.-Q., Zhu, Z.-L., Hua, H.-L., & Tao, Z.-Z. (2021). A Comprehensive Review on Radiomics and Deep Learning for Nasopharyngeal Carcinoma Imaging. *Diagnostics*, 11(9), 1523. <https://doi.org/10.3390/diagnostics11091523>
- Li, Y., Ma, X., Zhou, X., Cheng, P., He, K., & Li, C. (2021). Knowledge enhanced LSTM for coreference resolution on biomedical texts. *Bioinformatics*, 37(17), 2699–2705. <https://doi.org/10.1093/BIOINFORMATICS/BTAB153>
- Lima, A. N., Philot, E. A., Trossini, G. H. G., Scott, L. P. B., Maltarollo, V. G., & Honorio, K. M. (2016). Use of machine learning approaches for novel drug discovery. *Expert Opin Drug Discov*, 11(3), 225–239. <https://doi.org/10.1517/17460441.2016.1146250>
- Lin, E., Lin, C.-H., & Lane, H.-Y. (2020). Relevant Applications of Generative Adversarial Networks in Drug Design and Discovery: Molecular De Novo Design, Dimensionality Reduction, and De Novo Peptide and Protein Design. *Molecules*, 25(14), 3250. <https://doi.org/10.3390/molecules25143250>

- Lind, A. P., & Anderson, P. C. (2019). Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. *PLoS ONE*, *14*(7), e0219774. <https://doi.org/10.1371/journal.pone.0219774>
- Liu, Z.-P. (2015). Reverse engineering of genome-wide gene regulatory networks from gene expression data. *Curr Genom*, *16*(1), 3–22. <https://doi.org/10.2174/1389202915666141110210634>
- Ljosa, V., Caie, P. D., Ter Horst, R., Sokolnicki, K. L., Jenkins, E. L., Daya, S., Roberts, M. E., Jones, T. R., Singh, S., Genovesio, A., Clemons, P. A., Carragher, N. O., & Carpenter, A. E. (2013). Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *J Biomol Screen*, *18*(10), 1321–1329. <https://doi.org/10.1177/1087057113503553>
- Maltarollo, V. G., Kronenberger, T., Espinoza, G. Z., Oliveira, P. R., & Honorio, K. M. (2019). Advances with support vector machines for novel drug discovery. *Expert Opin Drug Discov*, *14*(1), 23–33. <https://doi.org/10.1080/17460441.2019.1549033>
- 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. *Front Genet*, *9*(JUL), 242. <https://doi.org/10.3389/fgene.2018.00242>
- Manisha, Dhull, S. K., & Singh, K. K. (2020). ECG Beat Classifiers: A Journey from ANN to DNN. *Procedia Computer Science*, *167*, 747–759. <https://doi.org/10.1016/J.PROCS.2020.03.340>
- Matlock, M. K., Hughes, T. B., & Swamidass, S. J. (2015). XenoSite server: A web-available site of metabolism prediction tool. *Bioinformatics*, *31*(7), 1136–1137. <https://doi.org/10.1093/bioinformatics/btu761>
- Mayr, A., Klambauer, G., Unterthiner, T., & Hochreiter, S. (2016). DeepTox: Toxicity prediction using deep learning. *Frontiers in Environmental Science*, *3*(FEB). <https://doi.org/10.3389/FENV.2015.00080/FULL>
- Meftahi, N., Walker, M. L., Enciso, M., & Smith, B. J. (2018). Predicting the Enthalpy and Gibbs Energy of Sublimation by QSPR Modeling. *Scientific Reports*, *8*(1), 9779. <https://doi.org/10.1038/s41598-018-28105-6>
- Millen, N., Kovačević, A., Khera, L., Djuriš, J., & Ibrić, S. (2019). Machine learning modelling of wet granulation scale-up using compressibility, compactibility and manufacturability parameters. *Hemijaska Industrija.*, *73*(3), 155–168. <https://doi.org/10.2298/hemind190412017m>
- Mujumdar, A., Robi, P., ... M. M.-I. J. for, & 2007, undefined. (2007). Artificial neural network (ANN) model for prediction of mixing behavior of granular flows. *Taylor & Francis*, *8*(3), 149–158. <https://doi.org/10.1080/15502280701252495>
- 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>
- Nagy, B., Galata, D. L., Farkas, A., & Nagy, Z. K. (2022). Application of Artificial Neural Networks in the Process Analytical Technology of Pharmaceutical Manufacturing—

- a Review. *The AAPS Journal* 2022 24:4, 24(4), 1–18. <https://doi.org/10.1208/S12248-022-00706-0>
- Närhi, M., Salmela, L., Toivonen, J., Billet, C., Dudley, J. M., & Genty, G. (2018). Machine learning analysis of extreme events in optical fibre modulation instability. *Nat Commun*, 9(1), 4923. <https://doi.org/10.1038/s41467-018-07355-y>
- Paengjuntuek, W., Thanasinthana, L., & Arpornwichanop, A. (2012). Neural network-based optimal control of a batch crystallizer. *Neurocomputing.*, 83, 158–164. <https://doi.org/10.1016/j.neucom.2011.12.008>
- Park, E. J., Myint, P. K., Ito, A., Appiah, M. G., Darkwah, S., Kawamoto, E., & Shimaoka, M. (2020). Integrin-Ligand Interactions in Inflammation, Cancer, and Metabolic Disease: Insights Into the Multifaceted Roles of an Emerging Ligand Irisin. *Frontiers in Cell and Developmental Biology*, 8, 588066. <https://doi.org/10.3389/FCELL.2020.588066/BIBTEX>
- Patidar, K., Deshmukh, A., Bandaru, S., Lakkaraju, C., Girdhar, A., Gutlapalli, V. R., Banerjee, T., Nayariseri, A., & Singh, S. K. (2016). Virtual screening approaches in identification of bioactive compounds Akin to delphinidin as potential HER2 inhibitors for the treatment of breast cancer. *Asian Pac J Cancer Prev*, 17(4), 2291–2295. <https://doi.org/10.7314/apjcp.2016.17.4.2291>
- Pesapane, F., Tantrige, P., Patella, F., Biondetti, P., Nicosia, L., Ianniello, A., Rossi, U. G., Carrafiello, G., & Ierardi, A. M. (2020). Myths and facts about artificial intelligence: why machine- and deep-learning will not replace interventional radiologists. *Medical Oncology*, 37(5), 40. <https://doi.org/10.1007/s12032-020-01368-8>
- Petrović, J., Chansanroj, K., Meier, B., Ibrić, S., & Betz, G. (2011). Analysis of fluidized bed granulation process using conventional and novel modeling techniques. *Eur J Pharm Sci.*, 44(3), 227–234. <https://doi.org/10.1016/j.ejps.2011.07.013>
- Pogány, P., Arad, N., Genway, S., & Pickett, S. D. (2019). De Novo Molecule Design by Translating from Reduced Graphs to SMILES. *Journal of Chemical Information and Modeling*, 59(3), 1136–1146. <https://doi.org/10.1021/acs.jcim.8b00626>
- Poggio, T., Banburski, A., & Liao, Q. (2020). Theoretical issues in deep networks. *Proceedings of the National Academy of Sciences*, 117(48), 30039–30045. <https://doi.org/10.1073/pnas.1907369117>
- Qi, Y., Bar-Joseph, Z., & Klein-Seetharaman, J. (2006). Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins*, 63(3), 490–500. <https://doi.org/10.1002/prot.20865>
- Rao, V. S., Srinivas, K., Sujini, G. N., & Kumar, G. N. S. (2014). Protein-Protein Interaction Detection: Methods and Analysis. *International Journal of Proteomics*, 2014, 1–12. <https://doi.org/10.1155/2014/147648>
- Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J., & Sanner, M. F. (2015). AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS Comput Biol*, 11(12), e1004586. <https://doi.org/10.1371/journal.pcbi.1004586>

- Reddy, K. K., & Singh, S. K. (2014). Combined ligand and structure-based approaches on HIV-1 integrase strand transfer inhibitors. *Chem Biol Interact*, *218*, 71–81. <https://doi.org/10.1016/j.cbi.2014.04.011>
- Rodrigues, T., Werner, M., Roth, J., Da Cruz, E. H. G., Marques, M. C., Akkapeddi, P., Lobo, S. A., Koeberle, A., Corzana, F., Da Silva Júnior, E. N., Werz, O., & Bernardes, G. J. L. (2018). Machine intelligence decrypts beta-lapachone as an allosteric 5-lipoxygenase inhibitor. *Chem Sci*, *9*(34), 6899–6903. <https://doi.org/10.1039/c8sc02634c>
- Ruan, Z., Liu, C., Guo, Y., He, Z., Huang, X., Jia, X., & Yang, T. (2021). SARS-CoV-2 and SARS-CoV: Virtual screening of potential inhibitors targeting RNA-dependent RNA polymerase activity (NSP12). *Journal of Medical Virology*, *93*(1), 389. <https://doi.org/10.1002/JMV.26222>
- Sahoo, M., Jena, L., Daf, S., & Kumar, S. (2016). Virtual Screening for Potential Inhibitors of NS3 Protein of Zika Virus. *Genomics & Informatics*, *14*(3), 104. <https://doi.org/10.5808/GI.2016.14.3.104>
- Sajjia, M., Shirazian, S., Kelly, C. B., Albadarin, A. B., & Walker, G. (2017). ANN Analysis of a roller compaction process in the pharmaceutical industry. *Chem Eng Technol.*, *40*(3), 487–492. <https://doi.org/10.1002/ceat.201600229>
- Sakellaropoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., Polyzos, A., Moss, T. J., Piha-Paul, S., Zhou, H., Kardala, E., Damianidou, E., Alexopoulos, L. G., Aifantis, I., Townsend, P. A., Panayiotidis, M. I., Sfikakis, P., Bartek, J., Fitzgerald, R. C., Thanos, D., ... Gorgoulis, V. G. (2019). A Deep Learning Framework for Predicting Response to Therapy in Cancer. *Cell Reports*, *29*(11), 3367-3373.e4. <https://doi.org/10.1016/j.celrep.2019.11.017>
- Sakkiah, S., Selvaraj, C., Gong, P., Zhang, C., Tong, W., & Hong, H. (2017). Development of estrogen receptor beta binding prediction model using large sets of chemicals. *Oncotarget*, *8*(54), 92989–93000. <https://doi.org/10.18632/oncotarget.21723>
- Salami, H., McDonald, M. A., Bommarius, A. S., Rousseau, R. W., & Grover, M. A. (2021). In situ imaging combined with deep learning for crystallization process monitoring: application to cephalixin production. *Org Process Res Dev.*, *25*(7), 1670–1679. <https://doi.org/10.1021/acs.oprd.1c00136>
- Sampat, C., & Ramachandran, R. (2021). Identification of granule growth regimes in high shear wet granulation processes using a physics-constrained neural network. *Processes.*, *9*(5), 737. <https://doi.org/10.3390/pr9050737>
- Scheeder, C., Heigwer, F., & Boutros, M. (2018). Machine learning and image-based profiling in drug discovery. *Curr Opin Syst Biol*, *10*, 43–52. <https://doi.org/10.1016/j.coisb.2018.05.004>
- Schenone, M., Dančik, V., Wagner, B. K., & Clemons, P. A. (2013). Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol*, *9*(4), 232–240. <https://doi.org/10.1038/nchembio.1199>
- Schmidt, T., Bergner, A., & Schwede, T. (2014). Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today*, *19*(7), 890–897. <https://doi.org/10.1016/j.drudis.2013.10.027>

- Selvaraj, C., Chandra, I., & Singh, S. K. (2021). Artificial intelligence and machine learning approaches for drug design: challenges and opportunities for the pharmaceutical industries. *Molecular Diversity* 2021 26:3, 26(3), 1893–1913. <https://doi.org/10.1007/S11030-021-10326-Z>
- Selvaraj, C., Dinesh, D. C., Panwar, U., Boura, E., & Singh, S. K. (2021). High-Throughput Screening and Quantum Mechanics for Identifying Potent Inhibitors Against Mac1 Domain of SARS-CoV-2 Nsp3. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(4), 1262–1270. <https://doi.org/10.1109/TCBB.2020.3037136>
- Selvaraj, C., Selvaraj, G., Mohamed Ismail, R., Vijayakumar, R., Baazeem, A., Wei, D.-Q., & Singh, S. K. (2021). Interrogation of Bacillus anthracis SrtA active site loop forming open/close lid conformations through extensive MD simulations for understanding binding selectivity of SrtA inhibitors. *Saudi Journal of Biological Sciences*, 28(7), 3650–3659. <https://doi.org/10.1016/j.sjbs.2021.05.009>
- Selvaraj, C., Vierra, M., Dinesh, D. C., Abhirami, R., & Singh, S. K. (2021). Structural insights of macromolecules involved in bacteria-induced apoptosis in the pathogenesis of human diseases. *Adv Protein Chem Struct Biol*, 126, 1–38. <https://doi.org/10.1016/bs.apcsb.2021.02.001>
- Shaikh, R., Shirazian, S., & Walker, G. M. (2021). Application of artificial neural network for prediction of particle size in pharmaceutical cocrystallization using mechanochemical synthesis. *Neural Comput Appl.*, 33(19), 12621–12640. <https://doi.org/10.1007/s00521-021-05912-z>
- Shalf, J. (2020). The future of computing beyond Moores Law. *Philosophical Transactions of the Royal Society A*, 378(2166). <https://doi.org/10.1098/RSTA.2019.0061>
- Sharma, K., Patidar, K., Ali, M. A., Patil, P., Goud, H., Hussain, T., Nayariseri, A., & Singh, S. K. (2018). Structure-based virtual screening for the identification of high affinity compounds as potent VEGFR2 inhibitors for the treatment of renal cell carcinoma. *Curr Top Med Chem*, 18(25), 2174–2185. <https://doi.org/10.2174/1568026619666181130142237>
- Singh, R. K., Lee, J.-K., Selvaraj, C., Singh, R., Li, J., Kim, S.-Y., & Kalia, V. C. (2018). Protein engineering approaches in the post-genomic era. *Curr Protein Pept Sci*, 19(1), 5–15. <https://doi.org/10.2174/1389203718666161117114243>
- Skrede, O. J., De Raedt, S., Kleppe, A., Hveem, T. S., Liestøl, K., Maddison, J., Askautrud, H. A., Pradhan, M., Nesheim, J. A., Albrechtsen, F., Farstad, I. N., Domingo, E., Church, D. N., Nesbakken, A., Shepherd, N. A., Tomlinson, I., Kerr, R., Novelli, M., Kerr, D. J., & Danielsen, H. E. (2020). Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *The Lancet*, 395(10221), 350–360. [https://doi.org/10.1016/S0140-6736\(19\)32998-8](https://doi.org/10.1016/S0140-6736(19)32998-8)
- Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol Rev*, 66(1), 334–395. <https://doi.org/10.1124/pr.112.007336>

- Šmak, P., Chandrabose, S., Tvaroška, I., & Koča, J. (2021). Pan-selectin inhibitors as potential therapeutics for COVID-19 treatment: in silico screening study. *Glycobiology*, *31*(8), 975–987. <https://doi.org/10.1093/glycob/cwab021>
- Smith, J. S., Isayev, O., & Roitberg, A. E. (2016). ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science*, *8*(4), 3192–3203. <https://doi.org/10.1039/C6SC05720A>
- Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M. M., Greenspan, H., & Klang, E. (2019). Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. *Radiology*, *290*(3), 590–606. <https://doi.org/10.1148/RADIOL.2018180547>
- Stokes, J. M., Yang, K., Swanson, K., Jaakkola, T. S., Barzilay, R., Correspondence, J. J. C., 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., ... Collins, J. J. (n.d.). *A Deep Learning Approach to Antibiotic Discovery Article A Deep Learning Approach to Antibiotic Discovery*. <https://doi.org/10.1016/j.cell.2020.01.021>
- Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights*, *14*, 1177932219899051. <https://doi.org/10.1177/1177932219899051>
- Suryanarayanan, V., Kumar Singh, S., Kumar Tripathi, S., Selvaraj, C., Konda Reddy, K., & Karthiga, A. (2013). A three-dimensional chemical phase pharmacophore mapping, QSAR modelling and electronic feature analysis of benzofuran salicylic acid derivatives as LYP inhibitors. *SAR and QSAR in Environmental Research*, *24*(12), 1025–1040. <https://doi.org/10.1080/1062936X.2013.821421>
- Swinney, D. C., & Lee, J. A. (2020). Recent advances in phenotypic drug discovery. *F1000Research*, *9*. <https://doi.org/10.12688/f1000research.25813.1>
- Takeuchi, I. (2002). The future of Lisp. *Systems and Computers in Japan*, *33*(6), 10–18. <https://doi.org/10.1002/SCJ.10000>
- Tewari, J., Strong, R., & Boulas, P. (2017). At-line determination of pharmaceuticals small molecule's blending end point using chemometric modeling combined with Fourier transform near infrared spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc.*, *173*, 886–891. <https://doi.org/10.1016/j.saa.2016.10.013>
- Tkatchenko, A. (2020). Machine learning for chemical discovery. *Nature Communications*, *11*(1), 4125. <https://doi.org/10.1038/s41467-020-17844-8>
- Turkoglu, M., Aydin, I., Murray, M., & Sakr, A. (1999). Modeling of a roller-compaction process using neural networks and genetic algorithms. *Eur J Pharm Biopharm.*, *48*(3), 239–245. [https://doi.org/10.1016/s0939-6411\(99\)00054-5](https://doi.org/10.1016/s0939-6411(99)00054-5)
- U.S. Food and Drug Administration. (2021). Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan. Retrieved from <https://www.fda.gov/media/122535/download>
- Valentini, G., Paccanaro, A., Caniza, H., Romero, A. E., & Re, M. (2014). An extensive analysis of disease-gene associations using network integration and fast kernel-based gene prioritization methods. *Artif Intell Med*, *61*(2), 63–78. <https://doi.org/10.1016/j.artmed.2014.03.003>

- Valizadeh, H., Pourmahmood, M., Mojarrad, J. S., Nemati, M., & Zakeri-Milani, P. (2009). Application of artificial intelligent tools to modeling of glucosamine preparation from exoskeleton of shrimp. *Drug Dev Ind Pharm.*, 35(4), 396–407. <https://doi.org/10.1080/03639040802422088>
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Trans Neural Netw.*, 10(5), 988–999. <https://doi.org/10.1109/72.788640>
- Vasanth Kumar, K., Martins, P., & Rocha, F. (2008). Modelling of the batch sucrose crystallization kinetics using artificial neural networks: comparison with conventional regression analysis. *Ind Eng Chem Res.*, 47(14), 4917–4923. <https://doi.org/10.1021/ie701706v>
- Vatansever, S., Schlessinger, A., Wacker, D., Kaniskan, H. Ü., Jin, J., Zhou, M., & Zhang, B. (2021). Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. *Medicinal Research Reviews*, 41(3), 1427–1473. <https://doi.org/10.1002/med.21764>
- Velásco-Mejía, A., Vallejo-Becerra, V., Chávez-Ramírez, A. U., Torres-González, J., Reyes-Vidal, Y., & Castañeda-Zaldivar, F. (2016). Modeling and optimization of a pharmaceutical crystallization process by using neural networks and genetic algorithms. *Powder Technol.*, 292, 122–128. <https://doi.org/10.1016/j.powtec.2016.01.028>
- Vink, G., Nebel, J.-C., & Wren, S. P. (2021). *In silico* design of bioisosteric modifications of drugs for the treatment of diabetes. *Future Medicinal Chemistry*, 13(8), 691–700. <https://doi.org/10.4155/fmc-2020-0374>
- Volk, M. J., Lourentzou, I., Mishra, S., Vo, L. T., Zhai, C., & Zhao, H. (2020). Biosystems design by machine learning. *ACS Synth Biol*, 9(7), 1514–1533. <https://doi.org/10.1021/acssynbio.0c00129>
- Wang, B., Liu, G., Liu, S., Fei, Q., & Ren, Y. (2009). Orthogonal projection to latent structures combined with artificial neural network for quantitative analysis of phenoxymethylpenicillin potassium powder. *Vib Spectrosc.*, 51(2), 199–204. <https://doi.org/10.1016/j.vibspec.2009.04.007>
- Wang, T., Yuan, X., Wu, M.-B., Lin, J.-P., & Yang, L.-R. (2017). The advancement of multidimensional QSAR for novel drug discovery - where are we headed? *Expert Opinion on Drug Discovery*, 1–16. <https://doi.org/10.1080/17460441.2017.1336157>
- What is Machine learning and why is it gaining so much popularity? - SQLRelease.* (n.d.). Retrieved June 29, 2023, from <https://sqlrelease.com/what-is-machine-learning-and-why-is-it-gaining-so-much-popularity>
- Wilton, D. J., Harrison, R. F., Willett, P., Delaney, J., Lawson, K., & Mullier, G. (2006). Virtual screening using binary kernel discrimination: analysis of pesticide data. *J Chem Inf Model*, 46(2), 471–477. <https://doi.org/10.1021/ci050397w>
- Wong, W. C., Chee, E., Li, J., & Wang, X. (2018). Recurrent neural network-based model predictive control for continuous pharmaceutical manufacturing. *Mathematics.*, 6(11), 242. <https://doi.org/10.3390/math6110242>
- Yamashita, F., Wanchana, S., & Hashida, M. (2002). Quantitative structure/property relationship analysis of Caco-2 permeability using a genetic algorithm-based partial

- least squares method. *Journal of Pharmaceutical Sciences*, *91*(10), 2230–2239. <https://doi.org/10.1002/jps.10214>
- Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. <https://doi.org/10.1007/s13244-018-0639-9>
- Yang, X., Wang, Y., Byrne, R., Schneider, G., & Yang, S. (2019). Concepts of artificial intelligence for computer-assisted drug discovery. *Chem Rev*, *119*(18), 10520–10594. <https://doi.org/10.1021/acs.chemrev.8b00728>
- You, J., McLeod, R. D., & Hu, P. (2019). Predicting drug-target interaction network using deep learning model. *Comput Biol Chem*, *80*, 90–101. <https://doi.org/10.1016/j.compbiolchem.2019.03.016>
- Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications*, *10*(1), 3770. <https://doi.org/10.1038/s41467-019-11786-6>
- Zakharov, A. V., Varlamova, E. V., Lagunin, A. A., Dmitriev, A. V., Muratov, E. N., Fourches, D., Kuz'min, V. E., Poroikov, V. V., Tropsha, A., & Nicklaus, M. C. (2016). QSAR Modeling and Prediction of Drug–Drug Interactions. *Molecular Pharmaceutics*, *13*(2), 545–556. <https://doi.org/10.1021/acs.molpharmaceut.5b00762>
- Zhang, M., Su, Q., Lu, Y., Zhao, M., & Niu, B. (2017). Application of machine learning approaches for protein-protein interactions prediction. *Med Chem*, *13*(6), 506–514. <https://doi.org/10.2174/1573406413666170522150940>
- Zhavoronkov, A., Vanhaelen, Q., & Oprea, T. I. (2020). Will artificial intelligence for drug discovery impact clinical pharmacology? *Clin Pharmacol Ther*, *107*(4), 780–785. <https://doi.org/10.1002/cpt.1795>
- Zhou, X., Li, C., Rahaman, M. M., Yao, Y., Ai, S., Sun, C., Wang, Q., Zhang, Y., Li, M., Li, X., Jiang, T., Xue, D., Qi, S., & Teng, Y. (2020). A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks. *IEEE Access*, *8*, 90931–90956. <https://doi.org/10.1109/ACCESS.2020.2993788>
- Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, *331*(6158), 679–684. <https://doi.org/10.1038/331679A0>