



TRABAJO DE FIN DE GRADO

LA INTELIGENCIA ARTIFICIAL EN SÍNTESIS ORGÁNICA



Nombre de la alumna: Ángela Romero Caballero

Nombre del tutor: José Ignacio Candela Lena

Departamento: Química orgánica y farmacéutica

Grado en Farmacia: Facultad de Farmacia. Universidad de Sevilla

Curso: 2022/2023



LA INTELIGENCIA ARTIFICIAL EN SÍNTESIS ORGÁNICA

Trabajo de fin de grado de carácter bibliográfico

Realizado por Ángela Romero Caballero. Grado en Farmacia.

Tutor: José Ignacio Candela Lena

Departamento de química orgánica

Facultad de Farmacia. Universidad de Sevilla.



Sevilla, junio de 2023.

¿Por qué esta magnífica tecnología científica, que ahorra trabajo y nos hace la vida más fácil, nos aporta tan poca felicidad? La respuesta es esta, simplemente: porque aún no hemos aprendido a usarla con acierto. (Albert Einstein)

Resumen

Cada vez es mayor el desafío al que se enfrentan los investigadores, ya que cada vez son mayores los requerimientos exigidos por el ámbito de la salud: Nuevas enfermedades, resistencia a fármacos, enfermedades conocidas, pero sin tratamiento o con un tratamiento ineficaz, exigencia de nuevas técnicas de diagnóstico y exploración...

El descubrimiento y desarrollo de fármacos es un proceso costoso, complejo y prolongado en el tiempo. Cada fármaco debe ser sometido a diferentes pruebas que pongan de manifiesto su seguridad y eficacia, así como los requerimientos fisicoquímicos y bioquímicos que deben superar. La inteligencia artificial (IA) ha supuesto un gran avance dentro de este ámbito, ya que, gracias a los sistemas de redes neuronales, diseñados por los investigadores y los algoritmos de alta complejidad que integra esta tecnología, se han obtenido datos claros y fiables acerca de nuevas moléculas, con un menor costo de producción y en un menor tiempo. Asimismo, la IA ha permitido la optimización de todos los procesos necesarios en la síntesis orgánica, lo que conlleva a obtener un rendimiento mucho más elevado en comparación con los procesos que hasta ahora se realizaban de forma manual. El diseño de nuevas moléculas utilizando el conocimiento, además de la metodología que nos ofrece esta disciplina, ha supuesto un gran avance para la ciencia.

Los avances en la IA, un sistema capaz de “suplantar” la inteligencia humana, nos brindan una excelente oportunidad para el desarrollo de la química orgánica, lo que conllevará a un gran avance de múltiples disciplinas.

En la actualidad, este campo está en dinámica evolución. El diseño de un sistema basado en la inteligencia humana, infalible y aplicable a múltiples situaciones, sin importar la complejidad, es un enorme esfuerzo que conlleva una gran inversión tanto en tiempo como en capital, sin embargo, la transformación que ha sufrido a lo largo del tiempo es colosal.

Palabras clave: Inteligencia artificial, química orgánica, molécula, síntesis, evolución

Glosario de siglas y abreviaturas

- AAE**: Adversarial autoencoders
- ADME**: Administración, distribución, metabolismo y transporte
- AE**: Atom environments
- ANN**: Artificial neural networks
- CNN**: Convolutional neural networks
- ConvGNN**: Convolutional graph neural network
- DL**: Deep learning
- GAAN**: Gated adversarial attention networks
- GAE**: Graph auto-encoders
- GAN**: Generative adversarial networks
- GAT**: Generative adversarial networks
- GCN**: Graph convolutional network
- GGN**: Graph generative networks
- GNN**: Graph neural network
- IA**: Inteligencia artificial
- LSTM**: Long short-term memory
- ML**: Machine learning
- PE**: Process element
- QSAR**: Relaciones cuantitativas estructura-actividad
- RecGNN**: Recurrent graph neural networks
- RNN**: Recurrent neural networks
- SMILES**: Simplified molecular input line entry system
- STGCN**: Spatial-temporal graph convolutional networks
- VAE**: Variational autoencoder

ÍNDICE

1. Introducción	4
1.1 Inteligencia artificial	4
• Bases de la Inteligencia Artificial	5
1.2 Aplicaciones de la IA en química	6
• Síntesis de nuevas moléculas prometedoras	8
• Selección de candidatos	10
• Retrosíntesis	11
• Verificación de la retrosíntesis	14
• Optimización de las reacciones químicas	14
2. Objetivos	15
3. Metodología	15
4. Resultados y discusión	16
• Aplicación de las técnicas de IA en la síntesis de nuevas moléculas.	16
• Predicción de las propiedades moleculares mediante IA.....	20
• Proceso retrosintético mediante IA.....	24
• Verificación de la retrosíntesis propuesta	26
• Análisis de las condiciones químicas de las reacciones propuestas	28
• Optimización de las reacciones químicas	30
5. Conclusiones	31
6. Bibliografía	33

1. Introducción

1.1 Inteligencia artificial

La IA es una rama de las ciencias de la computación que se basa en una serie de algoritmos lógicos, los cuales actúan simulando a la mente humana en la resolución de decisiones y problemas (Figura 1). La adquisición de datos por parte de una maquina en contraste con la humana, no es progresiva y no depende de la exposición a ciertas experiencias, lo que agiliza muchísimo la adquisición de datos, del mismo modo, la memoria humana tiene una capacidad reducida en comparación a la memoria digital (big data). Asimismo, la toma de decisiones por parte de los humanos conlleva tiempo y posibilidad de fracasar, mientras que estos pensamientos pueden ser transcritos a una serie de algoritmos, que las máquinas son capaces de interpretar y agilizar el proceso. En definitiva, la IA pretende emular la inteligencia humana. Combinando datos con posibles combinaciones, descubriendo así las pautas para la resolución de problemas, que permitan resolver con éxito situaciones complejas (Russell y Norvig, 2004). Dentro de este campo distinguimos dos ramas: el aprendizaje automático (ML) basado en algoritmos de regresión y el aprendizaje profundo (DL) basado en un mecanismo de redes neuronales (Naveja et al., 2022).

Sistemas que piensan como humanos	Sistemas que piensan racionalmente
«El nuevo y excitante esfuerzo de hacer que los computadores piensen... máquinas con mentes, en el más amplio sentido literal». (Haugeland, 1985) «[La automatización de] actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje...» (Bellman, 1978)	«El estudio de las facultades mentales mediante el uso de modelos computacionales». (Charniak y McDermott, 1985) «El estudio de los cálculos que hacen posible percibir, razonar y actuar». (Winston, 1992)
Sistemas que actúan como humanos	Sistemas que actúan racionalmente
«El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia». (Kurzweil, 1990) «El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor». (Rich y Knight, 1991)	«La Inteligencia Computacional es el estudio del diseño de agentes inteligentes». (Poole <i>et al.</i> , 1998) «IA... está relacionada con conductas inteligentes en artefactos». (Nilsson, 1998)

Figura 1. Definiciones de la Inteligencia artificial organizadas en cuatro secciones
(Russell y Norvig, 2004).

- **Bases de la Inteligencia Artificial**

La IA se basa en las conexiones neuronales del cerebro humano. La base de nuestra inteligencia reside en el cerebro humano, y este se basa en la información transmitida a través de nuestras neuronas, por lo que la IA parte de este concepto, basando su núcleo en un sistema de redes neuronales artificiales, compuesta por unos elementos que guardan semejanzas con las neuronas biológicas, denominados cada uno de ellos como "elemento procesador" (PE). Estos elementos tienen un funcionamiento que se asimila al de las neuronas, constando así de unos elementos de entrada (dendritas) que recogen los impulsos de entrada, que son integrados en el cuerpo y generan una respuesta. Las salidas de estos elementos (axón) se pueden conectar a las entradas de otros, mediante conexiones con una eficacia similar a la de la sinapsis neuronal (Figura 2). Estos PE están organizados en una serie de niveles denominados capas, y el conjunto de estas recibe el nombre de "redes neuronales artificiales" (ANN). Las ANN constituyen una poderosa herramienta dentro de la IA (Bajorath, 2022).

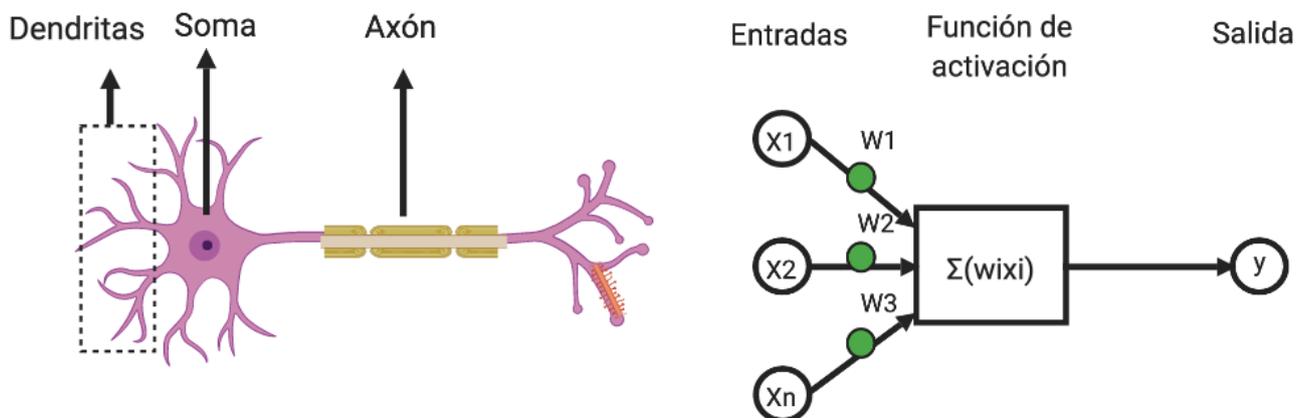


Figura 2. Comparación aclarativa de una neurona biológica y una neurona artificial
(García, 2019).

Existen dos métodos que permiten la expansión de información de la máquina: el primero se basa en la participación de un humano, el cual establece las relaciones correctas e incorrectas, y por tanto, etiqueta y categoriza los datos de entrada y establece un algoritmo específico para la toma de decisiones, para así generar diferentes salidas; y el segundo método se basa en dotar a la máquina de experiencias para que así sea capaz, de forma independiente, de ir aprendiendo y tomando decisiones complejas (Avila et al., 2020).

Los investigadores siguieron dos líneas de investigación, una de ellas consistía en el estudio en profundidad del cerebro y la segunda de ellas, buscó crear sistemas cuyo comportamiento se llevará a cabo de forma automática, tal y como lo realizaría una persona (Munera Salazar, 2010).

1.2 Aplicaciones de la IA en química

La investigación y desarrollo de fármacos se enfrenta a muchos desafíos, como el alto costo de comercialización, el gran porcentaje de fracasos en los ensayos clínicos y tiempos de ciclos prolongados (Yang et al., 2019). Estos ciclos de investigación pueden ser realizados a baja escala por un profesional, no obstante, cuando se trata de grupos muy grandes, se requiere del uso de algoritmos que realicen el trabajo de forma automática y en un menor tiempo (David et al., 2020). La IA ha permitido digitalizar un enorme volumen de datos complejos y de diferentes campos (Figura 3), englobando así medicina, educación, e incluso política, según últimas fuentes (Morrison, 2023). No obstante, esta novedosa tecnología requiere de la integración y cooperación de diversas ramas de conocimientos, tales como informática, matemática, estadística, química, economía... Esta herramienta tan eficaz ha permitido obtener numerosos avances en las ciencias de la salud y en farmacología.

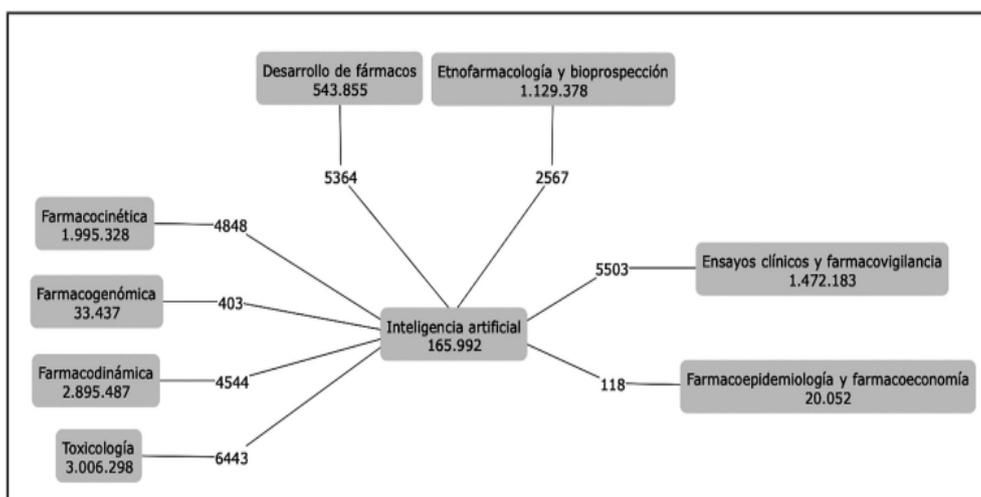


Figura 3. Diversificación de aplicaciones de la IA en diferentes ámbitos. Este esquema representa la variedad de publicaciones (números) en el portal PubMed en cada uno de los temas explorados. Correspondiendo los números de las líneas conectoras con el volumen de artículos de cada tema expuesto (Matsingos et al., 2021).

La farmacología es una ciencia de la salud que estudia los efectos y propiedades de los fármacos en el cuerpo humano. Para poder obtener estos fármacos hace falta el eslabón principal, una molécula que proporcione la actividad biológica, denominada principio activo. De la búsqueda y perfeccionamiento de este principio activo se encarga la química orgánica (Matsingos et al., 2021). Esta rama no sólo se limita al descubrimiento de fármacos, como se especifica en una revisión reciente (Butler et al., 2018), sino que sus aplicaciones se extienden a un mayor número de procesos (<https://chemintelligence.com/ai-for-chemistry>).

En la búsqueda de nuevas moléculas, la IA nos permite determinar rutas sintéticas, pero a su vez, también proporciona una predicción del producto y el rendimiento de una reacción, lo cual permite agilizar con creces dicho proceso. Además, esta tecnología brinda conocimientos sobre nuevas reacciones, ya que, gracias a esta, se pueden probar varias combinaciones de reactivos para encontrar reacciones nuevas y reproducibles. Estas reacciones se realizan teniendo en cuenta las condiciones óptimas para que puedan reproducirse de forma manual si se obtienen los resultados requeridos (Miljković et al., 2021).

- **Síntesis de nuevas moléculas prometedoras**

La búsqueda de nuevos fármacos se estructura en varias etapas. La primera fase será en la que tenga un mayor impacto la IA, ya que esta tecnología permite analizar y seleccionar los mejores candidatos (Tubert, 2001). En esta etapa se emplearán varios algoritmos de aprendizaje profundo para diseñar nuevas moléculas, entre los que podemos destacar los siguientes:

-Codificadores automáticos variacionales (VAE)

Estos autocodificadores combinan dos técnicas fundamentales: los autocodificadores (autoencoders) y los modelos generativos probabilísticos.

Un VAE es un autocodificador que consta de dos partes: el codificador, entrenado para generar a partir de los datos de entrada, una distribución probabilística de las propiedades del espacio latente; y el decodificador, el cual genera nuevos datos a partir de los datos representados en el espacio latente.

Este programa se somete a un entrenamiento previo, para optimizar los parámetros del codificador y el decodificador, para así obtener una correcta reconstrucción de los datos de entrada, y una estrecha relación entre los parámetros reales y los obtenidos (Rocca, 2019).

-Codificadores automáticos adversos (AAE)

Esta tecnología es similar a la anteriormente descrita, excepto por el empleo de redes generativas antagónicas (GAN). Estas redes están entrenadas con un nuevo elemento denominado discriminador. Esta incorporación tiene el objetivo de discriminar entre los datos reales introducidos en el programa, y los datos generados por el decodificador. Como resultado, a medida que se entrena el modelo, el codificador y el autocodificador generan muestras cada vez más similares a las insertadas, obteniendo resultados más realistas y similares (Rashad, 2020).

-Redes neuronales recurrentes (RNN)

Las RNN están diseñadas para trabajar con datos secuenciales utilizando conexiones recurrentes. Estas conexiones permiten almacenar los datos, simulando una especie de “memoria”, y generar así, a partir de salidas anteriores, nuevas entradas. De esta forma, este tipo de metodología emplea un mecanismo de retroalimentación (Saeed, 2022).

-Redes neuronales convolucionales (CNN)

Las CNN están diseñadas, para extraer información de grandes volúmenes de datos, analizando las características locales a través de un sistema jerárquico de capas convolucionales:

- Capa convolucional: es la encargada de aplicar la convolución a las imágenes de entrada, extrayendo patrones que serán analizados en las siguientes capas.
- Capa de agrupación: lleva a cabo una disminución de la dimensionalidad.
- Capa completamente conectada: esta capa realiza la clasificación de los parámetros obtenidos en las capas anteriores, para así generar las salidas deseadas (<https://www.intel.es/content/www/es/es/company-overview/company-overview.html>).

Estos algoritmos deben entrenarse con una gran variedad de moléculas, para construir una distribución estadística. Se generarán, por tanto, estructuras moleculares, ya sea como cadenas SMILES o directamente como gráficos (<https://chemintelligence.com/ai-for-chemistry>).

Las cadenas SMILES son un tipo de nomenclatura utilizada para representar estructuras químicas. Posee la ventaja de ser interpretada de forma computacional para la representación de estructuras, pero al mismo tiempo puede ser interpretada de forma manual. La forma de indicar un elemento químico utilizando esta metodología sería la siguiente:

Átomo: [<masa> símbolo<quiralidad> <hidrógenos> <signo<carga>>]

Cuando es necesario añadir información extra, se indica dentro de los corchetes (Hornos, 2020).

Además, al utilizar este tipo de algoritmos, deberá comprobarse que las moléculas generadas sean diferentes a las moléculas introducidas en los algoritmos (Figura 4).

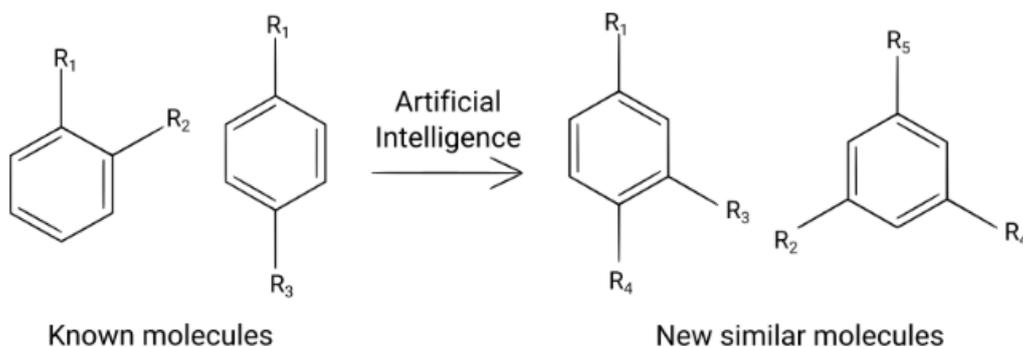


Figura 4. Síntesis de nuevas moléculas a partir de algoritmos de aprendizaje profundo (<https://chemintelligence.com/ai-for-chemistry>).

➤ Selección de candidatos

Una vez generado el banco de nuevas moléculas, se seleccionarán las que posean interés desde el punto de vista de sus propiedades. Existe una gran variedad de algoritmos, los cuales pueden ser empleados para predecir las propiedades de las nuevas moléculas obtenidas. Gracias al conocimiento de estos parámetros, podemos predecir el comportamiento del fármaco en el cuerpo humano, y de esta forma, podemos optimizarlo para obtener los mejores resultados (Matsingos et al., 2021). Además, permitirán predecir propiedades como bioactividad, toxicidad, solubilidad, puntos de fusión y muchos otros tipos (Figura 5).

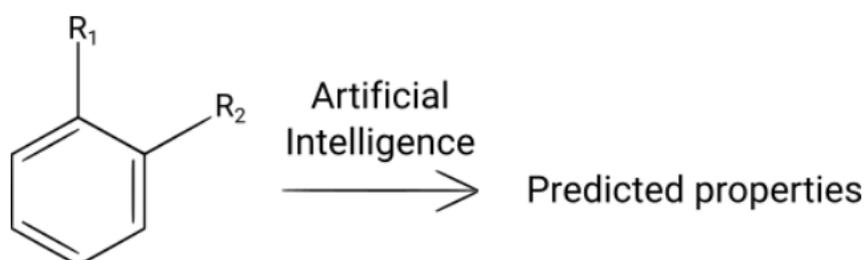


Figura 5. Predicción de las propiedades de las nuevas moléculas obtenidas por IA

(<https://chemintelligence.com/ai-for-chemistry>).

Actualmente existen distintas metodologías que establecen la relación entre diferentes variables fisicoquímicas y los fenómenos de absorción, distribución, metabolismo y excreción (ADME), así como la relación de estas con las relaciones cuantitativas estructura-actividad (QSAR). La IA actuará en este campo impulsando la eficacia de todos estos procesos, manejando así un gran volumen de datos en un tiempo reducido (Tubert, 2001).

Cuando los científicos diseñan nuevas moléculas de forma manual, es necesario que las sintetizen para poder comprobar todas sus propiedades bioquímicas y fisicoquímicas. En el caso de que estas moléculas no posean las propiedades requeridas, para ser administradas y que ejerzan su acción, los científicos diseñarán nuevas hasta obtener aquellas que cumplan sus requisitos. Este es un proceso que requiere mucho tiempo y dinero.

Ser capaz de predecir con precisión las propiedades de moléculas hipotéticas, evitará la generación de compuestos derivados de la molécula inicial para la comparación de parámetros, evitando así costes de producción y trabajo. (<https://chemintelligence.com/ai-for-chemistry>).

➤ **Retrosíntesis**

Partiendo de este punto, una vez que se seleccionan las moléculas objetivo, hay que realizar un proceso de retrosíntesis, en el cual, a partir de una serie de moléculas básicas e iniciales, a las que se les somete a una serie de reacciones químicas utilizando los reactivos adecuados, se generará la molécula requerida. De esta forma, se obtendrá un árbol retrosintético (Figura 6) que representa todos los pasos a seguir (Tubert, 2001).

Este proceso realizado de forma manual por químicos supone una gran tasa de errores y un proceso alargado en el tiempo. Además, a este inconveniente hay que sumarle los

enormes costes de materiales necesarios y el trabajo requerido por los profesionales, debido al gran control de todas las variables que hay que tener en cuenta en cada una de las reacciones llevadas a cabo (<https://chemintelligence.com/ai-for-chemistry>).

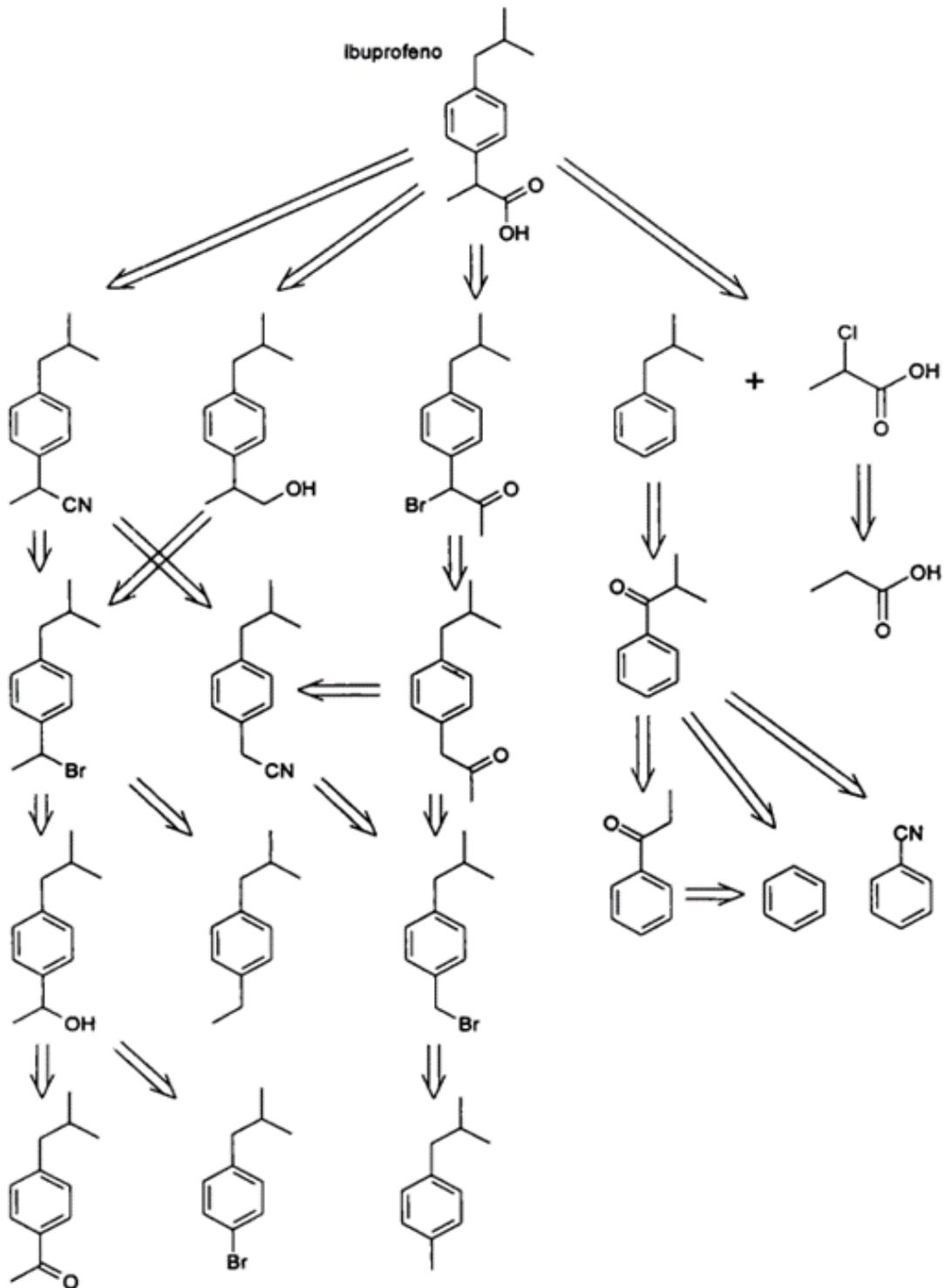


Figura 6. Árbol retrosintético del ibuprofeno (Tubert, 2001).

Recientemente se han desarrollado algoritmos de aprendizaje automático de alto rendimiento, para realizar este proceso a través de programas informáticos, entrenados con millones e incluso decenas de millones de reacciones químicas. En estos programas se ingresan las moléculas objetivo y se genera un árbol retrosintético esquematizado (Figura 7), en el que nos muestra los materiales de partida y los intermediarios (<https://chemintelligence.com/ai-for-chemistry>).

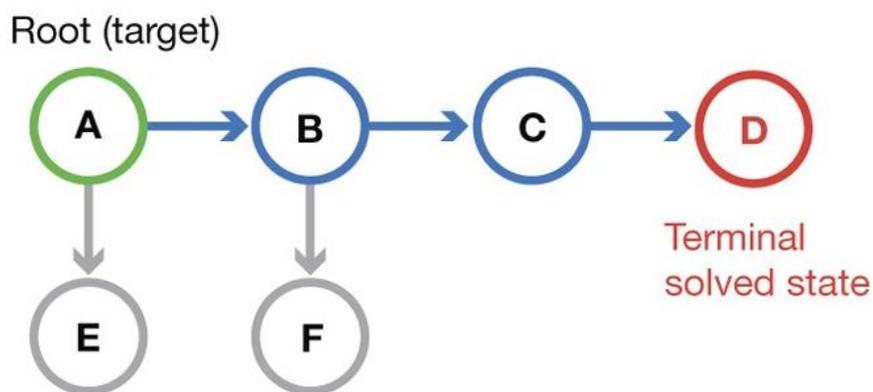


Figura 7. Representación del árbol de búsqueda diseñado con IA. Los nodos del árbol representan los intermediarios de síntesis y cada hoja es un material de partida disponible comercialmente (Segler et al., 2018).

Estas técnicas de retrosíntesis, basadas en algoritmos de aprendizaje automático, poseen limitaciones: en primer lugar, su eficacia en moléculas complejas es cuestionable, debido a la menor cantidad de moléculas disponibles para entrenar los algoritmos; en segundo lugar, no son capaces de predecir de forma fiable la estereoquímica de los productos de la reacción; y en último lugar, la literatura insertada en estos programas puede contener errores por lo que es necesaria la comprobación de las reacciones propuestas (<https://chemintelligence.com/ai-for-chemistry>).

No obstante, se ha demostrado que el rendimiento en las predicciones de retrosíntesis elaboradas por IA, es equivalente al obtenido por los químicos cuando realizan estos procesos de forma manual (Segler et al., 2018)

➤ Verificación de la retrosíntesis

Debido a las limitaciones descritas anteriormente, es necesaria la verificación de las reacciones propuestas por los programas informáticos. Estas comprobaciones pueden ser realizadas por programas de aprendizaje automático, en los cuales se insertan los reactivos y el programa predice el producto principal y los subproductos de la reacción. Ahora bien, esta técnica posee una limitación, no tienen en cuenta las condiciones de la reacción, como la temperatura, disolvente, enzimas... (<https://chemintelligence.com/ai-for-chemistry>).

➤ Optimización de las reacciones químicas

Una vez completados los procesos anteriores, es necesario identificar las condiciones a las que se debe ejecutar cada reacción. Este proceso puede basarse en la comparación con reacciones similares y la experiencia previa. Sin embargo, para comprobar dichas condiciones sería necesario hacerlo de forma experimental, lo que supone el control de distintos parámetros complejos y sensibles, y una gran duración.

En contraposición a este proceso, se puede hacer uso de redes neuronales de aprendizaje profundo, las cuales poseen una mayor sensibilidad que los algoritmos previos (Figura 8). Gracias a este proceso, se puede simular rápidamente una gran variedad de parámetros para encontrar una solución factible.

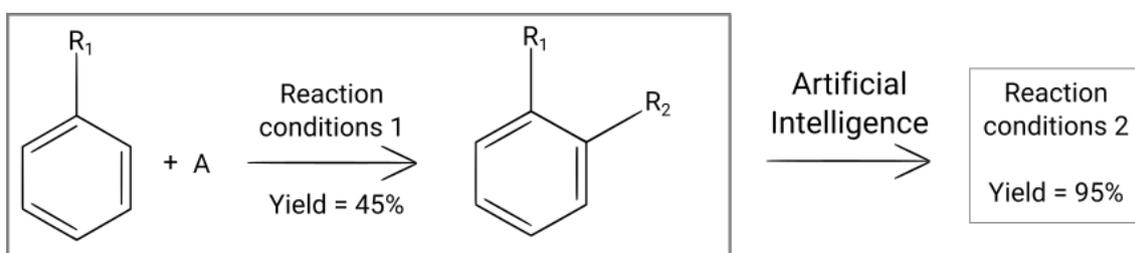


Figura 8. Mejora de las condiciones de la reacción para obtener un mayor rendimiento (<https://chemintelligence.com/ai-for-chemistry>).

2. Objetivos

El objetivo de este proyecto es, realizar una revisión bibliográfica sobre la evolución de la IA en el ámbito de la disciplina de química orgánica, en concreto de la síntesis orgánica, así como conocer su metodología, aplicaciones y sus objetivos a largo plazo. En segundo lugar, este trabajo pretende conocer en mayor profundidad en que consiste la aplicación de la IA dentro de la síntesis orgánica, así como los avances que ha supuesto y los hitos que se han conseguido a raíz del descubrimiento de esta metodología. Por último, en esta revisión se pretende aclarar la importancia de continuar con de la investigación dentro de este campo, para así poder seguir evolucionando a lo largo del tiempo, ya que, aunque el desarrollo de esta metodología está muy avanzado, todavía queda mucho por explorar.

3. Metodología

Para la realización de esta revisión bibliográfica se ha realizado una búsqueda partiendo de una gran variedad de bases de datos, buscadores y otros recursos electrónicos, abarcando así el empleo de artículos, revistas, libros y tesis.

De manera general, se ha ido recopilando toda la información posible, integrando el conocimiento otorgado a través de todos los recursos empleados, para así poder contrastar toda la información, seleccionando la información con más relevancia, con el objetivo de llevar a cabo una categorización esquematizada de todo el material recolectado. De esta forma, se han empleado tanto reviews como artículos de investigación, con el objetivo de generar la información más versátil posible.

Los artículos empleados se han obtenido de diferentes bases de datos entre las que podemos destacar: Mendeley, PubMed, Research gate, Scifinder, Medes y Scopus. No obstante, el empleo del motor de búsqueda, Google Scholar, también ha sido una fuente de artículos relevante.

El empleo de libros obtenidos a través de Fama, Academia Edu y Google Scholar, así como los datos obtenidos gracias a la red de revistas científicas Redalyc, han sido de vital importancia para la elaboración de múltiples puntos dentro de esta investigación.

Para terminar, las tesis que han servido como punto de apoyo para poder concluir toda la información requerida, han sido extraídas del portal de difusión conocido como Dialnet, así como del buscador Google Scholar.

Todos estos recursos han jugado un papel crucial en la realización de este trabajo, siendo el anteriormente nombrado, Google Scholar, el motor de búsqueda que ha proporcionado una mayor variedad de información.

Como criterios generales de selección de toda la información recolectada, se ha empleado el uso de un repertorio de palabras claves, dentro del cual podemos destacar las siguientes: artificial intelligence, organic synthesis, optimisation, prediction, properties, design, molecular, retrsynthesis, reaction... Así como la preferencia de elección de la bibliografía más reciente (a partir del año 2000), con el fin de generar la información más actualizada posible.

4. Resultados y discusión

La aplicación de la IA en síntesis orgánica se puede clasificar en las diversas etapas del proceso sintético. A continuación, se destacarán las tecnologías más avanzadas y eficientes dentro de este ámbito, según su aplicación en química.

➤ Aplicación de las técnicas de IA en la síntesis de nuevas moléculas.

El aprendizaje automático, en concreto el aprendizaje profundo, ha sido ampliamente utilizado en la generación de moléculas novedosas. Los enfoques basados en IA han demostrado su capacidad para generar estructuras químicas viables y mejorar la diversidad de los compuestos sintetizados (Gupta et al., 2017).

En un artículo de investigación (Gómez et al., 2018), se reporta un método reciente basado en redes neuronales profundas, este modelo permite la generación de nuevas moléculas. Estas redes neuronales, están entrenadas con un autocodificador para convertir moléculas representadas como cadenas SMILES.

El autocodificador se compone de dos redes profundas: una red codificadora para convertir cada cadena en un vector, y una red decodificadora para convertir vectores de nuevo en cadenas, tal y como se esquematiza a continuación (Figura 9).

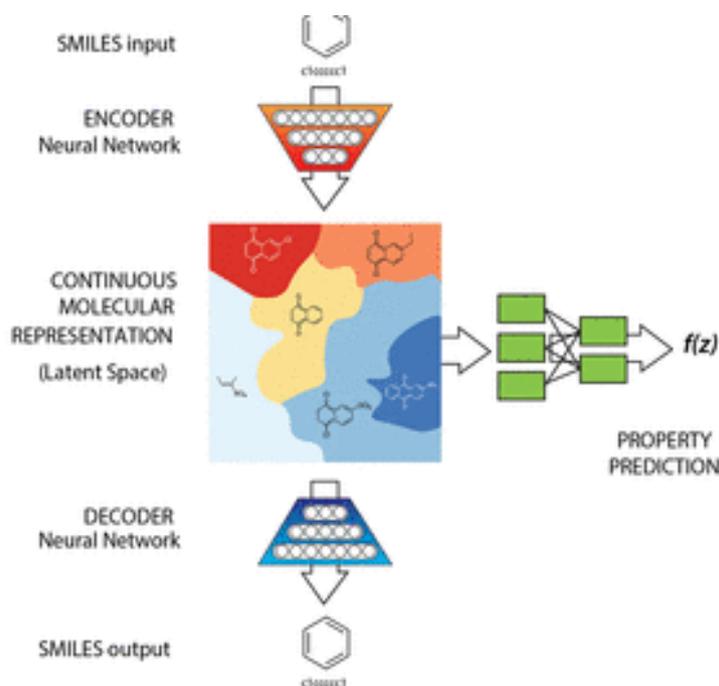


Figura 9. Modelo del autocodificador utilizado para el diseño molecular (Gómez et al., 2018).

Para poner a prueba este método se realizó un ensayo a partir del ibuprofeno, para sintetizar moléculas con propiedades similares (Figura 10). Estas estructuras se vuelven menos similares al aumentar la distancia del espacio latente. Para garantizar que los puntos en esta zona correspondan a moléculas válidas, se eligió un VAE.

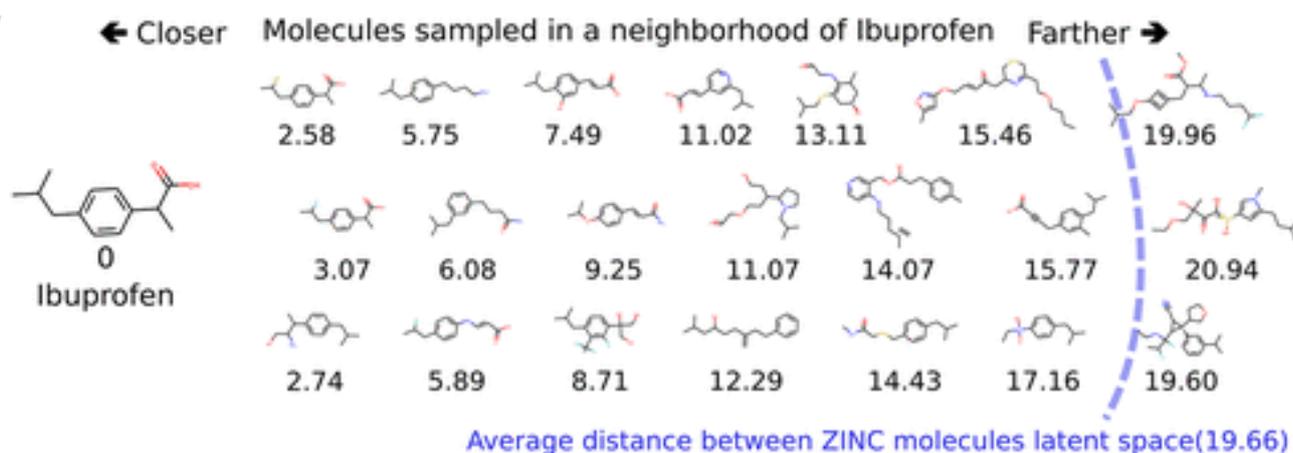


Figura 10. Moléculas sintetizadas a partir del ibuprofeno. Los valores que se muestran en la parte inferior de las moléculas corresponden a la distancia en el espacio latente entre la molécula decodificada y la molécula inicial. Cuanto menor sea este valor, mayor será la similitud entre ambas estructuras (Gómez et al., 2018).

No obstante, el espacio químico es demasiado extenso para poder ser analizado en su totalidad con el objetivo de generar nuevas moléculas. De esta forma, un nuevo método propone el análisis de un grupo de medicamentos con actividades conocidas contra objetivos particulares, para la generación de nuevas moléculas.

Este nuevo enfoque emplearía la metodología de aprendizaje profundo RNN basado en células LSTM (memoria larga a corto plazo). La aplicación de este tipo de células permite recordar los estados previos y decidir los siguientes, resolviendo el problema de la pérdida de datos que experimentan los RNN.

A partir de este modelo, se realizó un estudio, estructurado en dos partes. La primera parte de este estudio se llevó a cabo en dos fases. La primera fase, consistía en el entrenamiento de esta tecnología para generar bibliotecas de cadenas SMILES precisas y válidas. En segundo lugar, se hizo uso del aprendizaje de transferencia para afinar este modelo, generando así moléculas estructuralmente similares a fármacos comercializados. Se demostró, por tanto, que este modelo es capaz de generar moléculas en situaciones en las que se disponen de pocos datos.

En la segunda parte de este análisis, haciendo uso del modelo generativo, se generó una biblioteca de compuestos mediante el crecimiento de fragmentos (Figura 11).

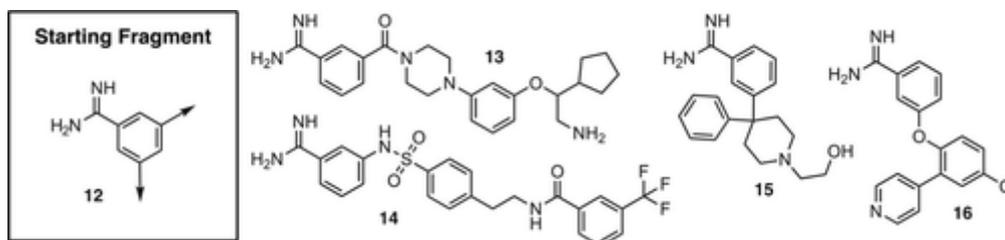


Figura 11. Generación de moléculas a partir de un fragmento principal (Gupta et al., 2017)

Además, este modelo también nos proporciona la opción de realizar el proceso de optimización hit-to-lead, permitiendo así la generación de estructuras químicamente válidas. De esta forma, este modelo de aprendizaje profundo permite la generación de bibliotecas de compuestos con cantidades bajas y altas de entrenamiento, el descubrimiento de fármacos basado en fragmentos y la optimización hit-to-lead (Gupta et al., 2017).

Existen múltiples estudios sobre el uso de redes generativas en la síntesis de nuevas moléculas, hasta ahora este tipo de modelos se han elaborado como enfoques quimiocéntricos, sin tener en cuenta la biología resultante de la interacción entre el ligando y la diana objetivo.

Otro ejemplo dentro de este tipo de IA sugiere la combinación de los modelos generativos con la biología de sistemas y el diseño molecular. Utilizando este sistema, podemos diseñar automáticamente moléculas que tengan una alta probabilidad de producir un perfil transcriptómico deseado. Este método, se basa en la síntesis de moléculas desde cero utilizando la información de la expresión génica para guiar el proceso de síntesis de nuevas moléculas.

Dentro de este estudio, se realizaron diversos experimentos para validar la actividad biológica de las moléculas generadas, demostrando así como la combinación de la IA

con las firmas de expresión génica pueden conducir con éxito a la síntesis de nuevas moléculas prometedoras (Figura 12).

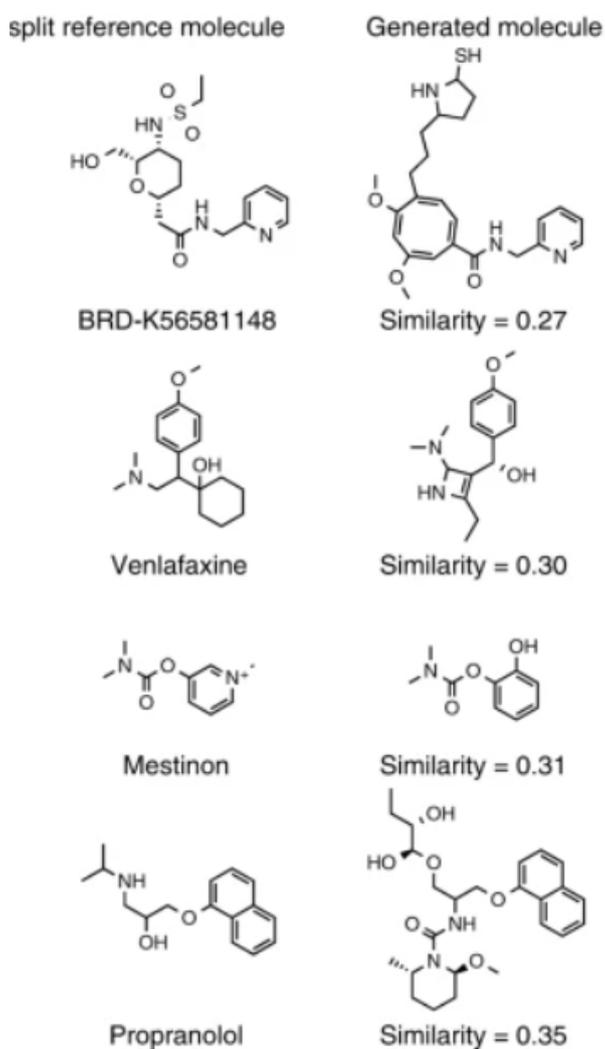


Figura 12. Ejemplos de moléculas generadas a partir de su compuesto de referencia (Méndez et al., 2020).

➤ Predicción de las propiedades moleculares mediante IA

En la industria farmacéutica, la comercialización de nuevos fármacos está sujeta a su eficacia clínica y su seguridad en humanos. Estos requisitos están determinados por las propiedades moleculares de cada fármaco, así como su objetivo terapéutico. Sin

embargo, la predicción de estas propiedades de forma precisa es un desafío, debido a la complejidad estructural de cada una de las moléculas estudiadas (Goh et al., 2017).

El método de aprendizaje profundo descrito en el punto anterior incluye la predicción de las propiedades moleculares. Las estructuras codificadas deben correlacionarse con las propiedades objetivas, por tanto, este modelo incluye un autocodificador que predice las propiedades de la representación del espacio latente. Para ello, se agrega un perceptrón multicapa que predice las propiedades moleculares a partir de la representación generada por el codificador (Gómez et al., 2018). Un perceptrón multicapa es una red neuronal unidireccional, compuesta por tres tipos de capas (Figura 13):

-Capa de entrada: recibe la señal de entrada del exterior y la transfiere a las siguientes capas.

-Capas ocultas: realizan el procesamiento de los datos a través de la información recibida por las capas anteriores.

-Capa de salida: recibe los valores de la última capa oculta y los transforma en valores de salida.

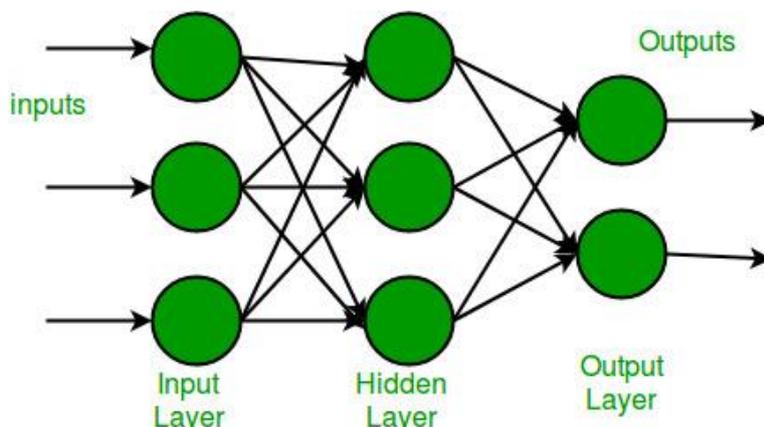


Figura 13. Diagrama perceptrón multicapa (Pathak, 2022).

Posteriormente, las representaciones latentes pueden decodificarse en cadenas SMILES (Figura 14).

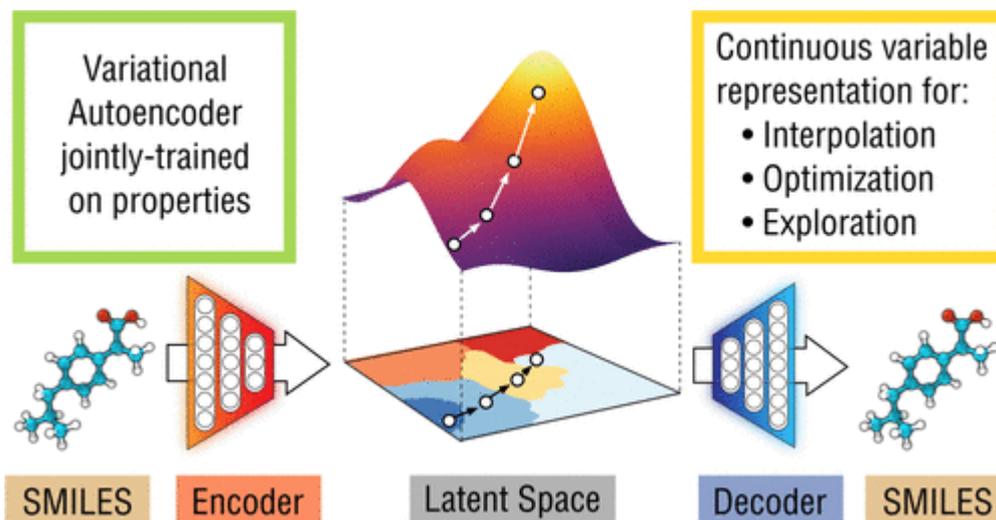


Figura 14. Representación de las propiedades basada en gradientes en el espacio latente (Gómez et al., 2018).

Otro ejemplo dentro de este ámbito, nos lo proporciona un estudio reciente (Zeng et al., 2022), en el que se describe una nueva metodología para la predicción de las propiedades moleculares. Este modelo está basado en un sistema de aprendizaje profundo de preentrenamiento no supervisado, llamado ImageMol, preentrenado con 10 millones de moléculas bioactivas, similares a fármacos, el cual permite predecir de forma precisa las propiedades moleculares (Figura 15).

Esta metodología, con conciencia química, emplea imágenes moleculares para representar estructuras moleculares, en lugar de las representaciones moleculares tradicionales. ImageMol combina un marco de procesamiento de imágenes con un amplio conocimiento en química para obtener así características moleculares de una forma de computación visual.

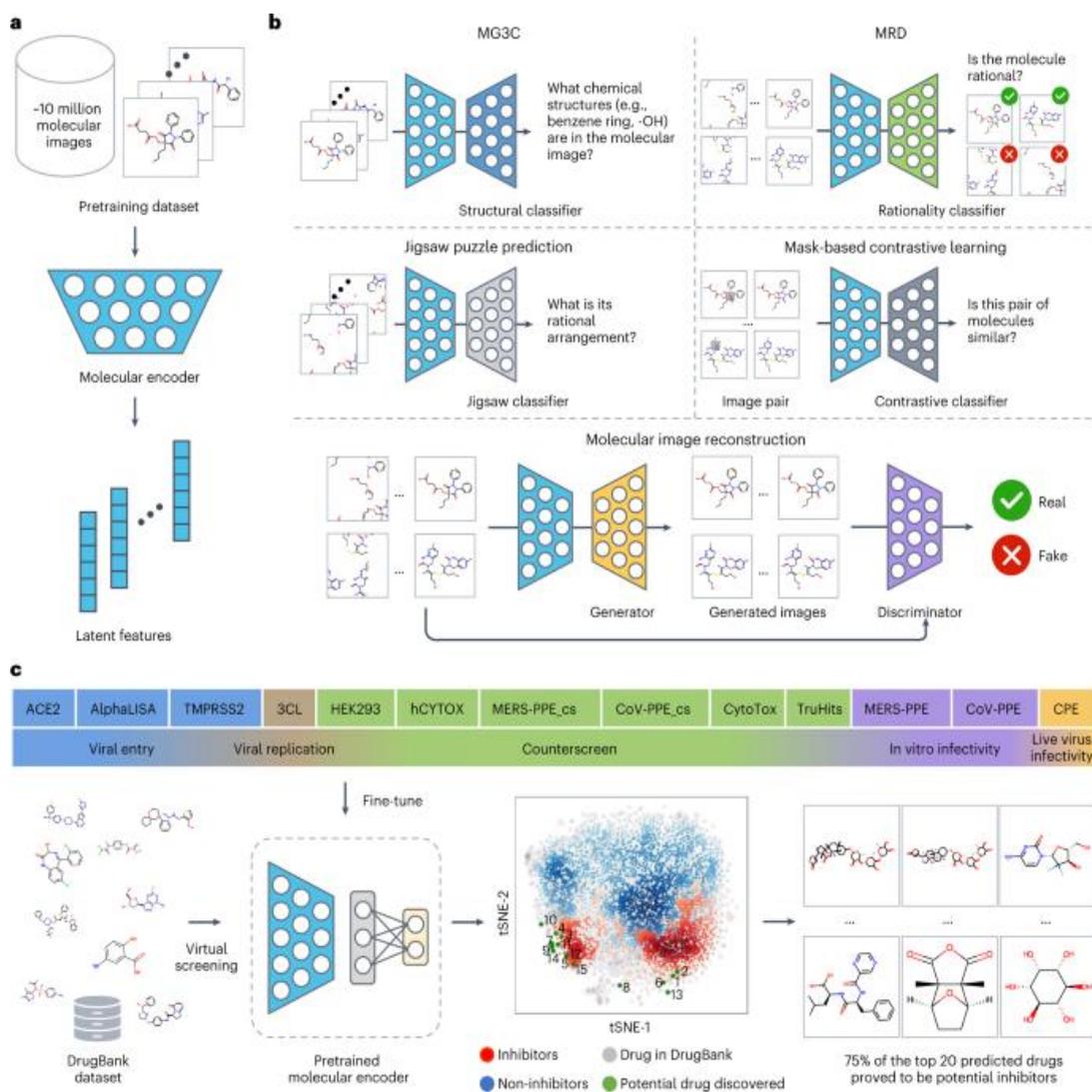


Figura 15. Diagrama ImageMol. A) Un codificador molecular diseñado para extraer características de las imágenes moleculares; B) Empleo de 5 estrategias para entrenar el codificador molecular; y C) Codificador molecular, entrenado con anterioridad, para mejorar el rendimiento del diseño (Zeng et al., 2022).

Al poner a prueba esta metodología, se demostró la alta precisión del software ImageMol en múltiples conjuntos de datos biomédicos de referencia, al realizar diversas tareas relacionadas con el descubrimiento de fármacos. En concreto, se logró identificar candidatos anti-SARS-CoV-2, los cuales se validaron mediante datos clínicos y experimentales en desarrollo, llevados a cabo en un total de 13 ensayos biológicos anti-SARS-CoV-2.

No obstante, esta tecnología posee limitaciones entre las que podemos destacar la escasez de imágenes moleculares, lo que puede afectar a los resultados obtenidos. Sin embargo, permanece en estudio diferentes mejoras para poder optimizar ImageMol y conseguir mitigar sus limitaciones (Zeng et al., 2022).

➤ **Proceso retrosintético mediante IA**

La predicción del proceso retrosintético de moléculas orgánicas, es una parte fundamental dentro del descubrimiento de nuevos fármacos. Este proceso ha ido evolucionando en los últimos años, surgiendo así nuevas técnicas basadas en IA, las cuales han demostrado ser una herramienta muy efectiva en el desempeño de estos procedimientos. A continuación, se abordarán varias técnicas utilizadas en la actualidad para la predicción de la retrosíntesis (Jiang et al., 2022).

Las moléculas además de ser representadas como cadenas SMILES, pueden ser codificadas como gráficos. La investigación de redes neuronales de grafos (GNN) ha proporcionado soluciones para representar las moléculas en grafos, para así poder predecir las transformaciones moleculares en las reacciones (Gori et al., 2005). Los grafos están compuestos por un conjunto de nodos (vértices) y un conjunto de aristas (bordes), que conectan los nodos. Este sistema permite extraer la información local de cada nodo, y a su vez las interacciones globales dentro del grafo (Zhou et al., 2020).

Se han desarrollado diferentes tipos de GNN, el primero de ellos fue las GNN recurrentes (RecGNN), en las que la información se propaga a las redes neuronales próximas para así actualizar la representación del nodo objetivo (Scarselli et al., 2008).

Posteriormente se desarrolló la red neuronal de grafos convolucionales (ConvGNN). Las convoluciones en grafos implican la propagación de información a través de las conexiones entre los nodos del grafo. Existen dos categorías principales dentro de ConvGNN: enfoques basados en el espectro y enfoques basados en el espacio. Mientras que los enfoques basados en el espectro se basan en el procesamiento de señales de grafos para adoptar filtros (Shuman et al., 2012), los enfoques basados en el espacio se basan en la propagación de la información (Micheli, 2009).

Más tarde, un grupo de investigadores desarrollaron la red convolucional de grafos (GCN), para cerrar la brecha entre los enfoques anteriores. Los GCN utilizan un tipo de propagación basada en el espacio. Este modelo ha demostrado buen rendimiento en la clasificación y predicción de propiedades moleculares (Defferrard et al., 2016).

En la actualidad las GCN han seguido evolucionando, desarrollando métodos basados en mecanismos de atención, entre los que destacamos dos tipos:

-Redes de atención de grafos (GAT): este modelo establece la importancia de cada nodo y posteriormente llevan a cabo un mecanismo de agregación de nodos (Veličković et al., 2017).

-Redes de atención generativas adversarias (GAAN): este tipo de redes neuronales se basan en la integración del generador y el discriminador de las redes GAN, con mecanismos de atención, los cuales les permiten extraer la información más relevante (Li et al., 2016).

A partir de estos trabajos se han desarrollado muchos enfoques alternativos de GNN, incluidos los autocodificadores de grafos (GAE) (Cao et al., 2016), las redes generativas de grafos (GGN) (De Cao et al., 2018) y las redes convolucionales de grafos espacio-temporales (STGCN) (Yan et al., 2018).

No obstante, la mayoría de predictores de la síntesis se basan en la arquitectura Transformer. Transformer es un modelo de traducción automática neuronal sin plantillas, que depende únicamente del mecanismo de atención (Vaswani et al., 2017). Molecular Transformer fue la primera técnica con SMILES desarrollada a partir de la arquitectura Transformer. No obstante, el desarrollo de predictores de la retrosíntesis utilizando SMILES, posee diferentes inconvenientes, como la gran tasa de errores que posee este método (Lin et al., 2020)

En contraposición, un artículo reciente (Ucak et al., 2022) señala la utilización de entornos atómicos (AE), en lugar de los enfoques convencionales basados en SMILES. Los AE son fragmentos que consisten en un átomo central y sus elementos unidos covalentemente con un radio establecido, considerándose la base de síntesis de las moléculas (Hähnke et al., 2015). Este tipo de modelo se denominó RetroTrae. A continuación se muestra un ejemplo representativo de las predicciones de retrosíntesis

elaboradas por este método (Figura 16). RetroTrae proporciona una predicción de las rutas retrosintéticas rápida y fiable. Demostrando así, que los AE son descriptores prometedores para desarrollar nuevas técnicas prometedoras.

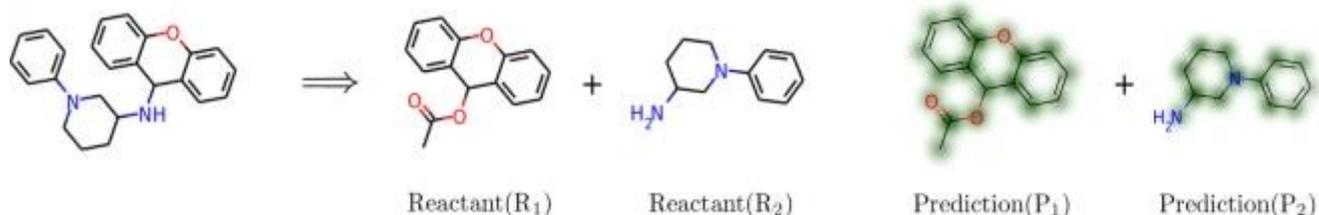


Figura 16. Reactivos resultantes de la retrosíntesis mediante RetroTrae. (Ucak et al., 2022).

➤ Verificación de la retrosíntesis propuesta

La planificación de la retrosíntesis, dentro del proceso de síntesis de nuevas moléculas, es un proceso arduo y que ha luchado por conseguir una evolución satisfactoria dentro del campo de la IA, debido a la gran complejidad que supone este proceso. Además, estas predicciones, al ser realizadas en el laboratorio, suelen fallar, a pesar de que inicialmente parecían fiables.

Por este motivo es de vital importancia la comprobación de las reacciones propuestas en la retrosíntesis, para así poder asegurarnos de que los resultados previstos de forma computacional coinciden con los resultados obtenidos de forma experimental.

En un estudio reciente (Coley et al., 2017), se muestra un nuevo modelo para predecir los resultados de las reacciones. Este modelo integra el uso de plantillas de reacción con el reconocimiento de patrones que ofrecen las redes neuronales. Para ello, dicho programa se entrena con 15.000 registros experimentales de reacciones.

Esta estrategia incluye varios enfoques adicionales: un aumento de datos, complementada con ejemplos de reacciones químicamente elogiadas; una

representación de la reacción centrada fundamentalmente en la obtención de los productos; y la implementación de un sistema ML, basado en redes neuronales, que aprende cuando es más probable que ocurran ciertos tipos de reactividad (Coley et al., 2017).

Este modelo predice el resultado de una reacción química en dos pasos. El primer paso es la aplicación de plantillas de reacción a un grupo de reactivos, con el objetivo de generar un conjunto de productos. El segundo paso, estima el producto principal de la reacción empleando ML (Figura 17) (Christ et al., 2012).

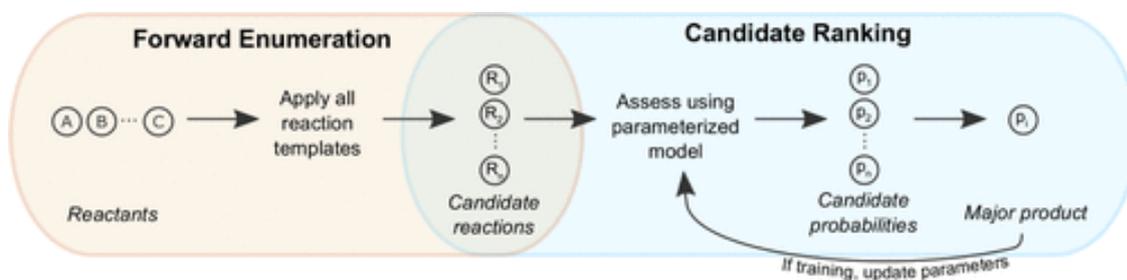


Figura 17. Fases del proceso de predicción de reacciones químicas (Coley et al., 2017).

Este modelo de ML se entrenó siguiendo un enfoque de aprendizaje supervisado, donde se integraron pares de entrada-salida de reacciones orgánicas previamente conocidas. Estos pares consistían en la descripción de la reacción y los productos de cada reacción química. Tomando de base esta información, este modelo fue capaz de aprender a reconocer patrones y relaciones entre las variables de entrada y salida.

Para poder poner a prueba esta técnica se realizaron experimentos de validación cruzada, para evaluar la eficacia en la predicción de las reacciones de este programa. Los resultados obtenidos fueron gratificantes, este modelo basado en redes neuronales fue capaz de predecir con precisión los resultados obtenidos en las reacciones orgánicas, en la mayoría de los casos.

De forma complementaria se realizó un análisis comparativo de la precisión entre este programa informático y otros modelos de predicción tradicionales. Este estudio comparativo demostró que el enfoque basado en el aprendizaje automático generaba un mayor rendimiento y una mayor tasa de precisión, en contraste a los otros métodos estudiados.

➤ **Análisis de las condiciones químicas de las reacciones propuestas**

La predicción de las condiciones a las que deben realizarse las reacciones químicas es una parte fundamental dentro del proceso de síntesis, debido a que el conocimiento de estas es necesario para llevar a cabo la validación experimental, y según las condiciones propuestas puede variar el resultado de las reacciones.

En numerosas ocasiones se ha intentado llevar a cabo este proceso de forma computacional sin embargo, la mayoría de los trabajos existentes se centran en la determinación de una sola condición química necesaria para que se produzca la reacción, como puede ser el disolvente, el catalizador, la temperatura...

Uno de estos métodos, consistía en un sistema experto para predecir el tipo de catalizador y solvente utilizado para las adiciones de Michael, siendo ineficaz para predecir el resto de condiciones (Marcou et al., 2015).

Dentro de este campo vamos a exponer una metodología basada en redes neuronales entrenada con 10 millones de ejemplos. Este modelo está entrenado para predecir un catalizador, dos solventes, dos reactivos y la temperatura para una determinada reacción química.

Este modelo se puso a prueba para poder evaluar su eficacia seleccionando una variedad de reacciones orgánicas comunes como: hidrólisis, reducción, oxidación, esterificación...

Dentro de cada tipo de reacción se emplearon alrededor de cinco ejemplos de reacciones. A continuación se incluye un esquema de algunas de las reacciones llevadas a cabo con esta metodología (Figura 18) (Gao et al., 2018).

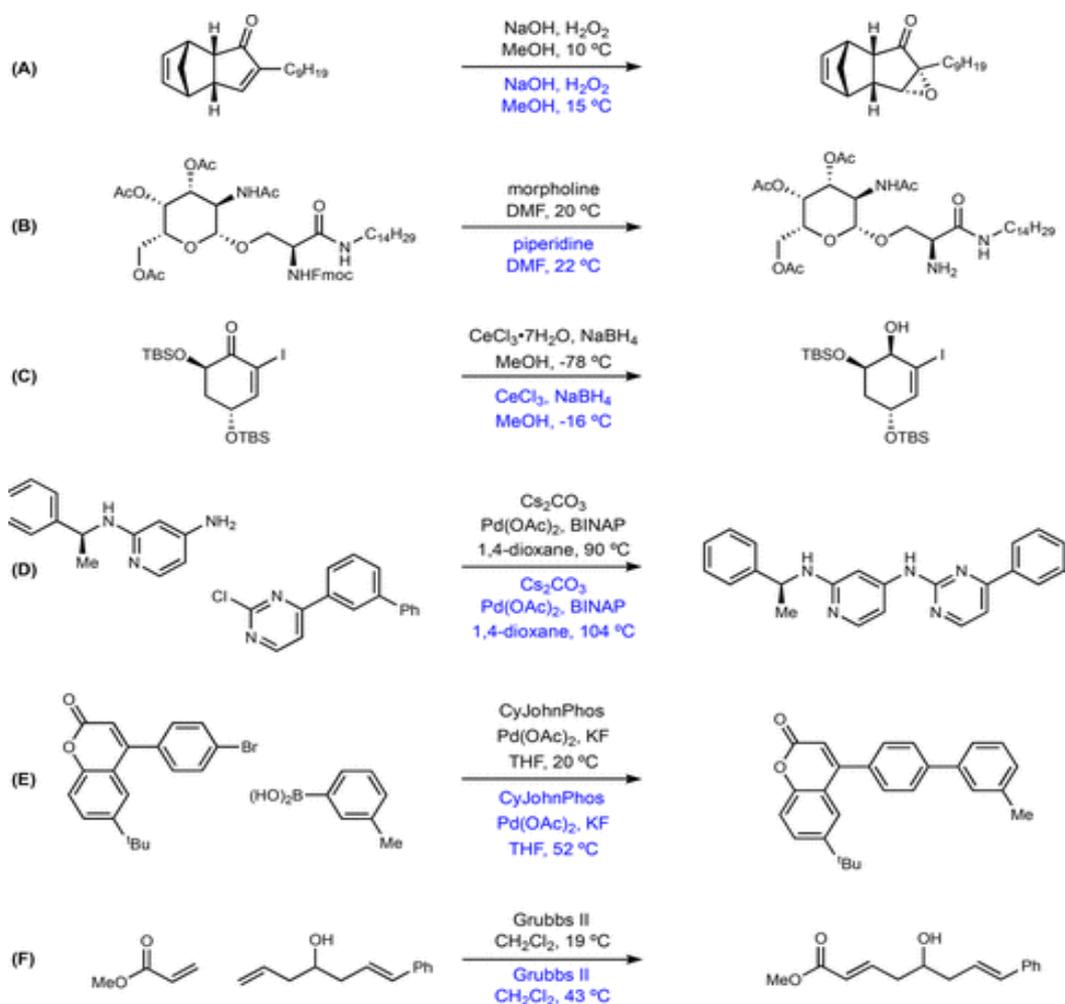


Figura 18. Predicciones de las condiciones de cada reacción llevadas a cabo con el modelo estudiado (El texto negro representa las condiciones registradas y el texto azul representa las condiciones predichas). a) Epoxidación nucleofílica, b) Reacción de desprotección de Fmoc, c) Reducción de Luche, d) Aminación de arilo de Buchwald-Hartwig, e) Reacción de acoplamiento Suzuki-Miyaura y f) Metátesis de Grubbs (Gao et al., 2018).

Analizando los resultados obtenidos en cada reacción (Figura 18) podemos comprobar que: el modelo predice correctamente las condiciones para la reacción a); en la reacción b) el reactivo propuesto (piperidina), es muy similar estructural y funcionalmente al reactivo registrado (morfolina), lo que demuestra la capacidad de este modelo para reconocer la similitud química (Gao et al., 2018); en la reacción c) el modelo reconoce el tipo de reacción y por tanto, la necesidad de un catalizador ácido para reducir

selectivamente sólo el grupo carbonilo, sugiriendo cerio (III) (Wang y Yu, 2015); los ligandos, átomos metálicos, base y disolventes de la reacción d) son correctamente propuestas; y por último, las condiciones de las últimas tres reacciones son correctas.

Además de las reacciones analizadas en este proyecto, se estudiaron un volumen de reacciones variado y abundante, en las cuales el modelo propuso con éxito las condiciones de reacción, salvo algunas excepciones sobre las cuales se realizó un análisis de errores.

Tras este análisis se llegó a la conclusión de que los errores cometidos no eran irracionales y se debían a problemas de calidad en los datos de entrenamiento. Por lo que el modelo descrito puede predecir condiciones exactas, o bien, predecir condiciones semejantes con la misma funcionalidad, otorgando un método más rápido, eficiente y rentable. No obstante, es necesario el perfeccionamiento de esta metodología aumentando la calidad de los datos insertados, para poder así minimizar los errores cometidos (Gao et al., 2018).

➤ **Optimización de las reacciones químicas**

Una vez que los programas informáticos nos dan la información acerca de las reacciones químicas idóneas para sintetizar las moléculas buscadas, tenemos la misión de averiguar las condiciones idóneas de cada reacción y a su vez, optimizar este proceso para obtener el máximo rendimiento. Estos procesos son muy importantes, debido a que las reacciones químicas no optimizadas conllevan grandes costes de producción y una pérdida de tiempo.

Para ello, el aprendizaje profundo nos servirá como una potente herramienta en este ámbito. Este modelo registra los resultados de las reacciones químicas y elige nuevas condiciones experimentales, como la temperatura, la presión y el tiempo de reacción para así mejorar los resultados de las reacciones elegidas.

Haciendo uso de este tipo de aprendizaje se desarrolló un modelo denominado Deep Reaction Optimizer (DRO). Este programa informático permite evaluar el rendimiento de cada reacción química, teniendo en cuenta diferentes condiciones experimentales.

Además, este modelo permite también la optimización de diversas funciones como la selectividad, la pureza o el costo (Zhou et al., 2017).

Haciendo uso de esta metodología se estudiaron las condiciones óptimas a las que debían someterse cuatro reacciones químicas, obteniéndose los resultados buscados en 30 minutos (Yan et al., 2016).

Para poner a prueba esta metodología se realizó un análisis comparativo entre estos métodos y varios algoritmos de optimización tradicionales. Como resultado de este análisis se concluyó que este método producía una mayor optimización de las reacciones químicas, reduciendo los pasos necesarios en un 71%.

Además, esta metodología demostró que el entrenamiento de las reacciones de este programa mejoraba el rendimiento obtenido, demostrando de esta forma, su capacidad de aprendizaje (Zhou et al., 2017).

5. Conclusiones

La aplicación de la IA dentro de la síntesis orgánica ha mostrado ser una herramienta muy eficaz en cada una de las fases del proceso de descubrimiento de nuevos fármacos.

Esta tecnología emergente nos permite sintetizar una gran variedad de moléculas químicas, a partir de un conjunto de moléculas ya descubiertas y conocidas. Este proceso brindará la oportunidad de superar el potencial de las moléculas ya descubiertas, obteniendo unos mejores resultados en su acción farmacológica.

A partir de aquí, proporciona un gran avance ya que, no sólo permite la generación de nuevas moléculas, sino que solventará muchos de los problemas a los que se enfrenta la industria farmacéutica. Entre ellos, la producción de estas moléculas de forma eficiente, debido a que, una vez que seleccionamos nuestra molécula objetivo, debemos de realizar una serie de procesos para averiguar el conjunto de reacciones y reactivos necesarios para producir estas moléculas.

Gracias a todas estas técnicas anteriormente descritas, este sistema permitirá no solo las mejoras anteriores, sino también la optimización de todas las reacciones y todas las condiciones necesarias para que se lleven a cabo estas reacciones.

La optimización de todas estas fases empleando las metodologías mencionadas supone, no sólo un enorme avance para la ciencia, sino también la prueba de que la investigación está generando cada vez programas informáticos más avanzados y precisos, con unos mejores resultados, poniendo de manifiesto la necesidad de continuar con la investigación dentro de este ámbito.

No podemos conformarnos con los avances que ya tenemos, ya que el éxito es fruto de la persistencia, y hoy en día hemos logrado esta evolución gracias a la constancia de los investigadores y su creencia de que siempre hay algo mejor, algo nuevo que permitirá mejorar la calidad de vida de las personas, en definitiva, una razón por la que seguir investigando.

6. Bibliografía

- Avila-Tomás J.F, Mayer-Pujadas M.A, Quesada-Varela V.J. La inteligencia artificial y sus aplicaciones en medicina I: introducción antecedentes a la IA y robótica. Atención primaria. 2020; 52(10): 778-784.
- Bajorath J. Deep machine learning for computer-aided drug design. Frontiers in drug discovery. 2022; 2.
- Butler K.T, Davies D.W, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018; 559(7715): 547-555.
- Cao S, Lu Wei, Xu Q. Deep neural networks for learning graph representations. Proceedings of the AAAI conference on artificial intelligence. 2016; 30(1): 1145-1152.
- ChemIntelligence SAS. AI for chemistry: a concise state of the art [en línea]. [Consultado en Mayo de 2023]. Disponible en: <https://chemintelligence.com/ai-for-chemistry>
- Christ C.D, Zentgraf M, Kriegl J.M. Mining electronic laboratory notebooks: analysis, retrosynthesis, and reaction based enumeration. Journal of chemical information and modeling. 2012; 52(7): 1745-56.
- Coley C.W, Barzilay R, Jaakkola T.S, Green W.H, Jensen K.F. Prediction of organic reaction outcomes using machine learning. ACS Central Science. 2017; 3(5): 434-443.
- David L., Thakkar A, Mercado R., Engkvist O. Molecular representations in AI-driven drug discovery: a review and practical guide. Journal of Cheminformatics (BioMed Central). 2020; 12(1):56.
- De Cao N, Kipf T. MolGAN: an implicit generative model for small molecular graphs. ICML workshop on theoretical foundations and applications of deep generative models. 2018; 1.
- Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. Advances in neural information processing systems. 2016; 1(1): 3844-3852.

- Gao H, Struble T.J, Coley C.W, Wang Y, Green W.H, Jensen K.F. Using machine learning to predict suitable conditions for organic reactions. ACS Central Science. 2018; 4(11): 1465-1476.
- García U. Introducción a las redes neuronales pt.1. Future lab [en línea]. [Consultado en mayo de 2023]. Disponible en: <https://futurelab.mx/redes%20neuronales/inteligencia%20artificial/2019/09/07/intro-a-redes-neuronales-pt-2/>
- GeeksforGeeks. Multi-Layer perceptron learning in tensorflow [en línea]. [Consultado en mayo de 2023]. Disponible en: <https://www.geeksforgeeks.org/multi-layer-perceptron-learning-in-tensorflow/>
- Goh G.B, Siegel C, Vishnu A, Hodas N.O, Baker N. Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models. ArXiv. 2017
- Gómez-Bombarelli R, Wei J.N, Duvenaud D, Hernández-Lobato J.M, Sánchez-Lengeling B, Sheberla D et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Central Science. 2018; 4(2), 268-276.
- Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains. Proceedings IEEE international joint conference on neural networks. 2005; 2: 729-734.
- Gupta A, Müller A.T, Huisman B.J, Fuchs J.A., Schneider P, Schneider G. Generative recurrent networks for de novo drug design. Molecular informatics. 2017; 37(1-2).
- Hähnke V.D, Bolton E.E. Bryant S.H. PubChem atom environments. Journal of cheminformatics. 2015; 7(41): 1-37.
- Hornos L. Introducción a SMILES: dibujando moléculas en el bloc de notas. Medium. 2020. [Consultado en abril de 2023]. Disponible en <https://medium.com/@hexagono/introducci%C3%B3n-a-smiles-dibujando-mol%C3%A9culas-en-el-bloc-de-notas-3c0df0d09c01>
- Intel shapes the future of technology. Redes neuronales convolucionales (CNN), aprendizaje profundo y visión informática [en línea]. [Consultado en mayo de

- 2023]. Disponible en: <https://www.intel.es/content/www/es/es/company-overview/company-overview.html>
- Jiang Y, Yu Y, Kong M, Mei Y, Yuan L, Huang Z et al. Artificial intelligence for retrosynthesis prediction. *Engineering*. 2022; 24(5).
 - Li Y, Tarlow D, Brockschmidt M, Zemel R. Gated graph sequence neural networks. *ICLR*. 2016; 1.
 - Lin K, Xu Y, Pei J, Lai L. Automatic retrosynthetic route planning using template-free models. *Chemical science*. 2020; 11(12): 3355-3564.
 - Marcou G, Aires de Sousa J, Latino D.A.R.S, De Luca A, Horvath D, Rietsch V et al. Expert system for predicting reaction conditions: the Michael reaction case. *Journal of chemical information and modeling*. 2015; 55(2): 239-250.
 - Matsingos C, Urdaneta A.M, Hernández J.C, Peña Silva R.A. Aplicaciones de la inteligencia artificial en la farmacología básica y clínica. *Medicina*. 2021; 43(4): 652-667.
 - Méndez-Lucio O, Baillif B, Clevert D.A., Rouquié D, Wichard J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nature communications*. 2020; 11(10).
 - Micheli A. Neural network for graphs: a contextual constructive approach. *IEEE trans neural netw*. 2009; 20 (3): 498-511.
 - Miljković F, Rodríguez-Pérez R, Bajorath J. Impact of artificial intelligence on compound discovery, design, and synthesis. *ACS Omega*. 2021; 6(49):33293-33299.
 - Morrison D. The good, the bad and the algorithmic: what impact could artificial intelligence have on political communications and democracy? The forum network. 2023. [Consultado en mayo de 2023]. Disponible en: <https://www.oecd-forum.org/posts/the-good-the-bad-and-the-algorithmic-what-impact-could-artificial-intelligence-have-on-political-communications-and-democracy>
 - Munera Salazar L.E. Inteligencia artificial y sistemas expertos. *Icesi*. 2010; (38).

- Naveja-Romero J.J, Saldívar-González F.I, Prado-Romero D.L, Ruiz-Moreno A.J, Velasco-Velázquez M, Miranda Quintana R.A, et al. ViSAS for entering chemical space: virtual screening of analog series and related advances. ChemRxiv. 2022.
- Pathak A. Redes neuronales convolucionales (CNN): una introducción. Geekflare. 2022. [Consultado en abril de 2023]. Disponible en: <https://geekflare.com/es/convolutional-neural-networks/>
- Rashad F. Adversarial auto encoder (AAE). ViTrox. 2020. [Consultado en abril de 2023]. Disponible en: <https://medium.com/vitrox-publication/adversarial-auto-encoder-aae-a3fc86f71758>
- Rocca J. Understanding variational autoencoders (VAEs). Towards data science. 2019. [Consultado en abril de 2023]. Disponible en: <https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73>
- Russell S, Norvig P. Inteligencia artificial: un enfoque moderno. 2ª ed. Madrid: pearson educación; 2004.
- Saeed M. An Introduction to recurrent neural networks and the math that powers them. Machine learning mastery. 2022. [Consultado en abril de 2023]. Disponible en: <https://machinelearningmastery.com/an-introduction-to-recurrent-neural-networks-and-the-math-that-powers-them/>
- Scarselli F, Gori M, Tsoi A.C, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE transactions on neural networks. 2008; 20(1): 61-80.
- Segler M, Preuss M, Waller M. Planning chemical syntheses with deep neural networks and symbolic AI. Nature. 2018; 555 (7698). 604-610.
- Shuman D.I, Narang S.K, Frossard P, Ortega A, Vandergheynst P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE signal processing magazine. 2012; 30(3): 83-98.
- Tubert Brohman E. Síntesis orgánica asistida por computadora (Tesis doctoral). Monterrey: universidad Virtual del Instituto Tecnológico y de Estudios Superiores de Monterrey; 2001.

- Ucak U.V, Ashyrmamatov I, Ko J, Lee J. Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments. *Nature communications*. 2022; 13(1): 1186.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A et al. Attention is all you need. In *advances in neural information processing systems*. ArXiv. 2017.
- Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. ArXiv. 2017; 1.
- Wang H, Yu S. Synthesis of isoquinolones using visible-light-promoted denitrogenative alkyne insertion of 1, 2, 3-benzotriazinones. *Organic letters*. 2015; 17(17): 4272-4275.
- Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition. ArXiv. 2018; (912): 7444-7452.
- Yan X, Bain R.M, Cooks R.G. Organic reactions in microdroplets: reaction acceleration revealed by mass spectrometry. *ACS Cent Sci*. 2016; 55(42): 12960-12972.
- Yang X, Wang Y, Byrne R, Schneider G, Yong Yang S. Concepts of artificial intelligence for computer-assisted drug discovery. *Chemical reviews*. 2019; 119(18): 10520-10594.
- Zeng X, Xiang H, Yu L, Wang J, Li K, Nussinov R et al. Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nature machine intelligence*. 2022; 4(11): 1004-1016.
- Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z et al. Graph neural networks: a review of methods and applications. *KeAi*. 2020; 1: 57-81.
- Zhou Z, Li X, Zare R.N. Optimizing chemical reactions with deep reinforcement learning. *ACS Cent Sci*. 2017; 3(12): 1337-1344.