



## **FACULTAD DE FARMACIA**

Grado en Farmacia

# **INTELIGENCIA ARTIFICIAL EN EL DESARROLLO DE TERAPIAS CONTRA EL CÁNCER DE PULMÓN. REVISIÓN SISTEMÁTICA DE ALCANCE.**

Memoria de Trabajo Fin de Grado

Sant Joan d'Alacant

Junio 2024

Autora: Elena Carmona Lozoya

Modalidad: Revisión bibliográfica

Tutor/es: Jose Antonio Picó Monllor

## ÍNDICE

RESUMEN .....	1
INTRODUCCIÓN .....	3
OBJETIVO.....	8
METODOLOGÍA.....	9
RESULTADOS.....	11
DISCUSIÓN .....	17
CONCLUSIÓN .....	24
ABREVIATURAS.....	25
BIBLIOGRAFÍA .....	27



## **RESUMEN**

**Antecedentes:** La inteligencia artificial (IA) ha revolucionado significativamente la sociedad contemporánea, su influencia abarca una gran variedad de campos, y la investigación clínica no es una excepción, así como, el diseño y descubrimiento de nuevos fármacos. El cáncer de pulmón es considerado uno de los cánceres con mayor morbi-mortalidad y el empleo de la IA proporciona un nuevo enfoque para abarcar las dificultades que presenta el tratamiento a nivel farmacológico de esta enfermedad.

**Objetivo:** El objetivo de esta revisión es conocer los conceptos básicos del funcionamiento de la inteligencia artificial y sus aplicaciones en el desarrollo terapéutico y el tratamiento del cáncer de pulmón.

**Métodos:** Se recuperaron 130 referencias de diferentes bases de datos, de las cuales se seleccionaron 12 artículos para un análisis más profundo.

**Conclusión:** Los modelos de aprendizaje automático utilizados en los estudios han demostrado ser eficaces en sus predicciones, aunque todos ellos señalan la importancia de validar estos hallazgos en estudios posteriores.

**Palabras clave:** inteligencia artificial, cáncer de pulmón, desarrollo de fármacos, aprendizaje automático, cáncer de pulmón de células no microcíticas.

## **ABSTRACT**

**Background:** Artificial intelligence (AI) has significantly revolutionized contemporary society, extending its influence across a wide variety of fields, including clinical research and the design and discovery of new drugs. Lung cancer is considered one of the neoplasms with the highest morbidity and mortality rates, and the use of AI provides a new approach to addressing the challenges presented by the pharmacological treatment of this disease.

**Objective:** The main objective of this review is to understand the basic concepts of how artificial intelligence works and its applications in the therapeutic development and treatment of lung cancer.

**Methods:** 130 references were retrieved from several databases, out of which 12 articles were selected for a more rigorous understanding.

**Conclusion:** The machine learning models used in the studies have proven to be effective in their predictions, although all of them highlight the importance of validating these findings in subsequent studies.

**Keywords:** Artificial intelligence, lung cancer, drug development, machine learning, non-small cell lung cancer.

## INTRODUCCIÓN

El cáncer es considerado una de las enfermedades con mayor mortalidad a nivel mundial. Según los datos del *Global Cancer Observatory* (GCO), elaborado por la *International Agency for Research on Cancer* (IARC) de la Organización Mundial de la Salud (OMS), se estima que en 2022 se produjeron casi 10 millones de muertes por cáncer. Principalmente, debido al cáncer de pulmón con 1,8 millones de fallecimientos, seguido del cáncer de colon con 900 mil de defunciones <sup>(1)</sup>. Concretamente, en el cáncer de pulmón, la supervivencia de los pacientes a los 5 años de diagnóstico no supera el 20%, ya que la gran parte de los casos se diagnostican cuando existe un proceso de metástasis, es decir, en etapas avanzadas del cáncer (estadios III-IV), donde las opciones de tratamiento curativas son prácticamente inexistentes <sup>(2,3)</sup>.

El cáncer de pulmón se clasifica en dos subtipos según su histopatología: cáncer de pulmón no microcítico o de células no pequeñas (*non-small cell lung cancer*, NSCLC) y cáncer de pulmón microcítico o de células pequeñas (*small cell lung cancer*, SCLC), siendo este último el de mayor malignidad, ya que tiene una elevada capacidad de metástasis. La gran mayoría de los casos se diagnostican como NSCLC, incluyendo el adenocarcinoma, el carcinoma de células escamosas, el carcinoma de células grandes y el adenocarcinoma escamoso. El diagnóstico se realiza principalmente a través de tomografía computarizada (TC) y biopsia del tejido, aunque en ocasiones resulta de gran dificultad para los oncólogos asociar los nódulos pulmonares en etapas tempranas con futuros cambios neoplásicos <sup>(2)</sup>.

En el tratamiento del cáncer de pulmón existen distintas opciones terapéuticas: la radioterapia, la cirugía y el tratamiento farmacológico, siendo la cirugía la principal alternativa para curar la enfermedad <sup>(4)</sup>. En cuanto a la farmacoterapia, la quimioterapia puede ser de inducción o neoadyuvante, para reducir en primera instancia el tamaño del tumor; o adyuvante, de forma complementaria tras la cirugía. En la quimioterapia adyuvante el tratamiento de elección es el cisplatino <sup>(5)</sup>, que junto con la radioterapia ofrecen un aumento de la supervivencia de los pacientes. Sin embargo, en pacientes con un cáncer

avanzado, no suele ser suficiente y se recurre a la introducción de terapias dirigidas molecularmente e inmunoterapia.

La terapia dirigida actúa específicamente sobre aquellas moléculas que participan en el crecimiento, el avance y la propagación del tumor. En el caso del cáncer NSCLC, se emplean inhibidores de la tirosina quinasa (*tyrosine kinase inhibitor*, TKI) para aquellos genes que han sufrido algún tipo de mutación, como EGFR, ALK o ROS-1 <sup>(6)</sup>. Por otro lado, la inmunoterapia potencia la capacidad del sistema inmune para reconocer y destruir las células tumorales, bloqueando moléculas como CTLA-4 o PD-1/PD-L1, que intervienen en la resistencia inmune adaptativa. En condiciones fisiológicas normales, estas moléculas se encargan de frenar la respuesta del sistema inmune para evitar la autoreactividad. Para su bloqueo se emplean inhibidores de puntos de control inmunitario (*immunecheckpoint inhibitors*, ICI), como los anticuerpos monoclonales <sup>(7)</sup>.

Tabla 1. Resumen de las dianas moleculares y fármacos empleados en terapia dirigida e inmunoterapia en cáncer de pulmón <sup>(6,7)</sup>.

<b>Terapia dirigida</b>	
Gen del receptor del factor de crecimiento epidérmico (EGFR)	Erlotinib o gefitinib (1ª gen), afatinib o dacomitinib (2ª gen), osimertinib (3ª gen), o erlotinib + ramucirumab.
Quinasa del linfoma anaplásico (ALK)	Crizotinib (1ª gen), ceritinib o alectinib (2ª gen)
BRAF	Dabrafenib/trametinib
Genes de fusión ROS1	Crizotinib, lorlatinib, repotrectinib, entrectinib
<b>Inmunoterapia</b>	
Proteína 4 asociada a linfocitos T citotóxicos (CTLA-4)	Ipilimumab
Inhibidores de PD-1/PD-L1	Nivolumab, pembrolizumab (antiPD-1) Atezolizumab, durvalumab (antiPD-L1)

Sin embargo, a pesar de los numerosos avances de las últimas décadas en la identificación de nuevos objetivos terapéuticos y el desarrollo de nuevos fármacos, el pronóstico de los tratamientos convencionales en etapas avanzadas es desfavorable <sup>(8)</sup>. Además, la heterogeneidad de los tumores, la aparición de metástasis, la ubicación de los tumores, la relación compleja entre el sistema inmune, las células tumorales y el microambiente tumoral y el alto costo de las pruebas diagnósticas, pueden dificultar la efectividad de la terapia antineoplásica. Todo ello hace que el cáncer de pulmón siga liderando las estadísticas de mortalidad a nivel mundial.

Los avances en la tecnología, como la inteligencia artificial (IA), están abriendo nuevas líneas de investigación, ofreciendo así nuevas perspectivas que ayuden a dilucidar el abordaje de esta patología. En el desarrollo de fármacos, la IA promete mejorar los resultados en investigación y desarrollo de los mismos mediante la identificación de nuevas dianas terapéuticas, la selección de fármacos candidatos y su reutilización, la predicción de estructuras de proteínas, el diseño de compuestos y desarrollo de nuevos biomarcadores predictivos u optimización de ensayos clínicos. En el cáncer de pulmón, resulta de gran interés la identificación de biomarcadores de predicción temprana, ya que aumentaría la probabilidad de éxito a los tratamientos, aumentando así la supervivencia general <sup>(9)</sup>. Además, la respuesta a los tratamientos innovadores de TKI e ICI han demostrado una falta de efectividad en algunos pacientes, por lo que buscar biomarcadores que predigan qué casos presentarán una mayor respuesta, también es un punto interesante que abordar <sup>(10)</sup>.

A través de diferentes algoritmos computacionales y en función de los datos disponibles, la inteligencia artificial es capaz de crear un razonamiento de forma lógica que simule y amplíe el pensamiento de un cerebro humano. Este enfoque se emplea para implementar tareas de clasificación, agrupación y predicción. No obstante, se requieren grandes cantidades de datos (conocidos como *big data*) para la construcción de los modelos para que mediante el entrenamiento y la validación, sea posible identificar problemas de sobreajuste o subajuste en la optimización de dichos modelos y garantizar su fiabilidad. El sobreajuste se produce cuando el modelo se ajusta únicamente a los datos de

entrenamiento y pierde su capacidad de generalización. Por el contrario, el subajuste surge cuando el modelo, debido a su falta de especialización o excesiva simplicidad, no logra buenos resultados en sus predicciones <sup>(11)</sup>.

El aprendizaje automático (ML o *Machine Learning*) es una rama de la inteligencia artificial que trata de identificar patrones de clasificación y extraer características comunes. Entre ellos encontramos algoritmos del árbol de decisión o los algoritmos de aprendizaje profundo (DL o *Deep Learning*) <sup>(12)</sup>. Por ejemplo, el estudio de Kuenzi et al. <sup>(13)</sup>, diseñó un modelo de aprendizaje profundo llamado DrugCell que predice medicamentos eficaces y combinaciones de estos según diferentes líneas de células tumorales. Otro modelo de aprendizaje profundo desarrollado fue el trabajo de Senior et al. <sup>(14)</sup> denominado AlphaFold, el cual utiliza secuencias de aminoácidos para predecir de forma precisa el mecanismo de plegamiento de una proteína.

En el ámbito de las ciencias de la salud, la mayor parte de las aplicaciones de descubrimiento de fármacos mediante IA se centran en el aprendizaje automático, especialmente en el aprendizaje profundo. Las técnicas de ML se dividen en: aprendizaje supervisado, no supervisado y por refuerzo <sup>(15)</sup>.

- Aprendizaje supervisado: implica trabajar con un conjunto de datos que incluye la variable que se desea predecir, permitiendo entrenar el modelo con esta información.
- Aprendizaje no supervisado: el modelo se entrena con datos no clasificados y el algoritmo se encarga de identificar patrones y clasificar los datos según sus similitudes. Permite la agrupación de datos.
- Aprendizaje por refuerzo: se basa en el principio prueba y error, similar al proceso de aprendizaje en humanos y animales. El algoritmo recibe una recompensa positiva cuando su acción produce un resultado favorable, y una sanción en caso contrario. Implica aprender nuevos patrones a través de la experiencia y la retroalimentación del entorno.



Además, los algoritmos de ML se agrupan en tres modelos <sup>(16)</sup>:

- Modelos lineales. Intentan trazar una línea que se “adapte” adecuadamente a la dispersión de puntos presentados. Destacan modelos de regresión logística y regresión lineal.
- Modelos de árbol. A diferencia de los anteriores, pueden representar variable son lineales. Destacan los árboles aleatorios (RF o *Random Forest*) y los árboles de decisión.
- Redes neuronales. Son una forma de aprendizaje profundo (DL). Tratan de imitar las interacciones entre las neuronas de un cerebro humano mediante capas de entrada; donde se introducen los datos de entrenamiento, capas de salida, que representan las predicciones realizadas; y capas ocultas, que se encuentran entre estas dos capas e identifica los patrones. Destacan las redes neuronales convolucionales (CNN) o las redes generativas antagónicas (GAN).

Una de las grandes ventajas que presenta la implementación de estos modelos de inteligencia artificial en la investigación científica es la aceleración de las diferentes etapas frente a las del proceso tradicional. Según las estadísticas, el periodo medio de investigación y desarrollo (I+D) de un medicamento suele abarcar entre 10 y 17 años <sup>(8)</sup>, además de un alto costo de comercialización y un éxito limitado de los ensayos clínicos. Gracias a la capacidad para analizar grandes volúmenes de datos y automatizar tareas repetitivas, se ha observado una considerable reducción del tiempo en dichos procesos.

## **OBJETIVO**

En base a la información previamente presentada, los objetivos de esta revisión sistemática de alcance son:

- Analizar el papel de la inteligencia artificial en el desarrollo de terapias contra el cáncer de pulmón.
- Explorar posibles enfoques futuros que puedan beneficiarse de esta tecnología.



## METODOLOGÍA

### Diseño y obtención de datos:

Para la realización de la presente memoria, se revisó la literatura publicada entre 2019 hasta el 19 de abril de 2024. Algunas de las bases de datos que se revisaron vía Internet fueron: MEDLINE (vía PubMed), Scopus, EMBASE y Google Académico.

Con el objetivo de establecer los términos *Thesaurus*, se procedió a la búsqueda en DeCS (Descriptores en Ciencias de la Salud) de los descriptores que formaron parte de la ecuación de investigación y, así, obtener el MeSH (*Medical Subject Headings*). Los resultados de los descriptores seleccionados se muestran en la tabla 2.

Tabla 2. Descriptores utilizados según MeSH.

<b>Término (DeCS)</b>	<b>MeSH</b>
Inteligencia artificial	Artificial Intelligence
Cáncer de pulmón	Lung Neoplasms
Desarrollo de fármacos	Drug Development

La ecuación de búsqueda empleada en MEDLINE (PubMed) utilizando los conectores booleanos pertinentes, quedó de la siguiente forma: “(((“Artificial Intelligence”[Mesh] OR “Machine Learning”[Mesh] OR “Deep Learning”[Mesh] OR “Natural Language Processing”[Mesh] OR “Expert Systems”[Mesh]) AND (“Drug Development”[Mesh] OR “Antineoplastic Agents”[Mesh] OR “Drug Design”[Mesh])) AND (“Lung Neoplasms”[Mesh])).”

### Selección de final de los artículos:

Los artículos que se incluyeron cumplieron los siguientes criterios de inclusión: uso de modelos de aprendizaje automático en cualquiera de sus modalidades, estudios experimentales ex-vivo e in-vivo dirigidos al cáncer de pulmón, redactados en inglés o español y disponer del texto completo de forma

gratuita. Mientras que los artículos basados únicamente en el diagnóstico, así como revisiones del tema a tratar, quedaron excluidos.

Extracción de datos:

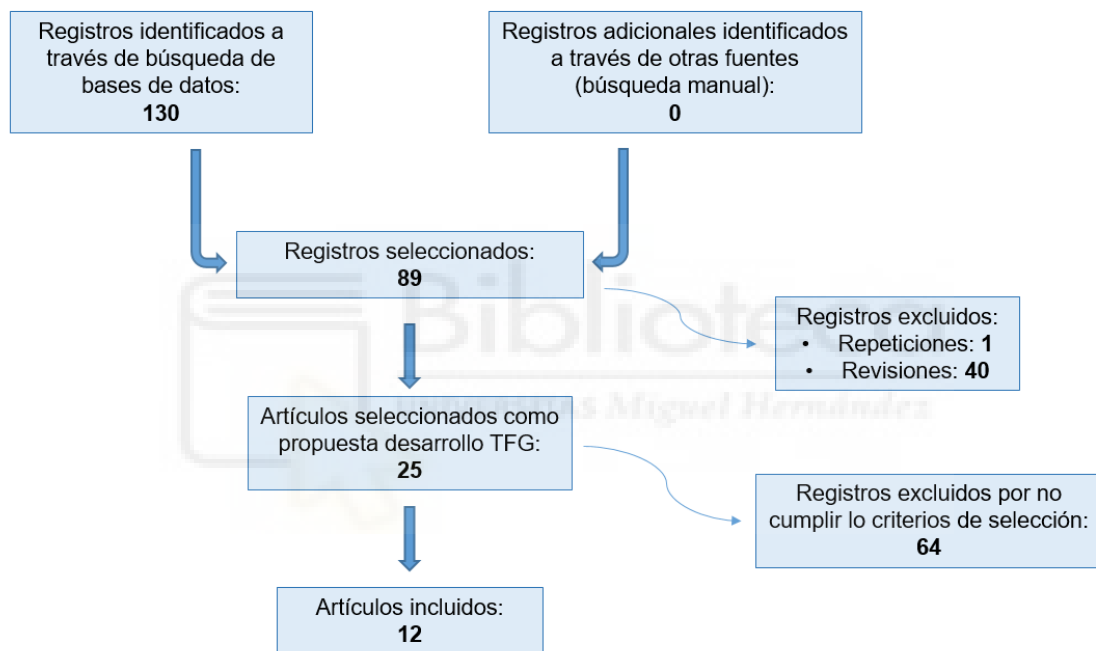
Los datos recopilados en los diferentes artículos se agruparon en la siguiente Tabla 3, con el fin de proporcionar una visión general de los resultados y facilitar la comprensión de estos, en las que se refleja: nombre del primer autor, año, tipo de estudio, tipo de aprendizaje automático, arquitectura del modelo, intervención y resultado.



## RESULTADOS

Una vez establecidos los criterios de búsqueda, se recuperaron 130 referencias: 33 en PubMed, 51 en Scopus y 46 en Embase. Tras eliminar los artículos duplicados, aplicar los criterios de inclusión y consultar los listados bibliográficos, fue posible seleccionar 12 documentos para llevar a cabo esta revisión y realizar un análisis crítico.

Figura 1. Flujograma de selección de los estudios incluidos. Modificación de la versión de la Declaración PRISMA <sup>(17)</sup>.



Finalmente, 25 artículos fueron seleccionados como propuesta para el desarrollo TFG y tras aplicar los criterios de selección e exclusión, solo 12 se incluyeron para poder hacer un análisis en profundidad de cada uno de ellos.

Tabla 3. Características principales de los estudios seleccionados.

Autor, año	Tipo de estudio	Tipo de aprendizaje automático	Arquitectura del modelo	Intervención	Resultados
Maliazurina B Saad, 2023 <sup>(18)</sup>	Observacional retrospectivo	Mixto. Basado en redes neuronales y árboles	Redes neuronales convolucionales (CNN) y bosque aleatorio	Desarrollo de un modelo basado en IA (Deep-CT) a partir de TC de 976 pacientes con NSCLC metastásico tratados con ICI.	Predicción de forma eficaz de la supervivencia general y supervivencia libre de progresión en pacientes tratados con ICI.
Mubashir Aziz, 2022 <sup>(19)</sup>	Observacional con bases de datos	Basado en redes neuronales	Red neuronal de paso de mensajes (MPNN) y combinación de red MPNN-CNN.	Realización de un cribado virtual basado en estructuras (SBVS) de 1200 análogos estructurales del dabrafenib utilizando el acoplamiento molecular y modelos de aprendizaje automático.	Los modelos predijeron mejor afinidad de unión del compuesto 762 entre todos los derivados. También predijo unas propiedades adecuadas de ADME para este compuesto.
Jiangdian Song, 2020 <sup>(20)</sup>	Observacional retrospectivo	Basado en red neuronales	Red generativa antagónica bidireccional a gran escala (BigBiGAN)	Se emplearon 6204 imágenes de TC de los pacientes de las cohortes y se introdujeron en una red neuronal.	Creación de una firma mediante IA para la estratificación de pacientes que pueden beneficiarse de EGFR-TKI.
Bingrui Li, 2020 <sup>(21)</sup>	Observacional con bases de datos	Basado en redes neuronales	Redes neuronales profundas (DNN)	Se utilizaron 3985 muestras para el análisis de 978 genes representativos y de 164 vías de señalización como respuesta celular a 75 fármacos. Las muestras fueron clasificadas por DNN en 6 categorías de uso terapéutico.	La pimozida mostró un fuerte efecto supresor en líneas celulares de NSCLC.

Elisabeth A. Coker, 2022 <sup>(22)</sup>	Observacional con bases de datos	Mixto. Basado en lineal y árboles	Red elástica y bosque aleatorio	Se midieron las respuestas dinámicas de señalización en diferentes líneas celulares de NSCLC expuestas durante una hora a una única concentración de siete fármacos anticancerígenos. Los resultados se introdujeron para el entrenamiento de diferentes modelos.	El modelo fue capaz de predecir tanto las respuestas altas como bajas a los fármacos.
Jun Cheng, 2022 <sup>(23)</sup>	Observacional retrospectivo	Basado en lineal	Regresión logística	Se emplearon tres cohortes de pacientes para diferenciación de la neumonitis por radiación y la neumonitis asociada a ICI en pacientes con NSCLC mediante diferentes algoritmos basados en IA.	El modelo fue capaz de diferenciar la procedencia de la neumonía con una eficacia estadísticamente significativa.
Han-Jing Jiang, 2020 <sup>(24)</sup>	Observacional con bases de datos	Mixto. Basado en redes neuronales y árboles.	Red neuronal profunda y bosque de rotación.	El conjunto de datos de 663 medicamentos, 409 enfermedades y 2532 asociaciones medicamento-enfermedad se introdujeron en el modelo (SAEROF).	Se identificaron 20 medicamentos asociados con neoplasias de pulmón.
Changhee Park, 2020 <sup>(25)</sup>	Observacional retrospectivo	Basado en redes neuronales	Red neuronal convolucional (CNN)	Desarrollo de un biomarcador basado en aprendizaje profundo para estimar el perfil inmunológico del adenocarcinoma.	El modelo propuesto fue capaz de predecir la actividad citolítica del tumor de manera precisa.

Fahsai Nakarin, 2022 <sup>(26)</sup>	Observacional con bases de datos	Mixto. Basado en redes neuronales	Método de agregación de vecindario principal (PNA) y red neuronal profunda (DNN)	A través de los valores experimentales de IC50 recopilados de miles de ligandos con afinidad por los receptores de tirosinas quinasas en NSCLC, se creó un modelo multitarea.	El modelo superó los modelos de referencia previos utilizados en el modelado molecular.
Konstantinos Rounis, 2021 <sup>(27)</sup>	Observacional prospectivo	Basado en lineal	Máquinas de vectores de soporte (SVM)	Determinación de características asociadas con la respuesta al tratamiento con ICI en pacientes con NSCLC mediante IA.	El modelo JADBios fue capaz de predecir la probabilidad de estilización de la enfermedad.
Francesca Grisoni, 2019 <sup>(28)</sup>	Observacional con bases de datos	Basado en redes neuronales	Redes neuronales artificiales de contrapropagación (CPANN)	Se recopiló conjuntos de datos de péptidos dirigidos a líneas celulares del cáncer de mama y de pulmón y se introdujeron en el algoritmo CPANN.	CPANN identificó patrones de secuenciación y se predijeron 14 péptidos activos para estas líneas celulares.
Geunho Choi, 2021 <sup>(29)</sup>	Observacional con bases de datos	Basado en redes neuronales	Red neuronal recurrente (RNN) y aprendizaje por transferencia (TL) y red neuronal convolucional (CNN)	Se seleccionaron 139 inhibidores de quinasas que fueron introducidos en el modelo para que este generara ligandos contra la mutación EGFR.	Se obtuvieron 360 ligandos prometedores para esta mutación.



A partir del año 2020, se observó un notable incremento en la cantidad de estudios publicados, siendo su punto máximo ese mismo año.

La gran mayoría de los artículos seleccionados tomaron como referencia el cáncer de pulmón de células no microcíticas (NSCLC), ya que es el cáncer de pulmón con mayor incidencia. Los estudios que tomaron cohortes de pacientes están diagnosticados con NSCLC <sup>(18,20,23,25,27)</sup>, mientras que en el resto de los estudios se emplean moléculas implicadas en esta enfermedad, o bien porque se encuentran incluidos en las bases de datos utilizadas por los modelos computacionales <sup>(29)</sup> o para comparar los resultados obtenidos con estas moléculas como parte de la validación de los resultados <sup>(21)</sup>. En el estudio realizado por Choi et al. <sup>(29)</sup>, se menciona la línea celular A549 asociada al cáncer de pulmón que, aunque no se especifique de forma expresa, también está relacionada con NSCLC. Sin embargo, en el estudio desarrollado por Jiang et al. <sup>(24)</sup>, se refiere a las neoplasias de pulmón como un concepto general sin entrar en especificaciones.

Los conjuntos de datos empleados en los estudios seleccionados que provenían de bases de datos públicas, utilizaron PubChem <sup>(19,24)</sup>, DrugBank <sup>(24)</sup>, *Gene Expression Omnibus* (GEO) <sup>(24,25)</sup>, *Online Mendelian Inheritance in Man* (OMIM) <sup>(24)</sup>, Cdataset <sup>(24)</sup>, *The Cancer Imaging Archive* (TCIA) <sup>(25)</sup>, ChEMBL <sup>(27,29)</sup>, BindingDB <sup>(25)</sup> y CancerCPP <sup>(27)</sup>. Algunos estudios emplearon datos pertenecientes a instituciones privadas para construir las diferentes cohortes de pacientes <sup>(18,20,23,27)</sup>.

Todos los estudios emplearon algoritmos de aprendizaje automático (ML), de los cuales nueve (75%) utilizaron modelos basados en redes neuronales <sup>(18-21,24-26,28,29)</sup>. La clasificación de las diferentes arquitecturas se realizó en base a la división descrita por A. Altuhaifa et al. <sup>(30)</sup> en su revisión. En algunos estudios se emplearon diferentes algoritmos, quedando reflejado en la Tabla 3 el modelo que mostró un mayor rendimiento y, por tanto, fue el seleccionado finalmente para el estudio. Por ejemplo, Cheng et al. <sup>(23)</sup>, entrenaron tres modelos diferentes en base a cada una de las características seleccionadas para la realización del estudio: bosque aleatorio, SVM y regresión logística. El mejor rendimiento lo logró este último. En el caso de Nakarin et al. <sup>(26)</sup>, se entrenaron diferentes

algoritmos de manera independiente utilizando los mismos datos (RF, GBRT, DNN, GAT, PNA, PNA+DNN). El modelo multimodal fue el seleccionado ya que se observó un aumento en el rendimiento tras la integración de las diferentes arquitecturas. También, Saad et al. <sup>(18)</sup>, utilizaron cuatro subredes de CNN que, posteriormente, combinaron entre sí mediante algoritmos de integración basados en árboles.

Entre los doce estudios seleccionados, siete (58%) utilizaron validación cruzada k veces, dos (17%) emplearon división de datos (cohortes de pacientes para el entrenamiento y la validación) y dos (17%) no proporcionan información acerca del método de validación. La técnica de validación cruzada más empleada fue la validación cruzada de 10 veces <sup>(21,23-28)</sup> y, uno de ellos, empleó la validación cruzada de 100 veces <sup>(22)</sup>. En este tipo de validaciones se dividen los datos en partes iguales de forma aleatoria utilizando cada una de las partes como conjunto de entrenamiento y las restantes como datos de prueba. Normalmente, los datos de prueba constituyen alrededor del 20% del total de los datos disponibles, aunque las proporciones pueden variar (90%-10% u 85%-15%). Además, cinco estudios <sup>(18,20,25,26)</sup> realizaron una validación interna y externa de los modelos, es decir, aplicaron el modelo de aprendizaje automático en datos de diferente procedencia, mientras que uno solo realiza una validación interna <sup>(23)</sup> y el resto, no mencionan nada al respecto. El entrenamiento y la validación de los modelos de inteligencia artificial son etapas clave en su desarrollo.

Para la evaluación del rendimiento de los modelos se emplean diferentes medidas: Área bajo la curva ROC (AUC) <sup>(20,22-25,27)</sup>, índice de concordancia de Antolini (índice C) <sup>(18)</sup>, puntuación F1 <sup>(19,21,24)</sup>, R<sup>2</sup> (coeficiente de determinación predictivo cuadrático) y RMSE (método estándar para calcular errores de predicción cuantitativa a partir de distancias normalizadas entre los valores observados y los valores predichos) <sup>(19,26)</sup> y la tasa de no error (NER) <sup>(29)</sup>. Estos datos añaden valor estadístico a los resultados obtenidos. La evaluación de los modelos se centra en las métricas derivadas de los datos de entrenamiento, que reflejan cómo responde el modelo a los datos de entrada. Por otro lado, la capacidad predictiva del modelo se evalúa utilizando el conjunto de datos de validación.

## DISCUSIÓN

Tras una revisión de la literatura en relación a la inteligencia artificial en el campo de desarrollo de nuevas terapias en el cáncer de pulmón, los artículos hallados abarcan diferentes áreas de estudio que tiene especial relevancia en este proceso.

Tabla 4. Resumen de los modelos de aprendizaje automático utilizados en los artículos incluidos.

Saad et al. <sup>(18)</sup>	Cuatro redes neuronales convolucionales (CNN) integradas con un algoritmo de bosque aleatorio.
Aziz et al. <sup>(19)</sup>	Red neuronal de paso de mensajes (MPNN) y otra red compuesta de MPNN y CNN.
Song et al. <sup>(20)</sup>	Red generativa antagónica bidireccional a gran escala (BigBIGAN)
Li et al. <sup>(21)</sup>	Redes neuronales profundas (DNN)
Coker et al. <sup>(22)</sup>	Combinación de red elástica y bosque aleatorio
Cheng, <sup>(23)</sup>	Regresión logística
Jiang et al. <sup>(24)</sup>	Combinación de una red neuronal profunda (DNN) y bosque aleatorio.
Park et al. <sup>(25)</sup>	Red neuronal convolucional (CNN)
Nakarin et al. <sup>(26)</sup>	Combinación de método de agregación de vecindario principal (PNA) y red neuronal profunda (DNN)
Rounis et al. <sup>(27)</sup>	Máquinas de soporte de vectores (SVM)
Grisoni et al. <sup>(28)</sup>	Redes neuronales de contrapropagación (CPANN)
Choi et al. <sup>(29)</sup>	Aprendizaje por transferencia (TL) basada en un red neuronal recurrente (RNN) y otra red neuronal convolucional (CNN)

La inmunoterapia y la terapia dirigida han demostrado una mejora notable en la supervivencia de los pacientes con cáncer de pulmón. Sin embargo, algunos pacientes no responden de forma eficaz al tratamiento con estos fármacos. Los estudios de Saad et al.<sup>(18)</sup> y Song et al.<sup>(20)</sup> permitieron extraer características relevantes de las imágenes médicas para la predicción de la respuesta al tratamiento con inhibidores de puntos de control (ICI) o inhibidores de quinasas (TKI). Ambos emplean algoritmos de aprendizaje profundo a partir

de imágenes de tomografía computarizada (TC): Deep-CT en la predicción de respuesta a ICI y BigBIGAN en la predicción de respuesta a TKI, en concreto a los TKI del receptor de crecimiento epidérmico (EGFR-TKI). Deep-CT está compuesto por diferentes subredes convolucionales (CNN), mientras que BigBIGAN, es una red antagónica regenerativa (GAN) de última generación. En otras revisiones realizadas acerca del empleo de los algoritmos de CNN <sup>(31)</sup> y GAN <sup>(32)</sup> para el análisis de imágenes médicas, han demostrado su potencial para mejorar tanto la interpretación de imágenes como la representación y clasificación de las mismas.

Además, Rounis et al. <sup>(27)</sup> implementaron a su estudio JADBio, un modelo de máquinas de soporte (SVM) para aportar información adicional acerca de las características asociadas con una mejor respuesta a la inmunoterapia dirigida al eje PD-1/PD-L1 en pacientes que se estudiaron previamente de manera independiente. JADBio identificó cinco de las características clínicas (administración prolongada a antibióticos, metástasis óseas y hepáticas e IMC < 25 kg/m<sup>2</sup>) y creó una firma que predijo la estabilización de la enfermedad tras la administración de ICI con una precisión del 81% (AUC = 0,806). Estos modelos propuestos han mostrado el potencial de la inteligencia artificial en la predicción de la eficacia de los medicamentos y sus enfoques pueden servir como punto de partida para investigaciones futuras.

También, el estudio realizado por Cheng et al. <sup>(23)</sup>, propone un modelo de regresión logística que utiliza características de BoW de las imágenes computarizadas (TC) para diferenciar entre la neumonitis como reacción adversa del tratamiento con ICI y la neumonitis causada a raíz de la radiación que reciben algunos pacientes de manera concomitante para el tratamiento de NSCLC. Aunque se trate de una reacción adversa muy poco común a este tipo de fármacos, es potencialmente peligrosa <sup>(33)</sup> y un diagnóstico más certero permitiría una mejora del manejo clínico de estos pacientes, principalmente en cuanto al mantenimiento o la retirada de ICI. Se entrenaron diferentes modelos (SVM, regresión logística y bosque aleatorio) que incorporaron TC de grupos pacientes que recibieron solo inmunoterapia (ICI) y solo radioterapia (RT). Una vez optimizados, el modelo de mayor rendimiento, que fue el de regresión logística, se aplicó en el conjunto de datos de los pacientes que recibieron tanto ICI como

RT (ICI+RT), mostrando una capacidad de discriminación con una eficacia del 89%. Estos tres mismos modelos también fueron entrenados en el estudio de Zheng et al. <sup>(34)</sup> para analizar las diferentes características de TC de tumores paratiroides benignos y malignos. Sin embargo, en esta ocasión fue el modelo SVM el que mostró una mayor estabilidad frente a los demás. Este estudio sugiere que los datos para este tipo de tumores podrían ser no lineales o ser linealmente inseparables<sup>1</sup>, lo que explicaría la menor eficiencia del modelo de regresión logística al ser un algoritmo lineal.

En el estudio llevado a cabo por Aziz et al. <sup>(19)</sup>, se propone el gen NEK7 como futuro objetivo en el tratamiento del NSCLC. Para ello, se llevó a cabo un cribado virtual basado en estructuras (SBVS)<sup>2</sup> de derivados del dabrafenib y, se empleó una red neuronal de paso de mensajes combinada con una red convolucional (MPNN-CNN) para predecir la afinidad de unión y una MPNN para la predicción de los perfiles farmacológicos de absorción, distribución, metabolismo y excreción (ADME) de aquellos compuestos que mostraron mejores propiedades de reactividad química tras la detección virtual. Los resultados de la predicción respaldaron los hallazgos de la detección virtual. En la revisión llevada a cabo por Carpenter et al. <sup>(35)</sup> indica que las redes neuronales se encuentran entre los mejores clasificadores para el cribado virtual durante la fase inicial en la obtención de nuevos compuestos químicos.

Los estudios de Nakarin et al. <sup>(26)</sup> y Choi et al. <sup>(29)</sup> se centraron en la búsqueda de nuevos ligandos que actúan modificando la respuesta de los receptores tirosinas quinasas que han demostrado en estudios previos intervenir en el NSCLC, como EGFR, ALK o ROS1. Este último solo lo hizo para el receptor EFGR. Ambos extrajeron moléculas de la base de datos de ChEMBL que tuvieran afinidad para estos receptores, y así introducirlos en los modelos correspondientes empleados en cada estudio. El modelo multitarea con una arquitectura DNN + PDA de Nakarin et al., predijo la capacidad inhibitoria frente a las TKI estudiadas, excepto para la TKI ERBB4, donde el rendimiento del modelo no alcanzó los estándares deseados. En el caso del modelo usado por

---

<sup>1</sup> En términos matemáticos, cuando se habla de datos “no lineales o linealmente inseparables”, se hace referencia a la dificultad de trazar una línea que separe los datos en diferentes grupos.

<sup>2</sup> El cribado virtual basado en estructuras utiliza la estructura 3D de los compuestos para identificar aquellas moléculas químicas que tienen una mayor probabilidad de unirse al objetivo.

Choi et al., el modelo de aprendizaje por transferencia (TL) basado en una estructura de red neuronal recurrente (RNN) fue capaz de generar más de 10 mil SMILES<sup>3</sup> asociados con moléculas TKI a partir de los datos introducidos previamente en el modelo. Además, en este estudio también se empleó otro modelo de aprendizaje profundo, llamado DeepDTA, para predecir las afinidades de unión de los compuestos que continuaron el estudio según los criterios establecidos, obteniendo finalmente 360 ligandos con propiedades prometedoras para el gen objetivo. A día de hoy, existen numerosos modelos de aprendizaje profundo que utilizan diferentes bases de datos para la predicción de la afinidad de unión proteína-ligando <sup>(36)</sup>.

Tanto Aziz et al. <sup>(19)</sup> como el de Li et al. <sup>(21)</sup>, emplearon la reutilización de fármacos como concepto base para la realización de sus estudios. El primer trabajo empleó las estructuras de bencenosulfamida derivadas del dabrafenib como referencia, mientras que el estudio de Li et al. <sup>(21)</sup> utilizó redes neuronales profundas (DNN) para la clasificación de 75 fármacos basándose en sus perfiles transcripcionales en seis categorías diferentes. Posteriormente, compararon la similitud de estructura química de estos fármacos con medicamentos aprobados por la Administración de Alimentos y Medicamentos de los Estados Unidos (*Food and Drug Administration*, FDA) para el NSCLC. La pimozida exhibió un efecto inhibitor en las vías asociadas con NSCLC, y los ensayos in vitro realizados con líneas celulares de este tipo de cáncer respaldaron este hallazgo. Además, en el estudio de Jiang et al. <sup>(24)</sup> el modelo de última generación que combina un codificador automático disperso y un bosque de rotación (SAEROF) consiguió predecir 20 medicamentos asociados con las neoplasias de pulmón, de los cuales, al compararlos con la base de datos CTD, 16 demostraron esta asociación. Entre estos se encontraban la piridoxina, el etopósido o el felbamato.

Si bien la reutilización de fármacos permite acortar el tiempo y los costos asociados con el desarrollo de nuevos medicamentos, la síntesis de novo de moléculas ofrece la posibilidad de diseñar compuestos específicamente dirigidos a una diana terapéutica particular, lo que puede aumentar la eficacia y reducir la citotoxicidad. El estudio de Grisoni et al. <sup>(28)</sup>, emplea redes neuronales de

---

<sup>3</sup> La nomenclatura SMILES (*Simplified Molecular Input Line Entry Specification*) es un método para representar estructuras químicas mediante secuencias de caracteres alfanuméricos en formato ASCII.

contrapropagación (CPANN) para diseñar péptidos anticancerígenos membranolíticos (membranolytic anticancer peptides, ACPs) contra la línea celular A549 del NSCLC. Se seleccionaron 14 péptidos para su síntesis y pruebas in vitro, de los cuales 6 mostraron propiedades anticancerígenas y 5, fueron efectivos contra esta línea celular. Paralelamente, Dauparas et al. <sup>(37)</sup> describieron un modelo basado en redes neuronales de paso de mensajes (MPNN) llamada ProteinMPNN para el diseño de nuevas secuencias de proteínas, cuyos resultados fueron los esperados.

Los biomarcadores basados en el genotipo se utilizan como predictores de sensibilidad a la terapia en el NSCLC, permitiendo la selección del tratamiento más adecuado en la clínica. Sin embargo, estos biomarcadores presentan limitaciones y el estudio de nuevos biomarcadores podría ayudar a entender el contexto tisular <sup>(23)</sup>. En los estudios de Coker et al. <sup>(22)</sup> y Park et al. <sup>(25)</sup>, se plantea la necesidad de encontrar nuevos biomarcadores para el abordaje del cáncer de pulmón. El primero de ellos, mide los cambios en la fosforilación de 52 fosfoproteínas presentes tanto en líneas celulares como en muestra de biopsias de pacientes con NSCLC en respuesta a siete fármacos anticancerígenos (gefitinib, trametinib, pictilisib, capivasertib, everolimus, vemurafenib y luminespib). Se desarrollaron métodos de aprendizaje automático para predecir la respuesta a los medicamentos y la efectividad de las combinaciones. Los métodos de aprendizaje automático predijeron con éxito qué células serían más o menos sensibles, con una precisión del 79% y 78% respectivamente. En cambio, las predicciones basadas en mutaciones genéticas comunes (EGFR, KRAS y PIKEC) no tuvieron éxito en predecir la sensibilidad a los medicamentos. En el segundo estudio <sup>(25)</sup> mencionado, se desarrolla un biomarcador que refleja el microambiente tumoral del adenocarcinoma de pulmón. El modelo se entrenó utilizando imágenes de tomografía por emisión de positrones con fluorodesoxiglucosa (FDG-PET) y secuenciación de ARN para predecir la actividad citolítica (CytAct), que se asocia a la actividad citotóxica de las células T CD8+ contra el tumor. Se seleccionó CytAct mínimo como medida representativa de cada paciente experimentales ya que mostró un desempeño significativamente mayor (AUC = 0,88), en comparación con CytAct medio y máximo (AUC = 0,54 y 0,74, respectivamente). De manera similar al estudio de



Park et al., Zeng et al. <sup>(38)</sup> desarrollaron un modelo de aprendizaje profundo utilizando la secuenciación de ARN para definir una firma de expresión genética en el cáncer.

Por último, cabe mencionar el número de artículos recopilados que llevan a cabo estudios incorporando bases de datos <sup>(19,21,22,24,26,28,29)</sup> en relación a los estudios que emplean cohortes de pacientes <sup>(18,20,23,25,27)</sup>, debido a la novedad del tema. Los estudios experimentales proporcionan una base sobre la que probar y validar nuevas técnicas y algoritmos de IA, así como para explorar nuevos enfoques en el desarrollo de nuevas terapias contra el cáncer de pulmón. Sin embargo, es esencial la realización de más estudios observacionales para proporcionar valiosa información a partir de la recopilación y análisis de datos existentes para esta patología, lo que contribuiría a una mejor comprensión de patrones, tendencias y correlaciones en grandes conjuntos de datos por parte de los modelos.





## **LIMITACIONES**

Además de las limitaciones inherentes a cada estudio per se, esta revisión tiene algunas restricciones que deben ser consideradas.

Al tratarse de un tema novedoso, en primera instancia fue difícil establecer una ecuación de búsqueda que fuera adecuada para resolver el objetivo de la investigación. Asimismo, debido a la heterogeneidad de los artículos, resultó complicado la comparación entre sí y establecer unas ideas claras y comunes a todos ellos.

Cabe mencionar la complejidad para una correcta comprensión del funcionamiento de los modelos analizados, ya que se encuentra fuera de nuestra competencia. La falta de conocimientos profundos en matemáticas aplicadas a la programación y, en algunos casos, el uso de un lenguaje técnico avanzado en este campo también representaron desafíos significativos.

Por último, a pesar de que los modelos han demostrado un desempeño adecuado para las tareas solicitadas, el éxito de estos sistemas depende en gran medida de la calidad y diversidad de los datos utilizados para su entrenamiento. La precisión de las predicciones mejora significativamente cuando los datos reflejan una variedad de situaciones y poblaciones. Además, los datos mal seleccionados o con errores pueden llevar a resultados inexactos, como el entrenar modelos con imágenes TC que proceden de máquinas mal calibradas (39).

## CONCLUSIÓN

A medida que aumenta la comprensión de los mecanismos moleculares que subyacen a la patogénesis del cáncer de pulmón y la aparición de nuevas terapias, se están recopilando datos masivamente. Esta abundancia de datos, junto a la creciente disponibilidad de información clínica almacenada electrónicamente, proporciona un entorno ideal para el desarrollo de modelos basados en inteligencia artificial para una comprensión integral del cáncer y la evolución de terapias personalizadas.

En esta revisión, se han explorado diferentes aplicaciones donde la IA interviene en diversas etapas del desarrollo terapéutico para el cáncer de pulmón. A la vista de los resultados, todos los modelos empleados en los estudios han demostrado tener una eficacia significativa para las tareas propuestas, desde la predicción de la respuesta a tratamientos hasta la identificación de nuevas dianas terapéuticas y el descubrimiento de nuevos ligandos. La IA ha sido capaz de detectar patrones y extraer características de grandes conjuntos de datos y generar nuevas predicciones.

Esta nueva tecnología posee un gran potencial para que el proceso de desarrollo de nuevos medicamentos y la optimización de terapias ya disponibles sea más eficiente. Sin embargo, para alcanzar el máximo rendimiento de esta tecnología y superar los desafíos actuales, es crucial realizar estudios que validen, refinan e integren estos modelos de manera efectiva. Solo a través de la investigación continua se puede asegurar que la IA se convierta en una herramienta confiable y beneficiosa en la lucha contra el cáncer de pulmón, así como en otras patologías.

## **ABREVIATURAS**

**GCO:** *Global Cancer Observatory*

**IARC:** *International Agency for Research on Cancer*

**OMS:** Organización Mundial de la Salud

**NSCLC:** *Non-Small Cell lung, Cancer*

**SCLC:** *Small Cell Lung Cancer*

**TC:** tomografía computarizada

**TKI:** *Tyrosine Kinase Inhibitor*

**EGFR:** *Epidermal Growth Factor Receptor*

**ALK:** *Anaplastic Lymphoma Kinase*

**ICI:** *Immunocheckpoint Inhibitors*

**Gen:** generación

**IA:** Inteligencia Artificial

**ML:** *Machine Learning*

**DL:** *Deep Learning*

**RF:** *Random Forest*

**CNN:** *Convolutional neural network*

**MPNN:** *Message Passing Neural Network*

**DNN:** *Deep Neural Network*

**GAN:** *Generative Adversarial Networks*

**CPANN:** *Counter-Propagation Artificial Neural Network*

**SVM:** *Support Vector Machine*

**RNN:** *Recurrent Neural Network*

**GAT:** *Graph Attention Network*

**GBRT:** *Gradient Boosting Regression Trees*

**TL:** *Transferring Learning*

**FDA:** *Food and Drug Administration*

**FDG-PET:** *Fluorodeoxyglucose Positron Emission Tomography*



## BIBLIOGRAFÍA

1. Cancer Today [Internet]. [citado 21 de mayo de 2024]. Disponible en:  
<https://gco.iarc.who.int/today/>
2. Pei Q, Luo Y, Chen Y, Li J, Xie D, Ye T. Artificial intelligence in clinical applications for lung cancer: diagnosis, treatment and prognosis. Clin Chem Lab Med CCLM. 1 de noviembre de 2022;60(12):1974-83.
3. Camacho Sosa K, Alonso Lemus L, Ramírez Rodríguez D, Carreño Rolando IE, Mendoza Jorge E, García Soto J, et al. Supervivencia de pacientes con cáncer de pulmón de células no pequeñas en estadios avanzados. Matanzas. Rev Médica Electrónica. febrero de 2021;43(1):2795-807.
4. Cáncer de pulmón. Una guía práctica. Disponible en:<https://www.contraelcancer.es/sites/default/files/migration/actualidad/publicaciones/documentos/guia-ca-pulmon.pdf>
5. Gómez-López A, Revuelta-Salgado F, García-Luján R. Cáncer de pulmón de células no pequeñas. Med - Programa Form Médica Contin Acreditado. 1 de noviembre de 2022;13(67):3933-41.
6. Casado D, Bellido L, del Barco E, Cigarral B, Escalera E, Claros J, et al. Biomarcadores moleculares y celulares. Implicaciones en la terapia dirigida. Med - Programa Form Médica Contin Acreditado. 1 de febrero de 2021;13(25):1418-23.
7. Sánchez Céspedes M. Medicina de precisión en inmunoterapia y cáncer de pulmón. R Acad Nac Farm Internet. 2023;9-22.
8. Pandiyan S, Wang L. A comprehensive review on recent approaches for cancer drug discovery associated with artificial intelligence. Comput Biol Med. 1 de noviembre de 2022;150:106140.
9. UA-MED-PDI-003-2021.pdf [Internet]. [citado 22 de mayo de 2024]. Disponible en:  
<https://dspace.uniandes.edu.ec/bitstream/123456789/13908/1/UA-MED-PDI-003-2021.pdf>

10. Perfil epigenético para predecir la respuesta a inmunoterapia en pacientes con cáncer de pulmón no microcítico [Internet]. Instituto de Salud Carlos III; 2020 [citado 22 de mayo de 2024]. Disponible en: <http://hdl.handle.net/20.500.12105/12056>
11. Pérez Del Barrio A, Menéndez Fernández-Miranda P, Sanz Bellón P, Lloret Iglesias L, Rodríguez González D. Artificial Intelligence in Radiology: an introduction to the most important concepts. *Radiologia*. 2022;64(3):228-36.
12. You Y, Lai X, Pan Y, Zheng H, Vera J, Liu S, et al. Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct Target Ther*. 10 de mayo de 2022;7(1):1-24.
13. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell*. 9 de noviembre de 2020;38(5):672-684.e6.
14. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. enero de 2020;577(7792):706-10.
16. Solanet MA, Martí ML, editores. *Inteligencia artificial: una mirada multidisciplinaria*. Buenos Aires. Academia Nacional de Ciencias Morales y Políticas; 2021. 239-258p.
16. Sandoval LJ. *ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS*. 2018;
17. Urrútia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Med Clínica*. 9 de octubre de 2010;135(11):507-11.
18. Saad MB, Hong L, Aminu M, Vokes NI, Chen P, Salehjahromi M, et al. Predicting benefit from immune checkpoint inhibitors in patients with non-

- small-cell lung cancer by CT-based ensemble deep learning: a retrospective study. *Lancet Digit Health*. 1 de julio de 2023;5(7):e404-20.
19. Aziz M, Ejaz SA, Zargar S, Akhtar N, Aborode AT, A. Wani T, et al. Deep Learning and Structure-Based Virtual Screening for Drug Discovery against NEK7: A Novel Target for the Treatment of Cancer. *Molecules*. enero de 2022;27(13):4098.
  20. Song J, Wang L, Ng NN, Zhao M, Shi J, Wu N, et al. Development and Validation of a Machine Learning Model to Explore Tyrosine Kinase Inhibitor Response in Patients With Stage IV EGFR Variant–Positive Non–Small Cell Lung Cancer. *JAMA Netw Open*. 17 de diciembre de 2020;3(12):e2030442.
  21. Li B, Dai C, Wang L, Deng H, Li Y, Guan Z, et al. A novel drug repurposing approach for non-small cell lung cancer using deep learning. *PLOS ONE*. 11 de junio de 2020;15(6):e0233112.
  22. Coker EA, Stewart A, Ozer B, Minchom A, Pickard L, Ruddle R, et al. Individualized Prediction of Drug Response and Rational Combination Therapy in NSCLC Using Artificial Intelligence–Enabled Studies of Acute Phosphoproteomic Changes. *Mol Cancer Ther*. 1 de junio de 2022;21(6):1020-9.
  23. Cheng J, Pan Y, Huang W, Huang K, Cui Y, Hong W, et al. Differentiation between immune checkpoint inhibitor-related and radiation pneumonitis in lung cancer by CT radiomics and machine learning. *Med Phys*. 2022;49(3):1547-58.
  24. Jiang HJ, Huang YA, You ZH. SAEROF: an ensemble approach for large-scale drug-disease association prediction by incorporating rotation forest and sparse autoencoder deep neural network. *Sci Rep*. 18 de marzo de 2020;10(1):4972.
  25. Park C, Na KJ, Choi H, Ock CY, Ha S, Kim M, et al. Tumor immune profiles noninvasively estimated by FDG PET with deep learning correlate

- with immunotherapy response in lung adenocarcinoma. *Theranostics*. 29 de agosto de 2020;10(23):10838-48.
26. Nakarin F, Boonpalit K, Kinchagawat J, Wachiraphan P, Rungrotmongkol T, Nutanong S. Assisting Multitargeted Ligand Affinity Prediction of Receptor Tyrosine Kinases Associated Nonsmall Cell Lung Cancer Treatment with Multitasking Principal Neighborhood Aggregation. *Molecules*. enero de 2022;27(4):1226.
27. Rounis K, Makrakis D, Papadaki C, Monastiriotti A, Vamvakas L, Kalbakis K, et al. Prediction of outcome in patients with non-small cell lung cancer treated with second line PD-1/PDL-1 inhibitors based on clinical parameters: Results from a prospective, single institution study. *PLOS ONE*. 1 de junio de 2021;16(6):e0252537.
28. Grisoni F, Neuhaus CS, Hishinuma M, Gabernet G, Hiss JA, Kotera M, et al. De novo design of anticancer peptides by ensemble artificial neural networks. *J Mol Model*. 5 de abril de 2019;25(5):112.
29. Choi G, Kim D, Oh J. AI-Based Drug Discovery of TKIs Targeting L858R/T790M/C797S-Mutant EGFR in Non-small Cell Lung Cancer. *Front Pharmacol* [Internet]. 28 de julio de 2021 [citado 22 de mayo de 2024];12. Disponible en:  
<https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2021.660313/full>
30. Altuhaifa FA, Win KT, Su G. Predicting lung cancer survival based on clinical data using machine learning: A review. *Comput Biol Med*. 1 de octubre de 2023;165:107338.
31. Yang R, Yu Y. Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis. *Front Oncol* [Internet]. 9 de marzo de 2021 [citado 22 de mayo de 2024];11. Disponible en:  
<https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2021.638182/full>



32. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. *Med Image Anal.* 1 de diciembre de 2019;58:101552.
33. ES-ESMO-Guía-para-Pacientes-los-Efectos-Secundarios-Relacionados-con-la-Inmunoterapia-y-su-Manejo.pdf [Internet]. [citado 22 de mayo de 2024]. Disponible en:  
<https://www.esmo.org/content/download/133756/2490209/1/ES-ESMO-Gu%C3%ADa-para-Pacientes-los-Efectos-Secundarios-Relacionados-con-la-Inmunoterapia-y-su-Manejo.pdf>
34. Zheng Y, Zhou D, Liu H, Wen M. CT-based radiomics analysis of different machine learning models for differentiating benign and malignant parotid tumors. *Eur Radiol.* 1 de octubre de 2022;32(10):6953-64.
35. Carpenter KA, Cohen DS, Jarrell JT, Huang X. Deep Learning and Virtual Drug Screening. *Future Med Chem.* 1 de noviembre de 2018;10(21):2557-67.
36. Wang H. Prediction of protein–ligand binding affinity via deep learning models. *Brief Bioinform.* 1 de marzo de 2024;25(2):bbae081.
37. Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte R, Milles L, et al. Robust deep learning–based protein sequence design using ProteinMPNN. *Science.* 7 de octubre de 2022;378(6615):49-56.
38. Zeng WZD, Glicksberg BS, Li Y, Chen B. Selecting precise reference normal tissue samples for cancer research using a deep learning approach. *BMC Med Genomics.* 31 de enero de 2019;12(1):21.
39. Vega MÁ, Mora LMQ, Badilla MVC. Inteligencia artificial y aprendizaje automático en medicina. *Rev Medica Sinerg.* 1 de agosto de 2020;5(8):e557-e557.