



# Inteligencia artificial en esclerosis múltiple

## *Artificial intelligence in multiple sclerosis*

Carmen Tur<sup>1,2\*</sup> y Sara Collorone<sup>2</sup>

<sup>1</sup>Centro de Esclerosis Múltiple de Cataluña (CEMCAT), Departamento de Neurología/Neuroinmunología, Hospital Universitario Vall d'Hebron, Universidad Autónoma de Barcelona, Barcelona, España;

<sup>2</sup>NMR Unit, Queen Square MS Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, University College London, Londres, Reino Unido

### Resumen

La inteligencia artificial (IA) es la rama de la ciencia que tiene como objetivo crear algoritmos capaces de llevar a cabo tareas que normalmente requieren inteligencia humana. La medicina ha experimentado un notable aumento en las aplicaciones de la IA gracias a computadoras cada vez más potentes y a la aparición de grandes repositorios de datos. La esclerosis múltiple (EM) es una enfermedad autoinmune crónica que afecta al sistema nervioso central con una patogénesis compleja, un proceso diagnóstico desafiante y una alta variabilidad entre personas, en gran parte inexplicada. Las aplicaciones de la IA en la EM tienen un gran potencial para respaldar mejor el diagnóstico, encontrar marcadores pronósticos y ayudar a comprender los mecanismos de la enfermedad. Esta revisión tiene como objetivo resumir los avances recientes de la IA en la EM para ilustrar sus logros, limitaciones y direcciones futuras.

**Palabras clave:** Inteligencia artificial. *Deep learning*. Esclerosis múltiple. Resonancia magnética.

### Abstract

*Artificial intelligence (AI) is the branch of science aiming at creating algorithms able to carry out tasks that typically require human intelligence. In medicine, there has been a tremendous increase in AI applications thanks to increasingly powerful computers and the emergence of big data repositories. Multiple sclerosis (MS) is a chronic autoimmune condition affecting the central nervous system with a complex pathogenesis, a challenging diagnostic process, and a high and largely unexplained variability across people. Therefore, AI applications in MS have the great potential of better support the diagnosis, find markers for prognosis, and help us understand the mechanisms of the disease. This topical review aims to summarise the recent advances in AI in MS to illustrate its achievements, limitations, and future directions.*

**Keywords:** *Artificial intelligence. Deep learning. Multiple sclerosis. Magnetic resonance imaging.*

\* **Autora de correspondencia:**  
Carmen Tur  
E-mail: ctur@cem-cat.org

Fecha de recepción: 05-08-2023  
Fecha de aceptación: 12-08-2023  
DOI: 10.24875/KRANION.M23000065

Disponible en internet: 23-11-2023  
Kranion. 2023;18:139-44  
[www.kranion.es](http://www.kranion.es)

## INTRODUCCIÓN

La inteligencia artificial (IA) es la rama de la ciencia que tiene como objetivo crear algoritmos capaces de realizar tareas que normalmente requieren inteligencia humana<sup>1</sup>. En medicina, la aplicación de técnicas de IA ha llegado más tarde que en otras áreas de la ciencia, posiblemente debido a la complejidad de los datos y a los aspectos éticos. No obstante, se ha observado un notable aumento en los últimos años. La tendencia es similar en el campo de la esclerosis múltiple (EM), pasando de cinco publicaciones en PubMed relacionadas con la IA en 2013 a casi 200 en 2022.

Las metodologías de IA utilizadas para la investigación médica tienen principalmente dos objetivos. Por un lado, la IA puede extraer patrones de los datos para comprender su estructura interna o agrupar sus características. Para este propósito, los datos no deben estar etiquetados y la salida (p. ej., cómo se agruparán esos datos en clústeres o conjuntos) es desconocida (aprendizaje no supervisado). Por otro lado, la IA puede realizar predicciones a partir de datos disponibles etiquetados, la salida es conocida (p. ej., clasificación entre enfermos y sanos) y está bien caracterizada (aprendizaje supervisado) (Fig. 1). Esto es posible gracias al aprendizaje automático (*machine learning*), un conjunto de algoritmos matemáticos que permiten identificar patrones, elaborar predicciones y además aprenden (mejoran) con el entrenamiento. El aprendizaje profundo (*deep learning*) es un tipo de aprendizaje automático que usa modelos de redes neuronales artificiales con muchas capas ocultas y posibilita

procesar datos masivos de nivel superior, por ejemplo imágenes médicas, vídeo, voz, lenguaje, etc. El aprendizaje profundo tiene la ventaja de no necesitar variables preseleccionadas como entrada, sino que extrae las variables (o características [*features*]) necesarias durante el proceso de análisis por la red. El aprendizaje profundo también puede ser supervisado o no supervisado según sea la cuestión a estudio, el tipo de datos disponibles y la salida (*outcome*) deseada. Suele aplicarse al análisis de imágenes médicas como la resonancia magnética (RM), cuyas unidades, los píxeles, contienen información numérica y espacial.

La IA aplicada a la EM tienen el potencial de respaldar el diagnóstico clínico, encontrar marcadores pronósticos y, en última instancia, comprender los mecanismos de la enfermedad. Esta revisión tiene como objetivo resumir los avances recientes de la IA en el campo de la EM, con foco en los aspectos mencionados e ilustrando sus logros, limitaciones y futuras direcciones.

## DIAGNÓSTICO DE LA ESCLEROSIS MÚLTIPLE

La EM es una afección autoinmune crónica que afecta el sistema nervioso central y se caracteriza por una patogénesis compleja, marcada por una cascada dinámica de inflamación, desmielinización y daño neuroaxonal<sup>2</sup>. En la mayoría de las personas, la EM comienza con un episodio de síntomas neurológicos subagudos indicativos de una lesión desmielinizante en el sistema nervioso central, conocido como síndrome clínicamente

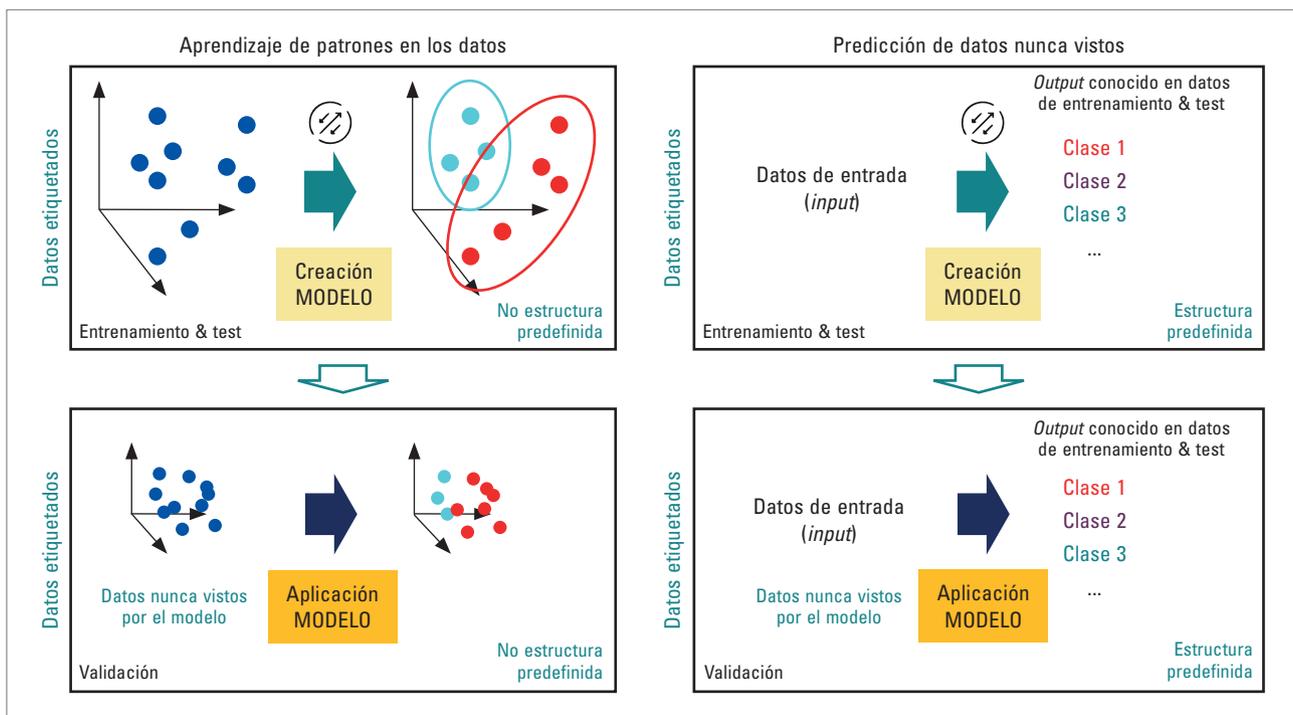


FIGURA 1. Principales tipos de análisis que pueden realizarse con modelos de inteligencia artificial y sus pasos fundamentales.

aislado (CIS, de sus siglas en inglés)<sup>3</sup>. Después de este primer episodio de desmielinización, suele haber una forma recurrente-remitente de la enfermedad caracterizada por episodios agudos o recaídas, tras las cuales puede haber una recuperación parcial o completa. Después de un periodo variable de tiempo puede observarse una progresión que ocurre de manera independiente a la presencia de recaídas. En alrededor del 10% de los casos la enfermedad puede comenzar con progresión clínica y se denomina EM primaria progresiva<sup>4</sup>.

El diagnóstico de la EM se basa en la integración de hallazgos clínicos, de imagen y de laboratorio para demostrar la presencia de diseminación en el espacio y el tiempo de la enfermedad, y excluir diagnósticos alternativos, especialmente en presencia de las llamadas señales de alarma<sup>5</sup>. En individuos con un CIS típico, la RM del cerebro y la médula espinal es la herramienta principal para respaldar el diagnóstico y excluir otras enfermedades. Sin embargo, el diagnóstico de la EM no carece de desafíos. Los neurólogos deben considerar cuidadosamente los signos, síntomas o resultados de pruebas diagnósticas, ya que otras enfermedades pueden imitar, clínica o radiológicamente, la EM. Además, existen poblaciones especiales (p. ej., personas no caucásicas, individuos con comorbilidades, etc.) en las que los criterios diagnósticos deben aplicarse con más cuidado<sup>5</sup>. Por lo tanto, es extremadamente importante que los médicos entiendan cuándo deben y cuándo no deben aplicar los criterios diagnósticos. Y es en esas últimas circunstancias donde el uso de algoritmos de IA para llegar al diagnóstico correcto puede ser más útil.

En este contexto, varios estudios han investigado el potencial de los modelos de IA para distinguir la EM de sus imitadores (y de individuos sanos) en función de datos de neuroimagen<sup>6-8</sup>. Por ejemplo, en 2016, Eshaghi et al.<sup>6</sup> utilizaron un modelo de aprendizaje automático para extraer 157 características de imágenes ponderadas en T1 de 144 participantes con EM, trastorno del espectro de la neuromielitis óptica (NMOSD, siglas en inglés) e individuos sanos de diferentes centros. El modelo de aprendizaje automático logró una precisión del 74%, abriendo nuevas posibilidades para utilizar el aprendizaje automático en el diagnóstico diferencial en la clínica de la EM. Más recientemente, Rocca et al. construyeron un modelo de aprendizaje profundo basado en redes neuronales convolucionales (CNN, sigla en inglés) para clasificar a los pacientes en diferentes grupos diagnósticos, a saber: EM (n = 70), NMOSD (n = 91), vasculitis del sistema nervioso central (n = 51) y migraña (n = 56)<sup>8</sup>. Luego compararon la precisión de ese modelo de aprendizaje profundo con la lograda por dos evaluadores humanos, neurorradiólogos experimentados, a la hora de clasificar a los pacientes en cada grupo diagnóstico. El modelo logró una precisión muy alta en todos los grupos, superando a los evaluadores humanos. Otros estudios basados en modelos de IA han logrado precisiones igualmente altas para clasificar EM

frente a otras enfermedades, como glioma de bajo grado<sup>9</sup> y microangiopatía crónica<sup>10</sup>, entre otras.

## PREDICCIÓN DE LA EVOLUCIÓN DE LA ESCLEROSIS MÚLTIPLE

La EM es una afección altamente heterogénea, lo que significa que las personas con EM en el momento del diagnóstico deben enfrentarse a una enorme incertidumbre, tanto en relación con la ocurrencia de futuras recaídas como en el desarrollo de discapacidad irreversible. En este contexto, la IA puede ser de gran ayuda para predecir tanto la aparición de brotes como la aparición de la discapacidad.

### Brotos

El número de estudios que han utilizado metodologías de IA para predecir el riesgo de futuras recaídas con datos clínicos y de RM en el inicio de los síntomas es limitado. Entre las publicaciones más relevantes se encuentra la de Wottschel et al.<sup>11</sup>, que aplicaron modelos de aprendizaje automático a medidas morfométricas extraídas de imágenes de RM cerebral de 400 pacientes en el momento del primer brote para predecir el desarrollo del segundo brote durante el primer año de la enfermedad. Utilizaron un algoritmo de *support vector machine* (SVM, máquina de vectores de soporte), un modelo de aprendizaje supervisado para tareas de clasificación y regresión. El modelo logró una precisión del 71% y mostró que una mayor carga de lesiones en todo el cerebro, un menor volumen de materia gris en el tálamo y la región del precuneus, y un córtex más delgado en el cuneus y la circunvolución temporal inferior fueron los mejores predictores de la actividad clínica futura.

### Progresión de la discapacidad

En los últimos años, predecir de la manera más precisa posible quiénes alcanzarán altos niveles de disfunción clínica a medio o largo plazo ha sido una prioridad para toda la comunidad científica<sup>12</sup>. La llegada de la RM para su uso en la práctica clínica ha significado un paso importante en términos de predicción clínica, y tanto las lesiones inflamatorias más extensas como la atrofia se han asociado con un peor pronóstico a largo plazo<sup>12-14</sup>. Más recientemente, la irrupción de la IA y, en particular, del aprendizaje profundo, ha aumentado exponencialmente el potencial de la RM para predecir la evolución de la enfermedad.

Entre las publicaciones enfocadas en predicciones a corto o medio plazo, un estudio digno de destacar es el trabajo de Tousignant et al.<sup>15</sup>, quienes desarrollaron un modelo de aprendizaje profundo utilizando datos de RM multimodal de 465 personas con EM provenientes de brazos placebo de ensayos clínicos controlados y aleatorizados. El modelo se diseñó para predecir la progresión

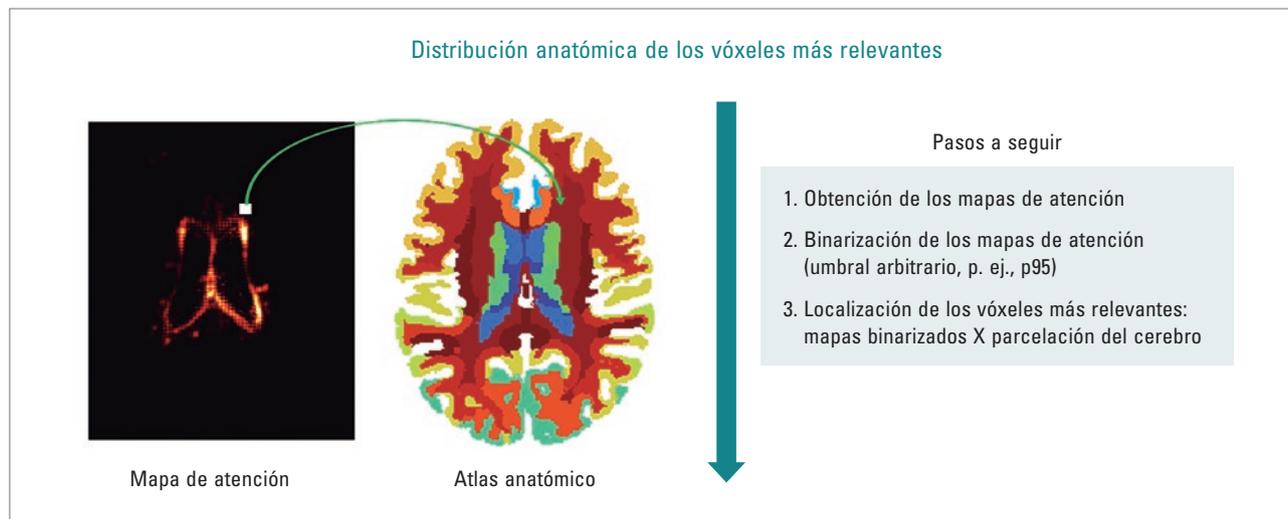
de la enfermedad a un año de seguimiento. Inicialmente, su modelo logró un rendimiento moderado con un área bajo la curva (AUC) de 0,66 ( $\pm$  0,055). Sin embargo, cuando incorporaron máscaras de lesiones de secuencias ponderadas en T2 y posgadolinio al modelo, el AUC mejoró a 0,701 ( $\pm$  0,027), resaltando la importancia de los datos de entrada de alta calidad para mejorar el rendimiento del modelo. En un estudio posterior, de 2020, Roca et al.<sup>16</sup> evaluaron el rendimiento de un modelo de CNN que utilizó la edad, el sexo y las exploraciones FLAIR del cerebro de una gran base de datos francesa multicentro ( $n = 971$ ) para predecir las puntuaciones de la *Expanded Disability Status Scale* (EDSS) en un seguimiento de dos años. El modelo CNN mostró un rendimiento general moderado, con un error promedio en las puntuaciones del EDSS de 1,7. Es importante destacar que la precisión del modelo pareció disminuir al predecir valores de EDSS muy bajos o muy altos. Este resultado posiblemente se debió a la variabilidad asociada con la estimación de las puntuaciones del EDSS cuando estas son bajas y a una distribución desequilibrada de las puntuaciones del EDSS en la población de estudio. Es relevante señalar que este estudio comparó el modelo CNN con dos modelos de aprendizaje automático no basados en aprendizaje profundo, que mostraron un rendimiento comparable<sup>16</sup>. Otro estudio realizado por Storelli et al.<sup>17</sup> también empleó un modelo de CNN para predecir la progresión de la discapacidad a dos años en 373 personas con EM. En este caso se utilizaron imágenes ponderadas en T1 y T2 como datos de entrada. La variable de salida para la discapacidad fue una medida binaria de progresión basada en el EDSS y puntuaciones del *Symbol Digit Modalities Test* (SDMT). El modelo CNN demostró una considerable precisión predictiva tanto para el empeoramiento del EDSS (83,3%) como para el empeoramiento del SDMT (67,7%). Sin embargo, la mayor precisión se logró al utilizar información de ambos test (85,7%)<sup>17</sup>.

En cuanto a las predicciones a largo plazo de la discapacidad utilizando modelos de IA, debe destacarse el trabajo realizado por Zhao et al.<sup>18</sup>. En su estudio, construyeron una serie de modelos de aprendizaje automático (tipo SVM) para predecir la progresión de la discapacidad (*outcome* binario) con un seguimiento de cinco años en 1.693 personas con EM. Utilizaron diferentes tipos de datos de entrada y compararon el rendimiento de todos estos modelos de aprendizaje automático con el de un modelo de regresión logística convencional que utilizó los mismos datos de entrada. Los modelos que solo utilizaron datos clínico-demográficos iniciales y a 12 meses proporcionaron muy poca precisión. En cambio, el modelo que incluyó datos clínico-demográficos y de RM iniciales y a 24 meses como datos de entrada mostró una precisión mayor (75%). Además, solo en este escenario el modelo de aprendizaje automático tuvo un mejor rendimiento que el de regresión logística.

## INVESTIGACIÓN DE LOS MECANISMOS DE LA ENFERMEDAD

Los procesos fisiopatológicos que subyacen a la progresión de la enfermedad en la EM no se comprenden completamente y se cree que son altamente heterogéneos entre personas y en las diferentes etapas de la enfermedad. En este contexto, el uso de la IA puede verse como una herramienta para ayudarnos a comprender los mecanismos subyacentes a la acumulación irreversible de discapacidad en esta enfermedad. Los estudios basados en IA diseñados con el principal propósito de comprender los mecanismos de la enfermedad han utilizado principalmente dos estrategias: a) modelos de aprendizaje automático no supervisado, y b) mapas de atención derivados de aprendizaje profundo (CNN), también llamados mapas de saliencia o de calor. Los mapas de atención representan aquellas partes (regiones) en los datos de entrada que llevan la información más importante para que el modelo de IA, en este caso la CNN<sup>19</sup>, realice la predicción final. En nuestro contexto, los mapas de atención representan aquellas regiones anatómicas cerebrales más relevantes para la tarea del modelo de aprendizaje profundo que se está evaluando<sup>20</sup>.

Entre los estudios que utilizan modelos de aprendizaje automático no supervisado, cabe destacar el trabajo realizado por Eshaghi et al.<sup>21</sup>, que aplicó un algoritmo de aprendizaje automático no supervisado (más específicamente, el modelo SuStaln<sup>22</sup>) en imágenes de RM de 6322 personas con EM para clasificarlas según diferentes patrones longitudinales de lesión cerebral. Identificaron tres patrones diferentes: uno dominado por el daño cortical, con un componente predominantemente neurodegenerativo; otro dominado por el daño en la sustancia blanca de apariencia normal, con un componente más inflamatorio crónico, y otro dominado por el daño derivado de las lesiones, con un componente inflamatorio más agudo. En su estudio, el subtipo dominado por lesiones tuvo el peor pronóstico, con una progresión más rápida de la discapacidad<sup>21</sup>. Este modelo fue posteriormente aplicado por Pontillo et al.<sup>23</sup> a una cohorte diferente de personas con EM ( $n = 425$ ). En este estudio utilizaron como datos de entrada al modelo de aprendizaje automático características volumétricas derivadas de la RM cerebral y el algoritmo de IA. Aunque encontraron dos patrones de enfermedad en lugar de tres, pudieron identificar, al igual que en el estudio de Eshaghi et al., un patrón dominado por el daño cortical con un componente neurodegenerativo caracterizado por la atrofia cortical como una de las principales características iniciales, y un patrón dominado por el daño en la sustancia gris profunda donde la atrofia de la sustancia gris profunda y la presencia de lesiones en T2 precederían al desarrollo de la atrofia cortical. Como hallazgo importante, encontraron que el patrón dominado por el daño en la sustancia gris profunda se asociaba a peor pronóstico que el patrón de daño cortical, en línea con los hallazgos de



**FIGURA 2.** Pasos que seguir para conocer la distribución anatómica de los vóxeles más relevantes de la imagen original de entrada en la decisión tomada por redes neuronales convolucionales (véase Coll et al.<sup>24</sup> para más detalles).

Eshaghi. En cualquier caso, este algoritmo de IA proporcionó, en ambos estudios, una estratificación precisa de los pacientes biológicamente fiable y significativa desde el punto de vista pronóstico<sup>21,23</sup>.

En cuanto a los estudios que han explotado los mapas de atención para comprender mejor los mecanismos de la enfermedad, destaca el trabajo de Eitel et al.<sup>7</sup>. En este estudio, los autores utilizaron mapas de atención derivados por CNN a partir de imágenes de RM de 76 personas con EM y 71 controles sanos para comprender qué áreas del cerebro eran determinantes para el diagnóstico de la EM. Descubrieron que estas áreas se encontraban en las regiones de la sustancia blanca periventricular posterior<sup>7</sup>. Más recientemente, Coll et al.<sup>24</sup> utilizaron mapas de atención derivados por CNN para descubrir las regiones cerebrales que el algoritmo de aprendizaje profundo había considerado más relevantes para la estratificación de personas con EM ( $n = 319$ ) en discapacitadas ( $EDSS \geq 3.0$ ) o no discapacitadas ( $EDSS < 3.0$ ) utilizando solo una imagen de RM estructural<sup>24</sup>. Su modelo de CNN logró una precisión promedio del 79%, superando a un modelo de regresión logística equivalente (77%). Los análisis de los mapas de atención (Fig. 2) revelaron el papel predominante de la corteza frontotemporal y el cerebelo en las decisiones de la CNN, lo que sugiere que los mecanismos que conducen a la acumulación de discapacidad podrían estar relacionados con el daño en esas regiones<sup>24</sup>.

## CONCLUSIONES

El uso de la IA para diagnosticar y predecir la evolución de la EM ha avanzado significativamente en los últimos años. Sin embargo, es esencial subrayar que la mayoría de los modelos de IA en la EM aún no han demostrado una clara superioridad sobre los modelos

estadísticos clásicos. A pesar de esto, el potencial de la IA para contribuir al diagnóstico y la predicción en la EM está adquiriendo un creciente reconocimiento. Actualmente, son varias las limitaciones probablemente responsables de que exista una brecha significativa entre los estudios de IA y su utilidad clínica. Sin embargo, es de esperar que tales limitaciones se aborden y superen en un futuro cercano, lo que pronto llevará a la implementación de algoritmos de IA en los entornos clínicos.

Uno de los principales desafíos en el uso de la IA para el diagnóstico y el pronóstico en la EM es la calidad de los datos de entrada. La recopilación precisa y completa de datos es esencial para desarrollar modelos de IA fiables. Otro desafío es garantizar la reproducibilidad de los métodos utilizados en los estudios basados en IA. La capacidad de reproducir y validar resultados es clave para aceptar e integrar la IA en la práctica clínica. Finalmente, es importante abordar adecuadamente los problemas éticos que puedan surgir de la implementación de la IA en los entornos clínicos en relación con la privacidad, la seguridad de los datos y los posibles sesgos en los algoritmos y sus decisiones.

A pesar de estos desafíos, existen numerosas oportunidades asociadas al uso de la IA en la investigación y la práctica clínica de la EM. El rápido desarrollo de los métodos de IA y los equipos de *hardware* posibilitará avances adicionales en este campo. La IA, especialmente el aprendizaje profundo, tiene un gran potencial para mejorar en un futuro próximo el diagnóstico y el pronóstico de la EM en el día a día con el paciente. Además, la IA proporcionará nuevos conocimientos para entender mejor los mecanismos subyacentes al aumento de la discapacidad en la EM, lo que podría derivar en un mejor diseño de los ensayos clínicos y, consecuentemente, en una mayor probabilidad de encontrar tratamientos más eficaces y efectivos para esta enfermedad.

## FINANCIACIÓN

La presente investigación no ha recibido ayudas específicas provenientes de agencias del sector público, sector comercial o entidades sin ánimo de lucro.

## CONFLICTO DE INTERESES

Las autoras declaran no tener conflicto de intereses.

## RESPONSABILIDADES ÉTICAS

### Protección de personas y animales

Las autoras declaran que para este trabajo no se han realizado experimentos en seres humanos ni en animales.

### Confidencialidad de los datos

Las autoras declaran que en este trabajo no aparecen datos de pacientes.

### Derecho a la privacidad y consentimiento informado

Las autoras declaran que en este trabajo no aparecen datos de pacientes.

### Uso de inteligencia artificial generativa

Las autoras declaran que no han utilizado ningún tipo de inteligencia artificial generativa en la redacción de este manuscrito ni para la creación de figuras, gráficos, tablas o sus correspondientes pies o leyendas.

## BIBLIOGRAFÍA

1. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. 1943. *Bull Math Biol.* 1990;52:99-115; discussion 73-97.
2. Kutzelnigg A, Lassmann H. Pathology of multiple sclerosis and related inflammatory demyelinating diseases. *Handb Clin Neurol.* 2014;122:15-58.
3. Miller DH, Chard DT, Ciccarelli O. Clinically isolated syndromes. *Lancet Neurol.* 2012;11:157-69.

4. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology.* 2014;83:278-86.
5. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 2018;17:162-73.
6. Eshaghi A, Wottschel V, Cortese R, Calabrese M, Sahraian MA, Thompson AJ, et al. Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology.* 2016;87:2463-70.
7. Eitel F, Soehler E, Bellmann-Strobl J, Brandt AU, Ruprecht K, Giess RM, et al. Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *NeuroImage Clin.* 2019;24:102003.
8. Rocca MA, Anzalone N, Storelli L, Del Poggio A, Cacciaguerra L, Manfredi AA, et al. Deep learning on conventional magnetic resonance imaging improves the diagnosis of multiple sclerosis mimics. *Invest Radiol.* 2021;56:252-260.
9. Eksi Z, Özcan ME, Çakıroğlu M, Öz C, Aralaşmak A. Differentiation of multiple sclerosis lesions and low-grade brain tumors on MRS data: machine learning approaches. *Neurol Sci.* 2021;42:3389-95.
10. Theoharakis P, Glotsos D, Kalatzis I, Kostopoulos S, Georgiadis P, Sifaki K, et al. Pattern recognition system for the discrimination of multiple sclerosis from cerebral microangiopathy lesions based on texture analysis of magnetic resonance images. *Magn Reson Imaging.* 2009;27:417-22.
11. Wottschel V, Chard DT, Enzinger C, Filippi M, Frederiksen JL, Gasperini C, et al. SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis. *NeuroImage Clin.* 2019;24:102011.
12. Tintore M, Rovira À, Río J, Otero-Romero S, Arrambide G, Tur C, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain.* 2015;138:1863-74.
13. Fisniku LK, Brex PA, Altmann DR, Miszkil KA, Benton CE, Lanyon R, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain.* 2008;131:808-17.
14. Eshaghi A, Prados F, Brownlee WJ, Altmann DR, Tur C, Cardoso MJ, et al. Deep gray matter volume loss drives disability worsening in multiple sclerosis. *Ann Neurol.* 2018;83:210-22.
15. Tousignant A, Lemaître P, Precup D, Arnold DL, Arbel T. Prediction of disease progression in multiple sclerosis patients using deep learning analysis of MRI data. *Proc Mach Learn Res.* 2019;102:483-92.
16. Roca P, Attye A, Colas L, Tucholka A, Rubini P, Cackowski S, et al. Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. *Diagn Interv Imaging.* 2020;101:795-802.
17. Storelli L, Azzimonti M, Gueye M, Vizzino C, Preziosa P, Tedeschi G, et al. A deep learning approach to predicting disease progression in multiple sclerosis using magnetic resonance imaging. *Invest Radiol.* 2022;57:423-32.
18. Zhao Y, Healy BC, Rotstein D, Guttmann CR, Bakshi R, Weiner HL, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS One.* 2017;12:e0174866.
19. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436-44.
20. Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One.* 2015;10:e0130140.
21. Eshaghi A, Young AL, Wijeratne PA, Prados F, Arnold DL, Narayanan S, et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat Commun.* 2021;12(1):2078.
22. Young AL, Marinescu R V, Oxtoby NP, Bocchetta M, Yong K, Firth NC, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nat Commun.* 2018;9:4273.
23. Pontillo G, Penna S, Cocozza S, Quarantelli M, Gravina M, Lanzillo R, et al. Stratification of multiple sclerosis patients using unsupervised machine learning: a single-visit MRI-driven approach. *Eur Radiol.* 2022;32:5382-91.
24. Coll L, Pareto D, Carbonell-Mirabent P, Cobo-Calvo A, Arrambide G, Vidal-Jordana A, et al. Deciphering multiple sclerosis disability with deep learning attention maps on clinical MRI. *NeuroImage Clin.* 2023;38:103376.