

# Inteligencia artificial ¿mejora el diagnóstico de la esclerosis múltiple?

Carlos Manuel Arocha Fuentes (1), Evelyn Omara González Lozano (1),  
Bárbara Teresa Catelani (1), María Eloísa Santos Armentia (1),  
Noelia Silva Priegue (1), Sergi Valverde (2),  
Roger Bramon (2), Marc Guirao (2), Alex Rovira (3)

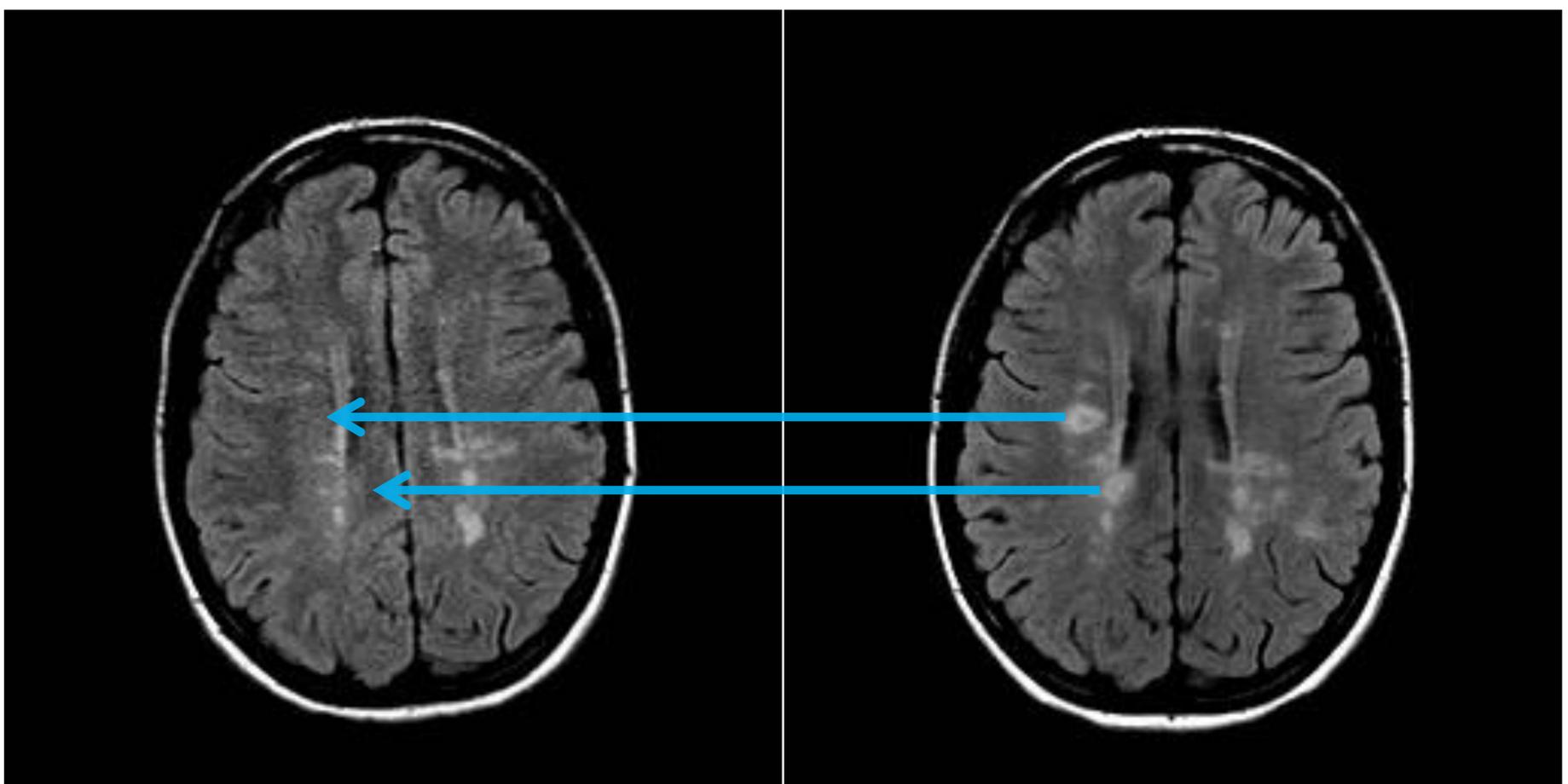
(1) Hospital Universitario Rivera Povisa

(2) Instituto de Investigación de Visión Artificial y Robótica. Universidad de Girona.

(3) Unidad de Resonancia Magnética. Departamento de Radiología, Hospital  
Universitario Vall d'Hebron.

La resonancia magnética (RM) cerebral es la técnica de imagen de referencia para el diagnóstico y seguimiento de la Esclerosis Múltiple.

En este escenario, el empleo de herramientas de postproceso basadas en la inteligencia artificial (IA) para analizar imágenes de RM ha emergido como una herramienta con gran potencial, que ayuda mejora la detección de lesiones desmielinizantes, facilita la monitorización y reduce del tiempo de evaluación del radiólogo.



## OBJETIVOS:

- Evaluar la validez de detección de nuevas lesiones de esclerosis múltiple utilizando herramientas de postproceso y de inteligencia artificial (IA) en imágenes de RM cerebral.
- Determinar nivel de acuerdo entre los evaluadores en la detección de datos de actividad de la enfermedad.
- Determinar si existen diferencias significativas en el número de lesiones identificadas con estas herramientas frente al análisis visual de los neurorradiólogos.
- Valorar si existen diferencias en la detección de lesiones según la adquisición de los estudios en diferentes máquinas de RM.

## MATERIAL Y MÉTODOS:

Estudio retrospectivo unicéntrico en un hospital de tercer nivel.

### PARTICIPANTES:

-2 Neurorradiólogos

-2 residentes de radiología de primer y segundo año

### MUESTRA

Se seleccionaron las RM cerebrales de pacientes diagnosticados con esclerosis múltiple (EM) en el mes de Noviembre de 2023 y se compararon con el estudio previo inmediatamente anterior que se le hubiera realizado a cada uno de ellos. Se analizaron únicamente las secuencias FLAIR-3D de las últimas dos RM de control.

### EQUIPOS:

Todas las RM se adquirieron en 3 equipos diferentes de 1,5 T (Philips Intera®, Prodiva® y Achieva®), según las necesidades del servicio en el momento de su adquisición. (**Tabla 1**).

**Tabla 1: Equipos de RM empleados en la adquisición de imágenes.**

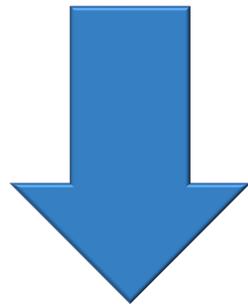
Equipos de RM	Tiempo de ECO (milisegundos)	Tiempo de Inversión (milisegundos)	Tiempo de repetición (milisegundos)
<b>Philips Intera</b>	<b>30 ms</b>	<b>1660 ms</b>	<b>4800 ms</b>
<b>Philips Prodiva</b>	<b>35 ms</b>	<b>1700 ms</b>	<b>5000 ms</b>
<b>Philips Achieva</b>	<b>40 ms</b>	<b>1700 ms</b>	<b>5000 ms</b>

Se recogió el equipo de RM utilizado para obtener las imágenes del estudio inicial y de control. Se dicotomizó en “igual equipo” si ambos estudios se realizaron en la misma máquina o “equipo diferente” en caso contrario.

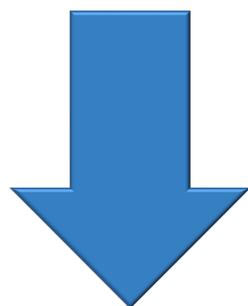
## **MATERIAL Y MÉTODOS:**

### PROTOCOLO DEL ESTUDIO

1º PASO: evaluación inicial: 2 neurorradiólogos informaron las RM con el protocolo habitual empleado en el hospital



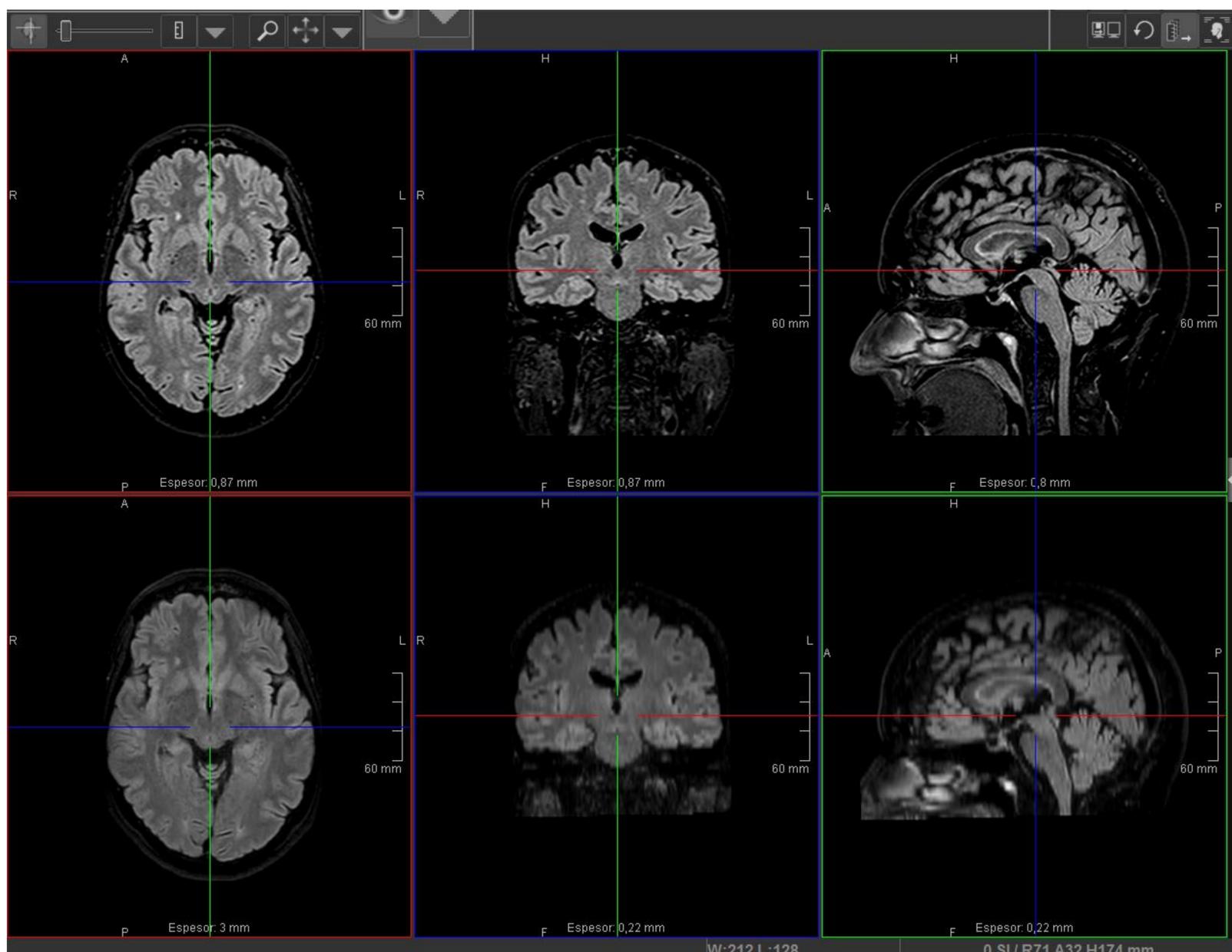
2º PASO: 2 lectores independientes evaluaron las imágenes de estas RM utilizando dos herramientas diferentes de postproceso de RM en EM de forma independiente y cegados a los resultados del paso 1



3º PASO: evaluación conjunta del neurorradiólogo con los resultados de las herramientas

## MATERIAL Y MÉTODOS: 1º PASO

Evaluación inicial: Las RM fueron informadas por dos neurorradiólogas de 17 y 13 años de experiencia en el momento de la realización de la última RM. Se visualizaron las imágenes en un sistema de almacenamiento de imágenes médicas (PACS) y se utilizó una herramienta de sustracción de imágenes (Voxar 3D® subtraction, Impax)(**Imagen 1**), sustrayendo la RM más reciente de la inmediatamente anterior. En caso de haber lesiones nuevas, se guardó una imagen documentando la localización exacta de cada una de ellas.



**Imagen .1:** Herramienta de sustracción de imágenes Voxar 3D® subtraction, Impax, empleada por los neurorradiólogos en la elaboración del informe.

## MATERIAL Y MÉTODOS: 2º PASO

Las secuencias FLAIR-3D de esas RM fueron evaluadas por dos observadores.

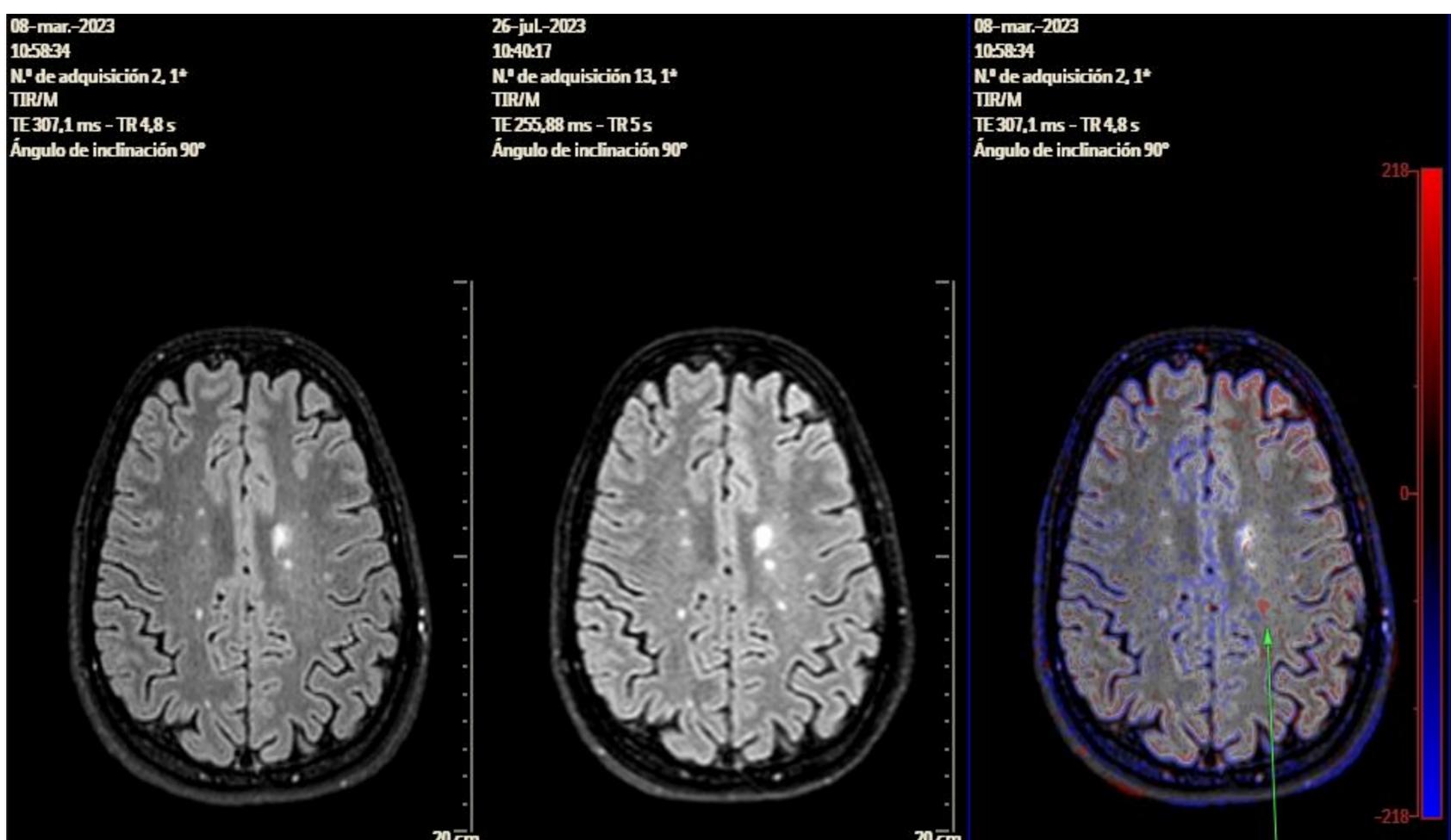
Para valorar la facilidad de uso de estas herramientas se seleccionaron dos observadores,

- un residente de primer año sin experiencia en RM ni en EM (Evaluador 1),
- un residente de segundo año (Evaluador 2); ambos sin experiencia en el uso de herramientas de postproceso ni de IA.

Ambos valoraron de forma independiente y cegados a los hallazgos de las neurorradiólogas la detección de nuevas lesiones de EM utilizando dos herramientas de postproceso de lesiones de EM en RM, **A** y **B**, diseñadas con diferentes algoritmos.

## MATERIAL Y MÉTODOS: 2º PASO

- **Herramienta A:** software que realiza automáticamente un corregistro, normalización y sustracción, vinculando las secuencias en el mismo nivel anatómico (Philips® Longitudinal Brain Imaging (LoBI) ®) (**Imagen 2**) Muestra la imagen de RM cerebral FLAIR más antigua a la izquierda, en el centro la imagen más reciente y a la derecha crea un mapa de superposición, donde resalta en rojo un aumento focal de la intensidad (las lesiones que han aparecido nuevas o que han aumentado de tamaño) y en azul una disminución de la intensidad de señal (las que han disminuido de tamaño). Sobre este mapa de colores los evaluadores hicieron un análisis visual, e hicieron imágenes marcando las lesiones que les parecían nuevas y las de dudoso significado.



**Imagen .2:** Interfaz de la herramienta A

## MATERIAL Y MÉTODOS: 2º PASO

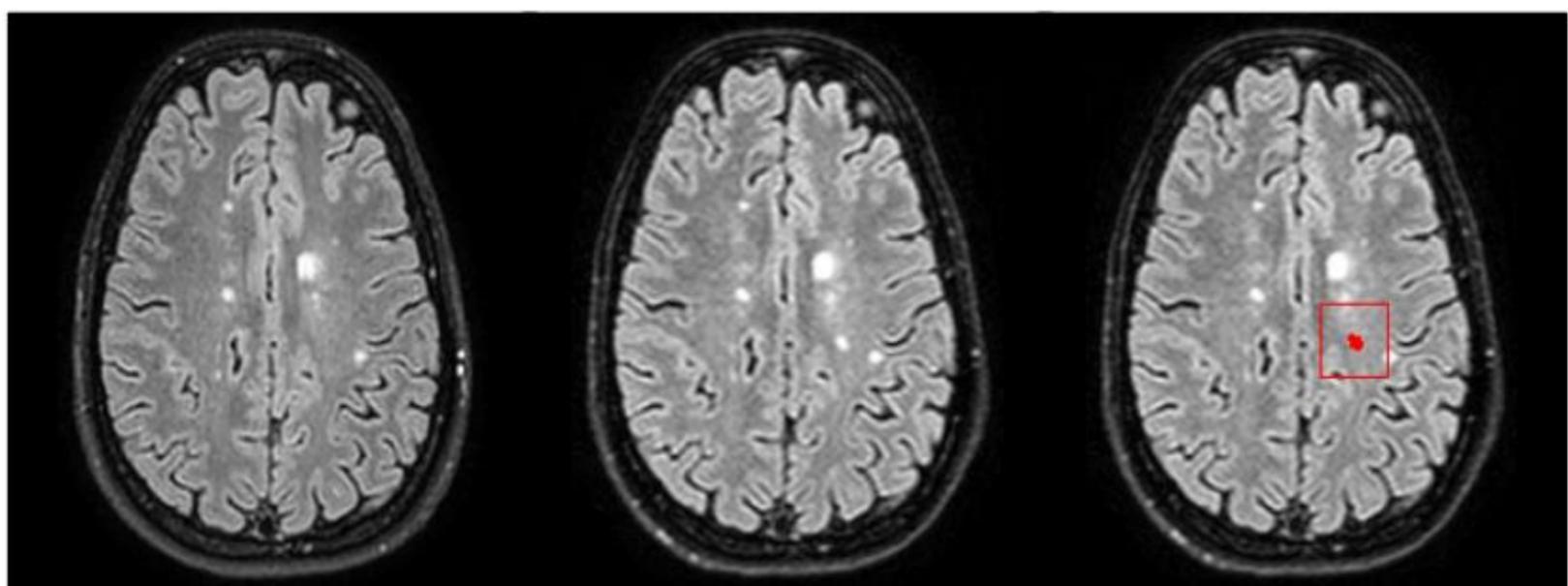
- **Herramienta B:** algoritmo de aprendizaje supervisado basado en la aplicación de redes neuronales convolucionales con técnicas de *deep learning* <sup>(2)</sup>. (TensorMedical®, Girona, Spain/Department of Computer Architecture and Technology). (**Imagen 3**)
- Esta herramienta muestra automáticamente un análisis de resultados con una pestaña a la derecha con los hallazgos de interés. Al ir a cada una de las imágenes clave, muestra a la izquierda la RM previa, en el centro la RM más reciente y a la derecha marca la lesión nueva que ha detectado en color rojo con un recuadro que permite identificar visualmente la lesión y la numeración del corte correspondiente. Tras validar los hallazgos genera un PDF señalando cada una de las lesiones nuevas.

### Analysis Results

Months between 4 months

New lesions 4

Lesion 1 (Axial Slice 194)



(a) Baseline

(b) Followup

(c) New Lesion

**Imagen .3:** Interfaz de la herramienta B

## **MATERIAL Y MÉTODOS: 2º PASO**

Se registró el tiempo que tardó cada evaluador en utilizar cada una de las herramientas y la usabilidad percibida por los evaluadores empleando la versión estándar traducida al español de la System Usability Scale (SUS). La escala se encuentra compuesta por 10 items con 5 opciones de respuesta, donde 1 indica completamente en desacuerdo y 5 completamente de acuerdo. Cada respuesta obtiene una puntuación de 0 a 4, con un rango de puntuación para el test de 0-40. La puntuación total obtenida se multiplica por 2.5 para obtener los valores de SUS estándar., clasificándose en tres grupos acorde el grado de aceptación:

- "Aceptable" valores por encima de 70.
- "Marginal" los de entre 50 y 70.
- "No aceptable" inferiores de 50

Se determinó el nivel de acuerdo en la detección de actividad de la enfermedad entre los dos evaluadores.

Se consideró el *gold standard* la valoración inicial de las neurorradiólogas. Se evaluó la existencia de correlación entre el total de lesiones detectadas por el *gold standard* y las erróneamente identificadas por los evaluadores con las herramientas de IA.

Se determinó si había diferencias significativas en la detección de lesiones de cada herramienta en función del equipo de RM utilizado en la obtención de los estudios.

Se consideró un nivel de significancia de  $p < 0.05$  para todas las pruebas estadísticas realizadas en el estudio.

## **MATERIAL Y MÉTODOS: 3º PASO**

### Interpretación conjunta:

Una neurorradióloga de 17 años de experiencia valoró los hallazgos de las dos herramientas y los comparó con los resultados iniciales obtenidos únicamente con la sustracción de imágenes con las que se realiza habitualmente la evaluación y el informe radiológico en el seguimiento de estos pacientes. Se tuvieron en cuenta los siguientes aspectos:

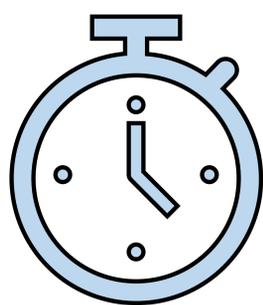
- Lesiones identificadas que coincidían.
- Lesiones identificadas por las herramientas, no detectadas por las neurorradiólogas en el informe inicial.
- Lesiones identificadas por las neurorradiólogas en el informe inicial, no detectadas por las herramientas.

## RESULTADOS

Se evaluaron un total de 24 pacientes:



Se registró el tiempo en segundos que tardó cada herramienta por separado en procesar los estudios y emitir un resultado, siendo la media de tiempo consumido por la **herramienta A** inferior (45,4 segundos) a la de la **herramienta B** (202,5 segundos).



**Herramienta A: 45,5'**



**Herramienta B: 202,5'**

## Resultados:

Mediante la escala SUS versión estándar se determinó la usabilidad percibida por los evaluadores para la **herramienta A** obteniendo un valor medio de **72.5** lo que indica un **“acceptable” (> 70)** grado de aceptación (**Tabla .2**).

System Usability Scale versión estándar		1	2	3	4	5
1.	Creo que me gustaría utilizar este sistema con frecuencia				X	
2.	Encontré el sistema innecesariamente complejo			X		
3.	Pensé que el sistema era fácil de usar					X
4.	Creo que necesitaría el apoyo de un técnico para poder utilizar este sistema		X			
5.	Me pareció que las diversas funciones en este sistema fueron bien integradas				X	
6.	Pensé que habría demasiada inconsistencia en este sistema		X			
7.	Me imagino que la mayoría de la gente aprende a utilizar este sistema rápidamente		X			
8.	El sistema me pareció muy complicado de usar	X				
9.	Me sentí muy seguro al utilizar el sistema					X
10.	Tuve que aprender un montón de cosas antes de poder utilizar este sistema			X		
<b>Valor total</b>		<b>72.5</b>				

**Tabla .2:** Escala SUS para la **herramienta A**, y valor final.

## Resultados:

El valor medio obtenido para la **herramienta B** fue de **85**, siendo un resultado “**acceptable**” y superior al grado de aceptación calculado para la **herramienta A** (Tabla .3).

System Usability Scale versión estándar		1	2	3	4	5
1.	Creo que me gustaría utilizar este sistema con frecuencia					X
2.	Encontré el sistema innecesariamente complejo		X			
3.	Pensé que el sistema era fácil de usar					X
4.	Creo que necesitaría el apoyo de un técnico para poder utilizar este sistema		X			
5.	Me pareció que las diversas funciones en este sistema fueron bien integradas				X	
6.	Pensé que habría demasiada inconsistencia en este sistema		X			
7.	Me imagino que la mayoría de la gente aprende a utilizar este sistema rápidamente					X
8.	El sistema me pareció muy complicado de usar		X			
9.	Me sentí muy seguro al utilizar el sistema					X
10.	Tuve que aprender un montón de cosas antes de poder utilizar este sistema		X			
<b>Valor total</b>		<b>85</b>				

**Tabla .3:** Escala SUS para la **herramienta B**, y valor final.

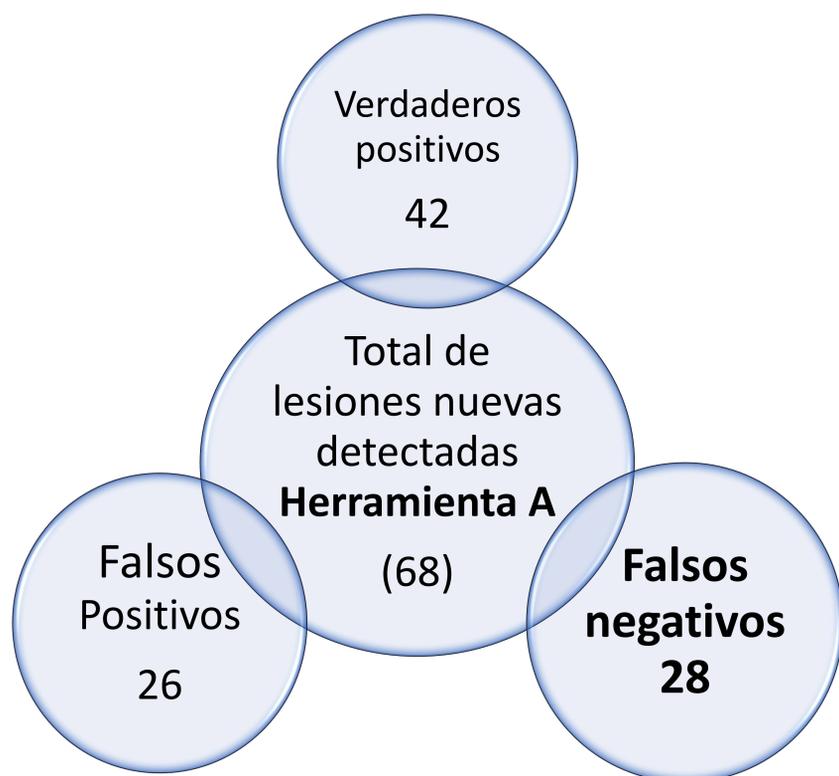
## RESULTADOS

El primer análisis visual identificó un total de 70 lesiones nuevas. La **herramienta A** detectó un total de 68 lesiones (**Esquema 1**):

- 42 coincidían con el análisis visual inicial de los radiólogo.
- 28 fueron no identificadas (falsos negativos).
- 26 lesiones no identificadas por las neurorradiólogas (falsos positivos).

La **herramienta B** detectó un total de 79 lesiones nuevas (**Esquema 2**):

- 51 coincidían con el análisis visual inicial de los radiólogos
- 19 fueron no identificadas (falsos negativos).
- 26 lesiones no identificadas por las neurorradiólogas (falsos positivos).



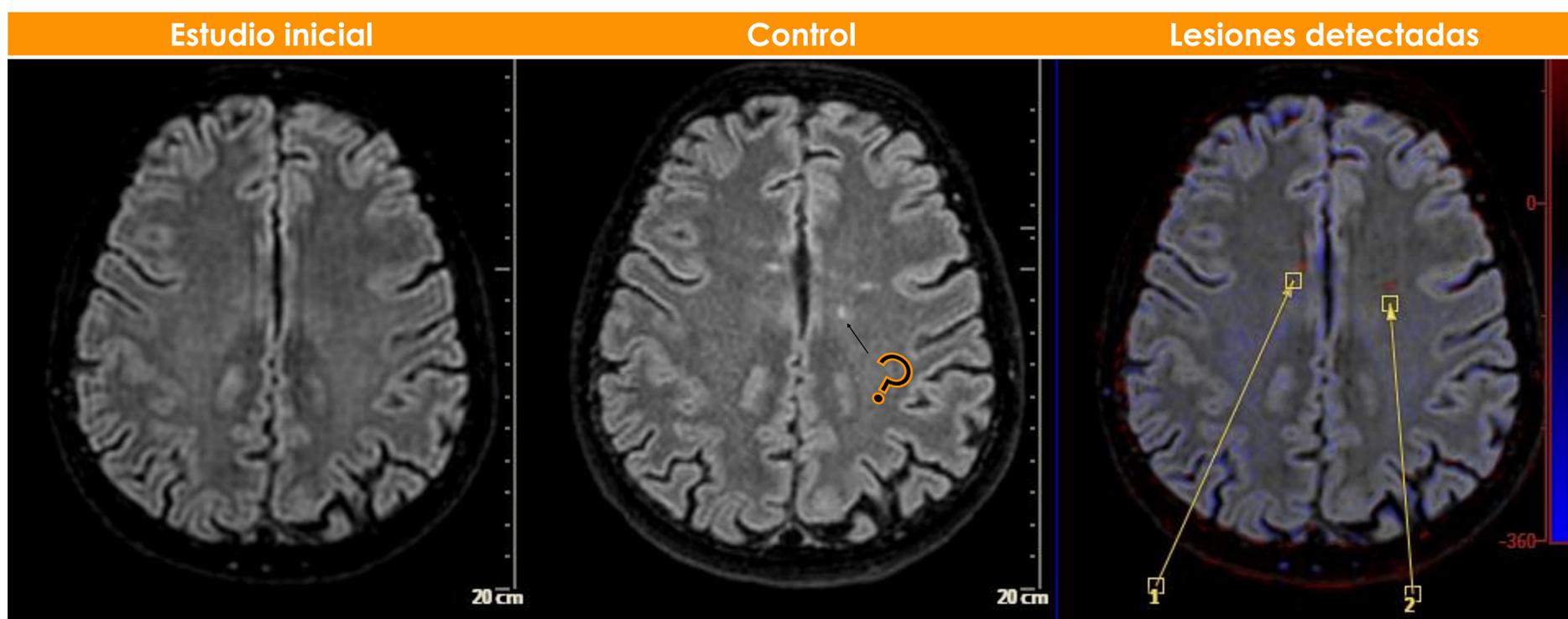
**Esquema 1**



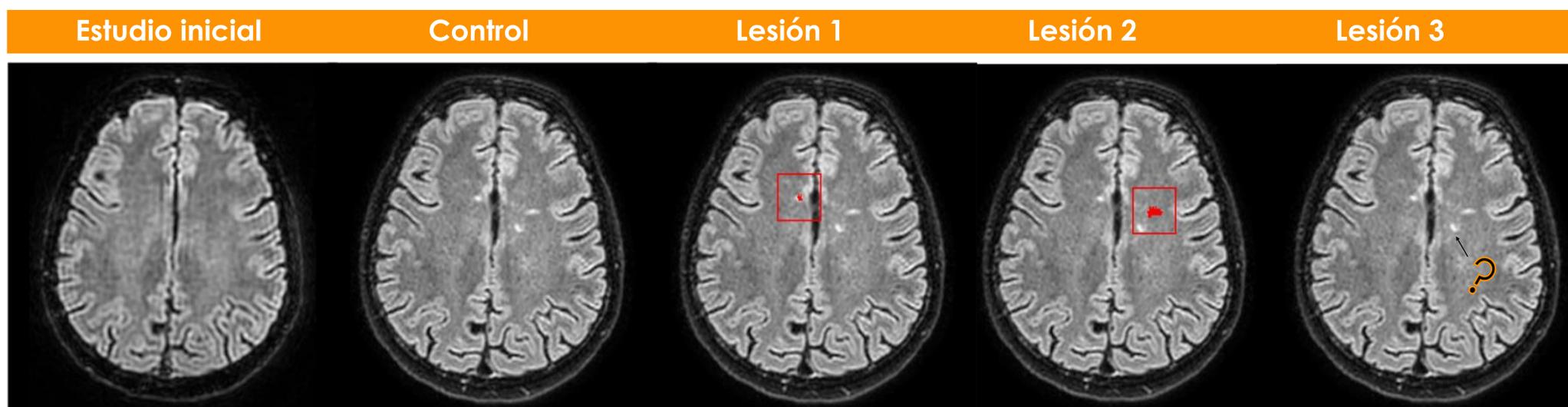
**Esquema 2**

## RESULTADOS: 2º PASO

No se detectaron un total de 47 lesiones siendo la **herramienta A** la de mayor cantidad de falsos negativos en ambos evaluadores con total de 28, por 19 de la **herramienta B** (ejemplos de falsos negativos **Imágenes 4 y 5**).



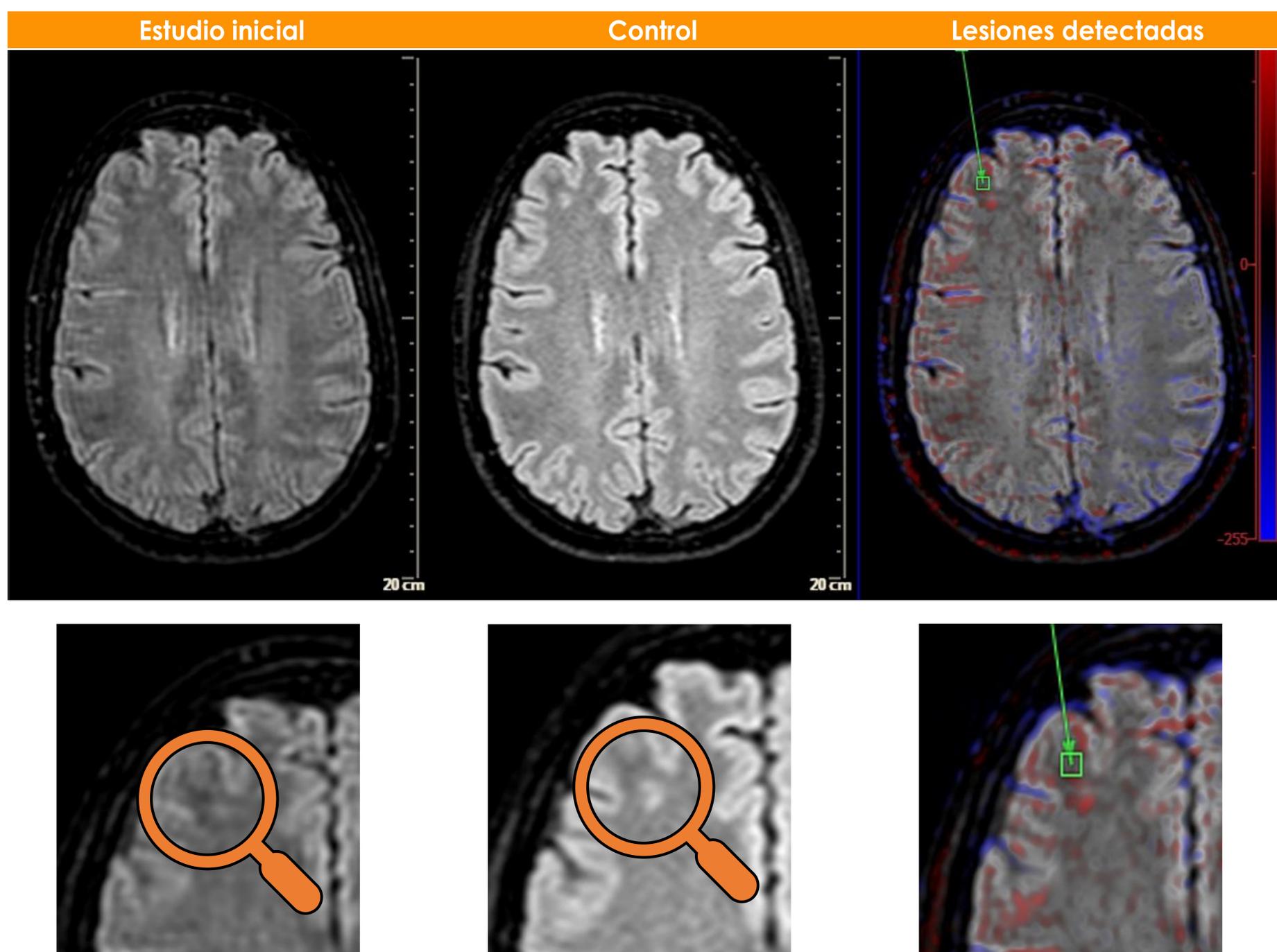
**Imagen 4:** RM Cerebral (secuencia FLAIR-3D) procesamiento de las imágenes por la **herramienta A** donde detecta dos lesiones desmielinizantes (señaladas en rojo), no detectando la tercera lesión más posterior (falso negativo).



**Imagen 5:** Corte axial de RM Cerebral (secuencia FLAIR-3D, igual corte que el mostrado en la **imagen 4**) procesado por la **herramienta B** donde detecta las lesiones desmielinizantes 1 y 2, no identificando la lesión 3 (falso negativo).

## RESULTADOS: 2º PASO

Se detectaron un total de 52 lesiones no identificadas en la evaluación inicial de los radiólogos, con distribución similar para ambas herramientas de IA. (Imágenes 6, 7, 8 y 9).

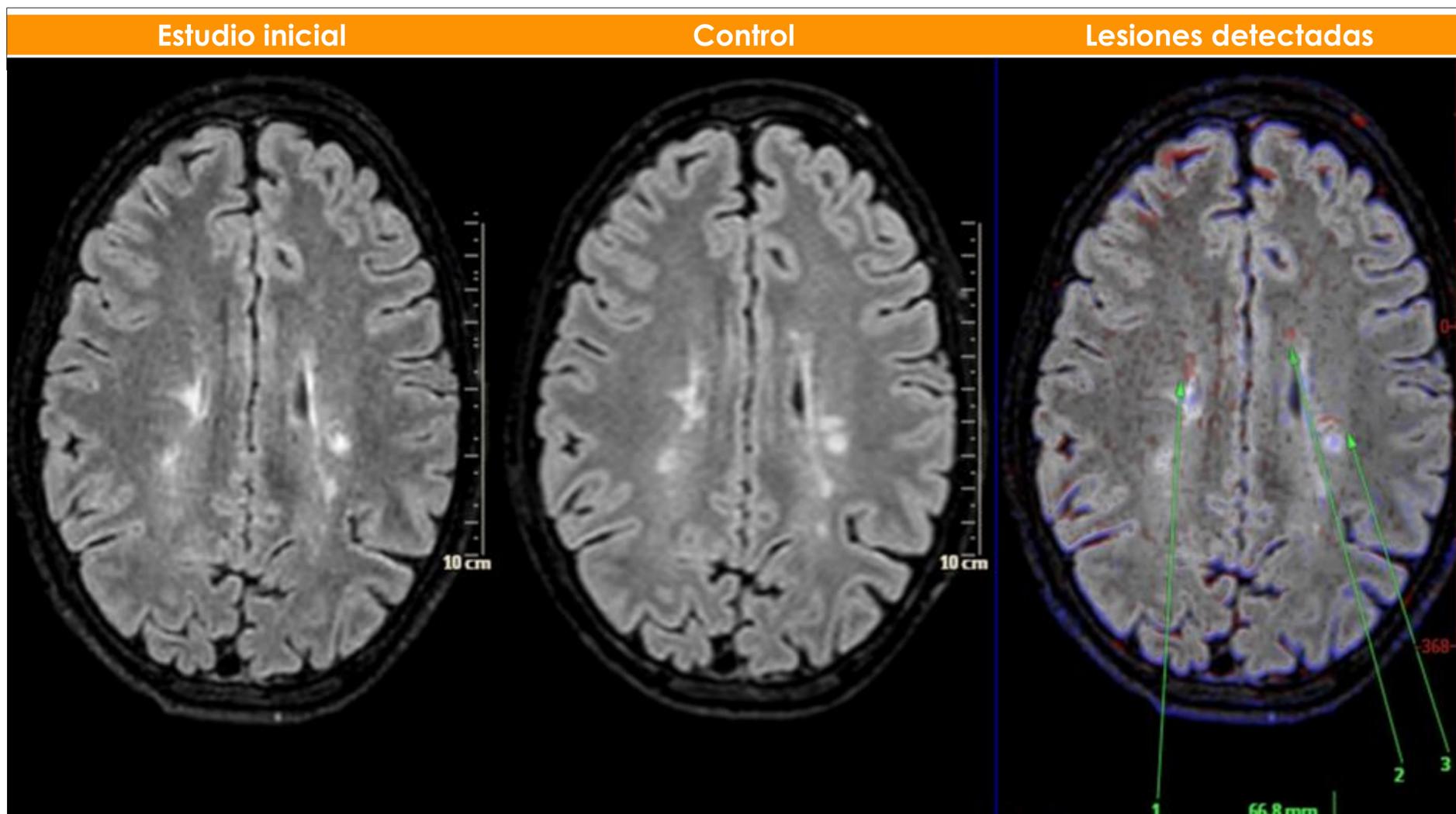


**Imagen .6:** RM Cerebral (secuencia FLAIR-3D), procesamiento de las imágenes por a **herramienta A** donde detecta 1 lesión desmielinizantes subcortical(señalada con la flecha) como nueva ya presente en el estudio inicial.

## RESULTADOS: 2º PASO

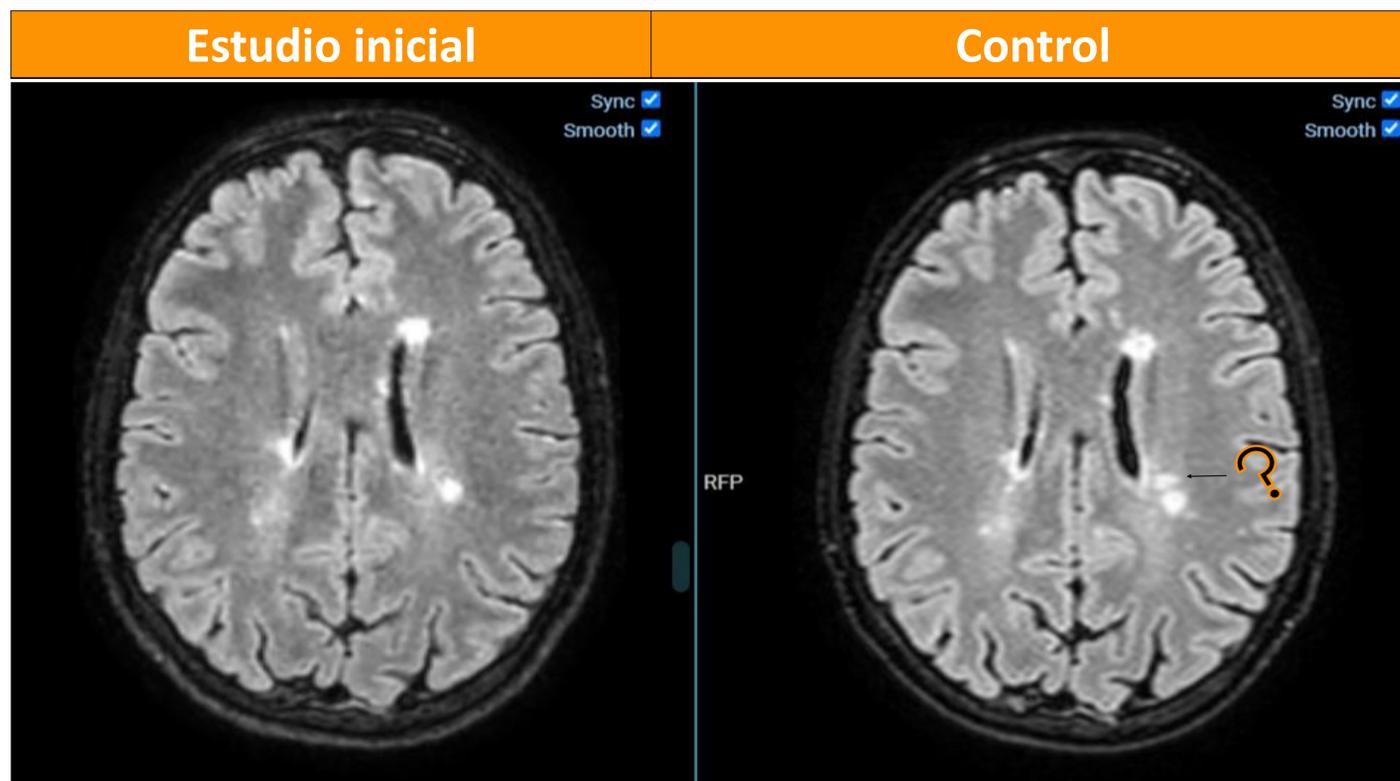


**Imagen 7:** Corte axial de RM Cerebral (secuencia FLAIR-3D) procesado por la **herramienta B** de inteligencia artificial donde detecta una lesión inexistente (falso positivo).

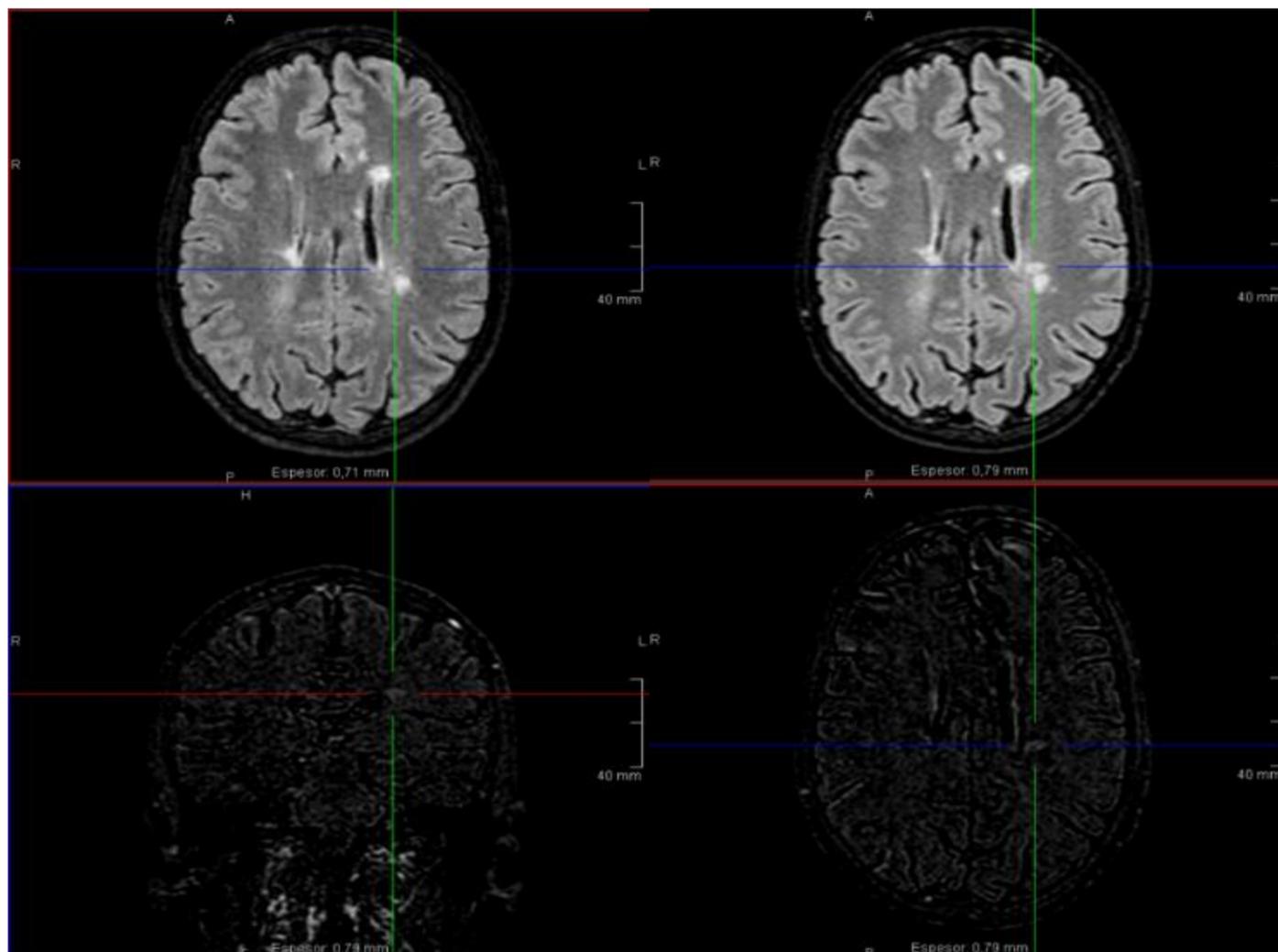


**Imagen 8:** Corte axial de RM Cerebral (secuencia FLAIR-3D) procesado por la **herramienta A** se identifican 3 lesiones. La lesión 1 resultó ser un falso positivo y las lesiones 2 y 3 correctamente identificadas.

## RESULTADOS: 2º PASO



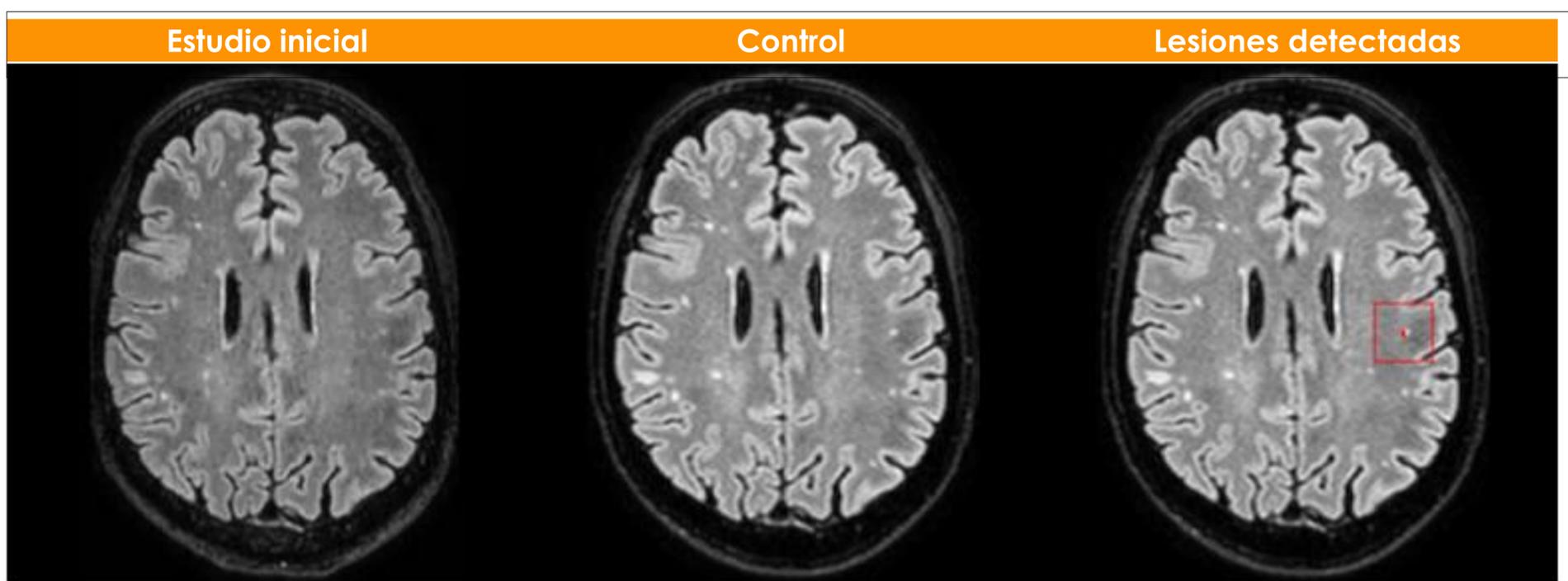
**Imagen 8:** Corte axial de RM Cerebral (secuencia FLAIR-3D) del paciente de la **Imagen 7** (anterior) procesado por la **herramienta B** donde no se identifica la lesión 3. (falso negativo).



**Imagen .9:** Corte axial de RM Cerebral (secuencia FLAIR-3D) del paciente de la **Imagen 7 y 8**. Muestra la lesión 3 identificada por las neurorradiólogas empleando la herramienta Voxar 3D® subtraction, Impax.

## RESULTADOS: 2º PASO

La **herramienta B** demostró una mayor sensibilidad en la detección de actividad de la enfermedad en comparación con la **herramienta A**, con valores de sensibilidad del 94,4% y 83.3%, respectivamente. Además, exhibió un porcentaje de aciertos superior, registrando un 92% frente al 81% de la **herramienta A**. (Imágenes .10, .11, )



**Imagen .10** : Corte axial de RM Cerebral (secuencia FLAIR-3D) procesado por la **herramienta B** se muestra lesión previa que ha aumentado de tamaño respecto al control anterior realizado en el año 2022. Llama la atención que a pesar de que la herramienta B esta diseñada limitada por diseño para detectar lesiones nuevas logra identificar esta placa.



**Imagen .11**: Corte axial de RM Cerebral (secuencia FLAIR-3D) procesado por la **herramienta B** se muestra una nueva lesión no identificada por el resto de métodos evaluados.

## RESULTADOS: 2º paso

Hubo mejor grado de acuerdo interevaluador para la **herramienta B** que con la **A**, con valores de **k** de **0,746** y **0,486** respectivamente (**Tablas 4 y 5**) con valores de p menores al 0.05. Este mayor grado de acuerdo interevaluador para la herramienta B sugiere una mayor confiabilidad en la detección de enfermedad activa lo que tiene implicaciones directas en el manejo de los pacientes con EM.

**Tabla .2 Nivel de acuerdo interevaluador herramienta A**

			Evaluador 2		Total
			E. estable	E. activa	
Evaluador 1	E. estable	% del total	45,8%	<b>12,5%</b>	58,3%
	E. activa	% del total	<b>12,5%</b>	29,2%	41,7%
Total		% del total	58,3%	41,7%	100,0%

**Kappa = 0,486 p=0,017**

**Tabla .4:** El grado de acuerdo interobservador calculado para la **herramienta A** fue moderado con un **25% de discrepancias** en los resultados obtenidos.

**Tabla .3 Nivel de acuerdo interevaluador herramienta B**

			Evaluador 2		Total
			E. estable	E. activa	
Evaluador 1	E. estable	% del total	50,0%	<b>4,2%</b>	54,2%
	E. activa	% del total	<b>8,3%</b>	37,5%	45,8%
Total		% del total	58,3%	41,7%	100,0%

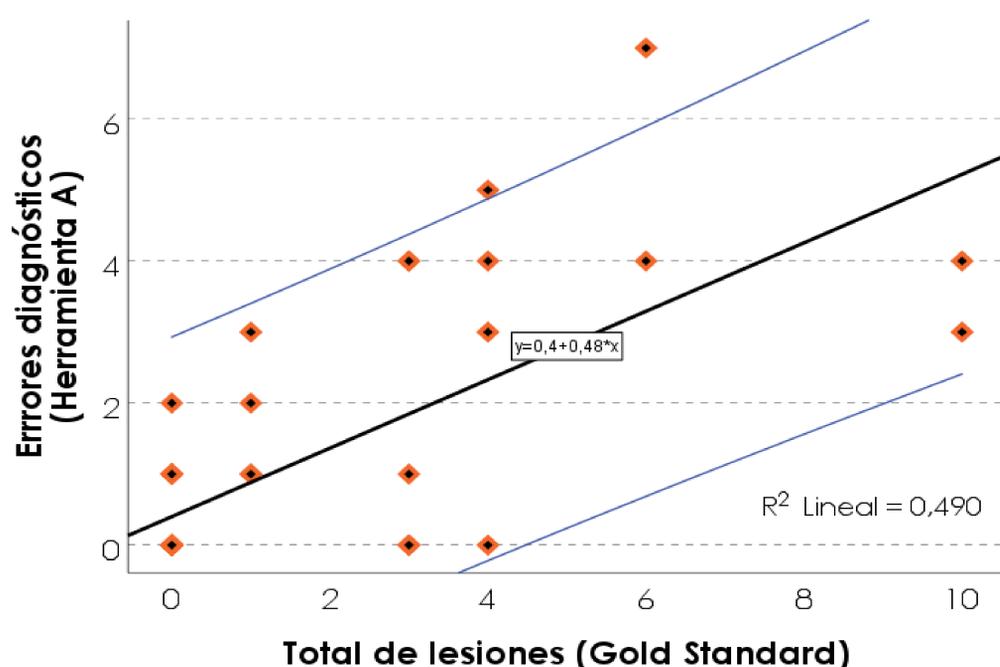
**Kappa = 0,746 p=0,001**

**Tabla .5:** El Kappa de Cohen mostró un grado de acuerdo considerable (**k=0.746**) entre los resultados obtenidos del Evaluador 1 y el Evaluador 2, empleando la **herramienta B**, observándose **discrepancias** en la detección de enfermedad activa en solo el **12,5%** de la muestra analizada.

## RESULTADOS: 2º PASO

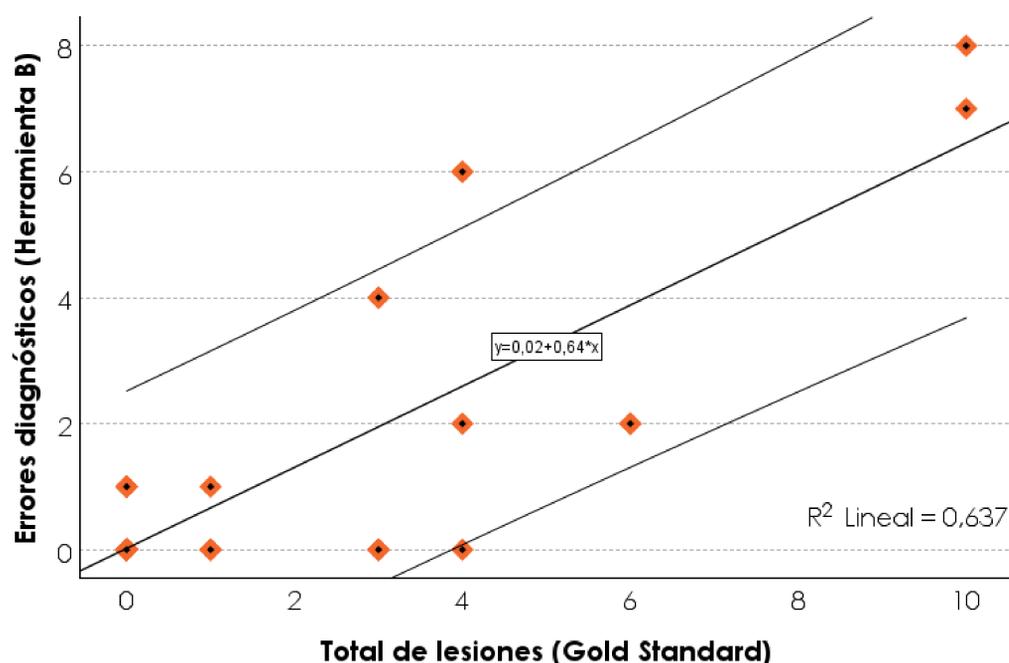
El coeficiente de correlación de Spearman (**Rho**) fue de **0,676** para la **herramienta A** y de **0,613** para **B**. Ambos resultados fueron estadísticamente significativos con un nivel de significancia bilateral de  $< 0,01$ . La **herramienta A** mostró una correlación ligeramente mayor que la **herramienta B**, según el coeficiente de correlación de Spearman. (Gráficos .1 y.2)

**Gráfico .1:** Correlación entre total de lesiones detectadas por los radiólogos y el número de errores diagnósticos de la **herramienta A**.



El gráfico muestra una correlación positiva entre las variables estudiadas con un coeficiente de determinación  $R^2$  lineal de 0.49.

**Gráfico .2:** Correlación entre total de lesiones detectadas por los radiólogos y el número de errores diagnósticos de la **herramienta B**.



Se observa una correlación positiva entre las variables estudiadas con un coeficiente de determinación  $R^2$  lineal de 0.637.

## RESULTADOS: 2º PASO

Para ambas herramientas (**A** y **B**), no hubo diferencias significativas en el número de lesiones no detectadas (falsos negativos) o lesiones detectadas no encontradas en el análisis visual, independientemente del equipo de RM utilizado. (**Tablas 6 y 7**).

**Tabla .6: Comparación del número de errores diagnósticos entre los grupos, según el equipo de resonancia magnética para la herramienta A.**

	Equipo de RM	N	Rango promedio	Suma de rangos
Errores Diagnósticos	Igual	18	23,42	421,50
	Diferente	30	25,15	754,50
	Total	48		

U de Mann-Whitney= 250,50      **Sig. Asin. (bilateral)= 0,642**

**Tabla .7: Comparación del número de errores diagnósticos entre los grupos, según el equipo de resonancia magnética para la herramienta B**

	Equipo de RM	N	Rango promedio	Suma de rangos
Errores Diagnósticos	Igual	18	23,58	424,50
	Diferente	30	25,05	751,50
	Total	48		

U de Mann-Whitney= 253,50      **Sig. Asin. (bilateral)= 0,661**

## RESULTADOS: 3º PASO

Las **Tablas 8 y 9** muestra la evaluación realizada por las neurorradiólogas comparando los hallazgos del análisis visual inicial con las herramientas de IA.

La **herramienta A** fue inferior en la detección de lesiones con un total de 23 lesiones, en comparación con las descritas en el informe radiológico (**Tabla 8**).

En la **Tabla 9** se muestran los resultados comparativos en la detección de lesiones entre la **herramienta B** y las neurorradiólogas observándose un mejor desempeño de la herramienta B en la detección de lesiones, pero aún con mayor número de lesiones no detectadas (**10**).

En general la valoración realizada por las neurorradiólogas fue superior a la detección de lesiones empleando herramientas de IA, sin embargo presentaron dificultades identificando nuevas lesiones en aquellos paciente con mayor número de lesiones preexistentes y mayor número de lesiones nuevas.

**Tabla 8: Comparación del número lesiones detectadas por neurorradiólogos y la herramienta A**

Nº	Lesiones nuevas identificadas		Segunda Evaluación			
	Informe Radiológico (Evaluación inicial)	Herramienta A	Coincidencia	No detecta Radiólogo	No detecta IA	Total de lesiones
1	3	3	●			3
2	3	2	●			3
3	1	0	●			1
4	3	1	●			3
5	0	0	●			0
6	1	0	●			1
7	0	0	●			0
8	0	0	●			0
9	0	0	●			0
10	4	0	●			6
11	0	0	●			0
12	0	0	●			0
13	6	3	●			7
14	0	0	●			0
15	0	0	●			0
16	0	0	●			0
17	0	0	●			0
18	0	0	●			0
19	0	0	●			0
20	10	7	●			14
21	0	0	●			0
22	0	0	●			0
23	0	0	●			0
24	4	3	●			4

**Leyenda:**

- Coincide detección de lesiones entre IA y radiólogo.
- No coincide detección de lesiones entre IA y radiólogo.
- | Lesión no detectada

**Tabla 9: Comparación del número lesiones detectadas por neurorradiólogos y la herramienta B**

Nº	Lesiones nuevas identificadas		Segunda Evaluación			
	Informe Radiológico (Evaluación inicial)	Herramienta B	Coinciden	No detecta Radiólogo	No detecta IA	Total de lesiones
1	3	3	●			3
2	3	3	●			3
3	1	1	●			1
4	3	1	●			3
5	0	0	●			0
6	1	1	●			1
7	0	0	●			0
8	0	0	●			0
9	0	0	●			0
10	4	3	●			6
11	0	0	●			0
12	0	0	●			0
13	6	4	●			7
14	0	0	●			0
15	0	0	●			0
16	0	0	●			0
17	0	0	●			0
18	0	0	●			0
19	0	0	●			0
20	10	13	●			14
21	0	0	●			0
22	0	0	●			0
23	0	0	●			0
24	4	3	●			4

**Leyenda:**

- **Coincide detección de lesiones entre IA y radiólogo.**
- **No coincide detección de lesiones entre IA y radiólogo.**
- | **Lesión no detectada (falso negativo)**

## DISCUSIÓN

El presente estudio retrospectivo evaluó la eficacia de dos herramientas en la detección de nuevas lesiones de esclerosis múltiple en imágenes de resonancia magnética cerebral.

-La **herramienta A** fue más rápida que la **B** en el procesamiento de los datos suministrados con una media de tiempo 4 veces inferior.

- Sin embargo en cuanto a términos de usabilidad valorados acorde a la escala SUS versión estándar<sup>(1)</sup>, la **herramienta B** fue superior, mostrando una interfaz intuitiva, fácil de operar y simplificando la elaboración del informe radiológico a la hora de identificar las lesiones de nueva aparición.

- Ambas mostraron valores adecuados de sensibilidad y de porcentaje de acierto, coincidiendo con estudios previos que informaron valores de precisión entre el 83,7% y el 90,0%.<sup>(2,3)</sup>

- El grado de acuerdo interobservador mostró un mejor grado de acuerdo para la **herramienta B** en comparación con la **A**, lo que sugiere una mayor consistencia y confiabilidad en la detección de enfermedad activa.

Una posible explicación de esta diferencia podría residir en las técnicas de desarrollo de cada algoritmo, siendo particularmente la utilización de técnicas de Deep Learning en la IA **B** lo que permitió un mejor rendimiento de la herramienta en el ámbito clínico<sup>(4,5)</sup>. El grado de acuerdo interobservador de la **herramienta B** (**k=0.746**) fue superior al observado en estudios previos donde evaluaban el grado de acuerdo entre neurorradiólogos<sup>(6)</sup>.

## DISCUSIÓN

- No demostramos diferencias significativas en el número de errores diagnósticos entre las herramientas de IA según el equipo de RM utilizado, lo que sugiere un desempeño consistente independientemente de las variaciones en las condiciones de adquisición de imágenes.

- Tras realizar la segunda evaluación se observó que las neurorradiólogas presentaron dificultades en la identificación de nuevas lesiones en aquellos pacientes con mayor carga lesional, detectando menos lesiones que las herramientas. Este hallazgo puede ser debido a que la comparación convencional lado a lado es agotadora, propensa a errores y con alta variabilidad interobservador, dependiendo directamente del nivel de entrenamiento entre los evaluadores <sup>(8)</sup>, y es interesante comentar que el sesgo de satisfacción al encontrar varias lesiones nuevas probablemente sea el responsable de que cuando el número de lesiones nuevas es alto, los radiólogos detectaron menos lesiones que las herramientas

## LIMITACIONES DEL ESTUDIO

- **El tamaño muestral:** el estudio incluyó una muestra pequeña de 24 pacientes lo que limita la generalización de los resultados.
- **La herramienta de IA:** para hacer este estudio la **herramienta B** de IA se utilizó la versión on-line, lo cual probablemente haga que los tiempos de evaluación de la IA de las imágenes sean superiores, por lo que no sería comparable con la **herramienta A**.
- **La experiencia de los observadores:** se eligieron dos evaluadores que no tenían ninguna experiencia en EM en RM para valorar la facilidad de uso de las herramientas por parte de personal no entrenado para ver si eran capaces de manejarlas de forma sencilla; sin embargo, esto podría afectar a los resultados aquí mostrados, como por ejemplo el tiempo empleado en la valoración de las lesiones con cada una de las herramientas de IA
- **Especificación del tipo de lesiones:** No se especificó el tipo de lesiones que se identificaron. Un análisis más detallado del tipo de lesiones ayudaría a comprender mejor la relación entre las lesiones erróneamente diagnosticadas y el resto de variables.

## CONCLUSIONES

Las herramientas basadas en IA para el seguimiento mediante RM de pacientes con EM demostraron ser de gran ayuda al proporcionar un análisis rápido y detallado de las imágenes, fáciles de utilizar y de interpretar, consistentes y con un gran potencial como complemento en el proceso de evaluar las imágenes, mejorando la eficiencia en el diagnóstico, ahorrando tiempo y mejorando la detección de lesiones. Sin embargo, la interpretación precisa de estas imágenes todavía necesita la intervención de radiólogos.

## Referencias Bibliográficas:

- 1- James R. Lewis (2018): The System Usability Scale: Past, Present, and Future, *International Journal of Human-Computer Interaction*, DOI: 10.1080/10447318.2018.1455307
- 2- Filippi M, Preziosa P, Arnold DL, Barkhof F, Harrison DM, Maggi P, Mainero C, Montalban X, Sechi E, Weinshenker BG, Rocca MA. Present and future of the diagnostic work-up of multiple sclerosis: the imaging perspective. *J Neurol*. 2023 Mar;270(3):1286-1299. doi: 10.1007/s00415-022-11488-y. Epub 2022 Nov 24. PMID: 36427168; PMCID: PMC9971159.
- 3- Rovira A, Corral JF, Auger C, Valverde S, Vidal-Jordana A, Oliver A, de Barros A, Ng Wong YK, Tintoré M, Pareto D, Aymerich FX, Montalban X, Lladó X, Alonso J. Assessment of automatic decision-support systems for detecting active T2 lesions in multiple sclerosis patients. *Mult Scler*. 2022 Jul;28(8):1209-1218. doi: 10.1177/13524585211061339. Epub 2021 Dec 3. PMID: 34859704.
- 4- Nabizadeh F, Masroui S, Ramezannezhad E, Ghaderi A, Sharafi AM, Soraneh S, Naser Moghadasi A. Artificial intelligence in the diagnosis of multiple sclerosis: A systematic review. *Mult Scler Relat Disord*. 2022 Mar;59:103673. doi: 10.1016/j.msard.2022.103673. Epub 2022 Feb 7. PMID: 35180619.
- 5- Moazami F, Lefevre-Utile A, Papaloukas C, Soumelis V. Machine Learning Approaches in Study of Multiple Sclerosis Disease Through Magnetic Resonance Images. *Front Immunol*. 2021 Aug 11;12:700582. doi: 10.3389/fimmu.2021.700582. PMID: 34456913; PMCID: PMC8385534.
- 6- Rocca MA, Anzalone N, Storelli L, Del Poggio A, Cacciaguerra L, Manfredi AA, Meani A, Filippi M. Deep Learning on Conventional Magnetic Resonance Imaging Improves the Diagnosis of Multiple Sclerosis Mimics. *Invest Radiol*. 2021 Apr 1;56(4):252-260. doi: 10.1097/RLI.0000000000000735. PMID: 33109920.
- 7-Kaka H, Zhang E, Khan N. Artificial Intelligence and Deep Learning in Neuroradiology: Exploring the New Frontier. *Can Assoc Radiol J*. 2021 Feb;72(1):35-44. doi: 10.1177/0846537120954293. Epub 2020 Sep 18. PMID: 32946272.
- 8-Hagens MH, Burggraaff J, Kilsdonk ID, Ruggieri S, Collorone S, Cortese R, Cawley N, Sbardella E, Andelova M, Amann M, Lieb JM, Pantano P, Lissenberg-Witte BI, Killestein J, Oreja-Guevara C, Wuerfel J, Ciccarelli O, Gasperini C, Lukas C, Rovira A, Barkhof F, Wattjes MP; MAGNIMS Study Group. Impact of 3 Tesla MRI on interobserver agreement in clinically isolated syndrome: A MAGNIMS multicentre study. *Mult Scler*. 2019 Mar;25(3):352-360. doi: 10.1177/1352458517751647. Epub 2018 Jan 12. PMID: 29327668; PMCID: PMC6393953.
- 9- Galletto Pregliasco A, Collin A, Guéguen A, Metten MA, Aboab J, Deschamps R, Gout O, Duron L, Sadik JC, Savatovsky J, Lecler A. Detección mejorada de nuevas lesiones de MS durante el seguimiento Usando un método automatizado de registro de MR-Fusion. *AJNR Am J Neuroradiol*. 2018 Jul;39(7):1226-1232. doi: 10.3174/ajnr.A5690. Epub 2018 Jun 7. PMID: 29880479; PMCID: PMC7655423.